



(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)



1-0028088

(51)⁷ C12N 1/21; C12N 15/70; C12P 7/40;
C12N 15/63

(13) B

(21) 1-2014-02592

(22) 29/01/2013

(86) PCT/US2013/023690 29/01/2013

(87) WO/2013/116244 08/08/2013

(30) 61/632,777 30/01/2012 US

(45) 25/04/2021 397

(43) 26/01/2015 322A

(73) PTT Global Chemical Public Company Limited (TH)

555/1, Energy Complex, Building A, 14th-18th Floor, Vibhavadi Rangsit Road,
Chatuchak, Chatuchak, Bangkok 10900 Thailand

(72) YOCUM, R., Rogers (US); GONG, Wei (CN); DOLE, Sudhanshu (IN); SILLERS,
Ryan (US); GANDHI, Meghal (IN); PERO, Janice, G. (US).

(74) Công ty Luật TNHH T&G (TGVN)

(54) VI SINH VẬT ĐƯỢC THIẾT KẾ DI TRUYỀN ĐỂ TỔNG HỢP RA AXIT
MUCONIC

(57) Sáng chế đề cập đến vi sinh vật được thiết kế di truyền để tổng hợp ra axit muconic và quy trình tổng hợp nguyên liệu hóa học có thể thu hồi được bằng cách sử dụng chất xúc tác sinh học đã được thiết kế di truyền làm gia tăng khả năng chuyển hóa các nguồn cacbon có thể thu hồi được thành hợp chất hữu ích. Cụ thể hơn, sáng chế đề xuất quy trình tổng hợp axit muconic từ các nguồn cacbon có thể thu hồi được bằng cách sử dụng vi sinh vật được thiết kế di truyền.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến vi sinh vật được thiết kế di truyền để tổng hợp ra axit muconic và quy trình tổng hợp nguyên liệu hóa học có thể thu hồi được bằng cách sử dụng chất xúc tác sinh học đã được thiết kế di truyền làm gia tăng khả năng chuyển hóa các nguồn cacbon có thể thu hồi được thành hợp chất hữu ích. Cụ thể, sáng chế đề xuất quy trình tổng hợp dạng đồng phân của axit muconic từ các nguồn cacbon có thể thu hồi được bằng cách sử dụng chất xúc tác sinh học đã được thiết kế di truyền.

Tình trạng kỹ thuật của sáng chế

Axit adipic là hóa chất thể tích lớn được sử dụng trong tổng hợp Nylon 66. Axit adipic hiện nay được điều chế từ hóa chất dầu mỏ, nhưng việc tổng hợp này không thân thiện với môi trường (Niu et al., 2002). Theo một cách khác, axit adipic có thể được tạo thành từ đồng phân bất kỳ trong số ba dạng đồng phân của axit muconic (cis, cis; cis, trans; trans, trans) bằng cách hydro hóa hóa học. Mong muốn là điều chế axit muconic từ các nguồn có thể thu hồi được bằng cách lên men với vi sinh vật, sau đó hydro hóa thành axit adipic, vì con đường điều chế axit adipic này thân thiện với môi trường hơn con đường điều chế từ hóa chất dầu mỏ truyền thống.

Những nỗ lực nhằm tổng hợp axit muconic bằng vi sinh vật có thể được chia thành ba nhóm sau: (1) Con đường phân hủy thơm để tổng hợp axit muconic, trong đó các hợp chất thơm khác nhau được đưa vào và một phần vòng benzen của các hợp chất thơm bị phân cắt bằng cách oxy hóa; (2) con đường tổng hợp muconat, trong đó khung phân tử axit muconic được tạo ra từ các hợp chất có 2, 3, 4 nguyên tử cacbon hoặc lysin; và (3) con đường axit muconic sinh tổng hợp từ axit amin thơm, trong đó axit muconic được tạo ra từ 3-dehydroshikimat, một sản phẩm trung gian trong con đường sinh tổng hợp axit amin thơm ở nhiều sinh vật.

Nhiều vi sinh vật có khả năng phân hủy các hợp chất thơm chứa vòng benzen, như phenol, catechol, và axit benzoic, bằng cách sử dụng con đường phân cắt nhân

như phenol, catechol, và axit benzoic, bằng cách sử dụng con đường phân cắt nhân thơm để thu được hợp chất cuối hoặc trung gian là các hợp chất không thơm như axit cis, cis-muconic, or axit 3-carboxy-cis, cis-muconic (Niu et al., 2002; Perez-Pantoja et al., 2008). Trước đây, một số nhóm đã nỗ lực sử dụng khả năng này của vi sinh vật để tổng hợp axit cis, cis-muconic trên quy mô công nghiệp (Mizuno et al, 1988; Yoshikawa et al, 1990; Choi et al, 1997). Cuối những năm 80, Celgen Corporation of USA ADN Mitsubishi Chemical Industries của Nhật Bản đã phát triển quy trình tổng hợp axit muconic lần lượt từ toluen và axit benzoic, được chứng minh bằng một số patent Hoa Kỳ và Nhật Bản đã được cấp bằng trong lĩnh vực này.

Một số vi sinh vật được báo cáo là tổng hợp axit cis, cis-muconic sử dụng toluen, axit benzoic, benzen hoặc catechol. Ví dụ, với catechol làm nguồn cacbon, quá trình tổng hợp axit cis, cis-muconic có thể đạt được hiệu suất chuyển hóa gần như là 100%mol bằng cách sử dụng các tế bào *E. coli* tái tổ hợp biểu hiện gen *catA*, mã hóa *Pseudomonas putida* mt-2 catechol 1,2-dioxyaza chịu trách nhiệm xúc tác sự phân cắt ở vị trí ortho của catechol, làm chất xúc tác sinh học (Kaneko et al, 2011). Các thiết bị phản ứng sinh học để tổng hợp liên tục axit muconic sử dụng hệ này đã được mô tả.

Phương pháp tổng hợp axit cis, cis-muconic bằng vi sinh vật sử dụng hợp chất cacbon vòng có 6 nguyên tử cacbon chưa bao giờ trở thành thực tế trong công nghiệp; tuy nhiên, vẫn tiếp tục nghiên cứu để tìm hiểu về chức năng của enzym này trong con đường phân hủy để tổng hợp axit muconic ở vi sinh vật.

Gần đây, công bố đơn quốc tế (WO 2011/017560) mô tả chất xúc tác sinh học có con đường muconat và phương pháp tổng hợp axit muconic bằng cách sử dụng chất xúc tác sinh học này. Một cách vắn tắt, công bố đơn quốc tế này đã mô tả bốn con đường khác nhau để tổng hợp axit muconic. Con đường thứ nhất để tổng hợp axit muconic bắt đầu từ *suxinyl-CoA* và *axetyl-CoA*. Con đường thứ hai để tổng hợp axit muconic bắt đầu từ *pyruvat* và *malonat semialdehyt*. Con đường thứ ba để tổng hợp axit muconic bắt đầu từ *pyruvat* và *suxinic semialdehyt*. Con đường thứ tư để tổng hợp axit muconic bắt đầu từ *lysine*. Tất cả các con đường để tổng hợp axit muconic được đề xuất trong đơn sáng chế này đều dựa trên mô hình máy tính và

chưa xem xét liệu chất xúc tác sinh học này có thể được tạo ra với năng suất và hiệu suất chấp nhận được về mặt thương mại đối với axit muconic hay không.

Con đường lên men để tạo ra axit cis, cis-muconic bằng cách sử dụng hệ *E. coli* được thiết kế di truyền đã được mô tả trong tài liệu chuyên ngành (Niu et al., 2002) và tài liệu patent (US 5,616,496; US 5,487,987; WO 2011/085311 A1), nhưng quy trình của các giải pháp đã biết cần cải thiện đáng kể về hiệu giá, hiệu suất, và tính thích hợp để tổng hợp trên quy mô thương mại lớn để tiết kiệm về mặt kinh tế. Có hai báo cáo về nấm men *Saccharomyces cerevisiae* được thiết kế di truyền để tổng hợp axit cis, cis-muconic từ glucoza, nhưng hiệu giá công bố chỉ là 1,5mg/l và 140mg/l (Weber et al., 2012; Curran et al., 2012). Ở hiệu giá này, tất cả các quy trình lên men bằng nấm men này đều không hấp dẫn để tổng hợp trên quy mô thương mại. Sáng chế đề cập đến quy trình tổng hợp axit cis, cis-muconic hoặc cis, trans-axit muconic, bằng cách lên men về cơ bản được cải thiện so với các quy trình đã công bố đã biết trong lĩnh vực này về tính thích hợp để tổng hợp trên quy mô thương mại lớn.

Một trong những mục đích của sáng chế được bộc lộ ở đây là tổng hợp axit cis, cis-muconic bằng cách lên men vi sinh vật bắt đầu từ nguồn cacbon không thơm thu hồi được, như đường hoặc các hợp chất cacbon đơn giản khác có thể thu được từ thực vật quang hợp, sử dụng sinh vật được thiết kế di truyền thích hợp để tổng hợp trên quy mô thương mại lớn.

Năm 2002, Niu et al đã công bố con đường “không có benzen” để tổng hợp axit adipic mà được sử dụng trong quy trình lên men để tổng hợp axit cis, cis-muconic, và sau đó là quy trình hóa học để chuyển hóa axit cis, cis-muconic tạo ra axit adipic. Quy trình này đã được cấp bằng, nhưng như tác giả sáng chế cho biết, quy trình này không được thương mại hóa (US 5,487,987; US 5,616,496). Phần lên men của quy trình đã công bố các chủng *E. coli* được thiết kế di truyền, chủng tốt nhất là WN1/pWN2,248. Con đường này sử dụng một phần của con đường sinh tổng hợp axit amin thơm (còn biết đến là “con đường axit shikimic”, “con đường shikimat”, “con đường chorismat”, “con đường thơm chung”, “con đường thơm trung tâm”, hoặc đơn giản là “con đường thơm”, một phần của chúng được minh

họa trên Fig.1. Theo sáng chế, bước sinh hóa xuôi dòng bất kỳ từ nguồn cacbon được cấp cho vi sinh vật, ví dụ, glucoza, và trực tiếp hoặc gián tiếp tạo ra chorismat, được xem là một phần của con đường axit shikimic, gồm, ví dụ, các bước được xúc tác bởi, Glf, Glk, Zwf, TktA, TalB, và Pps. Ngoài ra, quy trình đã công bố sử dụng ba enzym khác loại chuyển hóa 3-dehydroshikimat (sản phẩm trung gian trong con đường thơm) thành axit cis, cis-muconic qua sản phẩm sản phẩm trung gian protocatechuat và catechol. Chúng chủ đã được thiết kế là dẫn xuất của *E. coli* K-12 kiểu gen của *aroE353*, *serA::(aroB, aroZ)*, *lacZ::(tktA, aroZ)*, trong đó *aroB* mã hóa 3-dehydroquinat synthaza (dưới đây là AroB), *tktA* mã hóa transketolaza, và *aroZ* là gen khác loại từ *Klebsiella pneumoniae* mã hóa 3-dehydroshikimat dehydrataza (dưới đây là “AroZ”). Chúng được thiết kế chứa plasmit nhiều bản sao, pWN2.248, thu nhận được từ pBR322 và chứa catxet gen để biểu hiện *catA*, *catX*, *aroY*, *aroF* (kháng phản hồi), *serA*, *lacI^q*, và tính kháng ampicilin. Các gen khác loài *catA* và *catX* là từ *Acinetobacter calcoaceticus* và mã hóa catechol 1,2-dioxyaza (dưới đây là “CatA”), và protein (“CatX”) có chức năng chưa được biết có thể tăng cường hoạt tính của CatA. Trong tài liệu chuyên ngành, gen *catX* còn được gọi là “orf1” (Neidle and Ornston, 1986). Theo sáng chế, chúng tôi đề cập đến cặp gen *catA* và *catX* từ *Acinetobacter* là “catAX.” Gen khác loài *aroY* từ *Klebsiella pneumoniae* và mã hóa protocatechuat decarboxylaza (dưới đây là “AroY”).

Đơn yêu cầu cấp patent gần đây, đề cập đến các patent nêu trên, đã được công bố (WO 2011/085311 A1). Trong các đơn này, chúng giống như trên, WN1/pWN2.248, đã được sử dụng để tổng hợp axit cis, cis-muconic, sau đó, được đồng phân hóa thành axit cis, trans-muconic.

Tuy nhiên, chủng WN1/pWN2.248 không thích hợp để tổng hợp trên quy mô thương mại lớn, do đó cần có quy trình được cải thiện nhiều. Sáng chế đề xuất chất xúc tác sinh học được cải thiện cho quy trình tổng hợp axit cis, cis-muconic.

Quy trình được mô tả trong tài liệu WO 2011/085311 A1 có một số dấu hiệu khác làm cho quy trình không có khả năng thực hiện trên quy mô thương mại lớn. Đột biến *aroE353* được đưa vào chất xúc tác sinh học WN1/pWN2.248 được sử dụng trong quy trình lên men để tổng hợp axit cis, cis-muconic để chặn dòng

cacbon vào phần dưới của con đường shikimat để tối đa dòng vào con đường mong muốn tạo ra axit cis, cis-muconic. Tuy nhiên, đột biến *aroE* là đột biến “triệt tiêu” (đột biến làm cho gen bất hoạt đối với tất cả các mục đích thực hiện), có tác dụng biến chủng này thành sinh vật dinh dưỡng thụ động đối với axit amin thơm (phenylalanin, tycolophan và tryptophan) và vitamin thơm hoặc sản phẩm trung gian giống vitamin được tạo ra từ con đường shikimat (axit p-hydroxy benzoic, axit p-amino benzoic và axit 2,3-dihydroxy benzoic). Các axit amin thơm tương đối đắt và việc cần các axit amin này thêm một gánh nặng lớn vào chi phí tổng hợp axit cis, cis-muconic. Do vậy, vẫn cần quy trình không đòi hỏi các chất dinh dưỡng đắt tiền này để bổ sung vào môi trường sinh trưởng.

Một vấn đề khác nữa liên quan đến chất xúc tác sinh học sẵn có cho quy trình tổng hợp axit cis, cis-muconic đề cập đến nhu cầu duy trì plasmit nhiều bản sao để biểu hiện một số gen cần thiết (Niu et al., 2002). Plasmit nhiều bản sao thường rất không ổn định để sử dụng trong quy trình công nghiệp trên quy mô lớn. Hơn nữa, ít nhất một trong số các gen trên plasmit là được biểu hiện từ trình tự khởi đầu, P_{tac} , đòi hỏi isopropylthiogalactosit (IPTG) hoặc lactoza để gây cảm ứng, và hai hóa chất này quá đắt để quy trình có hiệu quả về mặt kinh tế. Do vậy, cần có chủng ổn định hơn có catxet biểu hiện được gắn ổn định vào nhiễm sắc thể của chủng tổng hợp và cần có mức biểu hiện cao từ trình tự khởi đầu cơ định để giảm bớt nhu cầu đối với tác nhân gây cảm ứng hóa học đối với trình tự khởi đầu này.

Bản chất kỹ thuật của sáng chế

Sáng chế đề xuất các vi sinh vật được thiết kế di truyền mà tổng hợp axit cis, cis-muconic từ nguồn cacbon không thơm. Các vi sinh vật được thiết kế di truyền không cần chứa plasmit ngoại sinh bất kỳ để tổng hợp axit muconic, mặc dù chúng có một số gen ngoại sinh hoặc các gen khác loài cần thiết để tạo ra kiểu hình mong muốn. Các gen ngoại sinh được đưa vào vi sinh vật được gắn ổn định vào nhiễm sắc thể ADN. Kết quả của việc gắn vào nhiễm sắc thể ADN của các gen ngoại sinh này loại bỏ hoàn toàn nhu cầu cần sử dụng chất kháng sinh hoặc phương pháp chọn lọc khác để duy trì các plasmit mang ADN ngoại sinh. Ngoài ra, trình tự khởi đầu mạnh không cần tác nhân gây cảm ứng hóa học được sử dụng để biểu hiện các gen cần

thiết đối với con đường bắt đầu từ nguồn cacbon, như glucoza, để tổng hợp axit cis, cis-muconic.

Theo một phương án của sáng chế, hoạt tính của yếu tố điều hòa âm tính quá trình tổng hợp sinh học axit amin thơm được thao tác về mặt di truyền. Theo một khía cạnh của sáng chế, hoạt tính của yếu tố điều hòa âm tính TyrR về cơ bản giảm bằng cách kiểm soát sự biểu hiện của gen *tyrR* mã hóa protein TyrR. Theo khía cạnh khác của sáng chế, hoạt tính của yếu tố điều hòa âm tính TyrR hoàn toàn bị loại bỏ bằng cách loại bỏ hoặc bất hoạt gen *tyrR* ra khỏi nhiễm sắc thể ADN của vi sinh vật.

Theo phương án khác của sáng chế, sự ức chế phản hồi của một số enzym nhất định trong con đường axit amin thơm do một số chất chuyển hoá gây ra được khắc phục nhờ các thao tác di truyền. Ở hầu hết các chủng *E. coli* kiểu dại, deoxyarabino-heptulosonat 7-phosphat (DAHP) synthaza xuất hiện ở ba dạng isozym khác nhau được biết đến được mã hóa bởi ba gen khác nhau là *aroG*, *aroF* và *aroH*. Protein được mã hóa bởi một trong số ba gen này chịu sự ức chế phản hồi bởi một hoặc nhiều chất chuyển hoá của con đường axit shikimic chịu trách nhiệm sinh tổng hợp axit amin thơm. Theo một khía cạnh của sáng chế, gen *aroG* kiểu dại được thay thế bởi gen *aroG* biến đổi mã hóa protein AroG kháng với sự ức chế phản hồi bằng một hoặc nhiều chất chuyển hoá của con đường axit amin thơm trong tế bào vi sinh vật. Theo khía cạnh khác của sáng chế, gen *aroF* kiểu dại được thế bởi gen *aroF* mã hóa protein AroF kháng lại sự ức chế phản hồi bằng một hoặc nhiều chất chuyển hoá của con đường axit amin thơm trong tế bào vi sinh vật. Theo khía cạnh khác của sáng chế, gen *aroH* kiểu dại được thế bởi gen *aroH* mã hóa protein AroH kháng lại sự ức chế phản hồi bằng một hoặc nhiều chất chuyển hoá của con đường axit amin thơm trong tế bào vi sinh vật. Theo một phương án khác nữa của sáng chế, chất xúc tác sinh học được chọn để tổng hợp axit cis, cis-muconic trên quy mô thương mại lớn có thể có nhiều hơn một isozym kháng phản hồi đối với deoxyarabino-heptulosonat 7-phosphat (DAHP) synthaza.

Theo phương án khác của sáng chế, hoạt tính của một hoặc nhiều của enzym liên quan đến dòng cacbon qua con đường axit amin thơm trong tế bào vi sinh vật

được tăng cường. Theo một khía cạnh của sáng chế, việc gia tăng hoạt tính của một hoặc nhiều enzym liên quan đến hoạt động của con đường axit amin thơm và/hoặc con đường axit muconic được thực hiện bằng các thao tác di truyền. Theo một khía cạnh được ưu tiên của sáng chế, sự biểu hiện của một hoặc nhiều của các gen mã hóa enzym hoặc protein AroF, AroG, AroH, AroB, TktA, TalB, AroZ, QutC, qa-4, asbF, QuiC, AroY, Rpe, Rpi, Pps, CatA và CatX hoặc chất tương đồng hoặc chất tương tự của chúng gia tăng dẫn đến làm gia tăng hoạt tính của các enzym này. Rpe là ribuloza-5-phosphat epimeraza, Rpi là ribuloza-5-phosphat isomeraza, và Pps là phosphoenol pyruvat synthetaza (Neidhardt và Curtiss, 1996). Nếu chủng chủ là nấm men, ví dụ *Saccharomyces cerevisiae*, hoặc nấm mốc dạng sợi, ví dụ, *Neurospora crassa*, một số enzym xúc tác các phản ứng trong con đường shikimat có thể kết hợp thành một protein lớn hoặc polypeptit, gọi là Aro1p, được mã hóa bởi gen ARO1 trong trường hợp *S. cerevisiae*. Aro1p kết hợp các chức năng của AroB, AroD, AroE, AroK (or AroL), và AroA). Được hiểu theo cách thông thường, đối với mục đích của sáng chế, Aro1p, và ARO1, hoặc một phần của chúng, có thể được dùng làm phần thay thế hoặc bổ sung cho AroB, AroD, AroE, AroK, và/hoặc AroA.

Theo một phương án khác nữa của sáng chế, dòng qua erythroza-4-phosphat trong tế bào vi khuẩn được gia tăng bằng cách biểu hiện quá mức enzym trong hoạt động của con đường pentoza phosphat. Theo một khía cạnh của sáng chế, việc tổng hợp quá mức enzym transaldolaza được mã hóa bởi gen *talB* hoặc gen *talA* được bộc lộ. Theo khía cạnh khác của sáng chế, sự biểu hiện của các gen mã hóa cả transaldolaza và transketolaza được gia tăng bằng các thao tác di truyền. Theo khía cạnh khác nữa của sáng chế, sự biểu hiện của các gen mã hóa ribuloza-5-phosphat epimeraza hoặc ribuloza-5-phosphat isomeraza hoặc cả hai được gia tăng bằng các thao tác di truyền.

Theo phương án khác của sáng chế, PEP (phosphoenol pyruvat) có sẵn để tham gia vào con đường axit amin thơm được gia tăng thông qua thao tác di truyền. Theo một khía cạnh của sáng chế, sự cạnh tranh đối với nhóm PEP được giảm bằng cách loại bỏ và/hoặc bổ sung hệ phosphotransferaza phụ thuộc PEP (PEP-dependent phosphotransferase system: PTS) đối với sự hấp thu glucoza bằng hệ độc lập PEP

đối với sự hấp thu glucoza. Theo một phương án khác nữa của sáng chế, độ sinh khả dụng của PEP được gia tăng bằng cách gia tăng sự biểu hiện của gen mã hóa PEP synthetaza, như *pps*.

Mô tả vắn tắt các hình vẽ

FIG. 1. Con đường để tổng hợp axit amin thơm ở *E. coli*.

FIG. 2. Con đường tổng hợp axit muconic ở *E. coli*.

FIG. 3. Phổ sắc ký thể hiện các mẫu chuẩn được sử dụng để phân tích HPLC axit muconic và các sản phẩm hóa sinh trung gian.

FIG. 4. Phổ sắc ký thể hiện các mẫu chuẩn được sử dụng để phân tích HPLC các dạng đồng phân của axit muconic.

FIG. 5. Hiệu giá đối với việc tổng hợp DHS ở chủng *E. coli* MYR34 được biến đổi bằng plasmit pCP32AMP, pCP14 và pCP54. Chủng *E. coli* MYR34 có một đoạn khuyết ở gen *aroE*. Plasmit pCP32AMP biểu hiện gen *aroG* mã hóa DAHP synthaza. Plasmit pCP14 biểu hiện gen *aroB* mã hóa DHQ synthaza. Plasmit pCP54 biểu hiện cả gen *aroB* lẫn gen *aroG*.

FIG. 6. Hiệu giá đối với việc tổng hợp DHS ở chủng *E. coli* MYR34 và MYR170 được biến đổi bằng plasmit pCP32AMP và pCP54. Chủng *E. coli* MYR34 có một đoạn khuyết ở gen *aroE*. Chủng MYR170 có một đoạn khuyết của gen *aroE* và bản sao thứ hai của gen *aroB* dưới sự kiểm soát của trình tự khởi đầu P₁₅ được gắn ở locus *ack* của nhiễm sắc thể ADN chủ. Plasmit pCP32AMP biểu hiện gen *aroG* mã hóa DAHP synthaza. Plasmit pCP54 biểu hiện cả gen *aroB* lẫn gen *aroG*.

FIG. 7. Hiệu giá đối với việc tổng hợp axit cis, cis-muconic ở chủng *E. coli* MYR34 và MYR170 được biến đổi bằng một mình plasmit pMG37 hoặc bằng cả plasmit pMG37 lẫn pCP32AMP. Chủng *E. coli* MYR34 có một đoạn khuyết ở gen *aroE*. Chủng MYR170 có một đoạn khuyết ở gen *aroE* và bản sao thứ hai của gen *aroB* dưới sự kiểm soát của P₁₅ trình tự khởi đầu được gắn ở locus *ack* của nhiễm sắc thể ADN chủ. Plasmit pCP32AMP biểu hiện gen *aroG* mã hóa DAHP synthaza. Plasmit pMG37 biểu hiện *aroZ*, *aroY*, và *catAX*, các gen mã hóa protein tham gia vào con đường axit muconic.

FIG. 8. Hiệu giá đối với việc tổng hợp DHS ở chủng *E. coli* MYR170 được biến đổi bằng plasmit biểu hiện một mình gen *aroG* (pCP32AMP) hoặc biểu hiện đồng thời các gen *aroG* và *tktA* (pCP50). Chủng MYR170 có một đoạn khuyết ở gen *aroE* và bản sao thứ hai của gen *aroB* dưới sự kiểm soát của gen tăng cường P₁₅ được gắn ở locus *ack* của nhiễm sắc thể ADN chủ.

FIG. 9. Hiệu suất DHS từ chủng *E. coli* MYR34 và MYR170 được biến đổi bằng plasmit pCP32AMP và pCP50. Hiệu suất DHS được tính theo số gam DHS được tạo ra trên số gam glucoza tiêu thụ. Plasmit pCP32AMP biểu hiện gen *aroG* trong khi pCP50 biểu hiện gen *aroG* và *tktA*. Chủng vi khuẩn MYR34 có một đoạn khuyết ở gen *aroE*. Chủng *E. coli* MYR170 thu nhận từ MYR34 và có gen bổ sung *aroB* được gắn ở locus *ack* trên nhiễm sắc thể ADN.

FIG. 10. Hiệu giá DHS của chủng *E. coli* MYR170 và MYR261 được biến đổi bằng plasmit pCP32AMP và pCP50. Plasmit pCP32AMP biểu hiện gen *aroG* trong khi pCP50 biểu hiện các gen *aroB* và *tktA*. Chủng MYR170 có một đoạn khuyết ở gen *aroE* và bản sao thứ hai của gen *aroB* dưới sự kiểm soát của trình tự khởi đầu P₁₅ được gắn ở locus *ack* của nhiễm sắc thể ADN chủ. Chủng *E. coli* MYR261 thu nhận được từ chủng *E. coli* MYR170. Chủng *E. coli* MYR261 có bản sao thứ hai của gen *tktA* gen với trình tự khởi đầu nguyên thể của nó được gắn ở locus *poxB* của nhiễm sắc thể ADN.

FIG. 11. Quá trình tổng hợp axit muconic và axit axetic ở chủng *E. coli* MYR170, MYR261 và MYR305 được biến đổi bằng plasmit pCP32AMP biểu hiện *aroG* mã hóa DAHP synthaza trong con đường sinh tổng hợp axit shikimic và plasmit pMG37 biểu hiện các gen *aroZ*, *aroY* và *catAX* mã hóa protein có chức năng trong con đường axit muconic. Chủng MYR170 có một đoạn khuyết ở gen *aroE* và bản sao bổ sung của gen *aroB* dưới sự kiểm soát của trình tự khởi đầu P₁₅ được chèn ở locus *ack* ở nhiễm sắc thể ADN của vật chủ. MYR261 và MYR305 là dẫn xuất của chủng MYR170. MYR261 có bản sao bổ sung của gen *tktA* được gắn ở locus *poxB* trên nhiễm sắc thể ADN chủ trong khi MYR305 có phân xóa ở locus *poxB* trên nhiễm sắc thể ADN chủ.

FIG. 12. Sự chuyển hoá DHS nội sinh được tạo ra bởi chủng *E. coli* MYR34

thành axit muconic. Chủng *E. coli* MYR34 có một đoạn khuyết ở gen *aroE* mã hóa shikimat dehydroaza. Kết quả là có sự tích lũy DHS. Khi chủng MYR34 được biến đổi bằng plasmit biểu hiện *aroZ*, *aroY* và *catAX*, các gen mã hóa protein tham gia vào con đường axit muconic, có sự chuyển hoá DHS thành axit muconic. Tuy nhiên, không có sự chuyển hoá DHS thành axit muconic xảy ra khi chủng MYR34 được biến đổi bằng vật truyền plasmit trống (pCL1921) không có gen ngoại sinh bất kỳ.

FIG.13. So sánh các thể tương đồng *aroZ* về khả năng chuyển hóa DHS thành axit muconic. Ba thể tương đồng *aroZ* khác nhau, là *quiC* của *Acinetobacter sp. ADP1*, *asbF* của *Bacillus thuringiensis*, và *qa-4* của *Neurospora crassa* được tách dòng dưới sự kiểm soát của trình tự khởi đầu P₂₆ ở plasmit có ít bản sao cũng biểu hiện các gen *catAX* và *aroY* của P₁₅ và trình tự khởi đầu lambda P_R. Ba cấu trúc plasmit khác nhau được biểu hiện ở MYR34 thông qua sự biến nạp và lượng axit muconic tạo ra được xác định.

FIG. 14. Bản sao đơn của *catAX*, *aroY* và *quiC* được gắn vào nhiễm sắc thể của chủng *E. coli* MYR170 (Δ *aroE*, Δ *ack::P₁₅-aroB*) thu được MYR352 (SEQ ID No. 41). MYR170 cũng được biến đổi bằng plasmit có ít bản sao pMG37 mang tất cả các gen cần thiết đối với hoạt động của con đường axit muconic tạo ra chủng MYR219. Cả MYR352 lẫn MYR219 được biến đổi bằng YEp24 (vật truyền trống có số bản sao trung bình) hoặc pCP32AMP (plasmit biểu hiện *aroG* có số bản sao trung bình) hoặc pCP50 (plasmit biểu hiện *tkiA* và *aroG* có số bản sao trung bình) và hàm lượng PCA, catechol và axit muconic được tạo ra được định lượng bằng cách sử dụng phương pháp HPLC.

FIG. 15. Việc loại bỏ sự tích lũy catechol ở MYR352 bằng cách gia tăng biểu hiện của *catAX*. MYR352 được biến đổi bằng plasmit biểu hiện một *aroY* (pMG27) hoặc plasmit biểu hiện một *quiC* (pMG39) hoặc plasmit biểu hiện cả ba gen trong con đường axit muconic là *catAX*, *aroY* và *quiC* (pMG37) hoặc plasmit biểu hiện hai gen trong con đường axit muconic là *catAX* và *aroY* (pMG33). Sự biểu hiện quá mức của một *catAX* là đủ để ngăn ngừa sự tích lũy catechol.

FIG. 16. Sự phát triển của chủng sử dụng các hệ khác nhau để hấp thu

glucoza. Việc loại bỏ *ptsHI* và *galP* (MYR31) dẫn đến không có sự phát triển trong môi trường glucoza tối thiểu, trong khi việc cài các gen *glf* và *glk* (MYR217) dẫn đến sự phát triển ngược. Chúng kiểm soát MYR34 là $\Delta aroE$, theo cách khác là kiểu dại. Ba axit amin thơm và ba vitamin thơm được bổ sung vào môi trường để chúng đặc dưỡng phát triển.

FIG. 17. Quá trình tổng hợp DHS ở chủng *E. coli* MYR34 và MYR217. Khi được biến đổi bằng plasmit dẫn đến việc tổng hợp DHS, MYR217, sử dụng *glf-glK* để nhận glucoza, tạo ra hiệu giá DHS cao hơn các thể biến đổi của MYR34, sử dụng hệ phosphotransferaza (PTS)

FIG. 18. Quá trình tổng hợp axit muconic ở chủng *E. coli* MYR428 trong thiết bị lên men dung tích 7l. Chủng *E. coli* MYR261 có kiểu gen $\Delta aroE \Delta ackA::P_{15-aroB} \Delta poxB::tktA$ được biến đổi bằng plasmit pCP32AMP và pMG37 tạo ra chủng *E. coli* MYR428.

Mô tả chi tiết sáng chế

Như được sử dụng theo sáng chế, thuật ngữ “ví dụ” hoặc “như” có nghĩa là có nhiều hơn một phương pháp, cách thức, giải pháp hoặc thành phần của chất đối với đối tượng sắp nói đến, và ví dụ đưa ra không nhằm giới hạn ở ví dụ đó.

Thuật ngữ “khác loài” chỉ gen hoặc protein không có trong sinh vật một cách tự nhiên hoặc được cho là nguyên thể, mà có thể được đưa vào sinh vật bằng thao tác di truyền như bằng cách biến đổi, ghép đôi, hoặc tải nạp. Gen khác loài có thể được gắn (được chèn) vào nhiễm sắc thể hoặc có trên plasmit. Thuật ngữ “ngoại sinh” chỉ gen hoặc protein đã được đưa vào hoặc được biến đổi, trong sinh vật nhằm mục đích tăng, giảm, hoặc khử hoạt tính, bằng cách thao tác di truyền, như bằng cách biến đổi, ghép đôi, tải nạp, hoặc gây đột biến. Gen hoặc protein ngoại sinh có thể khác loài hoặc có thể là gen hoặc protein nguyên thể đối với sinh vật chủ, nhưng được biến đổi bằng một hoặc nhiều phương pháp, ví dụ, đột biến, làm mất đoạn, thay trình tự khởi đầu, thay gen kết thúc, sao chép, hoặc chèn một hoặc nhiều bản sao bổ sung vào nhiễm sắc thể hoặc plasmit. Do đó, ví dụ, nếu bản sao thứ hai của gen *aroB* được chèn vào một vị trí trên nhiễm sắc thể khác với vị trí nguyên bản, bản sao thứ hai này là ngoại sinh.

Thuật ngữ “vi sinh vật” theo sáng chế bao gồm vi khuẩn, khuẩn cổ, nấm men, tảo và nấm sợi có thể được sử dụng để tổng hợp axit cis, cis-muconic trên quy mô thương mại lớn thông qua quá trình lên men.

Về danh pháp, gen hoặc vùng mã hóa thường được đặt tên bằng các chữ thường in nghiêng, ví dụ “*aroZ*”, còn enzym hoặc protein được mã hóa bởi gen có thể được đặt tên bằng các chữ cái giống như vậy, nhưng chữ cái đầu tiên dưới dạng chữ in hoa và không in nghiêng, ví dụ “AroZ” hoặc “Aro1p”, tên thứ hai là ví dụ về quy ước được sử dụng trong nấm men để đặt tên cho enzym hoặc protein. “p” là chữ viết tắt cho protein, được mã hóa bởi gen được đặt tên. Enzym hoặc protein này cũng có thể được mô tả bằng một tên cụ thể hơn, ví dụ, AroZ cũng có thể được mô tả là 3-dehydroshikimat dehydrataza. Gen hoặc vùng mã hóa mã hóa một enzym có hoạt tính xúc tác cụ thể có một số tên khác nhau do có nguồn gốc lịch sử khác nhau hoặc do gen bắt nguồn từ các loài khác nhau. Ví dụ, gen mã hóa 3-dehydroshikimat dehydrataza của *Bacillus thuringiensis* hoặc *Bacillus anthracis* có thể được đặt tên là *asbF* thay cho *aroZ*, gen họ hàng của *Aspergillus nidulans* có thể được đặt tên là *qutC*, gen họ hàng của *Neurospora crassa* có thể được đặt tên là *qa-4*, và gen họ hàng của *Acinetobacter baylyi* (còn được biết đến là *Acinetobacter calcoaceticus* và *Acinetobacter* Sp. ADP1) có thể được đặt tên là *quiC*.

“Plasmit” là phân tử ADN mạch vòng hoặc mạch thẳng về cơ bản nhỏ hơn nhiễm sắc thể, tách ra từ nhiễm sắc thể hoặc các nhiễm sắc thể của vi sinh vật, và sao chép một cách độc lập từ nhiễm sắc thể hoặc các nhiễm sắc thể. “Plasmit” có mặt ở dạng một bản sao/tế bào hoặc nhiều hơn một bản sao/tế bào. Việc duy trì plasmit trong tế bào vi sinh vật thường cần đến việc chọn lọc chất kháng sinh, nhưng việc bổ sung các sinh vật dinh dưỡng thụ động cũng có thể được sử dụng.

Thuật ngữ “nhiễm sắc thể” hoặc “ADN của nhiễm sắc thể” như được sử dụng trong sáng chế liên quan đến tế bào vi khuẩn là phân tử ADN mạch vòng về cơ bản lớn hơn plasmit và không cần đến sự chọn lọc chất kháng sinh bất kỳ.

“Catxet biểu hiện” là trình tự ADN có thể là một phần nhiễm sắc thể hoặc plasmit mà chứa ít nhất một trình tự khởi đầu và gen hoặc vùng mã hóa enzym hoặc protein khác, để vùng mã hóa này được biểu hiện bởi trình tự khởi đầu, và enzym

hoặc protein được tạo ra bởi tế bào chủ chứa trình tự ADN này. “Catxet biểu hiện” có thể được tổng hợp ít nhất một phần hoặc được tạo ra bằng phương pháp thiết kế di truyền, để vùng mã hóa này được biểu hiện bởi trình tự khởi đầu không liên quan một cách tự nhiên đến vùng mã hóa. Tùy ý, “catxet biểu hiện” có thể chứa gen kết thúc phiên mã có thể hoặc không phải là gen kết thúc có liên quan một cách tự nhiên đến vùng mã hóa. “Catxet biểu hiện” có thể có vùng mã hóa cho nhiều hơn một protein, trong trường hợp này nó được gọi là operon hoặc operon tổng hợp.

“Biểu hiện quá mức” của gen hoặc vùng mã hóa có nghĩa là việc enzym hoặc protein được mã hóa bởi gen hoặc vùng mã hóa đó được tạo ra ở vi sinh vật chủ ở mức cao hơn mức nhận thấy ở vi sinh vật chủ kiểu dại trong cùng điều kiện phát triển hoặc trong các điều kiện phát triển tương tự. Việc này có thể thực hiện được bằng, ví dụ, một hoặc nhiều của phương pháp sau: 1) cải trình tự khởi đầu mạnh hơn, 2) cải vị trí gắn ribosom mạnh hơn, như trình tự ADN 5'-AGGAGG, ở từ bốn đến mười bazơ xuôi dòng với codon khởi đầu dịch mã, 3) cải gen kết thúc hoặc gen kết thúc mạnh hơn, 4) cải thiện việc chọn codon ở một hoặc nhiều vị trí trong vùng mã hóa, 5) cải thiện tính ổn định của ARN thông tin, và 6) gia tăng số lượng bản sao của gen, bằng cách đưa nhiều bản sao vào nhiễm sắc thể hoặc đặt catxet trên plasmid nhiều bản sao. Enzym hoặc protein được tạo ra từ gen biểu hiện quá mức được gọi là “được tổng hợp quá mức”. Gen mà “biểu hiện quá mức” hoặc protein được “tổng hợp quá mức” có thể là nguyên thể đối với vi sinh vật chủ, hoặc nó có thể được cấy ghép bằng các phương pháp thiết kế di truyền từ các sinh vật khác nhau thành vi sinh vật chủ, trong trường hợp này, enzym hoặc protein và gen hoặc vùng mã hóa mã hóa enzym hoặc protein được gọi là “khác loài”. Các gen và protein khác loài là nằm trong định nghĩa về sự biểu hiện quá mức và tổng hợp quá mức vì chúng không có mặt trong sinh vật chủ không được thiết kế.

“Thể tương đồng” của gen, trình tự ADN, hoặc protein thứ nhất là gen, trình tự ADN, hoặc protein thứ hai thực hiện chức năng sinh học tương tự của gen, trình tự ADN hoặc protein thứ nhất và có ít nhất 25% mức tương đồng trình tự (khi so sánh với trình tự protein hoặc so sánh với trình tự protein thu được từ trình tự gen) với gen hoặc protein thứ nhất này, như được xác định bằng phần mềm máy tính BLAST dùng để so sánh trình tự (Altschul et al., 1990; Altschul et al., 1997), và cho

phép xóa và chèn. Ví dụ về thể tương đồng của gen *aroG* của *E. coli* là gen *aroG* của chủng *Salmonella typhimurium*.

“Thể tương tự” của gen, trình tự ADN, hoặc protein thứ nhất là gen, trình tự ADN, hoặc protein thứ hai thực hiện chức năng sinh học tương tự của gen, trình tự ADN hoặc protein thứ nhất và có ít hơn 25% mức tương đồng trình tự (khi so sánh với trình tự protein hoặc so sánh với trình tự protein thu được từ trình tự gen) với gen hoặc protein thứ nhất này, như được xác định bằng phần mềm máy tính BLAST dùng để so sánh trình tự (Altschul et al., 1990; Altschul et al., 1997), và cho phép xóa và chèn. Ví dụ về thể tương tự của protein AroZ *Klebsiella pneumoniae* là protein QutC *Aspergillus nidulans*, vì cả hai protein này là enzym xúc tác phản ứng 3-dehydroshikimat dehydrataza, nhưng không có sự tương đồng trình tự đáng kể giữa hai enzym hoặc các gen tương ứng của chúng. Chuyên gia trong lĩnh vực này sẽ nhận thấy rằng có nhiều enzym và protein có chức năng sinh học cụ thể, ví dụ DAHP synthaza hoặc 3-dehydroshikimat dehydrataza, có thể có ở nhiều sinh vật khác nhau, ở dạng thể tương đồng hoặc chất tương tự, và vì các thành viên của họ enzym hoặc protein này có cùng chức năng, mặc dù chúng có thể khác nhau một chút hoặc về cơ bản là khác nhau về cấu trúc, các thành viên khác nhau của một họ trong nhiều trường hợp có thể được sử dụng để thực hiện cùng chức năng sinh học bằng cách sử dụng phương pháp thiết kế gen hiện nay. Ví dụ, enzym AroZ và QutC cùng xúc tác cho phản ứng, DHS dehydrataza, một trong hai enzym này sẽ dẫn đến việc tổng hợp axit cis, cis-muconic với lượng thích hợp và việc chọn một trong hai enzym này có thể thực hiện bằng cách chọn enzym nào có hiệu giá axit cis, cis-muconic cao hơn trong các điều kiện lên men giống nhau.

“Nguồn cacbon không thơm” hoặc “hợp chất không thơm” là hợp chất chứa cacbon có thể được dùng để cung cấp cho vi sinh vật theo sáng chế làm nguồn cacbon và/hoặc năng lượng, trong đó hợp chất không chứa vòng 6 cạnh liên quan đến benzen. Ví dụ về nguồn cacbon không thơm gồm glucoza, xyloza, lactoza, glyxerol, axetat, arabinoza, galactoza, mannoza, maltoza, hoặc sucroza. “Hợp chất thơm” là hợp chất mà chứa một hoặc nhiều vòng 6 cạnh liên quan đến benzen. Ví dụ về hợp chất thơm là catechol, hoặc 1,2-dihydroxy benzen.

“Trình tự khởi đầu cơ định mạnh” là trình tự ADN thường là nằm ngược chiều (với mặt 5' của gen khi được thể hiện theo hướng từ 5' đến 3' thông thường), của trình tự ADN hoặc gen được phiên mã bởi ARN polymeaza và làm cho trình tự ADN hoặc gen biểu hiện bằng cách phiên mã bởi ARN polymeaza ở mức độ dễ dàng phát hiện trực tiếp hoặc gián tiếp bằng quy trình thử nghiệm thích hợp bất kỳ. Ví dụ về quy trình thử nghiệm thích hợp gồm 1) định lượng enzym transcriptaza ngược cùng với PCR, 2) thử nghiệm enzym đối với enzym đã được mã hóa, 3) gel protein được nhuộm bằng Coomassie Blue hoặc 4) Tạo ra chất chuyển hoá có thể xác định được mà chất chuyển hoá này được tạo ra một cách gián tiếp là kết quả của quá trình phiên mã này và quá trình phiên mã xác định được này xảy ra với sự có mặt hoặc không có mặt của protein điều hòa mức phiên mã, chất chuyển hoá, hoặc hóa chất cảm ứng. Ví dụ về trình tự khởi đầu không phải là “trình tự khởi đầu cơ định mạnh” là trình tự khởi đầu P_{lac} của *E. coli*, vì gen này bị kìm hãm bởi gen kìm hãm với sự vắng mặt của lactoza hoặc chất gây cảm ứng IPTG. Bằng cách sử dụng các phương pháp đã biết trong lĩnh vực này, “trình tự khởi đầu cơ định mạnh” có thể được dùng thay cho trình tự khởi đầu nguyên thể (trình tự khởi đầu tồn tại một cách tự nhiên ngược chiều với trình tự ADN hoặc gen), dẫn đến catxet biểu hiện có thể được đặt trong plasmit hoặc nhiễm sắc thể và tạo ra mức biểu hiện trình tự ADN hoặc gen mong muốn cao hơn mức của trình tự khởi đầu nguyên thể. Trình tự khởi đầu cơ định mạnh có thể đặc hiệu đối với loài hoặc giống nhưng thông thường trình tự khởi đầu cơ định mạnh của vi khuẩn có chức năng tốt ở các vi khuẩn có liên quan. Ví dụ, trình tự khởi đầu của *Bacillus subtilis* hoặc thể thực khuẩn thường phát triển trên *B. subtilis* có thể có chức năng tốt ở *E. coli*. “Trình tự khởi đầu cơ định mạnh” về cơ bản khác với trình tự khởi đầu cảm ứng, như P_{tac} , đã được sử dụng trong phương pháp tổng hợp axit cis, cis-muconic đã biết và thường đòi hỏi hóa chất đắt tiền hoặc cần sự thay đổi môi trường khác đối với chức năng mong muốn (Niu et al., 2002). Ví dụ về trình tự khởi đầu cơ định mạnh là P_{15} , P_{26} , của thể thực khuẩn *Bacillus subtilis* SP01, và thể thực khuẩn coli lambda P_R (các SEQ ID No. 1, 2, và 3).

“Con đường muconic” hoặc “con đường axit muconic” chỉ con đường hóa sinh từ DHS đến PCA đến catechol đến axit cis, cis-muconic, và “gen của con

đường muconic” hoặc “gen của con đường axit muconic” là gen mã hóa enzym xúc tác một bước trong con đường muconic hoặc mã hóa một chức phụ được sử dụng để gia tăng hoạt tính của một trong số các enzym này, ví dụ, *aroZ*, *aroY*, *catA*, *catX*, và *qutC*. DHS là chữ viết tắt của 3-dehydroshikimat, và PCA là chữ viết tắt của protocatechuic axit. “Plasmit muconic” là plasmit mà có chứa một hoặc nhiều gen của con đường muconic.

Một số thao tác di truyền được sử dụng trong sáng chế được đặt ở giữa quanh con đường thông thường để sinh tổng hợp axit amin thơm có trong các tế bào vi khuẩn như được thể hiện trên Fig.1. Con đường thông thường để sinh tổng hợp axit amin thơm như được thể hiện trên Fig.1 từ DAHP synthaza đến chorismat synthaza cũng được đề cập là “con đường axit shikimic”.

Đã có một lượng đáng kể các tài liệu đã được công bố về việc thiết kế gen của vi sinh vật để tổng hợp axit amin thơm, phenylalanin, tycolophan và tryptophan (US 4,681,852, US 4,753,883, US 6,180,373, EP 86300748,0). Các phương pháp bao gồm việc sử dụng các hỗn hợp enzym kháng phản hồi (*AroF*, *AroG*, *PheA*, *TyrA*), làm mất điều hòa sự kìm hãm phiên mã (*tyrR*), gia tăng độ mạnh của trình tự khởi đầu (P_{tac} , P_{lac}) và gia tăng số bản sao của một hoặc nhiều gen (*tklA*). Tuy nhiên, nhiều hỗn hợp đặc hiệu của các phương pháp trên không được thử nghiệm, do có quá nhiều hỗn hợp để thử nghiệm hoặc do thiếu kiến thức để xem xét hỗn hợp nào là hỗn hợp tốt nhất. Quan trọng hơn, mức độ thích hợp của hỗn hợp bất kỳ trong số các hỗn hợp thao tác di truyền này đối với việc phát triển chất xúc tác sinh học để tổng hợp công nghiệp axit muconic bằng cách sử dụng nguồn cacbon không thơm thu hồi được là chưa được biết đến.

Con đường sinh tổng hợp axit amin thơm là đã biết đối với nhiều vi sinh vật, đặc biệt là *E. coli* (Neidhardt và Curtiss, 1996). Trong tế bào kiểu đại, con đường này được điều hòa nhẹ bằng cách ức chế phản hồi và kìm hãm sự phiên mã. Bước đầu tiên được xúc tác bởi enzym deoxy-arabino-heptulosonat 7-phosphat (DAHP) synthaza, enzym này có ba isozym được mã hóa bởi *aroF*, *aroG*, và *aroH*. Ba isozym này, *AroF*, *AroG*, và *AroH*, ức chế phản hồi các sản phẩm của con đường sinh tổng hợp amin thơm lần lượt là tyrosin, phenylalanin, và tryptophan. Thế đột

biến kháng phản hồi của cả ba isozym là đã biết (Draths et al., 1992; Lutke-Eversloh và Stephanopoulos, 2007; Hu et al., 2003; Shumilin et al., 1999). Một khía cạnh theo sáng chế đề cập đến việc sử dụng các alen kháng phản hồi của các gen *aroF*, *aroG*, và *aroH* để biểu hiện protein enzym AroF, AroG và AroH kháng lại sự ức chế phản hồi bằng các sản phẩm của con đường sinh tổng hợp axit amin thơm. Sự phiên mã một số operon liên quan đến con đường thơm được điều hòa bằng gen kìm hãm mã hóa bởi gen *tyrR* hoặc gen kìm hãm mã hóa bởi gen *trpR* hoặc cả hai gen này (Neidhardt et al., 1996). Sự điều hòa âm sự phiên mã của gen *aroG* và *aroF* bằng protein TyrR khi nó được gắn kết với một hoặc nhiều của axit amin thơm là đặc biệt quan trọng. Một khía cạnh của sáng chế đề cập đến việc loại bỏ điều hòa âm ở gen *tyrR* hoặc *trpR* bằng cách loại bỏ các gen này ra khỏi nhiễm sắc thể của chủng vi khuẩn chủ.

Mục đích của sáng chế là tạo ra hỗn hợp mới gồm các catxet được thiết kế di truyền có các yếu tố được thiết kế di truyền mới để làm gia tăng các thông số lên men và tính thích hợp để tổng hợp trên quy mô thương mại lớn của chủng tổng hợp axit cis, cis-muconic. Cụ thể, giải pháp đã biết đối với việc tổng hợp axit cis, cis-muconic không đề cập đến một số hỗn hợp nhất định có các yếu tố gen này, ví dụ, nhưng không giới hạn ở, các hỗn hợp khác nhau gồm gen AroG kháng phản hồi tổng hợp quá mức, AroF kháng phản hồi tổng hợp quá mức, *tktA* biểu hiện quá mức, *talA* biểu hiện quá mức, catxet được gắn vào nhiễm sắc thể để biểu hiện *aroZ*, *aroY*, và *catAX* (hoặc chất tương tự hoặc thể tương đồng của chúng) từ trình tự khởi đầu cơ định mạnh và alen rò *aroE*, mà được định nghĩa là gen mã hóa enzym AroE có khả năng nguyên dưỡng đối với axit amin thơm và vitamin, nhưng không làm tiết đáng kể các hợp chất thơm.

Tất cả các ví dụ cụ thể về việc xây dựng các chủng được mô tả ở đây được thu nhận từ chủng *E. coli* C kiểu đại (ATCC 8739), hoặc chủng *E. coli* K-12 (YMC9 hoặc MM294) nhưng các yếu tố di truyền được mô tả ở đây có thể được tổ hợp ở chủng *E. coli* thích hợp khác và catxet biểu hiện hoặc thể tương tự và thể tương đồng thích hợp của yếu tố di truyền được mô tả ở đây được tổ hợp trong vi sinh vật thích hợp bất kỳ như vi khuẩn, khuẩn cổ, nấm men, tảo, và nấm sợi khác có thể được sử dụng để tổng hợp trên quy mô thương mại axit cis, cis-muconic thông qua

quy trình lên men.

Ở chủng *E. coli*, con đường sinh tổng hợp axit amin thơm từ glucoza bắt đầu bằng nhánh không oxy hóa của con đường pentoza phosphat (PPP). Bốn enzym then chốt đối với con đường pentoza phosphat không oxy hóa là transketolaza, transaldolaza, ribuloza-5-phosphat epimeraza và ribuloza-5-phosphat isomeraza. Các enzym này xúc tác phản ứng tạo ra erythroza 4-phosphat (E4P) từ đường hexoza hoặc pentoza. Để gia tăng độ sinh khả dụng của E4P ở chủng *E. coli*, gen *tktA* mã hóa transketolaza có thể biểu hiện quá mức (Niu et al., 2002). Tương tự, sự biểu hiện quá mức của gen transaldolaza cũng được cho là làm gia tăng độ sinh khả dụng của E4P trong một số trường hợp (Bongaerts et al., 2001). Theo khía cạnh khác của sáng chế, sự biểu hiện của cả gen transketolaza lẫn transaldolaza gia tăng bằng các thao tác di truyền dẫn đến sự gia tăng hoạt tính của enzym transketolaza và transaldolaza. Theo khía cạnh khác của sáng chế, dòng qua nhanh không oxy hóa của PPP được gia tăng bằng cách tổng hợp quá mức ribuloza-5-phosphat epimeraza và ribuloza-5-phosphat isomeraza.

Bước đầu tiên và phản ứng được điều hòa một cách chặt chẽ nhất trong con đường axit amin thơm thông thường là phản ứng ngưng tụ phosphoenolpyruvat (PEP) và E4P tạo ra deoxyarabino-heptulosonat 7-phosphat (DAHP) bằng enzym DAHP synthaza (mã hóa bởi *aroG*, *aroF*, và *aroH*). D-glucoza được *E. coli* sử dụng được đưa vào quy trình sinh tổng hợp thơm một phần qua PPP, và một phần qua quy trình thủy phân glucoza. Dòng glucoza đi vào con đường thơm gia tăng đáng kể khi transketolaza (*tktA*) và isozym của DAHP synthaza (*aroG*) được khuếch đại bằng cách biến đổi bằng plasmid làm gia tăng sự biểu hiện của chúng bằng cách gia tăng số lượng bản sao (Niu et al., 2002). Theo khía cạnh được ưu tiên của sáng chế, gen *aroG* và *tktA* ngoại sinh được gắn vào ADN nhiễm sắc thể để khuếch đại hoạt tính của enzym transketolaza và DAHP synthaza.

Theo phương án khác của sáng chế, dòng qua PEP trong tế bào vi sinh vật được cải thiện bằng cách gia tăng PEP có sẵn để tổng hợp DAHP. Nhiều gen của tế bào vi khuẩn sử dụng PEP để vận chuyển glucoza qua màng tế bào bằng cách sử dụng hệ phosphotransferaza (PTS) trong đó một phân tử PEP được sử dụng cho mọi

phân tử glucoza được vận chuyển qua màng ngoài của vi khuẩn. Bằng cách thế hoặc bổ sung PTS phụ thuộc PEP bằng cơ chế hấp thu glucoza không phụ thuộc PEP (độc lập với PEP), có thể làm gia tăng cỡ nhóm PEP có sẵn đối với con đường sinh tổng hợp axit amin thơm ở tế bào vi sinh vật. Ví dụ, hệ PTS để hấp thu đường có thể được thế hoặc bổ sung bằng hệ hấp thu đường dựa trên GalP hoặc hệ vận chuyển đường dựa trên protein Glf/Glk (ChADNran et al., 2003; Yi et al., 2003). Theo khía cạnh được ưu tiên của sáng chế bên cạnh việc loại bỏ hệ PTS để hấp thu đường nhằm bảo toàn nhóm PEP trong tế bào vi sinh vật, hệ hấp thu đường dựa trên GalP cũng bị bất hoạt nhằm bảo toàn ATP trong tế bào vi sinh vật. Ở tế bào vi sinh vật không có chức năng của cả hệ PTS và hấp thu đường dựa trên Gal-P (Δ PTS/ Δ galP), sự hấp thu đường có thể thực hiện được bằng cách đưa gen ngoại sinh mã hóa Glf, hoặc các gen ngoại sinh mã hóa cả protein Glf (protein khuếch tán glucoza qua trung gian) và Glk (glucokinaza). Theo sáng chế, thuật ngữ protein protein khuếch tán glucoza qua trung gian chỉ protein Glf bất kỳ cũng như protein bất kỳ khác có chức năng tương đương với Glf và chức năng vận chuyển đường vào tế bào vi sinh vật bằng cách khuếch tán. Theo một khía cạnh của sáng chế, gen mã hóa protein khuếch tán glucoza qua trung gian Glf được đưa vào tế bào vi sinh vật là Δ PTS/ Δ galP và glucoza được vận chuyển vào tế bào vi sinh vật được phosphoryl hoá bằng glucoza kinaza nội sinh. Theo khía cạnh khác của sáng chế, các gen mã hóa cả hai protein Glf và Glk được đưa vào tế bào vi sinh vật là Δ PTS/ Δ galP. Theo khía cạnh được ưu tiên của sáng chế, các gen *glf* và *glk* ngoại sinh được đưa vào tế bào vi sinh vật được gắn vào ADN của nhiễm sắc thể chủ.

Theo phương án khác của sáng chế, khi nguồn cacbon để sinh trưởng và năng lượng cần thiết để tổng hợp glucoza (ví dụ, nguồn cacbon là axetat hoặc suxinat), nhóm PEP có thể gia tăng bằng cách gia tăng hoạt tính của enzym cacboxy hóa có sẵn trong tế bào, ví dụ PEP carboxykinaza, được mã hóa bởi *pep* ở *E. coli*, hoặc bằng cách đưa enzym cacboxy hóa ngoại sinh vào. Theo phương án được ưu tiên, gen ngoại sinh được đưa vào mã hóa enzym cacboxy hóa có thể được gắn ổn định vào nhiễm sắc thể chủ. Các gen mã hóa enzym cacboxy hóa có thể thu nhận được từ các loại vi khuẩn khác nhau. Các gen mã hóa enzym cacboxy hóa còn có thể được đưa vào các thao tác di truyền để sự biểu hiện của enzym cacboxy hóa trong chất

xúc tác sinh học để tổng hợp axit cis, cis-muconic gia tăng đáng kể.

Theo một phương án khác nữa của sáng chế, nhóm PEP bên trong tế bào vi sinh vật gia tăng bằng cách giảm hoặc khử hoạt tính của enzym pyruvat kinaza như PykA và PykF sử dụng PEP làm cơ chất.

Từ DAHP, con đường axit amin thơm thực hiện qua một số sản phẩm trung gian đến chorismat (CHA), điểm phân nhánh của quy trình sinh tổng hợp ba axit amin thơm là L-Tycolophan (L-Tyr), L-Phenylalanin (L-Phe), và L-Tryptophan (L-Trp).

Trong các giai đoạn ban đầu của con đường axit amin thơm thông thường, 3-dehydroquinat (DHQ) synthaza (AroB) loại nhóm phosphat khỏi DAHP tạo thành DHQ. Enzym DHQ dehydrataza (AroD) loại phân tử nước khỏi DHQ tạo ra 3-dehydroshikimat (DHS) sau đó hợp chất được khử thành shikimat (SHK) bằng shikimat dehydroaza (AroE). Shikimat kinaza I/II (AroK, AroL) phosphoryl hóa shikimat thành shikimat 3-phosphat (S3P). S3P ngưng tụ với PEP tạo ra 5-enolpyruvoylshikimat 3-phosphat (EPSP). Việc tạo ra EPSP được điều tiết bởi EPSP synthaza (AroA). Nhóm phosphat được loại khỏi EPSP bằng chorismat synthaza (AroC) tạo ra chorismat (CHA).

Như được thể hiện trên Fig.2, con đường axit amin thơm có thể bị phong bế ở giai đoạn chuyển hoá 3-dehydroshikimat (DHS) thành shikimat (SHK) bằng cách đột biến gen *aroE* dẫn đến sự tích lũy DHS (Niu et al., 2002). Việc đưa gen *aroZ* ngoại sinh vào để chuyển hóa DHS thành protocatechuat (PCA). Sau đó, PCA chuyển hóa thành catechol qua phản ứng khử carboxyl hóa qua trung gian enzym AroY. Cuối cùng, catechol được chuyển hóa thành cis-cis axit muconic (ccMuA) nhờ tác dụng của sản phẩm gen *catA*. ccMuA có thể có tác dụng nhờ maleyl acetoacetat isomeraza để tạo ra trans-trans axit muconic (ttMuA). Con đường sinh tổng hợp từ DHS đến ccMuA và ttMuA được đề cập là con đường axit muconic. Ba gen khác nhau chịu trách nhiệm đối với sự chuyển hoá DHS thành ccMuA có thể thu được từ các loài vi sinh vật khác nhau và được đưa vào vi sinh vật được chọn để tổng hợp axit muconic như *Escherichia coli*. Theo một khía cạnh được ưu tiên của sáng chế, các gen ngoại sinh mã hóa protein liên quan đến con đường axit muconic

được gắn thành vào ADN nhiễm sắc thể chủ.

Liên quan đến con đường axit amin thơm để tổng hợp axit cis, cis-muconic sự đột biến của gen *aroE* là cần thiết. Gen *aroE* có thể bị bất hoạt hoàn toàn dẫn đến sự phong bế hoàn toàn quy trình sinh tổng hợp axit amin thơm như đối với chủng *E. coli* WN1/pWN2.248 được mô tả để tổng hợp axit muconic (Niu et al., 2002). Một hạn chế của chủng *E. coli* WN1/pWN2.248 và các chủng liên quan là do sự bất hoạt hoàn toàn của gen *aroE*, chủng này trở nên dinh dưỡng thụ động đối với axit thơm như phenylalanin, tycolophan và tryptophan, và vitamin thơm hoặc hợp chất giống vitamin nêu trên. Kết quả là, trong quá trình phát triển, để tổng hợp axit cis, cis-muconic chủng này cần bổ sung ngoại sinh sáu hợp chất này (hoặc sản phẩm trung gian thông thường như shikimat), do đó làm tăng đáng kể chi phí tổng hợp trên quy mô thương mại lớn axit cis, cis-muconic khi sử dụng chủng này. Một phương pháp mới khắc phục sự phụ thuộc vào nguồn axit amin thơm ngoại sinh này là sử dụng chủng có sự đột biến rò ở gen *aroE*. Thể đột biến rò của gen *aroE* cho phép dòng cacbon giới hạn đến axit shikimic trong khi tích lũy đáng kể một lượng DHS có sẵn để sau đó chuyển hoá thành PCA nhờ tác dụng của enzym AroZ. Việc sử dụng thể đột biến rò của *aroE* làm giảm sự phụ thuộc vào axit amin thơm ngoại sinh, trong khi vẫn chuyển hóa dòng cacbon thành axit cis, cis-muconic.

Các gen mã hóa đối với sự tổng hợp protein AroZ, AroY và CatA cần thiết đối với sự chuyển hoá DHS thành axit cis, cis-muconic có thể thu nhận được từ nhiều loài vi khuẩn. Theo một phương án, các gen ngoại sinh này được gắn vào nhiễm sắc thể chủ của chất xúc tác sinh học được phát triển. Theo phương án được ưu tiên, biểu hiện của các gen ngoại sinh này trong chất xúc tác sinh học được điều khiển bằng trình tự khởi đầu cơ định mà không cần đến tác nhân gây cảm ứng bất kỳ.

Enzym 3-dehydroshikimat dehydrataza (AroZ; EC 4,2,1,118) là cần thiết cho quy trình sinh tổng hợp hợp chất protocatechuat trung gian. Theo sáng chế, “AroZ” chỉ enzym bất kỳ xúc tác phản ứng 3-dehydroshikimat dehydrataza. Trong các giải pháp đã biết, enzym này được biểu hiện từ gen *aroZ* của chủng *Klebsiella pneumoniae* A170-40 (ATCC25597) (Niu et al., 2002; Draths and Frost, 1995). Tuy nhiên, hoạt tính đặc hiệu của AroZ thay đổi nhiều giữa các sinh vật, từ 0,1 đến 261

micromol/phút/mg (Wheeler et al, 1996; Fox et al, 2008; Pflieger et al, 2008), đến mức có thể có sự cải thiện đáng kể bằng cách biểu hiện gen *aroZ* còn được biết đến là *asbF* (Fox et al, 2008; Pflieger et al, 2008), *qutC* (Wheeler et al, 1996), *qa-4* (Rutledge, 1984), và *quiC*, của sinh vật có hoạt tính đặc hiệu cao hơn *K. pneumoniae*, ví dụ *Acinetobacter baylyi*, *Aspergillus nidulans* (Wheeler et al, 1996), hiện nay còn được biết đến là *Emecricella nidulans*, hoặc *Neurospora crassa* (Rutledge, 1984; Stroman et al, 1978), hoặc *Podospora anserina*, còn được biết đến là *Podospora pauciseta* (Hansen et al, 2009).

Làm một ví dụ cụ thể, trình tự mã hóa đối với gen *qa-4* của *N. crassa* mã hóa 3-dehydroshikimat dehydrataza có thể thu được bằng phương pháp bất kỳ trong một vài phương pháp đã biết, ví dụ tổng hợp toàn bộ gen ADN, tách dòng cADN hoặc bằng cách kết hợp phương pháp tách dòng ADN của hệ gen và phương pháp PCR hoặc tổng hợp nhóm liên kết ADN tổng hợp. Vì không có intron ở gen *qa-4*, vùng mã hóa có thể thu được bằng phương pháp PCR từ ADN hệ gen (Rutledge, 1984). Trình tự protein của enzym *qa-4* (SEQ ID No. 4) và ADN trình tự của gen nguyên thể (SEQ ID No. 5) là đã biết.

Theo cách khác, catxet biểu hiện có thể được thiết kế cho 3-dehydroshikimat dehydrataza của *A. nidulans*. Trình tự mã hóa cho enzym QutC của *A. nidulans* có thể thu được bằng phương pháp bất kỳ trong số một vài phương pháp đã biết, ví dụ tổng hợp ADN toàn bộ gen, tách dòng cADN hoặc bằng cách kết hợp phương pháp tách dòng ADN hệ gen và phương pháp PCR hoặc tổng hợp nhóm liên kết ADN tổng hợp. Trình tự protein của QutC (SEQ ID No. 6) và trình tự ADN của gen nguyên thể, không có intron, là đã biết (SEQ ID No. 7; số truy cập GenBank là M77665,1). Catxet biểu hiện có thể thu được bằng cách tổng hợp ADN, hoặc bằng cách kết hợp phương pháp tách dòng bộ gen và PCR, để enzym QutC có thể được tạo ra một cách chính xác trong *E. coli*. Bằng cách biểu hiện trình tự mã hóa cho QutC từ trình tự khởi đầu cơ định mạnh ở *E. coli*, sự biểu hiện đủ có thể thu được từ một hoặc hai bản sao của gen được gắn trong nhiễm sắc thể, tránh cần phải duy trì nhiều hơn hai bản sao của catxet biểu hiện trên plasmit nhiều bản sao như được nêu trong giải pháp đã biết (Niu et al., 2002), và có thể dẫn đến sự mất ổn định. Phương pháp được nêu trên đây có thể thường được dùng để tạo ra trình tự ADN mã hóa

enzym mong muốn, và trình tự mã hóa này sau đó được sử dụng để xây dựng catxet biểu hiện đã được thiết kế cho chức năng ở *E. coli* hoặc vi khuẩn chủ thích hợp khác.

Hoạt tính đặc hiệu của AroZ cũng có thể được cải thiện bằng cách sử dụng trình tự protein từ các giải pháp đã biết (Niu et al., 2002) nhưng việc tạo ra catxet biểu hiện được cải thiện, ví dụ, trong đó trình tự khởi đầu và/hoặc vị trí gắn ribosom (RBS) mạnh hơn (RBS) có thể được cài ở trước vùng mã hóa, như được mô tả trong ví dụ 4.

Gen *aroZ* mã hóa AroZ (3-dehydroshikimat dehydrataza) của chủng *Klebsiella pneumoniae* A170-40 có thể thu được như được mô tả trong giải pháp đã biết. Trình tự ADN của gen và ADN bao quanh có thể được xác định bằng phương pháp đã biết trong lĩnh vực này. Gen khác loài theo sáng chế như *aroZ* có thể được gắn vào catxet biểu hiện bằng cách sử dụng trình tự ADN nguyên thể hoặc có thể được tổng hợp bằng trình tự tối ưu hóa codon cho sinh vật chủ dự định. Gen *aroZ* có thể được tách dòng như được mô tả (Draths and Frost, 1995) từ vi khuẩn bất kỳ khác mà có chứa gen *aroZ* hoạt tính, ví dụ, chủng *K. pneumoniae* 342, *Acinetobacter Sp. ADP1* (*Acinetobacter baylyi* ADP1), *Bacillus thuringiensis*, *Emecricella nidulans*, *Erwinia amylovora*, *Pseudomonas putida* W619, và nhiều chủng khác.

Enzym protocatechuat decarboxylaza (AroY; EC 4.1.1.63) là cần cho quy trình sinh tổng hợp sản phẩm trung gian catechol. Trong bản mô tả này, “AroY” chỉ enzym bất kỳ xúc tác phản ứng protocatechuat decarboxylaza. Trong các giải pháp đã biết, enzym này được biểu hiện từ gen *aroY* của chủng *Klebsiella pneumoniae* A170-40 (ATCC25597) trên plasmit nhiều bản sao (Niu et al., 2002). Tuy nhiên, một lần nữa có thể đạt được sự cải thiện trong quy trình này bằng cách tạo ra enzym này với lượng đủ từ một hoặc hai bản sao của catxet biểu hiện được gắn trong nhiễm sắc thể của sinh vật chủ. Việc này có thể thực hiện được bằng cách tạo ra gen *aroY* từ sinh vật tổng hợp enzym AroY một cách tự nhiên có hoạt tính đặc hiệu cao hơn hoạt tính đặc hiệu của enzym AroY từ *K. pneumoniae* từ các giải pháp kỹ thuật đã biết hoặc bằng cách gia tăng mức biểu hiện của *K. pneumoniae* AroY bằng cách tạo ra catxet biểu hiện, ví dụ, sử dụng trình tự khởi đầu cơ định mạnh và/hoặc RBS

mạnh như được nêu trong ví dụ 4. Trình tự protein đối với AroY từ chủng *K. pneumoniae* A170-40 được đưa ra trong SEQ ID No. 8. Gen tương ứng, *aroY*, có thể được tách dòng như nêu trên (Draths and Frost, 1995), hoặc dựa trên trình tự protein này, có thể được tổng hợp bằng codon tối ưu hóa đối với sinh vật chủ dự định.

Gen *aroY* có thể thu được từ vi sinh vật bất kỳ khác mà chứa thể tương đồng hoặc thể tương tự, ví dụ, chủng *K. pneumoniae* NCTC418 (ATCC15380), *Klebsiella pneumoniae* 342, và *Arxula adeninivorans* (Sietmann et al, 2010). Trình tự ADN của gen *aroY* của *Klebsiella pneumoniae* 342 và ADN xung quanh được nêu trong SEQ ID No. 9.

Enzym catechol 1,2-dioxyaza (CatA; EC 1.13.11.1) là cần thiết đối với bước cuối cùng trong quy trình sinh tổng hợp axit cis, cis-muconic. Trong bản mô tả này, “CatA” chỉ enzym bất kỳ xúc tác phản ứng catechol 1,2-dioxyaza. Trong các giải pháp đã biết, enzym này được biểu hiện từ gen *catA* của chủng *Acinetobacter calcoaceticus* ADP1 trên plasmit nhiều bản sao (Niu et al., 2002). Chủng nguồn, chủng *Acinetobacter calcoaceticus* ADP1, được đặt tên lại là *Acinetobacter Sp.* ADP1 và *Acinetobacter baylyi* ADP1 (Neidle và Ornston, 1986; Barbe et al, 2004; de Berardinis et al, 2008). Trong ví dụ về giải pháp đã biết, gen *catA* được biểu hiện từ trình tự khởi đầu P_{tac}, đòi hỏi lactoza hoặc IPTG (isopropylthiogalactosit) làm tác nhân cảm ứng. Các hợp chất này quá đắt để sử dụng trong quy trình lên men ở quy mô thương mại, vì vậy cần có sự cải thiện đáng kể trong quy trình này, để vừa không cần đến tác nhân gây cảm ứng đắt tiền vừa tạo ra chủng ổn định hơn bằng cách gắn catxet biểu hiện vào nhiễm sắc thể. Việc này có thể thực hiện được bằng cách xây dựng catxet biểu hiện đối với gen *catA* sử dụng trình tự khởi đầu cơ định mạnh, RBS mạnh và/hoặc ARN thông tin ổn định hơn như được nêu trên trong các Ví dụ khác.

Trình tự ADN của gen *catA* và trình tự xung quanh của *Acinetobacter baylyi* ADP1 được đưa ra trong SEQ ID No. 10. Trình tự protein của CatA từ cùng chủng được đưa ra trong SEQ ID No. 11. Theo phương án được ưu tiên, catxet biểu hiện *catA* chứa một hoặc hai khung đọc mở khác tồn tại tự nhiên xuôi dòng của *catA*, để

làm gia tăng mức biểu hiện của gen *cata* (Schirmer và Hillen, 1998). Nhiều sinh vật khác có thể là nguồn cung cấp gen *cata*, ví dụ *Pseudomonas arvilla*, *Pseudomonas fluorescens* (Nakazawa et al, 1967; Kojima et al, 1967), *Streptomyces* Sp. Strain 2065 (Iwagami et al, 2000), *Cupriavidus necator* 335T, và nhiều chủng khác (Perez-Pantoja et al, 2008).

Để cải thiện dòng cacbon hướng đến axit cis, cis-muconic, cần ngăn chặn một số con đường khác phân nhánh từ con đường axit amin thơm, ngoài việc làm giảm dòng cacbon từ DHS đến shikimat (SHK) bằng cách sử dụng thể đột biến rò *aroE*. Một số vi khuẩn, ví dụ ở giống *Acinetobacter* và *Pseudomonas*, chứa gen *pobA*, mã hóa enzym, p-hydroxybenzoat hydroxylaza, mà chuyển hóa DHS thành axit galic. Mặc dù thể tương đồng hoặc thể tương tự của PobA không được tìm thấy ở chủng *E. coli*, chủng *E. coli* được thiết kế để tạo ra DHS tiết ra axit galic với lượng xác định được (Li và Frost, 1999), như là enzym này tồn tại ở *E. coli*. Ngoài ra, PCA thu được từ DHS có thể được chuyển hoá thành axit galic nhờ tác dụng của enzym p-hydroxybenzoat hydroxylaza (PobA) được mã hóa bởi gen *pobA*. Sau đó, axit galic đã tạo ra có thể được chuyển hóa thành pyrogallol. Một cách để chặn dòng cacbon chuyển hóa thành axit galic và pyrogallol trong chất xúc tác sinh học được chọn đối với axit cis, cis-muconic được cải thiện là ức chế hoặc làm giảm hoạt tính của protein p-hydrobenzoat hydroxylaza (PobA) thông qua các thao tác di truyền. Một cách tương tự, DHQ, tiền chất để tạo ra DHS cũng có thể tác dụng dựa vào enzym shikimat dehydroaza được mã hóa bởi *aroE* để tổng hợp axit quinic. Theo một phương án của sáng chế, thể đột biến rò AroE của enzym này được chọn lọc hoặc sàng lọc thêm dựa trên tính không có khả năng hoặc khả năng chuyển hóa DHQ thành axit quinic giảm.

Có một số ưu điểm trong việc tổng hợp axit trans, trans- muconic thay cho axit cis, cis-muconic. Axit trans, trans-muconic được ưu tiên hơn axit cis, cis-muconic trong phản ứng Diels Alder với etylen để tạo ra axit terephthalic. Chất xúc tác sinh học cùng với con đường thơm được thao tác di truyền tạo ra axit cis, cis-muconic mà có thể được chuyển hoá thành axit trans, trans-muconic bên ngoài tế bào bằng cách sử dụng quy trình chuyển hoá hóa học. Mặt khác, bằng cách đưa enzym maleylaxetoaxetat isomeraza hoặc enzym isomeraza tương tự vào chất xúc tác sinh

học, có thể chuyển hóa axit cis, cis-muconic thành axit trans, trans- muconic trong chất xúc tác sinh học từ vi khuẩn.

Sáng chế đề xuất một số khía cạnh khác nhau của sáng chế liên quan đến việc xây dựng chủng vi khuẩn tổng hợp axit muconic một cách hiệu quả. Người có hiểu biết trung bình trong lĩnh vực này có thể cải biến các khía cạnh khác nhau của sáng chế để tạo ra chất xúc tác sinh học có tính hiệu quả cao trong việc tổng hợp axit muconic.

Ví dụ thực hiện sáng chế

Phần thử nghiệm

Chuẩn bị chủng và nguyên liệu cấy: Danh sách các chủng vi khuẩn và plasmit được sử dụng trong sáng chế được đưa ra trong bảng 1. Tất cả các Ví dụ cụ thể về cấu trúc của chủng được mô tả ở đây thu được từ chủng *E. coli* C kiểu đại (ATCC 8739), hoặc chủng *E. coli* K-12 (YMC9 hoặc MM294) nhưng các yếu tố di truyền được mô tả ở đây có thể được tổ hợp ở chủng *E. coli* thích hợp và catxet biểu hiện hoặc thể tương tự và thể tương đồng thích hợp của các yếu tố gen được mô tả ở đây có thể được lắp ráp vào vi sinh vật thích hợp bất kỳ khác, như loài vi khuẩn, khuẩn cổ, nấm men, tảo, và nấm sợi khác có thể được sử dụng để tổng hợp trên quy mô thương mại axit cis, cis-muconic nhờ quy trình lên men.

Chủng *E. coli* C có khả năng lên men glucoza 10% trong môi trường chất khoáng AM1. Môi trường AM1 chứa 2,63g/L $(\text{NH}_4)_2\text{HPO}_4$, 0,87g/L $\text{NH}_4\text{H}_2\text{PO}_4$, 1,5mM MgSO_4 , 1mM betain, và 1,5ml/l thành phần vi lượng. Thành phần vi lượng được điều chế ở dạng dung dịch gốc 1000 lần và chứa các thành phần sau: 1,6g/l FeCl_3 , 0,2g/l $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0,1g/l CuCl_2 , 0,2g/l $\text{ZnCl}_2 \cdot 4\text{H}_2\text{O}$, 0,2g/l NaMoO_4 , 0,05g/l H_3BO_3 , và 0,33g/l $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$. Độ pH của canh trường lên men được duy trì ở 7,0 bằng dung dịch KOH 1 – 10M hoặc dung dịch amoni hydroxit 1 – 9M.

Lên men:

Quá trình lên men được bắt đầu bằng việc cấy thành vạch lên đĩa NBS-2% glucoza tươi (Jantama et al., 2008a) từ dung dịch gốc glyxerol 40% chứa chủng *E. coli* được thiết kế di truyền và được bảo quản trong tủ đá ở nhiệt độ -80°C. Plasmit,

nếu có, được giữ bằng cách đưa (các) chất kháng sinh thích hợp vào đĩa thạch và môi trường lỏng. Ampixilin (muối natri) được sử dụng với hàm lượng 150mg/l, spectinomycin HCL với hàm lượng 100mg/l, tetracyclin HCl với hàm lượng 15mg/l, và kanamycin sulfat với hàm lượng 50mg/l. Sau 24 đến 48 giờ (ở nhiệt độ 37°C), một khuẩn lạc được lấy vào bình lắc chứa 25ml cùng môi trường. Sau khi lắc ở tốc độ 200 vòng/phút ở nhiệt độ 37°C cho đến khi các tế bào sinh trưởng đến OD₆₀₀ bằng khoảng 1,0, môi trường nuôi cấy được làm lạnh bằng đá và dung dịch glycerol 80% vô trùng với thể tích tương đương được thêm vào. Sau đó, 2ml phân uớc được làm lạnh đông ở nhiệt độ -80°C để sử dụng làm nguyên liệu cấy để lên men.

Sự phát triển của tế bào:

Sinh khối tế bào được đánh giá bằng cách đo mật độ quang ở bước sóng 550nm (OD₅₅₀) hoặc 600nm (OD₆₀₀) bằng cách sử dụng quang phổ kế Thermo Electronic Spectronic 20.

Phân tích các sản phẩm trung gian trong con đường axit shikimic và con đường axit muconic:

Tổng lượng axit muconic được tạo ra trong canh trường lên men, gồm axit cis, cis-muconic và axit cis, trans-muconic, và các sản phẩm hóa sinh trung gian khác được thử nghiệm bằng phương pháp sắc ký lỏng cao áp (HPLC) bằng thiết bị Waters Alliance và đo độ hấp thụ ở bước sóng 210nm hoặc chỉ số khúc xạ ở nhiệt độ 45° C, sử dụng chất chuẩn mua từ Sigma-Aldrich. Cột là BioRad Aminx HPX-87H chạy ở nhiệt độ 50°C sử dụng dung dịch axit sulfuric 8mM làm pha động ở tốc độ dòng là 0,6ml/phút trong 40 phút. Phổ sắc ký của chất chuẩn đã mua (Sigma-Aldrich) được thể hiện trên Fig.3. Để chuẩn bị cho HPLC, các mẫu lên men được pha loãng 10 hoặc 100 lần trong dung dịch đệm kali phosphat 0,05M, độ pH= 7,0, bảo toàn dạng cis, cis- của axit muconic khỏi quá trình đồng phân hóa thành dạng cis, trans.

Để tách các dạng đồng phân của axit muconic, mẫu được điều chế như trên được chạy trong hệ HPLC thứ hai. Thiết bị này là Agilent 1200 HPLC, cột là Agilent Eclipse XDB-C18, 4,6 x 150mm chạy ở nhiệt độ 30°C với pha động là dung

dịch 50mM KH_2PO_4 trong 30% metanol được điều chỉnh đến độ pH= 3 bằng dung dịch axit phosphoric. Tốc độ dòng là 1ml/phút trong 4 phút, với việc phát hiện bằng độ hấp thụ ở 278nm. Axit cis, trans-axit muconic chuẩn được tạo ra bằng cách hoà tan axit cis, cis-muconic trong nước và cho tham gia phản ứng đồng phân hóa tự phát được xúc tác bằng axit trong 2 giờ ở nhiệt độ trong phòng, cho đến khi đỉnh HPLC hoàn toàn dịch chuyển đến vị trí mới. Các chất chuẩn khác được mua từ Sigma-Aldrich. Phổ sắc ký thể hiện các chất chuẩn được nêu trên Fig.4.

Thành phần của môi trường tổng hợp axit muconic dùng cho quy trình lên men:

Mỗi lit môi trường lên men chứa 50ml/l 1M KH_2PO_4 , 10ml dung dịch axit xitric 200g/L + 25g/l Sắt (III) xitrat, 1,2ml dung dịch axit sulfuric 98% , và một giọt Antifoam 204. Các thành phần được trộn với nước với lượng đủ để bổ sung các thành phần khác dưới đây. Sau khi hấp, thành phần sau được thêm vào : 10, 20, 30 hoặc 40ml dung dịch glucoza 50% (để thu được nồng độ cuối là 5, 10, 15, hoặc 20g/l), 2ml dung dịch MgSO_4 1M, 1ml dung dịch CaCl_2 0,1M , 10ml dung dịch nguyên tố vết 1000X (Jantama et al. 2008a), 1, 2, 4, hoặc 8ml Phenylalanin 50g/L + Tyrosin 50g/l + Tryptophan 50g/l (để thu được nồng độ cuối là 0,5, 0,1, 0,2, hoặc 0,4g/l), 10ml dung dịch axit p-hydroxybenzoic 1g/l + axit p-aminobenzoic 1g/l + axit 2,3-dihydroxylbenzoic 1g/l, và nếu cần, 1ml dung dịch Ampixilin 150mg/ml (muối natri) và/hoặc 1ml Spectinomycin HCl 100 mg/ml.

Đối với quy trình lên men gián đoạn, chai nạp chứa glucoza khan 600g/l và 32ml/l dung dịch Phenylalanin 50g/l + Tycolophan 50g/l + Tryptophan 50g/l. Dung dịch NH_4OH 9M được sử dụng làm bazơ để duy trì độ pH của môi trường lên men.

Đối với bình lắc, muối NBS (Jantama et al. 2008a) cùng với dung dịch đệm MOPS 0,2M, độ pH= 7,4 được thay cho hỗn hợp được hấp sơ bộ nêu trên, nhưng glucoza và các chất phụ gia khác là giống nhau.

Quy trình lên men gián đoạn được thực hiện trong bình lên men chuyên dụng New Brunswick dung tích 7l với độ pH, nồng độ oxy hòa tan, nhiệt độ, glucoza, và tốc độ nạp được theo dõi bằng thiết bị theo dõi DCU hoặc phần mềm BiocommADN Software. Nhiệt độ được duy trì ở 37°C, độ pH được duy trì bằng 7 bằng dung dịch nước amoni 9N và nồng độ oxy hòa tan (dissolved oxy - DO) được

duy trì ở mức bão hòa khí 30% trong khi tốc độ của máy khuấy tuabin gia tăng từ 750vòng/phút đến 1200vòng/phút. Hàm lượng glucoza ban đầu trong môi trường mong muốn là khoảng từ 5 đến 25g/l. Bổ sung dung dịch glucoza vào bình lên men khi hàm lượng glucoza giảm xuống dưới 5g/l, và tốc độ nạp glucoza được kiểm soát bằng nồng độ oxy hòa tan. Tổng thời gian lên men là 48 giờ và hiệu giá cuối là 16g/l axit muconic.

Việc tạo ra plasmit biểu hiện các gen của con đường axit muconic: Ba gen khác loài cần cho sự chuyển hoá DHS thành axit muconic được tách dòng riêng biệt hoặc kết hợp thành plasmit có ít bản sao, pCL1921 (Lerner và Inouye, 1990). Trình tự ADN của pCL1921 được đưa ra trong SEQ ID No. 20 trong bảng 3. Một cách vắn tắt, trình tự mã hóa của thể tương tự hoặc thể tương đồng của *catAX*, *aroY* và *aroZ* được tối ưu hóa bằng codon để biểu hiện ở *E. coli* và được tổng hợp trên quy mô thương mại (GenArt, Invitrogen). Sau đó, các trình tự này được khuếch đại bằng kỹ thuật PCR sử dụng đoạn môi xuôi mang vị trí gắn ribosom duy nhất và đoạn môi ngược mang trình tự gen kết thúc duy nhất đối với mỗi gen. Mảnh PCR thu được được phân huỷ bằng enzym giới hạn và được tách dòng xuôi dòng của trình tự trình tự khởi đầu cơ định duy nhất bằng quy trình tách dòng phân tử chuẩn. Trình tự của trình tự khởi đầu được tách dòng bằng kỹ thuật khuếch đại PCR từ trình tự ADN nguồn nêu trên (Đơn yêu cầu cấp patent Mỹ số 20090191610; Patent Mỹ số 7244593), sau đó là phân cắt giới hạn và tách dòng phân tử chuẩn. Trình tự trình tự khởi đầu-RBS-trình tự mã hóa-gen kết thúc được kết hợp với một catxet biểu hiện. Sau đó, các catxet biểu hiện riêng lẻ được kết hợp để tạo ra plasmit biểu hiện một, hai hoặc cả ba gen của con đường axit muconic.

Ví dụ 1

Gia tăng biểu hiện của gen *aroG* và *aroF*.

Gen *tyrR* của *E. coli* có thể được gây đột biến bằng phương pháp bất kỳ trong số các phương pháp đã biết như phương pháp gây đột biến hóa học hoặc phóng xạ và sàng lọc (ví dụ bằng PCR và xác định trình tự ADN) hoặc chọn lọc tính kháng với thể tương tự (ví dụ, kháng 4-flotycolophan), gây đột biến hạt chuyển, gây đột biến thể thực khuẩn Mu, hoặc biến nạp. Theo phương án được ưu tiên, sự đột biến ở

gen *tyrR* là đột biến triệt tiêu (sự đột biến làm hoạt tính không thể xác định được), và theo phương án ưu tiên hơn, ít nhất một phần gen *tyrR* bị khuyết. Có thể thực hiện việc này, ví dụ, bằng cách sử dụng phương pháp biến nạp hai bước sử dụng phân tử ADN mạch thẳng (Jantama et al, 2008a; Jantama et al, 2008b). Trong bước thứ nhất, catxet *cam^R*, *sacB* được gắn ở locus *tyrR* để thế phần lớn hoặc tất cả khung đọc mở *tyrR* bằng tái tổ hợp kép và chọn lọc tính kháng chloramphenicol. Trong bước thứ hai, ADN mạch thẳng chứa thể khuyết đoạn của gen *tyrR* được gắn bằng tái tổ hợp kép, chọn lọc tính kháng với sucroza 5% trong môi trường giàu như LB. Đoạn khuyết chính xác được nhận diện và xác nhận bằng phản ứng chuỗi polymeaza chuẩn đoán (PCR). Mục đích của việc làm khuyết đoạn *tyrR* là làm gia tăng biểu hiện của *aroG* và *aroF*. Một cách tiếp cận khác để thu được kết quả tương tự là thay thế trình tự khởi đầu nguyên thể ở trước *aroG* và/hoặc *aroF* bằng trình tự khởi đầu cơ định mạnh và bổ sung, nếu cần, gen kết thúc phiên mã. Chi tiết về phương pháp này được đưa ra trong Ví dụ 4 dưới đây.

Phương pháp thứ hai của hai cách tiếp cận nêu trên để giải quyết sự kìm hãm hoạt tính của enzym AroG và AroF bằng protein TyrR là được ưu tiên, do việc làm khuyết đoạn *tyrR* có thể gây ra sự biểu hiện quá mức không mong muốn của các gen như *aroLM* (Neidhardt và Curtiss, 1996). Chi tiết về việc này được đưa ra trong Ví dụ 4 dưới đây.

Ví dụ 2

AroG và AroF kháng phản hồi

Các đột biến ở gen *aroG* tạo ra enzym AroG kháng phản hồi (3-deoxy-D-arabinoheptulosonat-7-phosphat synthaza hoặc DAHPS) là đã biết trong lĩnh vực này (Shumilin et al, 1999; Kikuchi et al, 1997; Shumilin et al, 2002). Phương pháp tạo ra, nhận diện và mô tả các đột biến này cũng đã được biết đến (Ger et al., 1994, Hu et al., 2003). Đột biến được ưu tiên là đột biến dẫn đến tính kháng hoàn toàn với sự ức chế bởi phenylalanin. Đột biến kháng phản hồi đã công bố bất kỳ có thể được đưa vào gen *aroG* có trong nhiễm sắc thể hoặc trên plasmit bằng phương pháp bất kỳ trong số các phương pháp đã biết, một ví dụ về phương pháp này là phương pháp PCR gây đột biến trong đó đột biến mong muốn được tổng hợp là một

phần của oligonucleotit mỗi PCR (Hu et al., 2003). Việc thiết lập chính xác đột biến được xác nhận bằng việc xác định trình tự ADN. Trình tự của gen *aroG* kiểu dại của *E. coli* C được đưa ra trong SEQ ID No. 18. Đột biến được ưu tiên là đột biến điểm làm thay đổi axit amin 150 của AroG từ prolin thành loxin, ví dụ bằng cách thay đổi codon 150 từ CCA thành CTA (Hu et al., 2003). Theo phương án được ưu tiên hơn, codon 150 được thay đổi từ CCA thành CTG, là codon được ưu tiên trong *E. coli*. Alen cụ thể này của *aroG* là được ưu tiên, vì DAHP synthaza mã hóa kháng hoàn toàn với sự ức chế bởi phenylalanin tối đa là 3mM, và có hoạt tính đặc hiệu tương tự với enzym kiểu dại (Hu et al., 2003).

Alen *aroG* kháng phản hồi khác có thể thu được bằng cách gây đột biến và chọn lọc tính kháng với một hoặc nhiều chất tương tự phenylalanin, như beta-2-thienylalanin, p-flophenylalanin, p-clophenylalanin, o-flophenylalanin, và o-clophenylalanin, sau đó chứng minh rằng đột biến gây ra tính kháng này gắn kết với gen *aroG* (Ger et al., 1994; Patent Hoa Kỳ 4,681,852). Sự gắn kết với gen *aroG* có thể được chứng minh trực tiếp bằng cách xác định trình tự ADN hoặc thử nghiệm enzym với sự có mặt và vắng mặt của phenylalanin, (Ger et al., 1994) hoặc gián tiếp bằng sự tải nạp qua trung gian thể thực khuẩn và chọn lọc gen đánh dấu di truyền ở hoặc ở gần vị trí *aroG* mà có thể được chọn để (Patent Hoa Kỳ 4,681,852). Gen đánh dấu di truyền này có thể là thể khuyết đoạn hoặc đột biến điểm của chính gen *aroG* hoặc đột biến ở gen có quan hệ gần thích hợp như gen *nadA* trong trường hợp *E. coli*. Ví dụ ở *E. coli*, sau khi gây đột biến và chọn lọc tính kháng với chất tương tự phenylalanin, các thể đột biến riêng biệt hoặc nhóm thể đột có thể được dùng làm thể cho để tải nạp qua trung gian P1 thành thể nhận nguyên thể bị khuyết đoạn cả ba gen DAHP synthaza, *aroG*, *aroF*, và *aroH*, và chọn lọc đối với sự phát triển trên môi trường tối thiểu thích hợp. Sau đó, thể tải nạp được làm giàu cho các đột biến ở (các) gen mong muốn. Theo một cách khác, sau khi gây đột biến và chọn lọc đối với tính kháng chất tương tự, các thể đột biến riêng biệt hoặc nhóm thể đột có thể được dùng làm thể cho để tải nạp qua trung gian P1 thành chủng nhận nguyên thể mà có chứa đột biến triệt tiêu ở gen *nadA*, chọn lọc lần nữa đối với sự phát triển trên môi trường tối thiểu thích hợp thiếu nicotinamit. Cách tiếp cận khác là chọn thể đột biến kháng ở chủng mà có chứa hạt chuyển, ví dụ Tn10, chèn gần gen *aroG*, như

trong gen *nadA*. Tải nạp P1 từ thể đột biến kháng chất tương tự thành nền chủng không chứa hạt chuyển này và chọn lọc đối với tính kháng kháng sinh tetracyclin hoặc chất kháng sinh thích hợp khác sẽ được làm giàu cho các đột biến *aroG* mong muốn. Trong tất cả các cách tiếp cận này, tính kháng phản hồi được xác nhận sau cùng bằng thử nghiệm enzym và xác định trình tự ADN của gen. *aroG** được dùng để chỉ alen của *aroG* kháng ức chế phản hồi.

Chủng WM191 ($\Delta tyrR$, $\Delta aroF$) thu nhận từ YMC9 (ATCC 33927). Phương pháp thể gen gồm hai bước (Jantama et al., 2008a) được sử dụng để thiết lập đoạn khuyết sạch ở cả hai gen *tyrR* và *aroF*, để thu được chủng WM191. Sau đó, alen *nadA::Tn10* được tải nạp ở CAG12147 (CGSC 7351, Coli Gentic Stock Center, Yale University) để thu được chủng WM189 ($\Delta tyrR$, $\Delta aroF$, *nadA::Tn10*). Chọn lọc trên môi trường Luria (Luria-Bertani: LB) được bổ sung tetracyclin HCl (15mg/l). Chủng RY890 ($\Delta tyrR::kan$, *aroF363*) thu nhận từ MM294 (ATCC 33625) ở ba bước bằng cách tải nạp P1. Chủng cho, lần lượt là JW1316-1 (CGSC 9179, Coli Gentic Stock Center, Yale University), NK6024 (CGSC 6178, Coli Gentic Stock Center, Yale University), và AB3257 (CGSC 3257, Coli Gentic Stock Center, Yale University), và ba bước chọn lọc trên môi trường lần lượt là LB bổ sung kanamycin sulfat (50mg/l), LB bổ sung tetracyclin hydroclorua (15mg/l), và môi trường NBS chứa glucoza tối thiểu (Jantama et al., 2008a) với thiamin HCl (5mg/l).

WM189 được gây đột biến bằng tia cực tím đến tỷ lệ sống khoảng 20% và dàn trên môi trường NBS chứa glucoza tối thiểu (Jantama et al., 2008a) chứa o-flophenylalanin (1mM), thiamin (5mg/l), và nicotinamit (1mM). Các khuẩn lạc từ mỗi đĩa được gom thành các nhóm riêng, và dịch tan P1vir được điều chế trên mỗi nhóm. Các dịch tan này được sử dụng để tải nạp cho WM191 tính kháng tetracyclin (15mg/l) trên môi trường LB và các khuẩn lạc được cấy bằng con dấu vào môi trường NBS chứa glucoza tối thiểu chứa o-flophenylalanin 1mM, thiamin (5mg/l), và nicotinamit (1mM). Các con dấu khuẩn lạc sống được trong môi trường chứa tetracyclin và chất tương tự được cho là có đột biến kháng phản hồi ở gen *aroG*. Tám khuẩn lạc riêng biệt từ 5 nhóm độc lập được chọn để xác định trình tự ADN. Vùng mã hóa gen *aroG* được khuếch đại bằng phản ứng chuỗi polymeaza và xác

định trình tự. Kết quả, được thể hiện trên bảng 4, cho thấy rằng mỗi chủng trong số tám chủng này chứa đột biến điểm ở gen *aroG* của chúng. Một số alen được nhận diện là giống với alen đã công bố, còn một số alen là mới.

Dịch tan P1vir của một trong số các nhóm nêu trên được sử dụng để tải nạp cho RY890 (có alen kiểu dại của *aroG*) tính kháng tetracyclin và tính kháng o-flophenylalanin (0,3mM) bằng cách cấy bằng con đầu nêu trên. Bốn khuẩn lạc, RY893, RY897, RY899, và RY901, được thu nhặt để xác định trình tự ADN (bảng 4), và một lần nữa, hai trong số các alen này là giống với alen đã công bố, còn hai alen còn lại là mới. Chủng RY902, cùng gen với bốn chủng sau, nhưng chứa gen *aroG* kiểu dại được tạo ra làm đối chứng, bằng cách tải nạp từ CAG12147. Năm chủng này phát triển qua đêm trong bình lắc trong 25ml môi trường chứa glucoza tối thiểu (15g/l) được bổ sung thiamin HCl (5mg/l) và nicotinamit (1mM). Các tế bào thu được được thu hoạch bằng cách ly tâm, tái tạo huyền phù để súc rửa bằng 10ml nước, ly tâm lại và tái tạo huyền phù trong dung dịch kali phosphat 50mM với lượng 0,5ml, độ pH= 7,0. Các tế bào huyền phù được phân giải bằng cách lắc xoáy cùng với ba giọt cloform, và dịch tan thô được thử nghiệm đối với hoạt tính DAHP synthaza sử dụng phương pháp tương tự với phương pháp được mô tả trong tài liệu (Hu et al., 2003), với các cải biến sau. Nồng độ dung dịch đệm phosphat là 50mM (nồng độ cuối cùng), độ pH =7,0, hàm lượng erythroza-4-phosphat cuối là 2mM, hàm lượng phosphoenol pyruvat cuối là 5mM, nhiệt độ ủ là 30°C, và dừng phản ứng sau 10 phút. 1mU được định nghĩa là hoạt tính mà tổng hợp ra 1nM DAHP/phút/miligram protein. Để kiểm tra tính kháng phản hồi, mỗi dịch tan thô được thử nghiệm cùng với hoặc không cùng với phenylalanin ở nồng độ cuối là 18mM. Kết quả của thử nghiệm được chỉ ra trong bảng 5. Enzym thể hiện hoạt tính đặc hiệu và tính kháng phenylalanine thay đổi, nhưng tất cả các thể đột biến đã chọn được thử nghiệm đều kháng nhiều hơn đáng kể so với đối chứng kiểu dại.

Các alen *aroG* của RY893, RY899, RY901, và RY902, được nêu trên đây, được đưa vào nền chủng tổng hợp axit muconic như sau. Dịch tan của chủng cho P1vir *aroG** và *aroGwt* được sử dụng để tải nạp cho MYR219 (*E. coli* C, Δ *aroE*, Δ *ack::P₁₅-aroB*, pMG37) tính kháng tetracyclin HCl (15mg/l), để lần lượt thu được

các chủng mới RY903, RY909, RY911, và RY912. Sau đó, mỗi chủng này được tải nạp tính kháng kanamycin sulfat (50mg/l) bằng cách sử dụng dịch tan P1vir của JW1316-1, để đưa alen $\Delta tyrR::kan$ vào, để lần lượt thu được các chủng RY913, RY919, RY921, và RY922. Tính chọn lọc spectinomycin được duy trì từ đầu đến cuối để duy trì plasmit muconic. Bốn chủng thu được được phát triển trong 48 giờ ở nhiệt độ 37°C trong bình lắc trong 25ml môi trường NBS tối thiểu (Jantama et al., 2008a) được bổ sung 20g/l glucoza, dung dịch đệm MOPS 0,2M, độ pH= 7,4, nicotimamit (1mM), phenylalanin (100mg/l), tycolophan (100mg/l), tryptophan (100mg/l), axit p-hydroxybenzoic (1mg/l), axit p-aminobenzoic (1mg/l), axit 2,3-di-hydroxybenzoic (1mg/l), phenol đỏ (10mg/l), và amoni sulfat (1g/l). Độ pH được duy trì xung quanh 7 như được xác định bằng mắt dựa vào màu của phenol đỏ, dựa vào độ pH =7,0 chuẩn, bằng cách bổ sung bằng tay các phân ước 1ml dung dịch KOH 1M như đối với bình lắc. Axit muconic thu được được thử nghiệm bằng phương pháp sắc ký lỏng cao áp (HPLC) như được nêu ở trên và kết quả được thể hiện trong bảng 6. Cả ba chủng mà có alen *aroG** có tính kháng phản hồi tạo ra axit muconic nhiều hơn chủng cùng gen alen *aroG* kiểu dại. Trong thử nghiệm riêng biệt được mô tả ở đây, chủng MYR205, chứa *aroG* trên plasmit nhiều bản sao pCP32AMP, tạo ra 1,5g/l axit muconic trong bình lắc. Do vậy, tác giả sáng chế chứng minh rằng việc kết hợp $\Delta tyrR$ và *aroG** sắc thể của một bản sao có thể tổng hợp axit muconic trong bình lắc tốt hơn so với chủng chứa plasmit *aroG* cùng gen. Độ ổn định di truyền của alen sắc thể cao hơn alen plasmit, cùng với làm giảm nhu cầu đối với môi trường chọn lọc để duy trì trong plasmit, tạo ra chủng mới được mô tả ở đây thích hợp hơn để lên men thương mại trên quy mô lớn. Ngoài ra, tác nhân gây cảm ứng hóa học là cần cho biểu hiện của các gen trong con đường axit muconic. Do đó, chủng theo sáng chế cải thiện hơn so với các chủng trong các giải pháp đã biết (Niu et al., 2002), tất cả các chủng này đều chứa gen biểu hiện quá mức DAHP synthaza trên plasmit nhiều bản sao không mong muốn.

Theo cách thức tương tự như được mô tả đối với AroG, đột biến làm cho isozym AroF hoặc AroH kháng ức chế phản hồi bởi tycolophan có thể được tạo ra trên plasmit hoặc trong nhiễm sắc thể. Đột biến được ưu tiên là đột biến điểm thay đổi axit amin 148 của AroF từ prolin thành loxin, ví dụ bằng cách thay đổi codon

148 từ CCG thành CTG (Weaver et al., 1990), để thu được gen *aroF**. Alen khác của *aroF** có thể được phân lập dựa vào tính kháng với chất tương tự tyrosin (ví dụ, o-flotyrosin, m-flotyrosin, p-flophenylalanin, v.v.) theo cách thức tương tự như đối với alen *aroG**. Alen *aroF** có thể được chọn, làm giàu và tải nạp liên kết với hạt chuyển hoặc chèn tính kháng kanamycin, ví dụ ở $\Delta yfiR::kan$ được liên kết một cách chặt chẽ như ở chủng JW2584 (CGSC 10051, Coli Gentic Stock Center, Yale University).

Ví dụ 3

Xóa *aroE* ra khỏi ADN nhiễm sắc thể và tổng hợp axit muconic

Trong ví dụ này, ảnh hưởng của sự biểu hiện quá mức *aroB* và *aroG* trên plasmit nhiều bản sao cũng như biểu hiện của các gen mã hóa các protein có vai trò trong con đường axit muconic được nghiên cứu. Chủng MYR34 chứa đoạn khuyết ở gen *aroE* mã hóa shikimat dehydroaza được sử dụng làm chủng gốc trong nghiên cứu này. Việc xóa bản sao nhiễm sắc thể của *aroE* được thực hiện theo cách thức tương tự với cách thức được nêu trong Ví dụ 1. Khi MYR34 được biến đổi bằng plasmit pCP32AMP quá biểu hiện gen *aroG* mã hóa protein DAHP synthaza có vai trò trong con đường axit shikimic, sự tích lũy DHS gia tăng đáng kể. Khi MYR34 được biến đổi bằng plasmit biểu hiện *aroB* từ trình tự khởi đầu cơ định, nhận thấy không có sự gia tăng đáng kể tích lũy DHS. Tuy nhiên, khi chủng *E. coli* MYR34 được biến đổi bằng plasmit biểu hiện cả gen *aroB* lẫn *aroG*, sự tích lũy DHS gia tăng nhiều hơn so với sự gia tăng quan sát được với chủng MYR34 được biến đổi bằng một mình gen *aroG* chứng tỏ rằng *aroB* là một trở ngại thứ yếu trong quy trình tổng hợp DHS (Fig.5).

Trong các thử nghiệm được thể hiện trong Fig.6, ảnh hưởng của bản sao khác của gen *aroB* được gắn vào ADN nhiễm sắc thể chủ được nghiên cứu. ở chủng *E. coli* MYR170 thu nhận từ MYR34, một bản sao khác của gen *aroB* được gắn vào nhiễm sắc thể chủ tại vị trí *ack* dưới sự kiểm soát của trình tự khởi đầu P₁₅. Khi chủng MYR170 được biến đổi bằng plasmit pCP32AMP, sự tích lũy DHS tăng nhẹ so với sự tích lũy DHS được xác định ở chủng MYR34 được biến đổi bằng cùng plasmit. Sự gia tăng nhẹ tích lũy DHS ở chủng MYR170 có thể là do bản sao khác

của gen *aroB* được gắn vào ADN nhiễm sắc thể chủ. Khi MYR170 được biến đổi bằng pCP54 biểu hiện cả gen *aroB* lẫn *aroG*, có sự gia tăng thêm tích lũy DHS chứng tỏ rằng *aroB* là trở ngại thứ yếu trong quy trình tổng hợp DHS.

Fig.7 thể hiện các kết quả tổng hợp axit muconic bằng chủng *E. coli* MYR34 và MYR170. Được chứng minh rằng ở chủng MYR34 và MYR170 khuyết đoạn *aroE*, với sự biểu hiện quá mức của gen *aroB* và *aroG*, có sự tích lũy DHS, các nỗ lực được thực hiện để xem liệu sự biểu hiện của các gen “con đường muconic” mã hóa các protein có vai trò trong con đường tổng hợp axit muconic có dẫn đến sự chuyển hóa DHS thành axit cis, cis-muconic hay không. Trong các thử nghiệm này, chủng *E. coli* MYR34 và MYR170 được biến đổi bằng một mình plasmit pMG37 hoặc bằng cả hai plasmit pMG37 và pCP32AMP. Plasmit pMG37 biểu hiện các gen *aroZ*, *aroY* và *catAX* mã hóa protein có vai trò trong con đường axit muconic. Mức tổng hợp axit muconic ở cả hai chủng MYR34 và MYR170 gia tăng khi các chủng vi khuẩn này được biến đổi bằng cả hai plasmit pCP32AMP và pMG37 so với mức tổng hợp axit muconic ở hai chủng chỉ được biến đổi bằng plasmit pMG37 cho thấy rằng ở các chủng này, sự biểu hiện *aroB* là trở ngại trong việc tổng hợp axit cis, cis-muconic.

Ví dụ 4

Biểu hiện quá mức của *tktA*

Transketolaza được mã hóa bởi *tktA* là enzym then chốt trong con đường pentoza phosphat và được cho là giới hạn đối với việc tổng hợp erythroza-4-phosphat, một trong số các sản phẩm trung gian quan trọng trong quy trình tổng hợp axit muconic. Biểu hiện quá mức của *tktA*, mã hóa transketolaza, bằng cách cài gen này bằng trình tự khởi đầu nguyên thể của nó trên plasmit nhiều bản sao (Sprenger et al, 1995, 1995a), là đã biết để cải thiện dòng đi vào con đường thơm (Draths et al., 1992). Tuy nhiên, các plasmit này không ổn định và thường đòi hỏi chất kháng sinh chọn lọc để duy trì. Cách tiếp cận khác trong các giải pháp đã biết là bổ sung một bản sao bổ sung của gen *tktA* vào nhiễm sắc thể của chủng chủ (Niu et al., 2002). Tuy nhiên, một bản sao bổ sung của *tktA* có trình tự khởi đầu nguyên thể của nó là không đủ để thực hiện con đường thơm với erythroza-4-phosphat, vì trình tự

khởi đầu nguyên thể của nó là không giống với mô hình lý tưởng. Như vậy, quy trình này cần cải thiện đáng kể.

Có thể cải thiện sự biểu hiện quá mức của *tktA* bằng cách thể trình tự khởi đầu *tktA* nguyên thể trong nhiễm sắc thể bằng trình tự khởi đầu cơ định mạnh, ví dụ, trình tự khởi đầu P₁₅ hoặc P₂₆ của thể thực khuẩn *Bacillus subtilis* SPO1 (lần lượt là SEQ ID No. 1 và SEQ ID No. 2), hoặc trình tự khởi đầu P_R của thể thực khuẩn lambda (SEQ ID No. 3). Việc này được thực hiện ở hai bước như được mô tả trong Ví dụ 1, ngoại trừ catxet *cam^R*, *sacB* được sử dụng thay cho trình tự khởi đầu *tktA* nguyên thể trong nhiễm sắc thể ở bước thứ nhất. Trong bước thứ hai, trình tự khởi đầu cơ định mạnh được cài bằng cách biến đổi ADN mạch thẳng chứa trình tự khởi đầu cơ định mạnh, nằm bên sườn là ít nhất 50 bazơ có đầu tận cùng 5' của vùng mã hóa *tktA* trên sườn xuôi dòng và ít nhất 50 cặp bazơ của thể tương đồng ngược dòng của trình tự khởi đầu *tktA* nguyên thể trên sườn ngược dòng của trình tự khởi đầu cơ định mạnh, và chọn lọc đối với tính kháng sucroza. Cải thiện biểu hiện của catxet biểu hiện này cũng được thực hiện bằng cách gia tăng độ ổn định của ARN thông tin sao chép từ catxet biểu hiện này. Việc cải thiện độ ổn định của ARN thông tin được thực hiện bằng cách bổ sung cấu trúc vòng thân ở đầu tận cùng 5' của ARN thông tin, đầu tận cùng 3' của ARN thông tin, hoặc cả hai vị trí trên. Cấu trúc vòng thân thường có mặt ở đầu tận cùng của ARN thông tin được kết thúc một cách tự nhiên bằng gen kết thúc phiên mã độc lập rho, nhưng nếu không phải là gen này, thì gen kết thúc phiên mã độc lập rho có thể được thêm vào trình tự ADN bằng phương pháp thao tác di truyền đã biết (phương pháp nối, PCR, v.v.). Gen kết thúc này có thể gồm đoạn lặp được đảo ngược chứa từ 4 đến 20 bazơ trên mỗi đoạn lặp, được phân tách bằng "vòng" chứa 3 bazơ hoặc nhiều hơn, và tiếp đó là vùng chứa một hoặc nhiều bazơ được làm giàu bằng T (deoxythymidin). Đoạn lặp ngược giàu G's và C's (deoxyguanidin và deoxycytidin). Tương tự, vòng thân có thể được đưa vào đầu tận cùng 5' của ARN thông tin bằng cách chèn trình tự ADN ngay sau điểm khởi đầu của sự phiên mã, nhưng trước vị trí gắn kết ribosom, mà có chứa vòng thân nêu trên, nhưng không chứa vùng giàu T. Ví dụ về việc này được đưa ra trong phần liên quan đến trình tự khởi đầu P₁₅ (SEQ ID No. 1).

Trong các phân tích về ảnh hưởng của sự biểu hiện quá mức của gen *tktA* đối

với dòng cacbon qua con đường axit shikimic, chủng *E. coli* MYR170 được sử dụng làm chủng gốc. MYR170 có đoạn khuyết ở gen *aroE* mã hóa enzym shikimat dehydroaza và một bản sao bổ sung của gen *aroB* ở locus *ack*.

Trong các thử nghiệm được mô tả trong các Fig.8, 9 và 10, hai plasmit khác nhau là pCP32AMP và pCP50 được sử dụng. Plasmit pCP32AMP chỉ biểu hiện gen *aroG* mã hóa DAHP synthaza từ trình tự khởi đầu nguyên thể của nó và plasmit pCP50 biểu hiện gen *tktA* mã hóa transketolaza từ trình tự khởi đầu nguyên thể của nó cùng với gen *aroG*. MYR170, có đoạn khuyết *aroE* và một bản sao bổ sung của gen *aroB* dưới sự kiểm soát của trình tự khởi đầu P₁₅ được gắn ở locus *ack* của ADN nhiễm sắc thể, được biến đổi riêng biệt bằng plasmit pCP32AMP và pCP50. Như được thể hiện trên Fig.8, sự tích lũy DHS gia tăng với biểu hiện của gen *aroG* cùng với gen *tktA* khi so sánh với tế bào *E. coli* chỉ biểu hiện gen *aroG*.

Fig.9 cung cấp số liệu về hiệu suất DHS ở hai chủng khác nhau là MYR34, MYR170 được biến đổi bằng plasmit pCP32AMP hoặc pCP50. Chủng MYR34 có đoạn khuyết ở gen *aroE* tạo ra 0,1g DHS/g glucoza sử dụng. Hiệu suất DHS ở chủng MYR34 tăng đến 0,15g DHS/g glucoza sử dụng khi chủng này được biến đổi bằng plasmit pCP32AMP biểu hiện quá mức gen *aroG*. MYR170 có bản sao bổ sung của gen *aroB* được chèn vào ở locus *ack*. Kết quả của sự có mặt của bản sao bổ sung của gen *aroB* này là hiệu suất tổng hợp DHS ở chủng MYR170 được biến đổi bằng pCP32AMP cao hơn một chút so với hiệu suất tổng hợp DHS ở chủng MYR34 được biến đổi bằng pCP32AMP. Do đó, sự có mặt của bản sao bổ sung của gen *aroB* ở chủng MYR170 làm dòng cacbon qua con đường axit shikimic tăng. Sự gia tăng hiệu suất DHS cũng được quan sát khi MYR170 được biến đổi bằng plasmit pCP50 biểu hiện cả hai gen *aroG* và *tktA*. Như vậy, sự có mặt của bản sao bổ sung của gen *tktA* được cho là làm gia tăng dòng cacbon qua con đường axit shikimic. Cụ thể hơn, ảnh hưởng của sự có mặt của các gen *aroB* và *tktA* bổ sung gây ra ảnh hưởng khác đối với hiệu suất DHS.

MYR261 được sử dụng trong các thử nghiệm được mô tả trong Fig.10 được thiết kế để gắn bản sao bổ sung của gen *tktA* vào ADN nhiễm sắc thể của MYR170 ở locus *poxB*. Việc thể gen mong muốn (*poxB::tktA*) ở chủng MYR261 được xác

nhận bằng phương pháp PCR. MYR261 được biến đổi bằng plasmit pCP32AMP (quá biểu hiện *aroG*) hoặc plasmit pCP50 (quá biểu hiện *aroG* và *tktA*). MYR170 được biến đổi bằng plasmit pCP32AMP được sử dụng làm đối chứng. Kết quả được thể hiện trên Fig.10 cho thấy, sự có mặt của bản sao bổ sung của gen *tktA* ở ADN nhiễm sắc thể của MYR261 làm gia tăng hiệu giá tổng hợp DHS bằng plasmit pCP32AMP khi so sánh với hiệu giá tổng hợp DHS được quan sát ở chủng MYR170 được biến đổi bằng cùng plasmit. Sự gia tăng mức transketolaza ở chủng MYR261 khi được biến đổi bằng plasmit pCP50 quá biểu hiện transketolaza làm gia tăng hiệu giá tổng hợp DHS. Enzym mã hóa bởi *poxB*, PoxB, hoặc pyruvat oxidaza, tạo ra axetat là sản phẩm của phản ứng. Như vậy, sự khuyết đoạn *poxB* thu được từ việc chèn *tktA* như được mô tả ở đây làm mất đi con đường có khả năng tạo ra axetat. Tương tự, việc chèn đồng thời *P_{15aroB}* và làm mất đoạn *ackA*, mã hóa AckA, hoặc axetat kinaza, như được mô tả trong ví dụ dưới đây, làm mất đi con đường có khả năng tạo ra axetat. Trong quy trình lên men, thường không mong muốn tạo ra axetat (Jantama et al., 2008b). Như vậy, các khuyết đoạn này là hữu ích trong việc làm giảm sự tích lũy axit axetic.

Fig.11 cung cấp hiệu giá đối với việc tổng hợp axit muconic và axit axetic ở chủng *E.coli* MYR170, MYR261 và MYR305 sau khi biến đổi bằng plasmit pCP32AMP và pMG37. MYR305 thu nhận được từ MYR170 bằng cách xóa gen *poxB* ra khỏi ADN nhiễm sắc thể còn MYR261 thu được từ MYR170 trong đó gen *poxB* được bất hoạt bằng cách chèn bản sao bổ sung của gen *tktA* vào. Như nêu trên, plasmit pCP32AMP biểu hiện gen *aroG* mã hóa protein DAHP synthaza có vai trò trong con đường sinh tổng hợp axit shikimic dẫn đến sự tích lũy DHS do khuyết đoạn gen *aroE* ở chủng *E. coli* MYR170, MYR261 và MYR305. Với biểu hiện của các gen trong con đường muconic là *aroZ*, *aroY* và *catAX* trên plasmit pMG37, DHS được chuyển hóa thành axit cis, cis-muconic như được minh họa trên Fig.2. Với sự có mặt của bản sao bổ sung của *aroB* gen và *tktA* ở chủng MYR261, việc tổng hợp axit muconic gia tăng nhẹ cùng với sự tích lũy axit axetic giảm.

Ví dụ 5

Sự biểu hiện quá mức của *talA* hoặc *talB*

Gen *talB* mã hóa transaldolaza chiếm ưu thế ở chủng *E. coli*, còn gen *talA* mã hóa transaldolaza phụ. Tổng hợp quá mức transaldolaza là đã biết để cải thiện dòng đi vào con đường thơm (Lu và Liao, 1997; Sprenger, 1995; Sprenger et al, 1995b). Trong giải pháp đã biết, việc này được thực hiện bằng cách biểu hiện quá mức gen *tal* (hiện nay được biết đến là gen *talB*) trên plasmid nhiều bản sao từ trình tự khởi đầu nguyên thể của nó (Lu et al., 1997, Sprenger et al., 1995b). Tuy nhiên, plasmid này không ổn định và cần chọn lọc chất kháng sinh để duy trì. Do đó, cần cải thiện quy trình. Biểu hiện của *talB* có thể được cải thiện, ví dụ, bằng cách thể trình tự khởi đầu *talB* nguyên thể trong nhiễm sắc thể bằng trình tự khởi đầu cơ định mạnh, ví dụ, trình tự khởi đầu P₁₅ hoặc P₂₆ của thực khuẩn thể *Bacillus subtilis* SPO1 (lần lượt là SEQ ID No. 1 và SEQ ID No. 2), hoặc trình tự khởi đầu P_R của thể thực khuẩn lambda (SEQ ID No. 3). Điều này được tiến hành theo hai bước như được mô tả trong Ví dụ 1, ngoại trừ catxet *cam*^R, *sacB* được sử dụng thay cho trình tự khởi đầu *talB* nguyên thể trong nhiễm sắc thể ở bước thứ nhất. Trong bước thứ hai, trình tự khởi đầu cơ định mạnh được cài đặt bằng cách biến đổi bằng ADN mạch thẳng chứa trình tự khởi đầu cơ định mạnh, ở bên sườn ít nhất là 50 bazơ ở đầu tận cùng 5' của vùng mã hóa *talB* ở phía dưới và ít nhất 50 cặp bazơ của thể tương đồng ngay trên trình tự khởi đầu *talB* nguyên thể ở phía trên trình tự khởi đầu cơ định mạnh, và chọn lọc tính kháng sucroza. Gen *talA* cũng có thể được biểu hiện quá mức bằng phương pháp tương tự nhưng sự biểu hiện quá mức của gen *talB* là được ưu tiên, vì nó mã hóa hoạt tính chiếm ưu thế (Sprenger, 1995; Sprenger et al, 1995b). Xem Ví dụ 4 về chi tiết việc tạo ra catxet biểu hiện được thiết kế để biểu hiện quá mức.

Ví dụ 6

Biểu hiện của các gen *aroZ*, *aroY* và *catAX*

Để thực hiện chuyển hoá DHS nội sinh được tạo ra bởi *E. coli* thành axit muconic, các gen khác loài *catAX* của *Acinetobacter sp. ADP1*, *aroY* của *Klebsiella pneumoniae*, và *quiC* của *Acinetobacter sp. ADP1*, được tách dòng dưới sự kiểm soát của trình tự khởi đầu cơ định mạnh (lần lượt là P₁₅, P_R, và P₂₆) trong plasmid có ít bản sao, pCL1921 (Lerner và Inouye, 1990) tạo ra 'plasmid muconic' pMG37.

Phát triển dẫn xuất của chủng MYR34 mang vật truyền trống (pCL1921) hoặc pMG37 ở nhiệt độ 37°C trong 17 giờ trong môi trường bình lắc (môi trường tối thiểu NBS được bổ sung axit amin thơm và vitamin) chứa 2% glucoza. Dịch nổi được gom và phân tích bằng HPLC. Khác với MYR34/pCL1921 tích lũy DHS, MYR34/pMG37 tổng hợp axit muconic (Fig.12). Ở chủng thứ hai, hàm lượng DHS hoặc sản phẩm trung gian như PCA và catechol không đáng kể, gợi ý rằng các gen khác loài được biểu hiện từ pMG37 là có chức năng và là đủ.

Ví dụ 7

So sánh các thể tương đồng *aroZ*

Ba thể tương đồng và chất tương tự *aroZ* khác nhau được so sánh (Fig.13) đối với hoạt tính của chủng làm DHS hướng trạch sang con đường tổng hợp axit muconic. *quiC* của *Acinetobacter* sp. ADP1, *asbF* của *Bacillus thuringiensis*, và *qa-4* của *Neurospora crassa*, được báo cáo là mã hóa cho protein có hoạt tính giống AroZ (Elsemore và Ornston, 1995; Fox et al, 1995; Rutledge, 1984). Mỗi gen này được tối ưu hóa bằng codon để biểu hiện ở *E. coli* và được tổng hợp bằng GeneArt (Invitrogen), và được tách dòng dưới sự kiểm soát của trình tự khởi đầu cơ định mạnh P₂₆ trong 'plasmid muconic' số có ít bản sao cũng biểu hiện các gen *catAX* và *aroY* từ các trình tự khởi đầu P₁₅ và P_R. MYR34/pCL1921, MYR34/pMG37 (plasmid muconic có *quiC* làm *aroZ*), MYR34/pMG47 (plasmid muconic có *asbF* làm *aroZ*), và MYR34/pMG70 (plasmid muconic có *asbF* làm *aroZ*) được phát triển ở nhiệt độ 37°C trong 48 giờ trong bình lắc chứa môi trường tối thiểu chứa 2% glucoza, axit amin thơm và vitamin thơm. Dịch nổi được gom và phân tích bằng HPLC. Như mong đợi, vật truyền trống MYR34 đã biến đổi tích lũy DHS và không tổng hợp axit muconic. Hai thể tương đồng *aroZ* và một thể tương tự đã được thử nghiệm có chức năng trong việc làm chệch DHS vào con đường tổng hợp axit muconic, ở các mức khác nhau. Dẫn xuất MYR34 biểu hiện gen *quiC* là mạnh nhất và thể hiện chuyển hoá gần 100% DHS thành axit muconic với lượng DHS còn lại không đáng kể. Dẫn xuất MYR34 biểu hiện thể tương đồng *aroZ* của nấm, *qa-4*, đứng sau với khoảng 80% DHS chuyển hoá thành axit muconic và còn lại 20% DHS. Cuối cùng, dẫn xuất MYR34 biểu hiện gen *asbF* thể hiện chỉ có 50% DHS

chuyển hóa thành axit muconic và còn lại 50% DHS. Trong các điều kiện thử nghiệm trong bình lắc, biểu hiện và/hoặc hoạt tính của gen *quiC* tỏ ra là cao nhất so với các thể tương đồng của *aroZ*.

Ví dụ 8

Gắn *catAX*, *aroY* và *quiC* vào nhiễm sắc thể

Axit muconic có thể được tạo ra bởi chủng mà chỉ chứa một bản sao của *catAX*, *aroY* và *quiC* được gắn vào nhiễm sắc thể biểu hiện từ trình tự khởi đầu cơ định mạnh ở locus *adhE*.

MYR170 ($\Delta aroE$, $\Delta ack::P_{15-aroB}$), chủng tổng hợp DHS cao, là chủng chủ được sử dụng để gắn các gen trong con đường axit muconic vào nhiễm sắc thể ở locus *adhE* (SEQ ID No. 41). Chủng thu được MYR352 được biến đổi bằng plasmit YEp24 (có số bản sao trung bình, vật truyền trống), pCP32AMP (có số bản sao trung bình, *aroG* được biểu hiện từ trình tự khởi đầu nguyên thể), hoặc pCP50 (có số bản sao trung bình, *aroG* và *tklA* được biểu hiện từ trình tự khởi đầu nguyên thể tương ứng của chủng) để tạo ra chủng dẫn xuất. Hai plasmit sau được sử dụng để làm gia tăng sự tổng hợp DHS. Các chủng được sinh trưởng ở nhiệt độ 37°C trong 72 giờ trong môi trường bình lắc chứa 2% glucoza như được nêu trên. Dịch nổi được gom ở thời điểm 72 giờ và phân tích bằng HPLC. Như mong đợi, dẫn xuất MYR352 đã biến nạp *aroG* và *aroG/tklA* thể hiện sự gia tăng chung ở tổng sản phẩm tạo thành so với đối chứng vật truyền trống (Fig.14). Tất cả các thể biến đổi của MYR352 tạo ra hiệu giá axit muconic có thể xác định được, chứng minh rằng lần đầu tiên axit muconic có thể được tạo ra bởi chủng chỉ chứa gen trong “con đường muconic” đã được gắn kết và không cần tác nhân gây cảm ứng hóa học để biểu hiện gen.

Không phải tất cả DHS được tạo ra ở chủng dẫn xuất bất kỳ trong số các chủng dẫn xuất MYR352 này được chuyển hóa thành sản phẩm cuối là axit muconic. Thay vào đó, có một lượng đáng kể catechol tích lũy (Fig.14), cho thấy rằng biểu hiện hoặc hoạt tính của *catAX* là một trở ngại khi nó được biểu hiện từ một bản sao đơn trên nhiễm sắc thể. Vì sản phẩm trung gian tích lũy chủ yếu là catechol, nên có thể sự biểu hiện và/hoặc hoạt tính của gen *quiC* và *aroY* là đủ ở

chủng gốc MYR352 để tổng hợp axit muconic.

Chủng dẫn xuất MYR352 được so sánh song song với dẫn xuất chủng MYR219 tương tự. Chủng MYR219 là giống như chủng MYR170 nhưng chứa plasmit có ít bản sao pMG37 biểu hiện các gen của con đường axit muconic. Do đó, sự khác nhau chính giữa chủng MYR352 và MYR219 liên quan đến việc sử dụng các gen của con đường axit muconic (1 bản sao so với lần lượt 5 bản sao). Khác với dẫn xuất chủng MYR352, dẫn xuất chủng MYR219 tích lũy rất ít catechol hoặc sản phẩm trung gian khác và tổng hợp thành công sản phẩm axit muconic cuối. Đồng thời, các kết quả này chỉ ra cần phải gia tăng hoạt tính *catAX* ở các chủng như MYR352.

Ví dụ 9

Biểu hiện của *catAX*

Sự tích lũy catechol và việc tổng hợp axit muconic không hiệu quả ở chủng MYR352 là do hạn chế liều lượng và/hoặc hoạt tính của (các) sản phẩm gen *catAX*. Như được nêu trên, MYR352 chứa $\Delta aroE$, $\Delta ack::P_{15-aroB}$ và các bản sao đơn được gắn vào nhiễm sắc thể của các gen *catAX*, *aroY* và *quiC* dưới sự kiểm soát của trình tự khởi đầu cơ định mạnh. Chủng này được biến đổi bằng đối chứng vật truyền trống có số bản sao trung bình (YEp24) hoặc plasmit biểu hiện *aroG/tktA* (pCP50) làm gia tăng dòng cacbon vào con đường tổng hợp axit amin thơm và tạo ra một lượng lớn DHS. Phát triển chủng đã biến đổi ở nhiệt độ 37°C trong 72 giờ trong môi trường của bình lắc được bổ sung 2% glucoza như được nêu trên dẫn đến sự tích lũy sản phẩm trung gian catechol. Kết quả này gợi ý rằng hoạt tính của *catAX* có thể không đủ ở chủng MYR352. Để xác nhận giả thuyết này, khả năng làm giảm sự tích lũy catechol ở MYR352/pCP50 của một hoặc nhiều gen trong con đường axit muconic được biểu hiện từ plasmit có ít bản sao được thử nghiệm (Fig.15). Cụ thể, MYR352/pCP50 được biến đổi thêm bằng đối chứng vật truyền trống có ít bản sao (pCL1921) hoặc plasmit biểu hiện cả ba các gen, hai gen, hoặc một gen của con đường tổng hợp axit muconic. Chủng dẫn xuất này được thử nghiệm trong thử nghiệm bình lắc như nêu trên. Trong khi liều lượng của một mình *aroY* (của pMG27) hoặc một mình *quiC* (của pMG39) không làm giảm sự tích lũy catechol,

biểu hiện của tất cả các gen của con đường axit muconic (của pMG37) hoặc *catAX* cùng với *aroY* (của pMG33), dẫn đến sự chuyển hoá thành công catechol thành axit muconic. Ngoài ra, biểu hiện của một mình *catAX* (của pMG31) là đủ để tổng hợp axit muconic và ngăn chặn sự tích lũy catechol.

Ví dụ 10

Tạo ra đột biến rò của *aroE*

Trong giải pháp đã biết về quy trình tổng hợp axit cis, cis-muconic, chủng chủ chứa đột biến ở gen *aroE* là *aroE353*, là đột biến triệt tiêu. Kết quả là, chủng này cần sự cung cấp các axit amin thơm (phenylalanin, tycolophan, và tryptophan) và vitamin thơm được tạo ra từ từ con đường shikimat (axit p-hydroxy benzoic, axit p-amino benzoic và axit 2,3-di-hydroxy benzoic). Axit amin thơm quá đắt để được cấp vào quy trình trên quy mô thương mại lớn. Do vậy, các giải pháp đã biết cần cải thiện đáng kể. Việc này có thể được thực hiện bằng cách tạo ra thể đột biến rò của gen *aroE*, mà chúng ta gọi là *aroE**. Đột biến rò được tạo ra bằng cách tạo ra đột biến nhằm nghĩa thay đổi một axit amin trong trình tự mã hóa của *aroE* thu được kiểu hình triệt tiêu. Có thể thực hiện việc này bằng cách gây đột biến ở dạng bất kỳ hoặc sàng lọc sinh vật dinh dưỡng-sinh trưởng đồng thời đối với sáu các hợp chất thơm nêu trên. Phương pháp được ưu tiên là tạo ra nhóm gen *aroE* đột biến bằng cách gây đột biến bằng PCR dễ mắc lỗi, sử dụng *Taq* ADN polymeaza, sử dụng ADN thuộc hệ gen C của *E. coli* kiểu đại làm khuôn mẫu và sử dụng đoạn mỗi oligonucleotit PCR lai hóa khoảng 1000 cặp bazơ ngược dòng và 1000 cặp bazơ xuôi dòng của vùng mã hóa *aroE*. Nhóm phân tử ADN mạch thẳng thu được được sử dụng để biến đổi dẫn xuất của chủng *E. coli* C mà tổng hợp axit cis, cis-muconic, và chứa catxet *cam^R*, *sacB* đã được gắn thay cho vùng mã hóa *aroE* (xem Ví dụ 4), và chọn lọc tính kháng sucroza. Sau đó, các thể biến đổi được sàng lọc đối với sinh vật dinh dưỡng-sinh trưởng không còn tính kháng chloramphenicol và cần sáu các hợp chất thơm nêu trên. Một số sinh vật dinh dưỡng sinh trưởng độc lập được thu nhặt và thử nghiệm đối với tính hồi biến bằng cách dàn mỏng khoảng 10^7 , 10^8 , hoặc 10^9 tế bào (được rửa trong môi trường glucoza tối thiểu) trên đĩa glucoza tối thiểu không chứa sáu các hợp chất thơm. Thể hồi biến phát triển thành khuẩn lạc trên các

đã được thu nhận và thử nghiệm đối với việc tổng hợp axit cis, cis-muconic, nhưng không tổng hợp axit amin thơm với lượng đáng kể. Trong số các thể hội biến này sẽ là chủng mang một hoặc nhiều các đột biến ở gen *aroE*, để enzym AroE cung cấp đủ lượng axit amin thơm và vitamin để phát triển, nhưng không dư thừa các hợp chất thơm này. Phương pháp khác để thu được thể đột biến rò *aroE* là cài đặt một trong số các thể đột biến cổ điển *aroE* có thể hội biến được như *aroE353* và *aroE24* (cả hai có sẵn từ Coli Gentic Stock Center at Yale University, New Haven, CT, USA), thành chủng tổng hợp axit cis, cis-muconic và chọn lọc các thể hội biến như nêu trên.

Ví dụ 11

Hấp thụ glucoza bởi hệ khuếch tán qua trung gian

Một cơ chất của bước thứ nhất trong con đường thơm là phosphoenolpyruvat (PEP). PEP cũng là nguồn phosphat và năng lượng để hấp thụ glucoza và một số đường khác bởi hệ phosphotransferaza của vi khuẩn (PTS). Do vậy, khi vi khuẩn phát triển trên đường phụ thuộc PTS, có sự cạnh tranh giữa PTS và con đường thơm đối với PEP. Như vậy, sự cải thiện đáng kể làm gia tăng dòng vào con đường thơm có thể đạt được bằng cách làm mất đoạn PTS và tạo ra con đường khác để hấp thụ đường. Một giải pháp cho vấn đề này là thay thế PTS bằng GalP permeaza của *E. coli*, thể đồng vận chuyển làm việc khá tốt để hấp thụ glucoza (Patent Hoa Kỳ 6,692,794). Tuy nhiên, thể đồng vận chuyển proton vẫn sử dụng năng lượng để duy trì gradien proton cần thiết để điều khiển permeaza. Do đó, vẫn cần cải thiện thêm quy trình này.

Một số đường, như xyloza, có thể được hấp thụ bằng protein vận chuyển nhận năng lượng từ quá trình thủy phân ATP (adenosin triphosphat). Một lần nữa, nếu thể vận chuyển phụ thuộc vào năng lượng có thể được thế bằng thể vận chuyển đòi hỏi ít năng lượng, sau đó thực hiện việc cải thiện, vì năng lượng sẵn có trong ATP có thể được bảo toàn để dùng cho mục đích có lợi khác.

Có thể thu được sự cải thiện đáng kể bằng cách sử dụng thể vận chuyển khuếch tán qua trung gian, không sử dụng năng lượng để hấp thụ đường (Parker et al, 1995; Snoep et al, 1994). Ví dụ, chất trung gian glucoza của *Zymomonas mobilis*,

được mã hóa bởi gen *glf*, có thể được dùng thay cho, hoặc bổ sung cho, PTS ở chủng tổng hợp 3-dehydroshikimat (DHS) (Yi et al., 2003). Tuy nhiên, các chủng này vẫn phụ thuộc ít nhất một phần vào GalP để hấp thu glucoza. Vì GalP đòi hỏi năng lượng dưới dạng gradien proton để hấp thu glucoza, cần phải cải thiện hiệu quả hấp thu glucoza ở chủng tổng hợp muconic.

Catxet biểu hiện gen *glf* và gen glucokinaza, *glk*, also từ *Z. mobilis*, có thể được tổ hợp với trình tự khởi đầu cơ định mạnh, ví dụ P₂₆. Sau đó, catxet này có thể được gắn vào hệ gen của chủng chủ ở vị trí không tham gia vào việc tổng hợp hợp chất mong muốn, trong trường hợp này là axit cis, cis-muconic. Ví dụ về vị trí này trong nhiễm sắc thể *E. coli* nhiễm sắc thể là operon phân hủy threonin, *tdcABCDEFGF*. Nếu môi trường phát triển không chứa threonin, thì không cần tới hoặc biểu hiện operon này, để việc chèn catxet biểu hiện ở operon này không tham gia chuyển hóa.

Để đạt được sự cải thiện nêu trên, một hoặc nhiều của gen mã hóa chức PTS bị xóa, bằng cách sử dụng phương pháp tương tự với phương pháp được nêu trong Ví dụ 1. Ví dụ, một hoặc nhiều đoạn *ptsH*, *ptsI*, *corr*, hoặc *ptsG* bị xóa. Sau đó, *galP* bị xóa. Sau đó, catxet P₂₆-*glf*, *glk* có thể được cài ở hai bước, tương tự như được mô tả trong Ví dụ 1. ở bước thứ nhất, catxet *cam^R*, *sacB* được gắn ở operon *tdc*, sử dụng ADN mạch thẳng thu được từ pAC21 (SEQ ID No. 15), và chọn lọc đối với tính kháng chloramphenicol (30mg/l). ở bước thứ hai, catxet P₂₆-*glf*, *glk* được gắn ở operon *tdc*, sử dụng ADN mạch thẳng thu được từ pAC19 (SEQ ID No. 15), chọn lọc đối với tính kháng sucroza và sàng lọc đối với tính nhạy cảm chloramphenicol và trong trường hợp này, sự phát triển được cải thiện trên môi trường glucoza tối thiểu.

Để kiểm tra xem hệ khuếch tán glucoza qua trung gian có thể thay thế cho hệ hấp thu glucoza thông thường ở *E. coli* hay không, các gen *ptsHI* và gen *galP* bị xóa ra khỏi MYR34 (Δ *aroE*), và sau đó các catxet P₂₆-*glf*, *glk* được gắn ở operon *tdc*, bằng cách sử dụng ADN mạch thẳng thu được từ pAC19 (SEQ ID No. 14), tạo ra chủng MYR217. MYR217 phát triển khá tốt trên môi trường glucoza tối thiểu được bổ sung ba axit amin thơm cần thiết và ba hợp chất giống vitamin thơm (Fig.16).

Tuy nhiên, chủng MYR31, chứa các khuyết đoạn *ptsHI* và *galP*, nhưng không chứa catxet *glf*, *glk* không thể hiện sự phát triển có thể xác định được bất kỳ (Fig.16). Do đó, hệ khuếch tán dễ dàng là đủ để thay thế cho hai hệ hấp thu glucoza thông thường ở chủng nền của chúng tôi.

Để kiểm tra xem hệ khuếch tán qua trung gian có hữu ích để tổng hợp hợp chất thu được từ con đường thơm hay không, MYR34 và MYR217 được biến đổi bằng pCP54 (*aroG*, *aroB*) và pCP55 (*aroG*, *aroB*, *tktA*). Việc tổng hợp sản phẩm trung gian 3-dehydroshikimat (DHS) trong bình lắc của hai chủng này được so sánh (Fig.17). Bằng pCP54 hoặc pCP55, chủng sử dụng hệ khuếch tán qua trung gian tổng hợp DHS bằng hoặc nhiều hơn chủng sử dụng hệ hấp thu glucoza thông thường. Việc tổng hợp DHS là đại diện tốt cho việc tổng hợp axit muconic ở chủng *E. coli* được thiết kế, do đó, có thể kết luận rằng hệ khuếch tán glucoza qua trung gian là một cải thiện hữu ích để tổng hợp axit muconic.

Ví dụ 12

Biểu hiện quá mức của gen *aroB*

Biểu hiện của gen *aroB* được báo cáo là làm hạn chế mức độ tổng hợp axit cis, cis-muconic (Niu et al., 2002). Trong giải pháp kỹ thuật đã biết, việc này được giải quyết bằng cách gắn bản sao thứ hai của gen *aroB* với trình tự khởi đầu nguyên thể của nó. Tuy nhiên, việc này là chưa đủ để làm giảm hạn chế này của *aroB*, vì trình tự khởi đầu nguyên thể và vị trí gắn kết ribosom của gen *aroB* khác xa so với lý tưởng. Do đó, quy trình này cần được cải thiện đáng kể.

Có thể cải thiện sự quá biểu hiện của gen *aroB*, ví dụ, bằng cách thế trình tự khởi đầu nguyên thể của *aroB* trong nhiễm sắc thể bằng trình tự khởi đầu cơ định mạnh, ví dụ, trình tự khởi đầu P₁₅ hoặc P₂₆ của thể thực khuẩn *Bacillus subtilis* SPO1 (lần lượt là SEQ ID No. 1 và SEQ ID No. 2), hoặc trình tự khởi đầu P_R của thể thực khuẩn lambda (SEQ ID No. 3). Việc này có thể thực hiện ở hai bước như được mô tả trong Ví dụ 4, ngoại trừ catxet *cam^R*, *sacB* được sử dụng thay cho trình tự khởi đầu nguyên thể của *aroB* trong nhiễm sắc thể và/hoặc vị trí gắn ribosom ở bước thứ nhất. Ở bước thứ hai, trình tự khởi đầu cơ định mạnh được cài bằng cách biến nạp bằng ADN mạch thẳng chứa trình tự khởi đầu cơ định mạnh, tiếp đó là vị

trí gắn ribosom và ít nhất 50 bazơ từ đầu tận cùng 5' end của trình tự mã hóa *aroB*, gồm codon bắt đầu ATG, về phía xuôi dòng và ít nhất 50 cặp bazơ của thể tương đồng ngay trước trình tự khởi đầu nguyên thể của *aroB* trên trình tự khởi đầu về phía ngược dòng của trình tự khởi đầu cơ định mạnh, và chọn lọc đối với tính kháng sucroza. Ngoài hoặc thay cho, việc cài trình tự khởi đầu mạnh, sử dụng phương pháp tương tự, vị trí gắn kết ribosom mạnh, ví dụ, AGGAGG, có thể được cài khoảng từ 4 đến 10 cặp bazơ ngược dòng của codon bắt đầu dịch mã ATG của *aroB*. Một bản sao của catxet tổng hợp này, ví dụ, catxet *P₁₅-aroB*, có thể được gắn vào nhiễm sắc thể ở vị trí khác với vị trí nguyên thể của *aroB*, ví dụ, ở locus *ack*. Việc làm mất đoạn đồng thời gen *ack*, cũng như mất đoạn gen *poxB* như trong Ví dụ 4 có thể làm giảm sự tạo thành axetat không mong muốn trong quá trình lên men.

Ví dụ 13

Giảm dòng qua nhánh oxy hóa của con đường pentoza phosphat

Erythroza-4-phosphat cần thiết cho bước đầu tiên trong con đường thom thu nhận được từ phần không oxy hóa của con đường pentoza phosphat (PPP). Có hai con đường khác nhau mà bằng con đường này cacbon có thể đi vào PPP. Con đường thứ nhất bắt đầu từ glucoza-6-phosphat nhờ các enzym glucoza-6-phosphat dehydroaza (mã hóa bởi gen *zwf*), 6-phosphogluconolactonaza (mã hóa bởi gen *pgl*), và 6-phosphogluconat dehydroaza (mã hóa bởi gen *gnd*), tạo ra ribuloza-5-phosphat. Trong ba bước cuối này, một cacbon không còn nữa là CO₂. Con đường tạo ra PPP này gọi là nhánh oxy hóa của PPP. Sau đó, ribuloza-5-phosphat được chuyển hóa thành các phosphat của đường khác nhờ tác dụng của isomeraza, epimeraza, transketolaza, và transaldolaza. Nhóm phản ứng thuận nghịch này, bắt đầu từ ribuloza-5-phosphat, gọi là nhánh không oxy hóa của PPP. Con đường thứ hai mà bằng con đường này cacbon có thể đi vào PPP là qua fructoza-6-phosphat và glyxeraldehyt-3-phosphat (cả hai hợp chất này thu được từ con đường Embden-Myerhoff, còn được biết đến là sự thủy phân glucoza), được kết hợp và được sắp xếp lại bằng transaldolaza và transketolaza để thu được các loại phosphat khác nhau của đường khác, một trong số này là erythroza-4-phosphat. Nếu cacbon đi vào PPP

qua con đường thứ hai này, thì CO₂ không bị mất. Để cải thiện hiệu suất tổng hợp axit cis, cis-muconic từ glucoza, việc mất CO₂ có thể được ngăn ngừa bằng cách phong bế nhánh oxy hóa của PPP, để tất cả cacbon đi vào PPP phải qua con đường không oxy hóa từ fructoza-6-phosphat và glyxeraldehyt-3-phosphat. Thực hiện phong bế nhánh oxy hóa của PPP bằng cách làm mất đoạn gen *zwf*, sử dụng phương pháp gồm hai bước tương tự với phương pháp được nêu trong Ví dụ 1 đối với việc làm mất đoạn gen *tyrR*.

Ví dụ 14

Gia tăng dòng vào và qua PEP đến con đường thơm

Mong muốn đảm bảo rằng PEP không là sản phẩm trung gian làm giảm tốc độ trên con đường tổng hợp axit cis, cis-muconic. Việc này được thực hiện, ví dụ, bằng cách gia tăng tuần hoàn pyruvat đến PEP nhờ enzym PEP synthetaza, được thực hiện bằng cách gắn catxet biểu hiện quá mức của gen *pps* như được nêu trong các ví dụ khác. Phương pháp khác là hạn chế tiêu thụ PEP bởi pyruvat kinaza, ở *E. coli* enzym này được mã hóa bởi các gen *pykA* và *pykF*. Trong trường hợp này, phương pháp này làm giảm hoạt tính của (các) enzym. Việc này được thực hiện bằng cách làm mất đoạn một hoặc nhiều các gen mã hóa pyruvat kinaza (như được mô tả trong Ví dụ 1 đối với *tyrR*), hoặc giảm cường độ biểu hiện của một hoặc nhiều gen này, ví dụ, bằng cách gây đột biến trình tự khởi đầu, vị trí gắn ribosom, hoặc trình tự mã hóa, để hoạt tính của pyruvat kinaza giảm. Ví dụ, RBS ở trước gen *pykA* của *E. coli* là 5'CGGAGTATTACATG. Codon khởi đầu dịch mã ATG được gạch dưới. Trình tự này có thể được gây đột biến thành CaGAGTATTACATG, CaaAGTATTACATG, CaatGTATTACATG, CaataTATTACATG, và v.v., để trình tự RBS tạo ra ít giống với RBS liên ứng AGGAGG bằng cách thay một bazơ ở một thời điểm. Sau đó, mỗi thể đột biến được đưa vào nhiễm sắc thể ở locus *pykA*, thể kiểu dại, và mức tổng hợp axit cis, cis-muconic được xác định để cải thiện.

Ví dụ 15

Liên quan đến sự phát triển trên sucroza

Chủng thu được từ *E. coli* C không phát triển trên sucroza làm nguồn cacbon

duy nhất. Tuy nhiên, chúng có thể được được thiết kế di truyền để thực hiện việc này như được nêu trong đơn yêu cầu cấp patent quốc tế số PCT/US11/064598 được đưa vào đây bằng cách viện dẫn. Do vậy, chủng tổng hợp axit cis, cis-muconic có thể được thiết kế để phát triển trên sucroza như được nêu trong đơn nêu trên.

Ví dụ 16

Chủng tổng hợp axit cis, cis-muconic được cải thiện

Tất cả các dấu hiệu được nêu trong Ví dụ 1 – 15 có thể được kết hợp ở một chủng *E. coli* bằng cách tạo dấu hiệu này sau dấu hiệu khác. Chủng thu được gồm chủng tổng hợp axit cis, cis-muconic được cải thiện. Sau đó, chủng có thể được cải thiện thêm bằng cách gắn bản sao thứ hai của mỗi catxet biểu hiện quá mức nêu trên, một ở thời điểm, ở vị trí tách biệt so với vị trí của bản sao thứ nhất. Ví dụ về vị trí thông thường và an toàn là vị trí giới hạn *BsrB1* ngay sau gen kết thúc *rrfF*, mã hóa ARN ribosom. Catxet mong muốn được nối ở dạng ADN mạch thẳng cùn vào vị trí *BsrB1* duy nhất của plasmit pMH17F (SEQ ID No. 17). Một ví dụ là nối catxet biểu hiện *catAX* để thu được plasmit *pcatAX*. Song song, catxet *cam^R*, *sacB* được nối ở dạng mảnh cùn vào pMH17F để thu được pMH28F (SEQ ID No. 19). ADN mạch thẳng thu được từ pMH28 bằng PCR hoặc bằng cách cắt enzym giới hạn được sử dụng để đặt catxet *cam^R*, *sacB* vào vị trí *rrfF*. Tiếp đó, ADN mạch thẳng thu được từ *pcatAX* bằng PCR hoặc bằng cách cắt enzym giới hạn được sử dụng để cài bản sao thứ hai của catxet *catAX* ở locus *rrfF*, sử dụng tính chọn lọc trên sucroza. Sau đó, so sánh chủng thu được với chủng ông bà đối với việc tổng hợp axit cis, cis-muconic để xác định rằng *catAX* là bước giới hạn. Theo phương pháp tương tự, mỗi catxet từ Ví dụ 2 – 15 được thử nghiệm đối với bước giới hạn tốc độ. Nếu một bước được cho là giới hạn tốc độ, một hoặc nhiều bản sao bổ sung của catxet liên quan được gắn ở các vị trí thích hợp khác trong nhiễm sắc thể, làm cải thiện thêm trong quy trình tổng hợp axit cis, cis-muconic, mà không cần plasmit hoặc trình tự khởi đầu gây cảm ứng.

Ví dụ 17

Tổng hợp axit cis, cis-muconic bằng cách lên men

Axit cis, cis-muconic có thể được tạo ra bởi các vi sinh vật được thiết kế di truyền được nêu trong các Ví dụ 1-15 trên. Môi trường phát triển có thể thay đổi trong phạm vi rộng và có thể là môi trường bất kỳ để vi sinh vật phát triển một cách thích hợp. Môi trường được ưu tiên là môi trường tối thiểu chứa các muối khoáng và nguồn cacbon không thơm, như glucoza, xyloza, lactoza, glyxerol, axetat, arabinoza, galactoza, mannoza, maltoza, hoặc sucroza (xem phần ví dụ về môi trường phát triển tối thiểu được ưu tiên). Đối với mỗi tổ hợp gồm vi sinh vật và môi trường phát triển, các điều kiện thích hợp để tổng hợp axit cis, cis-muconic được xác định bằng thử nghiệm thông thường trong đó các thông số lên men thay đổi một cách có hệ thống, như nhiệt độ, độ pH, tốc độ sục khí và hợp chất hoặc các hợp chất được sử dụng để duy trì độ pH. Khi axit cis, cis-muconic được tạo ra, một hoặc nhiều hợp chất phải được nạp vào thiết bị lên men để độ pH không xuống quá thấp. Hợp chất được ưu tiên để trung hòa axit gồm muối kiềm như oxit, hydroxit, cacbonat, và bicacbonat của amoni, natri, kali, canxi, magie, hoặc hỗn hợp chứa hai hoặc nhiều muối kiềm này.

Việc tổng hợp axit muconic bởi chủng *E. coli* MYR428 trong thùng lên men dung tích 7l được thể hiện trên Fig.18. Chủng *E. coli* MYR261 có kiểu gen $\Delta aroE \Delta ackA::P_{15-aroB} \Delta poxB::tklA$ được biến đổi bằng plasmit pCP32AMP và pMG37 tạo ra MYR428. MYR428 được phát triển trong thùng lên men dung tích 7l như nêu trên với việc nạp glucoza trong 48 giờ. Hiệu giá axit muconic cuối là 16g/l (xem Fig.18).

Sau khi quy trình lên men hoàn thành, các tế bào được loại bỏ bằng cách keo tụ, ly tâm và/hoặc lọc và sau đó, axit cis, cis-muconic được tinh chế từ canh trường đã được làm sạch bằng cách kết hợp một hoặc nhiều bước tuần tự, ví dụ, kết tủa, kết tinh, thẩm tách bằng điện, sắc ký (trao đổi ion, kỵ nước ái lực, và/hoặc dựa trên kích thước), vi lọc, lọc màng nano, thẩm thấu ngược và làm bay hơi.

Bảng 1. Chủng vi khuẩn chủng và plasmit được sử dụng theo sáng chế	
Chủng vi khuẩn / Plasmit	Đặc điểm
Chủng vi khuẩn	

ATCC8739	<i>Escherichia coli</i> “C” kiểu dại
MYR34	ATCC8739 $\Delta aroE$
MYR170	ATCC8739 $\Delta aroE$, $\Delta ackA::P_{15aroB}$
MYR261	ATCC8739 $\Delta aroE$, $\Delta ackA::P_{15aroB}$, $\Delta poxB::tktA$
MYR305	ATCC8739 $\Delta aroE$, $\Delta ackA::P_{15aroB}$, $\Delta poxB$
MYR31	ATCC8739 $\Delta ptsHI$, $\Delta galP$
MYR217	ATCC8739 $\Delta ptsHI$, $\Delta galP$, $\Delta tdc::glf-glK$, $\Delta aroE$
MYR352	ATCC8739 $\Delta aroE$, $\Delta ackA::P_{15aroB}$, $\Delta adhE::P_{15-catAX}$, P_R-aroY , $P_{26-quiC}$
RY903	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-893$
RY909	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-899$
RY911	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-901$
RY912	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroGwt$
RY913	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-893$, $\Delta tyrR::kan$
RY919	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-899$, $\Delta tyrR::kan$
RY921	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroG^*20-901$, $\Delta tyrR::kan$
RY922	$\Delta aroE$, $\Delta ackA::P_{15aroB}$, pMG37, $aroGwt$, $\Delta tyrR::kan$
Plasmit	
YE _p 24	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R
pCP32AMP	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R , <i>aroG</i>
pCP14	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R , <i>P_{15aroB}</i>
pCP54	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R , <i>P_{15aroB}</i> , <i>aroG</i>
pCP50	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R , <i>aroG</i> , <i>tktA</i>
pCP55	2 μ nấm men ban đầu, <i>URA3</i> , Tc ^R , replicon pMB1, Ap ^R , <i>aroG</i> , <i>aroB</i> , <i>tktA</i>
pCL1921	replicon pSC101, Spc ^R
pMG27	replicon pSC101, Spc ^R , <i>P_{R-aroY}</i>
pMG31	replicon pSC101, Spc ^R , <i>P_{15-catAX}</i>
pMG33	replicon pSC101, Spc ^R , <i>P_{15-catAX}</i> , <i>P_{R-aroY}</i>

pMG37	replicon pSC101, Spc^R , $P_{15-catA-CatX}$, P_R-aroY , $P_{26-quiC}$
pMG39	replicon pSC101, Spc^R , $P_{26-quiC}$
pMG47	replicon pSC101, Spc^R , $P_{15-catAX}$, P_R-aroY , $P_{26-asbF}$
pMG70	replicon pSC101, Spc^R , $P_{15-catAX}$, P_R-aroY , $P_{26-qa-4}$

Bảng 2 Thông tin trình tự		
Số	Tên	Mô tả
1	SEQ ID No. 1	Trình tự khởi đầu P_{15} của thực khuẩn thể <i>Bacillus subtilis</i> SP01, với thân và vòng được thêm vào ngay sau vị trí khởi đầu phiên mã.
2	SEQ ID No. 2	Trình tự khởi đầu P_{26} của thực khuẩn thể <i>Bacillus subtilis</i> SP01
3	SEQ ID No. 3	Trình tự khởi đầu P_R của thực khuẩn thể <i>Escherichia coli</i>
4	SEQ ID No. 4	Trình tự protein của 3-dehydroshikimat dehydrataza từ <i>Neurospora crassa</i> được mã hóa bởi gen <i>qa-4</i> .
5	SEQ ID No. 5	Trình tự ADN của hệ gen của <i>qa-4</i> gen của <i>Neurospora crassa</i> cùng với trình tự xung quanh.
6	SEQ ID No. 6	Trình tự protein của 3-dehydroshikimat dehydrataza của <i>Aspergillus nidulans</i> được mã hóa bởi gen <i>qutC</i>
7	SEQ ID No. 7	Trình tự ADN của hệ gen của gen <i>qutC</i> của <i>Aspergillus nidulans</i> cùng với trình tự xung quanh
8	SEQ ID No. 8	Trình tự protein của protocatechuat decarboxylaza (AroY) của <i>Klebsiella pneumoniae</i> ATCC25597
9	SEQ ID No. 9	Trình tự ADN của gen <i>aroY</i> của <i>Klebsiella pneumoniae</i> 342 cùng với 2 kilobazơ trình tự ADN xung quanh
10	SEQ ID No. 10	Trình tự ADN của gen <i>catA</i> của <i>Acinetobacter baylyi</i> ADP1, gồm 410 bazơ của trình tự ngược dòng và hai khung đọc mở xuôi dòng
11	SEQ ID No. 11	Trình tự protein của CatA (catechol 1,2-dioxyaza) của <i>Acinetobacter baylyi</i> ADP1
12	SEQ ID No. 12	Trình tự ADN của gen <i>quiC</i> (3-dehydroshikimat dehydrataza) của <i>Acinetobacter sp. ADP1</i>
13	SEQ ID No. 13	Trình tự ADN được tối ưu hóa bởi codon của gen <i>quiC</i> (3-dehydroshikimat dehydrataza) của <i>Acinetobacter sp. ADP1</i>
14	SEQ ID No. 14	Trình tự protein của QuiC (3-dehydroshikimat dehydroaza)

		của <i>Acinetobacter</i> sp. ADP1
15	SEQ ID No. 15	Trình tự ADN của plasmit pAC21
16	SEQ ID No. 16	Trình tự ADN của plasmit pAC19
17	SEQ ID No. 17	Trình tự ADN của plasmit pMH17F
18	SEQ ID No. 18	Trình tự ADN của vùng mã hóa của gen <i>aroG</i> kiểu dại
19	SEQ ID No. 19	Trình tự ADN của plasmit pMH28F

Bảng 3. Thông tin trình tự –tiếp tục		
No.	Tên	Mô tả
20	SEQ ID No. 20	Trình tự ADN của plasmit pCL1921
21	SEQ ID No. 21	Trình tự ADN của plasmit pMG27
22	SEQ ID No.2 2	Trình tự ADN của plasmit pMG31
23	SEQ ID No. 23	Trình tự ADN của plasmit pMG33
24	SEQ ID No. 24	Trình tự ADN của plasmit pMG37
25	SEQ ID No. 25	Trình tự ADN của plasmit pMG39
26	SEQ ID No. 26	Trình tự ADN của plasmit pMG47
27	SEQ ID No. 27	Trình tự ADN của plasmit pMG70
28	SEQ ID No. 28	Trình tự ADN của plasmit pCP32AMP
29	SEQ ID No. 29	Trình tự ADN của plasmit pCP14
30	SEQ ID No. 30	Trình tự ADN của plasmit pCP50
31	SEQ ID No. 31	Trình tự ADN của plasmit pCP54
32	SEQ ID No. 32	Trình tự ADN của plasmit pCP55
33	SEQ ID No. 33	Trình tự ADN của plasmit YEP24
34	SEQ ID No. 34	Trình tự ADN của vùng <i>aroE</i> khuyết đoạn
35	SEQ ID No. 35	Trình tự ADN của catxet $\Delta ack::P_{15}aroB$ được gắn
36	SEQ ID No. 36	Trình tự ADN của vùng $\Delta poxB$
37	SEQ ID No. 37	Trình tự ADN của catxet $\Delta poxB::tklA$ được gắn
8	SEQ ID No. 38	Trình tự ADN của vùng $\Delta ptsHI$
39	SEQ ID No. 9	Trình tự ADN của catxet $\Delta tdc::glf-glK$ được gắn
40	SEQ ID No. 40	Trình tự ADN của vùng $\Delta galP$
41	SEQ ID No. 41	MYR352 $\Delta adhE::P_{15-catAX}$, P_R-aroY , $P_{26-quiC}$

Bảng 4. Alen đột biến <i>aroG</i> * dẫn đến tính kháng với sự ức chế phản hồi phenylalanin			
Chủng	Số alen	Đột biến nucleotit	Đột biến axit amin
RY893	<i>aroG</i> *20-893	C449T	Pro150Leu
RY897	<i>aroG</i> *20-897	C449T	Pro150Leu
RY899	<i>aroG</i> *20-899	T538C	Ser180Pro
RY901	<i>aroG</i> *20-901	C438T	Pro150Ser
MYR450	<i>aroG</i> *111	C55T	Pro19Ser
MYR451	<i>aroG</i> *211	G533A	Gly178Glu
MYR452	<i>aroG</i> *212	C540T	Ser180Phe
MYR453	<i>aroG</i> *311	Khuyết đoạn từ cặp bazơ 36 đến 44bp	Khuyết đoạn Glu-Ile-Lys
MYR454	<i>aroG</i> *312	C632T	Ser211Phe
MYR455	<i>aroG</i> *411	T29A	Ile10Asn
MYR456	<i>aroG</i> *412	G533A	Gly178Glu
MYR457	<i>aroG</i> *511	C448T	Pro150Ser

Bảng 5. Xác định hoạt tính của AroG ở dịch chiết thô từ các chủng <i>E. coli</i> tái tổ hợp khác nhau			
Chủng	Alen <i>aroG</i> *	Hoạt tính đặc hiệu mU (1mU = 1nM sản phẩm được tạo ra/miligam protein/phút)	% hoạt tính kháng phenylalanin
RY893	<i>aroG</i> *20-893	62	34
RY897	<i>aroG</i> *20-897	55	77
RY899	<i>aroG</i> *20-899	92	113
RY901	<i>aroG</i> *20-901	78	76
RY902	<i>aroG</i> kiểu dại	38	6
RY890	<i>aroG</i> kiểu dại	54	7

Bảng 6. Tổng hợp axit muconic trong bình lắc bằng chủng chứa alen <i>aroG</i> * kháng phản hồi
--

Chủng	Hiệu giá axit muconic g/l
RY913	3,04
RY919	3,11
RY921	2,99
RY922	1,45

Tài liệu tham khảo

Patent Hoa Kỳ số 4,480,034

Patent Hoa Kỳ số 4,535,059

Patent Hoa Kỳ số 4,588,688

Patent Hoa Kỳ số 4,608,338

Patent Hoa Kỳ số 4,681,852

Patent Hoa Kỳ số 4,753,883

Patent Hoa Kỳ số 4,833,078

Patent Hoa Kỳ số 4,968,612

Patent Hoa Kỳ số 5,168,056

Patent Hoa Kỳ số 5,272,073

Patent Hoa Kỳ số 5,487,987

Patent Hoa Kỳ số 5,616,496

Patent Hoa Kỳ số 6,600,077

Patent Hoa Kỳ số 6,180,373

Patent Hoa Kỳ số 6,210,937

Patent Hoa Kỳ số 6,472,169

Patent Hoa Kỳ số 6,613,552

Patent Hoa Kỳ số 6,962,794

Patent Hoa Kỳ số 7,244,593

Patent Hoa Kỳ số 7,638,312

Patent Hoa Kỳ số 7,790,431

Công bố đơn yêu cầu cấp patent Hoa Kỳ số US 2009/0191610

Công bố đơn yêu cầu cấp patent Hoa Kỳ số US 2010/0314243 A1

Công bố đơn yêu cầu cấp patent châu Âu số 86300748.0

Công bố đơn quốc tế số WO 2011/017560

Công bố đơn quốc tế số WO 2011/085311

Công bố đơn quốc tế số WO 2011/123154

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *J Mol Biol* 215, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: new generation of protein database search programs, *Nucleic Acid Res* 25, 3389-3402.

Barbe, V., Vallenet, D., Fonknechten, N., Kreimeyer, A., Oztas, S., Labarre, L., Cruveiller, S., Robert, C., Duprat, S., Wincker, P., Ornston, L. N., Weissenbach, J., Marliere, P., Cohen, G. N., and Medigue, C. (2004) Unique features revealed by genome sequence of *Acinetobacter* sp. ADP1, versatile and naturally transformation competent bacterium, *Nucleic Acid Res* 32, 5766-5779.

Bird, J. A. and Cain, R. B. (1968) cis-cis-muconate, product inducer of catechol 1,2-oxygenase in *Pseudomonas aeruginosa*. *Biochem. J.* 109, 479-481.

Bongaerts, J., Kramer, M., Muller, U., Raven, L. and Wubbolts, M. (2001) Metabolic engineering for microbial production of aromatic acid and derived compounds. *Met. Eng.* 3, 289-300.

Chandran, S. S., Yi, J., Draths, K. M., von Daeniken, R., Weber, W. and Frost, J. W. (2003) Phosphoenolpyruvate availability and the biosynthesis of shikimic acid. *Biotechnol. Prog.* 19, 808-814.

Chen, R., Hatzimanikatis, V., Yap, W. M. G. J., Potma, P. W. and Bailey, J. E. (1997) Metabolic consequences of phosphotransferase (PTS) mutation in phenylalanine-producing recombinant *Escherichia coli*. *Biotechnol. Prog.* 13, 768-775.

Chen, K., Dou, J., Tang, S., Yang, Y., Wang, H., Fang, H. and Zhou, C. (2012) Deletion of *aroK* gene is essential for high shikimic acid accumulation through shikimate in *E. coli*. *Biorenguồn Technol*, 119, 141-147.

Choi, W. J., Lee, E. Y., Cho, M. H., and Choi, C. Y. (1997) Enhanced production of cis, cis-muconate in cell-recycle bioreactor. *J. Fermentation and Bioengineering*. 84, 70-76.

Curran, K. A., Leavitt, J. M., Karim, A. S. and Alper, H. S. (2012) Metabolic engineering of production acid muconic in *Saccharomyces cerevisiae*. *Metabol. Engineer.* 15, 55-66.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., Durot, M., Kreimeyer, A., Le Fevre, F., Schachter, V., Pezo, V., Doring, V., Scarpelli, C., Medigue, C., Cohen, G. N., Marliere, P., Salanoubat, M., and Weissenbach, J. (2008) A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1, *Mol Syst Biol* 4, 174.

Draths, K. M., Pompliano, D. L., Conley, D. L., Frost, J. W., Berry, A., Disbrow, G. L., Staversky, R. J., and Lievens, J. C. (1992) Biocatalytic Synthesis of Aromatics from D-Glucose - Role of Transketolase, *JouARNI của American Chemical Society* 114, 3956-3962.

Draths, K. M., and Frost, J. W. (1995) Environmentally Compatible Synthesis of Catechol from D-Glucose, *JouARNI of American Chemical Society* 117, 2395-2400.

Elsemore, D. A., and Ornston, L. N. (1995) Unusual ancestry of dehydratases associated with quinate catabolism in *Acinetobacter calcoaceticus*, *J Bacteriol* 177, 5971-5978.

Escalante, A., Calderon, R., Valdiva, A., de ADNa, R., Hernandez, G., Ramirez, O. T., Gosset, G. and Bogan, F. (2010) Metabolic engineering for the production shikimic acid in evolved *Escherichia coli* strain lacking phosphoenolpyruvate: carbohydrate phosphotransferase system. *Microbial Cell Factories* 9, 21-33.

Flores, N., Xiao, J., Berry, A., Bolivar, F. and Valle, F. (1996) Pathway engineering for the production of aromatic compounds in *Escherichia coli*. *Nature Biotechnol.* 14, 620 – 623.

Fox, D. T., Hotta, K., Kim, C. Y., and Koppisch, A. T. (2008) The missing link in petrobactin biosynthesis: *asbF* encodes (-)-3-dehydroshikimate dehydratase, *Biochemical* 47, 12251-12253.

Ger, Y., Chen, S., Chiang, H., and Shiuan, D. (1994) A Single Ser-180 Mutation Desensitizes Feedback Inhibition of the Phenylalanine-Sensitive 3-Deoxygen-D-Arabino-Hepulosonate 7-Phosphate (DAHP) Synthetase in *Escherichia coli*, *J Biochem* 116, 986-990.

Grant, D. J., and Patel, J. C. (1969) The non-oxidative decarboxylation of p-hydroxybenzoic acid, gentisic acid, protocatechuic acid gallic acid by *Klebsiella aerogenes* (*Aerobacter aerogenes*), *Antonie Van Leeuwenhoek* 35, 325-343.

Hansen, E. H., Moller, B. L., Kock, G. R., Bunner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S., and Hansen, J. (2009) De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*), *Appl Environ Microbiol* 75, 2765-2774.

Hu, C., Jiang, P., Xu, J., Wu, Y., and Huang, W. (2003) Mutation analysis of the feedback inhibition site of phenylalanine-sensitive 3-deoxygen-D-arabino-heptulosonate 7-phosphate synthase of *Escherichia coli*, *J Basic Microbiol* 43, 399-406.

Iwagami, S. G., Yang, K., and Davies, J. (2000) Characterization of protocatechuic acid catabolic gene cluster from *Streptomyces* sp. strain 2065, *Appl Environ Microbiol* 66, 1499-1508.

Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., Moore, J. C., Shanmugam, K. T., and Ingram, L. O. (2008a) Combining metabolic engineering and metabolic evolution to develop nonrecombinant strain of *Escherichia coli* C that produce succinate and malate, *Biotechnol Bioeng* 99, 1140-1153.

Jantama, K., Zhang, X., Moore, J. C., Shanmugam, K. T., Svoronos, S. A., and Ingram, L. O. (2008b) Eliminating side products and increasing succinate yields in engineered strain of *Escherichia coli* C, *Biotechnol Bioeng* 101, 881-893.

Kaneko, A., Ishii, Y., and Kirimura, K. (2011) High-yield production of cis, cis-muconic acid from catechol in aqueous solution by biocatalyst. *Chem. Lett.* 40, 381-383.

Kikuchi, Y., Tsujimoto, K., and Kurahashi, O. (1997) Mutational analysis of feedback sites of phenylalanine-sensitive 3-deoxygen-D-arabino-heptulosonate-7-phosphate synthase of *Escherichia coli*, *Appl Environ Microbiol* 63, 761-762.

Kojima, Y., Fujisawa, H., Nakazawa, A., Nakazawa, T., Kanetsuna, F., Taniuchi, H., Nozaki, M., and Hayaishi, O. (1967) Studies on pyrocatechase. I. Purification and spectral properties, *J Biol Chem* 242, 3270-3278.

Kramer, M., Bongaerts, J., Bovenberg, R., Kremer, S., Muller, U., Orf, S., Wubbolts, M. and Raeven, L. (2003) Metabolic engineering for microbial production shikimic acid. *Metabol. Eng.* 5, 277-283.

Lerner, C. G., and Inouye, M. (1990) Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability, *Nucleic Acid Res* 18, 4631.

Li, K. and Frost, J.W. (1999) Microbial synthesis of 3-dehydroshikimic acid: A comparative analysis D-xylose, L-arabinose, and D-glucose carbon sources. *Biotechnol. Prog.* 15, 876-883.

Lu, J. L., and Liao, J. C. (1997) Metabolic engineering and control analysis for production aromatics: Role of transaldolase, *Biotechnol Bioeng* 53, 132-138.

Lu, J., Tang, J., Liu, Y., Zhu, X. (2012) Combinatorial modulation of galP and glk gene expression for improves alternative glucose utilization. *Appl. Microbiol. Biotechnol.* 93, 2455-2462

Lutke-Eversloh, T., and Stephanopoulos, G. (2007) L-tyclopheane production by deregulated strain of *Escherichia coli*, *Appl Microbiol Biotechnol* 75, 103-110.

Mizuno, S., Yoshikawa, N., Seki, M., Mikawa, T., and Imada, Y. (1988) Microbial production cis, cis- muconic acid from benzoic acid. *Appl Microbiol Biotechnol.* 28, 20-25.

Nakasewa, A., Kojima, Y., and Taniuchi, H. (1967) Purification and properties of pyrocatechase from *Pseudomonas fluorescens*, *Biochim Biophys Acta* 147, 189-199.

Neidhardt, F. C., and Curtiss, R. (1996) *Escherichia coli and Salmonella : cellular and molecular biology*, 2nd ed., ASM Press, Washington, D.C.

Neidle, E. L., and Ornston, L. N. (1986) Cloning and expression of *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase structural gene *catA* in *Escherichia coli*, *J Bacteriol* 168, 815-820.

Niu, W., Draths, K. M., and Frost, J. W. (2002) Benzene-free synthesis of adipic acid, *Biotechnol Prog* 18, 201-211.

Parker, C., Barnell, W. O., Snoep, J. L., Ingram, L. O., and Conway, T. (1995) Characterization of *Zymomonas mobilis* glucose facilitator gene product (*glf*) in recombinant *Escherichia coli*: examination of transport mechanism, kinetics and role of glucokinase in glucose transport, *Mol Microbiol* 15, 795-802.

Parsek, M. R., Shinabarger, D. .L., Rithmel, R. K. and Chakrabarty, A. M. (1992) Roles of *CatR* and *cis*, *cis*-Muconate in activation of *catBC* operon, which is involved in benzoate degradation in *Pseudomonas putida*. *J. Bacteriol.* 174, 7798-7806.

Patnaik, R. and Liao, J. C. (1994) Engineering of *Escherichia coli* central metabolism for aromatic metabolite with near theoretical yield. *App. Env. Microbiol.* 60, 3903-3908.

Pfleger, B. F., Kim, Y., Nusca, T. D., Maltseva, N., Lee, J. Y., Rath, C. M., Scaglione, J. B., Janes, B. K., Anderson, E. C., Bergman, N. H., Hanna, P. C., Joachimiak, A., and Sherman, D. H. (2008) Structural and functional analysis *AsbF*: origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis, *Proc Natl Acad Sci U S A* 105, 17133-17138.

Perez-Pantoja, D., De la Iglesia, R., Pieper, D. H., and Gonzalez, B. (2008) Metabolic reconstruction of aromatic compounds degradation from the genome of amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134, *FEMS*

Microbiol Rev 32, 736-794.

Perez-Pantoja, D., Donoso, R., Agullo, L., Cordova, M., Seeger, M., Pieper, D. H., and Gonzalez, B. (2011) Genomic analysis of potential for aromatic compounds biodegradation in Burkholderiales, *Environ Microbiol*.

Pittard, J. and Wallace, B. J. (1966) Distribution and function of genes concerned with aromatic biosynthesis in *Escherichia coli*. *J. Bacteriol.* 91, 1494-1508.

Polen, T., Spelberg, M. and Bott, M. (2012) toward biotechnological production of adipic acid and precursors from biorenewables, *J. Biotechnol*, <http://dx.doi.org/10.1016/j.biotech.2012-07.008>.

Rutledge, B. J. (1984) Molecular characterization of qa-4 gene of *Neurospora crassa*, *Gene* 32, 275-287.

Schirmer, F., and Hillen, W. (1998) The *Acinetobacter calcoaceticus* NCIB8250 mop operon mRNA is differentially degraded, resulting in a higher level of 3' CatA-encoding segment than of 5' phenolhydroxylase-encoding portion, *Mol Gen Genet* 257, 330-337.

Shumilin, I. A., Kretsinger, R. H., and Bauerle, R. H. (1999) Crystal structure of phenylalanine-regulated 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*, *Structure* 7, 865-875.

Shumilin, I. A., Zhao, C., Bauerle, R., and Kretsinger, R. H. (2002) Allosteric inhibition of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase alters coordination of both substrates, *J Mol Biol* 320, 1147-1156.

Shumilin, I. A., Bauerle, R., Wu, J., Woodard, R. W., and Kretsinger, R. H. (2004) Crystal structure of the reaction complex of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Thermotoga maritima* refines catalytic mechanism and indicates new mechanism of allosteric regulation, *J Mol Biol* 341, 455-466.

Shumkova, E. S., Solyanikova, I. P., Plotnikova, E. G. and Golovleva, L. A. (2009) Phenol degradation by *Rhodococcus opacus* Strain 1G. *App. Biochem.*

Microbiol. 45, 43-49.

Sietmann, R., Uebe, R., Boer, E., Bode, R., Kunze, G., and Schauer, F. (2010) Novel metabolic routes during the oxidation of hydroxylated aromatic acid by yeast *Arxula adeninivorans*, *J Appl Microbiol* 108, 789-799.

Smith, M. R. and Ratledge, C. (1989) Quantitative biotransformation of catechol to cis, cis –muconate. *Biotech. Lett.* 11, 105-110.

Snoep, J. L., Arfman, N., Yomano, L. P., Fliege, R. K., Conway, T., and Ingram, L. O. (1994) Reconstruction of glucose uptake and phosphorylation in glucose-negative mutant of *Escherichia coli* by using *Zymomonas mobilis* genes encoding the glucose facilitator protein and glucokinase, *J Bacteriol* 176, 2133-2135.

Sprenger, G. A. (1995) Genetics of pentose-phosphate pathway enzyme of *Escherichia coli* K-12, *Arch Microbiol* 164, 324-330.

Sprenger, G. A., Schorken, U., Sprenger, G., and Sahm, H. (1995a) Transketolase A of *Escherichia coli* K12. Purification and properties of enzyme from recombinant strains, *Eur J Biochem* 230, 525-532.

Sprenger, G. A., Schorken, U., Sprenger, G., and Sahm, H. (1995b) Transaldolase B of *Escherichia coli* K-12: cloning of its gene, *talB*, and characterization of enzyme from recombinant strains, *J Bacteriol* 177, 5930-5936.

Stroman, P., Reinert, W. R., and Giles, N. H. (1978) Purification and characterization of 3-dehydroshikimate dehydratase, an enzyme in the inducible quinic acid catabolic pathway of *Neurospora crassa*, *J Biol Chem* 253, 4593-4598.

Tang, J., Zhu, X., Lu, J. and Liu, P. (2012) Recruiting alternative glucose utilization pathways for improving succinate production. *App Microbiol Biotechnol* DOI 10, 1007/s00253-012-434.1

Tateoka, T., and Yasuda, I. (1995) 3-Dehydroshikimate dehydratase in mung bean cultured cells, *Plant Cell Reports* 15, 212-217.

Weaver, L. M., and Herrmann, K. M. (1990) Cloning of an *aroF* allele

encoding a tyrosine-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, *J Bacteriol* 172, 6581-6584.

Weber, C., Bruckner, C., Weinreb, S., Lehr, C., Essl, C. and Bole, E. (2012) Biosynthesis of cis, cis-muconic acid and its aromatic precursors catechol and proteocatechuic acid, từ renewable feedstocks by *Saccharomyces cerevisiae*, *App Environ Microbiol.* 78, 8421-8430.

Wheeler, K. A., Lamb, H. K., and Hawkins, A. R. (1996) Control of metabolic flux through quinate pathway in *Aspergillus nidulans*, *Biochem J* 315 (Pt 1), 195-205.

Wu, C-M., Wu, C-C., Su, C-C., Lee, S-N., Lee, Y-A. and Wu, J-Y. (2006) Microbial synthesis of cis,cis-muconic acid form benzoate by *Sphingobacterium* sp. Mutants. *Biochem. Eng. J.* 29, 35-40.

Xie, N., Tang, H., Feng, J., Tao, F., Ma, C. and Xu, P. (2009) Characterization of benzoate degradationby newly isolated bacterium *Pseudomonas* sp. XP-M2. *Biochem. Eng. J.* 46, 79-82.

Yi, J., Draths, K. M., Li, K. and Frost, J. W. (2003) Altered Glucose Transport and Pathway shikimate Product Yields in *E. coli*. *Biotechnol. Prog.* 2003, 19, 1450-1459

Yoshikawa, N., Mizuno, S., Ohta, K., and Suzuki, M. (1990) Microbial production cis, cis-muconic acid. *J. Biotechnol.* 14, 203-210

YÊU CẦU BẢO HỘ

1. Vi sinh vật được thiết kế di truyền để tổng hợp ra axit cis, cis-muconic bắt đầu từ nguồn cacbon không thơm chứa gen mã hóa enzym shikimat dehydrogenaza rò, trong đó enzym shikimat dehydrogenaza rò này có khả năng nguyên dưỡng đối với axit amin thơm và vitamin, nhưng không làm tiết đáng kể các hợp chất thơm.
2. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó vi sinh vật được thiết kế di truyền này là vi khuẩn.
3. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó vi sinh vật được thiết kế di truyền này là *Escherichia coli*.
4. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó vi sinh vật này còn chứa một hoặc nhiều gen ngoại sinh được chọn từ nhóm bao gồm *aroZ*, *qa-4*, *asbF*, *aroY*, *quiC* và *catAX*, trong đó các gen ngoại sinh này mã hóa protein có liên quan về mặt chức năng đến con đường axit muconic.
5. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó vi sinh vật này còn chứa một hoặc nhiều gen ngoại sinh được chọn từ nhóm gồm *aroB*, *aroD*, *aroF*, *aroG*, *aroH*, *tktA*, *talB*, *rpe* và *rpi*, trong đó các gen ngoại sinh này mã hóa protein có liên quan về mặt chức năng đến con đường axit shikimic.
6. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó hoạt tính của protein điều hòa âm của quy trình sinh tổng hợp axit amin thơm được mã hóa bởi gen *tyrR* hoặc thể tương đồng của nó bị giảm đáng kể hoặc bị loại trừ.
7. Vi sinh vật được thiết kế di truyền theo điểm 1, trong đó vi sinh vật này còn chứa gen *aroG** mã hóa enzym DAHP synthaza về cơ bản là kháng lại sự ức chế bởi phenylalanin.
8. Vi sinh vật được thiết kế di truyền theo điểm 7, trong đó vi sinh vật này chứa gen *aroG** mã hóa enzym DAHP synthaza mà được chọn từ nhóm bao gồm *aroG*20-893*, *aroG*20-897*, *aroG*20-899*, *aroG*20-901*, *aroG*111*, *aroG*211*, *aroG*212*, *aroG*311*, *aroG*312*, *aroG*411*, *aroG*412* và *aroG*511*.

Fig.1

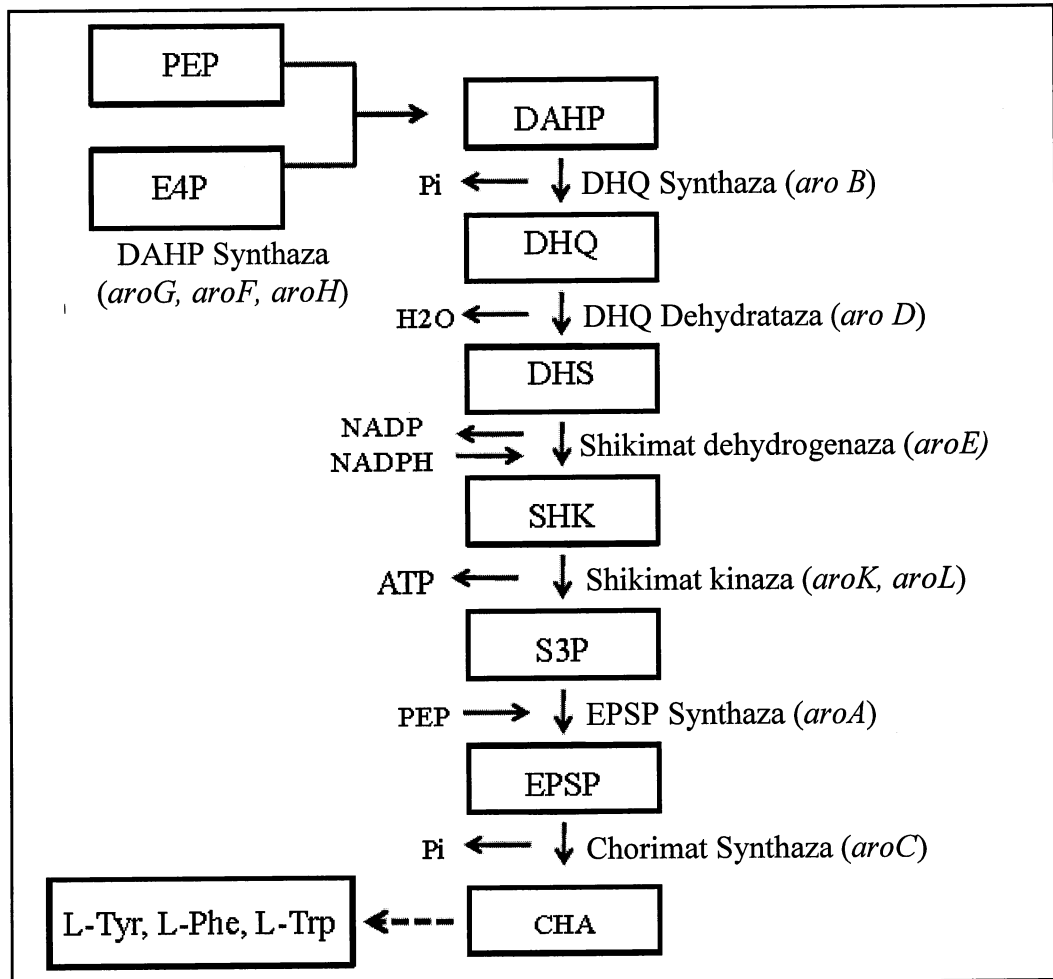


Fig.2

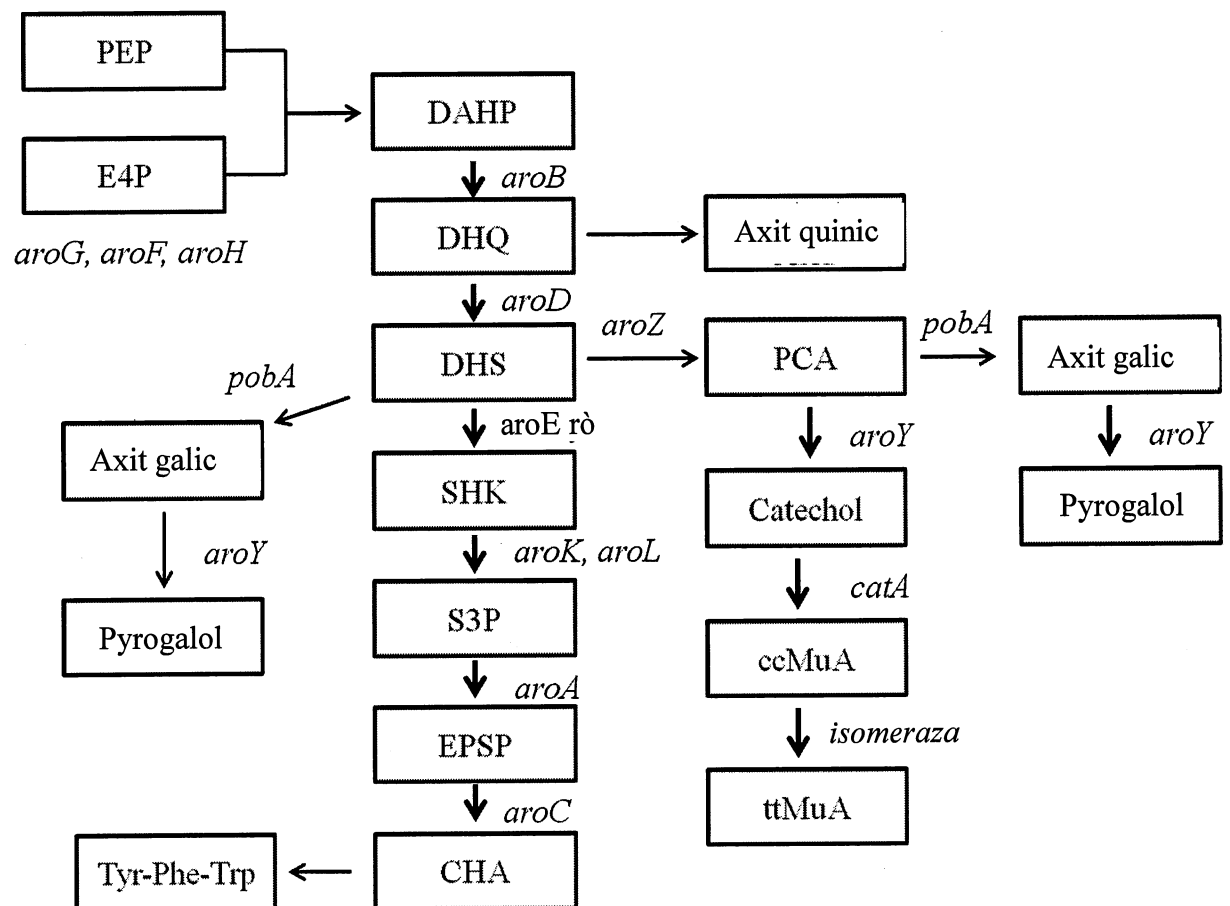


Fig.3

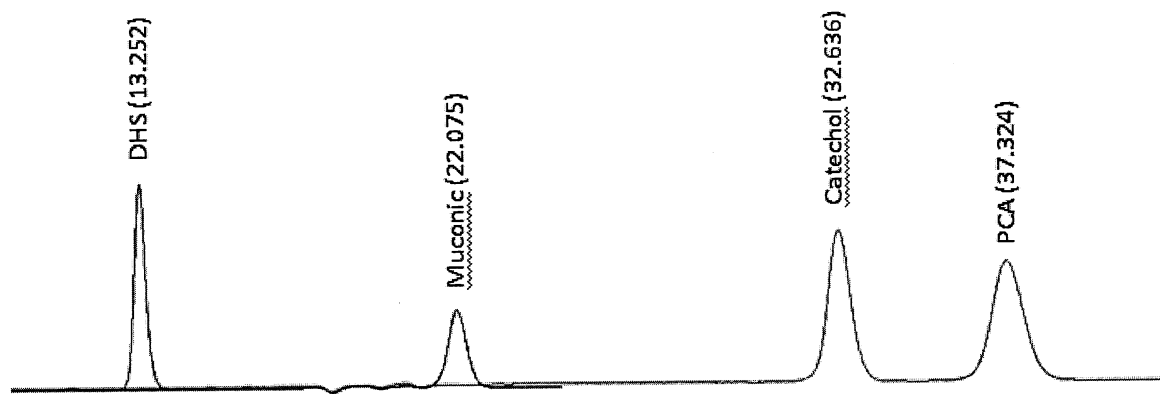


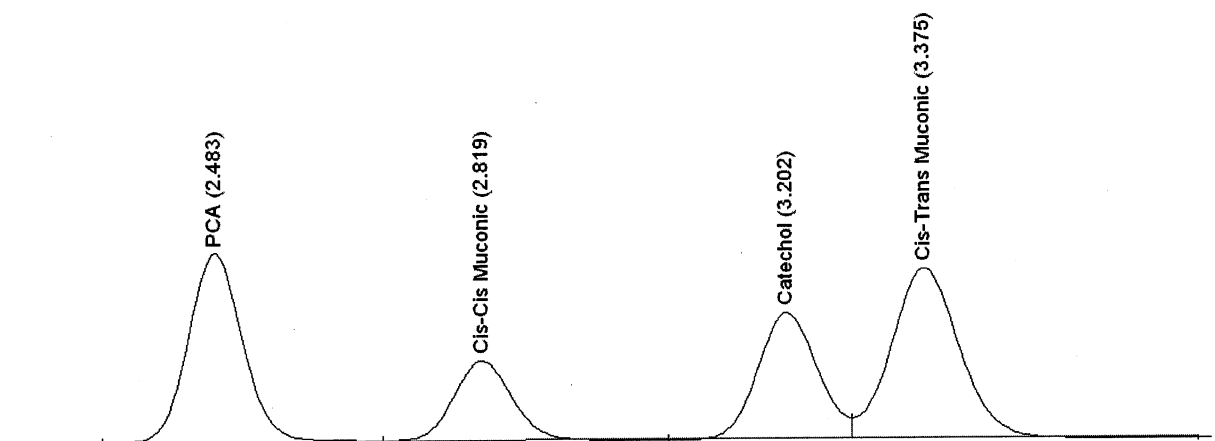
Fig.4

Fig.5

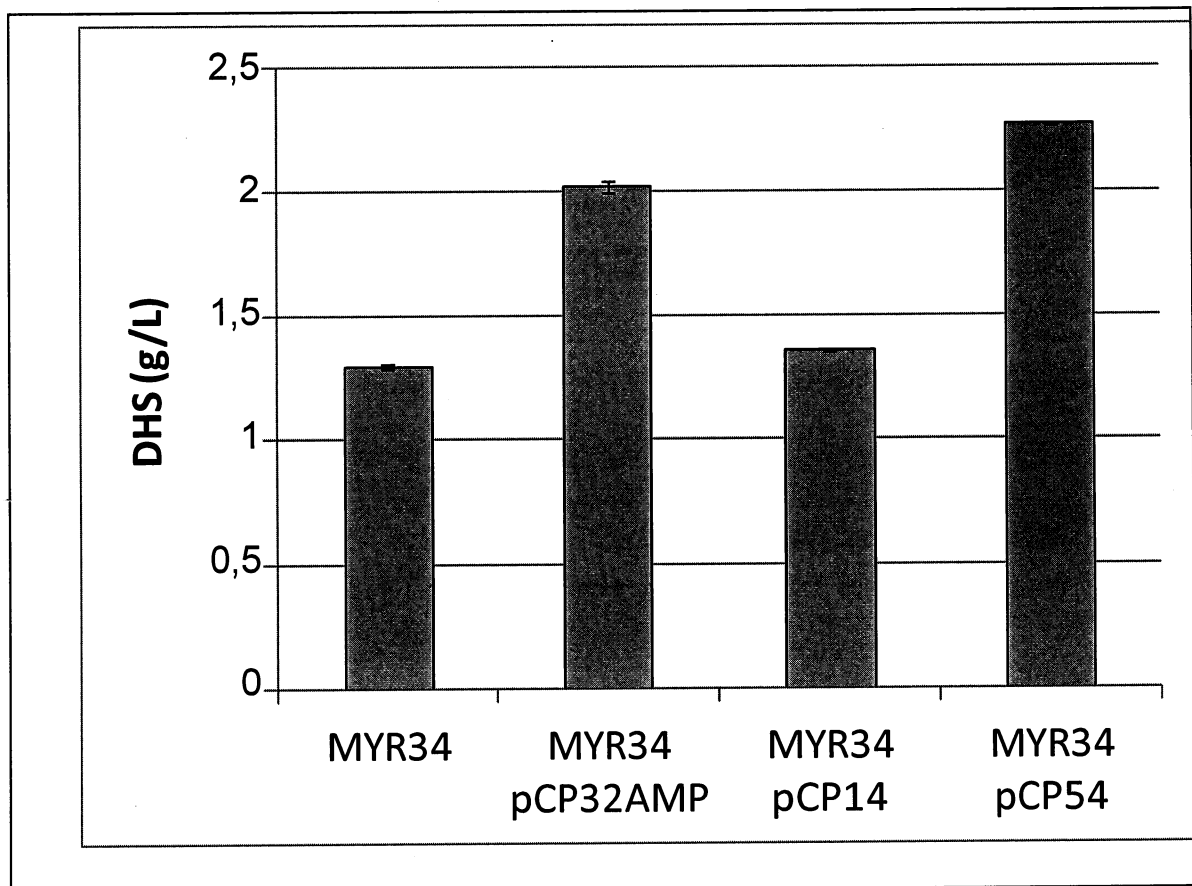


Fig.6

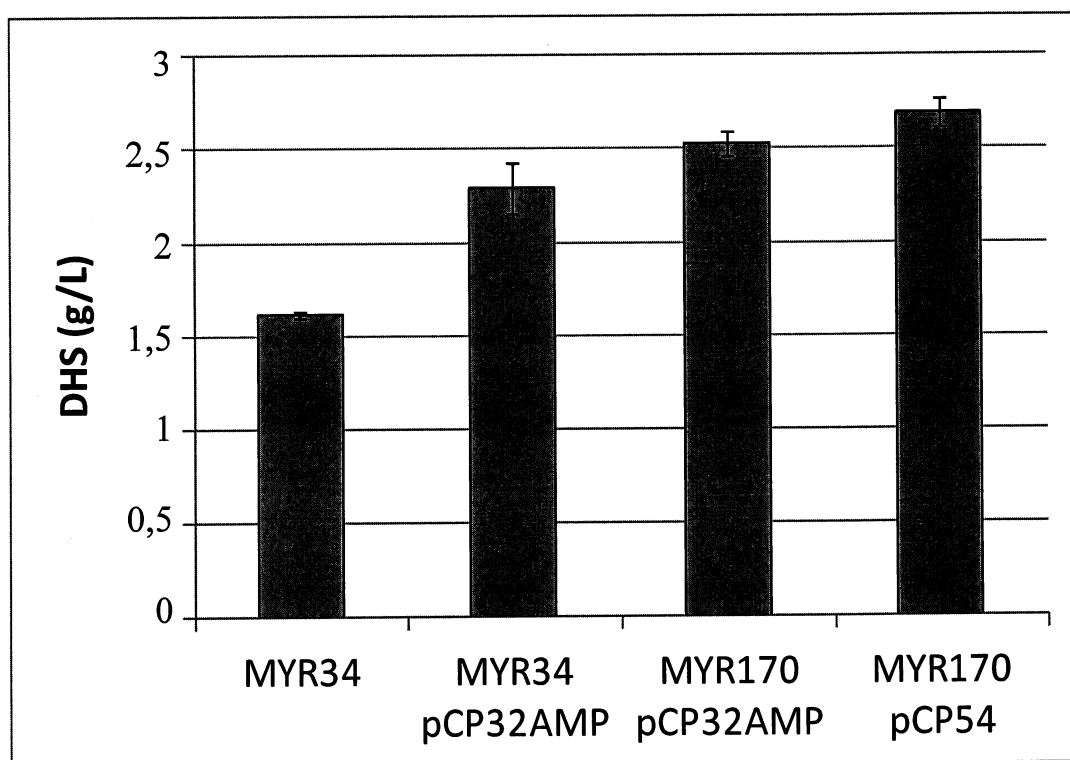


Fig.7

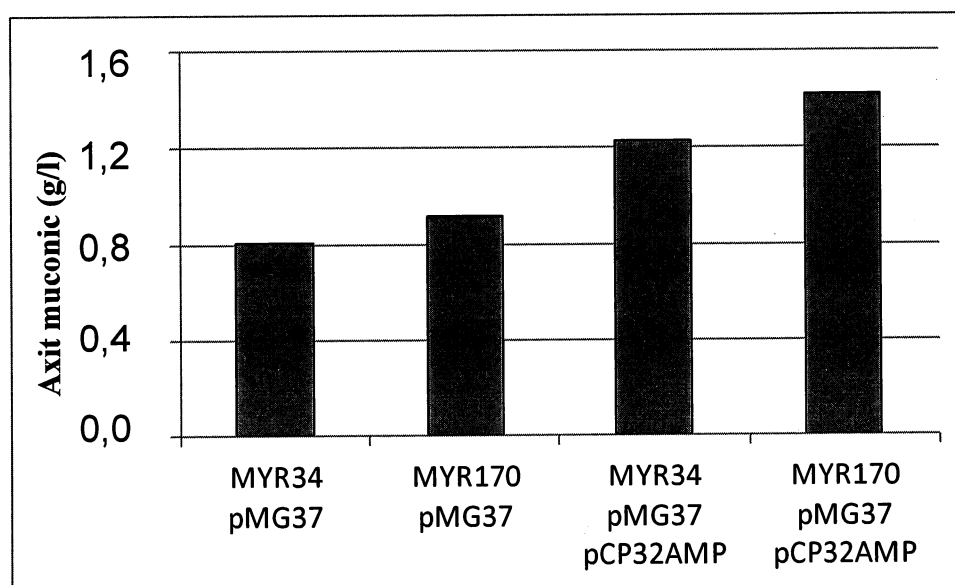


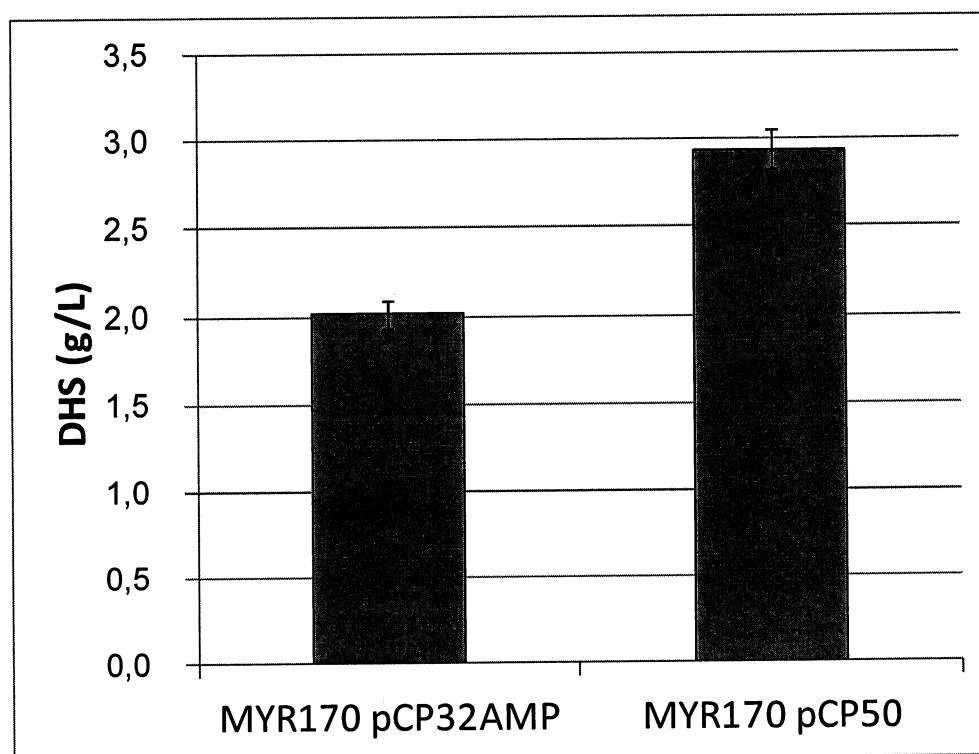
Fig.8

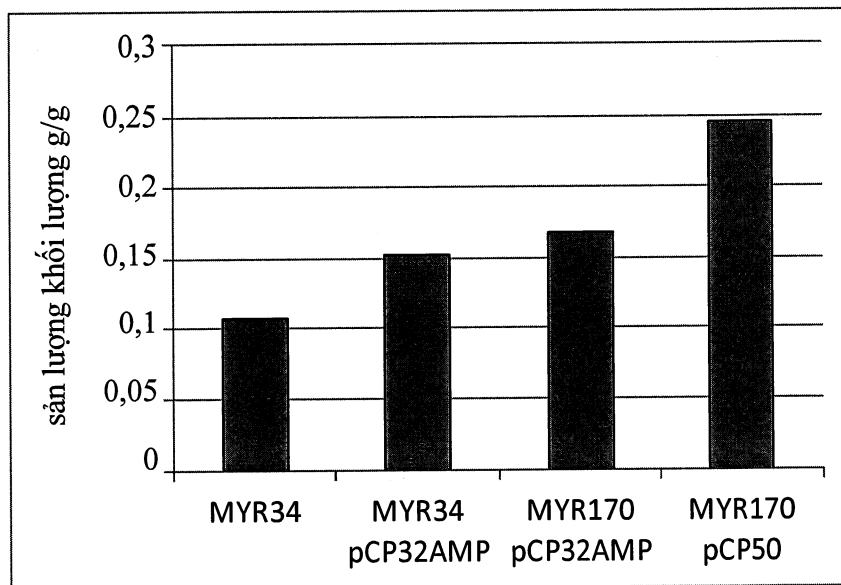
Fig.9

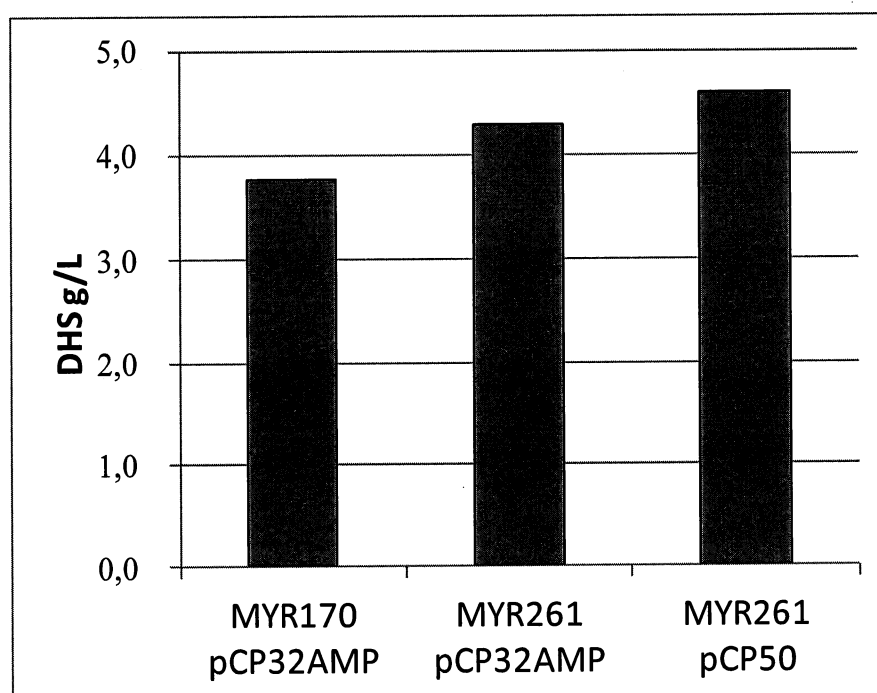
Fig.10

Fig.11

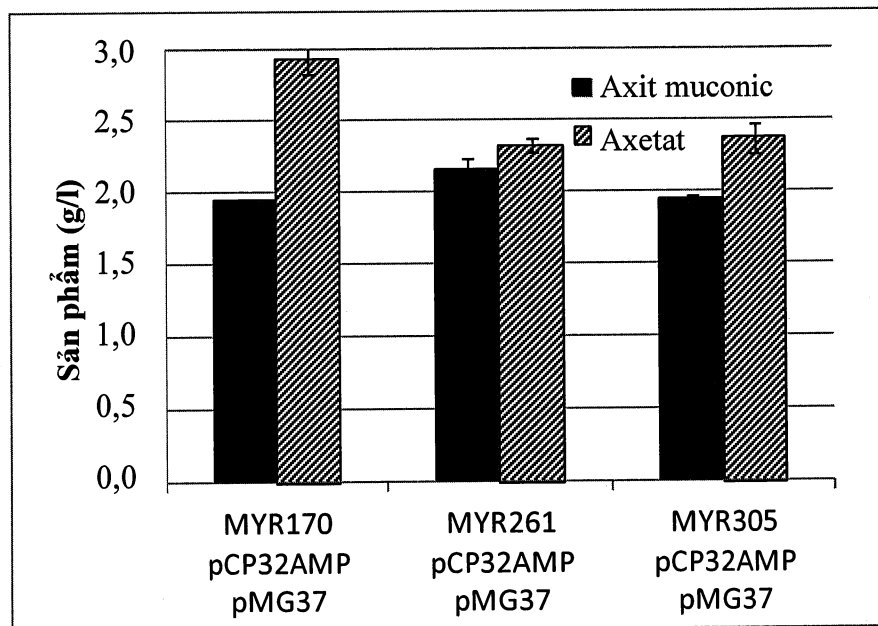


Fig.12

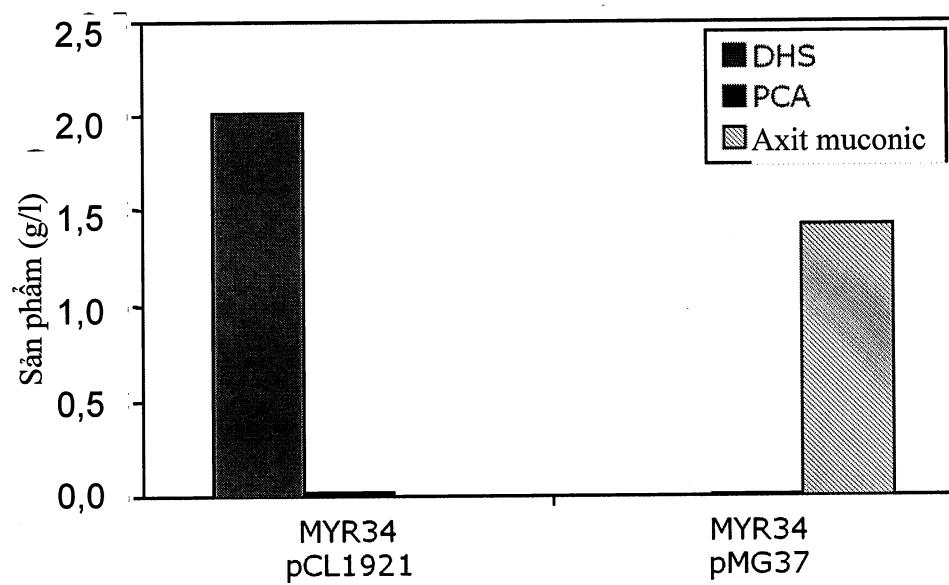


Fig.13

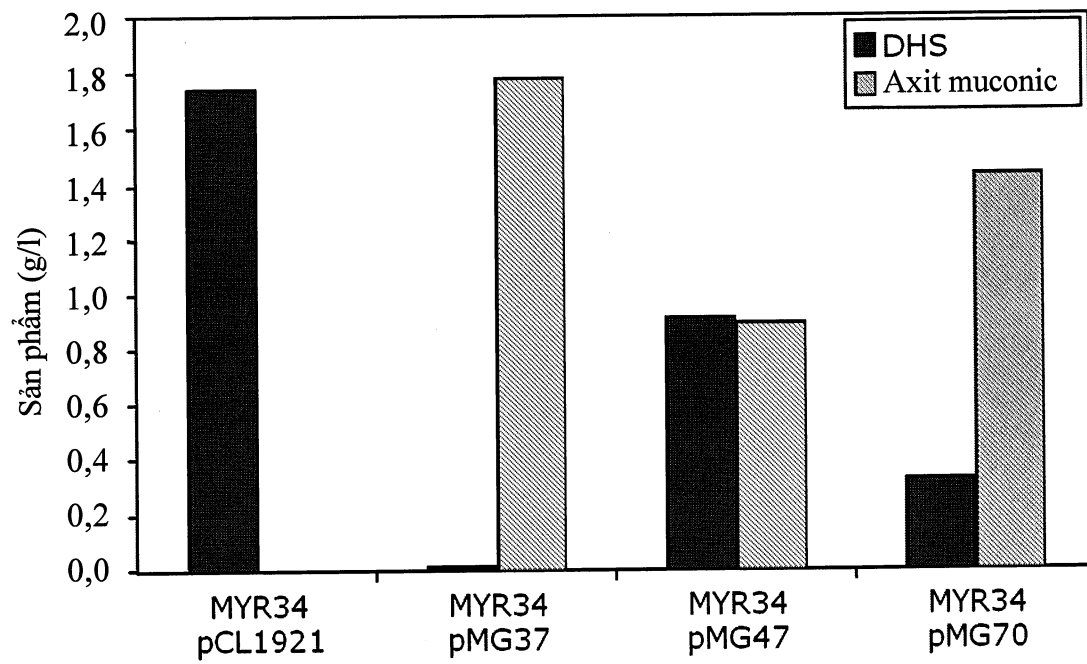


Fig.14

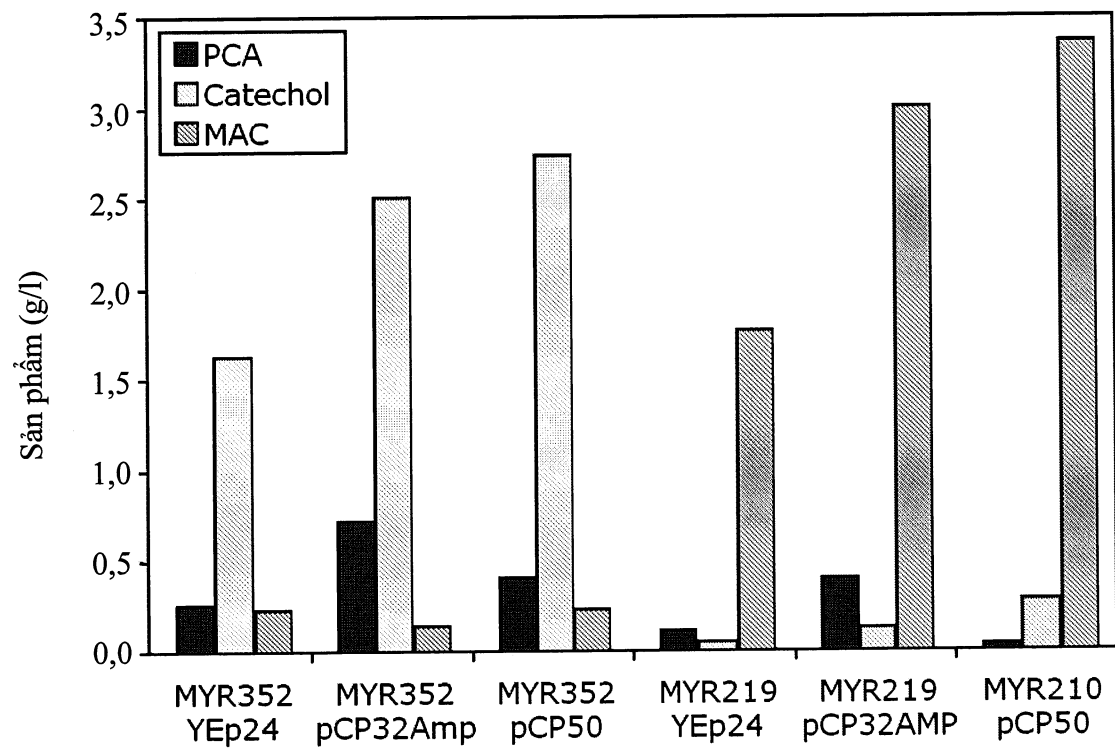


Fig.15

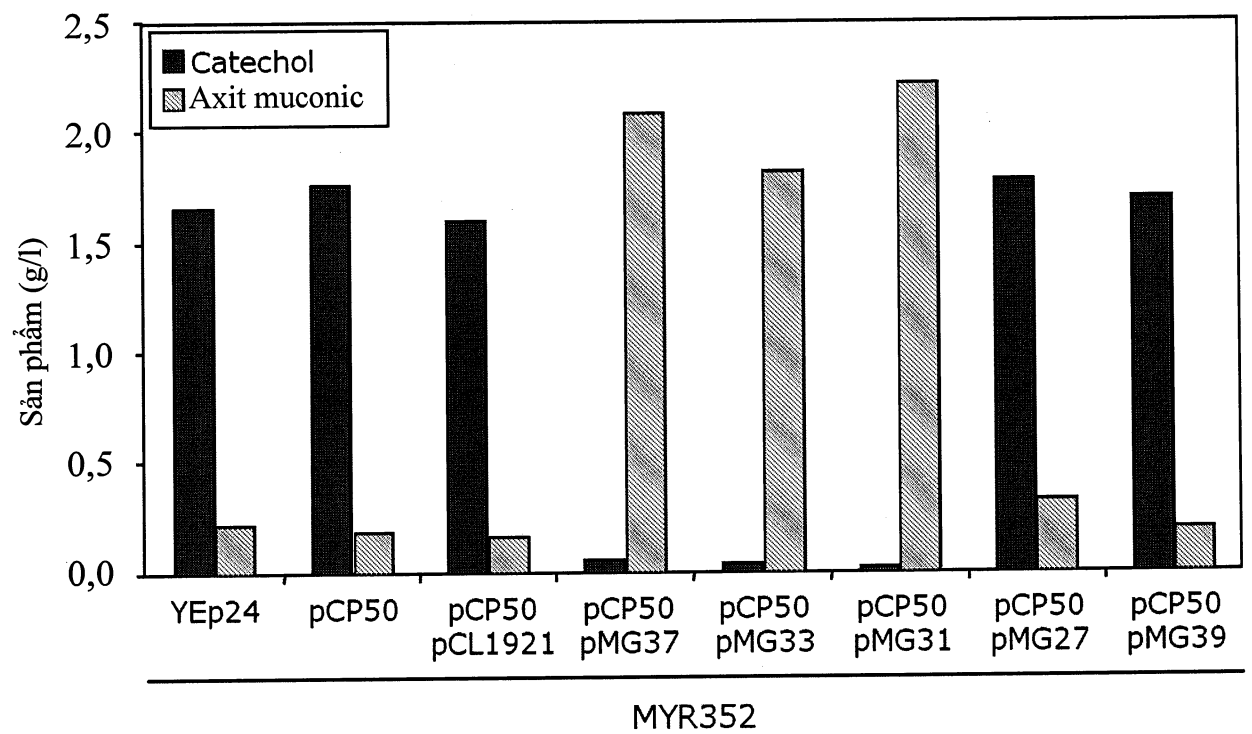


Fig.16

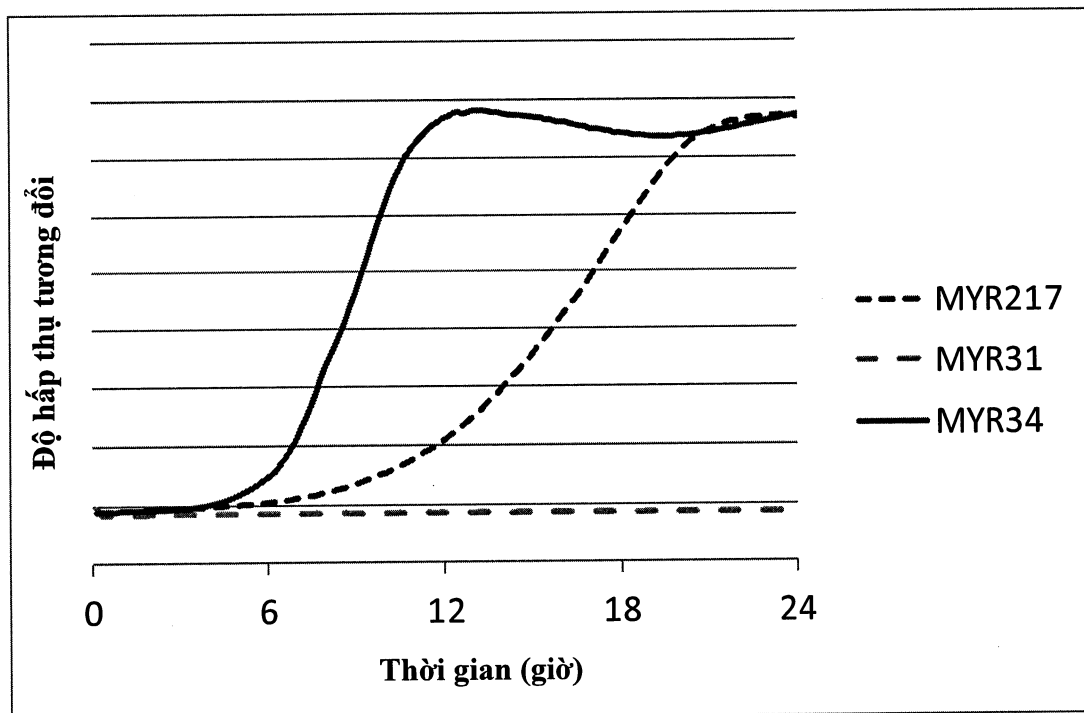


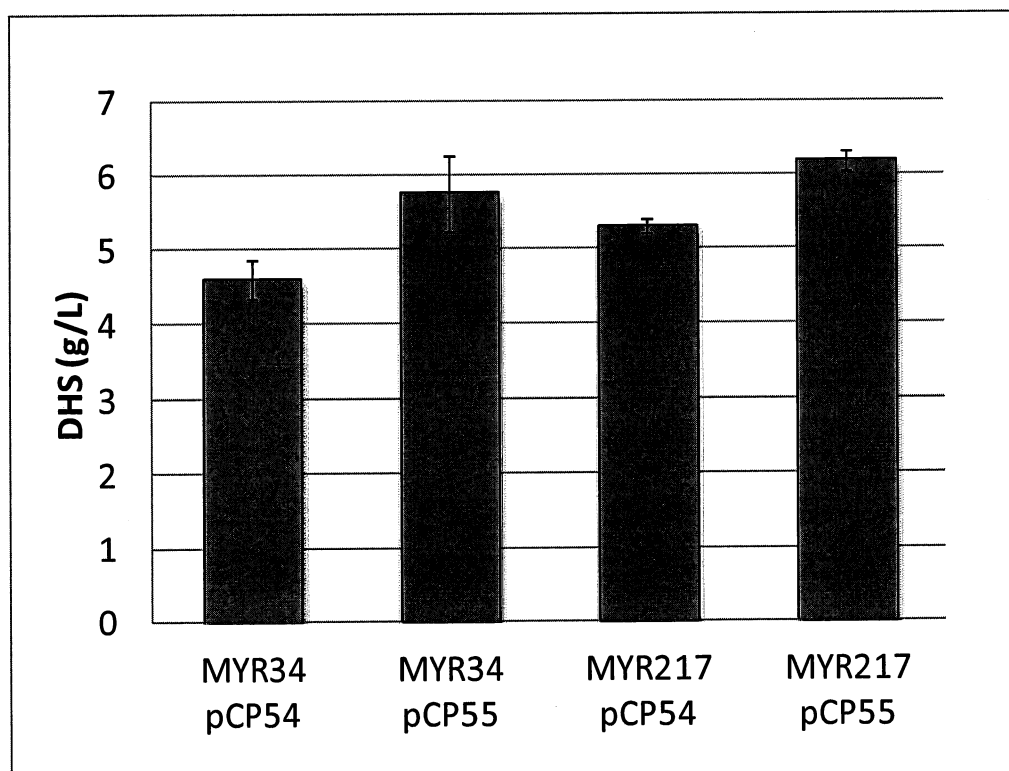
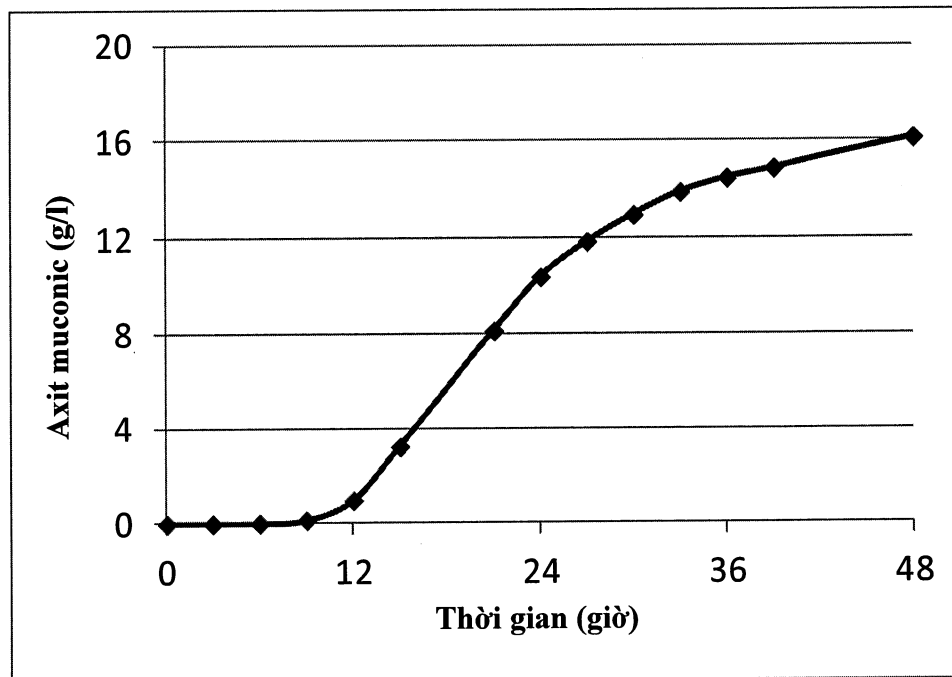
Fig.17

Fig.18



DANH MỤC TRÌNH TỰ

- <110> MYRIANT CORPORATION
Yocum, R. Rogers
Gong, Wei
Dole, Sudhanshu
Sillers, Ryan
Gandhi, Meghal
Pero, Janice G.
- <120> Vi sinh vật được thiết kế di truyền để tổng hợp ra axit muconic
- <130> MC2013-01PCT
- <150> US 61/632,777
<151> 2012-01-30
- <160> 41
- <170> PatentIn phiên bản 3.5
- <210> 1
<211> 195
<212> ADN
<213> trình tự khởi đầu P15;
- <400> 1
gctattgacg acagctatgg ttcactgtcc accaaccaaa actgtgctca gtaccgccaa 60
tattttctccc ttgaggggta caaagagggtg tccctagaag agatccacgc tgtgtaaaaa 120
ttttacaaaa aggtattgac tttccctaca ggggtgtgtaa taatttaatt acaggcgggg 180
gcaaccccg cgtgtt 195
- <210> 2
<211> 164
<212> ADN
<213> trình tự khởi đầu P26;
- <400> 2
gcctacctag cttccaagaa agatataccta acagcacaag agcggaaaga tgttttgttc 60
tacatccaga acaacctctg ctaaaattcc tgaaaaattt tgcaaaaagt tgttgacttt 120
atctacaagg tgtggtataa taatcttaac aacagcagga cgct 164
- <210> 3
<211> 91
<212> ADN
<213> trình tự khởi đầu PR;
- <400> 3
acgttaaadc tatcaccgca aggataaat atctaaccac gtgcgtgttg actattttac 60
ctctggcggt gataatgggt gcatgtacaa g 91
- <210> 4

<211> 359

<212> PRT

<213> 3-dehydroshikimat dehydratza được mã hóa bởi gen qa-4;

<400> 4

```

Met Pro Ser Lys Leu Ala Ile Ser Ser Met Ser Leu Gly Arg Cys Phe
1          5          10          15

Ala Gly His Ser Leu Asp Ser Lys Leu Asp Ala Ala Gln Arg Tyr Gly
          20          25          30

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

His Leu Ser Asn Glu Arg Pro Ser Pro Glu Gly Pro Phe Val Glu Ala
          50          55          60

Gln Ile Ala Ala Ala Arg His Ile Leu Gln Met Cys Gln Ala Arg Gly
65          70          75          80

Leu Glu Val Val Cys Leu Gln Pro Phe Met His Tyr Asp Gly Leu Asn
          85          90          95

Asp Arg Ala Glu His Glu Arg Arg Leu Glu Lys Leu Ala Leu Trp Ile
          100          105          110

Glu Leu Ala His Glu Leu His Thr Asp Ile Ile Gln Ile Pro Ala Asn
          115          120          125

Phe Leu Pro Ala Asn Gln Val Ser Asp Asn Leu Asp Leu Ile Val Ser
          130          135          140

Asp Leu Cys Lys Val Ala Asp Ile Gly Ala Gln Ala Leu Pro Pro Ile
145          150          155          160

Arg Phe Ala Tyr Glu Ser Leu Cys Trp Ser Thr Arg Val Asp Leu Trp
          165          170          175

Glu Arg Cys Trp Asp Ile Val Gln Arg Val Asp Arg Pro Asn Phe Gly
          180          185          190

Ile Cys Leu Asp Thr Phe Asn Ile Leu Gly Arg Ile Tyr Ala Asp Pro
          195          200          205

Thr Ser Pro Ser Gly Arg Thr Pro Asn Ala Lys Glu Ala Val Arg Lys
          210          215          220

Ser Ile Ala Asn Leu Val Ser Arg Val Asp Val Ser Lys Val Phe Tyr
225          230          235          240

Val Gln Val Val Asp Ala Glu Arg Leu Ser Lys Pro Leu Leu Pro Gly
          245          250          255

His Pro Tyr Tyr Asn Pro Glu Gln Pro Ala Arg Met Ser Trp Ser Arg
          260          265          270

Asn Cys Arg Leu Phe Tyr Gly Glu Thr Glu Tyr Gly Ala Tyr Leu Pro
          275          280          285

Val Lys Glu Val Ala Arg Ala Leu Phe His Gly Ile Gly Phe Glu Gly

```

290	295	300
Trp Val Ser Leu Glu Leu Phe Asn Arg Arg Met Ser Glu Glu Gly Pro		
305	310	315 320
Glu Val Pro Glu Glu Leu Ala Met Arg Gly Ala Ile Ser Trp Ala Lys		
	325	330 335
Leu Val Gln Asp Leu Arg Ile Pro Val Glu Gly Pro Leu Val Thr Met		
	340	345 350
Pro Arg Val Ser Ala Ser Leu		
	355	

<210> 5

<211> 2160

<212> ADN

<213> gen qa-4 từ *Neurospora crassa* cung với trình tự bao quanh;

<400> 5

gaattcggga aatggaatct tacctgggaa ccgaaatcac agtccgggta gggttatagag	60
catatagtga actgtcaaag ttctagacct ggaccagcca cttggagtcg ttgttttagt	120
tataacctaca ttcactcact gttgactttc aatcatactt acttagacgg agcaacgcgc	180
cagaatccaa attgttgcac agttgcggta tcaccaagtg gcttcccata atagtttgcc	240
attcgatgag acagctaact ggaagaccgg tactcgcagg ttgcacgatt acacggaagg	300
attcgggtatt ccgtgtttca tctgtcaaag tccctttcca tatgaatccg aggtactatg	360
actggatctc gatacaagct ggccagcgag gtgcctgcct tgacaggctg tcaactgcgg	420
gacggccggc taagtgttta acacgcaagg gtggaagatg tctcgtcccg tcatccaaga	480
ccgtcaacat tcgaggccat ctgatcggtg aagagatgct aaatcttggtg aaacgctcat	540
aggtegetta ccttcggccc acccgtaat gctttattcc gctgagcaaa cttcggcttc	600
catcccgcgg ttcaccgttt acatcactta tcgttgccgg tattggccga ttcttcgcaa	660
accgaaacga tgacatcccg aatatctgca atacaccgcc acggccggcg tctttatcac	720
acctcctatg ggagacgaaa gtgccttgat acccctagtc atttgaagat tcaggatggg	780
agacggctgg ccgcttgccg agttacgttc gagtcttggt cgcaggaacg cttgccgtat	840
tgaatgagac cccgagaagg tcaaatcaaa tcttggaaga cccaactgc ttcctcattg	900
ccttcactcc ccatatcaat ggggcacatc ctgtgactac cttggtgctt tatttcctca	960
ccatttggcg atacaagctc aaggacaccg aggtgatata cagttcttca aggacactat	1020
ctcacctcaa tatcaagaac cagtctcatc atctcttatt tctccaggat cccccacca	1080
acaacatcgg cttttttttt tcccctattc tcaagaccca tcaagacgct cacttcgctg	1140
agcctttcgc catgccgtca aagctagcca ttagttccat gtccttaggg cgtgctttg	1200

ccggccactc tctggacagc aagcttgatg ccgctcaacg atacggctat cttggtatcg 1260
 agcttttttta tgaggatctg gtcgacgttg cagagcattt gtcgaacgag cgtccctctc 1320
 ccgaaggccc ttttgtcgaa gctcagatag ccgccgctcg tcatattctc cagatgtgtc 1380
 aagccagggg gcttgaggtc gtctgcctcc agcctttcat gcactacgac ggccttaacg 1440
 acagggcaga acatgagcgt cgtctggaga agctagcact atggattgag ctcgctcatg 1500
 agcttcacac cgacatcatt cagatcccag ccaacttcct ccctgccaac caagtcagtg 1560
 acaacctcga cctgattgtc tcagatcttt gcaaggtggc cgatattgga gctcaagctt 1620
 tgccccctat ccgctttgcc tacgagagtc tttgctggag cacccggtgc gacctctggg 1680
 agcgctgctg ggacatcgta caacgcgttg accgccccaa ctttggcatt tgccttgaca 1740
 ctttcaacat cctcggccgc atctatgccg accctacatc tcctagcggg aggacacca 1800
 acgcaaaaga ggcagtcagg aagtccatcg ccaacttggg ctcgcgcgtg gatgtctcca 1860
 aagtcttcta cgtccagggtg gttgacgccg agaggctgag caagccacta ctgcccggtc 1920
 acccgattta caatccagag cagccggcga ggatgagctg gtcgcgcaat tgtagactgt 1980
 tctacggcga aacagaatat ggtgcgtatc ttcccgtgaa ggagggtgct cgagcccttt 2040
 tccacggcat tggtttcgag ggctgggtca gtttgagct tttcaaccgc agaattgtctg 2100
 aggagggacc tgaagtgccg gaggaacttg ccatgagagg cgctatctcg tgggccaagt 2160

<210> 6

<211> 348

<212> PRT

<213> QutC từ *Aspergillus nidulans* được mã hóa bởi gen qutC;

<400> 6

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

Pro Gly Leu His Ser Leu Asp His Lys Leu Arg Ser Ala Ala His Gly
20 25 30

Phe Ala Gly Ile Glu Leu Phe Ile Asp Asp Leu Ser His Phe Ala Ser
35 40 45

Ser Ser Phe Asn Gly Ser Leu Thr Gln Ala Ala Lys Tyr Ile Ser Ser
50 55 60

Leu Ala Lys Gln Leu Asn Leu Thr Phe Ile Cys Leu Gln Pro Phe Gly
65 70 75 80

Phe Tyr Glu Gly Leu Val Asp Thr Asn Gln Ser Thr Tyr Leu Leu Thr
85 90 95

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

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

<210> 7

<211> 3298

<212> ADN

<213> gen qutC từ *Aspergillus nidulans* cùng với trình tự bao quanh;

<400> 7

aagcttggtt tcaagtgatg atatatagtt atgaggatat aatatgaacc gaaagacgat	60
gtttcttggtg aatattttacg tgatagttgt ctgtctaata tggtacagca gtagaacaac	120
tacatacgggt cactacttac agccctagtc attccctccc tcgattgcct accatttata	180
cactttgaac atccacaggc ttgcctcctt ccatactctc cctaacagct tgaacaactc	240
tgagcgcctt caccctatct tcaacgccac agccaactcc tcgctcgcca tccctaccct	300

ttccactaac aacatcaaca aaatatcccc actgcgcatc aaacggcctc acatcagcat	360
ccttcactga aatctgctgc atcgccaact ccgtattcca gcctttctct tccccctgtc	420
cacaagaaac atgggtcatag ctccagcgcg tcatatcagg cacactgaga ctcgctctgg	480
ttccaagaat tcgataacag tcgcttgac tcttggaagg agcgggcgga atcgtaggat	540
tctcgcccgt tcctgtttca aagttcaacg gcgaaggcgt cgcgtcgcag atgagaaatg	600
tgcctactat ccagacgca aagcgcagcg tcacagcaca gccttcctcg gcggtatggt	660
ccgggttctg gcgcatgcg tgcaggagtg taccctccgc gtagacccta ctgacgggcc	720
caaacagaaa ctgcagcacg tcgatatcgt ggataagatt aatccccagc acgccgcct	780
tcttcttata tgcgcgcaa gaaccgagcg gcggcgcgaa gtaagaggcc ggcttcagaa	840
gtgtccagag gccgttact gcaacgacgg tgccgagtga gtctgtctct aacaaagact	900
ttgtggtttg gatgtacgga ttgaagcggc ggtggtgggc gatctggatg ttgatcttcg	960
catccttgcc tctctttctca tctttacatt tctgttcctt gacggtagcg aggaggtgct	1020
cggccgactc cagatcgtca ctgatcgggt tctcaaggag gatattgcgg attccgttct	1080
ccagcagctg gagcgtgacg tccacgtcg tgtgattggg cgtgctcacg atcgccgcgt	1140
ctggtttccc ggttgtctta ccgacaacgt ctaacataga cgtaatagaa tcatagcaag	1200
gaacgccaaa tgattctgcg accgggattg cagagggatga agggtaaca aaagcgatca	1260
gctgggttcg tgggtgtcgt tgcacggatt gtgcgtgacg gggccaata agtccggcac	1320
cgacgatgac aatgaggata ttcttgtcct tttccttgct gcagggcacc attgtgcatg	1380
tcggtggctg gaaataaaca gaacagggat atgggtcaagt cggagaatcg gtgcaggata	1440
gaccggctac ttgatgtagg acgacagtcg cgatctaccg agagcgtgag attcactgtg	1500
ggactgattt atgtaatttg aggcgcagca gacttaggga cttgaaatgt ggctgtctgt	1560
ggatgcattt gcggggatg gagtacagag tgcatacagc tgtgtatatg gagttcctta	1620
cggagagggt gacctggtat ggggagaacg ggcaaaatgc tcaccggca acctctcaaa	1680
gcgtttaccc ggtatactcc tctgatatca atatttccaa tcagcaccta tatcatcacg	1740
acgtctctct gaggattccg tagctaaccg ccctggatcc tacattaata aataagccat	1800
ttgctttttc tgctgcgagt gtgattctca atacgattac gtatcacatg cagattgcct	1860
ttacttcagc tgcatttgat cagccacagc tctaagagca aacataccct acctacctac	1920
ctacttcgcc tagggtacat aatcaccgcc atctcctcct cgatcagtct tcaactcaat	1980
cagctcattc attctattct taatataata tataccttta gatctccagc agagacccga	2040
agagtcggca attcaaaatg cccgcaaacc tcaaaatcgg tatcccaacc gtgtccctgt	2100

```

caaaaccggg cctgcactct cttgaccata agctccgctc ggccgctcat ggcttcgcgg 2160
ggatcgagct gtttattgat gacctctccc atttcgcctc atcgtcattc aatggctccc 2220
tcactcaagc ggcaaagtat atctcctcgc tcgccaagca acttaacctc acatttatct 2280
gcctgcaacc attcgggtttc tacgaggggc tgggtggacac aaatcagtcg acgtacctgc 2340
tcactgagaa actcccgcctc tggtttgcca tcgcccgcct tataggcaca gatctcatcc 2400
aaatccccgc aaatttcctc cagaatgacc ctgtcaccgg ggctgcacga acaagcggcg 2460
acataaggct tatcgctctca gatctgcaga cgatcgcaga tatcgggtgta aagcagggct 2520
tccgctttgt gtacgagggc ctctgctggc cgacgcatgt cgatacatgg gaagcagcgt 2580
ggaatgtcgt caagctgggt gatagagaga atttcgggat ctgcctggat agcttcaaca 2640
cgcggaaccc gcttccgtca ctgggaagac gccggatgct gagcaagccg tggccaagtc 2700
catggagacg ctccgttctc tcgtctccag tggagaactg gacatcagga aaatcttcta 2760
catccagctt gtcgatggcg agcgggtgtc ggcgccgctg gacgagaagc acccctttca 2820
tgtggagggc caacccccga agaattgagct ggagtcgcaa tgcgcgggta ttcccctgtg 2880
aagaggagag ggggtgggtat cttcctgtgt tggagatcgc gagggcggtc tttgaaatcg 2940
ggttcgaggg gtgggtgagt ctagagctgt tttcaaggac gtgtaatgat cccgatgtga 3000
acacgggtgg ggagcatgcg agacgtggga tggatagaag gaggaggggt gttgcggcgc 3060
taggactcga tgttgaggtg ccagcacgta actgtgaatg ttagcatgaa cggcaaggag 3120
aggggtggagg tgcaggtgca ggaggagctg gctgttcagc atcggctgta ggtagtggtg 3180
tcttgaaagg acgatagggt ttgatctaga gatttttatt ttgtctaatt actggtaatg 3240
atggcctcat gcacgctgtt gaacacgctg tacaacatca ctgttgaaga tgatacct 3298

```

<210> 8

<211> 502

<212> PRT

<213> AroY từ *Klebsiella pnemoniae* ATCC25597;

<400> 8

```

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

```

```

Gln His Asp Asn Gln Tyr Leu Glu Thr Asp His Pro Val Asp Pro Asn
20           25           30

```

```

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

```

```

Lys Arg Pro Thr Arg Ile Gly Pro Ala Met Met Phe Asn Asn Ile Lys
50           55           60

```

```

Gly Tyr Pro His Ser Arg Ile Leu Val Gly Met His Ala Ser Arg Gln

```

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

Val Asp Glu Asp Val Asp Ile Phe Asp Ser Asp Asp Ile Leu Trp Ala
 405 410 415

Met Thr Thr Arg Met Gln Gly Asp Val Ser Ile Thr Thr Ile Pro Gly
 420 425 430

Ile Arg Gly His Gln Leu Asp Pro Ser Gln Thr Pro Glu Tyr Ser Pro
 435 440 445

Ser Ile Arg Gly Asn Gly Ile Ser Cys Lys Thr Ile Phe Asp Cys Thr
 450 455 460

Val Pro Trp Ala Leu Lys Ser His Phe Glu Arg Ala Pro Phe Ala Asp
 465 470 475 480

Val Asp Pro Arg Pro Phe Ala Pro Glu Tyr Phe Ala Arg Leu Glu Lys
 485 490 495

Asn Gln Gly Ser Ala Lys
 500

<210> 9

<211> 5502

<212> ADN

<213> gen aroY của *Klebsiella pneumoniae* 342 và 2 kilobazo của trình tự ADN bao quanh;

<400> 9

gcgacgccga ctgggcgatc cgtgaactgc tggcgcgatat gaccacagcgt ctgcagggct 60

gtgaaaccat agaggatgtg attaaggtgg cggagctggt cgcgccgaac atcgccccga 120

cgatccccgg taaactgtat attctggata ccgatccatg gcagatgcgc tgcgtggcgc 180

agtggctgtc gcccgcggg gagacgacgt cctttgctcc cgacgactgc tgggcgatac 240

ggcggggact cagccatccg ccggtgcagg gtgagccga tatcacctgc tatcatctgc 300

cggaggcgca cgccggccag tcgctctgcg taccgctcat cgcccagggc gaagcgatcg 360

gtctgtgag ctttcagaac gtcaccgcca gtgacgcccc ttcccgggct tacctggagc 420

tgatggccga agcgctggg ctggcgctcg ccaatcagcg ttacgcagc gccctgctgg 480

aaaaagcgtt gttcgattcg ctgaccggcc tgcgtaaccg ccatcatctt gatgaagcgc 540

tgcaactcga gatggcgctg gcggtccata cccacacccc gctgagctgc ctgatgatcg 600

acatcgatca cttcaaagcc atcaatgacc gctacggcca tgaagccggg gatctgggtga 660

ttaagagcgt cgcgaccatt gtgcagcgcg cgggtgcgca tatcggcatg gctttccgct 720

acggcggcga ggagttttta gtgctgctcc ccgggattga cgaagccggg gcgcaccagt 780

gcgccagcga gatctacacc caggtgcaca atatgacgct gcgcgatggc ctgacggaga 840

taggccaggt ggatgtgtcg attggcatcg ccagctaccc gcagcacacc caaagcgaca 900

gcctgctgcg cgacgcggac gccgcgctgt accggggcgaa agagctgggc cgttcaagga 960

ttgtcagctt tggccgcctg aagacccgct aagcgggatt attgctcagc ggcattaagc	1020
agcgagataa ctttccgcac caccgccgaa cggaaatggc ggtggtaaac catgctcagc	1080
tcactctccg ccagccgatc ctgaagatcg atatacacca cgttatcaag gcgtaaggcc	1140
ctcgccgaga ccggcaccag cgccaccccc accccggcgg agaccaggct aatcatcgag	1200
gtgacatcgt taattcgctg caccacctgc ggcgtaaaac ccgcgacgcg acaggcgctca	1260
ataaatacct gctccagtcc ggtgccctgc ggatcgtcaa gcgagatcca gttgtcagtg	1320
cgcaacgagg ccagattgag cgccccacg cctgccagcg gatgctggtg ataaagcgcc	1380
aggcaaagtt tttcccgcac aaatggcctg accaccagcg cgtccggcgg tgacgccagc	1440
ggcgcgcgga tgatggcgat atccagacgc agatccagca gcgcttcgta gagcatttgc	1500
acatccccct gcaccagcga cagctcaatc cccggccagt cagcgcgcag ctgcgcgagg	1560
agccccggca gtttgctgtc atacatcgca ctggagacat agcccagatg caatcgcccc	1620
tgctcgctc gtgcggtgcg ctgggcgtcc aggaccgct gatcggccat ctccagcgcc	1680
agccgcgtct tctgcaagaa ggcctcgccc gcggcggtga gggtcaggcg ccggttagcg	1740
cgggagaaga gcaccacgcc caggcgctgc tcgagttgtt taatctgctg gctgagggcg	1800
ggctgggcga tatgtaaccg ctctgccgcc cgatgcatac gtagttcttc agcaacgacc	1860
acaaaatggc gtaacgctcg caaggacatg gccggactcc gcggagtaaa ttgataataa	1920
aaatgttatc aataaagcat gaatgatgca attgataacc attagcctgc gagcatactg	1980
tgcgcatcga cacgctaagg agaacatcat gaccgcaccg attcaggatc tgcgcgacgc	2040
tatcgcgctg ctgcaacagc atgacaatca gtacctcgaa accgatcatc cggttgaccc	2100
taacgctgag ctggccggcg tctatcgcca catcggcgcg ggcggcaccg tgaagcgccc	2160
cacgcgcatac ggcccggcga tgatgtttaa caatattaag ggctatccgc actcgcgcac	2220
tctggtgggc atgcacgcca gccgccagcg ggccgcgctg ctgctgggct gcgaagcctc	2280
acagctggcg ctggaggtag gcaaagcggc gaaaaaacgc gtcgcgccgg tggtcggtcc	2340
ggccagcagc gccccctgtc aggaacaggc ctttctggcc gacgatccgg attttgatth	2400
gcgcaccctg ctcccggcgc ccaccaacac cccgatcgac gccggtccct tcttctgcct	2460
gggcctggcg ctggccagcg atccccagca cgcctcgctc accgacgtca ccatccaccg	2520
cttgtgcgtc cagggccggg atgagctgtc gatgtttctc gccgctggcc gccatatcga	2580
agtgtttcgt cagaaagccg aggccgctgg caaacgcgtg ccgataacca tcaatatggg	2640
actcgatccg gctatctata tcggcgctg ctttgaagcg ccgaccacac gtttggtat	2700
aacgaactgg gcgtcgccgg tgcgctgctc cagcgtccgg tagagctggc acagggcgctc	2760

agcgtccccg agaaagccat cgctcgcgcc gagatcgtaa tcgaagggga actgctgccg	2820
ggggtacgcg tcagagaaga tcagcacaca acagcggcca tgcgatgccg gaatttcctg	2880
gttactgcgg cggcgccaat ccgtcgtgc cggtcattaa agtcaaagcg gtgaccatgc	2940
gaaacaatgc gattctgcag acgctggtag ggccgggcga agagcatacc accctcgccg	3000
gattgccaac ggaagccagt atctggaatg ctgtcgaggc tgctatcccc ggctttttac	3060
aaaatgtcta cgcccacacc gcgggcggcg gtaaattcct cgggatccctg cagggtgaaaa	3120
aacgccagcc cgccgacgaa gggcgtcagg ggcaggccgc gttgctggcg ctggcgacct	3180
attccgagct gaaaaatatc attctggtcg atgaagatgt cgatatcttt gacagcgacg	3240
atatcctgtg ggccatgacc acccgcatgc agggggatgt cagcatcacg acgatcccag	3300
gcattcgtgg tcaccagctg gatccttccc agaccccgcc ctacagcccc tcgatccgcg	3360
gagaggggat cagttgcaag acgattttcg attgcacggc gccgtgggcg ctaaaatcac	3420
acttcgagcg cgcaccgttt gccgatgtcg atccgcgtcc gtttgcgccg gagtattttg	3480
cccggctgga aaaaaaccac ggtcagtaaa atcaggtgat agccgccgga gcacggcgcc	3540
atcttccggg ccagcatcac ctgcagcggg tggctgacgc agggttagtt gatcgcgccg	3600
gagaggtctt ttttcacctg ctacgcgtgc tcgggggtca acacctggct cagctcgaag	3660
tagtatttca cacgataata ccgcacctgc tgggtccagct ggccaaaggc ggccagctgc	3720
tgtttgacct tagcgtcatc ccatttcccc gagtgaataa cgtctgccag ggcaccatcc	3780
tgatagccgc tgatttaatc tggcttacat tgttttcgaa tccctgacgc agcgccctgga	3840
ttttggcgac ctgctcttca ctacgcttca ggtgctggac gaccggatcc tgcgagacag	3900
acggtatatc ggccggaggtc gacgcctggc tggctgccgt aaagcaggtg gtcagcgcaa	3960
tggcgagcag ggtgttacgc aagcgagtat tcacagtga tgatccttca aaaaagaaaa	4020
tgagaggcga ttatcactgc gctaataaag actatctgta acaaagggtt aatttaaaac	4080
tggataaaaa aaggatggta agaaacagaa atcagatccc gggtcagcag cacagaaaga	4140
tatattcatc cttccagtaa cggccctgtc caatgatatc cccggcgccg ctgattaact	4200
gtttttgctt ttggtttcaa tcccctcaac gatcacatgg ctggtcaggg tatgaataga	4260
ttgcaacagc ccgggaaaag cggggtcgtt ttcttttatcc cagaagtaat ctttatccac	4320
tttgacgcaa tcgaagcgga agcgctccag cagcgggaag gacgtcgtcc gcggccaaaa	4380
tcatccagcc agaccgggca aagcgccgcc agcgtgctca gcgccgtcag ctcacgtccg	4440
gcgataaact cgtgaaagtt ctcatattatt tccagagcaa tgtgtttaca ggagcgcagg	4500
aaatcacaga gatagcgatc cgtcagaata aagtggctta ataaatcatc aatattcagc	4560
gatatcgggt tattatcgac ccgggcgaaa tcaaatacag aaagtaactc gatttgccga	4620

ataaatagag caagcttatc gcgttcggta agcgtggaga aagaaaacgt cgacgcagag 4680
 gtggtatatt gtgcgggtgc tataatatct ttagtaagca actcccacga gtggtagctg 4740
 ccgtcatcgc taatagcggg ctccagaacg aagcgataag aggtatatttt cacgttttct 4800
 tcaaccattt aaaaaatacc aaaaataaga aagggttaag catgtcatat attttcgcc 4860
 aacaaaaata gtttaaagt atcgataata atcattcgat agttaaaaac tatcaagata 4920
 taatttattg atcggtaa at tgaattaata taaattagcc actgccgtaa ctccctctga 4980
 aaagtcaatt aaaatattgt ttcaaaccag ccagttacca gagtattctg cgtaaagcct 5040
 ggtcgtctca cgctttgtgc tgccaggtaa aaaaagagag gggtataaaa aatgaaaaat 5100
 acaagccgcc agtttttagtc atatcattat gccgaatatg aataacgctg cgctgaggcg 5160
 ccgcttcgcc tggcatgcca tgagtcctca acaaaaaagt gtgactcagt cgacaaaacg 5220
 tcatattttc ccgctatcct gcagcgaaga agagtgaagt ggatgacagg cagtgaaaaa 5280
 aataaacgtg attccgctgg ggctgatgct attcatgctc atcgccagcg catggctggg 5340
 ccctgcgccg cggcacaccg gcagcatgca gtgcgtttgg tttgacgggg caatggtgag 5400
 ctgcctgccg aagcaacgac tgggcgaagg ctgcgccat catttactgg tcagacgata 5460
 aaccggtact cgccgggtgg tgttgaacag attatcgctg gc 5502

<210> 10

<211> 4629

<212> ADN

<213> gen catA từ *Acinetobacter baylyi* ADP1, bao gồm 410 bazo của trình tự ngược dòng và hai khung đọc mở xuôi dòng;

<400> 10

atctgctcga ccatagtaat gatcacatta tgagctaaat ttacttttta aaatttaaatt 60
 atattatata tatttgaatt ttattgtttt attttaattt ttagcttaga agtttttatt 120
 aagatttatt tttaaattag atgtcgaaaa aattagtata ccaaaaaagc atgaaaacat 180
 actctcttag gaattggagt cgccatgagt ttcagatata gttgatcagt atggaaggta 240
 tagaaacgac tatcgaaata aataagtttg tgggtgtgtga agcaaggtaa agctcaaggc 300
 tgaggcaaac caagcaaagg ttaattgaac cgatatgcac aacacattca acgatagcgt 360
 cgacagataa gtttatcaaa tgatgttttg gcgatttcaa ggagaaagcc atggaagtta 420
 aaatattcaa tactcaggat gtgcaagatt ttttacgtgt tgcaagcgga cttgagcaag 480
 aagggtggcaa tccgcgtgta aagcagatca tccatcgctg gctttcagat ttatataaag 540
 ccattgaaga tttgaatatc acttcagatg aatactgggc aggtgtggca tatttaaatc 600
 agctaggtgc caatcaagaa gctgggttac tctcgccagg cttgggtttt gaccattacc 660

tcgatatgcg tatggatgcc gaagatgccg cactaggtat tgaaaatgcg acaccacgta	720
ccattgaagg cccgctatac gtggcaggtg cgcctgaatc ggtaggttat gcgcgcatgg	780
atgacggaag tgatccaaat ggtcataccc tgattctaca tggcacgatac tttgatgcag	840
atggaaaacc tttacccaat gccaaagttg aaatctggca tgccaatacc aaaggctttt	900
attcacactt cgacccaaca ggcgagcagc aggcgttcaa tatgcgccgt agtattatta	960
ccgatgaaaa cggtcagtat cgcgttcgta ccattttgcc tgcgggttat ggttgcccac	1020
cagaaggtcc aacgcaacag ttgctgaatc agttgggccg tcatggtaac cgccctgcgc	1080
acattcacta ttttgtttct gccgatggac accgcaaact aactacgcaa attaatgtgg	1140
ctggcgatcc gtacacctat gacgactttg cttatgcaac ccgtgaaggc ttggtggttg	1200
atgcagtgga acacaccgat cctgaagcca ttaaggccaa tgatgttgaa ggcccatcgc	1260
ctgaaatggc tttcgatcta aaattgacgc gtttggttga tgggtgtgat aaccaagttg	1320
ttgatcgtcc acgtctagcg gtgtaataca ccaaaatggc tcaaaattat caggcgagtg	1380
atcatgatca ctggcctgtt tttatttcag ggaaggggtg agacaattac gtggacaatc	1440
aaatcattca ggaaaccgta gataaaattt taagcgtatt gccgaatcag gctgggcaat	1500
tggcacgctt ggttcgtctg atgcagtttg cttgtgacct caccattacc gtcattggta	1560
aatataatca tggtaaaagc cgactactca atgagctgat cgggacagat attttttctg	1620
ttgccgataa acgagagacg attcaactgg ccgaacataa acaagatcag gtgcgttggt	1680
tggatgcacc cggactcgat gcagatgttg cggcagtgga tgatcgtcat gcttttgaag	1740
cagtctggac acaggcagat attcgctttt ttgtgcattc agtccgagaa ggcgaactcg	1800
atgcaactga gcatcatctt ttacaacaac ttattgaaga tgcagaccat agccggcgcc	1860
aaaccatact ggtcttgacc cagatagatc agataccgga tcagacaatt ttaaccaga	1920
ttaaaacctc aattgcacag caggtaccca aactcgatat ttgggctgtt tcggccactc	1980
gccaccgtca gggatttgaa aatggaaaaa cttgtctgat cgaaaaaagt ggaatcggcg	2040
cgttacgaca tacacttgag caggcacttg ctcaggttcc atctgcacga acgtatgaaa	2100
agaatagatt gctgtctgac ttgcatcatc aacttaagca gttattactc gatcaaaaac	2160
atgtacttca gcaactacaa cagacacagc agcagcaatt gcatgacttt gatacaggac	2220
tcatcaacat actcgataag attcgagtag atcttgagcc cattgtaaat atagatggtc	2280
aagaccaagc actcaatcca gattcatttg ccacgatgtt taaaaataca gcagccaagc	2340
agcaacgtgc caaagtgcag attgcttact cacgtgcctg tatcgagatc aatagccatc	2400
tcatacgtca cgggtgtggtg ggtttacccg cagagcaaca aaccacgata aaaagtattg	2460
atacgggtcat tgttgcggtt tttggaattt cagtgaaatt tcgcgatcag ctacgtgcat	2520

tgttttatac cgataccgaa cgacaacgct tgcaaagaga gtttcgattt tattttgaaa	2580
agtcagcagg ccgaatgatt ttagctgcca agattgagca gacaatgcgg cagcaagggg	2640
gtattcaaaa tgcaatgatg gcgttgcaac agatggagag tgcagcatga ccagcggcgg	2700
acacattcaa ttgtttatcg aacacacccg gcagattgcg actgcccag gggatataca	2760
gttggcattg caatcgatgc agcaatggcg cgaagcattt gctacagcat taaaacaaaa	2820
tacctttgat ttaacgggct ggtcaccgca gacaaagatc gccaatcaac tcaagcaatt	2880
taaccataag cttacaacgc atgtatcgaa ttgggatacc gaatggcata cttttagtgc	2940
tgctcaatcg gttgcagaag tatttcatga tcgggtgatg ttgcttgat tcggtaagtt	3000
taatgccgga aagagttcat tgtgtaactt actggccgaa tgctttcggt ctcacgaaca	3060
aaccgtgcaa tattttcatg ttcaaaatga acagatattt tataccgaat ctcacttacg	3120
cgaaggtgca accgagacga cagcgcaact acagggcgta tgtctgggtg aaaaacttat	3180
tttgctagat acaccagggt tgcattctgg tactcagaaa aatgcagcgc tcacacaaaa	3240
atttatcgac agtgcagatg gtgtgctgtg gctcagtagc gcaacttcac cgggtcaggt	3300
gcaagagcta gatgcactgg ggcgcgagtt aaagcgtcat aaacctttat ttctgtttat	3360
taccggaagc gattttgtcg aagaagatga aattgatggg gagctatgta cagtgccttg	3420
caataaaaaat tcagaacaac gtgcgttgca agagtctgat gtattgatgc gtgcgaaaga	3480
aaaactgcac atatgcaagt ggatgtgagt ttattaaagc cgcccgtgtc cgtttcaact	3540
caaatggcgc gtgaagcaga tatgaacca caagccatga acgaggctgg ttttgagcga	3600
ttatttgagc cacttttggc tcttattgag cctgctttgc gctataagca gcgtaaacct	3660
gccgaagtat tgttgcattt tttgcaagaa catatcattg aaggtttaag gttttacctg	3720
caacccgatc tagagcaa atacaacaggac ctcaaacagg ctcaagatga ttacgacag	3780
ctacacaccg atttagccga ggcagtctgg cgtagcgtat tgcctgagct accacaactt	3840
cttgagcaac atgcaagtac acaaaatatt gatgccgtag tgaacagttt gaacgagtgg	3900
ataaacgtcg cattcgaaca acagcttgca attcagcttg atgcttatgg tttaaatttg	3960
gattcgctta gcaagatcga aaaaaccgaa aaaatgcagt atgaacgcat tgcgggaatg	4020
gtggtgcatg atggcttgta cacgactctc acgcagcaga ttcaacaagc tgtcaaagct	4080
tctacgagtg aattgattga tcagtgtcag gctcaacttg agcagtcaat caaacatgtt	4140
caaacactcg atgaaacctt catcgattac agcgcagcac tcgatcaact cagccaagcg	4200
ctacgcattg aataaagagc agtaaatttt tcagacatat ttatttcgat gagtggcctg	4260
atatgggtgcg ttgcaaacac ctctgtaca caggcgagaa ttttaggaat gtaattactg	4320

tgggtccatat ttcgcaccgc gaggtaaaatt gggctatagg catcatcatc taaaattgga 4380
 atataaagta gattcttcac cccaatatcc atggcagacg ccggtacgat gcagacgcct 4440
 tcacctgctg ccaccaagcc gagggccagt tgaatttctc gaatttcggt gaggttggat 4500
 ggtactaggc ctagttcggg aaagagtgac tgaataaagg tcgcaaaaatt gggcttttga 4560
 gagactgggt acagcagcat cggttcatca ataatttgag agagatgaac ccctgttgct 4620
 gcaaaactga 4629

<210> 11
 <211> 311
 <212> PRT
 <213> Cata từ *Acinetobacter baylyi* ADP1;

<400> 11

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

Ile His Tyr Phe Val Ser Ala Asp Gly His Arg Lys Leu Thr Thr Gln
225 230 235 240

Ile Asn Val Ala Gly Asp Pro Tyr Thr Tyr Asp Asp Phe Ala Tyr Ala
245 250 255

Thr Arg Glu Gly Leu Val Val Asp Ala Val Glu His Thr Asp Pro Glu
260 265 270

Ala Ile Lys Ala Asn Asp Val Glu Gly Pro Phe Ala Glu Met Val Phe
275 280 285

Asp Leu Lys Leu Thr Arg Leu Val Asp Gly Val Asp Asn Gln Val Val
290 295 300

Asp Arg Pro Arg Leu Ala Val
305 310

<210> 12
<211> 1461
<212> ADN
<213> quiC từ Acinetobacter sp. ADP1;

<400> 12
atgaaattaa cttctttacg cgtatcttta ttggcgctgg gcttggtaac atcaggtttt 60
gctgcggcag aaacttatac tgtagatcgt tatcaggatg atagtga aaa aggctctttg 120
cg ttgggcaa ttgaacaatc taatgcaaat agcgcacaa agaatcagat tctgattcag 180
gctgttggtg aggcacctta tgtgatcaag gtggataaac cg ttaccacc gattaaatca 240
tctgtaaaaa ttattggtac agaatgggat aaaacgggcg aatttattgc gattgatggt 300
tcaaactata tcaagggcga aggcgaaaaa gcatgtccag gtgcaaattcc aggacaatat 360
ggtaccaatg ttcgtaccat gactttacca gg tttggttc tacaagatgt caatgggtgtg 420
accctgaaag gtcttgatgt tcatcgcttc tgtattggtg tactggtaaa tcgttcaagc 480
aataatttga ttcagcataa ccgtatttca aataattacg gtggcgctgg tgcattgatc 540
acgggtgatg atggtaaagg taaccaacg tctaccacca ccaataacaa caaagtattg 600
gataatgtgt ttattgacaa tggcgatggt cttgaactga cgcgtggagc agcattcaac 660
ctgattgcta acaatctggt tacatcgacc aaagccaatc cagagccgtc tcaaggcatt 720
gaaattcttt gggggaatga caatgcagtg gtgggtaaca aatttgaaaa ctattcagat 780
ggtctacaaa tcaactgggg taaacgtaat tacatcgctt ataacgaatt gaccaataac 840
tctttggggt tcaatcttac aggtgatgga aacatcttcg atagtaacaa agtgcattggc 900
aatcgtattg gtatcgcaat tcgttctgaa aaagatgcaa atgcacgtat cacacttacc 960
aaaaatcaga tttgggataa tggtaaagat atcaaacgct gtgaggctgg tggttcatgt 1020

gttccaaacc aacgttttagg tgcaattgta tttggtgttc ctgcgcttga gcatgaaggt 1080
 tttgtaggct ctcggtggtg cgggtgtagtc attgaacctg caaaattaca aaaaacatgt 1140
 acacagccaa atcaacaaaa ctgtaatgcc attccgaacc aaggatttca ggcacctaaa 1200
 ctgactgtca gtaaaaaaca acttacagtt gaagttaaag gaacaccaaa ccagcggttac 1260
 aacgtagaat tttttggaaa tcgtaatgca tcttcttccg aagctgagca atatttaggt 1320
 tcaattgttg tagtgacaga tcatcaaggt cttgcaaaag caaactgggc accaaaagtc 1380
 agcatgccat ctgttactgc gaatgtaact gatcacttgg gcgccacttc agagttaagt 1440
 tctgcagtga aaatgagata a 1461

<210> 13

<211> 1461

<212> ADN

<213> gen quiC được tối ưu hóa codon từ *Acinetobacter* sp. ADP1;

<400> 13

atgaaactga ccagcctgcg tgttagcctg ctggcactgg gtctgggttac cagcgggtttt 60
 gcagcagcag aaacctatac cgttgatcgt tatcaggatg atagcgaaaa aggtagcctg 120
 cgttgggcaa ttgaacagag caatgcaaat agcgcacaaag aaaaccagat tctgattcag 180
 gcagttggta aagcaccgta tgttatcaaa gttgataaac cgctgcctcc gattaaaagc 240
 agcggttaaaa tcattggcac cgagtgggat aaaaccggtg aatttattgc aattgatggc 300
 agcaactata tcaaaggcga aggtgaaaaa gcatgtccgg gtgcaaatacc gggtcagtat 360
 ggcaccaatg ttcgtaccat gaccctgcct ggtctggttc tgcaagatgt taatgggtgtt 420
 accctgaaag gtctggatgt tcatcgtttt tgtattgggtg ttctgggttaa tcgcagcagc 480
 aataacctga ttcagcataa tcgtatcagc aacaattatg gtggtgccgg tgttatgatt 540
 accggtgatg atggtaaagg taatccgacc agcaccacca ccaataataa caaagttctg 600
 gataacgtgt tcatcgataa tggatgatgg ctggaactga cccgtgggtgc agcatttaat 660
 ctgattgcaa ataacctgtt taccagcaca aaagccaatc cggaaccgag ccaggggtatt 720
 gaaattctgt ggggtaatga taatgccgtg gtgggtaaca aattcgaaaa ctattcagat 780
 ggctgcaaaa tcaattgggg taaacgtaac tatatcgctt ataacgaact gaccaataac 840
 agcctggggt tcaatctgac aggtgatggg aacatttttc acagcaataa agtgcattgt 900
 aaccgtattg gtattgccat tcgtagtga aaagatgcca atgcacgtat taccctgacc 960
 aaaaatcaga tttgggataa cggcaaagat atcaaacgtt gtgaagccgg tggtagctgt 1020
 gttccgaatc agcgtctggg tgcaattgtt tttggtgttc cggcactgga acatgaaggt 1080
 tttgttggtg gccgtggcgg tgggtgtgtt attgaaccgg caaaactgca aaaaacctgc 1140

acccagccga accagcagaa ttgtaatgca attcctaatac aggggtattca ggcaccgaaa 1200
 ctgacagtta gcaaaaaaca gctgaccgtt gaagttaaag gcaccacctaa tcagcgttat 1260
 aatgtggaat tttttggcaa tcgtaatgcc agcagcagcg aagcagaaca gtatctgggt 1320
 agcattgttg ttgttaccga tcatcagggt ctggcaaaag caaattgggc tccgaaagtt 1380
 agcatgccga gcgttaccgc aaatgtgaca gatcatctgg gtgcgaccag cgaactgagc 1440
 agcgcagtta aaatgcgtta a 1461

<210> 14
 <211> 486
 <212> PRT
 <213> QuiC từ Acinetobacter sp. ADP1;

<400> 14

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

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

<210> 15
 <211> 9462
 <212> ADN
 <213> Plasmit pAC21;

<220>
 <221> dấu hiệu misc

<222> (5924)..(5924)

<223> n là a, c, g, hoặc t

<400> 15

```

gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtgggtcag actggaaaat cagagggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgccca taaaacgccc      180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcac gaatccataa      240
aaggcgccctg tagtgccatt taccgccatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt aggggtttta ggtctgtttt      420
gtagaggagc aaacagcgct tgcgacatcc ttttgaata ctgcggaact gactaaagta      480
gtgagttata cacagggctg ggatctatcc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggtcta gcggaattta      600
cagaggggtct agcagaattt acaagttttc cagcaaaggt ctagcagaat ttacagatac      660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata      720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa      780
atgaactagc gattagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct      840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt      900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat      960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta      1020
aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat      1080
tagtttttag tgaagagata ttgccttatc ttttccagtt aaaaaaattc ataaaatata      1140
atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggttat      1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat      1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg      1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggttgata      1380
agcgaggccg cccgactgat acgttgatgt tccaagttga actagataga caaatggatc      1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca      1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg      1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg      1620
atctcaatgg ttcgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac      1680
tggtctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca      1740

```

agactaaca	acaaaagtag	aacaactggt	caccgttaga	tatcaaaggg	aaaactgtcc	1800
atatgcacag	atgaaaacgg	tgtaaaaaag	atagatacat	cagagctttt	acgagttttt	1860
ggtgcattta	aagctgttca	ccatgaacag	atcgacaatg	taacagatga	acagcatgta	1920
acacctaata	gaacaggtga	aaccagtaaa	acaaagcaac	tagaacatga	aattgaacac	1980
ctgagacaac	ttgttacagc	tcaacagtca	cacatagaca	gcctgaaaca	ggcgatgctg	2040
cttatcgaat	caaagctgcc	gacaacacgg	gagccagtga	cgcctcccgt	ggggaaaaaa	2100
tcatggcaat	tctggaagaa	atagcgcttt	cagccggcaa	acctgaagcc	ggatctgcca	2160
ttctgataac	aaactagcaa	caccagaaca	gcccgtttgc	gggcagcaaa	acccgttatg	2220
cttgtaaacc	gttttgtgaa	aaaattttta	aaataaaaaa	ggggacctct	agggtcccca	2280
attaattagt	aatataatct	attaaaggtc	attcaaaagg	tcatccaccg	gatcaattcc	2340
cctgctcgcg	caggtctggg	gccaaagctct	cgggtaacat	caaggcccga	tccttgaggc	2400
ccttgccctc	ccgcacgatg	atcgtgccgt	gatcgaaatc	cagatccttg	acccgcagtt	2460
gcaaaccctc	actgatccgc	atgcccgttc	catacagaag	ctgggcgaac	aaacgatgct	2520
cgccttccag	aaaaccgagg	atgcgaacca	cttcatccgg	ggtcagcacc	accggcaagc	2580
gccgcgacgg	ccgaggtctt	ccgatctcct	gaagccaggg	cagatccgtg	cacagcacct	2640
tgccgtagaa	gaacagcaag	gccgccaatg	cctgacgatg	cgtggagacc	gaaaccttgc	2700
gctcgttcgc	cagccaggac	agaaatgcct	cgacttcgct	gctgccaag	gttgccgggt	2760
gacgcacacc	gtggaaacgg	atgaaggcac	gaaccagtg	gacataagcc	tgttcggttc	2820
gtaagctgta	atgcaagtag	cgtatgcgct	cacgcaactg	gtccagaacc	ttgaccgaac	2880
gcagcggtgg	taacggcgca	gtggcggttt	tcatggcttg	ttatgactgt	ttttttgggg	2940
tacagtctat	gcctcgggca	tccaagcagc	aagcgcgtta	cgccgtgggt	cgatgtttga	3000
tgttatggag	cagcaacgat	gttacgcagc	agggcagtcg	ccctaaaaca	aagttaaaca	3060
tcatgagggg	agcggtgatc	gccgaagtat	cgactcaact	atcagaggta	gttggcgctc	3120
tcgagcgcca	tctcgaaccg	acgttgctgg	ccgtacattt	gtacggctcc	gcagtggatg	3180
gcggcctgaa	gccacacagt	gatattgatt	tgctggttac	ggtgaccgta	aggcttgatg	3240
aaacaacgcg	gcgagctttg	atcaacgacc	ttttggaaac	ttcggcttcc	cctggagaga	3300
gcgagattct	ccgcgctgta	gaagtcacca	ttgttgtgca	cgacgacatc	attccgtggc	3360
gttatccagc	taagcgcgaa	ctgcaatttg	gagaatggca	gcgcaatgac	attcttgccg	3420
gtatcttcga	gccagccacg	atcgacattg	atctggctat	cttgctgaca	aaagcaagag	3480
aacatagcgt	tgcccttggt	ggtccagcgg	cggaggaact	ctttgatccg	gttcctgaac	3540

aggatctatt	tgaggcgcta	aatgaaacct	taacgctatg	gaactcgccg	cccgactggg	3600
ctggcgatga	gcgaaatgta	gtgcttacgt	tgtcccgcat	ttgggtacagc	gcagtaaccg	3660
gcaaaatcgc	gccgaaggat	gtcgctgccg	actgggcaat	ggagcgccctg	ccggcccagt	3720
atcagcccgt	catacttgaa	gctagacagg	cttatcttgg	acaagaagaa	gategcttgg	3780
cctcgcgcg	agatcagttg	gaagaatttg	tccactacgt	gaaaggcgag	atcaccaagg	3840
tagtcggcaa	ataatgtcta	acaattcggt	caagccgacg	ccgcttcgcg	gcgcggttta	3900
actcaagcgt	tagatgcact	aagcacataa	ttgctcacag	ccaaactatc	aggtcaagtc	3960
tgcttttatt	atttttaagc	gtgcataata	agccctacac	aaattgggag	atatatcatg	4020
aaaggctggc	tttttcttgt	tatcgcaata	gttggcggaag	taatcgcaac	atccgcatta	4080
aatctagcg	agggttttac	taagctgata	cgggtggatga	ccttttgaat	gacctttaat	4140
agattatatt	actaattaat	tggggacctt	agaggtcccc	ttttttatct	taaaaatttt	4200
ttcacaaaac	ggttttacaag	catacggttg	ccgattcatt	aatgcagctg	gcacgacagg	4260
tttcccgact	ggaaaagcggg	cagtgcgcgc	aacgcaatta	atgtgagtta	gctcactcat	4320
taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tgttggtgtg	aattgtgagc	4380
ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgccaagct	tgcatgcctg	4440
caggtcgact	ctagaggatc	cccccgccg	ccgacagagt	aatagggttt	acttaatagc	4500
tcttcctgtc	ccttccaggc	agtgatccgc	attccgttct	catggcgagg	caacatttcg	4560
ggatggaaga	taatgttctt	tgctacagga	aatcaacaa	tatgcgcacc	agatgccact	4620
ggcagccgcc	cgctgcgcgt	tactaactct	ataaatgcag	ggatctcatc	aatgacaaca	4680
tcctgcggac	tgtttcctgc	cagtcccatg	atgatggcga	catccgtggc	atggcctttg	4740
cccgtcagtg	acaacgacct	gtacagatcg	accacaatat	ggctcgtcgc	ggttaataag	4800
ccgctacttt	ccagccgatc	aataaaactt	tttccggcat	tcattggccc	cacgggtatgc	4860
gaactggagg	gaccaatccc	aattttgaaa	atatcgaatg	cactaatcat	gtgacggaag	4920
atcacttcgc	agaataaata	aatcctgggt	tcctgtttga	taccgggaag	ccctgggcca	4980
acttttggcg	aaaatgagac	gttgatcggc	acgtaagagg	ttccaacttt	caccataatg	5040
aaataagatc	actaccgggc	gtattttttg	agttatcgag	attttcagga	gctaagggaag	5100
ctaaaatgga	gaaaaaatc	actggatata	ccaccgttga	tatatcccaa	tggcatcgta	5160
aagaacattt	tgaggcattt	cagtcagttg	ctcaatgtac	ctataaccag	accgttcagc	5220
tggtatattac	ggccttttta	aagaccgtaa	agaaaaataa	gcacaagttt	tatccggcct	5280
ttattcacat	tcttgcccgc	ctgatgaatg	ctcatccgga	attccgtatg	gcaatgaaag	5340
acggtgagct	ggtgatatgg	gatagtgttc	acccttgttt	caccgttttc	catgagcaaa	5400

ctgaaacggtt ttcacgcgtc tggagtgaat accacgacga tttccggcag tttctacaca	5460
tatattcgca agatgtggcg tgttacggtg aaaacctggc ctatttccct aaagggttta	5520
ttgagaatat gtttttcgtc tcagccaatc cctgggtgag tttcaccagt tttgatttaa	5580
acgtggccaa tatggacaac ttcttcgccc ccgttttcac catgggcaaa tattatacgc	5640
aaggcgacaa ggtgctgatg ccgctggcga ttcaggttca tcatgccgtt tgtgatggct	5700
tccatgtcgg cagaatgctt aatgaattac aacagtactg cgatgagtgg cagggcgggg	5760
cgtaattttt ttaaggcagt tattggtgcc cttaaacgcc tgggtgctacg cctgaataag	5820
tgataataag cggatgaatg gcagaaattc gaaagcaaat tcgacccggt cgtcggttca	5880
gggcagggtc gttaaatagc cgcttatgtc tattgctggt ttantcggta cccggggatc	5940
gcggccgcgg accggatccc atcacatata cctgccgttc actattattht agtgaaatga	6000
gatattatga tatttttctga attgtgatta aaaaggcaac tttatgccca tgcaacagaa	6060
actataaaaa atacagagaa tgaaaagaaa cagatagatt ttttagttct ttaggccgt	6120
agtctgcaaa tccttttatg attttctatc aaacaaaaga ggaaaataga ccagttgcaa	6180
tccaaacgag agtctaatag aatgagggtcg aaaagtaaat cgcgcgggtt tgttactgat	6240
aaagcaggca agacctaaaa tgtgtaaagg gcaaagtgtg tactttggcg tcaccctta	6300
catatttttag gtctttttttt attgtgcgta actaacttgc catcttcaaa caggagggt	6360
ggaagaagca gaccgctaac acagtacata aaaaaggaga catgaacgat gaacatcaaa	6420
aagtttgcaa aacaagcaac agtattaacc tttactaccg cactgctggc aggaggcgca	6480
actcaagcgt ttgcgaaaga aacgaaccaa aagccatata aggaaacata cggcatttcc	6540
catattacac gccatgatat gctgcaaatc cctgaacagc aaaaaaatga aaaatatcaa	6600
gttcttgaat tcgattcgtc cacaattaaa aatatctctt ctgcaaaagg cctggacgtt	6660
tgggacagct ggccattaca aaacgctgac ggcactgtcg caaactatca cggctaccac	6720
atcgtctttg cattagccgg agatcctaaa aatgcggatg acacatcgat ttacatgttc	6780
tatcaaaaag tcggcgaaac ttctattgac agctggaaaa acgctggccg cgtctttaaa	6840
gacagcgaca aattcgatgc aaatgattct atcctaaaag accaaacaca agaatggtca	6900
ggttcagcca catttacatc tgacggaaaa atccgtttat tctacactga tttctccggt	6960
aaacattacg gcaaacaac actgacaact gcacaagtta acgtatcagc atcagacagc	7020
tctttgaaca tcaacggtgt agaggattat aaatcaatct ttgacggtga cggaaaaacg	7080
tatcaaaatg tacagcagtt catcgatgaa ggcaactaca gctcaggcga caaccatacg	7140
ctgagagatc ctactacgt agaagataaa ggccacaaat acttagtatt tgaagcaaac	7200

actggaactg aagatggcta ccaaggcgaa gaatctttat ttaacaaagc atactatggc	7260
aaaagcacat cattcttccg tcaagaaagt caaaaacttc tgcaaagcga taaaaaacgc	7320
acggctgagt tagcaaacgg cgctctcggt atgattgagc taaacgatga ttacacactg	7380
aaaaaagtga tgaaaccgct gattgcatct aacacagtaa cagatgaaat tgaacgcgcg	7440
aacgtcttta aaatgaacgg caaatggtac ctgttctactg actcccgcgg atcaaaaatg	7500
acgattgacg gcattacgtc taacgatatt tacatgcttg gttatgtttc taattcttta	7560
actggcccat acaagccgct gaacaaaact ggccttgtgt taaaaatgga tcttgatcct	7620
aacgatgtaa cctttactta ctcacacttc gctgtacctc aagcgaaagg aaacaatgtc	7680
gtgattacaa gctatatgac aaacagagga ttctacgcag acaaacaaatc aacgtttgcg	7740
ccgagcttcc tgctgaacat caaaggcaag aaaacatctg ttgtcaaaga cagcatcctt	7800
gaacaaggac aattaacagt taacaaataa aaacgcaaaa gaaaatgcc aatcctatt	7860
ggcattttct tttatttctt ccattttaa at ggatgcatgc gctagcggag tgtatactgg	7920
cttactatgt tggcactgat gaggggtgtca gtgaagtgtc tcagcctcgt gagcgggacg	7980
gtcgtaaggc cgttccgctc cacttctactg aacggcaatc cgaggggtgtg gatccaatta	8040
aggccacgct gtcattttaa ttcggttttt ccagttcaaa tgcaattgcc ttcaatgcac	8100
cttcgtagct gtggtgagcc agcgggtgctg gctctcccc atttacggat aagaatgcat	8160
tttccgagtt aataccgtcg gcaataacctg acattaatac ttcacagtcg ctggcatcga	8220
gtacggaaaa cttaatcgaa gacgaaccac agttaataac caaaacaacc ggaaattcat	8280
tcactctttt tctcatcctg agttacggat taaaacagtt tgtatacgat gttcaggatg	8340
gtcagcagac caatcacggc aacaaacacg ttatccagac gaccacggta tttcgccaga	8400
gacggcgctt tacggatggc atacatcggc aacaggcaca gcagggatgc gataatcggc	8460
gcgcccattg cttcaatcag gtcgaggatg ttcgggttgg cgtaggcaac aaccagggtg	8520
gagcccatga tgaagatcat gctgagagta ttcagtttac ccagcgacac tttgggtttg	8580
tcacctttat aaccgaactt cagaatcaga ccattcaagc cttccagcgt cccagatag	8640
tgaccgaaga aagatttgaa gatagccacg agtgcgatga tggaagccgc atattccagt	8700
gtaatcgca acgttgtttt ggtaccggtc atggacgcaa agtggttagc cagataagaa	8760
agcactggaa tattctgcgc tttggcttcc gccatgttgg ccggagacag agtaaacagg	8820
cagctaaagg caaagaacat caccactgca accatcagca tgctggcacg agaaatgatt	8880
tggaacatt tacgttcggc gaagtcgca ccgaagtctt tctcactc ttcacgttta	8940
gaaaccacga aggaagagac gattggcgag aagttaaagg agaaaaccat gatggaaatc	9000
cccagccaga cagtgatcag gataccgtca tgaccggtta acgacagcga accgagggtca	9060

```

acctggtcga taactgcaga gttccagtaa gggatcagcg acaaagaaat cagcaccagg      9120
ctggcgataa acggccatac caggtagctc aggggtaccga gctcgaattc actggccgtc      9180
gttttacaac gtcgtgactg ggaaaaccct ggcgttacct aacttaatcg ccttgacgca      9240
catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cctttcccaa      9300
cagttgcgca gcctgaatgg cgaatggcgc ctgatgcggt atttttctct tacgcatctg      9360
tgcggtatatt cacaccgcat atgggtgcaat ctgagtacaa tctgctctga tgccgcatag      9420
ttaagccagc cccgacaccc gccaacaccc gctgacgaat tc                          9462

```

```

<210> 16
<211> 9430
<212> ADN
<213> Plasmid pAC19;

```

```

<400> 16
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agaccgcga taaaacgccc      180
tgagaagccc gtgacgggct tttcttgat tatgggtagt ttccttgcat gaatccataa      240
aaggcgctg tagtgccatt taccctcatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcagggt cctgatggtc ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt      420
gtagaggagc aaacagcggt tcgcacatcc ttttgtaata ctgcggaact gactaaagta      480
gtgagttata cacagggtcg ggatctattc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggctca gcggaattta      600
cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac      660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattggtata      720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa      780
atgaactagc gattagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct      840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt      900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatggtgtat      960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggta      1020
aaggccttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat      1080
tagtttttag tgaagagata ttgccttatt ttttccagtt aaaaaattc ataaaatata      1140
atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtggttat      1200

```

taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgttctca tggctcaccg aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactgtt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccgt ggggaaaaaa	2100
tcattggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtcccca	2280
attaattagt aatataatct attaaaggtc attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctc cgggtaacat caaggcccga tccttgagac	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgcatgt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgcttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgt gctgccaag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgt cagcgaactg gtccagaacc ttgaccgaac	2880
gcagcggtgg taacggcgca gtggcggtt tcatggcttg ttatgactgt ttttttggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000

tggttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgagggga agcgggtgatc gccgaagtat cgactcaact atcagaggta gttggcgtca	3120
tcgagcgcca tctcgaaccg acgttgctgg cegtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttgtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgag	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcg agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aatctagcg agggctttac taagctgatc cggtgatga ccttttgaat gacctttaat	4140
agattatatt actaattaat tggggaccct agaggcccc ttttttattt taaaaatttt	4200
ttcacaaaac ggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccgaact ggaaagcggg cagtgcgcgc aacgcaatta atgtgagtta gctcactcat	4320
taggcacccc aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc	4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgcattgctg	4440
caggtcgact ctagaggatc cccccgccg ccgacagagt aataggtttt acttaatagc	4500
tcttcctgtc ccttcaggc agtgatccgc attccgttct catggcgagg caacatttcg	4560
ggatggaaga taatgttctt tgctacagga aaatcaacaa tatgcgcacc agatgccact	4620
ggcagccgcc cgctgcgcgt tactaactct ataaatgcag ggatctcatc aatgacaaca	4680
tcctgcggac tgtttcctgc cagtcccatg atgatggcga catccgtggc atggcctttg	4740
cccgtcagtg acaacgacct gtacagatcg accacaatat ggctcgtcgc ggttaataag	4800
ccgctacttt ccagccgatc aataaaaactt tttccggcat tcattggccc cacgggtatgc	4860

gaactggagg gaccaatccc aattttgaaa atatcgaatg cactaatcat atccacaccc	4920
tcggattgcc gttcagtcaa gtggagcgga acgaccttac gaccgtcccg ctcacgagggc	4980
tttacgcact acgtactgcg atggcttcaa tttccagcgg gagggcggat ccactaatac	5040
aaaatatatc aaaagttaat aataatatta ttcttactta agactttttt gtcttcattt	5100
tttagtaaaa aatataaaaa aggccacctc ccgattttat cggaaggcag cctcttaaat	5160
tcagttcata atattaaaaa atattattca acttcagaat atttgttggc ataggcagct	5220
gccgcaccca acagtccagg ctgcggataa gtaatcaact taaccggaat cttggacatg	5280
acgcgttcaa agcgtccttt tgaaacaaag cgctgacgga aaccagattc tggcaaattg	5340
gaagcgatac gaagaccgac accaccgcca ataacaacac tggttgcacc ctgtgccaaa	5400
gcaagatcac cagcgatagc gccaaaggctc aagcagaagc gatccaaagc ggcttcagca	5460
aggttgtctt taccttccaa agccatctgc cataatttaa tatcatccag caagctgaac	5520
ggaacgcctt caatggcagc cagtgtctcg tagatattac caagaccggg gccagaaata	5580
atgcgttcga tagaaacgcg gcggaaacgt tcacgtaaac gtgccagaat tttgtcttca	5640
agtctgtcaa gcggagcaaa gtcgatatga ccgccttcag tttcgatgac gaaataacgg	5700
ccttcagtcc gcaacagatg ggcaacaccc aagcccgttc ccggaccaag aatagtata	5760
acaccatcgc taggaagcgc ttcacagga ccacaaatat gatccagata agaagaatcc	5820
atatgcgcaa ccgcgtgggc aaccgcgccg aagtcattga tcagaacatg cgtatcgatg	5880
tccagctttt cattcagagt agctggctt aatacccaag gggtattggg aagttttaa	5940
acttcacat gaaccgggccc agcccatgca atagctgcgg cacgtggcag aggacgaccc	6000
agtttttcac cgaaacgttc ccaagctaac tgcaagctag catgttctgc cgttttaaaa	6060
gttgtttctt ctccaagaga aagaacccga ccattgctta cttccgcaat agagaaacgc	6120
gcatgcgttc caccgatgtc aatcgcaaca atttcataa taattccttt ctgaaatcag	6180
aaggctaccc aacaggtaaa ataagtccgc ccgctttata ccatcgttgt aaacaaaaag	6240
tataattggg taagacttat ctaaaaaaga caaaaggatt cagccaaagc aagtttaact	6300
acttctggga gcgccacatc tctctgattt catccaggct ccgaccttg gtttccggca	6360
cgaagcgagc aacaatcaag ccacctaaga tacttaatgc tgcgaaaacg agataggaga	6420
aaccgtgggt gaaagtctga ttcaatgctg gagaaccatc ggcaacctta aacaggaagt	6480
taaccaagat attagctaac cattgtccgg taacagcgat aggcatagct gcgccttga	6540
tggaactcgg gaacatttct gacagaacaa cccagcagac agggcccat gacataccaa	6600
agactgcaat ataaagaagc acagaagcca aaggcaaac accaccgact ttgaaccaga	6660

aacagcagcc taaaacagcc atcattgcag ccataccgag agcaccceaa ataagcagag	6720
gtttacggcc gaagcgggtca acaacacggg aagcaatcat ggtgaagatg aagttcacia	6780
caccgataga gatggtctgc aataatgccg tatcagctcc aaaacctaaa ttctggaaca	6840
tctgcgggtgc ataatacagc acggcggttaa taccgactaa ctgctggaag gcagcaacgg	6900
atacaccggc aaaaacaacg gtgataccaa aagcaaacia acctgcgctg cttttgtcca	6960
tggctttatc aaagccagct ttaatctttt gaatcgtcag attaggatcg gcttgcgggt	7020
ccagacgagc aaggattttg ctagcctcgg aatgacgtcc cttcatcacc aaccaatgcg	7080
gcgtatccgg tgcgggttaac agcagcaata agaaggcaat accgatcagg ccttctgaag	7140
ccggagacca gcaccaacca ctggcattaa cccaatcgat agaaccgaaa tgagccagta	7200
accaggtaaa gatataaccg gttaaagcac ccgtcacaat ggccatctgc tgaccagaaa	7260
ccatctgacc acgtttgtct ggcgagcaa tttcagcaat ataggttggg gtcaaggttg	7320
aaacgacacc gatacctaaa ccggcaagaa accggaaaaa gcaaaaaatt tgtaaagccg	7380
aaccaccggt tccaaataat ttttcgggtta acgcagcacc aaaaccggcg gcgacgaaac	7440
aatggaact catcaacaat ccgccgcgac gaccgaagcg aataccaatc cagccagaca	7500
gcaaagaacc ggtaacacia ccgacaaaaa cagcaacaac gaccatccca gaaagggag	7560
ccgcagccgt agcagacagg tgacgagggg caataaaatg gatatcaacc ggtgtaccga	7620
ttgcagcgat aaccgctgaa tcgtaaccga aaagcaagcc gcctatagca gcgattaggg	7680
ctagtccgct gactagaccc tgactacttt cagaactcat ggcgattcct ctccctctag	7740
agcgtcctgc tgttggttaag attattatac cacaccttgt agataaagtc aacaactttt	7800
tgcaaaattt ttcaggaatt ttagcagagg ttgttctgga tgtagaacia aacatctttc	7860
cgctcttgtg ctgttaggat atctttcttg gaagctaggt aggcctcgag ttatggcagt	7920
tggttaaaag gaaacaaaaa gaccgttttc acacaaaacg gtcttttttcg atttcttttt	7980
acagtacag ccacttttgc accaattaag gccacgtgt catttaaact ccgtttttcc	8040
agttcaaatg caattgcctt caatgcacct tcgtagctgt ggtgagccag cgggtgctggc	8100
tctcccccat ttacggataa gaatgcattt tccgagttaa taccgtcggc aatacctgac	8160
attaatactt cacagtcgct ggcatcgagt acggaaaact taatcgaaga cgaaccacag	8220
ttaataacca aaacaaccgg aaattcatte atctcttttc tcatcctgag ttacggatta	8280
aaacagtttg tatacgatgt tcaggatggt cagcagacca atcacggtaa caaacacgtt	8340
atccagacga ccacggtatt tcgccagaga cggcgcttta cggatggcat acatcggcaa	8400
caggcacagc agggatgcga taatcgggtc gcccatggct tcaatcaggt cgaggatgtt	8460
cgggttgggc taggcaacia cccaggtgga gcccatgatg aagatcatgc tgagagtatt	8520

```

cagtttacct agcgacactt tggttttgtc acctttataa ccgaacttca gaatcagacc      8580
attcaagcct tccagcgtcc ccagatagtg accgaagaaa gatttgaaga tagccacgag      8640
tgcgatgatg gaagccgcat attccagtgt aatcgcgaaac gttgttttgg taccgggtcat      8700
ggacgcaaag tggtttagcca gataagaaag cactggaata ttctgcgctt tggcttccgc      8760
catgttggcc ggagacagag taaacaggca gctaaaggca aagaacatca ccactgcaac      8820
catcagcatg ctggcacgag aaatgatttg ggaacattta cgttcgggtga agtcgcgacc      8880
gaagtctttc tcatactctt cacgtttaga aaccacgaag gaagagacga ttggcgagaa      8940
gttaaaggag aaaaccatga tggaaatccc cagccagaca gtgatcagga taccgtcatg      9000
accggttaac gacagcgaac cgagggtcaac ctggtcgata actgcagagt tccagtaagg      9060
gatcagcgac aaagaaatca gcaccaggct ggcgataaac ggccatacca ggtagctcag      9120
ggtaccgagc tcgaattcac tggccgctgt tttacaacgt cgtgactggg aaaaccctgg      9180
cgttacccaa cttaatcgcc ttgcagcaca tcccccttcc gccagctggc gtaatagcga      9240
agaggccccg accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggcgctt      9300
gatgcggtat tttctcctta cgcattctgt cggtatttca caccgcatat ggtgcactct      9360
cagtacaatc tgctctgatg ccgcatagtt aagccagccc cgacacccgc caacacccgc      9420
tgacgaattc                                     9430

```

```

<210> 17
<211> 5768
<212> ADN
<213> Plasmid pMH17F;

```

```

<400> 17
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtgggtcag actggaaaat cagaggggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgccca taaaacgccc      180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcat gaatccataa      240
aaggcgcttg tagtgccatt taccocccatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcagggtg cctgatgggt ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt      420
gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta      480
gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaagggtcta gcggaattta      600
cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac      660

```

ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattggtata	720
gtgattaataa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa	780
atgaactagc gattagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct	840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt	900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatggtgtat	960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttggtta	1020
aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat	1080
tagtttttag tgaagagata ttgccttata ttttccagtt aaaaaaattc ataaaatata	1140
atctggaaca tgttaagtct ttgaaaaca aatactctat gaggatttat gagtggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt ttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttctgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactggt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccggt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtccca	2280
attaattagt aatataatct attaaagggt attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggcccga tccttgagc	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgagtt	2460

gcaaaccctc	actgatccgc	atgcccgttc	catacagaag	ctgggcgaac	aaacgatgct	2520
cgccttccag	aaaaccgagg	atgcgaacca	cttcattccg	ggtcagcacc	accggcaagc	2580
gccgcgacgg	ccgaggtctt	ccgatctcct	gaagccaggg	cagatccgtg	cacagcacct	2640
tgccgtagaa	gaacagcaag	gccgccaatg	cctgacgatg	cgtggagacc	gaaaccttgc	2700
gctcgttcgc	cagccaggac	agaaatgcct	cgacttcgct	gctgcccagg	gttgccgggt	2760
gacgcacacc	gtggaaacgg	atgaaggcac	gaacccagt	gacataagcc	tggtcgggtc	2820
gtaagctgta	atgcaagtag	cgtatgcgct	cacgcaactg	gtccagaacc	ttgaccgaac	2880
gcagcgggtg	taacggcgca	gtggcggttt	tcattggctt	ttatgactgt	ttttttgggg	2940
tacagtctat	gcctcgggca	tccaagcagc	aagcgcgtta	cgccgtgggt	cgatgtttga	3000
tgttatggag	cagcaacgat	gttacgcagc	agggcagtcg	ccctaaaaca	aagttaaaca	3060
tcattgagga	agcggtgatc	gccgaagtat	cgactcaact	atcagaggta	gttggcgtca	3120
tcgagcgcca	tctcgaaccg	acgttgctgg	ccgtacattt	gtacggctcc	gcagtggatg	3180
gcggcctgaa	gccacacagt	gatattgatt	tgctggttac	ggtgaccgta	aggcttgatg	3240
aaacaacgcg	gcgagctttg	atcaacgacc	ttttggaaac	ttcggcttcc	cctggagaga	3300
gcgagattct	ccgcgctgta	gaagtcacca	ttgttggtga	cgacgacatc	attccgtggc	3360
gttatccagc	taagcgcgaa	ctgcaatttg	gagaatggca	gcgcaatgac	attcttgacg	3420
gtatcttcga	gccagccacg	atcgacattg	atctggctat	cttgctgaca	aaagcaagag	3480
aacatagcgt	tgctttggta	ggtccagcgg	cggaggaact	ctttgatccg	gttcctgaac	3540
aggatctatt	tgaggcgcta	aatgaaacct	taacgctatg	gaactcgccg	cccgactggg	3600
ctggcgatga	gcgaaatgta	gtgcttacgt	tgtcccgcat	ttggtacagc	gcagtaaccg	3660
gcaaaatcgc	gccgaaggat	gtcgtcgccg	actgggcaat	ggagcgcctg	ccggcccagt	3720
atcagcccgt	catacttgaa	gctagacagg	cttatcttgg	acaagaagaa	gatcgcttgg	3780
cctcgcgcgc	agatcagttg	gaagaatttg	tccactacgt	gaaaggcgag	atcaccaagg	3840
tagtcggcaa	ataatgtcta	acaattcggt	caagccgacg	ccgcttcgcg	gcgcggctta	3900
actcaagcgt	tagatgcact	aagcacataa	ttgctcacag	ccaaactatc	aggtcaagtc	3960
tgctttttatt	atttttaagc	gtgcataata	agccctacac	aaattgggag	atatatcatg	4020
aaaggctggc	tttttcttgt	tatcgcaata	gttggcgaag	taatcgcaac	atccgcatta	4080
aaatctagcg	agggctttac	taagctgatc	cggtggatga	ccttttgaat	gacctttaat	4140
agattatatt	actaattaat	tggggaccct	agaggtcccc	ttttttattt	taaaaatttt	4200
ttcacaaaac	ggtttacaag	catacgttgg	ccgattcatt	aatgcagctg	gcacgacagg	4260
tttcccgaact	ggaaagcggg	cagtgcgcgc	aacgcaatta	atgtgagtta	gctcactcat	4320

taggcacccc aggcctttaca ctttatgctt ccggctcgta tgttggtgtgg aattgtgagt	4380
tacctagaga gggtgagaat tgccgaacat gcgcataagt ttcccggaca gatttcaggt	4440
ggtcagcagc aacgcgttgc cattgcgcgt tcgctgtgta tgaagccgaa aattatgttg	4500
tttgatgagc caacgtcggc gctcgatcct gagatggtga aagagggtgt ggatacgtg	4560
attgggctgg cgcagtcggg tatgacaatg ttgtgtgtaa cacatgagat ggggtttgca	4620
cgaaccgtcg ctgaccgggt aatttttatg gatcgtgggg aaatagtgga gcaagctgca	4680
cctgatgaat tttttgcgca tcctaaatca gagcgtacga gggcattttt atcgcaggta	4740
atccattaat tgaatgttag ttcgaaaagc aaaaaggcca tcctttcgga tggcctttcg	4800
cttgatttga tgtctggcag tttatggcgg gcgtcctgcc cgccaccctc cgggccgttg	4860
cttcgcaacg ttcaaaccg ctcccggcgg atttgtccta ctcgggagag tgttcaccga	4920
caaacaacag ataaaaacaaa agggccagtc ttccgactga gccttttgtt ttatttgatg	4980
tctggcagtt ccctactctc gcatggggag accccacact accatcggcg ctacggcggt	5040
ttcactttctg agttcggcat ggggtcaggt gggaccaccg cgctactgcc gccagacaaa	5100
ttcttttcta atctgccgaa ctttaacct aaaaagtgggtg ctgataccca gagtcgaact	5160
ggggacctca cccttaccaa ggggtgcgctc taccaactga gccatatcag cacgctaaat	5220
ttgatgcctg gcagttccct actctcgcat ggggagaccc cacactacca tcggcgctac	5280
ggcgtttcac ttctgagttc ggcattgggg caggtgggac caccgcgcta cggccgccag	5340
gcaaattctg ttttatcaga ccgcttctgc gttctgattt aatctgtatc aggctgaaaa	5400
tcttctctca tccggataac aatttcacac aggaaacagc tatgaccatg attacgccaa	5460
gctcgagctc gaattcactg gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg	5520
ttacccaact taatcgcctt gcagcacatc cccctttcgc cagctggcgt aatagcgaa	5580
aggcccgcac cgatcgccct tcccaacagt tgccgagcct gaatggcgaa tggcgccctga	5640
tgccgtatct tctccttacg catctgtgcg gtatttcaca ccgcatatgg tgcaactctca	5700
gtacaatctg ctctgatgcc gcatagttaa gccagccccg acacccgccca acacccgtg	5760
acgaattc	5768

<210> 18

<211> 1053

<212> ADN

<213> Vùng mã hóa của gen aroG kiểu dại;

<400> 18

atgaattatc agaacgacga tttacgcac aaagaaatca aagagttact tcctcctgtc 60

gcattgctgg aaaaattccc cgctactgaa aatgccgcga atacggttgc ccatgcccca 120

```

aaagcgatcc ataagatcct gaaaggtaat gatgatcgcc tgttggttgt gattggccca 180
tgctcaattc atgatcctgt cgcggcaaaa gagtatgcc a ctcgcttgct ggcgctgcgt 240
gaagagctga aagatgagct ggaaatcgta atgcgcgtct attttgaaaa gccgcgtacc 300
acggtgggct ggaaagggtt gattaacgat ccgcatatgg ataatagctt ccagatcaac 360
gacggtctgc gtatagcccc 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
aatgccgccc gtgcgcccga ctgcttctg tccgtaacga aatgggggca ttcggcgatt 660
gtgaatacca gcggtaacgg cgattgccat atcattctgc gcggcggtaa agagcctaac 720
tacagcgcca agcacgttgc tgaagtgaag gaagggtga acaaagcagg cctgccagca 780
caggtgatga tcgatttcag ccatgctaac tcgtccaaac aattcaaaaa gcagatggat 840
gtttgtgctg acgtttgcc a gcagattgcc ggtggcgaaa aggccattat tggcgtgatg 900
gtggaaagcc atctggtgga aggcaatcag agcctcgaga gcggggagcc gctggcctac 960
ggtaagagca tcaccgatgc ctgcatcgcc tgggaagata ccgatgctct gttacgtcaa 1020
ctggcgaatg cagtaaaagc gcgtcgcggg taa 1053

```

```

<210> 19
<211> 8820
<212> ADN
<213> Plasmid pMH28F;

```

```

<220>
<221> dấu hiệu misc
<222> (5894)..(5894)
<223> n là a, c, g, hoặc t

```

```

<400> 19
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc 60
agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag 120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcca taaaacgccc 180
tgagaagccc gtgacgggct tttcttgat tatgggtagt ttccttgcat gaatccataa 240
aaggcgctg tagtgccatt taccgccatt cactgccaga gccgtgagcg cagcgaactg 300
aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca 360
gcgatttgcc cgagcttgcg agggtgctac ttaagccttt agggttttta ggtctgtttt 420
gtagaggagc aaacagcgtt tgcgacatcc ttttgtaata ctgcggaact gactaaagta 480

```

gtgagttata cacagggctg g gatctattc tttttatctt tttttattct ttctttattc	540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggtcta gcggaattta	600
cagagggctct agcagaattt acaagttttc cagcaaaggt ctagcagaat ttacagatac	660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattggtata	720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa	780
atgaactagc gattagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct	840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt	900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatggtgtat	960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttggtta	1020
aaggctttga gattttccag tggacaaact atgccaggt ctcaagcgaa aaattagaat	1080
tagtttttag tgaagagata ttgccttatt ttttccagtt aaaaaaattc ataaaatata	1140
atctggaaca tgttaagtct ttgaaaaca aatactctat gaggatttat gagtgggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggtgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgtttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaaca acaaaagtag aacaactggt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccggt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct agggccccca	2280

attaattagt aatataatct attaaaggtc attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagtct cgggtaacat caaggccccg tccttggagc	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgcagtt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctggggcgaa aaacgatgct	2520
cgccttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgcccagg gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cacgcaactg gtccagaacc ttgaccgaac	2880
gcagcgggtg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgaggga agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgta	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttggtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgacg	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcgc agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aaatctagcg agggctttac taagctgatc cggtggatga ccttttgaat gacctttaat	4140

agatttatatt actaattaat tggggaccct agaggtcccc ttttttattt taaaaatttt	4200
ttcacaaaac ggttttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccgaact ggaaagcggg cagtgagegc aacgcaatta atgtgagtta gtcactcat	4320
taggcacccc aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagt	4380
tacctagaga gggtgagaat tgccgaacat gcgcataagt tccccggaca gatttcaggt	4440
ggtcagcagc aacgcgttgc cattgcgcgt tcgctgtgta tgaagccgaa aattatgttg	4500
tttgatgagc caacgtcggc gctcgatcct gagatggtga aagaggtgct ggatacgatg	4560
attgggctgg cgcagtcggg tatgacaatg ttgtgtgtaa cacatgagat ggggtttgca	4620
cgaaccgtcg ctgaccgggt aatttttatg gatcgtgggg aaatagtgga gcaagctgca	4680
cctgatgaat tttttgcgca tcctaaatca gagcgtacga gggcattttt atcgcaggta	4740
atccattaat tgaatgtag ttcgaaaagc aaaaaggcca tcctttcgga tggcctttcg	4800
cttgatttga tgtctggcag tttatggcgg gcgtcctgcc cgccaccctc cgggccgttg	4860
cttcgcaacg ttcaaaccg gtgacggaag atcacttcgc agaataaata aatcctggtg	4920
tcctgttga taccgggaag ccctgggcca acttttggcg aaaatgagac gttgatcggc	4980
acgtaagagg ttccaacttt caccataatg aaataagatc actaccgggc gtatTTTTTg	5040
agttatcgag attttcagga gctaaggaag ctaaaatgga gaaaaaaatc actggatata	5100
ccaccgttga tatatcccaa tggcatcgta aagaacattt tgaggcattt cagtcagttg	5160
ctcaatgtac ctataaccag accgttcagc tggatattac ggccttttta aagaccgtaa	5220
agaaaaataa gcacaagttt tatccggcct ttattcacat tcttgccgc ctgatgaatg	5280
ctcatccgga attccgtatg gcaatgaaag acggtgagct ggtgatatgg gatagtgttc	5340
acccttgttt caccgttttc catgagcaaa ctgaaacgtt ttcatcgctc tggagtgaat	5400
accacgacga tttccggcag tttctacaca tatattcgca agatgtggcg tgttacggtg	5460
aaaacctggc ctatttccct aaagggttta ttgagaatat gtttttcgct tcagccaatc	5520
cctgggtgag tttcaccagt tttgatttaa acgtggccaa tatggacaac ttcttcgccc	5580
ccgtttttcac catgggcaaa tattatacgc aaggcgacaa ggtgctgatg ccgctggcga	5640
ttcaggttca tcatgccgtt tgtgatggct tccatgtcgg cagaatgctt aatgaattac	5700
aacagtactg cgatgagtgg cagggcgggg cgtaattttt ttaaggcagt tattggtgcc	5760
cttaaacgcc tgggtgctacg cctgaataag tgataataag cggatgaatg gcagaaattc	5820
gaaagcaaat tcgaccgggt cgtcggttca gggcagggtc gttaaatagc cgcttatgtc	5880
tattgctggg ttantcggta cccggggatc gcggccgcgg accggatccc atcacatata	5940

cctgccgttc actattat	ttt agtgaaatga gatattatga t	at	tttttctga attgtgatta	6000
aaaaggcaac tttatgcccc	tgcaacagaa actataaaaa	atacagagaa	tgaaaagaaa	6060
cagatagatt ttttagttct	ttaggcccg	agtc	tgcaaa tccttttatg attttctatc	6120
aaacaaaaga ggaaaataga	ccagttgcaa tccaaacgag	agtc	taataag aatgaggtcg	6180
aaaagtaa	at cgcgcggtt	tg	ttaactgat aaagcaggca agaccta	6240
aaa agtaaa	at cgcgcggtt	tg	ttaactgat aaagcaggca agaccta	6240
gcaaagtgt	at tactttggcg	tc	acccctta catat	6300
gcaaagtgt	at tactttggcg	tc	acccctta catat	6300
actaacttgc catcttcaaa	caggaggggt	ggaagaagca	gaccgctaac acagtacata	6360
aaaaaggaga catgaacgat	gaacatcaaa aagtttgcaa	aacaagcaac	agtattaacc	6420
tttactaccg cactgctggc	aggaggcgca actcaagcgt	ttgcgaaaga	aacgaaccaa	6480
aagccatata aggaaacata	cggcatttcc catattacac	gccatgat	at gctgcaa	6540
cctgaacagc aaaaaaatga	aaaatatcaa gttcctgaat	tcgattcgtc	cacaattaa	6600
aatatctctt ctgcaaaagg	cctggacgtt tgggacagct	ggccattaca	aaacgctgac	6660
ggcactgtcg caaactatca	cggctaccac atcgtctttg	cattagccgg	agatcctaaa	6720
aatgcggatg acacatcgat	ttacatgttc tatcaaaaag	tcggcgaaac	ttctattgac	6780
agctggaaaa acgctggccg	cgtcttttaa gacagcgaca	aattcgatgc	aatgattct	6840
atcctaaaag accaaacaca	agaatgggtca ggttcagcca	catttacatc	tgacggaaaa	6900
atccgtttat tctacactga	tttctccggt aaacattacg	gcaaacaac	actgacaact	6960
gcacaagtta acgtatcagc	atcagacagc tctttgaaca	tcaacggtgt	agaggattat	7020
aatcaatct ttgacggtga	cggaaaaacg tatcaaaatg	tacagcagtt	catcgatgaa	7080
ggcaactaca gctcaggcga	caaccatacg ctgagagatc	ctcactacgt	agaagataaa	7140
ggccacaaat acttagtatt	tgaagcaaac actggaactg	aagatggcta	ccaaggcgaa	7200
gaatctttat ttaacaaagc	atactatggc aaaagcacat	cattcttccg	tcaagaaagt	7260
caaaaacttc tgcaaagcga	taaaaaacgc acggctgagt	tagcaaacgg	cgctctcggt	7320
atgattgagc taaacgatga	ttacacactg aaaaaagtga	tgaaaccgct	gattgcatct	7380
aacacagtaa cagatgaaat	tgaacgcgcg aacgtcttta	aatgaacgg	caaatggtac	7440
ctgttcactg actcccgcgg	atcaaaaatg acgattgacg	gcattacgtc	taacgatatt	7500
tacatgcttg gttatgtttc	taattcttta actggcccat	acaagccgct	gaacaaaact	7560
ggccttgtgt taaaaatgga	tcttgatcct aacgatgtaa	cctttactta	ctcacacttc	7620
gctgtacctc aagcgaaagg	aaacaatgtc gtgattacaa	gctatatgac	aaacagagga	7680
ttctacgcag acaacaatc	aacgtttgcg ccgagcttcc	tgctgaacat	caaaggcaag	7740
aaaacatctg ttgtcaaaga	cagcatcctt gaacaaggac	aattaacagt	taacaaataa	7800

```

aaacgcaaaa gaaaatgcc aatcctatt ggcattttct tttattttctt ccattttaaat      7860
ggatgcatgc gctagcggag tgtatactgg ctactatgt tggcactgat gaggggtgtca      7920
gtgaagtgct tcctccccggc ggatttgtcc tactcgggag agtggttcacc gacaaacaac      7980
agataaaaca aaaggcccag tcttcgcact gagccttttg ttttatttga tgtctggcag      8040
ttccctactc tcgcatgggg agaccccaca ctaccatcgg cgctacggcg gtttcacttc      8100
tgagttcggc atgggggtcag gtgggaccac cgcgctactg ccgccagaca aattcttttc      8160
taatctgccc aactttaacc taaaaagtgg tgctgatacc cagagtcgaa ctggggacct      8220
cacccttacc aagggtgcgc tctaccaact gagccatata agcacgctaa atttgatgcc      8280
tggcagttcc ctactctcgc atggggagac cccacactac catcggcgct acggcgtttc      8340
acttctgagt tcggcatggg gtcaggtggg accaccgcgc tacggccgcc aggcaaattc      8400
tgttttatca gaccgcttct gcgttctgat ttaatctgta tcaggctgaa aatcttctct      8460
catccggata acaatttcac acaggaaaca gctatgacca tgattacgcc aagctcgagc      8520
tcgaattcac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg cgttacccaa      8580
cttaatcgcc ttgcagcaca tcccccttcc gccagctggc gtaatagcga agaggcccgc      8640
accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggcgctt gatgcggtat      8700
tttctcctta cgcactctgt cggtatttca caccgcatat ggtgcactct cagtacaatc      8760
tgctctgatg ccgcatagtt aagccagccc cgacaccgcg caacaccgcg tgacgaattc      8820

```

```

<210> 20
<211> 4774
<212> ADN
<213> Plasmid pCL1921;

```

```

<400> 20
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtgggtcag actggaaaaat cagaggggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agaccgcgca taaaacgccc      180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcat gaatccataa      240
aaggcgcttg tagtgccatt taccctcatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt      420
gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta      480
gtgagttata cacagggtcg ggatctattc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggctta gcggaattta      600

```

cagaggggtct agcagaattht acaagththt cagcaaaggt ctagcagaat ttacagatac	660
ccacaactca aaggaaaagg actagthaatt atcattgact agcccatctc aattgggtata	720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttgtt ttcaaagcaa	780
atgaactagc gatttagtcgc tatgacttaa cggagcatga aaccaagcta atthttatgct	840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt	900
tcacttataa ccaatacgtc cagatgatga acatcagtag ggaaaatgct tatgggtgat	960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttggtta	1020
aaggctthtga gattthtccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat	1080
tagththttag tgaagagata ttgccttata thttccagtt aaaaaaattc ataaaatata	1140
atctggaaca tgttaagtct thtgaaaaca aatactctat gaggatttat gagtggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagthcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
thttgaaacc aataagtaaa gattthaaaca cttacagcaa tatgaaattg gtggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga thctaccta cgtaacggac taagaaaaac actacacgat gctthtaactg	1560
caaaaattca gctcaccagt thtgaggcaa aaththtgag tgacatgcaa agtaagcatg	1620
atctcaatgg thcgtthtca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggthctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactgtt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagcttht acgagththt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac thgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccggt ggggaaaaaa	2100
tcatggcaat thtggagaa atagcgctth cagccggcaa acctgaagcc ggatctgcga	2160
thctgataac aaactagcaa caccagaaca gcccgthtgc gggcagcaaa acccgthtatg	2220
cttgtaaac gththtgtaa aaaaththta aaataaaaaa ggggacctct agggthccca	2280
attaattagt aatataatct attaaaggth attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctc cgggtaacat caaggcccg tctthggagc	2400

ccttgccctc	ccgcacgatg	atcgtgccgt	gatcgaaatc	cagatccttg	acccgcagtt	2460
gcaaaccctc	actgatccgc	atgcccgttc	catacagaag	ctgggcgaac	aaacgatgct	2520
cgccttccag	aaaaccgagg	atgcgaacca	cttcatccgg	ggtcagcacc	accggcaagc	2580
gccgcgacgg	ccgaggtctt	ccgatctcct	gaagccaggg	cagatccgtg	cacagcacct	2640
tgccgtagaa	gaacagcaag	gccgccaatg	cctgacgatg	cgtggagacc	gaaaccttgc	2700
gctcgttcgc	cagccaggac	agaaatgcct	cgacttcgct	gctgcccaag	gttgccgggt	2760
gacgcacacc	gtggaaacgg	atgaaggcac	gaaccagtg	gacataagcc	tgttcggttc	2820
gtaagctgta	atgcaagtag	cgtatgcgct	cacgcaactg	gtccagaacc	ttgaccgaac	2880
gcagcgggtg	taacggcgca	gtggcggttt	tcatggcttg	ttatgactgt	ttttttgggg	2940
tacagtctat	gcctcgggca	tccaagcagc	aagcgcgtta	cgccgtgggt	cgatgtttga	3000
tgttatggag	cagcaacgat	gttacgcagc	agggcagtcg	ccctaaaaca	aagttaaaca	3060
tcatgaggga	agcggtgatc	gccgaagtat	cgactcaact	atcagaggta	gttggcgctc	3120
tcgagcgcca	tctcgaaccg	acgttgctgg	ccgtacattt	gtacggctcc	gcagtggatg	3180
gcggcctgaa	gccacacagt	gatattgatt	tgctggttac	ggtgaccgta	aggcttgatg	3240
aaacaacgcg	gcgagctttg	atcaacgacc	ttttggaaac	ttcggttcc	cctggagaga	3300
gcgagattct	ccgcgctgta	gaagtcacca	ttgttggtgca	cgacgacatc	attccgtggc	3360
gttatccagc	taagcgcgaa	ctgcaatttg	gagaatggca	gcgcaatgac	attcttgtag	3420
gtatcttcga	gccagccacg	atcgacattg	atctggctat	cttgctgaca	aaagcaagag	3480
aacatagcgt	tgccctggta	ggtccagcgg	cggaggaact	ctttgatccg	gttcctgaac	3540
aggatctatt	tgaggcgcta	aatgaaacct	taacgctatg	gaactcgccg	cccgactggg	3600
ctggcgatga	gcgaaatgta	gtgcttacgt	tgtcccgcat	ttggtacagc	gcagtaaccg	3660
gcaaaatcgc	gccgaaggat	gtcgtgccc	actgggcaat	ggagcgcctg	ccggcccagt	3720
atcagcccgt	catacttgaa	gctagacagg	cttatcttgg	acaagaagaa	gatcgcttgg	3780
cctcgcgcgc	agatcagttg	gaagaatttg	tccactacgt	gaaaggcgag	atcaccaagg	3840
tagtcggcaa	ataatgtcta	acaattcggt	caagccgacg	ccgcttcgcg	gcgcggctta	3900
actcaagcgt	tagatgcact	aagcacataa	ttgctcacag	ccaaactatc	aggtcaagtc	3960
tgcttttatt	atttttaagc	gtgcataata	agccctacac	aaattgggag	atatatcatg	4020
aaaggctggc	tttttcttgt	tatcgcaata	gttggcgaag	taatcgcaac	atccgcatta	4080
aaatctagcg	agggctttac	taagctgac	cgggtggatga	ccttttgaat	gacctttaat	4140
agattatatt	actaattaat	tggggaccct	agaggtcccc	ttttttat	taaaaatttt	4200
ttcacaaaac	ggttttacaag	catacgttgg	ccgattcatt	aatgcagctg	gcacgacagg	4260

```

tttcccgact ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat 4320
taggcacccc aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc 4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgcattgctg 4440
caggctcgact ctagaggatc cccgggtacc gagctcgaat tcaactggccg tcgtttttaca 4500
acgtcgtgac tgggaaaacc ctggcggttac ccaacttaat cgccttgcag cacatcccc 4560
tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgccttcccc aacagttgcg 4620
cagcctgaat ggcgaaatggc gcctgatgcg gtattttctc cttacgcac tgtgcggtat 4680
ttcacaccgc atatggtgca ctctcagtac aatctgctct gatgccgcat agttaagcca 4740
gccccgacac ccgccaacac ccgctgacga attc 4774

```

```

<210> 21
<211> 6432
<212> ADN
<213> Plasmid pMG27;

```

```

<400> 21
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc 60
agtctgaatg acctgtcacg ggataatccg aagtggctcag actggaaaat cagagggcag 120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcca taaaacgccc 180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcac gaatccataa 240
aaggcgctg tagtgccatt tccccccatt cactgccaga gccgtgagcg cagcgaactg 300
aatgtcacga aaaagacagc gactcagggtg cctgatgggc ggagacaaaa ggaatattca 360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt aggggttttaa ggtctgtttt 420
gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta 480
gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc 540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggctca gcggaattta 600
cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac 660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattggtata 720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttgtt ttcaaagcaa 780
atgaactagc gattagtgcg tatgacttaa cggagcatga aaccaagcta attttatgct 840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt 900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatggtgtat 960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta 1020
aaggctttga gattttccag tggacaaaact atgccaaagt ctcaagcgaa aaattagaat 1080

```

tagtttttag tgaagagata ttgccttatc ttttccagtt aaaaaaattc ataaaatata	1140
atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgagggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaaca acaaaagtag aacaactgtt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgcctcccgt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtcccca	2280
attaattagt aatataatct attaaaggtc attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggcccga tccttgagc	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgcatgt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgccttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgcccaag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cagcgaactg gtccagaacc ttgaccgaac	2880

gcagcggtgg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgagggg agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgctca	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttgctga cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgag	3420
gtatcttcga gccagccagc atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcg agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aaatctagcg agggctttac taagctgatc cggtggatga ccttttgaat gacctttaat	4140
agattatatt actaattaat tggggaccct agaggtcccc ttttttattt taaaaatttt	4200
ttcacaaaac ggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccgact ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat	4320
taggcacccc aggcctttaca ctttatgctt ccggctcgta tgttggtgagc aattgtgagc	4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgcattgctg	4440
caggtcgacc gttaaactcta tcaccgcaag ggataaatat ctaacaccgt gcgtgttgac	4500
tattttacct ctggcgggtg taatggttgc atgtactaat ctagataagg aatatagcca	4560
tgaccgcacc gattcaggat ctgcgtgatg caattgccct gctgcaacag catgataatc	4620
agtatctgga aaccgatcat ccggttgatc cgaatgcaga actggcaggc gtttatcgtc	4680
atattggtgc cgggtggacc gttaaacgtc cgaccgctat tgggtccggca atgatgttta	4740

ataacattaa aggttatccg cacagccgta ttctgggttg tatgcatgca agccgtcagc	4800
gtgcagcact gctgctgggt tgtgaagcaa gtcagctggc actggaagtt ggtaaagcag	4860
ttaaaaaacc gggtgcaccg gtggttggtc cggaagcag cgcaccgtgt caagagcaga	4920
tttttctggc agatgatccg gatatttgatc tgcgtaccct gctgcctgca cataccaata	4980
ccccgattga tgcaggtccg tttttttgtc tgggtctggc cctggcaagc gatccgggtg	5040
atgcaagcct gaccgatgtt accattcatc gtctgtgtgt tcagggctcg gatgaactga	5100
gcatgttcct ggcagcaggt cgccatattg aagtttttcg tcagaaagca gaagcagcag	5160
gtaaaccgct gccgattacc attaatatgg gtctggacct agcaatctat attggcgcgt	5220
gttttgaagc accgaccacc ccgtttgggt ataatgaact ggggtgttgc ggtgcactgc	5280
gtcagcgtcc gggtgaactg gttcaggggtg ttagcgttcc ggaaaaagca attgcacgtg	5340
ccgaaattgt tattgaaggt gaactgctgc ctggtgttcg tgttcgtgaa gatcagcata	5400
ccaattcagg tcatgcaatg ccggaatttc cgggttattg tgggtggtgca aatccgagcc	5460
tgccggttat taaagttaaa gccgttacca tgcgcaataa cgcaattctg caaaccttgg	5520
ttggtccggg tgaagaacat accaccctgg caggtctgcc gaccgaagca agcatttgga	5580
atgcagttga agcagcaatt ccgggttttc tgcaaaatgt ttatgcccgt accgcaggcg	5640
gtggtaaatt tctgggtatt ctgcaagtga aaaaacgtca gcctgccgat gaaggctcgc	5700
agggtcaggc agccctgctg gcgctggcaa cctatagcga actgaaaaat atcattcttg	5760
tggatgagga tgtggacatt tttgatagt atgatattct gtgggcaatg accaccctga	5820
tgcaggggtga tgtagcatt accaccattc cgggtattcg cggtcacag ctggaccgca	5880
gccagacacc ggaatattca ccgagcattc gtggtaatgg tattagctgc aaaaccatct	5940
ttgattgtac cgttccgtgg gcactgaaaa gccattttga acgtgcaccg tttgcagatg	6000
ttgatccgcg tccgtttgca cctgaatatt ttgcacgtct ggaaaaaaat cagggcagcg	6060
caaaataagc taataacagg cctgctggta atcgcaggaa tttttatttg gatggatccc	6120
cgggtaccga gtcgaattc actggccgtc gttttacaac gtcgtgactg ggaaaaccct	6180
ggcgttaccc aacttaatcg ccttgacgca catccccctt tcgccagctg gcgtaatagc	6240
gaagaggccc gcaccgatcg cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc	6300
ctgatgcggg attttctcct tacgcattct tgcgggtattt cacaccgcat atgggtgact	6360
ctcagtacaa tctgctctga tgccgcatag ttaagccagc cccgacaccc gccaacaccc	6420
gctgacgaat tc	6432

<211> 7294
 <212> ADN
 <213> Plasmid pMG31;

<400> 22
 gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc 60
 agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag 120
 gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgccca taaaacgccc 180
 tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcac gaatccataa 240
 aaggcgccctg tagtgccatt taccctccatt cactgccaga gccgtgagcg cagcgaactg 300
 aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca 360
 gcgattttgcc cgagcttgcg aggggtgtac ttaagccttt agggttttaa ggtctgtttt 420
 gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta 480
 gtgagttata cacagggctg ggatctatct tttttatctt tttttattct ttctttattc 540
 tataaattat aaccacttga atataaaca aaaaaacaca caaaggctca gcggaattta 600
 cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac 660
 ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata 720
 gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa 780
 atgaactagc gattagtgcg tatgacttaa cggagcatga aaccaagcta attttatgct 840
 gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt 900
 tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat 960
 tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta 1020
 aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat 1080
 tagtttttag tgaagagata ttgccttatc ttttccagtt aaaaaattc ataaaatata 1140
 atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggttat 1200
 taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat 1260
 ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg 1320
 ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggttgata 1380
 agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc 1440
 tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca 1500
 ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg 1560
 caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg 1620
 atctcaatgg ttcgttctca tggtcacgc aaaaacaacg aaccacacta gagaacatac 1680

tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactggt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacagggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctccccgt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtccca	2280
attaattagt aatataatct attaaaggct attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggccccga tccttgagc	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgagtt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgcttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgcccag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cacgcaactg gtccagaacc ttgaccgaac	2880
gcagcggtgg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgaggga agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgta	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggccttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttgtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgag	3420
gtatcttcga gccagccagc atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540

aggatctatt	tgaggcgcta	aatgaaacct	taacgctatg	gaactcgccg	cccgactggg	3600
ctggcgatga	gcgaaatgta	gtgcttacgt	tgtcccgcat	ttggtacagc	gcagtaaccg	3660
gcaaaatcgc	gccgaaggat	gtcgctgccg	actgggcaat	ggagcgccctg	ccggcccagt	3720
atcagcccgt	catacttgaa	gctagacagg	cttatcttgg	acaagaagaa	gatcgcttgg	3780
cctcgcgcg	agatcagttg	gaagaatttg	tccactacgt	gaaaggcgag	atcaccaagg	3840
tagtcggcaa	ataatgtcta	acaattcggt	caagccgacg	ccgcttcgcg	gcgcggtta	3900
actcaagcgt	tagatgcact	aagcacataa	ttgtcacag	ccaaactatc	aggtcaagtc	3960
tgcttttatt	atttttaagc	gtgcataata	agccctacac	aaattgggag	atatatcatg	4020
aaaggctggc	tttttcttgt	tatcgcaata	gttggcgag	taatcgcaac	atccgcatta	4080
aatctagcg	agggtcttac	taagctgac	cgggtgatga	ccttttgaat	gacctttaat	4140
agattatatt	actaattaat	tggggaccct	agaggtcccc	ttttttat	taaaaatt	4200
ttcacaaaac	ggtttacaag	catacggttg	ccgattcatt	aatgcagctg	gcacgacagg	4260
tttcccgact	ggaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	gctcactcat	4320
taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tgttggtg	aattgtgagc	4380
ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgccaagct	tgctattgac	4440
gacagctatg	gttcactgtc	caccaaccaa	aactgtgctc	agtaccgcca	atatttctcc	4500
cttgaggggt	acaagaggt	gtccctagaa	gagatccacg	ctgtgtaaaa	attttacaaa	4560
aaggtattga	ctttccctac	agggtgtgta	ataatttaat	tacaggcggg	ggcaaccccg	4620
cctgttctgc	agaggaggaa	tatagccatg	gaagtgaaaa	tcttcaacac	ccaggatggt	4680
caggattttc	tgcgtgttgc	aagcggctctg	gaacaagagg	gtggtaatcc	gcgtgttaaa	4740
caaattattc	atcgtgttct	gagcgacctg	tataaagcaa	ttgaagatct	gaatatcacc	4800
agcgacgaat	attgggcagg	cgttgcatat	ctgaatcagc	tgggtgcaaa	tcaagaagca	4860
ggtctgctga	gtccgggtct	gggttttgat	cattatctgg	atatgcgtat	ggatgcagaa	4920
gatgcagcac	tgggtattga	aaatgcaaca	ccgcgtacca	ttgaagggtcc	gctgtatggt	4980
gcgggtgcac	cgaaagcgt	tggttatgca	cgcatggatg	atggtagcga	tccgaatggt	5040
cataccctga	ttctgcatgg	caccatTTTT	gatgcagatg	gtaaaccgct	gccgaatgca	5100
aaagttgaaa	tttggcatgc	aaacacccaaa	ggctttttata	gccattttga	tccgaccggt	5160
gaacagcagg	cctttaatat	gcgtcgtagc	attattaccg	atgagaatgg	tcagtatcgt	5220
gttcgtacca	ttctgcctgc	cggttatggt	tgtcctccgg	aagggtccgac	ccagcaactg	5280
ctgaaccaac	tgggtcgta	tggtaatcgt	ccggcacata	ttcattattt	tgttagcgca	5340

gatggtcacc gtaaactgac caccagatt aatgttgccg gtgatccgta tacctatgat	5400
gattttgcat atgccacccg tgaaggctctg gttgttgatg cagttgaaca taccgatccg	5460
gaagcaatta aagccaatga tgtggaaggt ccttttgccg aaatgggtgtt tgatctgaaa	5520
ctgacccgtc tggttgatgg tgttgataat cagggttgagg atcggtccgag tctggcagtt	5580
taatacacca aaatgggttca aaattatcag gcgagtgatc atgatcactg gcctgttttt	5640
atttcaggga aggggtggaga caattacgtg gataatcaga tcatccaaga aaccgtggat	5700
aaaattctga gcgttctgcc gaatcaggca ggtcagctgg cacgtctggg gcgtctgatg	5760
caatttgcat gcgatccgac cattaccgtt attggcaaat ataaccatgg taaaagccgt	5820
ctgctgaatg aactgattgg caccgatatc tttagcgttg cagataaacg tgaaaccatt	5880
cagctggccg aacataaaca ggatcaggtt cgttggttg atgcacctgg tctggatgcc	5940
gatgttgacg cagttgatga tcgtcatgca tttgaagcag tttggaccca ggcagatatt	6000
cgtctgtttg ttcatagcgt tcgtgaaggt gaactggatg caaccgaaca ccatctgctg	6060
caacagctga ttgaagatgc cgatcatagc cgtcgtcaga ccattctggg tctgaccag	6120
attgatcaga ttccggatca gaccatcctg acacagatta aaaccagcat tgcacagcag	6180
gttccgaaac tggatatattg ggcagttagc gcaaccgcgc atcggtcaggg cattgaaaac	6240
ggtaaaaccc tgctgatcga aaaaagcggg attggtgcac tgcgccatac cctggaacag	6300
gcactggcac aggtgccgag cgcacgtacc tatgaaaaaa atcggtctgct gtcagatctg	6360
caccatcagc tgaaacaact gctgctggat cagaaacatg ttctgcaaca actgcaacag	6420
acacagcaac agcagctgca tgattttgat accggtctga ttaacattct ggacaaaatt	6480
cgtgttgatc tggaaaccgat tgtgaatatt gatggtcagg atcaagcact gaatccggat	6540
agctttgcaa ccatgtttta aaacaccgca gcaaacagc agcgtgccaa agttcagatt	6600
gcatatagcc gtgcatgcat tgaaatcaac agccatctga ttcgccatgg tgttgttggt	6660
ctgcctgcgg aacagcagac caccattaaa agcattgata ccgtgattgt tgccgtgttt	6720
ggatcagcg ttaaatttcg tgatcagctg cgtgccctgt tttataccga taccgaacgt	6780
cagcgtctgc aacgtgaatt tcgtttctat tttgaaaaaa gtgccggctg catgattctg	6840
gcagcaaaaa ttgaacagac catgcgtcag cagggctgta ttcagaatgc catgatggca	6900
ctgcaacaaa tggaaagcgc agcataaaaa cacggacgcc gcaaacggcg tccgaatttc	6960
ttggtcgact ctagaggatc cccgggtacc gagctcgaat tcaactggccg tcgttttaca	7020
acgtcgtgac tgggaaaacc ctggcggttac ccaacttaat cgccttgag cacatcccc	7080
tttcgccagc tggcgtaata gcgaagaggc ccgcaccgat cgccttccc aacagttgag	7140
cagcctgaat ggcgaatggc gcctgatgag gtattttctc cttacgcac tgtgcggtat	7200

```

ttcacaccgc atatggtgca ctctcagtac aatctgctct gatgccgcat agttaagcca      7260
gccccgacac ccgccaacac ccgctgacga attc                                  7294

<210> 23
<211> 8952
<212> ADN
<213> Plasmid pMG33;

<400> 23
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcca taaaacgccc      180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcat gaatccataa      240
aaggcgcttg tagtgccatt taccgccatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt      420
gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta      480
gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaagggtcta gcggaattta      600
cagaggggtct agcagaattt acaagttttc cagcaaaggt ctagcagaat ttacagatac      660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata      720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa      780
atgaactagc gattagtgcg tatgacttaa cggagcatga aaccaagcta attttatgct      840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt      900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat      960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta      1020
aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat      1080
tagtttttag tgaagagata ttgccttatc ttttccagtt aaaaaaattc ataaaatata      1140
atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggttat      1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat      1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg      1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtggttgata      1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc      1440
tcgtaaccga acttgagaac aaccagataa aatgaatgg tgacaaaata ccaacaacca      1500

```

ttacatcaga	ttcctaccta	cgtaacggac	taagaaaaac	actacacgat	gctttaactg	1560
caaaaattca	gctcaccagt	tttgaggcaa	aattttttgag	tgacatgcaa	agtaagcatg	1620
atctcaatgg	ttcgtttctca	tggctcacgc	aaaaacaacg	aaccacacta	gagaacatac	1680
tggctaaata	cggaaggatc	tgaggttctt	atggctcttg	tatctatcag	tgaagcatca	1740
agactaacia	acaaaagtag	aacaactggt	caccgttaga	tatcaaaggg	aaaactgtcc	1800
atatgcacag	atgaaaacgg	tgtaaaaaag	atagatacat	cagagctttt	acgagttttt	1860
ggtgcattta	aagctgttca	ccatgaacag	atcgacaatg	taacagatga	acagcatgta	1920
acacctaata	gaacaggtga	aaccagtaaa	acaaagcaac	tagaacatga	aattgaacac	1980
ctgagacaac	ttgttacagc	tcaacagtca	cacatagaca	gcctgaaaca	ggcgatgctg	2040
cttatcgaat	caaagctgcc	gacaacacgg	gagccagtga	cgctctccgt	ggggaaaaaa	2100
tcatggcaat	tctggaagaa	atagcgcttt	cagccggcaa	acctgaagcc	ggatctgcga	2160
ttctgataac	aaactagcaa	caccagaaca	gcccgtttgc	gggcagcaaa	accggttatg	2220
cttgtaaacc	gttttgtgaa	aaaattttta	aaataaaaaa	ggggacctct	agggccccca	2280
attaattagt	aatataatct	attaaaggtc	attcaaaagg	tcattccaccg	gatcaattcc	2340
cctgctcgcg	caggctgggt	gccaaagctct	cgggtaacat	caaggccccga	tccttgggagc	2400
ccttgccctc	ccgcacgatg	atcgtgccgt	gatcgaaatc	cagatccttg	accgcagtt	2460
gcaaaccctc	actgatccgc	atgcccgttc	catacagaag	ctgggcgaac	aaacgatgct	2520
cgcttccag	aaaaccgagg	atgcgaacca	cttcatccgg	ggtcagcacc	accggcaagc	2580
gccgcgacgg	ccgaggtctt	ccgatctcct	gaagccaggg	cagatccgtg	cacagcacct	2640
tgccgtagaa	gaacagcaag	gccgccaatg	cctgacgatg	cgtggagacc	gaaaccttgc	2700
gctcgttcgc	cagccaggac	agaaatgcct	cgacttcgct	gctgccaag	gttgccgggt	2760
gacgcacacc	gtggaaacgg	atgaaggcac	gaaccagtg	gacataagcc	tgttcggttc	2820
gtaagctgta	atgcaagtag	cgtatgcgct	cacgcaactg	gtccagaacc	ttgaccgaac	2880
gcagcgggtg	taacggcgca	gtggcggttt	tcatggcttg	ttatgactgt	ttttttgggg	2940
tacagtctat	gcctcgggca	tccaagcagc	aagcgcgtta	cgccgtgggt	cgatgtttga	3000
tgttatggag	cagcaacgat	gttacgcagc	agggcagtcg	ccctaaaaca	aagttaaaca	3060
tcatgagggg	agcggtgatc	gccgaagtat	cgactcaact	atcagaggta	gttggcgctca	3120
tcgagcgcca	tctcgaaccg	acgttgctgg	ccgtacattt	gtacggctcc	gcagtggatg	3180
gcggcctgaa	gccacacagt	gatattgatt	tgctggttac	ggtgaccgta	aggcttgatg	3240
aaacaacgcg	gcgagctttg	atcaacgacc	ttttggaaac	ttcggttcc	cctggagaga	3300

gcgagattct ccgcgctgta gaagtcacca ttgttggtga cgacgacatc attccgtggc 3360
 gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgacg 3420
 gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag 3480
 aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac 3540
 aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg 3600
 ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg 3660
 gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgctg cgggccagc 3720
 atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg 3780
 cctcgcgcg agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg 3840
 tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggtta 3900
 actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc 3960
 tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg 4020
 aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta 4080
 aaatctagcg agggctttac taagctgatc cggtggtatga ccttttgaat gacctttaat 4140
 agattatatt actaattaat tggggaccct agaggtcccc ttttttattt taaaaatttt 4200
 ttcacaaaac ggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg 4260
 tttcccgact ggaaagcggg cagtgcgcgc aacgcaatta atgtgagtta gctcactcat 4320
 taggcacccc aggctttaca ctttatgctt ccggctcgta tgttggtggtg aattgtgagc 4380
 ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgctattgac 4440
 gacagctatg gttcactgtc caccaaccaa aactgtgctc agtaccgcca atattttctcc 4500
 cttgaggggt acaaagaggt gtccctagaa gagatccacg ctgtgtaaaa attttacaaa 4560
 aaggtattga ctttccctac aggggtgtgta ataatttaat tacaggcggg ggcaaccccg 4620
 cctgttctgc agaggaggaa tatagccatg gaagtgaana tcttcaacac ccaggatgtt 4680
 caggattttc tgcgtgttgc aagcggctct gaacaagagg gtggtaatcc gcgtgttaaa 4740
 caaattattc atcgtgttct gagcgacctg tataaagcaa ttgaagatct gaatatcacc 4800
 agcgacgaat attgggcagg cgttgcatat ctgaatcagc tgggtgcaaa tcaagaagca 4860
 ggtctgctga gtccgggtct gggttttgat cattatctgg atatgcgtat ggatgcagaa 4920
 gatgcagcac tgggtattga aaatgcaaca ccgcgtacca ttgaaggctc gctgtatgtt 4980
 gcgggtgcac cggaaagcgt tggttatgca cgcgtgatg atggtagcga tccgaatggt 5040
 cataccctga ttctgcatgg caccattttt gatgcagatg gtaaaccgct gccgaatgca 5100
 aaagttgaaa tttggcatgc aaacaccaaa ggctttttata gccattttga tccgaccggt 5160

gaacagcagg cctttaatat gcgtcgtagc attattaccg atgagaatgg tcagtatcgt	5220
gttcgtacca ttctgcctgc cggttatggt tgtcctccgg aagggtccgac ccagcaactg	5280
ctgaaccaac tgggtcgtca tggtaatcgt ccggcacata ttcattatTTt tgtagcgca	5340
gatggtcacc gtaaactgac caccagatt aatgttgccg gtgatccgta tacctatgat	5400
gattttgcat atgccacccg tgaaggctctg gttgttgatg cagttgaaca taccgatccg	5460
gaagcaatta aagccaatga tgtggaaggT ccttttgccg aaatggTgtt tgatctgaaa	5520
ctgaccctgc tggttgatgg tgttgataat caggttgtgg atcgtccgcg tctggcagtt	5580
taatacacca aaatggttca aaattatcag gcgagtgatc atgatcactg gcctgttttt	5640
atttcaggggagggtggaga caattacgtg gataatcaga tcatccaaga aaccgtggat	5700
aaaattctga gcgttctgcc gaatcaggca ggtcagctgg cacgtctggt gcgtctgatg	5760
caatttgcat gcgatccgac cattaccgtt attggcaaat ataaccatgg taaaagccgt	5820
ctgctgaatg aactgattgg caccgatatc tttagcgttg cagataaacg tgaaaccatt	5880
cagctggccg aacataaaca ggatcaggtt cgttggctgg atgcacctgg tctggatgcc	5940
gatgttgcaaggagttgatga tcgtcatgca tttgaagcag tttggacca ggcagatatt	6000
cgtctgtttg ttcatagcgt tcgtgaaggT gaactggatg caaccgaaca ccatctgctg	6060
caacagctga ttgaagatgc cgatcatagc cgtcgtcaga ccattctggt tctgaccag	6120
attgatcaga ttccggatca gaccatcctg acacagatta aaaccagcat tgcacagcag	6180
gttccgaaac tggatatttg ggcagttagc gcaaccctgc atcgtcaggg cattgaaaac	6240
ggtaaaaccc tgctgatcga aaaaagcggT attggtgcac tgcgccatac cctggaacag	6300
gcactggcac aggtgccgag cgcacgtacc tatgaaaaaa atcgtctgct gtcagatctg	6360
caccatcagc tgaaacaact gctgctggat cagaaacatg ttctgcaaca actgcaacag	6420
acacagcaac agcagctgca tgattttgat accggtctga ttaacattct ggacaaaatt	6480
cgtgttgatc tggaaccgat tgtgaatatt gatggtcagg atcaagcact gaatccggat	6540
agctttgcaa ccatgtttta aaacaccgca gcaaaacagc agcgtgccaa agttcagatt	6600
gcatatagcc gtgcatgcat tgaaatcaac agccatctga ttcgccatgg tgttgttggt	6660
ctgcctgcgg aacagcagac caccattaaa agcattgata ccgtgattgt tgccgtgttt	6720
ggtatcagcg ttaaatttcg tgatcagctg cgtgccctgt tttataccga taccgaacgt	6780
cagcgtctgc aacgtgaatt tcgtttctat tttgaaaaaa gtgccggctc catgattctg	6840
gcagcaaaaa ttgaacagac catgcgtcag cagggctgta ttcagaatgc catgatggca	6900
ctgcaacaaa tggaaagcgc agcataaaaa cacggacgcc gcaaacggcg tccgaatttc	6960

ttgggtcgacc	gttaaatacta	tcaccgcaag	ggataaatat	ctaacaccgt	gcgtgttgac	7020
tattttacct	ctggcggtga	taatggttgc	atgtactaat	ctagataagg	aatatagcca	7080
tgaccgcacc	gattcaggat	ctgcgtgatg	caattgccct	gctgcaacag	catgataatc	7140
agtatctgga	aaccgatcat	ccggttgatc	cgaatgcaga	actggcaggc	gtttatcgtc	7200
atattggtgc	cggtggcacc	gttaaacgtc	cgacccgtat	tgggtccggca	atgatgttta	7260
ataacattaa	aggttatccg	cacagccgta	ttctggttgg	tatgcatgca	agccgtcagc	7320
gtgcagcact	gctgctgggt	tgtgaagcaa	gtcagctggc	actggaagtt	ggtaaagcag	7380
ttaaaaaacc	ggttgcaccg	gtggttgttc	cggcaagcag	cgcaccgtgt	caagagcaga	7440
tttttctggc	agatgatccg	gattttgatc	tgcgtaccct	gctgcctgca	cataccaata	7500
ccccgattga	tgcagggtccg	tttttttgtc	tgggtctggc	cctggcaagc	gatccggtgg	7560
atgcaagcct	gaccgatgtt	accattcatc	gtctgtgtgt	tcagggtcgt	gatgaactga	7620
gcatgttcct	ggcagcaggt	cgccatattg	aagtttttcg	tcagaaagca	gaagcagcag	7680
gtaaaccgct	gccgattacc	attaatatgg	gtctggaccc	agcaatctat	attggcgcat	7740
gttttgaagc	accgaccacc	ccgtttgggt	ataatgaact	gggtgttgcc	ggtgcactgc	7800
gtcagcgtcc	ggttgaactg	gttcaggggtg	ttagcgttcc	ggaaaaagca	attgcacgtg	7860
ccgaaattgt	tattgaaggt	gaactgctgc	ctggtgttcg	tgttcgtgaa	gatcagcata	7920
ccaattcagg	tcatgcaatg	ccggaatttc	cgggttattg	tgggtggtgca	aatccgagcc	7980
tgccggttat	taaagttaaa	gccgttacca	tgcgcaataa	cgcaattctg	caaaccctgg	8040
ttgggtccggg	tgaagaacat	accaccctgg	caggctctgcc	gaccgaagca	agcatttgga	8100
atgcagttga	agcagcaatt	ccgggttttc	tgcaaaatgt	ttatgccccat	accgcaggcg	8160
gtggtaaatt	tctgggtatt	ctgcaagtga	aaaaacgtca	gcctgccgat	gaaggctgtc	8220
agggtcaggc	agccctgctg	gcgctggcaa	cctatagcga	actgaaaaat	atcattctgg	8280
tggatgagga	tgtggacatt	tttgatagtg	atgatattct	gtgggcaatg	accaccgta	8340
tgcaggggtga	tgtagcatt	accaccattc	cgggtattcg	cggtcatcag	ctggacccga	8400
gccagacacc	ggaatattca	ccgagcattc	gtggtaatgg	tattagctgc	aaaaccatct	8460
ttgattgtac	cgttccgtgg	gactgaaaa	gccattttga	acgtgcaccg	tttgcatg	8520
ttgatccgcg	tccgtttgca	cctgaatatt	ttgcacgtct	ggaaaaaaat	cagggcagcg	8580
caaaataagc	taataacagg	cctgctggta	atcgcaggaa	tttttatattg	gatggatccc	8640
cgggtaccga	gctcgaattc	actggccgtc	gttttacaac	gtcgtgactg	ggaaaaccct	8700
ggcgttaccc	aacttaatcg	ccttgacgca	catccccctt	tcgccagctg	gcgtaatagc	8760
gaagaggccc	gcaccgatcg	cccttcccaa	cagttgcgca	gcctgaatgg	cgaatggcgc	8820

ctgatgcggt attttctcct tacgcatctg tgcggtattht cacaccgcat atggtgcact 8880
 ctcagtacaa tctgctctga tgccgcatag ttaagccagc cccgacaccc gccaacaccc 8940
 gctgacgaat tc 8952

<210> 24
 <211> 10630
 <212> ADN
 <213> Plasmid pMG37;

<400> 24
 gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc 60
 agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag 120
 gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcca taaaacgccc 180
 tgagaagccc gtgacgggct tttcttgat tatgggtagt ttccttgcat gaatccataa 240
 aaggcgctg tagtgccatt taccctcatt cactgccaga gccgtgagcg cagcgaactg 300
 aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca 360
 gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt 420
 gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta 480
 gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc 540
 tataaattat aaccacttga atataaaca aaaaaacaca caaagggtcta gcggaattta 600
 cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac 660
 ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata 720
 gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa 780
 atgaactagc gattagtgcg tatgacttaa cggagcatga aaccaagcta attttatgct 840
 gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt 900
 tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat 960
 tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggta 1020
 aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat 1080
 tagtttttag tgaagagata ttgccttata ttttccagtt aaaaaaattc ataaaatata 1140
 atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggtat 1200
 taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat 1260
 ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg 1320
 ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggtgata 1380
 agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc 1440

tcgtaaccga acttgagaac aaccagataa aatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttcctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatTTTTgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactggt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccgt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct agggccccca	2280
attaattagt aatataatct attaaaggtc attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggccccga tccttgagac	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgcatgt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctggggcgaa aaacgatgct	2520
cgcttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgccccaa gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cacgcaactg gtccagaacc ttgaccgaac	2880
gcagcgggtg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgaggga agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgtca	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240

aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttgtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgcag	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgcctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcgc agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcgtt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aatctagcg agggctttac taagctgac cgggtggatga ccttttgaat gacctttaat	4140
agattatatt actaattaat tggggaccct agaggcccc ttttttattt taaaaatttt	4200
ttcacaaaac ggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccgact ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat	4320
taggcacccc aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc	4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgctattgac	4440
gacagctatg gttcactgtc caccaaccaa aactgtgctc agtaccgcca atattttctcc	4500
cttgaggggt acaaagagg tgccttagaa gagatccacg ctgtgtaaaa attttacaaa	4560
aaggatttga ctttccctac aggggtgtgta ataatttaac tacaggcggg ggcaaccccg	4620
cctgttctgc agaggaggaa tatagccatg gaagtgaana tcttcaacac ccaggatggt	4680
caggattttc tgcgtgttgc aagcggctctg gaacaagagg gtggtaatcc gcgtgttaaa	4740
caaattatcc atcgtgttct gagcgacctg tataaagcaa ttgaagatct gaatatcacc	4800
agcgacgaat attgggcagg cgttgcatat ctgaatcagc tgggtgcaaa tcaagaagca	4860
ggtctgctga gtccgggtct gggttttgat cattatctgg atatgcgtat ggatgcagaa	4920
gatgcagcac tgggtattga aaatgcaaca ccgcgtacca ttgaaggctc gctgtatggt	4980
gcggggtgcac cggaaagcgt tggttatgca cgcattggatg atggtagcga tccgaatggt	5040
cataccctga ttctgcatgg caccattttt gatgcagatg gtaaaccgct gccgaatgca	5100

aaagttgaaa tttggcatgc aaacacccaaa ggctttttata gccatttttga tccgaccggt	5160
gaacagcagg cctttaatat gcgtcgtagc attattaccg atgagaatgg tcagtatcgt	5220
gttcgtacca ttctgcctgc cggttatggt tgtcctccgg aaggctccgac ccagcaactg	5280
ctgaaccaac tgggtcgtca tggtaatcgt ccggcacata ttcattatgt tgttagcgca	5340
gatggtcacc gtaaaactgac caccagatt aatggtgccg gtgatccgta tacctatgat	5400
gattttgcat atgccacccg tgaaggctctg gttgttgatg cagttgaaca taccgatccg	5460
gaagcaatta aagccaatga tgtggaaggc ccttttgccg aaatgggtgtt tgatctgaaa	5520
ctgaccgcgc tgggttgatgg tgttgataat cagggtgtgg atcgctccgcg tctggcagtt	5580
taatacacca aaatggttca aaattatcag gcgagtgatc atgatcactg gcctgttttt	5640
atttcaggga aggggtggaga caattacgtg gataatcaga tcatccaaga aaccgtggat	5700
aaaattctga gcgttctgcc gaatcaggca ggtcagctgg cacgtctggt gcgtctgatg	5760
caatttgcat gcgatccgac cattaccgtt attggcaaata ataaccatgg taaaagccgt	5820
ctgctgaatg aactgattgg caccgatatc tttagcgttg cagataaacg tgaaaccatt	5880
cagctggccg aacataaaca ggatcagggt cgttggctgg atgcacctgg tctggatgcc	5940
gatgttgcat cagttgatga tcgtcatgca tttgaagcag tttggacca ggcagatatt	6000
cgtctgtttg ttcatagcgt tcgtgaaggc gaactggatg caaccgaaca ccatctgctg	6060
caacagctga ttgaagatgc cgatcatagc cgtcgtcaga ccattctggt tctgaccag	6120
attgatcaga ttccggatca gaccatcctg acacagatta aaaccagcat tgcacagcag	6180
gttccgaaac tggatatattg ggcagttagc gcaaccgcgc atcgctcaggg cattgaaaac	6240
ggtaaaaccc tgctgatcga aaaaagcggc attggtgcac tgcgccatac cctggaacag	6300
gcactggcac aggtgccgag cgcacgtacc tatgaaaaaa atcgctctgct gtcagatctg	6360
caccatcagc tgaaacaact gctgctggat cagaaacatg ttctgcaaca actgcaacag	6420
acacagcaac agcagctgca tgattttgat accggtctga ttaacattct ggacaaaatt	6480
cgtgttgatc tgggaaccgat tgtgaatatt gatggtcagg atcaagcact gaatccggat	6540
agctttgcaa ccatgtttta aaacaccgca gcaaaacagc agcgtgcaa agttcagatt	6600
gcatatagcc gtgcatgcat tgaaatcaac agccatctga ttcgccatgg tgttgttggt	6660
ctgcctgcgg aacagcagac caccattaaa agcattgata ccgtgattgt tgccgtgttt	6720
ggtatcagcg ttaaatttcg tgatcagctg cgtgccctgt tttataccga taccgaacgt	6780
cagcgtctgc aacgtgaatt tcgtttctat tttgaaaaaa gtgccggctc catgattctg	6840
gcagcaaaaa ttgaacagac catgcgtcag cagggtctga ttcagaatgc catgatggca	6900

ctgcaacaaa	tggaaagcgc	agcataaaaa	cacggacgcc	gcaaacggcg	tccgaatttc	6960
ttggtcgacc	gttaaatacta	tcaccgcaag	ggataaatat	ctaacaccgt	gcgtgttgac	7020
tattttacct	ctggcggtga	taatggttgc	atgtactaat	ctagataagg	aatatagcca	7080
tgaccgcacc	gattcaggat	ctgcgtgatg	caattgccct	gctgcaacag	catgataatc	7140
agtatctgga	aaccgatcat	ccggttgatc	cgaatgcaga	actggcaggc	gtttatcgtc	7200
atattggtgc	cgggtggcacc	gttaaacgtc	cgacccgtat	tgggtccggca	atgatgttta	7260
ataacattaa	aggttatccg	cacagccgta	ttctggttgg	tatgcatgca	agccgtcagc	7320
gtgcagcact	gctgctgggt	tgtgaagcaa	gtcagctggc	actggaagtt	ggtaaagcag	7380
ttaaaaaacc	ggttgaccgc	gtggttgttc	cggcaagcag	cgcaccgtgt	caagagcaga	7440
tttttctggc	agatgatccg	gattttgatc	tgcgtaacct	gctgcctgca	cataccaata	7500
ccccgattga	tgcagggtccg	tttttttgtc	tgggtctggc	cctggcaagc	gatccggtgg	7560
atgcaagcct	gaccgatgtt	accattcatc	gtctgtgtgt	tcagggtcgt	gatgaactga	7620
gcatgttcct	ggcagcaggt	cgccatattg	aagtttttcg	tcagaaagca	gaagcagcag	7680
gtaaaccgct	gccgattacc	attaatatgg	gtctggaccc	agcaatctat	attggcgcat	7740
gttttgaagc	accgaccacc	ccgtttgggt	ataatgaact	gggtgttgcc	ggtgcactgc	7800
gtcagcgtcc	ggttgaactg	gttcagggtg	ttagcgttcc	ggaaaaagca	attgcacgtg	7860
ccgaaattgt	tattgaaggt	gaactgctgc	ctggtgttcg	tgttcgtgaa	gatcagcata	7920
ccaattcagg	tcatgcaatg	ccggaatttc	cgggttattg	tgggtggtgca	aatccgagcc	7980
tgccggttat	taaagttaaa	gccgttacca	tgcgcaataa	cgcaattctg	caaaccctgg	8040
ttggtccggg	tgaagaacat	accaccctgg	caggctctgcc	gaccgaagca	agcatttgga	8100
atgcagttga	agcagcaatt	ccgggttttc	tgcaaaatgt	ttatgcccac	accgcaggcg	8160
gtggtaaatt	tctgggtatt	ctgcaagtga	aaaaacgtca	gcctgccgat	gaaggctcgc	8220
agggtcaggc	agccctgctg	gcgctggcaa	cctatagcga	actgaaaaat	atcattctgg	8280
tggatgagga	tgtggacatt	tttgatagtg	atgatattct	gtgggcaatg	accacccgta	8340
tgcagggtga	tgtagcatt	accaccattc	cgggtattcg	cggtcacag	ctggaccoga	8400
gccagacacc	ggaatattca	ccgagcattc	gtggtaatgg	tattagctgc	aaaaccatct	8460
ttgattgtac	cggtccgtgg	gcactgaaaa	gccattttga	acgtgcaccg	tttgcagatg	8520
ttgatccgcg	tccgtttgca	cctgaatatt	ttgcacgtct	ggaaaaaaat	cagggcagcg	8580
caaaataagc	taataacagg	cctgctggta	atcgcaggaa	tttttatttg	gatggatccg	8640
cctacctagc	ttccaagaaa	gatatcctaa	cagcacaaga	gcggaaagat	gttttggtct	8700
acatccagaa	caacctctgc	taaaattcct	gaaaaatttt	gcaaaaagtt	gttgacttta	8760

tctacaaggt gtggtataat aatcttaaca acagcaggac gctccccgggt tgaggaaaac	8820
ctaataaaac tgaccagcct gcgtgttagc ctgctggcac tgggtctggt taccagcgggt	8880
tttgcagcag cagaaaccta taccgttgat cggtatcagg atgatagcga aaaaggttagc	8940
ctgcggttggg caattgaaca gagcaatgca aatagcgcac aagaaaacca gattctgatt	9000
caggcagttg gtaaagcacc gtatgttatc aaagttgata aaccgctgcc tccgattaaa	9060
agcagcggtta aaatcattgg caccgagtgg gataaaaccg gtgaatttat tgcaattgat	9120
ggcagcaact atatcaaagg cgaaggtgaa aaagcatgtc cgggtgcaaa tccgggtcag	9180
tatggcacca atgttcgtac catgaccctg cctgggtctgg ttctgcaaga tggtaatgggt	9240
gttaccctga aaggtctgga tgttcacgtt ttttgtattg gtgttctggt taatcgcagc	9300
agcaataacc tgattcagca taatcgtatc agcaacaatt atgggtggtgc cgggtgttatg	9360
attaccggtg atgatggtaa aggtaatccg accagcacca ccaccaataa taacaaagtt	9420
ctggataacg tgttcacga taatgggtgat ggtctggaac tgaccctggg tgcagcattt	9480
aatctgattg caaataacct gtttaccagc acaaaagcca atccggaacc gagccaggggt	9540
attgaaattc tgtggggtaa tgataatgcc gtgggtgggt acaaattcga aaactattca	9600
gatggcctgc aaatcaattg gggtaaacgt aactatatcg cctataacga actgaccaat	9660
aacagcctgg gtttcaatct gacaggtgat ggtaacattt tcgacagcaa taaagtgcac	9720
ggtaaccgta ttggtattgc cattcgtagt gaaaaagatg ccaatgcacg tattaccctg	9780
acaaaaaatc agatttggga taacggcaaa gatatcaaac gttgtgaagc cgggtggtagc	9840
tgtgttccga atcagcgtct ggggtgcaatt gtttttgggt ttccggcact ggaacatgaa	9900
ggttttgttg gtagccgtgg cgggtggtgtt gttattgaac cggcaaaaact gcaaaaaacc	9960
tgcaccagc cgaaccagca gaattgtaat gcaattccta atcaggggat tcaggcaccg	10020
aaactgacag ttagcaaaaa acagctgacc gttgaagtta aaggcacccc taatcagcgt	10080
tataatgtgg aatttttttg caatcgtaat gccagcagca gcgaagcaga acagtatctg	10140
ggtagcattg ttgttggtac cgatcatcag ggtctggcaa aagcaaattg ggctccgaaa	10200
gttagcatgc cgagcggtac cgcaaattgt acagatcatc tgggtgagac cagcgaactg	10260
agcagcgcag ttaaaatgcg ttaaatgcat gcgcgccgcg ttcgcgcggc gctttttttt	10320
ggtagcagc tcgaattcac tggccgtcgt tttacaacgt cgtgactggg aaaaccctgg	10380
cgttacccaa cttaatcgcc ttgcagcaca tcccccttc gccagctggc gtaatagcga	10440
agaggccgc accgatcgcc cttcccaaca gttgcgcagc ctgaatggcg aatggcgcct	10500
gatgcggtat tttctccta cgcacgtgtg cggtatttca caccgcatat ggtgcactct	10560

cagtacaatc tgctctgatg ccgcatagtt aagccagccc cgacacccgc caacacccgc 10620
 tgacgaattc 10630

<210> 25
 <211> 6452
 <212> ADN
 <213> Plasmid pMG39;

<400> 25
 gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc 60
 agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag 120
 gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcca taaaacgccc 180
 tgagaagccc gtgacgggct tttcttgat tatgggtagt ttccttgcat gaatccataa 240
 aaggcgctg tagtgccatt taccgccatt cactgccaga gccgtgagcg cagcgaactg 300
 aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca 360
 gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt 420
 gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta 480
 gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc 540
 tataaattat aaccacttga atataaaca aaaaaacaca caaaggctca gcggaattta 600
 cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac 660
 ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata 720
 gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa 780
 atgaactagc gattagtgcg tatgacttaa cggagcatga aaccaagcta attttatgct 840
 gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acgggtatcgt 900
 tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat 960
 tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta 1020
 aaggctttga gatgttccag tggacaaaact atgccaagtt ctcaagcgaa aaattagaat 1080
 tagtttttag tgaagagata ttgccttatc ttttccagtt aaaaaaatc ataaaatata 1140
 atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggttat 1200
 taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat 1260
 ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg 1320
 ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggttgata 1380
 agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc 1440
 tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca 1500

ttacatcaga ttcctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatTTTTgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgtttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaaca acaaaagtag aacaactgtt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgcctcccggt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtcccca	2280
attaattagt aatataatct attaaaggct attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggcccga tccttgagac	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgagtt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgccttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgcccaag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cagcgaactg gtccagaacc ttgaccgaac	2880
gcagcggtgg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgagggg agcgggtgat gccgaagtat cgactcaact atcagaggta gttggcgctca	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctgggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttgtgca cgacgacatc attccgtggc	3360

gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgagc	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgcctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcgc agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcgtt caagccgacg ccgcttcgcg gcgcggtta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aatctagcg agggctttac taagctgacg cggtggtatga ccttttgaat gacctttaat	4140
agattatatt actaattaat tggggaccct agaggcccc ttttttattt taaaaatttt	4200
ttcacaaaac gggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccagct ggaaagcggg cagtgcgcgc aacgcaatta atgtgagtta gctcactcat	4320
taggcacccc aggcctttaca ctttatgctt ccggctcgta tgttggtgtg aattgtgagc	4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgcattgctg	4440
caggctcact ctagaggatc cgcctacctg gcttccaaga aagatatcct aacagcacia	4500
gagcggaaaag atgttttgtt ctacatccag aacaacctct gctaaaattc ctgaaaaatt	4560
ttgcaaaaag ttgttgactt tatctacaag gtgtggtata ataattctaa caacagcagg	4620
acgctcccgg gttgaggaaa acctaatgaa actgaccagc ctgcgtgtta gcctgctggc	4680
actgggtctg gttaccagcg gttttgcagc agcagaaacc tataccgttg atcgttatca	4740
ggatgatagc gaaaaaggta gcctgcgttg ggcaattgaa cagagcaatg caaatagcgc	4800
acaagaaaac cagattctga ttcaggcagt tggtaaagca ccgtatgtta tcaaagttga	4860
taaaccgctg cctccgatta aaagcagcgt taaaatcatt ggcaccgagt gggataaaac	4920
cgggtgaattt attgcaattg atggcagcaa ctatatcaaa ggcgaagggtg aaaaagcatg	4980
tccgggtgca aatccgggtc agtatggcac caatgttcgt accatgaccc tgcctgggtc	5040
ggttctgcaa gatgttaatg gtgttaccct gaaaggctctg gatgttcacg gtttttgtat	5100
tgggtgttctg gttaatcgca gcagcaataa cctgattcag cataatcgta tcagcaacaa	5160

```

ttatggtggt gccggtgtta tgattaccgg tgatgatggt aaaggtaatc cgaccagcac      5220
caccaccaat aataacaaag ttctggataa cgtgttcac c gataatggtg atggtctgga      5280
actgacccgt ggtgcagcat ttaatctgat tgcaataaac ctgtttacca gcacaaaagc      5340
caatccggaa ccgagccagg gtattgaaat tctgtggggg aatgataatg ccgtgggtggg      5400
taacaaattc gaaaactatt cagatggcct gcaaatcaat tggggtaaac gtaactatat      5460
cgcctataac gaactgacca ataacagcct ggggtttcaat ctgacagggtg atggtaacat      5520
tttcgacagc aataaagtgc atggtaaccg tattggtatt gccattcgta gtgaaaaaga      5580
tgccaatgca cgtattaccc tgaccaaaaa tcagatttgg gataacggca aagatatcaa      5640
acgttgtgaa gccggtggta gctgtgttcc gaatcagcgt ctgggtgcaa ttgtttttgg      5700
tgttccggca ctggaacatg aaggttttgt tggtagccgt ggcggtggtg ttgttattga      5760
accggcaaaa ctgcaaaaaa cctgcaccca gccgaaccag cagaattgta atgcaattcc      5820
taatcagggt attcaggcac cgaaactgac agttagcaaa aaacagctga ccgttgaagt      5880
taaaggcacc cctaatacgc gttataatgt ggaatTTTTT ggcaatcgta atgccagcag      5940
cagcgaagca gaacagtatc tgggtagcat tgttgttgtt accgatcatc agggctctggc      6000
aaaagcaaat tgggctccga aagttagcat gccgagcgtt accgcaaata tgacagatca      6060
tctgggtgcg accagcgaac tgagcagcgc agttaaatac cgttaaatac atgcgcgcgcg      6120
cgttcgcgcg gcgctTTTTT ttggtaccga gtcgaattc actggccgtc gttttacaac      6180
gtcgtgactg ggaaaaccct ggcgttacc c aacttaatac ccttgcagca catccccctt      6240
tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa cagttgcgca      6300
gcctgaatgg cgaatggcgc ctgatgcggt attttctcct tacgcatctg tgcggtatTT      6360
cacaccgcat atggtgcact ctcagtacaa tctgctctga tgccgcatag ttaagccagc      6420
cccgcacccc gccaacaccc gctgacgaat tc                                     6452

```

```

<210> 26
<211> 10012
<212> ADN
<213> Plasmid pMG47F;

```

```

<400> 26
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtgggtcag actggaaaat cagaggggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agaccgcga taaaacgccc      180
tgagaagccc gtgacgggct tttcttgtat tatgggtagt ttccttgcat gaatccataa      240
aaggcgctg tagtgccatt taccctcatt cactgccaga gccgtgagcg cagcgaactg      300

```

aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca	360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt aggggttttaa ggtctgtttt	420
gtagaggagc aaacagcgtt tgcgacatcc ttttgaata ctgcggaact gactaaagta	480
gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc	540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggtcta gcggaattta	600
cagaggggtct agcagaattt acaagttttc cagcaaaggt ctagcagaat ttacagatac	660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata	720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa	780
atgaactagc gatttagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct	840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt	900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat	960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta	1020
aaggctttga gattttccag tggacaaact atgccaagtt ctcaagcgaa aaattagaat	1080
tagtttttag tgaagagata ttgccttata ttttccagtt aaaaaattc ataaaatata	1140
atctggaaca tgttaagtct ttgaaaaca aatactctat gaggatttat gagtgggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtgggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttcttaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaaca acaaaagtag aacaactggt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgctcccgt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagcgcttt cagccggcaa acctgaagcc ggatctgcga	2160

ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtcccca	2280
attaattagt aatataatct attaaaggtc attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagtct cgggtaacat caaggcccga tccttgagac	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgcagtt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgccttccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgcccag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcggttc	2820
gtaagctgta atgcaagtag cgtatgcgct cacgcaactg gtccagaacc ttgaccgaac	2880
gcagcgggtg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tgttatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgaggga agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgctca	3120
tcgagcgcca tctcgaaccg acgttgctgg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttggtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcctgcag	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgcctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcgc agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960

tgcttttatt	atttttaagc	gtgcataata	agccctacac	aaattgggag	atatatcatg	4020
aaaggctggc	tttttcttgt	tatcgcaata	gttggcgaag	taatcgcaac	atccgcatta	4080
aaatctagcg	agggctttac	taagctgac	cgggtggatga	cctttttgaat	gaccttttaat	4140
agattatatt	actaattaat	tggggaccct	agaggtcccc	ttttttat	taaaaat	4200
ttcacaaaac	ggtttacaag	catacgttgg	ccgattcatt	aatgcagctg	gcacgacagg	4260
tttcccgact	ggaaagcggg	cagtgagcgc	aacgcaatta	atgtgagtta	gctcactcat	4320
taggcacccc	aggctttaca	ctttatgctt	ccggctcgta	tgttgtgtgg	aattgtgagc	4380
ggataacaat	ttcacacagg	aaacagctat	gaccatgatt	acgccaagct	tgctattgac	4440
gacagctatg	gttcactgtc	caccaaccaa	aactgtgctc	agtaccgcca	atatttctcc	4500
cttgaggggt	acaaagaggt	gtccctagaa	gagatccacg	ctgtgtaaaa	attttacaaa	4560
aaggtattga	ctttccctac	agggtgtgta	ataatttaat	tacaggcggg	ggcaaccccg	4620
cctgttctgc	agaggaggaa	tatagccatg	gaagtgtaaa	tcttcaacac	ccaggatggt	4680
caggattttc	tgcgtgttgc	aagcggctctg	gaacaagagg	gtggtaatcc	gcgtgttaaa	4740
caaattattc	atcgtgttct	gagcgacctg	tataaagcaa	ttgaagatct	gaatatcacc	4800
agcgacgaat	attgggcagg	cgttgcatat	ctgaatcagc	tgggtgtcaaa	tcaagaagca	4860
ggctctgctga	gtccgggtct	gggttttgat	cattatctgg	atatgcgtat	ggatgcagaa	4920
gatgcagcac	tgggtattga	aaatgcaaca	ccgcgtacca	ttgaagggtcc	gctgtatggt	4980
gcgggtgcac	cggaaagcgt	tggttatgca	cgcatggatg	atggtagcga	tccgaatggt	5040
cataccctga	ttctgcatgg	caccat	gatgcagatg	gtaaaccgct	gccgaatgca	5100
aaagttgaaa	tttggcatgc	aaacaccaaa	ggcttttata	gccattttga	tccgaccggt	5160
gaacagcagg	cctttaatat	gcgtcgtagc	attattaccg	atgagaatgg	tcagtatcgt	5220
gttcgtacca	ttctgcctgc	cggttatgg	tgtcctccgg	aagggtccgac	ccagcaactg	5280
ctgaaccaac	tgggtcgtca	tggtaatcgt	ccggcacata	ttcattat	tgttagcgca	5340
gatggtcacc	gtaaactgac	caccagatt	aatgttgccg	gtgatccgta	tacctatgat	5400
gattttgcat	atgccacccg	tgaaggctctg	gttgttgatg	cagttgaaca	taccgatccg	5460
gaagcaatta	aagccaatga	tgtggaagg	ccttttgccg	aaatgggtgtt	tgatctgaaa	5520
ctgacccgtc	tggttgatgg	tgttgataat	caggttgtgg	atcgtccg	tctggcagtt	5580
taatacacca	aaatggttca	aaattatcag	gcgagtgatc	atgatcactg	gcctgttttt	5640
atttcaggga	agggtggaga	caattacgtg	gataatcaga	tcatccaaga	aaccgtggat	5700
aaaattctga	gcgttctgcc	gaatcaggca	ggtcagctgg	cacgtctggt	gcgtctgatg	5760
caatttgcat	gcgatccgac	cattaccgtt	attggcaaat	ataaccatgg	taaaagccgt	5820

ctgctgaatg aactgattgg caccgatatc tttagcgttg cagataaacg tgaaccatt	5880
cagctggccg aacataaaca ggatcaggtt cgttggctgg atgcacctgg tctggatgcc	5940
gatgttgacg cagttgatga tcgtcatgca tttgaagcag tttggacca ggcagatatt	6000
cgtctgtttg ttcatagcgt tcgtgaaggt gaactggatg caaccgaaca ccatctgctg	6060
caacagctga ttgaagatgc cgatcatagc cgctcgcaga ccattctggt tctgaccag	6120
attgatcaga ttccggatca gaccatcctg acacagatta aaaccagcat tgcacagcag	6180
gttccgaaac tggatatattg ggcagttagc gcaaccgctc atcgtcaggg cattgaaaac	6240
ggtaaaacc tgcctgatcg aaaaagcggg attggtgcac tgcgccatac cctggaacag	6300
gcactggcac aggtgccgag cgcacgtacc tatgaaaaaa atcgctctgct gtcagatctg	6360
caccatcagc tgaacaact gctgctggat cagaaacatg ttctgcaaca actgcaacag	6420
acacagcaac agcagctgca tgattttgat accggtctga ttaacattct ggacaaaatt	6480
cgtgttgatc tggaaaccgat tgtgaatatt gatggtcagg atcaagcact gaatccggat	6540
agctttgcaa ccatgtttaa aaacaccgca gcaaacagc agcgtgcaa agttcagatt	6600
gcatatagcc gtgcatgcat tgaaatcaac agccatctga ttcgccatgg tgttgttggg	6660
ctgcctgcgg aacagcagac caccattaaa agcattgata ccgtgattgt tgccgtgttt	6720
ggatcagcg ttaaatttcg tgatcagctg cgtgcctgt tttataccga taccgaacgt	6780
cagcgtctgc aacgtgaatt tcgtttctat tttgaaaaaa gtgccggctc catgattctg	6840
gcagcaaaaa ttgaacagac catgcgtcag cagggctgta ttcagaatgc catgatggca	6900
ctgcaacaaa tggaaagcgc agcataaaaa cacggacgcc gcaaacggcg tccgaatttc	6960
ttggtcgacc gttaaactca tcaccgcaag ggataaatat ctaacaccgt gcgtgttgac	7020
tattttacct ctggcgggta taatggttgc atgtactaat ctagataagg aatatagcca	7080
tgaccgcacc gattcaggat ctgcgtgatg caattgccct gctgcaacag catgataatc	7140
agtatctgga aaccgatcat ccggttgatc cgaatgcaga actggcaggc gtttatcgtc	7200
atattgggtg cggtggcacc gttaaacgtc cgaccgtat tgggccggca atgatgttta	7260
ataacattaa aggttatccg cacagccgta ttctggttgg tatgcatgca agccgtcagc	7320
gtgcagcact gctgctgggt tgtgaagcaa gtcagctggc actggaagtt ggtaaagcag	7380
ttaaaaaacc gggtgcaccg gtggttggtc cggcaagcag cgcaccgtgt caagagcaga	7440
tttttctggc agatgatccg gatcttgatc tgcgtaccct gctgcctgca cataccaata	7500
ccccgattga tgcaggccg tttttttgtc tgggtctggc cctggcaagc gatccgggtg	7560
atgcaagcct gaccgatgtt accattcatc gtctgtgtgt tcagggtcgt gatgaactga	7620

gcatgttcct	ggcagcaggt	cgccatattg	aagtttttcg	tcagaaagca	gaagcagcag	7680
gtaaaccgct	gccgattacc	attaatatgg	gtctggaccc	agcaatctat	attggcgcat	7740
gttttgaagc	accgaccacc	ccgtttgggt	ataatgaact	gggtgttgcc	ggtgcactgc	7800
gtcagcgtcc	ggttgaactg	gttcaggggtg	ttagcgttcc	ggaaaaagca	attgcacgtg	7860
ccgaaattgt	tattgaaggt	gaactgctgc	ctgggtgttcg	tgttcgtgaa	gatcagcata	7920
ccaattcagg	tcatgcaatg	ccggaatttc	cggtttattg	tgggtggtgca	aatccgagcc	7980
tgccggttat	taaagttaaa	gccgttacca	tgcgcaataa	cgcaattctg	caaaccctgg	8040
ttggtccggg	tgaagaacat	accaccctgg	caggtctgcc	gaccgaagca	agcattttgga	8100
atgcagttga	agcagcaatt	ccgggttttc	tgcaaaatgt	ttatgccccat	accgcaggcg	8160
gtggtaaatt	tctgggtatt	ctgcaagtga	aaaaacgtca	gcctgccgat	gaaggctcgtc	8220
agggtcaggc	agccctgctg	gcgctggcaa	cctatagcga	actgaaaaat	atcattctgg	8280
tggatgagga	tgtggacatt	ttgatagtgt	atgatattct	gtgggcaatg	accacccgta	8340
tgcaagggtga	tgtagcatt	accaccattc	cgggtattcg	cggtcatcag	ctggaccgca	8400
gccagacacc	ggaatattca	ccgagcattc	gtggtaatgg	tattagctgc	aaaaccatct	8460
ttgattgtac	cgttccgtgg	gactgaaaa	gccattttga	acgtgcaccg	tttgcatatg	8520
ttgatccgcg	tccgtttgca	cctgaatatt	ttgcacgtct	ggaaaaaaat	cagggcagcg	8580
caaaataagc	taataacagg	cctgctggta	atcgcaggaa	tttttatattg	gatggatccg	8640
cctacctagc	ttccaagaaa	gatatcctaa	cagcacaaga	gcggaaagat	gttttgttct	8700
acatccagaa	caacctctgc	taaaattcct	gaaaaatttt	gcaaaaagtt	gttgacttta	8760
tctacaaggt	gtggtataat	aatcttaaca	acagcaggac	gtccccgggt	tgaggaaaac	8820
ctaataaaat	atagcctgtg	caccattagc	tttcgtcacc	agctgattag	ctttaccgat	8880
attgttcagt	ttgcctatga	aaacggcttt	gaaggatttg	aactgtgggg	cacccatgca	8940
cagaatctgt	atatgcaaga	atatgaaacc	accgaacgtg	aactgaattg	cctgaaagat	9000
aaaaccctgg	aaattaccat	gatcagcgat	tatctggata	ttagcctgag	cgcagatttt	9060
gaaaaaacca	tcgaaaaatg	tgaacagctg	gcaattctgg	ccaattgggt	taaaacgaac	9120
aaaattcgta	cctttgccgg	tcagaaaggt	agtgcagatt	ttagccagca	agaacgtcaa	9180
gagtatgtga	atcgtattcg	catgatttgt	gaactgtttg	cccagcataa	tatgtatgtt	9240
ctgctggaaa	cccatccgaa	taccctgacc	gataccctgc	cgagcaccct	ggaactgctg	9300
ggtgaagttg	atcatccgaa	tctgaaaatc	aacctggatt	ttctgcatat	ctgggaaagc	9360
ggtgcagatc	cgtttgatag	ctttcagcag	ctgcgtccgt	ggattcagca	ttatcacttt	9420
aaaaacatta	gcagcgcaga	ctatctgcat	gtgtttgaac	cgaataatgt	ttatgcagca	9480

```

gcaggtaatc gtaccggtat ggttccgctg tttgaaggca ttgttaacta tgatgaaatc      9540
atccaagaag tgcgcgatac cgatcatttt gcaagcctgg aatgggtttg tcataacgca      9600
aaagatattc tgaaagccga aatgaaagtg ctgaccaatc gtaatctgga agttgttacc      9660
agctaaatgc atgcgcgccg cgttcgcgcg gcgctttttt ttggtaccga gctcgaattc      9720
actggccgtc gttttacaac gtcgtgactg ggaaaaccct ggcgttacct aacttaatcg      9780
ccttgacgca catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg      9840
cccttcccaa cagttgcgca gcctgaatgg cgaatggcgc ctgatgcggt attttctcct      9900
tacgcatctg tgcggtatct cacaccgcat atgggtgact ctcagtacaa tctgctctga      9960
tgccgcatag ttaagccagc cccgacaccc gccaacaccc gctgacgaat tc              10012

```

```

<210> 27
<211> 10249
<212> ADN
<213> Plasmid pMG70;

```

```

<400> 27
gttgacagta agacgggtaa gcctgttgat gataccgctg ccttactggg tgcattagcc      60
agtctgaatg acctgtcacg ggataatccg aagtggtcag actggaaaat cagaggggcag      120
gaactgctga acagcaaaaa gtcagatagc accacatagc agacccgcc aaaaacgccc      180
tgagaagccc gtgacgggct tttcttgat tatgggtagt ttccttgcat gaatccataa      240
aaggcgctg tagtgccatt tccccccatt cactgccaga gccgtgagcg cagcgaactg      300
aatgtcacga aaaagacagc gactcaggtg cctgatggtc ggagacaaaa ggaatattca      360
gcgatttgcc cgagcttgcg aggggtgctac ttaagccttt agggttttta ggtctgtttt      420
gtagaggagc aaacagcggt tgcgacatcc ttttgtaata ctgcggaact gactaaagta      480
gtgagttata cacagggctg ggatctattc tttttatctt tttttattct ttctttattc      540
tataaattat aaccacttga atataaaca aaaaaacaca caaaggctca gcggaattta      600
cagaggggtct agcagaattt acaagttttc cagcaaaggc ctagcagaat ttacagatac      660
ccacaactca aaggaaaagg actagtaatt atcattgact agcccatctc aattgggtata      720
gtgattaaaa tcacctagac caattgagat gtatgtctga attagttggt ttcaaagcaa      780
atgaactagc gattagtcgc tatgacttaa cggagcatga aaccaagcta attttatgct      840
gtgtggcact actcaacccc acgattgaaa accctacaag gaaagaacgg acggtatcgt      900
tcacttataa ccaatacgct cagatgatga acatcagtag ggaaaatgct tatgggtgat      960
tagctaaagc aaccagagag ctgatgacga gaactgtgga aatcaggaat cctttgggtta      1020
aaggctttga gattttccag tggacaaact atgccaaagt ctcaagcgaa aaattagaat      1080

```

tagtttttag tgaagagata ttgccttata ttttccagtt aaaaaaattc ataaaatata	1140
atctggaaca tgttaagtct tttgaaaaca aatactctat gaggatttat gagtgggttat	1200
taaaagaact aacacaaaag aaaactcaca aggcaaatat agagattagc cttgatgaat	1260
ttaagttcat gttaatgctt gaaaataact accatgagtt taaaaggctt aaccaatggg	1320
ttttgaaacc aataagtaaa gatttaaaca cttacagcaa tatgaaattg gtggttgata	1380
agcgaggccg cccgactgat acgttgattt tccaagttga actagataga caaatggatc	1440
tcgtaaccga acttgagaac aaccagataa aaatgaatgg tgacaaaata ccaacaacca	1500
ttacatcaga ttctaccta cgtaacggac taagaaaaac actacacgat gctttaactg	1560
caaaaattca gctcaccagt tttgaggcaa aatttttgag tgacatgcaa agtaagcatg	1620
atctcaatgg ttcgttctca tggctcacgc aaaaacaacg aaccacacta gagaacatac	1680
tggctaaata cggaaggatc tgaggttctt atggctcttg tatctatcag tgaagcatca	1740
agactaacia acaaaagtag aacaactgtt caccgttaga tatcaaaggg aaaactgtcc	1800
atatgcacag atgaaaacgg tgtaaaaaag atagatacat cagagctttt acgagttttt	1860
ggtgcattta aagctgttca ccatgaacag atcgacaatg taacagatga acagcatgta	1920
acacctaata gaacaggtga aaccagtaaa acaaagcaac tagaacatga aattgaacac	1980
ctgagacaac ttgttacagc tcaacagtca cacatagaca gcctgaaaca ggcgatgctg	2040
cttatcgaat caaagctgcc gacaacacgg gagccagtga cgccctccgt ggggaaaaaa	2100
tcatggcaat tctggaagaa atagecgttt cagccggcaa acctgaagcc ggatctgcga	2160
ttctgataac aaactagcaa caccagaaca gcccgtttgc gggcagcaaa acccgttatg	2220
cttgtaaacc gttttgtgaa aaaattttta aaataaaaaa ggggacctct aggggtccca	2280
attaattagt aatataatct attaaaggct attcaaaagg tcatccaccg gatcaattcc	2340
cctgctcgcg caggctgggt gccaaagctct cgggtaacat caaggcccga tccttgagc	2400
ccttgccctc ccgcacgatg atcgtgccgt gatcgaaatc cagatccttg acccgagtt	2460
gcaaaccctc actgatccgc atgcccgttc catacagaag ctgggcgaac aaacgatgct	2520
cgccctccag aaaaccgagg atgcgaacca cttcatccgg ggtcagcacc accggcaagc	2580
gccgcgacgg ccgaggtctt ccgatctcct gaagccaggg cagatccgtg cacagcacct	2640
tgccgtagaa gaacagcaag gccgccaatg cctgacgatg cgtggagacc gaaaccttgc	2700
gctcgttcgc cagccaggac agaaatgcct cgacttcgct gctgccaag gttgccgggt	2760
gacgcacacc gtggaaacgg atgaaggcac gaaccagtg gacataagcc tgttcgggtc	2820
gtaagctgta atgcaagtag cgtatgcgct cagcgaactg gtccagaacc ttgaccgaac	2880

gcagcgggtgg taacggcgca gtggcggttt tcatggcttg ttatgactgt ttttttgggg	2940
tacagtctat gcctcgggca tccaagcagc aagcgcgtta cgccgtgggt cgatgtttga	3000
tggtatggag cagcaacgat gttacgcagc agggcagtcg ccctaaaaca aagttaaaca	3060
tcatgaggga agcggtgatc gccgaagtat cgactcaact atcagaggta gttggcgtca	3120
tcgagcgcca tctcgaaccg acgttgcttg ccgtacattt gtacggctcc gcagtggatg	3180
gcggcctgaa gccacacagt gatattgatt tgctggttac ggtgaccgta aggcttgatg	3240
aaacaacgcg gcgagctttg atcaacgacc ttttggaac ttcggcttcc cctggagaga	3300
gcgagattct ccgcgctgta gaagtcacca ttgttggtgca cgacgacatc attccgtggc	3360
gttatccagc taagcgcgaa ctgcaatttg gagaatggca gcgcaatgac attcttgacg	3420
gtatcttcga gccagccacg atcgacattg atctggctat cttgctgaca aaagcaagag	3480
aacatagcgt tgccttggtg ggtccagcgg cggaggaact ctttgatccg gttcctgaac	3540
aggatctatt tgaggcgcta aatgaaacct taacgctatg gaactcgccg cccgactggg	3600
ctggcgatga gcgaaatgta gtgcttacgt tgtcccgcat ttggtacagc gcagtaaccg	3660
gcaaaatcgc gccgaaggat gtcgctgccg actgggcaat ggagcgctg ccggcccagt	3720
atcagcccgt catacttgaa gctagacagg cttatcttgg acaagaagaa gatcgcttgg	3780
cctcgcgcgc agatcagttg gaagaatttg tccactacgt gaaaggcgag atcaccaagg	3840
tagtcggcaa ataatgtcta acaattcggt caagccgacg ccgcttcgcg gcgcggctta	3900
actcaagcgt tagatgcact aagcacataa ttgctcacag ccaaactatc aggtcaagtc	3960
tgcttttatt atttttaagc gtgcataata agccctacac aaattgggag atatatcatg	4020
aaaggctggc tttttcttgt tatcgcaata gttggcgaag taatcgcaac atccgcatta	4080
aatctagcg agggctttac taagctgatc cggtgatga ccttttgaat gacctttaat	4140
agattatatt actaattaat tggggaccct agaggcccc ttttttatth taaaaattht	4200
ttcacaaaac ggtttacaag catacgttgg ccgattcatt aatgcagctg gcacgacagg	4260
tttcccgaact ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gtcactcat	4320
taggcacccc aggcctttaca ctttatgctt ccggctcgta tgttggtggtg aattgtgagc	4380
ggataacaat ttcacacagg aaacagctat gaccatgatt acgccaagct tgctattgac	4440
gacagctatg gttcactgtc caccaaccaa aactgtgctc agtaccgcca atatttctcc	4500
cttgaggggt acaaagaggt gtccctagaa gagatccacg ctgtgtaaaa attttacaaa	4560
aaggtattga ctttccctac aggggtgtgta ataatttaat tacaggcggg ggcaaccccg	4620
cctgttctgc agaggaggaa tatagccatg gaagtgaaaa tcttcaacac ccaggatggt	4680
caggattttc tgcgtgttgc aagcggctctg gaacaagagg gtggtaatcc gcgtgttaaa	4740

caaattattc atcgtgttct gagcgacctg tataaagcaa ttgaagatct gaatatcacc	4800
agcgacgaat attgggcagg cgttgcatat ctgaatcagc tgggtgcaaa tcaagaagca	4860
ggtctgctga gtccgggtct gggttttgat cattatctgg atatgcgtat ggatgcagaa	4920
gatgcagcac tgggtattga aaatgcaaca ccgcgtacca ttgaagggtcc gctgtatggt	4980
gcgggtgcac cggaaagcgt tggttatgca cgcattggat atggtagcga tccgaatggt	5040
cataccctga ttctgcatgg caccattttt gatgcagatg gtaaaccgct gccgaatgca	5100
aaagttgaaa tttggcatgc aaacaccaaa ggcttttata gccattttga tccgaccggt	5160
gaacagcagg cctttaatat gcgtcgtagc attattaccg atgagaatgg tcagtatcgt	5220
gttcgtacca ttctgcctgc cggttatggt tgtcctccgg aagggtccgac ccagcaactg	5280
ctgaaccaac tgggtcgtca tggtaatcgt ccggcacata ttcattattt tgttagcgca	5340
gatggtcacc gtaaactgac caccagatt aatgttgccg gtgatccgta tacctatgat	5400
gattttgcat atgccacccg tgaagggtctg gttgttgatg cagttgaaca taccgatccg	5460
gaagcaatta aagccaatga tgtggaaggc ctttttgccg aaatgggtgtt tgatctgaaa	5520
ctgaccctgc tggttgatgg tgttgataat caggttgtgg atcgtccgcg tctggcagtt	5580
taatacacca aaatggttca aaattatcag gcgagtgatc atgatcactg gcctgttttt	5640
atttcagggg aggggtggaga caattacgtg gataatcaga tcatccaaga aaccgtggat	5700
aaaattctga gcgttctgcc gaatcaggca ggtcagctgg cacgtctggt gcgtctgatg	5760
caatttgcat gcgatccgac cattaccgtt attggcaaataa accatggg taaaagccgt	5820
ctgctgaatg aactgattgg caccgatatc tttagcgttg cagataaacg tgaaaccatt	5880
cagctggccg aacataaaca ggatcagggt cgttggctgg atgcacctgg tctggatgcc	5940
gatgttgcat cagttgatga tcgtcatgca tttgaagcag tttggacca ggcagatatt	6000
cgtctgtttg ttcatagcgt tcgtgaaggc gaactggatg caaccgaaca ccatctgctg	6060
caacagctga ttgaagatgc cgatcatagc cgtcgtcaga ccattctggt tctgaccag	6120
attgatcaga ttccggatca gaccatcctg acacagatta aaaccagcat tgcacagcag	6180
gttccgaaac tggatatttg ggcagttagc gcaaccctgc atcgtcaggg cattgaaaac	6240
ggtaaaaccc tgctgatcga aaaaagcggc attggtgcac tgcgccatac cctggaacag	6300
gcactggcac aggtgccgag cgcacgtacc tatgaaaaaa atcgtctgct gtcagatctg	6360
caccatcagc tgaaacaact gctgctggat cagaaacatg ttctgcaaca actgcaacag	6420
acacagcaac agcagctgca tgattttgat accggtctga ttaacattct ggacaaaatt	6480
cgtgttgatc tggaaccgat tgtgaatatt gatggtcagg atcaagcact gaatccggat	6540

agctttgcaa ccatgttttaa aaacaccgca gcaaaacagc agcgtgccaa agttcagatt	6600
gcatatagcc gtgcatgcat tgaaatcaac agccatctga ttcgccatgg tgttggtggt	6660
ctgcctgctg aacagcagac caccattaaa agcattgata ccgtgattgt tgccgtgttt	6720
ggtatcagcg ttaaatttcg tgatcagctg cgtgccctgt tttataccga taccgaacgt	6780
cagcgtctgc aacgtgaatt tcgtttctat tttgaaaaaa gtgccggctc catgattctg	6840
gcagcaaaaa ttgaacagac catgcgtcag cagggctgta ttcagaatgc catgatggca	6900
ctgcaacaaa tggaaagcgc agcataaaaa cacggacgcc gcaaacggcg tccgaatttc	6960
ttggtcgacc gttaaactta tcaccgcaag ggataaatat ctaacaccgt gcgtgttgac	7020
tattttacct ctggcggtga taatggttgc atgtactaat ctagataagg aatatagcca	7080
tgaccgcacc gattcaggat ctgcgtgatg caattgccct gctgcaacag catgataatc	7140
agtatctgga aaccgatcat ccggttgatc cgaatgcaga actggcaggc gtttatcgtc	7200
atattgggtgc cgggtggcacc gttaaactgc cgaccctgat tgggtccggca atgatgttta	7260
ataacattaa aggttatccg cacagccgta ttctggttgg tatgcatgca agccgtcagc	7320
gtgcagcact gctgctgggt tgtgaagcaa gtcagctggc actggaagtt ggtaaagcag	7380
ttaaaaaacc gggtgcaccg gtggttgttc cggcaagcag cgcaccgtgt caagagcaga	7440
tttttctggc agatgatccg gatcttgatc tgcgtaccct gctgcctgca cataccaata	7500
ccccgattga tgcaggctccg tttttttgtc tgggtctggc cctggcaagc gatccgggtg	7560
atgcaagcct gaccgatgtt accattcatc gtctgtgtgt tcagggtcgt gatgaactga	7620
gcatgttcct ggcagcaggc cgccatattg aagtttttcg tcagaaagca gaagcagcag	7680
gtaaaccgct gccgattacc attaatatgg gtctggacce agcaatctat attggcgcat	7740
gttttgaagc accgaccacc ccgtttgggt ataatgaact ggggtgttgc ggtgcactgc	7800
gtcagcgtcc gggtgaactg gttcaggggtg ttagcgttcc ggaaaaagca attgcacgtg	7860
ccgaaattgt tattgaaggc gaactgctgc ctggtgttcg tgttcgtgaa gatcagcata	7920
ccaattcagg tcatgcaatg ccggaatttc cgggttattg tgggtggtgca aatccgagcc	7980
tgccggttat taaagttaaa gccgttacca tgcgcaataa cgcaattctg caaaccttgg	8040
ttgggtccggg tgaagaacat accaccctgg caggtctgcc gaccgaagca agcatttgga	8100
atgcagttga agcagcaatt ccgggttttc tgcaaatgt ttatgcccat accgcaggcg	8160
gtggtaaatt tctgggtatt ctgcaagtga aaaaacgtca gcctgccgat gaaggctcgtc	8220
agggtcaggc agccctgctg gcgctggcaa cctatagcga actgaaaaat atcattctgg	8280
tggatgagga tgtggacatt tttgatagtg atgatattct gtgggcaatg accaccgta	8340
tgcagggtga tgtagcatt accaccattc cgggtattcg cggtcacag ctggaccgca	8400

gccagacacc ggaatattca ccgagcattc gtggtaatgg tattagctgc aaaaccatct	8460
ttgattgtac cgttccgtgg gcaactgaaa gccattttga acgtgcaccg tttgcagatg	8520
ttgatccgcg tccgtttgca cctgaatatt ttgcacgtct ggaaaaaaat cagggcagcg	8580
caaaataagc taataacagg cctgctggta atcgcaggaa tttttatttg gatggatccg	8640
cctacctagc ttccaagaaa gatataccta cagcacaaga gcggaaaagat gttttgttct	8700
acatccagaa caacctctgc taaaattcct gaaaaatttt gcaaaaagtt gttgacttta	8760
tctacaaggt gtggtataat aatcttaaca acagcaggac gctcccgggt tgaggaaaac	8820
ctaatagccg gcaaaactggc aattagcagc atgagcctgg gtcgttggtt tgcaggatcat	8880
agcctggata gtaaactgga tgcagcacag cgttatgggt atctgggtat tgaactgttt	8940
tatgaggatc tggttgatgt tgcagaacat ctgagcaatg aacgtccgag tccggaaggt	9000
ccgtttgttg aagcacagat tgcagcagca cgtcatattc tgcaaatgtg tcaggcacgt	9060
ggtctggaag ttgtttgtct gcaaccgttt atgcattatg atgggtctgaa tgatcgtgcc	9120
gaacatgaac gtcgtctgga aaaactggca ctgtggattg aactggcaca tgaactgcat	9180
accgatatta ttcagattcc ggcaaatttt ctgcctgcaa atcagggttag cgataatctg	9240
gatctgattg ttagcgatct gtgtaaagtt gcagatattg gtgcacaggc actgcctccg	9300
attcgttttg catatgaaag cctgtgttgg agcaccctg tttgatctgtg ggaacgttgt	9360
tgggatattg ttcagcgtgt ggatcgtccg aattttggta tttgtctgga tacctttaac	9420
atcctgggtc gcatttatgc agatccgacc agcccgagcg gtcgtacccc gaatgcaaaa	9480
gaagcagttc gtaaaagcat tgccaatctg gttagccgtg tggatgtag caaagttttt	9540
tatgttcagg ttgtggatgc cgaacgtctg agtaaaccgc tgctgcctgg tcatccgtat	9600
tataaccggt aacagcctgc acgtatgagc tggtcacgta attgtcgtct gttctatggt	9660
gaaaccgaat atggtgcata tctgccgggt aaagaagttg cacgcgcact gtttcatggt	9720
attggttttg aaggttgggt tagcctggaa ctgtttaatc gtcgtatgag cgaagaaggt	9780
ccggaagttc ctgaagaact ggccatgcgt ggtgcaatta gctgggcaaa actggttcag	9840
gatctgcgta ttccggttga aggtccgtg gttaccatgc ctggtgtag cgcaagcctg	9900
taaatgcatg cgcgccgctg tcgcgcggcg cttttttttg gtaccgagct cgaattcact	9960
ggcgcgtggt ttacaacgtc gtgactggga aaaccctggc gttacccaac ttaatgcct	10020
tgcagcacat ccccttttcg ccagctggcg taatagcgaa gagggccgca ccgatcgccc	10080
ttcccaacag ttgctcagcc tgaatggcga atggcgctg atgcggtatt ttctccttac	10140
gcatctgtgc ggtatttcac accgcataatg gtgcactctc agtacaatct gctctgatgc	10200

cgcatagtta agccagcccc gacacccgcc aacacccgct gacgaattc

10249

<210> 28

<211> 7623

<212> ADN

<213> Plasmid pCP32AMP;

<400> 28

acatgaatgg tcttcggttt ccgtgtttcg taaagtctgg aaacgcggaa gtcagcgccc	60
tgcaccatta tgttccggat ctgcatcgca ggatgctgct ggctaccctg tggaacacct	120
acatctgtat taacgaagcg ctggcattga ccctgagtga tttttctctg gtcccgccgc	180
atccataaccg ccagttgttt accctcacia cgttccagta accgggcatg ttcattcatca	240
gtaacccgta tcgtgagcat cctctctcgt ttcattcgga tcattacccc catgaacaga	300
aattccccct tacacggagg catcaagtga ccaaacagga aaaaaccgcc cttaacatgg	360
cccgttttat cagaagccag acattaacgc ttctggagaa actcaacgag ctggacgcgg	420
atgaacaggc agacatctgt gaatcgcttc acgaccacgc tgatgagctt taccgcagct	480
gcctcgcgcg tttcgggtgat gacgggtgaaa acctctgaca catgcagctc ccggagacgg	540
tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg	600
gtgttgggcg gtgtcggggc gcagccatga ccagtcacg tagcgatagc ggagtgtata	660
ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata tgcgggtgtga	720
aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcttccg cttcctcgct	780
cactgactcg ctgcgctcgg tcgttcgggt gcggcgagcg gtatcagctc actcaaaggc	840
ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg	900
ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg	960
ccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg	1020
actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac	1080
cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca	1140
tagctcacgc tgtaggtatc tcagttcggg ttaggtcggt cgctccaagc tgggctgtgt	1200
gcacgaaccc ccggttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc	1260
caaccgggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag	1320
agcgaggtat gtaggcgggtg ctacagagtt cttgaagtgg tggcctaact acggctacac	1380
tagaaggaca gtatttggtg tctgcgctct gctgaagcca gttaccttcg gaaaaagagt	1440
tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa	1500
gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg	1560

gtctgacgct	cagtggaacg	aaaactcacg	ttaagggatt	ttggatcatga	gattatcaaa	1620
aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaatcaa	tctaaagtat	1680
atatgagtaa	acttggtctg	acagttacca	atgcttaatc	agtgaggcac	ctatctcagc	1740
gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	gtcgtgtaga	taactacgat	1800
acgggagggc	ttaccatctg	gccccagtgc	tgcaatgata	ccgcgagacc	cacgctcacc	1860
ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	gccgagcgca	gaagtgggtcc	1920
tgcaacttta	tccgcctcca	tccagtctat	taattgttgc	cggaagcta	gagtaagtag	1980
ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	gcaggcatcg	tggtgtcacg	2040
ctcgtcgttt	ggtatggctt	cattcagctc	cggttcccaa	cgatcaaggc	gagttacatg	2100
atcccccatg	ttgtgcaaaa	aagcggttag	ctccttcggt	cctccgatcg	ttgtcagaag	2160
taagttggcc	gcagtgttat	cactcatggt	tatggcagca	ctgcataatt	ctcttactgt	2220
catgccatcc	gtaagatgct	tttctgtgac	tggtgagtac	tcaaccaagt	cattctgaga	2280
atagtgtatg	cggcgaccga	gttgcctctg	cccggcgta	acacgggata	ataccgcgcc	2340
acatagcaga	actttaaaag	tgctcatcat	tggaacacgt	tcttcggggc	gaaaactctc	2400
aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	actcgtgcac	ccaactgatc	2460
ttcagcatct	tttactttca	ccagcgtttc	tggtgagca	aaaacaggaa	ggcaaaatgc	2520
cgcaaaaaag	ggaataaggg	cgacacggaa	atggtgaata	ctcatactct	tcctttttca	2580
atattattga	agcatttatc	agggttattg	tctcatgagc	ggatacatat	ttgaatgtat	2640
ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	cgaaaagtgc	cacctgacgt	2700
ctaagaaacc	attattatca	tgacattaac	ctataaaaa	aggcgtatca	cgaggccctt	2760
tcgtcttcaa	gaattctgaa	ccagtcctaa	aacgagtaaa	taggaccggc	aattcttcaa	2820
gcaataaaca	ggaataccaa	ttattaaaag	ataacttagt	cagatcgtac	aataaagctt	2880
tgaagaaaaa	tgcgcttat	tcaatctttg	ctataaaaa	tgcccaaaa	tctcacattg	2940
gaagacattt	gatgacctca	tttctttcaa	tgaagggcct	aacggagtgt	actaatgttg	3000
tggaatattg	gagcgataag	cgtgcttctg	ccgtggccag	gacaacgtat	actcatcaga	3060
taacagcaat	acctgatcac	tacttcgcac	tagtttctcg	gtactatgca	tatgatccaa	3120
tatcaaagga	aatgatagca	ttgaaggatg	agactaatcc	aattgaggag	tggcagcata	3180
tagaacagct	aaagggtagt	gctgaaggaa	gcatacgata	ccccgcatgg	aatgggataa	3240
tatcacagga	ggtactagac	tacctttcat	cctacataaa	tagacgcata	taagtacgca	3300
tttaagcata	aacacgcact	atgccgttct	tctcatgtat	atatatatac	aggcaacacg	3360
cagatatagg	tgcgacgtga	acagtgagct	gtatgtgcgc	agctcgcgtt	gcattttcgg	3420

aagcgctcgt tttcggaaac gctttgaagt tcctattccg aagttcctat tctctagaaa	3480
gtataggaac ttcagagcgc ttttgaaaac caaaagcgc ctgaagacgc actttcaaaa	3540
aaccaaaaac gcaccggact gtaacgagct actaaaatat tgccaatacc gcttccacaa	3600
acattgctca aaagtatctc tttgctatat atctctgtgc tataatcccta tataacctac	3660
ccatccacct ttcgctcctt gaacttgcct ctaaactcga cctctacatt ttttatgttt	3720
atctctagta ttactcttta gacaaaaaaa ttgtagtaag aactattcat agagtgaatc	3780
gaaaacaata cgaaaatgta aacatttcct atacgtagta tatagagaca aaatagaaga	3840
aaccgttcat aattttctga ccaatgaaga atcatcaacg ctatcacttt ctgttcacaa	3900
agtatgcgca atccacatcg gtatagaata taatcgggga tgcctttatc ttgaaaaaat	3960
gcacccgcag cttcgctagt aatcagtaaa cgcggaagt ggagtcaggc ttttttatg	4020
gaagagaaaa tagacaccaa agtagccttc ttctaacttc aacggaccta cagtgcacaa	4080
agttatcaag agactgcatt atagagcgc caaaggagaa aaaaagtaat ctaagatgct	4140
ttgttagaaa aatagcgc tcgggatgca tttttgtaga acaaaaaaga agtatagatt	4200
ctttgttggt aaaatagcgc tctcgcgttg catttctgtt ctgtaaaaaat gcagctcaga	4260
ttctttgttt gaaaaattag cgctctcgcg ttgcattttt gttttacaaa aatgaagcac	4320
agattcttcg ttggtaaaat agcgctttcg cgttgcattt ctgttctgta aaaatgcagc	4380
tcagattctt tgtttgaaaa attagcgc tcgcgttgca tttttgttct acaaaatgaa	4440
gcacagatgc ttcgttaaca aagatatgct attgaagtgc aagatggaaa cgcagaaaat	4500
gaaccgggga tgcgacgtgc aagattacct atgcaataga tgcaatagtt tctccaggaa	4560
ccgaaataca tacattgtct tccgtaaagc gctagactat atattattat acaggttcaa	4620
atatactatc tgtttcaggg aaaactccca gggtcggatg ttcaaaattc aatgatgggt	4680
aacaagtacg atcgtaaata tgtaaaacag tttgtcggat attaggctgt atctcctcaa	4740
agcgatttcg aatatcattg agaagctgca gcgtcacatc ggataataat gatggcagcc	4800
attgtagaag tgccttttgc atttctagtc tctttctcgg tctagctagt ttactacat	4860
cgcgaagata gaatcttaga tcacactgcc tttgctgagc tggatcaata gagtaacaaa	4920
agagtggtaa ggcctcgtaa aaggacaagg acctgagcgg aagtgtatcg tacagtagac	4980
ggagtatact agtatagtct atagtccgtg gaattctcat gtttgacagc ttatcatcga	5040
taagcttttc aattcaattc atcatttttt ttttattctt ttttttgatt tcggtttctt	5100
tgaaattttt ttgattcggg aatctccgaa cagaaggaag aacgaaggaa ggagcacaga	5160
cttagattgg tatatatagc catatgtagt gttgaagaaa catgaaattg cccagtattc	5220

ttaacccaac tgcacagaac aaaaacctgc aggaaacgaa gataaatcat gtcgaaagct	5280
acataataagg aacgtgctgc tactcatcct agtcctgttg ctgccaaagct atttaatatc	5340
atgcacgaaa agcaaacaaa cttgtgtgct tcattggatg ttcgtaccac caaggaatta	5400
ctggagttag ttgaagcatt aggtcccaaa atttgtttac taaaaacaca tgtggatatc	5460
ttgactgatt tttccatgga gggcacagtt aagccgctaa aggcattatc cgccaagtac	5520
aatTTTTTtac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag	5580
tactctgcgg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacggtgtg	5640
gtgggcccag gtattgttag cggtttgaag caggcggcag aagaagtaac aaaggaacct	5700
agaggccttt tgatgttagc agaattgtca tgcaagggtc ccctatctac tggagaatat	5760
actaagggtta ctgttgacat tgcgaagagc gacaaagatt ttgttatcgg ctttattgct	5820
caaagagaca tgggtggaag agatgaaggt tacgattggt tgattatgac acccggtgtg	5880
ggtttagatg acaagggaga cgcattgggt caacagtata gaaccgtgga tgatgtggtc	5940
tctacaggat ctgacattat tattgttgga agaggactat ttgcaaaggg aagggatgct	6000
aaggtagagg gtgaacgtta cagaaaagca ggctgggaag catatttgag aagatgcggc	6060
cagcaaaact aaaaaactgt attataagta aatgcatgta tactaaactc acaaattaga	6120
gcttcaattt aattatatca gttattaccg gggaatctcg gtcgtaatga cttgaaataa	6180
ttaacaaaca aaggagttac agttagaaat tgtaggagag atctcgtttt tcgcgacaat	6240
ctggcgtttt tcttgctaatt tccaggatta atccgttcat agtgtaaaac cccgtttaca	6300
cattctgacg gaagatatag attggaagta ttgcattcac taagataagt atggcaacac	6360
tggaacagac atgaattatc agaacgacga tttacgcac aaagaaatca aagagttact	6420
tcctcctgtc gcattgctgg aaaaattccc cgctactgaa aatgccgcga atacggttgc	6480
ccatgcccga aaagcgatcc ataagatcct gaaaggtaat gatgatcgcc tgttggttgt	6540
gattggccca tgctcaattc atgatcctgt cgcggcaaaa gagtatgcca ctcgcttgct	6600
ggcgctgcgt gaagagctga aagatgagct ggaaatcgta atgcgcgtct attttgaaaa	6660
gccgcgtacc acggtgggct ggaaagggt gattaacgat ccgcatatgg ataatagctt	6720
ccagatcaac gacggtctgc gtatagcccc taaattgctg cttgatatta acgacagcgg	6780
tctgccagcg gcaggtgagt ttctcgatat gatcacccca caatatctcg ctgacctgat	6840
gagctggggc gcaattggcg cacgtaccac cgaatcgag gtgcaccgcg aactggcatc	6900
agggctttct tgtccggctg gcttcaaaaa tggcaccgac ggtacgatta aagtggctat	6960
cgatgccatt aatgccgcg gtgcgccgca ctgcttctg tccgtaacga aatgggggca	7020
ttcggcgatt gtgaatacca gcggtaacgg cgattgccat atcattctgc gcggcggtaa	7080

```

agagcctaac tacagcgca agcacgttgc tgaagtgaag gaagggctga acaaagcagg 7140
cctgccagca caggtgatga tcgatttcag ccatgctaac tcgtccaaac aattcaaaaa 7200
gcagatggat gtttgtgctg acgtttgcc gcagattgcc ggtggcgaaa aggccattat 7260
tggcgtgatg gtggaaagcc atctggtgga aggcaatcag agcctcgaga gcggggagcc 7320
gctggcctac ggtaagagca tcaccgatgc ctgcatcggc tgggaagata ccgatgctct 7380
gttacgtcaa ctggcgaatg cagtaaaagc gcgtcgcggg taagggttaa ttgtcggatg 7440
cgccgtcaga gtggcgtatc cgatgaatca ccacaggcct gataagtcgc gcagcgtcgc 7500
atcaggcaat gtgctccatt gttagcaaca aaaaagccga ctcaattgca gtcggctttc 7560
tcattttaaa cgaatgacgt ttacttcgct ttaccctggg ttgcaaccgc cgctgctttc 7620
gct 7623

```

```

<210> 29
<211> 7630
<212> ADN
<213> Plasmid pCP14;

```

```

<400> 29
ctcgaggcta ttgacgacag ctatggttca ctgtccacca accaaaaactg tgctcagtag 60
cgccaatatt tctcccttga ggggtacaaa gaggtgtccc tagaagagat ccacgctgtg 120
taaaaatttt aaaaaagggt attgactttc cctacagggt gtgtaataat ttaattacag 180
gcggggggcaa ccccgctgt tctagaggag gaggaatcgc catggagagg attgtcggtta 240
ctctcgggga acgtagttag ccaattacca tcgcatctgg tttgtttaat gaaccagctt 300
cattcttacc gctgaaatcg ggcgagcagg tcatgttggg caccaacgaa accctggctc 360
ctctgtatct cgataaggtc cgcggcgtag ttgaacaggc ggggtgttaac gtcgatagcg 420
ttatcctccc tgacggcgag cagtataaaa gcctggctgt actcgatacc gtctttacgg 480
cgttggttaca aaagccgcat ggtcgcgata ctacgctggg ggcgcttggc ggcggcgtag 540
tgggcgatct gaccggcttc gcggcgcgga gttatcagcg cgggtgttcgt ttcatccaag 600
tcccgaacgac gttactgtcg caggctgatt cctccgttgg cggcaaaaact gcggtcaacc 660
atccccctcg taaaaacatg attggcgcggt tctaccagcc tgcttcagtg gtggtggatc 720
tcgactgtct gaaaacgctt ccccgcggtg agttagcgtc ggggctggca gaagtcatca 780
aatacggcat tattcttgac ggtgcgtttt tcaactgggt ggaagagaat ctggatgcgt 840
tggtgcgtct ggacgggtccg gcaatggcgt actgtattcg ccgttggtgt gaactgaagg 900
cagaagttgt cgccgccgac gagcgcgaaa ccgggttacg tgctttactg aatctgggac 960
acacctttgg tcatgccatt gaagctgaaa tgggggtatgg caattgggtta catggtgaag 1020

```

cggtcgctgc	gggtatggtg	atggcggcgc	ggacgtcggg	acgtctcggg	cagtttagtt	1080
ctgccgaaac	gcagcgtatt	ataaccctgc	tcacgcgggc	tgggttaccg	gtcaatgggc	1140
cgcgcgaaat	gtccgcgcag	gcgtatttac	cgcatatgct	gcgtgacaag	aaagtccttg	1200
cgggagagat	gcgcttaatt	cttccggttg	caattggtaa	gagtgaagtt	cgcagcggcg	1260
tttcgcacga	gcttgttctt	aacgccattg	ccgattgtca	atcagcgtaa	tcatcgttca	1320
tgctgatgc	cgctatgtag	gccggataag	gcgttcacgc	cgcacccggc	aaccgatgcc	1380
tgatgcgacg	cggtcgcgtc	ttatcaggcc	tacaggtcga	tgccgatatg	tacatcgatat	1440
tcggcaatta	atacatagca	acatgaatgg	tcttcggttt	ccgtgtttcg	taaagtctgg	1500
aaacgcggaa	gtcagcgcgc	tgcaccatta	tgttccggat	ctgcatcgca	ggatgctgct	1560
ggctaccctg	tggaacacct	acatctgtat	taacgaagcg	ctggcattga	ccctgagtga	1620
tttttctctg	gtcccgcgcg	atccataccg	ccagttgttt	accctcacia	cgttccagta	1680
accgggcatg	ttcatcatca	gtaacccgta	tcgtgagcat	cctctctcgt	ttcatcggtg	1740
tcattacccc	catgaacaga	aattccccct	tacacggagg	catcaagtga	ccaaacagga	1800
aaaaaccgcc	cttaacatgg	cccgccttat	cagaagccag	acattaacgc	ttctggagaa	1860
actcaacgag	ctggacgcgg	atgaacaggc	agacatctgt	gaatcgcttc	acgaccacgc	1920
tgatgagctt	taccgcagct	gcctcgcgcg	tttcggtgat	gacggtgaaa	acctctgaca	1980
catgcagctc	ccggagacgg	tcacagcttg	tctgtaagcg	gatgccggga	gcagacaagc	2040
ccgtcagggc	gcgtcagcgg	gtggtggcgg	gtgtcggggc	gcagccatga	cccagtcacg	2100
tagcgatagc	ggagtgtata	ctggcttaac	tatgcggcat	cagagcagat	tgtactgaga	2160
gtgcaccata	tgcggtgtga	aataccgcac	agatgcgtaa	ggagaaaata	ccgcatcagg	2220
cgctcttccg	cttcctcgct	cactgaactcg	ctgcgctcgg	tcgttcggct	gcggcgagcg	2280
gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	aatcagggga	taacgcagga	2340
aagaacatgt	gagcaaaaagg	ccagcaaaaag	gccaggaacc	gtaaaaaggc	cgcgttgctg	2400
gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	ctcaagtcag	2460
aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	ttccccctgg	aagctccctc	2520
gtgcgctctc	ctgttccgac	cctgcgcgtt	accggatacc	tgtccgcctt	tctcccttcg	2580
ggaagcgtgg	cgttttctca	tagctcacgc	tgtaggtatc	tcagttcggt	gtaggtcggt	2640
cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	cgccttatcc	2700
ggtaactatc	gtcttgagtc	caacccggtg	agacacgact	tatcgccact	ggcagcagcc	2760
actggtaaca	ggattagcag	agcgaggtat	gtaggcggtg	ctacagagtt	cttgaagtgg	2820

tggcctaact	acggctacac	tagaaggaca	gtatTTtgga	tctgcgctct	gctgaagcca	2880
gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	2940
ggtggttttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	3000
cctttgatct	tttctacggg	gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	3060
ttggtcatga	gattatcaaa	aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	3120
tttaaataca	tctaaagtat	atatgagtaa	acttggtctg	acagttacca	atgcttaatc	3180
agtgaggcac	ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	3240
gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	gccccagtgc	tgcaatgata	3300
ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	3360
gccgagcgca	gaagtgggcc	tgcaacttta	tccgcctcca	tccagtctat	taattggtgc	3420
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	3480
gcaggcatcg	tggtgtcacg	ctcgtcgttt	ggatatggctt	cattcagctc	cggttcccaa	3540
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcgggtag	ctccttcggt	3600
cctccgatcg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggt	tatggcagca	3660
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tggtgagtag	3720
tcaaccaagt	cattctgaga	atagtgtatg	cggcgaccga	gttgctcttg	cccggcgcta	3780
acacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggaaaacgt	3840
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	3900
actcgtgcac	ccaactgata	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	3960
aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	4020
ctcatactct	tcctttttca	atattattga	agcattttatc	agggttattg	tctcatgagc	4080
ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	4140
cgaaaagtgc	cacctgacgt	ctaagaaacc	attattatca	tgacattaac	ctataaaaaat	4200
aggcgatatca	cgaggccctt	tcgtcttcaa	gaattctgaa	ccagtcctaa	aacgagtaaa	4260
taggaccggc	aattcttcaa	gcaataaaca	ggaataccaa	ttattaaaag	ataacttagt	4320
cagatcgtac	aataaagctt	tgaagaaaaa	tgcgcttat	tcaatctttg	ctataaaaaa	4380
tggcccaaaa	tctcacattg	gaagacattt	gatgacctca	tttctttcaa	tgaagggcct	4440
aacggagttg	actaatgttg	tgggaaattg	gagcgataag	cgtgcttctg	ccgtggccag	4500
gacaacgtat	actcatcaga	taacagcaat	acctgatcac	tacttcgcac	tagtttctcg	4560
gtactatgca	tatgatccaa	tatcaaagga	aatgatagca	ttgaaggatg	agactaatcc	4620
aattgaggag	tggcagcata	tagaacagct	aaagggtagt	gctgaaggaa	gcatacgata	4680

ccccgcatgg aatgggataa tatcacagga ggtactagac tacctttcat cctacataaa	4740
tagacgcata taagtacgca ttttaagcata aacacgcact atgccgttct tctcatgtat	4800
atatatatac aggcaacacg cagatatagg tgcgacgtga acagtgagct gtatgtgcgc	4860
agctcgcgtt gcattttcgg aagcgctcgt tttcggaaac gctttgaagt tcctattccg	4920
aagttcctat tctctagaaa gtataggaac ttcagagcgc ttttgaaaac caaaagcgct	4980
ctgaagacgc actttcaaaa aaccaaaaac gcaccggact gtaacgagct actaaaatat	5040
tgcgaatacc gcttccacaa acattgctca aaagtatctc tttgctatat atctctgtgc	5100
tatatcccta tataacctac ccatccacct ttcgctcctt gaacttgcac ctaaactcga	5160
cctctacatt ttttatgttt atctctagta ttactcttta gacaaaaaaa ttgtagtaag	5220
aactattcat agagtgaatc gaaaacaata cgaaaatgta aacatttcct atacgtagta	5280
tatagagaca aaatagaaga aaccgttcat aattttctga ccaatgaaga atcatcaacg	5340
ctatcacttt ctgttcacaa agtatgcgca atccacatcg gtatagaata taatcgggga	5400
tgcctttatc ttgaaaaaat gcaccgcag cttcgctagt aatcagtaaa cgcgggaagt	5460
ggagtcaggc tttttttatg gaagagaaaa tagacaccaa agtagccttc ttctaacctt	5520
aacggacctc cagtgcacaa agttatcaag agactgcatt atagagcgca caaaggagaa	5580
aaaaagtaat ctaagatgct ttgttagaaa aatagcgctc tcgggatgca tttttgtaga	5640
acaaaaaaga agtatagatt ctttgttggt aaaatagcgc tctcgcgttg catttctggt	5700
ctgtaaaaaat gcagctcaga ttctttgttt gaaaaattag cgctctcgcg ttgcattttt	5760
gttttacaaa aatgaagcac agattcttcg ttggtaaaat agcgccttcg cgttgcattt	5820
ctgttctgta aaaatgcagc tcagattctt tgtttgaaaa attagcgctc tcgcgttgca	5880
tttttgttct acaaaatgaa gcacagatgc ttcgttaaca aagatatgct attgaagtgc	5940
aagatggaaa cgcagaaaat gaaccgggga tgcgacgtgc aagattacct atgcaataga	6000
tgcaatagtt tctccaggaa ccgaaataca tacattgtct tccgtaaagc gctagactat	6060
atattattat acaggttcaa atatactatc tgtttcaggg aaaactccca ggttcggatg	6120
ttcaaaatcc aatgatgggt aacaagtacg atcgtaaadc tgtaaaacag tttgtcggat	6180
attaggctgt atctcctcaa agcgtattcg aatatcattg agaagctgca gcgtcacatc	6240
ggataataat gatggcagcc attgtagaag tgccttttgc atttctagtc tctttctcgg	6300
tctagctagt ttactacat cgcgaagata gaatcttaga tcacactgcc tttgctgagc	6360
tggatcaata gagtaacaaa agagtggtaa ggctcgtta aaggacaagg acctgagcgg	6420
aagtgtatcg tacagtagac ggagtatact agtatagtct atagtccgtg gaattctcat	6480

```

gtttgacagc ttatcatcga taagcttttc aattcaattc atcatttttt ttttattctt 6540
ttttttgatt tcggtttctt tgaaattttt ttgattcggg aatctccgaa cagaaggaag 6600
aacgaaggaa ggagcacaga cttagattgg tatatatacg catatgtagt gttgaagaaa 6660
catgaaattg cccagtattc ttaacccaac tgcacagaac aaaaacctgc aggaaacgaa 6720
gataaatcat gtcgaaagct acatataagg aacgtgctgc tactcatcct agtcctgttg 6780
ctgccaagct atttaatatc atgcacgaaa agcaaacaaa cttgtgtgct tcattggatg 6840
ttcgtaccac caaggaatta ctggagttag ttgaagcatt aggtcccaaa atttgtttac 6900
taaaaacaca tgtggatatc ttgactgatt tttccatgga gggcacagtt aagccgctaa 6960
aggcattatc cgccaagtac aattttttac tcttcgaaga cagaaaattt gctgacattg 7020
gtaatacagt caaattgcag tactctgcgg gtgtatacag aatagcagaa tgggcagaca 7080
ttacgaatgc acacggtgtg gtgggcccag gtattgttag cggtttgaag caggcggcag 7140
aagaagtaac aaaggaacct agaggccttt tgatgttagc agaattgtca tgcaagggct 7200
ccctatctac tggagaatat actaagggtta ctgttgacat tgcgaagagc gacaaagatt 7260
ttgttatcgg ctttattgct caaagagaca tgggtggaag agatgaagggt tacgattgggt 7320
tgattatgac acccggtgtg ggtttagatg acaagggaga cgcattgggt caacagtata 7380
gaaccgtgga tgatgtggtc tctacaggat ctgacattat tattgttgga agaggactat 7440
ttgcaaaggg aagggatgct aaggtagagg gtgaacgtta cagaaaagca ggctgggaag 7500
catatttgag aagatgcggc cagcaaaact aaaaaactgt attataagta aatgcatgta 7560
tactaaactc acaaattaga gcttcaattt aattatatca gttattaccc gggaatctcg 7620
gtcgtaatga 7630

```

```

<210> 30
<211> 10015
<212> ADN
<213> Plasmid pCP50;

```

```

<400> 30
cttgaaataa ttaacaaaca aaggagttac agttagaat tgtaggagag atctcgtttt 60
tcgcgacaat ctggcgtttt tcttgctaata tccaggatta atccgttcat agtgtaaaac 120
cccgtttaca cattctgacg gaagatatag attggaagta ttgcattcac taagataagt 180
atggcaacac tggaacagac atgaattatc agaacgacga tttacgcac aaagaaatca 240
aagagttact tcctcctgtc gcattgctgg aaaaattccc cgctactgaa aatgccgcga 300
atacggttgc ccatgcccga aaagcgatcc ataagatcct gaaaggtaat gatgatcgcc 360
tgttggttgt gattggccca tgctcaattc atgatcctgt cgcggcaaaa gagtatgcca 420

```

ctcgcttgct ggcgctgcgt gaagagctga aagatgagct ggaaatcgta atgcgcgtct	480
at tt tt gaaaa gccgctacc acggtgggct ggaaagggct gattaacgat ccgcatatgg	540
ataatagctt ccagatcaac gacggtctgc gtatagcccc taaattgctg cttgatatta	600
acgacagcgg tctgccagcg gcaggtgagt ttctcgatat gatcacccca caatatctcg	660
ctgacctgat gagctggggc gcaattggcg cacgtaccac cgaatcgag gtgcaccgcg	720
aactggcatc agggctttct tgtccggctc gcttcaaaaa tggcaccgac ggtacgatta	780
aagtggctat cgatgccatt aatgccgcg gtgcgccga ctgcttcctg tccgtaacga	840
aatgggggca ttcggcgatt gtgaatacca gcggtaacgg cgattgccat atcattctgc	900
gcggcggtaa agagcctaac tacagcgcga agcacgttgc tgaagtgaaa gaagggtga	960
acaaagcagg cctgccagca caggtgatga tcgatttcag ccatgctaac tcgtccaaac	1020
aattcaaaaa gcagatggat gtttgtgctg acgtttgcc a gcagattgcc ggtggcgaaa	1080
aggccattat tggcgtgatg gtggaaagcc atctgggtga aggcaatcag agcctcgaga	1140
gcggggagcc gctggcctac ggtaagagca tcaccgatgc ctgcatcggc tgggaagata	1200
ccgatgctct gttacgtcaa ctggcgaatg cagtaaaagc gcgtcgcggg taaggtttaa	1260
ttgtcggatg cgcgctcaga gtggcgtatc cgatgaatca ccacaggcct gataagtcgc	1320
gcagcgtcgc atcaggcaat gtgctccatt gttagcaaca aaaaagccga ctcacttgca	1380
gtcggctttc tcattttaaa cgaatgacgt ttacttcgct ttaccctggt ttgcaaccgc	1440
cgctgctttc gctacatgaa tggctcttcgg tttccgtgtt tcgtaaagtc tggaaacgcg	1500
gaagtcagcg ccctgcacca ttatgttccg gatctgcac gcaggatgct gctggctacc	1560
ctgtggaaca cctacatctg tattaacgaa gcgctggcat tgaccctgag tgatttttct	1620
ctggccccgc cgcattcata ccgccagttg tttaccctca caacgttcca gtaaccgggc	1680
atgttcatca tcagtaacct gtatcgtgag catcctctct cgtttcatcg gtatcattac	1740
ccccatgaac agaaattccc ctttacacgg aggcatacag tgaccaaaca ggaaaaaac	1800
gcccttaaca tggcccgtt tatcagaagc cagacattaa cgcttctgga gaaactcaac	1860
gagctggacg cggatgaaca ggcagacatc tgtgaatcgc ttcacgacca cgctgatgag	1920
ctttaccgca gctgcctcgc gcgtttcgg gatgacggtg aaaacctctg acacatgcag	1980
ctccccgaga cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag	2040
ggcgcgtcag cgggtgttg cgggtgtcgg ggcgcagcca tgaccagtc acgtagcgat	2100
agcggagtgt atactggctt aactatgcgg catcagagca gattgtactg agagtgcacc	2160
atatgcggtg tgaaataccg cacagatgcg taaggagaaa ataccgcac aggcgctctt	2220
ccgcttctc gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag	2280

ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	2340
tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcggtg	ctggcggtttt	2400
tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	2460
gaaacccgac	aggactataa	agataccagg	cgtttcccc	tggaagctcc	ctcgtgcgct	2520
ctcctgttcc	gaccctgccg	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	2580
tggcgctttc	tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	2640
agctgggctg	tgtgcacgaa	cccccgttc	agcccgaccg	ctgcgcctta	tccggttaact	2700
atcgtcttga	gtccaacccg	gtaagacacg	acttatcgcc	actggcagca	gccactggta	2760
acaggattag	cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tggtggccta	2820
actacggcta	cactagaagg	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	2880
tcggaaaaag	agttggtagc	tcttgatccg	gcaaacaaac	caccgctggg	agcgggtggtt	2940
tttttgtttg	caagcagcag	attacgcgca	gaaaaaaagg	atctcaagaa	gaccccttga	3000
tcttttctac	ggggtctgac	gctcagtgga	acgaaaactc	acgttaaggg	attttggtca	3060
tgagattatc	aaaaaggatc	ttcacctaga	tccttttaaa	ttaaaaatga	agttttaaat	3120
caatctaaag	tatatatgag	taaacttggg	ctgacagtta	ccaatgctta	atcagtgagg	3180
cacctatctc	agcgatctgt	ctatttcggt	catccatagt	tgccctgactc	cccgtcgtgt	3240
agataactac	gatacgggag	ggcttaccat	ctggccccag	tgctgcaatg	ataccgcgag	3300
accacgctc	accggctcca	gatttatcag	caataaacca	gccagccgga	agggccgagc	3360
gcagaagtgg	tcctgcaact	ttatccgcct	ccatccagtc	tattaattgt	tgccgggaag	3420
ctagagtaag	tagttcgcca	gttaatagtt	tgcgcaacgt	tgttgccatt	gctgcaggca	3480
tcgtgggtgc	acgctcgtcg	tttgggtatg	cttcattcag	ctccggttcc	caacgatcaa	3540
ggcgagttac	atgatcccc	atgttgtgca	aaaaagcggg	tagctccttc	ggtcctccga	3600
tcgttgtcag	aagtaagttg	gccgcagtgt	tatcactcat	ggttatggca	gcactgcata	3660
attctcttac	tgatcatgcca	tccgtaagat	gcttttctgt	gactgggtgag	tactcaacca	3720
agtcattctg	agaatagtgt	atgcggcgac	cgagttgctc	ttgcccggcg	tcaacacggg	3780
ataataccgc	gccacatagc	agaactttaa	aagtgtcat	cattggaaaa	cgttcttcgg	3840
ggcgaaaact	ctcaaggatc	ttaccgctgt	tgagatccag	ttcgatgtaa	cccactcgtg	3900
caccaactg	atcttcagca	tcttttactt	tcaccagcgt	ttctgggtga	gcaaaaacag	3960
gaaggcaaaa	tgccgcaaaa	aagggaataa	gggcgacacg	gaaatgttga	atactcatac	4020
tcttcctttt	tcaatattat	tgaagcattt	atcagggtta	ttgtctcatg	agcggataca	4080

tatttgaatg tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag	4140
tgccacctga cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta	4200
tcacgaggcc ctttcgtctt caagaattct gaaccagtcc taaaacgagt aaataggacc	4260
ggcaattctt caagcaataa acaggaatac caattattaa aagataactt agtcagatcg	4320
tacaataaag ctttgaagaa aaatgcgcct tattcaatct ttgctataaa aaatggccca	4380
aaatctcaca ttggaagaca tttgatgacc tcatttcttt caatgaaggg cctaacggag	4440
ttgactaatg ttgtgggaaa ttggagcgat aagcgtgctt ctgccgtggc caggacaacg	4500
tatactcatc agataacagc aatacctgat cactacttcg cactagtttc tcggtactat	4560
gcatatgatc caatatcaaa ggaaatgata gcattgaagg atgagactaa tccaattgag	4620
gagtggcagc atatagaaca gctaaagggt agtgctgaag gaagcatacg ataccccgca	4680
tggaatggga taatatcaca ggaggtacta gactaccttt catcctacat aaatagacgc	4740
atataagtac gcatttaagc ataaacacgc actatgccgt tcttctcatg tatatatata	4800
tacaggcaac acgcagatat aggtgcgacg tgaacagtga gctgtatgtg cgcagctcgc	4860
gttgcattht cggaagcgct cgttttcgga aacgctttga agttcctatt ccgaagttcc	4920
tattctctag aaagtatagg aacttcagag cgcttttgaa aacccaaaagc gctctgaaga	4980
cgcactttca aaaaaccaa aacgcaccgg actgtaacga gctactaaaa tattgcgaat	5040
accgcttcca caaacattgc tcaaaagtat ctctttgcta tatatctctg tgctatatcc	5100
ctatataacc taccatcca ctttcgctc cttgaacttg catctaaact cgacctctac	5160
atthttttatg tttatctcta gtattactct ttagacaaaa aaattgtagt aagaactatt	5220
catagagtga atcgaaaaca atacgaaaat gtaaacattt cctatacgta gtatatagag	5280
acaaaataga agaaaccgtt cataatthtc tgaccaatga agaatcatca acgctatcac	5340
tttctgttca caaagtatgc gcaatccaca tcggtataga atataatcgg ggatgccttt	5400
atcttgaaaa aatgcacccg cagcttcgct agtaatcagt aaacgcggga agtggagtca	5460
ggctthtttt atggaagaga aaatagacac caaagtagcc ttcttctaac cttaacggac	5520
ctacagtgca aaaagttatc aagagactgc attatagagc gcacaaaagga gaaaaaaagt	5580
aatctaagat gctttgttag aaaaatagcg ctctcgggat gcattthttgt agaacaaaaa	5640
agaagtatag attctthgtt ggtaaaatag cgctctcgcg ttgcatttct gttctgtaaa	5700
aatgcagctc agattctthg tttgaaaaat tagcgcctct gcgttgcat tttgtthttac	5760
aaaaatgaag cacagattct tcgttggtaa aatagcgctt tcgcgttgca tttctgttct	5820
gtaaaaatgc agctcagatt ctttgthttga aaaattagcg ctctcgcgtt gcattthttgt	5880
tctacaaaat gaagcacaga tgcttcgtta acaagatat gctattgaag tgcaagatgg	5940

aaacgcagaa aatgaaccgg ggatgcgacg tgcaagatta cctatgcaat agatgcaata	6000
gtttctccag gaaccgaaat acatacattg tcttccgtaa agcgctagac tatatattat	6060
tatacagggtt caaatatact atctgtttca gggaaaactc ccaggttcgg atgttcaaaa	6120
ttcaatgatg ggtaacaagt acgatcgtaa atctgtaaaa cagtttgtcg gatattaggc	6180
tgtatctcct caaagcgtat tcgaatatca ttgagaagct gcagcgtcac atcggataat	6240
aatgatggca gccattgtag aagtgccttt tgcatcttcta gtctctttct cggcttagct	6300
agttttacta catcgcgaa agatgaatctt agatcacact gcctttgctg agctggatca	6360
atagagtaac aaaagagtgg taaggcctcg ttaaaggaca aggacctgag cggaagtgt	6420
tcgtacagta gacggagtat actagtatag tctatagtc gtggaattct catgtttgac	6480
agcttatcat cgataagctt ttcaattcaa ttcattcttt tttttttatt cttttttttg	6540
atttcggttt ctttgaaatt tttttgattc ggtaatctcc gaacagaagg aagaacgaag	6600
gaaggagcac agacttagat tggatatatat acgcatatgt agtggtgaag aaacatgaaa	6660
ttgcccagta ttcttaaccc aactgcacag aacaaaaacc tgcaggaaac gaagataaat	6720
catgtcgaaa gctacatata aggaacgtgc tgctactcat cctagtcctg ttgctgccaa	6780
gctattttaat atcatgcacg aaaagcaaac aaacttgtgt gcttcattgg atgttcgtac	6840
caccaaggaa ttactggagt tagttgaagc attaggtccc aaaatttgtt tactaaaaac	6900
acatgtggat atcttgactg atttttccat ggagggcaca gttaagccgc taaaggcatt	6960
atccgccaa ag tacaattttt tactcttcga agacagaaaa tttgctgaca ttggtaatac	7020
agtcaaattg cagtactctg cgggtgtata cagaatagca gaatgggcag acattacgaa	7080
tgcacacggt gtggtgggcc caggtattgt tagcggtttg aagcaggcgg cagaagaagt	7140
aacaaaggaa cctagaggcc ttttgatgtt agcagaattg tcatgcaagg gctccctatc	7200
tactggagaa tatactaagg gtactgttga cattgcgaag agcgacaaag attttgttat	7260
cggcttttatt gctcaaagag acatgggtgg aagagatgaa ggttacgatt ggttgattat	7320
gacacccggt gtgggttttag atgacaaggg agacgcattg ggtcaacagt atagaaccgt	7380
ggatgatgtg gtctctacag gatctgacat tattattgtt ggaagaggac tatttgcaaa	7440
gggaagggat gctaaggtag aggggtgaacg ttacagaaaa gcaggctggg aagcatattt	7500
gagaagatgc ggccagcaaa actaaaaaac tgtattataa gtaaattgcat gtatactaaa	7560
ctcaciaaatt agagcttcaa tttaattata tcagttatta cccgggaatc tcggtcgtaa	7620
tgaaaggaaa agcgcaacgg acgggcgagt agattgcgca acatgcgagc atgatccaga	7680
gattttctgaa gcagcaaaag gatgttccat gtacatgacg cgcggcttgc ggtaaattgt	7740

tggcaaattt	tccggcgtag	cccaaaacgc	gctgtcgtca	agtcgttaag	ggcgtgccct	7800
tcatcatccg	atctggagtc	aaaatgtcct	cacgtaaaga	gcttgccaat	gctattcgtg	7860
cgctgagcat	ggacgcagta	cagaaagcca	aatccggtca	cccgggtgcc	cctatgggta	7920
tggctgacat	tgccgaagtc	ctgtggcggtg	atttcttgaa	acacaacccg	cagaatccgt	7980
cctgggctga	ccgtgaccgc	ttcgtgctgt	ccaacggcca	cggctccatg	ctgatctaca	8040
gcctgctgca	cctcaccggt	tacgatctgc	cgatggaaga	actgaaaaac	ttccgtcagc	8100
tgcactctaa	aactccgggc	cacccggaag	taggttatac	cgctgggtgtg	gaaaccacca	8160
ccggtccgct	gggtcaggggt	attgccaacg	cagtcggtat	ggcgattgca	gaaaaaacgc	8220
tggcggcgca	gtttaaccgt	ccaggtcacg	acattgtcga	ccactacacc	tacgccttca	8280
tgggcgacgg	ctgcatgatg	gaaggcatct	cccacgaagt	ttgctctctg	gcggggtacgc	8340
tgaagctggg	taaactgatt	gcgttctacg	atgacaacgg	tatctcaatc	gatggtcacg	8400
ttgaaggctg	gttcactgac	gacaccgcaa	tgcgtttcga	agcttacggc	tggcacgtta	8460
ttcgcgacat	cgacgggtcat	gacgcggcat	ccatcaaacg	cgcagtagaa	gaagcgcgcg	8520
cagtgactga	caaaccgtcc	ctgctgatgt	gcaaaaccat	catcggtttc	ggttccccga	8580
acaaagccgg	taccacgac	tcccacgggtg	cgccgctggg	cgacgctgaa	attgccctga	8640
cccgcgaaca	gctgggctgg	aaatacgcgc	cgttcgaaat	cccgtctgaa	atctatgctc	8700
agtgggatgc	gaaagaagca	ggccaggcga	aagaatctgc	atggaatgag	aagtttgccg	8760
cttacgcgaa	agcttatccg	caggaagcgg	ctgaatttac	ccgccgtatg	aaaggcgaaa	8820
tgccgtctga	cttcgacgcc	aaagcgaaag	agtttatcgc	taaactgcag	gctaataccgg	8880
cgaaaatcgc	cagccgtaaa	gcgtcgcaga	atgctatcga	agcgttcggc	ccgctgttgc	8940
ctgaattcct	cggcggctct	gctgacctgg	caccgtctaa	cctgacctg	tggctctggtt	9000
ctaaagcaat	caacgaagat	gctgcaggta	actacatcca	ctacggtggt	cgcgagtctg	9060
gtatgaccgc	gattgctaac	ggtatctccc	tgcacgggtg	tttcctgccg	tacacctcca	9120
ccttctgat	gttcgtggaa	tacgcacgta	acgccgtacg	tatggctgcg	ctgatgaaac	9180
agcgtcaggt	gatggtttac	accacgact	ccatcggtct	gggcgaagat	ggcccgactc	9240
accagccggt	tgagcaggtc	gcttctctgc	gcgtgacccc	gaacatgtct	acatggcgtc	9300
cgtgtgacca	ggttgaatcc	gcggtcgcgt	ggaaatacgg	cgttgagcgt	caggacggcc	9360
cgactgcgct	tatcctctcc	cgtcagaacc	tggcgcagca	ggaacgaact	gaagagcaac	9420
tggcaaacat	cgcgcgcgggt	ggttatgtgc	tgaaagactg	cgccggtcag	ccggaactga	9480
ttttcatcgc	taccggttca	gaagttgaac	tggctgttgc	tgcttacgaa	aaactgactg	9540
ccgaaggcgt	gaaagcgcgc	gtggtgtcca	tgccgtctac	cgacgcattt	gacaagcagg	9600

atgctgctta ccgatgaatcc gtactgccga aagcgggttac tgcacgcggtt gctgtagaag	9660
cgggtattgc tgactactgg tacaagtatg ttggcctgaa cgggtgctatc gtcggtatga	9720
ccaccttcgg tgaatctgct ccggcagagc tgctgtttga agagtccggc ttcactgttg	9780
ataacgttgt tgcgaaagca aaagaactgc tgtaattagc atttcgggta aaaaggctgc	9840
ttcggcgacc ttttttatta ccttgatatg tccgtttgcg gacaagcaat agataaagcg	9900
tggtgtagat cacaaatatt tatatgcaat aaatatcaat tatgtaatat gcatcacgat	9960
atgcgtattg acatttggtg ttataactat aactcaatgt tatataagaa attaa	10015

<210> 31

<211> 9065

<212> ADN

<213> Plasmid pCP54;

<400> 31

acatgaatgg tcttcggttt ccgtgtttcg taaagtctgg aaacgcggaa gtcagcgcgc	60
tgcaccatta tgttccggat ctgcatcgca ggatgctgct ggctaccctg tggaacacct	120
acatctgtat taacgaagcg ctggcattga ccctgagtga tttttctctg gtcccgccgc	180
atccataccg ccagttgttt accctcaca cgttccagta accgggcatg ttcacatca	240
gtaaccgta tcgtgagcat cctctctcgt ttcacggta tcattacccc catgaacaga	300
aattccccct tacacggagg catcaagtga ccaaacagga aaaaaccgcc cttaacatgg	360
cccgttttat cagaagccag acattaacgc ttctggagaa actcaacgag ctggacgcgg	420
atgaacaggc agacatctgt gaatcgcttc acgaccacgc tgatgagctt taccgcagct	480
gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc ccggagacgg	540
tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg	600
gtgttgggcg gtgtcggggc gcagccatga ccagtcacg tagcgatagc ggagtgtata	660
ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata tgcggtgtga	720
aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcttccg cttcctcgt	780
cactgactcg ctgcgctcgg tcgttcgggt gcggcgagcg gtatcagctc actcaaaggc	840
ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg	900
ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg	960
ccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg	1020
actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac	1080
cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca	1140
tagctcacgc tgtaggtatc tcagttcggg ttaggtcgtt cgctccaagc tgggctgtgt	1200

gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc	1260
caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag	1320
agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact acggctacac	1380
tagaaggaca gtattttgta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt	1440
tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa	1500
gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg	1560
gtctgacgct cagtggaaacg aaaactcacg ttaagggatt ttggtcatga gattatcaaa	1620
aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaataca tctaaagtat	1680
atatgagtaa acttgggtctg agtggcggtt ttcattggctt gttatgactg tttttttggg	1740
gtacagtcta tgccctcggc atccaagcag caagcgcgtt acgccgtggg tcgatgtttg	1800
atgttatgga gcagcaacga tgttacgcag cagggcagtc gccctaaaac aaagttaaac	1860
atcatgaggg aagcgggtgat cgccgaagta tcgactcaac tatcagaggt agttggcgtc	1920
atcgagcgcc atctcgaacc gacgttgctg gccgtacatt tgtacggctc cgcagtggat	1980
ggcggcctga agccacacag tgatattgat ttgctgggta cgggtgaccgt aaggcttgat	2040
gaaacaacgc ggcgagcttt gatcaacgac ctttttgaaa cttcggcttc ccctggagag	2100
agcgagattc tccgcgctgt agaagtcacc attgttgtgc acgacgacat cattccgtgg	2160
cgttatccag ctaagcgcga actgcaatth ggagaatggc agcgcaatga cattcttgca	2220
ggtatcttcg agccagccac gatcgacatt gatctggcta tcttgctgac aaaagcaaga	2280
gaacatagcg ttgccttggt aggtccagcg gcggaggaac tctttgatcc ggttcctgaa	2340
caggatctat ttgaggcgct aaatgaaacc ttaacgctat ggaactcgcc gcccgactgg	2400
gctggcgatg agcgaaatgt agtgcttacg ttgtcccga tttggtacag cgcagtaacc	2460
ggcaaaatcg cgccgaagga tgtcgctgcc gactgggcaa tggagcgcct gccggcccag	2520
tatcagcccg tcataactga agctagacag gcttatcttg gacaagaaga agatcgcttg	2580
gcctcgcgcg cagatcagtt ggaagaatth gtccactacg tgaaaggcga gatcaccaag	2640
gtagtcggca aataatgtct aacaattcgt tcaagccgac gccgcttcgc ggcgcggctt	2700
aactcaagcg ttagatgcac taagcacata attgctcaca gccaaactat cagaattctg	2760
aaccagtcct aaaacgagta aataggaccg gcaattcttc aagcaataaa caggaatacc	2820
aattattaaa agataactta gtcagatcgt acaataaagc tttgaagaaa aatgcgcctt	2880
attcaatctt tgctataaaa aatggcccaa aatctcacat tggaagacat ttgatgacct	2940
catttctttc aatgaagggc ctaacggagt tgactaatgt tgtgggaaat tggagcgata	3000

agcgtgcttc	tgccgtggcc	aggacaacgt	atactcatca	gataacagca	atacctgatc	3060
actacttcgc	actagtttct	cggtactatg	catatgatcc	aatatcaaag	gaaatgatag	3120
cattgaagga	tgagactaat	ccaattgagg	agtggcagca	tatagaacag	ctaaagggta	3180
gtgctgaagg	aagcatacga	taccccgcat	ggaatgggat	aatatcacag	gaggtactag	3240
actacctttc	atcctacata	aatagacgca	tataagtacg	catttaagca	taaacacgca	3300
ctatgccgtt	cttctcatgt	atatatatat	acaggcaaca	cgcagatata	ggtgcgacgt	3360
gaacagtgag	ctgtatgtgc	gcagctcgcg	ttgcattttc	ggaagcgctc	gttttcggaa	3420
acgctttgaa	gttcctattc	cgaagtccct	attctctaga	aagtatagga	acttcagagc	3480
gcttttgaaa	accaaagcg	ctctgaagac	gcactttcaa	aaaaccaaaa	acgcaccgga	3540
ctgtaacgag	ctactaaaat	attgcgaata	ccgcttcac	aaacattgct	caaaagtatc	3600
tctttgctat	atatctctgt	gctatatccc	tatataacct	acccatccac	ctttcgctcc	3660
ttgaacttgc	atctaaactc	gacctctaca	ttttttatgt	ttatctctag	tattactctt	3720
tagacaaaaa	aattgtagta	agaactattc	atagagtga	tcgaaaacaa	tacgaaaatg	3780
taaacatttc	ctatacgtag	tatatagaga	caaaatagaa	gaaaccgttc	ataattttct	3840
gaccaatgaa	gaatcatcaa	cgctatcact	ttctgttcac	aaagtatgcg	caatccacat	3900
cggtatagaa	tataatcggg	gatgccttta	tcttgaaaaa	atgcacccgc	agcttcgcta	3960
gtaatcagta	aacgcgggaa	gtggagtcag	gcttttttta	tggaagagaa	aatagacacc	4020
aaagtagcct	tcttctaacc	ttaacggacc	tacagtgcaa	aaagtatatca	agagactgca	4080
ttatagagcg	cacaaaggag	aaaaaaagta	atctaagatg	ctttgttaga	aaaatagcgc	4140
tctcgggatg	catttttgta	gaacaaaaaa	gaagtataga	ttctttgttg	gtaaaatagc	4200
gctctcgcg	tgcatctctg	ttctgtaaaa	atgcagctca	gattctttgt	ttgaaaaatt	4260
agcgctctcg	cgttgcattt	ttgtttttaca	aaaatgaagc	acagattctt	cgttggtaaa	4320
atagcgcttt	cgcgttgcat	ttctgttctg	taaaaatgca	gctcagattc	ttgttttgaa	4380
aaattagcgc	tctcgcgttg	catttttggt	ctacaaaatg	aagcacagat	gcttcgttaa	4440
caaagatatg	ctattgaagt	gcaagatgga	aacgcagaaa	atgaaccggg	gatgcgacgt	4500
gcaagattac	ctatgcaata	gatgcaatag	tttctccagg	aaccgaaata	catacattgt	4560
cttccgtaaa	gcgctagact	atatattatt	atacaggttc	aaatatacta	tctgtttcag	4620
ggaaaactcc	caggttcgga	tgttcaaaat	tcaatgatgg	gtaacaagta	cgatcgtaaa	4680
tctgtaaaac	agtttgctcg	atattaggct	gtatctcctc	aaagcgattt	cgaatatcat	4740
tgagaagctg	cagcgtcaca	tcggataata	atgatggcag	ccattgtaga	agtgcctttt	4800
gcatttctag	tctctttctc	ggtctagcta	gttttactac	atcgcgaa	tagaatctta	4860

gatcacactg cctttgctga gctggatcaa tagagtaaca aaagagtggg aaggcctcgt	4920
taaaggacaa ggacctgagc ggaagtgtat cgtacagtag acggagtata ctagtatagt	4980
ctatagtccg tggaaattctc atgtttgaca gcttatcatc gataagcttt tcaattcaat	5040
tcatcatctt ttttttattc ttttttttga tttcggtttc tttgaaattt ttttgattcg	5100
gtaatctccg aacagaagga agaacgaagg aaggagcaca gacttagatt ggtatatata	5160
cgcataatgta gtgttggaaga aacatgaaat tgcccagtat tcttaacca actgcacaga	5220
acaaaaacct gcaggaaacg aagataaatc atgtcgaaag ctacatataa ggaacgtgct	5280
gctactcatc ctagtcctgt tgctgccaag ctattttaata tcatgcacga aaagcaaaca	5340
aacttgtgtg cttcattgga tgttcgtacc accaaggaat tactggagtt agttgaagca	5400
ttaggtccca aaatttgttt actaaaaaca catgtggata tcttgactga tttttccatg	5460
gagggcacag ttaagccgct aaaggcatta tccgccaagt acaatttttt actcttcgaa	5520
gacagaaaat ttgctgacat tggtaataca gtcaaattgc agtactctgc ggggtgtatac	5580
agaatagcag aatgggcaga cattacgaat gcacacgggtg tgggtgggcc aggtattggt	5640
agcggtttga agcaggcggc agaagaagta acaaaggaac ctagaggcct tttgatgtta	5700
gcagaattgt catgcaaggg ctccctatct actggagaat atactaaggg tactgttgac	5760
attgcgaaga gcgacaaaga ttttgttatc ggctttattg ctcaaagaga catgggtgga	5820
agagatgaag gttacgattg gttgattatg acaccgggtg tgggtttaga tgacaaggga	5880
gacgcattgg gtcaacagta tagaaccgtg gatgatgtgg tctctacagg atctgacatt	5940
attattgttg gaagaggact atttgcaaag ggaagggatg ctaaggtaga ggggtgaacgt	6000
tacagaaaag caggctggga agcatatttg agaagatgcg gccagcaaaa ctaaaaaact	6060
gtattataag taaatgcatg tataactaaac tcacaaatta gagcttcaat ttaattatat	6120
cagttattac ccgggaatct cggtcgtaat gacttgaaat aattaacaaa caaaggagtt	6180
acagttagaa attgtaggag agatctcggt tttcgcgaca atctggcggt tttcttgcta	6240
attccaggat taatccgttc atagtgtaaa accccgttta cacattctga cggaagatat	6300
agattggaag tattgcattc actaagataa gtatggcaac actggaacag acatgaatta	6360
tcagaacgac gattttacgca tcaaagaaat caaagagtta cttcctcctg tcgcattgct	6420
ggaaaaattc cccgctactg aaaatgccgc gaatacgggt gcccatgccc gaaaagcgat	6480
ccataagatc ctgaaaggta atgatgatcg cctgttggtt gtgattggcc catgctcaat	6540
tcatgatcct gtcgcggcaa aagagtatgc cactcgcttg ctggcgctgc gtgaagagct	6600
gaaagatgag ctggaaatcg taatgcgcgt ctattttgaa aagccgcgta ccacggtggtg	6660

ctggaaaggg	ctgattaacg	atccgcata	ggataatagc	ttccagatca	acgacgggtct	6720
gcgtatagcc	cgtaaattgc	tgcttgatat	taacgacagc	ggctctgccag	cggcaggtga	6780
gtttctcgat	atgatcacc	cacaatatct	cgctgacctg	atgagctggg	gcgcaattgg	6840
cgcacgtacc	accgaatcgc	aggtgcaccg	cgaactggca	tcagggcttt	cttgtccggg	6900
cggcttcaaa	aatggcaccg	acggtaacgat	taaagtggct	atcgatgcc	ttaatgccgc	6960
cggcgcgcg	cactgcttcc	tgtccgtaac	gaaatggggg	cattcggcga	ttgtgaatac	7020
cagcggtaac	ggcgattgcc	atatcattct	gcgcggcggt	aaagagccta	actacagcgc	7080
gaagcacgtt	gctgaagtga	aagaagggct	gaacaaagca	ggcctgccag	cacaggtgat	7140
gatcgatttc	agccatgcta	actcgtccaa	acaattcaaa	aagcagatgg	atgtttgtgc	7200
tgacgtttgc	cagcagattg	ccggtggcga	aaaggccatt	attggcgtga	tggtggaaag	7260
ccatctggtg	gaaggcaatc	agagcctcga	gagcggggag	ccgctggcct	acggtaagag	7320
catcacgat	gcctgcatcg	gctgggaaga	taccgatgct	ctgttacgtc	aactggcgaa	7380
tgcagtaaaa	gcgcgtcgcg	ggtaagggtt	aattgtcgga	tcgcgccgtca	gagtggcgta	7440
tccgatgaat	caccacaggc	ctgataagtc	gcgcagcgtc	gcacagggca	atgtgctcca	7500
ttgttagcaa	caaaaaagcc	gactcacttg	cagtcggctt	tctcatttta	aacgaatgac	7560
gtttacttcg	ctttaccctg	gtttgcaacc	gccgctgctt	tcgctctcga	ggctattgac	7620
gacagctatg	gttcactgtc	caccaaccaa	aactgtgctc	agtaccgcca	atatttctcc	7680
cttgaggggt	acaaagaggt	gtccctagaa	gagatccacg	ctgtgtaaaa	attttacaaa	7740
aaggtattga	ctttccctac	aggggtgtgta	ataatttaat	tacaggcggg	ggcaaccccg	7800
cctgttctag	aggaggagga	atcgccatgg	agaggattgt	cgttactctc	ggggaacgta	7860
gttacccaat	taccatcgca	tctgggtttgt	ttaatgaacc	agcttcattc	ttaccgctga	7920
aatcgggcga	gcaggatcatg	ttggtcacca	acgaaaccct	ggctcctctg	tatctcgata	7980
aggtccgcgg	cgtacttgaa	caggcgggtg	ttaacgtcga	tagcggtatc	ctccctgacg	8040
gcgagcagta	taaaagcctg	gctgtactcg	ataccgtctt	tacggcggtg	ttacaaaagc	8100
cgcattggtc	cgatactacg	ctgggtggcg	ttggcggcgg	cgtagtgggc	gatctgaccg	8160
gcttcgcggc	ggcgagttat	cagcgcgggtg	ttcgtttcat	tcaagtcccc	acgacgttac	8220
tgctgcaggt	cgattcctcc	gttggcggca	aaactgcggg	caaccatccc	ctcggtaaaa	8280
acatgattgg	cgcgttctac	cagcctgctt	cagtgggtgg	ggatctcgac	tgtctgaaaa	8340
cgttcccccc	gcgtgagtta	gcgtcggggc	tggcagaagt	catcaaatac	ggcattattc	8400
ttgacgggtg	gtttttcaac	tggctggaag	agaatctgga	tcggttggtg	cgtctggacg	8460
gtccggcaat	ggcgtactgt	attcgccgtt	gttgtgaact	gaaggcagaa	gttgtcgccg	8520

```

ccgacgagcg cgaaaccggg ttacgtgctt tactgaatct gggacacacc tttgggtcatg      8580
ccattgaagc tgaaatgggg tatggcaatt ggttacatgg tgaagcgggc gctgcgggta      8640
tgggtgatggc ggcgcggacg tcggaacgtc tcgggcagtt tagttctgcc gaaacgcagc      8700
gtattataac cctgctcacg cgggctgggt taccggtcaa tgggccgcgc gaaatgtccg      8760
cgcaggcgta tttaccgcat atgctgcgtg acaagaaagt ccttgcgggg gagatgcgct      8820
taattcttcc gttggcaatt ggtaagagtg aagttcgcag cggcgtttcg cacgagcttg      8880
ttcttaacgc cattgccgat tgtcaatcag cgtaatcatc gttcatgcct gatgccgcta      8940
tgtaggccgg ataaggcggt cacgccgat ccggcaaccg atgcctgatg cgacgcggtc      9000
gcgtcttata aggctacag gtcgatgccg atatgtacat cgtattcggc aattaatata      9060
tagca                                          9065

```

```

<210> 32
<211> 11475
<212> ADN
<213> Plasmid pCP55;

```

```

<400> 32
acatgaatgg tcttcggttt ccgtgtttcg taaagtctgg aaacgcggaa gtcagcgcgc      60
tgcaccatta tgttccggat ctgcatcgca ggatgctgct ggctaccctg tggaacacct      120
acatctgtat taacgaagcg ctggcattga ccctgagtga tttttctctg gtcccgccgc      180
atccataaccg ccagttgttt accctcacia cgttccagta accgggcatg ttcacatca      240
gtaaccgta tcgtgagcat cctctctcgt ttcacggta tcattacccc catgaacaga      300
aattccccct tacacggagg catcaagtga ccaaacagga aaaaaccgcc cttaacatgg      360
cccgctttat cagaagccag acattaacgc ttctggagaa actcaacgag ctggacgcgg      420
atgaacaggc agacatctgt gaatcgcttc acgaccacgc tgatgagctt taccgcagct      480
gcctcgcgcg tttcggtgat gacggtgaaa acctctgaca catgcagctc ccggagacgg      540
tcacagcttg tctgtaagcg gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg      600
gtgttggcgg gtgtcggggc gcagccatga ccagtcacg tagcgatagc ggagtgtata      660
ctggcttaac tatgcggcat cagagcagat tgtactgaga gtgcaccata tgcgggtgtga      720
aataccgcac agatgcgtaa ggagaaaata ccgcatcagg cgctcttccg cttcctcgt      780
cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc      840
ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg      900
ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg      960
ccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg      1020

```

actataaaga	taccaggcgt	ttccccctgg	aagctccctc	gtgcgctctc	ctgttccgac	1080
cctgccgctt	accggatacc	tgtccgcctt	tctcccttcg	ggaagcgtgg	cgctttctca	1140
tagctcacgc	tgtaggtatc	tcagttcggg	gtaggtcggt	cgctccaagc	tggtctgtgt	1200
gcacgaaccc	cccgttcagc	ccgaccgctg	cgccttatcc	ggtaactatc	gtcttgagtc	1260
caacccggta	agacacgact	tatcgccact	ggcagcagcc	actggtaaca	ggattagcag	1320
agcgaggtat	gtaggcgggtg	ctacagagtt	cttgaagtgg	tggcctaact	acggctacac	1380
tagaaggaca	gtatttggtg	tctgcgctct	gctgaagcca	gttaccttcg	gaaaaagagt	1440
tggtagctct	tgatccggca	aacaaaccac	cgctggtagc	ggtgggtttt	ttgtttgcaa	1500
gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	cctttgatct	tttctacggg	1560
gtctgacgct	cagtggaaacg	aaaactcacg	ttaagggatt	ttgggtcatga	gattatcaaa	1620
aaggatcttc	acctagatcc	ttttaaatta	aaaatgaagt	tttaaataca	tctaaagtat	1680
atatgagtaa	acttgggtctg	acagttacca	atgcttaatc	agtgaggcac	ctatctcagc	1740
gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	gtcgtgtaga	taactacgat	1800
acgggagggc	ttaccatctg	gccccagtgc	tgcaatgata	ccgcgagacc	cacgctcacc	1860
ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	gccgagcgca	gaagtgggtcc	1920
tgcaacttta	tccgcctcca	tccagtcctat	taattgttgc	cgggaagcta	gagtaagtag	1980
ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	gcaggcatcg	tggtgtcacg	2040
ctcgtcgttt	ggtatggctt	cattcagctc	cggttcccaa	cgatcaaggc	gagttacatg	2100
atcccccatg	ttgtgcaaaa	aagcgggttag	ctccttcggg	cctccgatcg	ttgtcagaag	2160
taagttggcc	gcagtgttat	cactcatggg	tatggcagca	ctgcataatt	ctcttactgt	2220
catgccatcc	gtaagatgct	tttctgtgac	tggtgagtac	tcaaccaagt	cattctgaga	2280
atagtgtatg	cggcgaccga	gttgcctctg	cccggcgctc	acacgggata	ataccgcgcc	2340
acatagcaga	actttaaaaag	tgctcatcat	tggaaaacgt	tcttcggggc	gaaaactctc	2400
aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	actcgtgcac	ccaactgatc	2460
ttcagcatct	tttactttca	ccagcgtttc	tggttgagca	aaaacaggaa	ggcaaaatgc	2520
cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	ctcatactct	tcctttttca	2580
atattattga	agcatttatc	agggttattg	tctcatgagc	ggatacatat	ttgaatgtat	2640
ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	cgaaaagtgc	cacctgacgt	2700
ctaagaaacc	attattatca	tgacattaac	ctataaaaat	aggcgtatca	cgaggccctt	2760
tcgtcttcaa	gaattctgaa	ccagtcctaa	aacgagtaaa	taggaccggc	aattcttcaa	2820

gcaataaaca ggaataccaa ttattaaaag ataacttagt cagatcgtac aataaagctt	2880
tgaagaaaaa tgcgccttat tcaatctttg ctataaaaaa tggcccaaaa tctcacattg	2940
gaagacattt gatgacctca tttcttttcaa tgaagggcct aacggagttg actaatgttg	3000
tgggaaattg gagcgataag cgtgcttctg ccgtggccag gacaacgtat actcatcaga	3060
taacagcaat acctgatcac tacttcgcac tagtttctcg gtactatgca tatgatccaa	3120
tatcaaagga aatgatagca ttgaaggatg agactaatcc aattgaggag tggcagcata	3180
tagaacagct aaagggtagt gctgaaggaa gcatacgata ccccgcatgg aatgggataa	3240
tatcacagga ggtactagac tacctttcat cctacataaa tagacgcata taagtacgca	3300
tttaagcata aacacgcact atgccgttct tctcatgtat atatatatac aggcaacacg	3360
cagatatagg tgcgacgtga acagtgaagt gtatgtgcgc agctcgcgtt gcattttcgg	3420
aagcgctcgt tttcggaaac gctttgaagt tcctattccg aagttcctat tctctagaaa	3480
gtataggaac ttcagagcgc ttttgaaaac caaaagcgct ctgaagacgc actttcaaaa	3540
aaccaaaaac gcaccggact gtaacgagct actaaaatat tgcgaatacc gcttcacaa	3600
acattgctca aaagtatctc tttgctatat atctctgtgc tatatcccta tataacctac	3660
ccatccacct ttcgctcctt gaacttgcac ctaaacctga cctctacatt ttttatgttt	3720
atctctagta ttactcttta gacaaaaaaa ttgtagtaag aactattcat agagtgaatc	3780
gaaaacaata cgaaaatgta aacatttcct atacgtagta tatagagaca aaatagaaga	3840
aaccgttcat aattttctga ccaatgaaga atcatcaacg ctatcacttt ctgttcacaa	3900
agtatgcgca atccacatcg gtatagaata taatcgggga tgcctttatc ttgaaaaaat	3960
gcaccgcgag cttcgctagt aatcagtaaa cgcgggaagt ggagtcaggc tttttttatg	4020
gaagagaaaa tagacaccaa agtagccttc ttctaacctt aacggaccta cagtgcaaaa	4080
agttatcaag agactgcatt atagagcgca caaaggagaa aaaaagtaat ctaagatgct	4140
ttgttagaaa aatagcgctc tcgggatgca tttttgtaga acaaaaaaga agtatagatt	4200
ctttgttggt aaaatagcgc tctcgcgttg catttctgtt ctgtaaaaaat gcagctcaga	4260
ttctttgttt gaaaaattag cgctctcgcg ttgcattttt gttttacaaa aatgaagcac	4320
agattcttcg ttggtaaaaat agcgcttttcg cgttgcattt ctgttctgta aaaatgcagc	4380
tcagattctt tgtttgaaaa attagcgctc tcgcgttgca tttttgttct acaaaatgaa	4440
gcacagatgc ttcgttaaca aagatatgct attgaagtgc aagatggaaa cgcagaaaaat	4500
gaaccgggga tgcgacgtgc aagattacct atgcaataga tgcaatagtt tctccaggaa	4560
ccgaaataca tacattgtct tccgtaaagc gctagactat atattattat acaggttcaa	4620
atatactatc tgtttcaggg aaaactccca gggtcggatg ttcaaaattc aatgatgggt	4680

aacaagtacg atcgtaaadc tgtaaaacag tttgtcggat attaggctgt atctcctcaa	4740
agcgtattcg aatatcattg agaagctgca gcgtcacatc ggataataat gatggcagcc	4800
attgtagaag tgccttttgc atttctagtc tctttctcgg tctagctagt tttactacat	4860
cgcgaaagata gaatcctaga tcacactgcc tttgctgagc tggatcaata gagtaacaaa	4920
agagtggtaa ggcctcgtaa aaggacaagg acctgagcgg aagtgtatcg tacagtagac	4980
ggagtatact agtatagtct atagtccgtg gaattctcat gtttgacagc ttatcatcga	5040
taagcttttc aattcaattc atcatttttt ttttattctt ttttttgatt tcggtttctt	5100
tgaaattttt ttgattcggc aatctccgaa cagaaggaag aacgaaggaa ggagcacaga	5160
cttagattgg tatatatagc catatgtagt gttgaagaaa catgaaattg cccagtattc	5220
ttaacccaac tgcacagaac aaaaacctgc aggaaacgaa gataaatcat gtcgaaagct	5280
acataaagg aacgtgctgc tactcatcct agtcctgttg ctgccaagct atttaatatc	5340
atgcacgaaa agcaaacaaa cttgtgtgct tcattggatg ttcgtaccac caaggaatta	5400
ctggagttag ttgaagcatt aggtcccaaa atttgtttac taaaaacaca tgtggatatc	5460
ttgactgatt tttccatgga gggcacagtt aagccgctaa aggcattatc cgccaagtac	5520
aattttttac tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag	5580
tactctgcgg gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacgggtgtg	5640
gtgggcccag gtattgttag cggtttgaag caggcggcag aagaagtaac aaaggaacct	5700
agaggccttt tgatgttagc agaattgtca tgcaagggtc ccctatctac tggagaatat	5760
actaagggtg ctgttgacat tgcgaagagc gacaaagatt ttgttatcgg ctttattgct	5820
caaagagaca tgggtggaag agatgaaggc tacgattggc tgattatgac acccgggtgtg	5880
ggtttagatg acaagggaga cgcattgggt caacagtata gaaccgtgga tgatgtggtc	5940
tctacaggat ctgacattat tattgttgga agaggactat ttgcaaaggg aagggatgct	6000
aaggtagagg gtgaacgtta cagaaaagca ggctgggaag catatttgag aagatgcggc	6060
cagcaaaaact aaaaaactgt attataagta aatgcatgta tactaaactc acaaattaga	6120
gcttcaattt aattatatca gttattaccg gggaatctcg gtcgtaatga aaggaaaagc	6180
gcaacggacg ggcgagtaga ttgcgcaaca tgcgagcatg atccagagat ttctgaagca	6240
gcaaaaggat gttccatgta catgacgcgc ggcttgcggt aaattgttgg caaattttcc	6300
ggcgtagccc aaaacgcgct gtcgtcaagt cgtaagggc gtgcccttca tcatccgatc	6360
tggagtcaaa atgtcctcac gtaaagagct tgccaatgct attcgtgcgc tgagcatgga	6420
cgcagtacag aaagccaaat ccggtcaccc ggggtgccct atgggtatgg ctgacattgc	6480

cgaagtcctg	tggcgtgatt	tcctgaaaca	caaccgcgag	aatccgtcct	gggctgaccg	6540
tgaccgcttc	gtgctgtcca	acggccacgg	ctccatgctg	atctacagcc	tgctgcacct	6600
caccgggttac	gatctgccga	tggaagaact	gaaaaacttc	cgtcagctgc	actctaaaac	6660
tccggggccac	ccggaagtag	gttataccgc	tgggtgtggaa	accaccaccg	gtccgctggg	6720
tcagggtatt	gccaacgcag	tcgggtatgg	gattgcagaa	aaaacgctgg	cggcgcagtt	6780
taaccgtcca	ggtcacgaca	ttgtcgacca	ctacacctac	gccttcatgg	gcgacggctg	6840
catgatggaa	ggcatctccc	acgaagtttg	ctctctggcg	ggtagcgtga	agctgggtaa	6900
actgattgcg	ttctacgatg	acaacggtat	ctcaatcgat	ggtcacgttg	aaggctggtt	6960
cactgacgac	accgcaatgc	gtttcgaagc	ttacggctgg	cacgttattc	gcgacatcga	7020
cggtcgatgac	gcggcatcca	tcaaacgcgc	agtagaagaa	gcgcgcgcag	tgactgacaa	7080
accgtccctg	ctgatgtgca	aaaccatcat	cggtttcggt	tccccgaaca	aagccgggtac	7140
ccacgactcc	cacggtgctc	cgctggggcg	cgctgaaatt	gccctgaccc	gcgaacagct	7200
gggctggaaa	tacgcgccgt	tcgaaatccc	gtctgaaatc	tatgctcagt	gggatgcgaa	7260
agaagcaggc	caggcgaaag	aatctgcatg	gaatgagaag	tttgcggtt	acgcgaaagc	7320
ttatccgcag	gaagcggctg	aattttaccg	ccgtatgaaa	ggcgaaatgc	cgtctgactt	7380
cgacgccaaa	gcgaaagagt	ttatcgctaa	actgcaggct	aatccggcga	aaatcgccag	7440
ccgtaaagcg	tcgcagaatg	ctatcgaagc	gttcggcccc	ctggtgcctg	aattcctcgg	7500
cggctctgct	gacctggcac	cgtctaacct	gacctgtgg	tctggttcta	aagcaatcaa	7560
cgaagatgct	gcaggtaact	acatccacta	cgggtgttcg	gagttcggta	tgaccgcgat	7620
tgctaacggg	atctccctgc	acgggtggtt	cctgccgtac	acctccacct	tcctgatggt	7680
cgtggaatac	gcacgtaacg	ccgtacgtat	ggctgcgctg	atgaaacagc	gtcagggtgat	7740
ggtttacacc	cacgactcca	tcgggtctgg	cgaagatggc	ccgactcacc	agccgggtga	7800
gcaggctcgt	tctctgcgcg	tgacccccgaa	catgtctaca	tggcgtccgt	gtgaccaggt	7860
tgaatccgcg	gtcgcgtgga	aatacggcgt	tgagcgtcag	gacggccccg	ctgcgcttat	7920
cctctcccgt	cagaacctgg	cgcagcagga	acgaactgaa	gagcaactgg	caaacatcgc	7980
gcgcggtggt	tatgtgctga	aagactgcgc	cggtcagccg	gaactgattt	tcacgcctac	8040
cggttcagaa	gttgaactgg	ctggtgctgc	ctacgaaaaa	ctgactgccg	aaggcgtgaa	8100
agcgcgcgtg	gtgtccatgc	cgtctaccga	cgcatttgac	aagcaggatg	ctgcttaccg	8160
tgaatccgta	ctgccgaaag	cggttactgc	acgcgttgct	gtagaagcgg	gtattgctga	8220
ctactgggtac	aagtatgttg	gcctgaacgg	tgctatcgtc	ggatatgacca	ccttcgggtga	8280
atctgctccg	gcagagctgc	tgtttgaaga	gttcggcttc	actggtgata	acgttggtgc	8340

gaaagcaaaa gaactgctgt aattagcatt tcgggtaaaa aggtcgcttc ggcgaccttt	8400
tttattacct tgatatgtcc gtttgccggac aagcaataga taaagcgtgt tgtagatcac	8460
aaatatttat atgcaataaa tatcaattat gtaatatgca tcacgatatg cgtattgaca	8520
tttgttgtta taactataac tcaatgttat ataagaaatt aacttgaaat aattaacaaa	8580
caaaggagtt acagttagaa attgtaggag agatctcgctt ttccgcgaca atctggcgctt	8640
tttcttgcta attccaggat taatccgttc atagtgtaaa accccgttta cacattctga	8700
cggaagatat agattggaag tattgcattc actaagataa gtatggcaac actggaacag	8760
acatgaatta tcagaacgac gattttacgca tcaaagaaat caaagagtta cttcctcctg	8820
tcgcattgct ggaaaaattc cccgctactg aaaatgccgc gaatacgggt gcccatgccc	8880
gaaaagcgat ccataagatc ctgaaaggta atgatgatcg cctgttggtt gtgattggcc	8940
catgctcaat tcatgatcct gtcgcgccaa aagagtatgc cactcgcttg ctggcgctgc	9000
gtgaagagct gaaagatgag ctggaaatcg taatgcgcgt ctattttgaa aagccgcgta	9060
ccacggtggg ctggaaaggg ctgattaacg atccgcatat ggataatagc ttccagatca	9120
acgacggtct gcgtatagcc cgtaaattgc tgcttgatat taacgacagc ggtctgccag	9180
cggcaggtga gtttctcgat atgatcacc cacaatatct cgctgacctg atgagctggg	9240
gcgcaattgg cgcacgtacc accgaatcg aggtgcaccg cgaactggca tcagggcttt	9300
cttgctcggg cggttcaaaa aatggcaccg acggtacgat taaagtggct atcgatgcca	9360
ttaatgccgc cggcgcgcg cactgcttcc tgtccgtaac gaaatggggg cattcggcga	9420
ttgtgaatac cagcggtaac ggcgattgcc atatcattct gcgcggcggt aaagagccta	9480
actacagcgc gaagcacgtt gctgaagtga aagaagggt gaacaaagca ggcctgccag	9540
cacaggtgat gatcgatttc agccatgcta actcgtccaa acaattcaaa aagcagatgg	9600
atgtttgtgc tgacgtttgc cagcagattg ccggtggcga aaaggccatt attggcgtga	9660
tggtggaaag ccatctggtg gaaggcaatc agagcctcga gagcggggag ccgctggcct	9720
acggtaagag catcaccgat gcctgcatcg gctgggaaga taccgatgct ctgttacgtc	9780
aactggcgaa tgcagtaaaa gcgcgtcgcg ggtaagggtt aattgtcgga tgcgccgtca	9840
gagtggcgta tccgatgaat caccacaggc ctgataagtc gcgcagcgtc gcatcaggca	9900
atgtgctcca ttgttagcaa caaaaaagcc gactcacttg cagtcggctt tctcatttta	9960
aacgaatgac gtttacttcg ctttaccctg gtttgcaacc gccgctgctt tcgctctcga	10020
ggctattgac gacagctatg gttcactgtc caccaaccaa aactgtgctc agtaccgcca	10080
atattttctcc cttgaggggt acaaagaggt gtccctagaa gagatccacg ctgtgtaaaa	10140

```

atattacaaa aaggatttga ctttccttac aggggtgtgta ataattttaat tacaggcggg 10200
ggcaaccccg cctgttctag aggaggagga atcgccatgg agaggattgt cgttactctc 10260
ggggaacgta gttacccaat taccatcgca tctggtttgt ttaatgaacc agcttcattc 10320
ttaccgctga aatcgggcga gcaggtcatg ttggtcacca acgaaaccct ggctcctctg 10380
tatctcgata aggtccgcgg cgtacttgaa caggcgggtg ttaacgtcga tagcgttatc 10440
ctccctgacg gcgagcagta taaaagcctg gctgtactcg ataccgtctt tacggcggtg 10500
ttacaaaagc cgcattggtc cgatactacg ctgggtggcg ttggcgggcg cgtagtgggc 10560
gatctgaccg gcttcgcggc ggcgagttat cagcgcgggtg ttcgtttcat tcaagtcccg 10620
acgacgttac tgctgcaggt cgattcctcc gttggcgcca aaactgcggt caaccatccc 10680
ctcggtaaaa acatgattgg cgcgttctac cagcctgctt cagtgggtgt ggatctcgac 10740
tgtctgaaaa cgcttcccc gcgtgagtta gcgtcggggc tggcagaagt catcaaatac 10800
ggcattatct ttgacgggtc gtttttcaac tggctggaag agaactctga tgcgttgttg 10860
cgtctggacg gtccggcaat ggcgtactgt attcgccgtt gttgtgaact gaaggcagaa 10920
gttgctgcgg ccgacgagc cgaaaccggg ttacgtgctt tactgaatct gggacacacc 10980
tttggctcatg ccattgaagc tgaaatgggg tatggcaatt ggttacatgg tgaagcggtc 11040
gctgcgggta tggatgatgg ggcgcggacg tcggaacgtc tcgggcagtt tagttctgcc 11100
gaaacgcagc gtattataac cctgctcacg cgggctgggt taccggtcaa tgggcccgcg 11160
gaaatgtccg cgcaggcgta tttaccgcat atgctgcgtg acaagaaagt ccttgcggga 11220
gagatgcgct taattcttcc gttggcaatt ggtaagagtg aagttcgcag cggcggttctg 11280
cacgagcttg ttcttaacgc cattgccgat tgtcaatcag cgtaatcatc gttcatgcct 11340
gatgccgcta ttagggccgg ataaggcgtt cagccgcgat ccggcaaccg atgcctgatg 11400
cgacgcggtc gcgtcttctc aggcctacag gtcgatgccg atatgtacat cgtattcggc 11460
aattaataca tagca 11475

```

```

<210> 33
<211> 7769
<212> ADN
<213> Plasmid pYEP24;

```

```

<400> 33
gaattctgaa ccagtcctaa aacgagtaaa taggaccggc aattcttcaa gcaataaaca 60
ggaataccaa ttattaaaag ataacttagt cagatcgtag aataaagctt tgaagaaaaa 120
tgccgcttat tcaatctttg ctataaaaaa tggcccaaaa tctcacattg gaagacattt 180
gatgacctca tttctttcaa tgaagggcct aacggagttg actaatgttg tgggaaattg 240

```

gagcgataag cgtgcttctg ccgtggccag gacaacgtat actcatcaga taacagcaat	300
acctgatcac tacttcgcac tagtttctcg gtactatgca tatgatccaa tatcaaagga	360
aatgatagca ttgaaggatg agactaatcc aattgaggag tggcagcata tagaacagct	420
aaagggtagt gctgaaggaa gcatacgata ccccgcatgg aatgggataa tatcacagga	480
ggtactagac tacctttcat cctacataaa tagacgcata taagtaagca tttaagcata	540
aacacgcact atgccgttct tctcatgtat atatatatac aggcaacacg cagatatagg	600
tgcgacgtga acagtgagct gtatgtgcgc agctcgcgtt gcatttttcgg aagcgctcgt	660
tttcggaaac gctttgaagt tcctattccg aagttcctat tctctagaaa gtataggaac	720
ttcagagcgc ttttgaaaac caaaagcgct ctgaagacgc actttcaaaa aaccaaaaac	780
gcaccggact gtaacgagct actaaaatat tgcgaatacc gcttccacaa acattgctca	840
aaagtatctc tttgctatat atctctgtgc tatatcccta tataacctac ccatccacct	900
ttcgctcctt gaacttgcac ctaaactcga cctctacatt ttttatgttt atctctagta	960
ttactcttta gacaaaaaaa ttgtagtaag aactattcat agagtgaatc gaaaacaata	1020
cgaaaatgta aacatttcct atacgtagta tatagagaca aaatagaaga aaccgttcat	1080
aattttctga ccaatgaaga atcatcaacg ctatcacttt ctgttcacaa agtatgcgca	1140
atccacatcg gtatagaata taatcgggga tgcctttatc ttgaaaaaat gcaccgcgag	1200
cttcgctagt aatcagtaaa cgcgggaagt ggagtcaggc tttttttatg gaagagaaaa	1260
tagacaccaa agtagccttc ttctaacctt aacggaccta cagtgcacaaa agttatcaag	1320
agactgcatt atagagcgca caaaggagaa aaaaagtaat ctaagatgct ttgttagaaa	1380
aatagcgctc tcgggatgca tttttgtaga acaaaaaaga agtatagatt ctttggtggt	1440
aaaatagcgc tctcgcgttg catttctgtt ctgtaaaaat gcagctcaga ttctttgttt	1500
gaaaaattag cgctctcgcg ttgcattttt gttttacaaa aatgaagcac agattcttcg	1560
ttggtaaaat agcgctttcg cgttgcattt ctgttctgta aaaatgcagc tcagattctt	1620
tgtttgaaaa attagcgctc tcgcgttgca ttttgttctt acaaaatgaa gcacagatgc	1680
ttcgttaaca aagatatgct attgaagtgc aagatggaaa cgcagaaaaat gaaccgggga	1740
tgcgacgtgc aagattacct atgcaataga tgcaatagtt tctccaggaa ccgaaataca	1800
tacattgtct tccgtaaagc gctagactat atattattat acaggttcaa atatactatc	1860
tgtttcaggg aaaactccca ggttcggatg ttcaaaatc aatgatgggt aacaagtacg	1920
atcgtaaadc tgtaaaacag tttgtcggat attaggctgt atctcctcaa agcgtattcg	1980
aatatcattg agaagctgca gcgtcacatc ggataataat gatggcagcc attgtagaag	2040
tgccttttgc atttctagtc tctttctcgg tctagctagt ttactacat cgcgaagata	2100

gaatccttaga tcacactgcc tttgctgagc tggatcaata gagtaacaaa agagtggtaa	2160
ggcctcgтта aaggacaagg acctgagcgg aagtgtatcg tacagtagac ggagtatact	2220
agtatagtct atagtccgtg gaattctcat gtttgacagc ttatcatcga taagcttttc	2280
aattcaattc atcatttttt ttttattctt ttttttgatt tcggtttctt tgaaattttt	2340
ttgattcggт aatctccgaa cagaaggaag aacgaaggaa ggagcacaga cttagattgg	2400
tatatatacg catatgtagt gttgaagaaa catgaaattg cccagtattc ttaacccaac	2460
tgcacagaac aaaaacctgc aggaaacgaa gataaatcat gtcgaaagct acatataagg	2520
aacgtgctgc tactcatcct agtcctgttg ctgccaaгсt atttaatatc atgcacgaaa	2580
agcaaacaaa cttgtgtgct tcattggatg ttcgtaccac caaggaatta ctggagttag	2640
ttgaagcatt aggtcccaaa atttgtttac taaaaacaca tgtggatatc ttgactgatt	2700
tttccatgga gggcacagtt aagccgctaa aggcattatc cgccaagtac aattttttac	2760
tcttcgaaga cagaaaattt gctgacattg gtaatacagt caaattgcag tactctgcgg	2820
gtgtatacag aatagcagaa tgggcagaca ttacgaatgc acacgggtgtg gtgggcccag	2880
gtattgttag cggtttgaag caggcggcag aagaagtaac aaaggaacct agaggccttt	2940
tgatgttagc agaattgtca tgcaagggtc ccctatctac tggagaatat actaagggtta	3000
ctgttgacat tgcgaagagc gacaaagatt ttgttatcgg ctttattgct caaagagaca	3060
tgggtggaag agatgaaggт tacgattggт tgattatgac acccggtgtg ggtttagatg	3120
acaagggaga cgcattgggt caacagtata gaaccgtgga tgatgtggtc tctacaggat	3180
ctgacattat tattgttgga agaggactat ttgcaaaggг aagggatgct aaggtagagg	3240
gtgaacgtta cagaaaagca ggctgggaag catatttgag aagatgcggc cagcaaaact	3300
aaaaaactgt attataagta aatgcatgta tactaaactc acaaattaga gcttcaattt	3360
aattatatca gttattaccg gggaatctcg gtcgtaatga tttttataat gacgaaaaaa	3420
aaaaaattgg aaagaaaaag ctttaatgсg gtagtttatc acagttaaат tgctaacgca	3480
gtcaggcacc gtgtatgaaa tctaacaatg cgctcatcgt catcctcggc accgtcacc	3540
tggatgctgt aggcataggc ttggttatgc cggтactgcc gggcctcttg cgggatatcg	3600
tccattccga cagcatcgcc agtcactatg gcgtgctgct agcgctatat gcgttgatgc	3660
aatttctatg cgcacccgtt ctсggagcac tgtccgaccg ctttgгccgc cgcccagtcc	3720
tgctcgcttc gctacttgga gccactatcg actacgcgat catggcgacc acacccgtcc	3780
tgtggatcct ctacgccgga cgcatcgтg ccggcatcac cggcgccaca ggtgcggттg	3840
ctggcgцcta tatcgccgac atcaccgatg gggaagatcg ggctcgccac ttcgggctca	3900

tgagcgcttg	tttcggcgtg	ggtatggtgg	caggccccgt	ggccggggga	ctgttgggcg	3960
ccatctcctt	gcatgcacca	ttccttgcg	cggcggtgct	caacggcctc	aacctactac	4020
tgggctgctt	cctaatacag	gagtcgcata	agggagagcg	tcgaccgatg	cccttgagag	4080
ccttcaaccc	agtcagctcc	ttccggtggg	cgcggggcat	gactatcgtc	gccgcactta	4140
tgactgtctt	ctttatcatg	caactcgtag	gacaggtgcc	ggcagcgctc	tgggtcattt	4200
tcggcgagga	ccgctttcgc	tggagcgga	cgatgatcgg	cctgtcgctt	gcggtattcg	4260
gaatcttgca	cgccctcgct	caagccttcg	tactggtcc	cgccacaaaa	cgtttcggcg	4320
agaagcaggc	cattatcgcc	ggcatggcgg	ccgacgcgct	gggctacgtc	ttgctggcgt	4380
tcgcgacg	aggctggatg	gccttcccc	ttatgattct	tctcgcttcc	ggcggcacg	4440
ggatgcccgc	gttgacggcc	atgctgtcca	ggcaggtaga	tgacgaccat	cagggacagc	4500
ttcaaggatc	gctcgcggt	cttaccagcc	taacttcgat	cactggaccg	ctgatcgta	4560
cggcgattta	tgccgcctcg	gcgagcacat	ggaacgggtt	ggcatggatt	gtaggcgccg	4620
ccctatacct	tgtctgcctc	cccgcgttgc	gtcgcggtgc	atggagccgg	gccacctcga	4680
cctgaatgga	agccggcggc	acctcgctaa	cggattcacc	actccaagaa	ttggagccaa	4740
tcaattcttg	cggagaactg	tgaatgcgca	aaccaaccct	tggcagaaca	tatccatcgc	4800
gtccgccatc	tccagcagcc	gcacgcggcg	catctcgggc	agcgttgggt	cctggccacg	4860
ggtgcgcatg	atcgtgctcc	tgtcgttgag	gaccgggcta	ggctggcggg	gttgcccttac	4920
tggttagcag	aatgaatcac	cgatacgca	gcgaacgtga	agcgactgct	gctgcaaaac	4980
gtctgcgacc	tgagcaacaa	catgaatggt	cttcggtttc	cgtgtttcgt	aaagtctgga	5040
aacgcggaag	tcagcgccct	gcaccattat	gttccggatc	tgcacgcgag	gatgctgctg	5100
gctaccctgt	ggaacaccta	catctgtatt	aacgaagcgc	tggcattgac	cctgagtgat	5160
ttttctcttg	tccgcgcgca	tccataccgc	cagttgttta	ccctcacaac	gttccagtaa	5220
ccgggcatgt	tcatcatcag	taaccggtat	cgtgagcatc	ctctctcggt	tcatcggtat	5280
cattaccccc	atgaacagaa	attccccctt	acacggaggc	atcaagtgac	caaacaggaa	5340
aaaaccgccc	ttaacatggc	ccgctttatc	agaagccaga	cattaacgct	tctggagaaa	5400
ctcaacgagc	tggacgcgga	tgaacaggca	gacatctgtg	aatcgcttca	cgaccacgct	5460
gatgagcttt	accgcagctg	cctcgcgcg	ttcgggtgatg	acggtgaaaa	cctctgacac	5520
atgcagctcc	cggagacggg	cacagcttgt	ctgtaagcgg	atgccgggag	cagacaagcc	5580
cgtcagggcg	cgtcagcggg	tgttggcggg	tgtcggggcg	cagccatgac	ccagtcacgt	5640
agcgatagcg	gagtgtatac	tggcttaact	atgcggcatc	agagcagatt	gtactgagag	5700
tgcaccatat	gcggtgtgaa	ataccgcaca	gatgcgtaag	gagaaaatac	cgcacagggc	5760

gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggc	cgttcggctg	cggcgagcgg	5820
tatcagctca	ctcaaaggcg	gtaatacggc	tatccacaga	atcaggggat	aacgcaggaa	5880
agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	5940
cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	tcaagtcaga	6000
ggtggcgaaa	cccgcacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	6060
tgcgctctcc	tggtccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	6120
gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	taggtcggtc	6180
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	6240
gtaactatcg	tcttgagtcc	aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	6300
ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtggc	6360
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	6420
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	6480
gtgggttttt	tggttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	6540
ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	taagggattt	6600
tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	6660
ttaaatcaat	ctaaagtata	tatgagtaaa	cttggctctga	cagttaccaa	tgcttaatca	6720
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	tgactccccg	6780
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	ccccagtgc	gcaatgatac	6840
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	6900
ccgagcgcag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattggtgcc	6960
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgctg	7020
caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	7080
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	7140
ctccgatcgt	tgtcagaagt	aagttggccg	cagtgttatc	actcatgggt	atggcagcac	7200
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	7260
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	7320
cacgggataa	taccgcgcca	catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	7380
cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	7440
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	7500
aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tggtgaatac	7560

tcatactctt cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg 7620
gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttcgcgcg acatttcccc 7680
gaaaagtgcc acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata 7740
ggcgtatcac gaggcccttt cgtcttcaa 7769

<210> 34

<211> 2000

<212> ADN

<213> Trình tự của vùng aroE khuyết đoạn;

<400> 34

actacgtccg tcctctgaaa tcttcagcgg atggacatat cgtcaaagtt ctggaggggc 60
aggtttgccc tgcattgtggc gcaaactctgg tattacgccca gggacgcctt ggtatgttta 120
ttggttgcat taactaccct gaatgcgaac ataccgaact tatcgataaa ccggacgaaa 180
cagcaattac atgcccccaa tgtcggacgg gccatctggg ccagcgcgcg tcccgttatg 240
gcaaaacatt tcactcttgt gatcgctacc cggagtgtca atttgccatt aacttcaaac 300
ccatagctgg agaatgccct gagtgtcatt atccgctact catcgaaaag aaaaccgcgc 360
agggtgtaaa acacttttgt gccagtaaac aatgtggaaa gccggtttcg gcggaataat 420
aacgtgaata ataacctgca aagagacgct atcgcagctg cgatagatgt tctcaatgaa 480
gaacgtgtca tcgcctatcc aacggaagcc gttttcgggtg ttgggtgcga tcctgatagc 540
gaaacagcag tgatgcgact gttggagtta aaacagcgtc cggttgataa ggggctgatt 600
ttaatcgcag caaattacga gcagcttaaa ccctatattg atgacaccat gttgactgac 660
gtgcagcgtg aaaccatttt tcccgcgtgg ccaggctctg tcacctttgt ctttcccgcg 720
cctgcgacaa caccgcgtg gttgacgggc cgctttgatt cgcttgctgt acgagtcacc 780
gaccatccgt tgggtggttg tttgtgccag gcttatggta aaccgctggg ttctaccagt 840
gccaacttga gtggattgcc accttgctga acagtagacg aagttcgcgc acaatttggc 900
gcggcgttcc cggttggtgc tggtgaaacg gggggcggtt taaatccttc agaaatccgc 960
gatgccctga cgggtgaact gtttcgacag gggtaacata atcaggccat ccagtttccg 1020
gacaggggaag agtgggacga gaataaaaaa tgtgtatggt tcccgcgtct cgtgaatggt 1080
atgcaactga catgcgcgat ctctggcgag agtctggcgt atcgctttac tggagatacg 1140
ccagaacagt ggtagcgag ttttcgtcag catcgctggg acctggaaga agaagcggaa 1200
aacttaattc aggaacaaag tgaagatgat caaggctggg tctggttacc ctgatccaga 1260
tattcgtcct tccatttcac gtaattattc gcggaatagc gtaaccacgc cttctcttca 1320
tcacttaacg ggcggatctg tttgacgggg ctaccgagat acagatatcc gctctccagc 1380

```

cgtttatttt gtgggaccag actaccgca ccaatcatca catcatcttc tactattgcg      1440
ccatcaagta aaattgagcc catcccaacc aaaactcgat tgccaatggg gcagccgtgg      1500
agcatcacct tgtgaccaac agtgacatct tcgccaatgg ttaatggggt gccatctggg      1560
ttgtacgagg atttatgagt gacatgcaac atactgccat cctggatatt ggtgcgtgct      1620
ccgatctgta cataatgtac atctccacga atcacaacga gcggccagat ccccatcatca      1680
tcagccagac gaacgtcacc aatcacgaca ctgctatcgt cgatcattac gcgctgaccg      1740
at ttgtggaa aaagatcgcg gtatgggcgt aaaacatcag acatacttac ctgagcaata      1800
aatgatttac taatgacttt gggggcatta ttggccttgt gcaagtcttt tagtatgcaa      1860
aaaagcaccg ttttgtgtgc gattgcagca aaaagggtga aaaaacaaca aacagaaaaa      1920
aagatcaaaa aaatacttgt gcaaaaaatt gggatcccta taatgcgcct ccgttgagac      1980
gacaacgtga aacacttcac                                          2000

```

<210> 35

<211> 2460

<212> ADN

<213> Trình tự của catxet P15aroB được gắn ở locus ack khuyết đoạn;

<400> 35

```

gatcggcgggc ataaaacgga tcgcataacg cgtcatcttg ataacgcgat tttcgacaaa      60
gaccgggggca aggcgttttt ccagcggcca cgtctttgag taatgctgtc cccggcgaaa      120
caagctaaaa aaattaacag aacgattatc cggcgttgac atgcttcacc tcaacttcac      180
atataaagat tcaaaaattt gtgcaaattc acaactcagc gggacaacgt tcaaaacatt      240
ttgtcttcca taccactat caggtatcct ttagcagcct gaaggcctaa gtagtacata      300
ttcattgagt cgtcaaattc atatacatta tgccattggc tgaaaattac gcaaaatggc      360
atagactcaa gatatttctt ccatcatgca aaaaaaatt tgcagtgcac gatgttaatc      420
ataaatgtcg gtgtcatcat gcgctacgct ctatggctcc ctgacgtttt tttagccacg      480
tatcaattat aggtacttcc ctcgaggcta ttgacgacag ctatggttca ctgtccacca      540
acaaaaactg tgctcagtac cgccaatatt tctcccttga ggggtacaaa gaggtgtccc      600
tagaagagat ccacgctgtg taaaaatttt acaaaaagggt attgactttc cctacagggg      660
gtgtaataat ttaattacag gcggggggcaa cccgcctgt tctagaggag gaggaatcgc      720
catggagagg attgtcgta ctctcgggga acgtagttac ccaattacca tcgcatctgg      780
tttgtttaat gaaccagctt cattcttacc gctgaaatcg ggcgagcagg tcatgttggt      840
caccaacgaa accctggctc ctctgtatct cgataaggct cgcggcgtac ttgaacaggc      900
gggtgttaac gtcgatagcg ttatcctccc tgacggcgag cagtataaaa gcctggctgt      960

```

actcgatacc gtcttttacgg cgttggttaca aaagccgcat ggtcgcgata ctacgctggg 1020
 ggcgcttggc ggcggcgtag tgggcgatct gaccggcttc gcggcgccga gttatcagcg 1080
 cgggtgttcgt ttcattcaag tcccgacgac gttactgtcg caggtcgatt cctccgttgg 1140
 cggcaaaact gcggtcaacc atcccctcgg taaaaacatg attggcgcggt tctaccagcc 1200
 tgcttcagtg gtggtggatc tcgactgtct gaaaacgctt cccccgcgtg agttagcgtc 1260
 ggggctggca gaagtcatca aatacggcat tattcttgac ggtgcgtttt tcaactggct 1320
 ggaagagaat ctggatgcgt tgttgcgctc ggacgggtccg gcaatggcgt actgtattcg 1380
 ccgttggttg gaactgaagg cagaagttgt cgccgccgac gagcgcgaaa ccggggttacg 1440
 tgctttactg aatctgggac acacctttgg tcatgccatt gaagctgaaa tgggggtatgg 1500
 caattgggta catggtgaag cggtcgctgc gggatgggtg atggcgccgc ggacgtcgga 1560
 acgtctcggg cagtttagtt ctgccgaaac gcagcgtatt ataaccctgc tcacgcgggc 1620
 tggggtaccg gtcaatgggc cgccgcaaat gtccgcgcag gcgtatttac cgcataatgct 1680
 gcgtgacaag aaagtccttg cgggagagat gcgcttaatt cttccgttgg caattggtaa 1740
 gagtgaagtt cgcagcggcg tttcgcacga gcttggttctt aacgccattg ccgattgtca 1800
 atcagcgtaa tcatcggtca tgctgatgc cgctatgtag gccggataag gcgttcacgc 1860
 cgcateccggc aaccgatgcc tgatgcgacg cggtcgcgtc ttatcaggcc tacaggtcga 1920
 tgccgatatg tacatcgatc tcggcaatta atacatagca tttcacaccg ccagctcagc 1980
 tggcgggtgct gttttgtaac ccgccaaatc ggcggtaacg aaagaggata aaccgtgtcc 2040
 cgtattatta tgctgatccc taccggaacc agcgtcggtc tgaccagcgt cagccttggc 2100
 gtgatccgtg caatggaacg caaaggcgtt cgtctgagcg ttttcaaacc tatcgctcag 2160
 ccgcgtaccg gtggcgatgc gcccgatcag actacgacta tcgtgcgtgc gaactcttcc 2220
 accacgacgg ccgctgaacc gctgaaaatg agctacgttg aaggtctgct ttccagcaat 2280
 cagaaagatg tgctgatgga agagatcgtc gcaaactacc acgctaacac caaagacgct 2340
 gaagtcgttc tgggtgaagg tctgggtccc acacgtaagc accagtttgc ccagttctctg 2400
 aactacgaaa tcgctaaaac gctgaatgcg gaaatcgtct tcgttatgtc tcagggcact 2460

<210> 36

<211> 1000

<212> ADN

<213> Trình tự của vùng poxB khuyết đoạn;

<400> 36

gcggcccgcc tccgtatatg gattgggtag agcaggaagt gaaagcgctc ggcgtgacgc 60
 gtttctttta agagaaattc ttcaccccag tagcgggaagc agcgaccagc ggtctgaaat 120

tcaccaaact gcaaccggca cgagaatttt acgccccggt tggcaccacg ctactggagg	180
cgctggaaag caataacggt ccggttgctg ccgcctgccg tgcgggtggt tgcggctgct	240
gtaagacaaa agtggtttcc ggtgaatata cggtgagcag cacaatgacg ctgaccgacg	300
ccgaaatcgc tgaagggttac gtactggcct gtcctgccca tccgcagggg gatttggttc	360
tcgcataatc gccttatgcc cgatgatatt cctttcatcg ggctatttaa ccgttagtgc	420
ctcctttctc tcccatccct tccccctccg tcagatgaac taaacttggt accgttatca	480
cattcaggag atggagaacc aaaggggtggc atttcccgtc ataataagga catgccatga	540
ttgatttacg cagtgatacc gttacccgac cgagccgcgc catgctcgaa gcaatgatgg	600
ccgccccggt tggggacgac gtttacggag acgaccctac cgттаатgct ctgcaggact	660
acgccgcaga gctttccggt aaagaagccg ccatttttct gccgaccggc actcaggcca	720
acctggtcgc tctgctcagt cactgcgaac gtggcgaaga gtatattgtc ggtcaggccg	780
cgcataacta tctgtttgaa gccgggtggcg cagcgggtgct gggcagtatt cagccgcaac	840
ccatcgacgc ggctgccgac ggcacgctac cgctggataa agtggcgatg aaaatcaaac	900
ccgacgatat ccatttcgcc cgcaccaaat tactcagtct ggaaaacacc cacaacggca	960
aagtgctgcc gcgtgaatac ctgaaagaag catgggaatt	1000

<210> 37

<211> 3392

<212> ADN

<213> catxet tktA tại locus poxB khuyết đoạn;

<400> 37

gcgccccggc tccgtatatg gattgggtag agcaggaagt gaaagcgctc ggcgtgacgc	60
gtttctttta agagaaattc ttcacccag tagcgggaagc agcgaccagc ggtctgaaat	120
tcaccaaact gcaaccggca cgagaatttt acgccccggt tggcaccacg ctactggagg	180
cgctggaaag caataacggt ccggttgctg ccgcctgccg tgcgggtggt tgcggctgct	240
gtaagacaaa agtggtttcc ggtgaatata cggtgagcag cacaatgacg ctgaccgacg	300
ccgaaatcgc tgaagggttac gtactggcct gtcctgccca tccgcagggg gatttggttc	360
tcgcataatc gccttatgcc cgatgatatt cctttcatcg ggctatttaa ccgttagtgc	420
ctcctttctc tcccatccct tccccctccg tcagatgaac taaacttggt accgttatca	480
cattcaggag atggagaacc aaggaaaagc gcaacggacg ggcgagtaga ttgcgcaaca	540
tgcgagcatg atccagagat ttctgaagca gcaaaaggat gttccatgta catgacgcgc	600
ggcttgcggt aaattgttgg caaatTTTcc ggcgtagccc aaaacgcgct gtcgtcaagt	660
cgттаagggc gtgccttca tcatccgatc tggagtcaaa atgtcctcac gtaaagagct	720

tgccaatgct attcgtgcg	tgagcatgga cgcagtacag	aaagccaaat ccggtcaccc	780
gggtgcccc	atgggtatgg ctgacattgc	cgaagtccctg tggcgtgatt	840
caaccgcag aatccgtcct	gggctgaccg tgaccgcttc	gtgctgtcca acggccacgg	900
ctccatgctg atctacagcc	tgctgcacct caccggttac	gatctgccga tggaagaact	960
gaaaaacttc cgtcagctgc	actctaaaac tccgggccac	ccggaagtag gttataccgc	1020
tgggtgtggaa accaccaccg	gtccgctggg tcaggggtatt	gccaacgcag tcggtatggc	1080
gattgcagaa aaaacgctgg	cggcgcagtt taaccgtcca	ggtcacgaca ttgtcgacca	1140
ctacacctac gccttcatgg	gcgacggctg catgatggaa	ggcatctccc acgaagtttg	1200
ctctctggcg ggtacgctga	agctgggtaa actgattgcg	ttctacgatg acaacggtat	1260
ctcaatcgat ggtcacgttg	aaggctgggt cactgacgac	accgcaatgc gtttcgaagc	1320
ttacggctgg cacgttattc	gcgacatcga cggtcatgac	gcggcatcca tcaaacgcgc	1380
agtagaagaa gcgcgcgcag	tgactgacaa accgtccctg	ctgatgtgca aaaccatcat	1440
cggtttcggg tccccgaaca	aagccgggtac ccacgactcc	cacgggtgcgc cgctgggcga	1500
cgctgaaatt gccctgaccc	gcgaacagct gggctggaaa	tacgcgccgt tcgaaatccc	1560
gtctgaaatc tatgctcagt	gggatgcgaa agaagcaggc	caggcgaaag aatctgcatg	1620
gaatgagaag tttgcggctt	acgcgaaagc ttatccgcag	gaagcggctg aatttaccgc	1680
ccgtatgaaa ggcgaaatgc	cgtctgactt cgacgccaaa	gcgaaagagt ttatcgctaa	1740
actgcaggct aatccggcga	aaatcgccag ccgtaaagcg	tcgcagaatg ctatcgaagc	1800
gttcggcccc ctggtgctg	aattcctcgg cggctctgct	gacctggcac cgtctaacct	1860
gacctgtgg tctggttcta	aagcaatcaa cgaagatgct	gcaggtaact acatccacta	1920
cgggtgttcgc gagttcggt	tgaccgcgat tgctaacgg	atctccctgc acggtggttt	1980
cctgccgtac acctccacct	tctgatgtt cgtggaatac	gcacgtaacg ccgtacgtat	2040
ggctgcgctg atgaaacagc	gtcagggtgat ggtttacacc	cacgactcca tcggctctggg	2100
cgaagatggc ccgactcacc	agccggttga gcaggctcgt	tctctgcgcg tgaccccgaa	2160
catgtctaca tggcgtccgt	gtgaccagg	tgaatccgcg gtcgcgtgga	2220
tgagcgtcag gacggcccga	ctgcgcttat cctctcccg	cagaacctgg cgcagcagga	2280
acgaactgaa gagcaactgg	caaacatcgc gcgcgggtgg	tatgtgctga aagactgcgc	2340
cggtcagccg gaactgattt	tcatcgctac cggttcagaa	gttgaaactgg ctggtgctgc	2400
ctacgaaaaa ctgactgccg	aaggcgtgaa agcgcgcgtg	gtgtccatgc cgtctaccga	2460
cgcatttgac aagcaggatg	ctgcttaccg tgaatccgta	ctgccgaaag cggttactgc	2520
acgcgttgct gtagaagcgg	gtattgctga ctactggtac	aagtatgttg gcctgaacgg	2580

tgctatcgtc ggtatgacca ccttcggtga atctgctccg gcagagctgc tgtttgaaga	2640
gttcggcttc actggtgata acgttggtgc gaaagcaaaa gaactgctgt aattagcatt	2700
tcgggtaaaa aggtcgcttc ggcgaccttt tttattacct tgatatgtcc gtttgcggaac	2760
aagcaataga taaagcgtgt tgtagatcac aaatatattat atgcaataaa tatcaattat	2820
gtaatatgca tcacgatatg cgtattgaca tttgttggtta taactataac tcaatgttat	2880
ataagaaatt aaaaaggggtg gcattttcccg tcataataag gacatgccat gattgattta	2940
cgcagtata ccggttaccg accgagccgc gccatgctcg aagcaatgat ggccgccccg	3000
gttggggacg acgtttacgg agacgacctt accgttaatg ctctgcagga ctacgccgca	3060
gagctttccg gtaaagaagc cgccattttt ctgccgaccg gcaactcaggc caacctggtc	3120
gctctgctca gtcactgca acgtggcgaa gagtatattg tcggtcaggc cgcgcataac	3180
tatctgtttg aagccggttg cgcagcgggtg ctgggcagta ttcagccgca acccatcgac	3240
gcggctgccg acggcacgct accgctggat aaagtggcga tgaaaatcaa acccgacgat	3300
atccatttcg cccgcaccaa attactcagt ctggaaaaca cccacaacgg caaagtgctg	3360
ccgcgtgaat acctgaaaga agcatgggaa tt	3392

<210> 38

<211> 1045

<212> ADN

<213> trình tự ptsHI khuyết đoạn;

<400> 38

gaagatgaaa gctttaccaa caagaatatt gtggttattc taccatcatc gggtgagcgt	60
tatttaagca ccgcattggt tgccgatctc ttcactgaga aagaattgca acagtaatgc	120
cagcttggtta aaaatgcgta aaaaagcacc tttttaggtg cttttttgtg gcctgcttca	180
aactttcgcc cctcctggca ttgattcagc ctgtcggaac tggtatatta ccagactaat	240
tattttgatg cgcgaaatta atcgttacag gaaaagccaa agctgaatcg attttatgat	300
ttggttcaat tcttccttta gcggcataat gtttaatgac gtacgaaacg tcagcgggtca	360
acaccgcga gcaatggact gtattgcgct cttcgtgcgt cgcgtctggt aaaaactggc	420
gctaacaata caggctaaag tcgaaccgcc aggctagact ttagttccac aacactaaac	480
ctataagttg gggaaatata atgttccagc aagaagttac cattaccgct ccgacaatct	540
gctaattccac gagatgcggc ccaatttact gcttaggaga agatcatggg tttgttcgat	600
aaactgaaat ctctggtttc cgacgacaag aaggataccg gaactattga gatcattgct	660
ccgctctctg gcgagatcgt caatatcgaa gacgtgccgg atgtcgtttt tcgggaaaaa	720
atcgttgggtg atggtattgc tatcaaacca acgggtaaca aaatggtcgc gccagtagac	780

ggcaccattg gtaaaatctt tgaaaccaac cacgcattct ctatcgaatc tgatagcggc 840
 gttgaactgt tcgtccactt cggatatcgc accggtgaac tgaaaggcga aggcttcaag 900
 cgtattgctg aagaaggtca gcgcgtgaaa gttggcgata ctgtcattga atttgatctg 960
 ccgctgctgg aagagaaagc caagtctacc ctgactccgg ttgttatctc caacatggac 1020
 gaaatcaaag aactgatcaa actgt 1045

<210> 39

<211> 4595

<212> ADN

<213> catxet glf-glκ được gán tại locus tđc khuyết đoạn;

<400> 39

ctgatttctt tgtcgtgat cccttactgg aactctgcag ttatcgacca ggttgacctc 60
 ggttcgctgt cgtaaccgg tcatgacgg atcctgatca ctgtctggct ggggatttcc 120
 atcatggttt tctcctttaa cttctcgcca atcgtctctt ccttcgtggg ttctaaacgt 180
 gaagagtatg agaaagactt cggtcgcgac ttcaccgaac gtaaagtgtc ccaatcatt 240
 tctcgtgcca gcatgctgat ggttgcagt gtgatgttct ttgcctttag ctgcctgttt 300
 actctgtctc cggccaacat ggcggaagcc aaagcgcaga atattccagt gctttcttat 360
 ctggctaacc actttgcgtc catgaccgg accaaaacaa cgttcgcgat tacactggaa 420
 tatgcggctt ccatcatcgc actcgtgggt atcttcaaat ctttcttcgg tcactatctg 480
 gggacgctgg aaggcttgaa tggctctgatt ctgaagttcg gttataaagg tgacaaaacc 540
 aaagtgtcgc tgggtaaaact gaatactctc agcatgatct tcatcatggg ctccacctgg 600
 gttgttgctt acgccaacc gaacatcctc gacctgattg aagccatggg cgcaccgatt 660
 atcgcacccc tgctgtgcct gttgccgatg tatgccatcc gtaaagcgcc gtctctggcg 720
 aaataccgtg gtcgtctgga taacgtgttt gttaccgtga ttggtctgct gaccatcctg 780
 aacatcgtat acaaactgtt ttaatccgta actcaggatg agaaaagaga tgaatgaatt 840
 tccggttggt ttggttatta actgtgggtc gtcttcgatt aagttttccg tactcgatgc 900
 cagcgactgt gaagtattaa tgtcaggatg tgccgacggg attaactcgg aaaatgcatt 960
 cttatccgta aatgggggag agccagcacc gctggctcac cacagctacg aagggtgcatt 1020
 gaaggcaatt gcatttgaac tggaaaaacg gagtttaaat gacagcgtgg ccttaattgg 1080
 tgcaaaagtg gctgtgactg taaaaagaaa tcgaaaaaga ccgttttgtg tgaaaacggt 1140
 ctttttgttt ccttttaacc aactgccata actcgaggcc tacctagctt ccaagaaaga 1200
 tatictaaca gcacaagagc ggaaagatgt tttgttctac atccagaaca acctctgcta 1260
 aaattcctga aaaattttgc aaaaagttgt tgactttatc tacaaggtgt ggtataataa 1320

tcttaacaac agcaggacgc tctagagggga gaggaatcgc catgagttct gaaagtagtc	1380
agggtctagt cacgcgacta gccctaatcg ctgctatagg cggcttgctt ttcggttacg	1440
attcagcggg tategctgca atcgggtacac cgggtgatat ccattttatt gccctcgtc	1500
acctgtctgc tacggctgcg gcttcctttt ctgggatggg cgttggttgcg gttttggctg	1560
gttggtgttac cggttctttg ctgtctgggt ggattgggtat tcgcttcggg cgtcgcggcg	1620
gattgttgat gagttccatt tgtttcgtcg ccgccggttt tgggtgctgcg ttaaccgaaa	1680
aattatttgg aaccgggtgg tccggctttac aaattttttg ctttttccgg tttcttgccg	1740
gtttaggtat cgggtgctgt tcaaccttga cccaacctata tattgctgaa attgctccgc	1800
cagacaaacg tggtcagatg gtttctgggtc agcagatggc cattgtgacg ggtgctttaa	1860
ccggttatat ctttacctgg ttactggctc atttcgggtc tategattgg gttaatgcca	1920
gtgggtgggtg ctgggtctccg gcttcagaag gcctgatcgg tattgccttc ttattgctgc	1980
tgtaaccgc accggatagc ccgcattggg tgggtgatgaa gggacgtcat tccgaggcta	2040
gcaaaatcct tgctcgtctg gaaccgcaag ccgatcctaa tctgacgatt caaaagatta	2100
aagctggctt tgataaagcc atggacaaaa gcagcgcagg tttggttgct tttggtatca	2160
ccgttgtttt tgccgggtga tccgttgctg ccttcagca gttagtcggg attaacgccg	2220
tgctgtatta tgcaccgcag atgttccaga atttaggttt tggagctgat acggcattat	2280
tgcagaccat ctctatcggg gttgtgaact tcattctcac catgattgct tcccgtgttg	2340
ttgaccgctt cggccgtaaa cctctgctta tttgggggtgc tctcggtatg gctgcaatga	2400
tggctgtttt aggctgctgt ttctgggtca aagtcgggtg tgttttgcct ttggcttctg	2460
tgcttcttta tattgcagtc tttggtatgt catggggccc tgtctgctgg gttgttctgt	2520
cagaaatgtt cccgagttcc atcaagggcg cagctatgcc tategctgtt accggacaat	2580
ggttagctaa tatcttggtt aacttcctgt ttaagggtgc cgatggttct ccagcattga	2640
atcagacttt caaccacggg ttctcctatc tcgttttcgc agcattaagt atcttaggtg	2700
gcttgattgt tgctcgcttc gtgccggaaa ccaaaggctg gagcctggat gaaatcgagg	2760
agatgtggcg ctcccagaag tagttaaaact tgctttgggt gaatcctttt gtctttttta	2820
gataagtctt aaccaattat actttttgtt tacaacgatg gtataaagcg ggcggactta	2880
ttttacctgt tgggtagcct tctgatttca gaaaggaatt attatggaaa ttggtgcgat	2940
tgacatcggg ggaacgcag cgcgtttctc tattgcggaa gtaagcaatg gtcgggttct	3000
ttctcttgga gaagaaacaa cttttaaaac ggcagaacat gctagcttgc agttagcttg	3060
ggaacgtttc ggtgaaaaac tgggtcgtcc tctgccacgt gccgcagcta ttgcatgggc	3120

```

tggccccggtt catggtgaag ttttaaaact taccaataac ccttggggtat taagaccagc 3180
tactctgaat gaaaagctgg acatcgatac gcatgttctg atcaatgact tcggcgcggt 3240
tgcccacgcg gttgcgcata tggattcttc ttatctggat catatttggtg gtccctgatga 3300
agcgcttcct agcgatgggtg ttatcactat tcttgggtccg ggaacgggct tgggtgttgc 3360
ccatctgttg cggactgaag gccgttatatt cgtcatcgaa actgaaggcg gtcatatcga 3420
ctttgtcccg cttgacagac ttgaagacaa aattctggca cgtttacgtg aacgtttccg 3480
ccgcggtttct atcgaacgca ttattttctgg cccgggtctt ggtaatatct acgaagcact 3540
ggctgccatt gaaggcggtc cgttcagctt gctggatgat attaaattat ggagatggc 3600
tttggaagggt aaagacaacc ttgctgaagc cgctttggat cgcttctgct tgagccttgg 3660
cgctatcgct ggtgatcttg ctttggcaca gggtcgaacc agtggtgtta ttggcggtgg 3720
tgtcgggtctt cgtatcgctt cccatttgcc agaactctgtt ttccgtcagc gctttgtttc 3780
aaaaggacgc tttgaacgcg tcatgtccaa gattccggtt aagttgatta cttatccgca 3840
gcctggactg ttgggtgctg agctgcctat gccacaaat attctgaagt tgaataatat 3900
tttttaatat tatgaactga atttaagagg ctgccttccg ataaaatcgg gaggtggcct 3960
tttttatatt ttttactaaa aaatgaagac aaaaaagtct taagtaagaa taatattatt 4020
attaactttt gatataattt gtattagtgg atccgcctc cgcgtggaaa ttgaagccat 4080
cgcagtacgt agtgcgtaaa gcctcgtgag cgggacggtc gtaaggctgt tccgctccac 4140
ttcactgaac ggcaatccga ggggtgtggat atgattagt cttcgcata tttcaaaatt 4200
gggattggtc cctccagttc gcataccgtg gggccaatga atgccggaaa aagttttatt 4260
gatcggctgg aaagtagcgg cttattaacc gcgacgagcc atattgtgggt cgatctgtac 4320
gggtcgttgt cactgacggg caaaggccat gccacggatg tcgcatcat catgggactg 4380
gcaggaaaca gtccgcagga tgttgtcatt gatgagatcc ctgcatttat agagtttagta 4440
acgcgcagcg ggcggctgcc agtggcatct ggtgcgcata ttgttgattt tcctgtagca 4500
aagaacatta tcttccatcc cgaaatgttg cctcgccatg agaacggaat gcggatcact 4560
gcctggaagg gacaggaaga gctattaagt aaaac 4595

```

<210> 40

<211> 1069

<212> ADN

<213> vùng đoạn khuyết galP;

<400> 40

```

actttgggtcg tgaacatttc ccgtgggaaa aaaccgacaa agcgcagctg ctgcgcgatg 60
ctgccgggtct gaagtaatct ttcttcacct gcgttcaaag gccagcctcg cgctggcctt 120

```

```

tttcttttgg ataggcggttc acgccgcac cggcaaaaaa accgcccgc caataacatc 180
attcttctctg atcacgtttc accgcagatt atcatcacia ctgaaaccga ttacaccaac 240
cacaacagac aaagatttgt aatattttca tattattatt cggttttcac agttgttaca 300
tttcttttca gtaaagtctt aattgcagat aacagcggtt aatctatgat gatataactc 360
aattattttc atgcacttaa atcataacta agataaatgt tagtgtaagc gattacactg 420
atgtgatttg cttcacatct ttttacgtcg tactcaccta tcttaattca caataaaaaa 480
taaccatatt ggaggcgcatc atgcctgacg ctaaaaaaca ggggcggtca aacaaggcaa 540
tgacgtttga aataggcgct cagattaat ctccccaagc ttcctcccat cgcggaggaa 600
gccacctctt gcagtcacat tttcttcgt ctatcctctg ccgctatgaa aacatcccgt 660
ctccctatcg ccatccaaca ggccgttatg cgtcgctgc gggaaaaact cgcccaggcc 720
aacctgaagc tagggcgtaa ctaccggag ccaaaactct cttacacca gcgcggaacc 780
tccgccgga cggcctggct ggaaagctat gaaattcgcc tcaatcccgt tttgctgttg 840
gaaaacagtg aagcttttat tgaagaagt gtaccgcacg aactggcaca tttgctggta 900
tgaaacatt tcggccgcgt agcggcacat ggcaaagagt ggaagtggat gatggaaaac 960
gtgctgggtg ttcccgcccg tcgtacgcat cagttcgaac tgcaatccgt gcgtcgcaac 1020
accttcccct accgctgcaa gtgccaggag catcagctta ccgtacgcc 1069

```

<210> 41

<211> 6100

<212> ADN

<213> catxet p15-catAX +PR-aroY+P26-quiC tại locus adhE;

<400> 41

```

ttgattttca taggttaagc aaatcatcac cgcactgact atactctcgt attcgagcag 60
atgatttact aaaaaagttt aacattatca ggagagcatt agcttgctat tgacgacagc 120
tatggttcac tgtccaccaa ccaaaactgt gtcagtacc gccaatatct ctcccttgag 180
gggtacaaag aggtgtccct agaagagatc cagctgtgtg aaaaatttta caaaaaggta 240
ttgactttcc ctacaggggtg tgtaataatt taattacagg cgggggcaac cccgcctggt 300
ctgcagagga ggaatatagc catggaagtg aaaatcttca acaccagga tgttcaggat 360
tttctgcgtg ttgcaagcgg tctggaacaa gaggggtggt atccgcgtgt taaacaaatt 420
attcatcgtg ttctgagcga cctgtataaa gcaattgaag atctgaatat caccagcgac 480
gaatattggg caggcggtgc atatctgaat cagctgggtg caaatcaaga agcaggtctg 540
ctgagtcggg gtctgggttt tgatcattat ctggatatgc gtatggatgc agaagatgca 600
gcactgggta ttgaaaatgc aacaccgcgt accattgaag gtccgctgta tgttgccgggt 660

```

gcaccggaaa gcgttggtta tgcacgcatg gatgatggta gcgatccgaa tggtcatacc	720
ctgattctgc atggcaccat ttttgatgca gatggtaaac cgctgccgaa tgcaaaagtt	780
gaaatttggc atgcaaacac caaaggcttt tatagccatt ttgatccgac cgggtgaacag	840
caggccttta atatgcgtcg tagcattatt accgatgaga atggtcagta tcgtgttcgt	900
accattctgc ctgccggtta tggttgtcct ccggaaggtc cgaccagca actgctgaac	960
caactgggtc gtcatggtaa tcgtccggca catattcatt attttgtttag cgcagatggg	1020
caccgtaaac tgaccacca gattaatgtt gccggtgatc cgtataccta tgatgatttt	1080
gcatatgcca cccgtgaagg tctggttggt gatgcagttg aacataccga tccggaagca	1140
attaaagcca atgatgtgga aggtcctttt gccgaaatgg tgtttgatct gaaactgacc	1200
cgtctgggtg atgggtgtga taatcaggtt gtggatcgtc cgcgtctggc agtttaatac	1260
accaaaatgg ttcaaaatta tcaggcgagt gatcatgatc actggcctgt ttttatttca	1320
gggaaggggtg gagacaatta cgtggataat cagatcatcc aagaaaccgt ggataaaatt	1380
ctgagcggtc tgccgaatca ggcaggtcag ctggcacgtc tgggtcgtct gatgcaattt	1440
gcatgcgatc cgaccattac cgttattggc aaatataacc atggtaaaag ccgtctgctg	1500
aatgaactga ttggcaccga tatcttttagc gttgcagata aacgtgaaac cattcagctg	1560
gccgaacata aacaggatca ggttcggttg ctggatgcac ctgggtctgga tgccgatggt	1620
gcagcagttg atgatcgta tgcatttgaa gcagtttgga ccaggcaga tattcgtctg	1680
tttgttcata gcgttcgtga aggtgaactg gatgcaaccg aacaccatct gctgcaacag	1740
ctgattgaag atgccgatca tagccgtcgt cagaccattc tggttctgac ccagattgat	1800
cagattccgg atcagaccat cctgacacag attaaaacca gcattgcaca gcaggttccg	1860
aaactggata tttgggcagt tagcgcaacc cgtcatcgtc agggcattga aaacggtaaa	1920
accctgctga tcgaaaaaag cggatttggg gcactgcgcc ataccctgga acaggcactg	1980
gcacaggtgc cgagcgcacg tacctatgaa aaaaatcgtc tgctgtcaga tctgcaccat	2040
cagctgaaac aactgctgct ggatcagaaa catgttctgc aacaactgca acagacacag	2100
caacagcagc tgcattgatt tgataccggt ctgattaaca ttctggacaa aattcgtggt	2160
gatctggaac cgattgtgaa tattgatggg caggatcaag cactgaatcc ggatagcttt	2220
gcaaccatgt ttaaaaacac cgcagcaaaa cagcagcgtg ccaaagttca gattgcatat	2280
agccgtgcat gcattgaaat caacagccat ctgattcgcc atgggtgttg tggctctgcct	2340
gcggaacagc agaccaccat taaaagcatt gataccgtga ttggtgccgt gtttggtatc	2400
agcgttaaat ttcgtgatca gctgcgtgcc ctgttttata ccgataccga acgtcagcgt	2460
ctgcaacgtg aatttcgttt ctattttgaa aaaagtgccg gtcgcatgat tctggcagca	2520

aaaattgaac agaccatgcg tcagcagggc tgtattcaga atgccatgat ggcactgcaa	2580
caaatggaaa gcgcagcata aaaacacgga cgccgcaaac ggcgtccgaa tttcttggtc	2640
gaccgttaaa tctatcacccg caagggataa atatctaaca ccgtgcgtgt tgactatttt	2700
acctctggcg gtgataatgg ttgcatgtac taatctagat aaggaatata gccatgaccg	2760
caccgattca ggatctgcgt gatgcaattg ccctgctgca acagcatgat aatcagtatc	2820
tggaaccga tcatccggtt gatccgaatg cagaactggc aggcgtttat cgtcatattg	2880
gtgccggtgg caccgttaaa cgtccgaccc gtattggtcc ggcaatgatg ttttaataaca	2940
ttaaagggtta tccgcacagc cgtattctgg ttggtatgca tgcaagccgt cagcgtgcag	3000
cactgctgct gggttgtgaa gcaagtcagc tggcactgga agtttggtaaa gcagttaaaa	3060
aaccggttgc accggtggtt gttccggcaa gcagcgcacc gtgtcaagag cagatttttc	3120
tggcagatga tccggatttt gatctgcgta ccctgctgcc tgcacatacc aataccccga	3180
ttgatgcagg tccgtttttt tgtctgggtc tggccctggc aagcgatccg gtggatgcaa	3240
gcctgaccga tgttaccatt catcgtctgt gtgttcaggg tcgtgatgaa ctgagcatgt	3300
tcctggcagc aggtcgccat attgaagttt ttcgtcagaa agcagaagca gcaggtaaac	3360
cgctgccgat taccattaat atgggtctgg acccagcaat ctatattggc gcatgttttg	3420
aagcaccgac cccccgttt ggttataatg aactgggtgt tgccggtgca ctgcgtcagc	3480
gtccggttga actggttcag ggtgttagcg ttccggaaaa agcaattgca cgtgccgaaa	3540
ttgttattga aggtgaactg ctgcctggtg ttcgtgttcg tgaagatcag cataccaatt	3600
caggatcatgc aatgccggaa tttccgggtt attgtggtgg tgcaaatccg agcctgccgg	3660
ttattaaagt taaagccgtt accatgcgca ataacgcaat tctgcaaacc ctggttggtc	3720
cgggtgaaga acataccacc ctggcaggtc tgccgaccga agcaagcatt tggaatgcag	3780
ttgaagcagc aattccgggt tttctgcaaa atgtttatgc ccataccgca ggcggtggta	3840
aatttctggg tattctgcaa gtgaaaaaac gtcagcctgc cgatgaaggt cgtcagggtc	3900
aggcagccct gctggcgctg gcaacctata gcgaactgaa aaatatcatt ctggtggatg	3960
aggatgtgga cttttttgat agtgatgata ttctgtgggc aatgaccacc cgtatgcagg	4020
gtgatgttag cattaccacc attccgggtg ttcgcggtca tcagctggac ccgagccaga	4080
caccggaata ttcaccgagc attcgtggta atggtattag ctgcaaaaacc atctttgatt	4140
gtaccgttcc gtgggcactg aaaagccatt ttgaacgtgc accgtttgca gatgttgatc	4200
cgcgctccgtt tgcacctgaa ttttttgac gtctggaaaa aaatcagggc agcgcaaat	4260
aagctaataa caggcctgct ggtaatcgca ggaattttta tttggatgga tccgcctacc	4320

tagcttccaa gaaagatatc ctaacagcac aagagcggaa agatgttttg ttctacatcc	4380
agaacaacct ctgctaaaat tcctgaaaaa ttttgcaaaa agttgttgac tttatctaca	4440
aggtgtggta taataatctt aacaacagca ggacgctccc gggttgagga aaacctaattg	4500
aaactgacca gcctgcgtgt tagcctgctg gcactgggtc tggttaccag cggttttgca	4560
gcagcagaaa cctataccgt tgatcgttat caggatgata gcgaaaaagg tagcctgcgt	4620
tgggcaattg aacagagcaa tgcaaatagc gcacaagaaa accagattct gattcaggca	4680
gttggttaaag caccgtatgt tatcaaagtt gataaacccg tgccctccgat taaaagcagc	4740
gttaaaatca ttggcaccga gtgggataaa accggtgaat ttattgcaat tgatggcagc	4800
aactatatca aaggcgaagg tgaaaaagca tgtccgggtg caaatccggg tcagtatggc	4860
accaatgttc gtaccatgac cctgcctggt ctggttctgc aagatgttaa tgggtgttacc	4920
ctgaaaggtc tggatgttca tcgtttttgt attggtgttc tggttaatcg cagcagcaat	4980
aacctgattc agcataatcg tatcagcaac aattatgggt gtgccgggtg tatgattacc	5040
ggtgatgatg gtaaaggtaa tccgaccagc accaccacca ataataacaa agttctggat	5100
aacgtgttca tcgataatgg tgatggtctg gaactgacct gtggtgcagc atttaatctg	5160
attgcaaata acctgtttac cagcacaaaa gccaatccgg aaccgagcca gggatttgaa	5220
attctgtggg gtaatgataa tgccgtgggt ggtaacaaat tcgaaaacta ttcagatggc	5280
ctgcaaatac attggggtaa acgtaactat atcgccata acgaactgac caataacagc	5340
ctgggtttca atctgacagg tgatggtaac attttcgaca gcaataaagt gcatggtaac	5400
cgtattggta ttgccattcg tagtgaaaaa gatgccaatg cacgtattac cctgaccaa	5460
aatcagattt gggataacgg caaagatatc aaacgttgtg aagccggtgg tagctgtgtt	5520
ccgaatcagc gtctgggtgc aattgttttt ggtgttccgg cactggaaca tgaaggtttt	5580
gttggtagcc gtggcgggtg tgttgttatt gaaccggcaa aactgcaaaa aacctgcacc	5640
cagccgaacc agcagaattg taatgcaatt cctaatacagg gtattcaggc accgaaactg	5700
acagttagca aaaaacagct gaccgttgaa gttaaaggca cccctaatac gcgttataat	5760
gtggaatttt ttggcaatcg taatgccagc agcagcgaag cagaacagta tctgggtagc	5820
attgttggtt ttaccgatca tcagggtctg gcaaaagcaa attgggctcc gaaagttagc	5880
atgccgagcg ttaccgcaa tgtgacagat catctgggtg cgaccagcga actgagcagc	5940
gcagttaaaa tgcgttaa atgcgtgcgc cgcgttcgcg cggcgctttt ttttggtagc	6000
cagtagcgct gtctggcaac ataaacggcc ctttctgggc aatgccgatc agttaaggat	6060
tagttgaccg atccttaa ac tgaggcacta taacggcttc	6100