



(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)



1-0026612

(51)⁷ C07K 19/00; C07K 14/435; C12N
15/63; C12N 15/12; A01K 67/04

(13) B

(21) 1-2013-01306

(22) 28/09/2011

(86) PCT/US2011/053760 28/09/2011

(87) WO 2012/050919 19/04/2012

(30) 61/387,332 28/09/2010 US

(45) 25/12/2020 393

(43) 26/05/2014 314A

(73) THE UNIVERSITY OF NOTRE DAME (US)

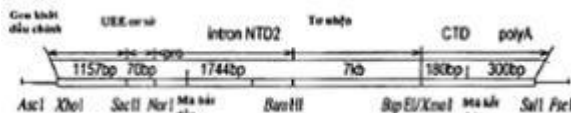
940 Grace Hall, Notre Dame, Indiana 46556, United States of America

(72) FRASER, Malcolm James (US); LEWIS, Randy (US); JARVIS, Don (US);
THOMPSON, Kimberly (US); HULL, Joseph (US); MIAO, Yun-Gen (CN); TEULE,
Florence (FR); SOHN, Bonghee (KR); KIM, Youngsoo (KR).

(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) AXIT NUCLEIC MÃ HÓA POLYPEPTIT TƠ NHỆN KHẢM, POLYPEPTIT TƠ NHỆN KHẢM, TỔ COMPOZIT CHỨA POLYPEPTIT TƠ NHỆN KHẢM VÀ PHƯƠNG PHÁP TẠO RA TẦM CHUYỂN GEN

(57) Sáng chế đề cập đến tầm chuyển gen được xử lý để có và biểu hiện gen protein tơ nhện khảm, gen protein tơ nhện khảm mã hoá protein tơ nhện khảm có các miền tơ nhện đặc trưng cho trình tự môtip đàn hồi của tơ nhện và/hoặc trình tự môtip bền của tơ nhện. Sáng chế cũng đề xuất tơ tầm được cải thiện có đặc tính về độ bền (độ bền kéo) và độ đàn hồi được tăng cường so với tơ tầm tự nhiên. Sáng chế này cũng đề xuất phương pháp cải tiến để tạo ra tơ khảm sử dụng tầm chuyển gen được bộc lộ trong bản mô tả này, phương pháp này sử dụng hệ vectơ dựa trên *piggyBac* và plasmit hỗ trợ. Catxet biểu hiện gen được đề xuất và được sử dụng để tạo nhiều trình tự mã hoá tơ nhện tổng hợp (nhện 2, nhện 4, nhện 6, nhện 8). Hệ thống vectơ *piggyBac* được sử dụng để biến nạp tầm đột biến, với sự có mặt của plasmit hỗ trợ, để hợp nhất gen tơ nhện khảm vào trong tầm để tạo ra thể biến nạp ổn định. Tầm chuyển gen này vì vậy tạo ra sinh vật tạo ra tơ nhện có hiệu quả thích hợp để sản xuất tơ thương mại.



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực sợi tơ, như sợi tơ nhện khảm có đặc tính về độ bền và độ dẻo được cải thiện. Ngoài ra, sáng chế đề cập đến lĩnh vực phương pháp sản xuất tơ khảm, như phương pháp sản xuất tơ được cải thiện (cụ thể, sợi tơ tằm/nhện khảm) sử dụng tằm chuyển gen được thao tác di truyền có các trình tự di truyền của tơ nhện đặc trưng (các trình tự môtip bền của tơ nhện và/hoặc dẻo của tơ nhện và/hoặc đàn hồi). Sáng chế cũng đề cập đến các sinh vật chuyển gen, như tằm chuyển gen được thao tác di truyền để chứa trình tự tằm khảm chứa các trình tự di truyền của tơ nhện đặc trưng cho các môtip dẻo và/hoặc đàn hồi của tơ nhện và các môtip bền của tơ nhện, và phương pháp tạo ra tằm chuyển gen này bằng cách sử dụng vector *piggyBac* được thiết kế đặc biệt. Sáng chế cũng đề xuất các phương pháp sản xuất thương mại tơ khảm bằng cách sử dụng tằm chuyển gen đã được mô tả.

Tình trạng kỹ thuật của sáng chế

Tơ đã được sử dụng trong nhiều năm làm chỉ khâu cho nhiều loại quy trình phẫu thuật quan trọng. Tơ nhỏ cần để làm chỉ khâu cho các phẫu thuật mắt, thần kinh và thẩm mỹ. Tơ cũng có triển vọng lớn làm vật liệu làm dây chằng nhân tạo, gân nhân tạo, băng đàn hồi dùng cho ghép da ở các bệnh nhân bị bỏng, và khung có thể cung cấp tác dụng đỡ và, trong một số trường hợp, chức năng tạm thời trong quá trình tái tạo các mô xương, mô bao quanh răng và mô liên kết. Sự phát triển của tơ làm vật liệu làm dây chằng và gân được dự kiến sẽ trở nên ngày càng quan trọng do tai nạn ở dây chằng hình chữ thập trước (anterior cruciate ligament: ACL) và các tổn thương khớp khác đòi hỏi khắc phục bằng phẫu thuật gia tăng trong nhóm dân số cao tuổi. Trong khi một tỷ lệ nhỏ tơ hiện thời được sử dụng làm chỉ khâu có nguồn gốc từ tơ tằm tự nhiên, hầu hết các tơ đều được sản xuất dưới dạng các polyme tổng hợp nhờ công nghiệp hoá học. Hạn chế chủ yếu của phương pháp này là ở chỗ nó chỉ có thể tạo ra tơ có phạm vi hẹp về các tính chất vật lý, như đường kính, độ bền và độ đàn hồi.

Nhiều hệ thống tái tổ hợp khác nhau gồm vi khuẩn (Lewis, et al. 1996), nấm men (Fahnestock ADN Bedzyk, 1997), tế bào côn trùng nhiễm virut baculo (Huemmerich, et

al. 2004), tế bào động vật có vú (Lazaris, et al. 2002) và thực vật chuyển gen (Scheller, et al. 2001) đã được sử dụng để tạo ra các protein tơ khác nhau. Tuy nhiên, không có hệ thống nào của các hệ thống này được thiết kế tự nhiên để kéo tơ và, vì vậy, không một hệ thống nào tạo ra một cách tin cậy tơ hữu dụng. Để tơ được xem là hữu dụng theo quan điểm thương mại, tơ phải có đủ các đặc tính chịu kéo (độ bền) và dẻo và/hoặc đàn hồi, và thích hợp để tạo thành tơ trong ứng dụng thương mại mong muốn. Vì vậy, vẫn cần hệ thống có thể dùng cho mục đích này.

Protein tơ nhện đã được sản xuất trong một số hệ thống sản xuất protein khác loại. Trong mỗi trường hợp, lượng protein được sản xuất thấp xa so với mức độ thương mại thực tế. Các hệ thống biểu hiện của thực vật và động vật chuyển gen có thể được tăng quy mô, nhưng ngay cả trong các hệ thống này, mức độ sản xuất protein tái tổ hợp cần phải được tăng đáng kể để hiệu quả về chi phí. Một vấn đề thậm chí còn khó khăn hơn là các nỗ lực sản xuất trước đây đã tạo được protein, nhưng không phải tơ. Vì vậy, protein phải được kéo thành tơ bằng cách sử dụng phương pháp sau sản xuất. Do các vấn đề về sản xuất và kéo sợi này, vẫn không có một ví dụ nào về hệ thống sản xuất protein tái tổ hợp có thể tạo ra tơ nhện đủ dài có ý nghĩa thương mại; nghĩa là, tơ “hữu dụng”.

Các cố gắng được báo cáo trước đây để sản xuất tơ sử dụng hệ thống tế bào động vật có vú để biểu hiện các gen mã hoá MaSp1, MaSp2 và các protein tơ có liên quan của nhện, *A. diadematus* (Lazaris, et al. 2002). Công trình này dẫn đến tạo ra được protein tơ nhện 60Kd, ADF-3, được làm sạch và được sử dụng để sản xuất tơ bằng phương pháp kéo sợi sau sản xuất. Tuy nhiên, hệ thống này không tạo được tơ hữu dụng một cách nhất quán. Hơn nữa, phương pháp này có vấn đề do cần hoà tan protein, phát triển điều kiện kéo sợi có kết quả và thực hiện kéo giãn sau kéo sợi để có được tơ có tính chất hữu dụng.

Lĩnh vực kỹ thuật này vẫn thiếu phương pháp thương mại để sản xuất nhất quán tơ có đặc tính chịu kéo và dẻo cần thiết để sử dụng trong sản xuất.

Bản chất kỹ thuật của sáng chế

Sáng chế khắc phục được các nhược điểm đã được mô tả trong lĩnh vực kỹ thuật này. Cụ thể, sáng chế đề xuất hệ thống sản xuất tằm chuyển gen có khả năng thích ứng với quy mô thương mại tránh được các vấn đề liên quan đến làm sạch, hoà tan và kéo sợi sau sản xuất nhân tạo protein, vì nó được trang bị tự nhiên để kéo tơ.

Theo hướng chung và tổng quát, sáng chế đề xuất phương pháp công nghệ sinh

học để sản xuất tơ nhện khảm bằng cách sử dụng tầm chuyển gen làm nền tảng để sản xuất protein tơ khác loại của tơ khảm hữu dụng thương mại có đặc tính chịu kéo và dẻo tốt hơn. Tơ khảm này có thể được thiết kế theo yêu cầu để tạo ra tơ có phạm vi tính chất vật lý mong muốn hoặc các tính chất định trước đặc trưng, được tối ưu hoá cho các ứng dụng y sinh mong muốn.

Protein tơ nhện/tầm và tơ nhện khảm:

Theo một khía cạnh, sáng chế đề xuất protein tơ nhện/tơ tầm khảm tái tổ hợp được mã hoá bởi trình tự chứa một hoặc nhiều trình tự motif/miền dẻo và/hoặc đàn hồi của tơ nhện và/hoặc một hoặc nhiều trình tự miền bền của tơ nhện. Theo một số phương án, protein tơ nhện/tơ tầm khảm còn được mô tả là mã hoá protein tơ nhện/tơ tầm khảm của nhện 2, nhện 4, nhện 6 hoặc nhện 8.

Hơn nữa, sáng chế đề xuất tơ nhện khảm được tạo ra từ các protein tơ tầm/nhện khảm. Theo các phương án cụ thể, tơ nhện khảm được mô tả là có độ bền kéo lớn hơn so với tơ tầm tự nhiên, và theo một số phương án, có độ bền kéo lớn hơn gấp 2 lần khi so với tơ tầm tự nhiên.

Tầm chuyển gen:

Theo phương án khác, sáng chế đề xuất các sinh vật chuyển gen, cụ thể là côn trùng tái tổ hợp và động vật chuyển gen. Theo một số phương án, sinh vật chuyển gen là tầm chuyển gen, như *Bombyx mori* chuyển gen. Theo các phương án cụ thể, tầm vật chủ được biến nạp để tạo ra tầm chuyển gen sẽ là tầm đột biến không có khả năng tạo ra tơ tự nhiên. Theo một số phương án, thể đột biến của tầm là pnd-w1.

Theo một số phương án, tầm đột biến (*B. mori*) sẽ được biến nạp bằng cách sử dụng hệ thống *piggyBac*, trong đó vectơ *piggyBac* được tạo ra bằng cách sử dụng catxet biểu hiện chứa trình tự protein tơ nhện tổng hợp được gắn bên cạnh bởi các mảnh đầu tận cùng N và C của protein fhc *B. mori*. Thông thường, sự biến đổi tầm bao gồm việc đưa hỗn hợp vectơ *piggyBac* và plasmid hỗ trợ, mã hoá *piggyBac* transposaza, vào trong phôi tiền phôi bì nhờ việc tiêm tế bào cho trứng tầm. Máy điều khiển kim tiêm của người máy Eppendorf đã được định chuẩn để chọc thủng màng đệm được sử dụng để tạo ra lỗ vi chèn mà xuyên qua đó mao quản thuỷ tinh được chèn vào và qua đó dung dịch ADN được tiêm vào trong trứng tầm. Trứng đã được tiêm sau đó được để cho trưởng thành, và phát triển nở thành ấu trùng. Ấu trùng được trưởng thành thành tầm, và kéo kén theo

vòng đời thông thường của tằm.

Sự lai giống các côn trùng chuyển gen này với nhau, hoặc với côn trùng/tằm không chuyển gen cũng được đề xuất là một phần của sáng chế.

Catxet biểu hiện di truyền của tơ nhện:

Theo phương án khác, sáng chế đề xuất các catxet biểu hiện tơ tằm/nhện khảm, catxet này chứa một hoặc nhiều môtip trình tự protein tơ nhện tương ứng với một hoặc nhiều trong số các trình tự môtip dẻo và/hoặc đàn hồi của tơ nhện cụ thể và/hoặc trình tự môtip bền của tơ nhện như được mô tả trong bản mô tả này. Theo khía cạnh khác, sáng chế đề xuất các phương pháp sản xuất protein tơ nhện/tằm khảm và tơ được đề xuất. Ít nhất tám (8) phiên bản khác nhau của catxet biểu hiện như được biểu thị trên Fig.5 được đề xuất, các catxet biểu hiện này mã hoá bốn protein tơ nhện tổng hợp khác nhau có hoặc không có EGFP được lồng trong khung giữa các trình tự tơ nhện abd NTD. Các trình tự này được nhận diện/xác định trong bản mô tả này là “nhện 2”, “nhện 4”, “nhện 6” và “nhện 8”.

Tằm chuyển gen:

Theo một khía cạnh khác nữa, sáng chế đề xuất tằm chuyển gen và các phương pháp sản xuất tằm chuyển gen. Theo một số phương án, phương pháp sản xuất tằm chuyển gen bao gồm: bước tạo ra catxet biểu hiện có trình tự chứa trình tự tằm, trình tự tơ nhện khảm mã hoá một hoặc nhiều trình tự môtip bền của tơ nhện và một hoặc nhiều trình tự môtip dẻo và/hoặc đàn hồi của tơ nhện, bước tách dòng phụ trình tự catxet nêu trên vào trong vector *piggyBac* (như vector *piggyBac* pBac[3xP3-DsRedaf], xem Fig.6, xem các Fig.10-11 về các plasmit cha mẹ, xem các Fig.12A-12E về các plasmit được tách dòng phụ từ các plasmit cha mẹ, bước đưa hỗn hợp của vector *piggyBac* và plasmit hỗ trợ mã hoá *piggyBac* transposaza, vào trong phôi tằm tiền phôi bì (ví dụ, nhờ việc tiêm tế bào cho trứng tằm), bước duy trì phôi tằm đã được tiêm trong điều kiện nuôi bình thường (nhiệt độ khoảng 28°C và độ ẩm 70%) cho đến khi ấu trùng nở và thu được tằm chuyển gen.

Tằm chuyển gen này có thể còn được cho giao phối để tạo ra phôi thế hệ F1 dùng để nhận dạng tiếp theo các thể biến nạp giả định, dựa vào sự biểu hiện của gen đánh dấu mắt đỏ S. Các thể biến nạp được và cái giả định được nhận dạng bằng phương pháp này sau đó được cho giao phối để tạo ra các dòng đồng hợp tử dùng để phân tích di truyền chi

tiết hơn. Cụ thể, sự biến nạp của tằm bao gồm việc tiêm hỗn hợp của vector *piggyBac* và ADN plasmit hỗ trợ vào trong trứng tằm của thể đột biến của tằm có biểu bì trong suốt, pnd-w1. Thể đột biến của tằm, pnd-w1, được mô tả trong Tamura, et al. 2000, tài liệu này được đưa vào bản mô tả này bằng cách viện dẫn. Thể đột biến này có sự thiếu hụt melanin hoá làm cho việc sàng lọc bằng cách sử dụng các gen huỳnh quang dễ dàng hơn. Khi các thể biến nạp F1 giả định có mắt đỏ được nhận dạng, các dòng đồng hợp tử được khẳng định bằng cách sử dụng thẩm tách Western các protein tuyến tơ và tơ kén đã được thu hoạch.

Phương pháp sản xuất tơ nhện/tơ tằm khám:

Theo một khía cạnh khác nữa, sáng chế đề xuất phương pháp sản xuất thương mại để tạo ra tơ nhện/tơ tằm khám trong tằm chuyển gen. Theo một phương án, phương pháp này bao gồm bước sản xuất tằm chuyển gen được mô tả trong bản mô tả này, và nuôi tằm chuyển gen trong điều kiện để tằm trưởng thành và tạo kén, thu hoạch kén, và thu được tơ nhện khám từ kén. Các kỹ thuật tiêu chuẩn để tháo tơ và/hoặc nói cách khác thu hoạch tơ từ kén tơ có thể được sử dụng.

Vật phẩm và phương pháp sử dụng vật phẩm này:

Theo các khía cạnh khác nữa, sáng chế đề xuất nhiều vật phẩm được làm từ tơ nhện khám theo sáng chế. Ví dụ, tơ nhện/tằm khám tái tổ hợp có thể được sử dụng làm vật liệu làm chỉ khâu y tế, các vật liệu băng vết thương và các vật liệu và thiết bị thay thế và phục hồi lại mô/khớp, miếng dán phân phối thuốc và/hoặc các hệ phân phối thuốc khác, quần áo bảo vệ (áo gi lê chống đạn và các vật phẩm khác), vật phẩm giải trí (lều, dù, phụ tùng cắm trại, v.v.), trong số các vật phẩm khác nữa.

Theo một khía cạnh khác, sáng chế đề xuất phương pháp sử dụng tơ nhện/tơ tằm khám tái tổ hợp trong các quy trình y tế khác nhau. Ví dụ, tơ có thể được dùng để tạo thuận lợi cho việc phục hồi mô, trong quá trình sinh trưởng hoặc tái sinh như khung trong cấu trúc tương hợp sinh học được thao tác di truyền của mô được tạo ra bằng tơ tái tổ hợp, hoặc để cung cấp protein hoặc chất chữa bệnh đã được thao tác di truyền vào trong tơ.

Trừ khi có quy định khác, tất cả các thuật ngữ kỹ thuật và khoa học được sử dụng trong bản mô tả này đều có ý nghĩa giống như thường được hiểu bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật. Mặc dù các phương pháp và các chất tương tự hoặc

tương đương với các phương pháp và các chất được mô tả trong bản mô tả này có thể được sử dụng trong việc thực hành hoặc thử nghiệm sáng chế, các phương pháp và các chất ưu tiên được mô tả dưới đây. Tất cả các tài liệu công bố, đơn yêu cầu cấp patent, patent và các tài liệu khác được nêu trong bản mô tả này được đưa vào bản mô tả này bằng cách viện dẫn. Ngoài ra, các chất, các phương pháp và các ví dụ này chỉ có tính chất minh họa và không nhằm để giới hạn. Trong trường hợp có sự mâu thuẫn, bản mô tả này, bao gồm các định nghĩa, sẽ có tính quyết định.

Mô tả vắn tắt các hình vẽ

Các mục đích và ưu điểm khác của sáng chế sẽ trở nên rõ ràng đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật này dựa vào việc đọc phần mô tả chi tiết các phương án ưu tiên dưới đây, dựa vào các hình vẽ đi kèm, trong đó các số chỉ dẫn giống nhau được sử dụng để ký hiệu các phần tử giống nhau, và trong đó:

Fig.1 thể hiện các trình tự axit amin (SEQ ID NO 18-23, lần lượt, theo thứ tự xuất hiện) của hai protein tơ ampullate lớn của các con nhện dệt tròn phân kỳ hoặc dệt tròn dẫn xuất (Gatesy, et al. 2001). Kết quả so sánh cho thấy mức độ bảo toàn trình tự cao, cụ thể là trong các trình tự môtip được mô tả ở trên, đã được duy trì trên 125 triệu năm kể từ khi loài này phân kỳ từ một loài khác. Các trình tự axit amin lặp liên ứng của protein tơ ampullate lớn ở loài dệt tròn khác nhau (-) cho biết axit amin không có mặt khi so với các trình tự khác. Các loài nhện là: *Nep.c.*, *Nephila clavipes*; *Lat.g.*, *Lactrodectus geometricus*; *Arg.t.*, *Argiope trifasciata*.

Fig.2 thể hiện các trình tự axit amin liên ứng (SEQ ID NO 24-26, lần lượt, theo thứ tự xuất hiện) của các protein tơ ampullate nhỏ của nhện dệt tròn. Ngay sau khi các trình tự protein tơ ampullate lớn đầu tiên được công bố, các cADN là các sản phẩm phiên mã protein tơ ampullate nhỏ (Mi) của *N. clavipes* được phân lập và được giải trình tự (Colgin ADN Lewis, 1998). Trình tự MiSp được cung cấp trên hình vẽ này có cả trình tự tương tự và khác nhau rõ ràng so với các protein MaSp. MiSp chứa các trình tự GGX và polyAla ngắn, nhưng các môtip polyAla dài hơn trong các MaSp được thay thế bằng các đoạn lặp (GA)_n. Các đoạn lặp liên ứng có cấu tạo tương tự nhưng số đoạn lặp GGX và GA khác nhau.

Fig.3 thể hiện các trình tự liên ứng cADN của protein tơ hình roi (SEQ ID NO 27-29, lần lượt, theo thứ tự xuất hiện). Các cADN của protein tơ này mã hoá protein tơ dạng

xoắn móc của tuyến hình roi *N. clavipes* (Fig.3; Hayashi and Lewis, 2000). Các cADN này chứa các trình tự mã hoá vùng không dịch mã 5' và peptit tín hiệu tiết, một số lớn đoạn lặp của năm môtip axit amin, và đầu tận cùng C. Phân tích thẩm tách Northern cho kích thước mARN xấp xỉ 15kb, mã hoá protein khoảng 500Kd. Trình tự axit amin được dự đoán từ trình tự gen gợi ý mô hình cấu trúc protein giúp giải thích cơ sở vật lý cho độ đàn hồi của tơ nhện, mà cũng phù hợp với tính chất của MaSp2 (được mô tả thêm trong bản mô tả này).

Fig.4 thể hiện mô hình trên máy tính của vòng xoắn β . Đây là mô hình của trình tự (SEQ ID NO: 1) (GPGGQGPGGY)₂ được tối thiểu hoá về năng lượng, có cấu hình bắt đầu là vòng β loại II ở mỗi trình tự năm mắt xích.

Fig.5 thể hiện một số thay đổi trên catxet biểu hiện chuỗi nặng fibron của tơ *Bombyx mori* cơ bản đã được xây dựng. Thiết kế này bao gồm tổ hợp các cấu trúc được thiết kế để biểu hiện thể khảm của chuỗi nặng fibroin (fibroin heavy chain: fhc)-tơ nhện, trong đó trình tự protein tơ nhện tổng hợp được gắn bên sườn bởi các mảnh đầu tận cùng N và C của protein fhc *B.mori*. Các yếu tố di truyền có liên quan về chức năng trong mỗi catxet biểu hiện, từ trái sang phải, bao gồm: trình tự khởi đầu chính, yếu tố tăng cường ngược dòng (upstream enhancer element: UEE), trình tự khởi đầu cơ bản, và miền đầu tận cùng N (N-terminal domain: NTD) của gen fhc *B. mori*, tiếp theo là các trình tự protein tơ nhện tổng hợp khác nhau nằm trong khung với vị trí bắt đầu dịch mã đặt ngược dòng trong NTD, tiếp theo là miền đầu tận cùng C fhc (C-terminal domain: CTD), gồm các vị trí kết thúc dịch mã và polyadenyl hóa ARN.

Fig.6 thể hiện sơ đồ tách dòng phụ các catxet vào trong *piggyBac*. Mỗi phiên bản trong số tám phiên bản khác nhau của catxet biểu hiện đã được vẽ được cắt khỏi plasmit mẹ bằng cách sử dụng AscI và FseI và được phân lập dòng vào trong các vị trí tương ứng của pBAC[3xP3-DSRedaf]. Bản đồ của vectơ *piggyBac* này được thể hiện.

Fig.7 thể hiện thẩm tách Western của tơ tằm chuyển gen. Tơ này được phân tích về sự có mặt của protein khảm của tơ nhện nhờ thẩm tách Western cả thành phần protein tuyến tơ tằm lẫn tơ của kén tằm chuyển gen bằng cách sử dụng kháng thể đặc trưng của tơ nhện. Trong cả hai trường hợp, tằm chuyển gen được kiểm tra khi sản xuất protein khảm, và các nghiên cứu chiết xuất vi phân cho thấy các protein này là các thành phần không thể thiếu của tơ chuyển gen của kén. Hơn nữa, sự biểu hiện của mỗi thể dung hợp

protein huỳnh quang màu lục khảm là rõ ràng trong cả tuyến tơ lẫn tơ nhờ việc kiểm tra trực tiếp tuyến tơ hoặc tơ bằng cách sử dụng kính hiển vi giải phẫu huỳnh quang. Trong hầu hết các trường hợp, lượng protein huỳnh quang trong tơ đủ cao để có thể nhìn thấy bằng mắt thường nhờ màu lục của kén dưới ánh sáng bình thường.

Fig.8 thể hiện plasmid mẹ pSL-nhện#4, kích thước 17.388bp. Plasmid mẹ này mang catxet protein tơ nhện khảm #4, tơ nhện (A4S8)x42.

Fig.9 thể hiện plasmid mẹ pSL-nhện#4 + GFP. GFP là protein huỳnh quang màu lục (green fluorescent protein: GFP). Vector này có kích thước 18.102bp. Plasmid mẹ này mang protein tơ nhện khảm #4 có protein đánh dấu, GFP, catxet, tơ nhện (A4S8)x42.

Fig.10 thể hiện plasmid mẹ pSL-nhện#6. Plasmid mẹ này có kích thước 12.516bp. Plasmid mẹ này mang catxet protein tơ nhện khảm #6, tơ nhện (A2S8)x14)x42.

Fig.11 thể hiện plasmid mẹ pSL-nhện#6 + GFP. GFP là protein huỳnh quang màu lục. Plasmid cha này có kích thước 13.230bp. Plasmid mẹ này mang protein tơ nhện khảm #6 có protein đánh dấu, GFP, catxet, tơ nhện (A2S8)x14.

Fig.12A - Fig.12B thể hiện các plasmid *piggyBac*. Fig.12A biểu thị cấu trúc pXLBacII-ECFP NTD CTD maspX16 có kích thước 10.458bp. Fig.12B biểu thị cấu trúc pXLBacII-ECFP NTD CTD maspX24, và có kích thước 11.250bp.

Fig.13 thể hiện trình tự đối với pSL-nhện#4 (SEQ ID NO: 30).

Fig.14 thể hiện trình tự đối với pSL-nhện#4+GFP (SEQ ID NO: 31)

Fig.15 thể hiện trình tự đối với pSL-nhện#6 (SEQ ID NO: 32).

Fig.16 thể hiện trình tự đối với pSL-nhện#6+GFP (SEQ ID NO: 33).

Fig.17 thể hiện các thiết kế vector *piggyBac*. Fig.17A. Gen tơ nhện tổng hợp A2S8₁₄; Fig.17B. Gen tơ tằm/nhện khảm của nhện 6; Fig.17C. Gen tơ tằm/nhện khảm của tơ nhện 6-GFP; Fig.17D. Các vector *piggyBac*; Fig.17E. Các ký hiệu dùng cho: Môtip đàn hồi hình roi (A2; 120bp); Spidroin ampullate lớn-2; trình tự khởi đầu chính fhc (1.157bp) môtip nhện (S8; 55bp), trình tự tăng cường fhc (70bp); trình tự khởi đầu cơ bản fhc, miền dịch mã 5' Hhc (Exon 1/intron/Exon 2; cds đầu tận cùng N fhc) = 1.744bp; EGF (720bp); trình tự tơ nhện A2SB₁₄. (2.462bp), cds đầu tận cùng C Fhc (180bp), tín hiệu polyadenyl hóa Fhc (300bp).

Fig.18 thể hiện sự biểu hiện của protein tầm/tơ nhện/EGFP khảm trong kén (18A), tuyến tơ (18B, 18C) và tơ (18D) của tầm nhện 6-GFP. Sự biểu hiện và định vị của protein tầm/tơ nhện khảm trong tuyến tơ tầm. Các tuyến tơ được cắt ra, bị bắn phá bằng các vectơ nhện 6 hoặc nhện 6-GFP *piggyBac*, và được kiểm tra dưới kính hiển vi huỳnh quang, như được mô tả ở phần Phương pháp.

Fig.19 thể hiện sự chiết tơ tuần tự. Kén được tạo ra bởi tầm pnd-w1 (các lần 3-6), nhện 6 (các lần 8-11), hoặc nhện 6-GFP (các lần 13-16) được khử keo và được cho qua quy trình chiết tuần tự, như được mô tả trong bản mô tả này. Protein được hoà tan trong từng bước chiết được phân tích bằng Độ lệch chuẩnS-PAGE và nhuộm xanh Coomassie (19A) hoặc thẩm tách miễn dịch (19B) bằng kháng huyết thanh đặc hiệu protein tơ nhện. M: Gen đánh dấu khối lượng phân tử. +: protein tơ nhện A2S814 được biểu hiện và được làm sạch trong *E. coli*. Các lần 3, 8 và 13: các phần chiết bằng nước muối. Các lần 4, 9 và 14: các phần chiết bằng Độ lệch chuẩnS. Các lần 5, 10 và 15: các phần chiết bằng LiSCN 8M/2% mercaptoetanol. Các lần 6, 11 và 16: các phần chiết bằng LiSCN 16M/5% mercaptoetanol. Các mũi tên đánh dấu protein tơ nhện khảm. Khối lượng phân tử biểu kiến ~75kDa đối với A2S814 của *E. coli*, ~106kDa đối với nhện 6, và ~130kDa và ~110kDa đối với nhện 6-GFP.

Fig.20 thể hiện so sánh các đặc tính cơ học tốt nhất được quan sát thấy đối với tơ composit của tầm chuyển gen, tơ tự nhiên của tầm cha mẹ và tơ nhện tự nhiên (tơ kéo) điển hình. Độ bền của tơ được xác định bởi diện tích dưới các đường cong ứng suất/biến dạng. Các tính chất cơ học của tơ tự nhiên và tơ composit đã được khử keo. Các đặc tính cơ học tốt nhất được xác định đối với tơ tầm tự nhiên (pnd-w1) và nhện điển hình (tơ kéo *N. clavipes*) được so sánh với các đặc tính cơ học thu được với tơ composit được tạo ra bởi tầm chuyển gen. Tất cả các tơ đều được thử nghiệm trong cùng một điều kiện. Các giá trị bền nhất là: nhện 6 dòng 7 (86,3MJ/m³); nhện 6-GFP dòng 1 (98,2MJ/m³), nhện 6-GFP dòng 4 (167,2MJ/m³); và tơ kéo *N. clavipes* (138,7MJ/m³), khi so với tầm tự nhiên pnd-w1 (43,9MJ/m³). Các số liệu này cho thấy tất cả các tơ composit của tầm chuyển gen đều bền hơn tơ tự nhiên của tầm không chuyển gen.

Fig.21 biểu thị trình tự axit nucleic của cấu trúc pXLBacII-ECFP NTD CTD masp1X16 (10.458pb) (SEQ ID NO: 34).

Fig.22 biểu thị trình tự axit nucleic của cấu trúc pXLBacII-ECFP NTD CTD maspX24 (11.250bp) (SEQ ID NO: 35).

Mô tả chi tiết sáng chế

Phương pháp chèn gen vào trong nhiễm sắc thể của tằm được sử dụng theo sáng chế phải cho phép gen này được đưa vào và được biểu hiện một cách ổn định trong nhiễm sắc thể, và được nhân giống ổn định đến các thế hệ con cháu, và cả, nhờ giao phối. Mặc dù phương pháp sử dụng tiêm tế bào vào trong trứng tằm hoặc phương pháp sử dụng súng bắn gen có thể được sử dụng, một phương pháp được ưu tiên sử dụng gồm tiêm tế bào vào trong trứng tằm bằng vectơ chứa gen đích để chèn gen ngoại sinh vào trong nhiễm sắc thể của tằm và plasmit hỗ trợ chứa gen nhảy (Nature Biotechnology 18, 81-84, 2000) một cách đồng thời.

Gen đích được chèn vào trong tế bào sinh sản của tằm tái tổ hợp đã nở và trưởng thành từ trứng tằm đã được tiêm tế bào. Con cháu của tằm tái tổ hợp thu được theo cách này có thể duy trì ổn định gen đích trong nhiễm sắc thể của chúng. Gen của tằm tái tổ hợp thu được theo sáng chế có thể được duy trì theo cách giống như tằm thông thường. Cụ thể là, tằm đến tuổi thứ năm có thể được nuôi nhờ ấp trứng trong điều kiện bình thường, thu ấu trùng đã nở để cho ăn nhân tạo và sau đó nuôi chúng trong điều kiện giống như tằm thông thường.

Tằm tái tổ hợp thu được theo sáng chế có thể được nuôi theo cách giống như tằm thông thường, và nó có thể tạo ra protein ngoại sinh nhờ nuôi trong điều kiện thông thường, để tối đa hoá sự phát triển và trưởng thành của tằm.

Tằm tái tổ hợp gen thu được theo sáng chế có thể hoá nhộng và tạo ra kén theo cách giống như tằm thông thường. Tằm đực và tằm cái được phân biệt trong giai đoạn nhộng, và sau khi đã biến đổi thành ngài, con đực và con cái giao phối và trứng được thu lượm vào ngày hôm sau. Trứng có thể được bảo quản theo cách giống như trứng tằm thông thường. Tằm tái tổ hợp gen theo sáng chế có thể được duy trì trong các thế hệ tiếp theo nhờ lặp lại sự nhân giống như được mô tả ở trên, và có thể tăng đến một số lượng lớn.

Mặc dù không có giới hạn cụ thể nào đối với trình tự khởi đầu được sử dụng trong bản mô tả này, và trình tự khởi đầu bất kỳ có nguồn gốc trong sinh vật bất kỳ có thể được sử dụng miễn là nó tác động có hiệu quả lên các tế bào của tằm, trình tự khởi đầu được

thiết kế để kích thích riêng protein của tuyến tơ tằm được ưu tiên. Các ví dụ về trình tự khởi đầu của protein tuyến tơ tằm bao gồm trình tự khởi đầu của chuỗi fibroin H, trình tự khởi đầu chuỗi fibroin L, trình tự khởi đầu p25 và trình tự khởi đầu serixin.

Theo sáng chế, "catxet gen để biểu hiện protein tơ nhện khảm" là bộ ADN cần để tổng hợp protein khảm trong trường hợp được chèn vào trong tế bào côn trùng. Catxet gen để biểu hiện protein tơ nhện khảm này chứa trình tự khởi đầu xúc tiến sự biểu hiện của gen mã hoá protein tơ nhện khảm. Thường thì, catxet gen này còn chứa trình tự kết thúc và vùng bổ sung poly A, và tốt hơn là chứa trình tự khởi đầu, gen cấu trúc protein ngoại sinh, trình tự kết thúc và vùng bổ sung poly A. Hơn nữa, nó cũng có thể chứa gen tín hiệu tiết ghép đôi giữa trình tự khởi đầu và gen cấu trúc protein ngoại sinh. Trình tự gen tùy ý cũng có thể được ghép đôi giữa trình tự bổ sung poly A và gen cấu trúc protein ngoại sinh. Hơn nữa, trình tự gen được thiết kế và được tổng hợp nhân tạo cũng có thể được ghép đôi.

Ngoài ra, "catxet gen để chèn gen tơ nhện/tằm khảm" là catxet gen để biểu hiện gen tơ nhện/tằm khảm có trình tự lặp đảo của cặp gen nhảy *piggyBac* ở cả hai bên, và gồm bộ ADN được chèn vào trong nhiễm sắc thể của tế bào côn trùng thông qua tác động của các gen nhảy *piggyBac*.

Vector theo sáng chế là vector có cấu trúc ADN vòng hoặc thẳng. Vector có khả năng sao chép trong *E. Coli* và có cấu trúc ADN vòng được đặc biệt ưu tiên. Vector này cũng có thể hợp nhất với gen đánh dấu như gen kháng kháng sinh hoặc gen protein huỳnh quang màu lục của sửa dùng cho mục đích giúp lựa chọn các thể biến nạp.

Mặc dù không có giới hạn nào đối với các tế bào côn trùng được sử dụng theo sáng chế, tốt hơn là chúng là tế bào côn trùng cánh vảy, tốt hơn nữa là tế bào *Bombyx mori*, và còn tốt hơn nữa là tế bào tuyến tơ tằm hoặc tế bào nằm trong trứng *Bombyx mori*. Trong trường hợp tế bào tuyến tơ, các tế bào tuyến tơ sau của ấu trùng tằm tuổi thứ năm được ưu tiên bởi vì có sự tổng hợp protein của fibroin mạnh và chúng được điều khiển dễ dàng.

Không có sự giới hạn cụ thể nào đối với phương pháp dùng để hợp nhất catxet gen để biểu hiện protein tơ nhện khảm nhờ các tế bào côn trùng. Các phương pháp sử dụng súng bắn gen và các phương pháp sử dụng tiêm tế bào có thể được sử dụng để hợp nhất vào trong tế bào côn trùng được nuôi, trong trường hợp hợp nhất vào trong tế bào tuyến tơ tằm, ví dụ, gen có thể được hợp nhất dễ dàng vào trong mô tuyến tơ sau được phân lập

khởi thân của ấu trùng tằm tuổi thứ năm bằng cách sử dụng súng bắn gen.

Hợp nhất gen vào trong tuyến tơ sau bằng cách sử dụng súng bắn gen có thể được thực hiện nhờ, ví dụ, việc bắn phá các hạt vàng được phủ bằng vector chứa catxet gen để biểu hiện protein ngoại sinh vào trong tuyến tơ sau được cố định trên đĩa thạch và sau đó bằng cách sử dụng súng bắn hạt (Bio-Rad, Model No. PDS-1000/He) với áp suất khí He từ 1.100 đến 1.800psi.

Trong trường hợp hợp nhất gen vào trong tế bào nằm trong trứng của *Bombyx mori*, phương pháp sử dụng tiêm tế bào được ưu tiên. Trong bản mô tả này, trong trường hợp thực hiện tiêm tế bào vào trong trứng, không nhất thiết tiêm tế bào trực tiếp vào trong tế bào của trứng, mà tốt hơn là gen có thể được hợp nhất một cách đơn giản nhờ tiêm tế bào vào trong trứng.

Tầm tái tổ hợp chứa "catxet gen để biểu hiện protein tơ nhện khảm" theo sáng chế trong các nhiễm sắc thể của nó có thể thu được nhờ tiêm tế bào vector có "hộp để chèn gen tơ nhện khảm" vào trong trứng của *Bombyx mori*. Ví dụ, tầm thế hệ thứ nhất (G1) thu được nhờ đồng thời tiêm tế bào vector có "catxet gen để lồng gen tơ nhện khảm" và plasmit trong đó gen transposaza *piggyBac* được bố trí dưới sự kiểm soát của trình tự khởi đầu actin của tầm vào trong trứng *Bombyx mori* theo phương pháp của Tamara, et al. (Nature Biotechnology 18, 81-84, 2000), tiếp theo là việc nhân giống ấu trùng đã nở và việc lai các côn trùng trưởng thành (G0) thu được trong cùng một nhóm. Tầm tái tổ hợp thường xuất hiện với tần suất từ 1 đến 2% trong thế hệ G1 này.

Việc lựa chọn tầm tái tổ hợp có thể được thực hiện nhờ PCR bằng cách sử dụng các đoạn mồi được thiết kế dựa vào trình tự gen protein ngoại sinh sau khi tách ADN của mô tầm thế hệ G1. Theo cách khác, tầm tái tổ hợp có thể được lựa chọn dễ dàng nhờ việc lồng gen mã hoá protein huỳnh quang màu lục được ghép đôi ở phía sau trình tự khởi đầu có khả năng được biểu hiện trong tế bào của tầm vào trong "catxet gen để chèn gen" trước, và sau đó lựa chọn các cá thể phát ra huỳnh quang màu lục dưới tia cực tím trong tầm thế hệ G1 trong giai đoạn tuổi thứ nhất.

Ngoài ra, trong trường hợp tiêm tế bào vector có "catxet gen để chèn gen" vào trong trứng *Bombyx mori* cho mục đích thu được tầm tái tổ hợp chứa "catxet gen để biểu hiện protein ngoại sinh" trong các nhiễm sắc thể của chúng, tầm tái tổ hợp có thể thu được theo cách giống như được mô tả ở trên nhờ việc đồng thời tiêm tế bào protein transposaza

piggyBac.

Gen nhảy *piggyBac* là yếu tố chuyển ADN có các trình tự đảo có 13 cặp bazơ ở cả hai đầu và ORF bên trong có khoảng 2,1k cặp bazơ. Mặc dù không có giới hạn nào đối với gen nhảy *piggyBac* được sử dụng theo sáng chế, các ví dụ về các gen nhảy có thể được sử dụng gồm các gen nhảy có nguồn gốc trong dòng tế bào *Trichoplusia ni* TN-368, *Autographa californica* NPV (AcNPV) và *Galleria mellonea* NPV (GmMNPV). Gen nhảy *piggyBac* có hoạt tính chuyển gen và ADN tốt hơn nếu có thể được điều chế bằng cách sử dụng các plasmit pHA3PIG và pPIGA3GFP có phần *piggyBac* có nguồn gốc trong dòng tế bào *Trichoplusia ni* TN-368 (Nature Biotechnology 18, 81-84, 2000). Cấu trúc của trình tự ADN có nguồn gốc trong *piggyBac* cần phải có một cặp trình tự kết thúc đảo chứa trình tự TTAA, và có gen ngoại sinh như gen xytokin được chèn giữa các trình tự ADN này. Tốt hơn nữa là sử dụng transposaza để chèn gen ngoại sinh vào trong nhiễm sắc thể của tằm bằng cách sử dụng trình tự ADN có nguồn gốc trong gen nhảy. Ví dụ, tần số mà ở đó gen được chèn vào trong nhiễm sắc thể của tằm có thể được cải thiện đáng kể nhờ lồng đồng thời ADN có khả năng biểu hiện transposaza *piggyBac* cho phép transposaza được sao chép và được dịch mã trong tế bào của tằm để nhận biết hai cặp trình tự kết thúc đảo, bị cắt khỏi mảnh gen giữa chúng, và chuyển nó vào nhiễm sắc thể của tằm.

Sáng chế có thể còn được hiểu rõ đầy đủ hơn nhờ sự mô tả dưới đây.

Protein tơ khảm trong lĩnh vực y sinh:

Tơ nhện khảm được cung cấp làm một phần vật liệu được sử dụng rộng rãi cho nhiều quy trình, như phẫu thuật mắt, phục hồi thần kinh và phẫu thuật tạo hình đòi hỏi tơ cực mảnh. Các ứng dụng phụ bao gồm vật liệu làm khung để tái tạo xương, dây chằng và gân cũng như vật liệu cung cấp thuốc.

Tơ nhện tái tổ hợp được tạo ra bởi các quy trình theo sáng chế có thể được sử dụng trong hàng loạt các ứng dụng y tế như các hệ thống băng bó vết thương, bao gồm các thiết bị phục hồi vết thương mạch máu, băng bó cầm máu, miếng dán và keo, chỉ khâu, cung cấp thuốc và trong các ứng dụng thao tác di truyền của mô, như, ví dụ, khung, thiết bị lắp dây chằng và các sản phẩm để cấy lâu dài hoặc có khả năng phân hủy sinh học vào trong cơ thể người. Khung được thao tác di truyền của mô được ưu tiên là lưới không dệt từ tơ được điều chế bằng tơ nhện/tằm tái tổ hợp được mô tả trong bản mô tả này.

Hơn nữa, tơ khảm tái tổ hợp theo sáng chế có thể được sử dụng cho các phương pháp phục hồi, thay thế hoặc tái tạo cơ quan mà có thể thu lợi từ các khung duy nhất này, bao gồm, nhưng không giới hạn ở, đĩa cột sống, mô sọ, màng cứng, mô thần kinh, gan, tuyến tụy, thận, bàng quang, lá lách, cơ tim, cơ xương, gân, dây chằng và mô ngực.

Theo một phương án khác của sáng chế, các vật liệu từ tơ nhện tái tổ hợp có thể chứa chất chữa bệnh. Để tạo ra các vật liệu này, chất chữa bệnh có thể được thiết kế di truyền vào trong tơ trước khi tạo ra vật liệu hoặc được đưa vào trong vật liệu sau khi vật liệu được tạo thành. Nhiều chất chữa bệnh khác nhau có thể được sử dụng kết hợp với tơ khảm tái tổ hợp theo sáng chế. Nói chung, chất chữa bệnh có thể được sử dụng thông qua các dược phẩm theo sáng chế bao gồm, nhưng không giới hạn ở: chất chống lây nhiễm như chất kháng sinh và chất chống virus; chất hóa trị liệu (nghĩa là, chất chống ung thư); chất chống đào thải; chất giảm đau và hỗn hợp chất giảm đau; chất chống viêm; hormon như steroid; yếu tố sinh trưởng (protein tạo hình xương (nghĩa là, bone morphogenic protein: BMP's 1-7), protein giống tạo hình xương (nghĩa là, GFD-5, GFD-7 và GFD-8), yếu tố sinh trưởng biểu bì (epidermal growth factor: EGF), yếu tố sinh trưởng nguyên bào sợi (nghĩa là, fibroblast growth factor: FGF 1-9), yếu tố sinh trưởng có nguồn gốc từ tiểu cầu (platelet derived growth factor: PDGF), yếu tố sinh trưởng giống insulin (insulin like growth factor: IGF-I và IGF-II), yếu tố sinh trưởng biến đổi (nghĩa là, transforming growth factor: TGF- β .I-III), yếu tố sinh trưởng màng trong mạch máu (vascular endothelial growth factor: VEGF)); và các protein có nguồn gốc từ tự nhiên hoặc được thao tác di truyền khác, polysaccharit, glycoprotein, hoặc lipoprotein. Các yếu tố sinh trưởng này được mô tả trong tài liệu: *The Cellular and Molecular Basis of Bone Formation and Repair* by Vicki Rosen and R. Scott Thies, xuất bản bởi R. G. Landes Company, tài liệu này được đưa vào bản mô tả bằng cách viện dẫn.

Tơ nhện/tằm tái tổ hợp chứa các chất có hoạt tính sinh học có thể được tạo thành bằng cách trộn một hoặc nhiều chất chữa bệnh với tơ dùng để làm ra vật liệu. Theo cách khác, chất chữa bệnh có thể được phủ lên tơ, tốt hơn là bằng chất mang dược dụng. Có thể sử dụng chất mang dược dụng bất kỳ không hoà tan tơ. Các chất chữa bệnh, có thể là chất lỏng, chất rắn được nghiền mịn, hoặc dạng vật lý thích hợp bất kỳ khác.

Lượng chất chữa bệnh sẽ tùy thuộc vào loại thuốc cụ thể được sử dụng và tình trạng bệnh y học được điều trị. Thường thì, lượng thuốc là khoảng 0,001 phần trăm đến khoảng 70 phần trăm, tiêu biểu hơn là khoảng 0,001 phần trăm đến khoảng 50 phần trăm,

tiêu biểu nhất là từ khoảng 0,001 phần trăm đến khoảng 20 phần trăm theo khối lượng của vật liệu. Khi tiếp xúc với dịch lỏng của mô hoặc của cơ thể, ví dụ, thuốc sẽ được giải phóng ra.

Các khung được thao tác di truyền của mô được làm bằng tơ nhện/tằm tái tổ hợp còn có thể được cải biến sau khi chế tạo. Ví dụ, khung có thể được phủ bằng các chất có hoạt tính sinh học có chức năng như các thụ thể hoặc các chất thu hút hóa học cho quần thể tế bào mong muốn. Việc phủ có thể được áp dụng qua liên kết hấp thụ hoặc hoá học.

Các chất phụ gia thích hợp để sử dụng theo sáng chế bao gồm các hợp chất có hoạt tính sinh học hoặc được học. Các ví dụ về các hợp chất có hoạt tính sinh học bao gồm các chất điều biến gắn tế bào, như peptit chứa các thay đổi của trình tự liên kết integrin “RGD” được biết để tác động vào sự gắn tế bào, các phối tử có hoạt tính sinh học, và chất tăng cường hoặc loại trừ các dạng cụ thể khác nhau của quá trình nội sinh trong tế bào hoặc mô. Các chất như vậy bao gồm, ví dụ, chất tạo xương, như protein tạo hình xương (BMP), yếu tố sinh trưởng biểu bì (EGF), yếu tố sinh trưởng nguyên bào sợi (FGF), yếu tố sinh trưởng có nguồn gốc từ tiểu cầu (PDGF), yếu tố sinh trưởng màng trong mạch máu (VEGF), yếu tố sinh trưởng giống insulin (IGF-I và II), TGF-, peptit YIGSR, glycosaminoglycan (GAG), axit hyaluronic (HA), integrin, selectin và cadherin.

Các khung được tạo hình thành các vật phẩm dùng cho các ứng dụng thao tác di truyền của mô và tái tạo mô, kể cả phẫu thuật phục hồi. Cấu trúc của khung cho phép sự nội sinh trưởng tế bào số lượng lớn, loại trừ nhu cầu gieo hạt trước của tế bào. Khung cũng có thể được đúc thành khung ngoài để hỗ trợ việc nuôi các tế bào *in vitro* để tạo thành bộ phận đỡ ngoài.

Khung hoạt động để bắt chước khuôn ngoại bào (extracellular matrice: ECM) của cơ thể. Khung hoạt động như cả giá đỡ vật lý lẫn chất kết dính cho các tế bào đã tách trong môi trường nuôi cấy *in vitro* và cấy tiếp theo. Khi quần thể tế bào đã được cấy phát triển và các tế bào hoạt động bình thường, chúng bắt đầu tạo ra giá đỡ ECM riêng của chúng.

Đối với việc phục hồi mô cấu trúc như sụn và xương, hình dạng mô gắn với chức năng, đòi hỏi việc đúc khung thành vật phẩm có chiều dày và hình dạng khác nhau. Các kẽ, lỗ hoặc cấu trúc tinh vi bất kỳ mang một cấu trúc ba chiều có thể được tạo ra bằng cách loại bỏ các phần của khuôn bằng kéo, dao, chùm laze hoặc dụng cụ cắt bất kỳ khác.

Các ứng dụng của khung bao gồm tái tạo mô như mô thần kinh, xương cơ, sụn, gân, gan, tụy, mắt, da, động mạch-tĩnh mạch, tiết niệu hoặc mô khác bất kỳ tạo thành cơ quan đặc hoặc rỗng.

Khung cũng có thể được sử dụng trong quá trình cấy ghép làm khuôn dùng cho các tế bào phân ly, ví dụ, tế bào sụn hoặc tế bào gan, để tạo mô hoặc cơ quan ba chiều. Loại tế bào bất kỳ có thể được bổ sung vào khung dùng để nuôi và có thể cấy, bao gồm tế bào của các hệ thống cơ và khung xương, như tế bào sụn, nguyên bào sợi, tế bào cơ và tế bào xương, tế bào nhu mô như tế bào gan, tế bào tụy (bao gồm tế bào Islet), các tế bào có nguồn gốc ruột, và tế bào khác như tế bào thần kinh, tế bào tuỷ xương, tế bào da, tế bào toàn năng và tế bào gốc, và tổ hợp của chúng, hoặc khi thu được từ thể cho, từ các dòng nuôi tế bào được thiết lập, hoặc thậm chí trước hoặc sau khi thao tác di truyền. Các mảnh mô cũng có thể được sử dụng, chúng có thể tạo ra nhiều loại tế bào khác nhau trong cùng một cấu trúc.

Các tế bào thu được từ thể cho thích hợp, hoặc bệnh nhân mà chúng được cấy vào trong đó, được phân ly bằng cách sử dụng các kỹ thuật tiêu chuẩn và được gieo lên trên và vào trong khung. Việc nuôi *in vitro* có thể được thực hiện tùy ý trước khi cấy. Cách khác, khung được cấy, được để cho tạo mạch, sau đó các tế bào được tiêm vào trong khung. Các phương pháp và các chất phản ứng để nuôi tế bào *in vitro* và cấy khung mô đã được biết đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật này.

Tơ nhện/tằm tái tổ hợp theo sáng chế có thể được khử trùng bằng cách sử dụng quy trình khử trùng thông thường như khử trùng bằng bức xạ (nghĩa là, tia gama), khử trùng bằng hoá chất (etylen oxit) hoặc các quy trình thích hợp khác. Tốt hơn là quy trình khử trùng bằng etylen oxit ở nhiệt độ từ 52 đến 55°C trong thời gian 8 giờ hoặc ngắn hơn. Sau khi khử trùng, các vật liệu sinh học có thể được bao gói trong bao bì chống ẩm vô trùng thích hợp để vận chuyển và sử dụng trong bệnh viện hoặc các cơ sở chăm sóc sức khoẻ khác.

Tơ khảm theo sáng chế cũng có thể được sử dụng để sản xuất các dạng quần áo thể thao và bảo vệ khác nhau, như trong sản xuất/chế tạo quần áo thể thao và gi lê chống đạn. Tơ nhện khảm được bộc lộ trong bản mô tả này cũng có thể dùng trong công nghiệp ô tô, như để chế tạo túi khí cải tiến. Túi khí sử dụng tơ khảm được bộc lộ có năng lượng

va đập lớn hơn khi đâm xe, rất giống mạng nhện hấp thu năng lượng của các côn trùng đang bay là nạn nhân của mạng nhện.

Định nghĩa

Như được sử dụng trong bản mô tả này, các phương pháp tương thích sinh học từ đó tơ hoặc vật liệu được điều chế là không độc, không gây đột biến, và gợi ra phản ứng viêm từ rất ít đến vừa phải. Polyme tương thích sinh học ưu tiên sử dụng theo sáng chế có thể bao gồm, ví dụ, polyetylen.oxit (PEO), polyetylen glycol (PEG), collagen, fibronectin, keratin, axit polyaspartic, polylysin, alginat, chitosan, chitin, axit hyaluronic, pectin, polycaprolacton, axit polylactic, axit polyglycolic, polyhydroxyalkanoat, dextran và polyanhydrit. Theo sáng chế, hai hoặc nhiều polyme tương thích sinh học có thể được bổ sung vào dung dịch nước.

Như được sử dụng trong bản mô tả này, trình tự miền và/hoặc môtip dẻo và/hoặc đàn hồi được xác định là trình tự di truyền có thể nhận dạng được của mảnh gen hoặc protein mà mã hoá tơ nhện liên quan đến việc truyền đặc tính đàn hồi và/hoặc dẻo cho vật liệu, như cho tơ. Ví dụ, các môtip mềm dẻo và/hoặc đàn hồi và/hoặc miền là GPGGA (SEQ ID NO: 2).

Như được sử dụng trong bản mô tả này, môtip bền được xác định là trình tự di truyền có thể nhận dạng được của mảnh gen hoặc protein mã hoá tơ nhện liên quan đến việc truyền đặc tính bền cho vật liệu, như tăng và/hoặc tăng cường độ bền kéo cho tơ. Ví dụ, một số môtip bền của tơ nhện này là: GGPSGPGS(A) 8 (trong đó A là trình tự poly adenin) (SEQ ID NO: 3).

Ví dụ thực hiện sáng chế

Sáng chế còn được đặc trưng bởi các ví dụ dưới đây, các ví dụ này được dự định để minh họa sáng chế.

Ví dụ 1 – Vật liệu và phương pháp

Ví dụ này được thực hiện để mô tả vật liệu và phương pháp/kỹ thuật được sử dụng để tạo ra tầm chuyển gen, các quy trình chung được sử dụng để tạo ra các cấu trúc di truyền được sử dụng, cũng như các bảng tham chiếu được dùng để đánh giá độ bền kéo của tơ nhện chuyển gen.

1. Trình tự gen được sử dụng. Các trình tự gen được sử dụng được đưa ra trên các

Fig.13-16 được cung cấp trong bản mô tả này. Các cải biến của các trình tự này cũng được xem là một phần của sáng chế, khi dự liệu là các phiên bản ngắn hơn và/hoặc dài hơn của các trình tự này có thể được sử dụng để có các thay thế bảo toàn, ví dụ, với các tính chất của protein tơ nhện khảm gần như tương tự.

2. Các protein tơ nhện khảm và các tơ thu được bằng các protein tơ khảm này sẽ được đánh giá về độ bền kéo. Bảng 1 đưa ra mẫu thông thường so với tơ nhện khảm sẽ được đánh giá. Tơ nhện khảm theo sáng chế được phát hiện có độ bền kéo và các đặc tính về độ bền cơ học khác tương tự như các đặc tính của tơ nhện tự nhiên.

Bảng 1: So sánh các tính chất cơ học của tơ nhện^a

Vật liệu	Độ bền (N m ⁻²)	Độ giãn (%)	Năng lượng phá huỷ (J kg ⁻¹)
Tơ kéo	4 x 10 ⁹	35	4 x 10 ⁵
Tơ ampullate phụ	1 x 10 ⁹	5	3 x 10 ⁴
Tơ hình roi	1 x 10 ⁹	>200	4 x 10 ⁵
Tơ hình ống	1 x 10 ⁹	20	1 x 10 ⁵
Chùm nang	0,7 x 10 ⁹	80	6 x 10 ⁹
KEVLAR	4 x 10 ⁹	5	3 x 10 ⁴
Cao su	1 x 10 ⁶	600	8 x 10 ⁴
Gân	1 x 10 ⁶	5	5 x 10 ³

^aSố liệu được trích từ (Gosline, et al. 1984).

Ví dụ 2 – Phân tích các tính chất về độ bền kéo của tơ tằm được biến nạp riêng rẽ

Tơ tằm chuyển gen được phân tích về sự có mặt của protein khảm của tơ nhện nhờ thẩm tách Western cả thành phần protein tuyến tơ tằm và tơ của kén tằm chuyển gen bằng cách sử dụng kháng thể đặc hiệu của tơ nhện. Trong cả hai trường hợp, tằm chuyển gen được xác nhận là tạo ra các protein khảm, và các nghiên cứu chiết vi phân cho thấy các protein này là các thành phần không thể thiếu của tơ chuyển gen của kén. Hơn nữa, biểu hiện của mỗi thể dung hợp protein huỳnh quang màu lục khảm là rõ ràng trong cả tuyến tơ lẫn tơ nhờ sự kiểm tra trực tiếp tuyến tơ hoặc tơ bằng cách sử dụng kính hiển vi giải phẫu huỳnh quang. Trong hầu hết các trường hợp, lượng protein huỳnh quang của tơ

đủ cao để nhìn thấy bằng mắt thường nhờ màu lục của kén dưới ánh sáng bình thường.

Bảng 2 thể hiện phân tích tơ chuyển gen được tạo ra từ các tầm chuyển gen riêng rẽ. Các phân tích này cho thấy chắc chắn là các dòng chuyển gen được biến nạp với các cấu trúc nhện-4 hoặc nhện-6 tạo ra tơ nhện/tơ tầm khảm có độ bền được cải thiện so với tơ của tầm không được biến nạp. Đặc biệt, các tơ này trong một số trường hợp chắc hơn tơ tầm tự nhiên gần gấp hai lần. Sự tăng gấp đôi về độ bền của tơ khảm của tầm/nhện gần đúng với sự cải thiện được cho là cần thiết để làm cho tơ tầm chắc và dẻo như tơ nhện. Vì vậy, các kết quả này chứng minh rằng tầm có thể được thao tác di truyền để tạo ra tơ nhện/tơ tầm khảm có thể cạnh tranh thuận lợi với tơ nhện tự nhiên bằng cách sử dụng các vectơ *piggyBac* mã hoá các miền bền và/hoặc dẻo xác định của tơ nhện để thiết kế protein khảm của tơ *Bombyx*/nhện.

Bảng 2: Phân tích độ bền kéo của tơ tằm chuyển gen so với pnd-w1 không được biến nạp và giống tằm thương mại.

Bảng 2: Phân tích độ bền kéo của tơ tằm chuyển gen so với pnd-w1 không được biến nạp và giống tằm thương mại.					
Mẫu số	Dòng tằm	Độ bền kéo bù (N)	Độ bền kéo biến đổi sang đơn vị CGS (dyn/21 denier)	Độ bền kéo biến đổi sang đơn vị CGS (dyn/denier)	Số lần tăng so với pnd-w1
1	đối chứng pnd –w1	0,531	53131,1	2530,1	1
2	P6+0	0,809	80947,7	3854,7	1,52
3	P6+1	0,552	55155,2	2626,4	1,03
4	P6+3	0,542	54218,2	2581,8	1,02
5	P6+4	0,815	81496,7	3880,8	1,53
6	P6+5	0,656	65594,1	3123,5	1,23
7	P4+1	0,965	96460,6	4593,4	1,82
8	P4+3	0,630	63000,0	3000,0	1,18
9	tằm thương mại của Triều Tiên	0,676	67584,5	3218,3	1,27

Ví dụ 3 – Catxet biểu hiện gen khảm của tằm và vector *piggyBac* dùng để biểu hiện protein tơ nhện/tằm khảm của tằm chuyển gen

Ví dụ này được thực hiện để chứng minh tính hữu dụng và phạm vi của sáng chế trong việc tạo ra nhiều loại catxet biểu hiện gen tơ nhện khảm của tằm. Ví dụ này cũng chứng minh sự hoàn thiện của các vector *piggyBac* được thể hiện để biến đổi tằm thành công, và dẫn đến sản xuất thành công các protein tơ nhện khảm hữu dụng về mặt thương mại để sản xuất tơ có chiều dài hữu dụng về mặt thương mại trong sản xuất.

Catxet biểu hiện.

Một số thay đổi của các catxet biểu hiện cơ bản được thể hiện dưới đây được thiết

kế. Các thiết kế này phản ánh sự gắn các thiết kế dành để biểu hiện thể khảm chuỗi nặng fibroin (fnc)-tơ nhện, trong đó trình tự protein tơ nhện tổng hợp được gắn bên sườn bởi các mảnh tận cùng N và C của protein fnc *B. mori*. Về mặt này, một số thay đổi về catxet biểu hiện chuỗi nặng fibroin tơ *Bombyx mori* cơ bản được thể hiện trên Fig.5 được thiết kế. Thiết kế bao gồm việc gắn các thiết kế dành để biểu hiện thể khảm chuỗi nặng fibroin (fnc)-tơ nhện, trong đó trình tự protein tơ nhện tổng hợp được gắn bên sườn bởi các mảnh đầu tận cùng N và C của protein fnc *B. mori*. Các yếu tố di truyền có liên quan về chức năng trong mỗi catxet biểu hiện, từ trái sang phải, bao gồm: trình tự khởi đầu chính, yếu tố tăng cường trước (UEE), trình tự khởi đầu cơ bản, và miền đầu tận cùng N (NTD) của gen fnc *B. mori*, tiếp theo là các trình tự protein tơ nhện tổng hợp khác nhau (xem dưới đây) nằm trong khung có vị trí bắt đầu dịch mã nằm trước trong NTD, tiếp theo là miền đầu tận cùng C fnc (CTD) chứa các vị trí kết thúc dịch mã và polyadenyl hóa ARN.

Có tám phiên bản khác nhau của catxet biểu hiện được vẽ trên Fig.5, chúng mã hoá bốn protein tơ nhện/tầm khảm khác nhau có hoặc không có EGFP được chèn trong khung giữa NTD và các trình tự tơ nhện. Các trình tự này được ký hiệu là “nhện 2”, “nhện 4”, “nhện 6”, và “nhện 8” và chúng được xác định như sau:

- a) Nhện 2: 7.104bp, cấu tạo từ (A4S8)24. A1 chỉ 4 bản sao của môtip đàn hồi của tơ hình roi giả định (GPGGA) (SEQ ID NO: 2); vì vậy A4 chỉ 16 bản sao của trình tự này. S8 chỉ môtip bền của tơ kéo giả định [GGPSGPGS(A)8] (SEQ ID NO: 3), còn được mô tả là trình tự “polyalanin nổi”. Kích thước gần đúng của protein dung hợp GFP (green fluorescent protein: GFP, protein huỳnh quang màu lục) là $161,9 + 50,4 = 212,3\text{Kd}$.
- b) Nhện 4: 7.386bp, cấu tạo từ (A2S8)42. A2 chỉ 8 bản sao của môtip đàn hồi của tơ hình roi giả định (GPGGA) (SEQ ID NO: 2). S8 chỉ môtip bền của tơ kéo giả định [GGPSGPGS(A)8] (SEQ ID NO: 3), như ở trên. Kích thước gần đúng của protein dung hợp GFP là $169,4 + 50,4 = 219,8\text{Kd}$.
- c) Nhện 6: 2.462bp, cấu tạo từ (A2S8)14. A2 chỉ 8 bản sao của môtip đàn hồi (GPGGA) (SEQ ID NO: 2) và S8 chỉ môtip bền [GGPSGPGS(A)8] (SEQ ID NO: 3), như ở trên. Kích thước gần đúng của protein dung hợp GFP là $56,4 + 50,4 = 106,8\text{Kd}$.
- d) Nhện 8: 4.924bp, cấu tạo từ (A2S8)28. A2 chỉ 8 bản sao của môtip đàn hồi (GPGGA) (SEQ ID NO: 2) và S8 chỉ môtip bền [GGPSGPGS(A)8] (SEQ ID NO: 3), như ở trên. Kích thước gần đúng của protein dung hợp GFP là $112,8 + 50,4 = 163,2\text{Kd}$.

Kích thước của NTD exon I & II (1625 + 15161); eGFP (27135); CTD (6470) = 50.391Kd.

Ví dụ 4 – Tách dòng phụ catxet biểu hiện vào trong *piggyBac*

Mỗi phiên bản trong số tám phiên bản khác nhau của catxet biểu hiện được vẽ trên Fig.5 (và được mô tả trong ví dụ 3) ở trên được cắt khỏi plasmid mẹ bằng cách sử dụng AscI và FseI và được tách dòng phụ vào trong các vị trí tương ứng của pBAC[3xP3-DSRedaf]. Bản đồ của vector *piggyBac* này được thể hiện trên Fig.6.

Tất cả các vector *piggyBac* được mô tả ở trên, có và không có EGFP, được thử nghiệm nhờ PCR cho các thành phần riêng rẽ và hiển thị các sản phẩm được định cỡ dự kiến.

Mỗi vector trong số tám vector *piggyBac* mã hoá các protein tơ nhện được dung hợp với EGFP được đánh giá về chức năng nhờ việc thử nghiệm khả năng của chúng kích thích biểu hiện EGFP trong tuyến tơ *B. mori*. Nói ngắn gọn, tuyến tơ được phân lập khỏi tằm và súng bắn hạt được sử dụng để bắn phá tuyến tơ bằng các hạt vonfram được phủ bằng ADN *piggyBac* (hoặc đối chứng). Các mô bị bắn phá sau đó được nuôi trong môi trường Grace trong các đĩa nuôi và kính hiển vi phẫu thuật được trang bị để phát huỳnh quang EGFP có trong phòng thí nghiệm của đồng nghiệp được sử dụng để kiểm tra tuyến tơ để biểu hiện EGFP hai và ba ngày sau đó. Mỗi vector được thể hiện để kích thích huỳnh quang EGFP.

Bộ bốn vector *piggyBac* mã hoá nhện 4 và 6 có và không có lòng EGFP được sử dụng để tạo ra tằm chuyển gen.

Ví dụ 5 – Tách tằm chuyển gen.

Thông thường, sự biến đổi tằm bao gồm việc đưa hỗn hợp của vector *piggyBac* và plasmid hỗ trợ, mã hoá transposaza *piggyBac*, vào trong phôi tiền phôi bì nhờ tiêm tế bào trứng tằm. Sự tạo thành phôi bì không xảy ra miễn là 4 giờ sau khi trứng được đẻ ra. Vì vậy, việc thu thập và tiêm các phôi có thể được thực hiện ở nhiệt độ trong phòng trong khoảng thời gian tương đối dài. Khó khăn kỹ thuật đối với tiêm tế bào là cần chọc thủng màng trứng, việc này gây ra rào cản lớn. Tamura và các đồng nghiệp đã hoàn thiện kỹ thuật tiêm tế bào đối với tằm nhờ việc chọc thủng màng bằng kim vonfram nhọn và sau đó đưa chính xác kim tiêm mao quản bằng thủy tinh vào trong lỗ thu được. Hiện nay kỹ

thuật này là quy trình tương đối bình thường, được thực hiện bằng máy điều khiển kim tiêm của người máy Eppendorf được định chuẩn để chọc thủng màng, lấy kim vonfram ra, lồng mao quản bằng thủy tinh, và tiêm dung dịch ADN. Trứng sau đó được bịt kín lại bằng cách sử dụng một giọt keo Krazy nhỏ và được duy trì trong điều kiện nuôi bình thường ở nhiệt độ 28°C và độ ẩm 70% cho đến khi ấu trùng nở. Các côn trùng được tiêm còn sống sau đó được cho giao phối để sinh ra phôi thế hệ F1 dùng để nhận dạng tiếp theo thể biến nạp giả định, dựa vào biểu hiện của gen đánh dấu mắt đỏ DS. Các thể biến nạp đực và cái giả định được nhận dạng bằng phương pháp này sau đó được cho giao phối để tạo ra các dòng đồng hợp tử dùng để phân tích di truyền chi tiết hơn.

Cụ thể, sự biến đổi tầm dùng cho dự án này bao gồm việc tiêm hỗn hợp của vector *piggyBac* và ADN plasmit hỗ trợ vào trong trứng của thể đột biến của tầm có biểu bì trong suốt, *Bombyx mori* pnd-w1. Tầm đột biến này được mô tả bởi Tamura, et al. 2000, tài liệu này được đưa vào bản mô tả này bằng cách viện dẫn. Thể đột biến này thiếu sự melanin hóa làm cho việc sàng lọc bằng cách sử dụng các gen huỳnh quang dễ dàng hơn nhiều. Một khi có sự biểu hiện của gen đánh dấu mắt đỏ, các thể biến nạp giả định F1 được nhận dạng, các dòng đồng hợp tử được thiết lập và các thể biến nạp thật được khẳng định bằng cách sử dụng thẩm tách Western của protein tuyến tơ và tơ kén thu hoạch được.

Ví dụ 6 – Phân tích sản xuất tơ nhện/tầm khảm bởi tầm chuyển gen

Tơ tầm chuyển gen được phân tích về sự có mặt của protein khảm của tơ nhện nhờ thẩm tách Western cả thành phần protein tuyến tơ tầm và tơ của kén tầm chuyển gen bằng cách sử dụng kháng thể đặc hiệu tơ nhện. Trong cả hai trường hợp, tầm chuyển gen được xác nhận là tạo ra các protein khảm, và các thử nghiệm chiết khác nhau cho thấy các protein này là các thành phần không thể thiếu của tơ chuyển gen của kén.

Hơn nữa, sự biểu hiện của mỗi thể dung hợp của protein huỳnh quang màu lục khảm là rõ ràng trong cả tuyến tơ và tơ nhờ sự kiểm tra trực tiếp tuyến tơ hoặc tơ bằng cách sử dụng kính hiển vi giải phẫu huỳnh quang. (Fig.7). Trong hầu hết các trường hợp lượng protein huỳnh quang của tơ đủ cao để được nhìn thấy bằng mắt thường nhờ màu lục của kén dưới ánh sáng bình thường.

Ví dụ 7 – Thiết kế vectơ *piggyBac*

piggyBac là vector chọn cho dự án này bởi vì nó có thể được sử dụng để biến nạp tằm^{4, 11, 43} một cách có hiệu quả. Các vector *piggyBac* đặc trưng được sử dụng trong dự án này được thiết kế để mang các gen có một số đặc điểm quyết định. Như được làm nổi bật trên Fig.17, các vector này chứa trình tự khởi đầu chuỗi nặng fibroin (fhc) *B. mori* có thể nhằm biểu hiện protein tơ nhện ngoại lai vào tuyến tơ sau^{91, 92}, và gen tăng cường fhc có thể tăng mức độ biểu hiện và tạo điều kiện gắn protein tơ ngoại lai vào trong tơ⁹³. Các vector *piggyBac* cũng mã hoá A2S814 (Fig.17A), là protein tơ nhện tổng hợp tương đối lớn có cả hai motif đàn hồi (GPGGA)₈ (SEQ ID NO: 4) và bền (alanin₈ liên kết) ("alanin₈" được bộc lộ như SEQ ID NO: 5). Trình tự protein tơ nhện tổng hợp được được gắn vào trong các trình tự mã hoá các miền đầu tận cùng N và C của protein fhc *Bombyx mori* (các Fig.17B-17C). Thiết kế tơ tằm/nhện khảm này đã được sử dụng trước đây để hướng sự hợp nhất protein ngoại lai vào trong tơ nội sinh mới sinh trong tuyến tơ *B. mori* và tạo ra tơ compozit^{91, 92}.

Một trong các vector *piggyBac* được thiết kế theo nghiên cứu này mã hoá protein tơ tằm/nhện khảm riêng rẽ (Fig.17B), trong khi các vector khác mã hoá cùng một protein này bằng nhãn protein huỳnh quang màu lục cải tiến (enhanced green fluorescent protein: EGFP) đầu tận cùng N (Fig.7C). Kết cấu sau giúp phân tích tơ được tạo ra bởi thể hệ con cháu đã được biến nạp và cũng được dùng cho các thử nghiệm bắn phá tuyến tơ *ex vivo* sơ bộ để kiểm tra sự biểu hiện protein tơ nhện khảm trong tuyến tơ, như được mô tả trong bản mô tả này.

Phương pháp:

Một số mảnh gen được phân lập nhờ các phản ứng chuỗi polymeraza (polymerase chain reaction: PCR) với ADN hệ gen được phân lập từ tuyến tơ của giống P50/Daizo *Bombyx mori* và các đoạn mỗi gen đặc trưng được thể hiện trên Fig.17. Các mảnh này chứa trình tự khởi đầu chính fhc và yếu tố tăng cường trước (major promoter ADN upstream enhancer element: MP-UEE), hai phiên bản của trình tự khởi đầu cơ bản fhc (basal promoter: BP) và miền đầu tận cùng N (NTD; exon 1/intron 1/exon 2) có các vị trí giới hạn gắn bên sườn ở vị trí 5' và 3' khác nhau, miền đầu tận cùng C fhc (CTD; trình tự mã hoá 3' và tín hiệu poly A), và EGFP. Trong mỗi trường hợp, các sản phẩm khuếch đại được làm sạch bằng gen, và các mảnh ADN có các kích thước mong muốn được cắt ra và được phục hồi. Tiếp theo, các mảnh MP-UEE fhc, CTD fhc, và EGFP được phân lập dòng vào trong pSLfa1180fa (pSL) (Y. Miao), hai mảnh NTD khác nhau được phân lập

dòng vào trong pCR4-TOPO (Invitrogen Corporation, Carlsbad, CA), và các thể biến nạp *E. coli* chứa các sản phẩm khuếch đại chính xác được nhận dạng nhờ lập bản đồ giới hạn và được xác nhận bằng cách xác định trình tự.

Các mảnh này sau đó được sử dụng để gắn các vector *piggyBac* được sử dụng trong nghiên cứu này như dưới đây. Trình tự tơ nhện A2S8₁₄ tổng hợp được cắt khỏi tiền chất plasmid pBluescript SKII+ (F. Teulé and R.V. Lewis) bằng BamHI và BspEI, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước CTD trong plasmid trung gian pSL được mô tả ở trên. Bước này tạo ra plasmid dành cho pSL-nhện 6-CTD. Mảnh NotI/BamHI sau đó được cắt khỏi một plasmid trong số các plasmid trung gian pCR4- TOPO-NTD được mô tả ở trên, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước trình tự nhện 6-CTD trong pSL nhện 6-CTD để tạo ra pSL-NTD-nhện 6-CTD. Đồng thời, mảnh NotI/XbaI được cắt khỏi plasmid trung gian pCR4-TOPO-NTD kia được mô tả ở trên, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước gen khuếch đại EGFP trong plasmid trung gian pSL-EGFP được mô tả ở trên. Việc này tạo ra plasmid chứa mảnh NTD-EGFP được cắt bằng NotI và BamHI và được tách dòng phụ vào trong các vị trí tương ứng trước các trình tự nhện 6-CTD trong pSL-nhện 6-CTD. Mảnh MP-UEE sau đó được cắt bằng SfiI và NotI khỏi plasmid trung gian pSL được mô tả ở trên, được làm sạch bằng gel, được phục hồi, và được phân lập dòng vào trong các vị trí tương ứng trước các trình tự NTD-nhện 6-CTD và NTD-EGFP-nhện 6-CTD trong hai plasmid pSL trung gian khác nhau được mô tả ở trên. Cuối cùng, các catxet MP-UEE-NTD- A2S8₁₄-CTD hoặc MP-UEE- NTD-EGFP-A2S8₁₄- CTD đã được gắn hoàn toàn được cắt bằng AscI và FseI khỏi các plasmid pSL cuối cùng tương ứng và được tách dòng phụ vào trong các vị trí tương ứng của pBAC[3XP3-DsRedaf]⁹⁸. Bước tách dòng phụ cuối cùng này tạo ra hai vector *piggyBac* riêng rẽ được gọi là nhện 6 và nhện 6-EGFP để chỉ sự không có hoặc có mặt của gen đánh dấu EGFP. Các vector này được dùng cho thử nghiệm bắn phá tuyến tơ *ex vivo* và truyền thông tin di truyền của tằm, như được mô tả dưới đây.

Kết quả:

Kết quả thử nghiệm *ex vivo* cho thấy vector *piggyBac* mã hoá protein tơ tằm/nhện khảm gắn nhãn GFP kích thích huỳnh quang màu lục trong vùng tuyến tơ sau. Các thử nghiệm thẩm tách miễn dịch bằng kháng thể đặc hiệu GFP còn chứng minh rằng tuyến tơ

bị bắn phá chứa protein phản ứng miễn dịch có khối lượng phân tử biểu kiến (M_r) ~116kDa. Chỉ lớn hơn một chút so với dự kiến (106kDa), kết quả này xác nhận thiết kế cơ bản của các vectơ *piggyBac* và thúc đẩy sự tách tầm chuyển gen bằng cách sử dụng các cấu trúc này.

Ví dụ 8 – Phân lập tầm chuyển gen

Mỗi vectơ *piggyBac* được trộn với plasmit mã hoá transposaza *piggyBac* và các hỗn hợp được tiêm tế bào độc lập vào trong trứng được phân lập từ pnd-w1⁴³ *Bombyx mori*. Giống tầm này được sử dụng bởi vì nó thiếu sự melanin hóa tạo ra phenotip biểu bì trong suốt giúp phát hiện protein tơ tầm - nhện khảm được gắn nhãn EGFP trong các thể biến nạp. Các thể biến nạp F1 giả định được nhận dạng ban đầu nhờ phenotip mắt đỏ thu được từ biểu hiện của DS-đỏ dưới sự kiểm soát của trình tự khởi đầu²⁷ 3XP3 trung tính đặc thù nằm trong mỗi vectơ *piggyBac* (Fig.17D). Các con vật này được sử dụng để thiết lập một số dòng tầm chuyển gen đồng hợp tử, như được mô tả trong phần phương pháp, chúng được gọi là nhện 6 và nhện 6-GFP, chỉ vectơ *piggyBac* dùng để biến nạp chúng.

Phương pháp:

Thử nghiệm bắn phá tuyến tơ *ex vivo*

Tầm pnd-w1 giống *Bombyx mori* sống bước vào ngày thứ ba của tuổi thứ năm được khử trùng nhờ ngâm chìm trong etanol 70% trong vài giây và được để trong NaCl 0,7% khối lượng/thể tích. Toàn bộ tuyến tơ sau đó được cắt vô trùng khỏi mỗi con tầm và được chuyển vào các đĩa Petri chứa môi trường Grace được bổ sung bằng các chất kháng sinh, ở đó chúng được giữ trước quy trình bắn phá ADN. Đồng thời, các vi hạt vonfram (1,7µm M-25 hạt vi tải; Bio-Rad Laboratories, Hercules, CA) được phủ bằng ADN dùng để bắn phá, như dưới đây. Các vi hạt được xử lý trước theo chỉ dẫn của nhà sản xuất và được giữ với lượng nhỏ 3mg/50µl trong 50% glyxerol ở -20°C. Ngay trước mỗi thử nghiệm bắn phá, các lượng nhỏ 3mg vi hạt được phủ bằng 5µg ADN *piggyBac* có liên quan trong thể tích tối đa 5µl, theo chỉ dẫn của nhà sản xuất. Một số lượng nhỏ vi hạt được phủ bằng nước cất để sử dụng làm các đối chứng ADN âm tính. Mỗi thử nghiệm bắn phá bao gồm sáu bản sao và mỗi bắn phá riêng rẽ bao gồm một cặp tuyến tơ nguyên vẹn. Đối với việc bắn phá, các tuyến được chuyển từ trạng thái nằm trong môi trường Grace trên các đĩa Petri 90mm chứa 1% khối lượng/thể tích thạch vô trùng và các đĩa Petri được để trong buồng của hệ thống phân phối hạt Bio-Rad Biolistic® PDS-1000/He.

Buồng được tạo chân không đến từ 20-22 trong Hg và các tuyến tơ bị bắn phá bằng các vi hạt vonfram được phủ sơ bộ bằng cách sử dụng áp suất heli 1.100psi ở khoảng cách 6cm cách nguồn hạt đến các mô đích, như được mô tả trước đây²⁶. Sau khi bắn phá, các tuyến tơ được đặt trong các đĩa Petri mới chứa môi trường Grace được bổ sung bởi 2 lần chất kháng sinh và được ủ ở nhiệt độ 28°C. Biểu hiện nhất thời của gen đánh dấu EGFP trong vectơ *piggyBac* nhận 6-GFP được đánh giá nhờ sử dụng kính hiển vi huỳnh quang tại thời điểm 48 và 72 giờ sau bắn phá. Các ảnh được chụp bằng kính hiển vi Olympus FSX100 với độ phóng đại 4,2 lần, pha 1/120 giây, và huỳnh quang màu lục 1/110 giây (chụp). Hơn nữa, biểu hiện nhất thời của các protein tơ tằm/nhện khảm gắn nhãn và không gắn nhãn EGFP được đánh giá nhờ thẩm tách miễn dịch các phần chiết của tuyến tơ bị bắn phá bằng các kháng huyết thanh đặc hiệu EGFP hoặc tơ nhện, như được mô tả dưới đây.

Sự biến nạp tằm

Trứng được thu thập 1 giờ sau khi ngài pnd-w1 đẻ và được xếp trên tấm kính của kính hiển vi. Vectơ và các plasmit hỗ trợ được huyền phù hoá lại trong chất đệm tiêm (natri phosphat 0,1mM, KCl 5mM, độ pH là 6,8) với nồng độ cuối 0,2µg/ul đối với mỗi loại, và 1-5nl được tiêm vào trong từng phôi tằm tiền phôi bì bằng cách sử dụng hệ thống tiêm cấu tạo từ bộ điều chỉnh áp suất World Precision Instruments PV820 (Hoa Kỳ), vi máy điều khiển Suruga Seiki M331 (Nhật Bản), và thiết bị giữ pipet kép Narishige HD-21 (Nhật Bản). Trứng bị chọc thủng được bịt kín bằng gel Helping Hand Super Glue (The Faucet Queens, Inc., USA) và sau đó được để trong buồng sinh trưởng ở nhiệt độ 25°C và độ ẩm 70% để phát triển phôi. Sau khi nở, ấu trùng được nuôi theo chế độ ăn nhân tạo (Nihon Nosan Co., Japan) và thu được các thể hệ tiếp theo nhờ việc cho giao phối cùng huyết thống trong cùng một dòng. Thể hệ con do truyền gen được nhận dạng thăm dò nhờ sự có mặt của gen đánh dấu mắt huỳnh quang đỏ Ds bằng cách sử dụng kính hiển vi Olympus SXZ12 (Tokyo, Nhật Bản) có các kính lọc giữa 550 và 700nm.

Kết quả:

Ngay cả bằng sự kiểm tra bằng mắt dưới ánh sáng trắng, không cần kích thích EGFP đặc biệt, biểu hiện EGFP được quan sát thấy ở kén được tạo ra bởi các thể biến nạp nhận 6-GFP (Fig.18A). Biểu hiện EGFP mạnh khi các tuyến tơ (các Fig.18B-18C) và các kén (Fig.18D) của các con tằm này được kiểm tra dưới kính hiển vi huỳnh quang

cũng được quan sát thấy. Kén có vẻ chứa ít nhất một số tơ có các tín hiệu EGFP được hợp nhất. Biểu hiện protein tơ tằm/nhện khảm gắn nhãn EGFP trong các tuyến tơ nhện 6-GFP và các kén được khẳng định nhờ thẩm tách miễn dịch các phần chiết tuyến tơ và kén bằng các kháng huyết thanh protein đặc hiệu EGFP và tơ nhện (Fig.19). Kết quả tương tự thu được với các phần chiết tuyến tơ và kén nhện 6 nhờ thẩm tách miễn dịch bằng kháng huyết thanh protein tơ nhện đặc hiệu (Fig.19). Kết quả này cho thấy các tác giả của sáng chế đã phân lập thành công tầm chuyển gen mã hoá các dạng gắn nhãn hoặc không gắn nhãn EGFP của protein tơ tằm/nhện khảm và các protein này có liên quan đến tơ được tạo ra bởi các con tầm chuyển gen này.

Ví dụ 9 – Phân tích tơ composit

Phương pháp chiết protein tuần tự được sử dụng để phân tích sự liên quan của các protein tơ tằm/nhện khảm với các tơ composit được tạo ra bởi tầm chuyển gen. Sau khi loại bỏ lớp serixin liên kết yếu, tơ đã khử keo được đưa qua dãy chiết mạnh tăng dần, như được mô tả trong phần Phương pháp.

Phương pháp:

Chiết tuần tự protein của kén tằm

Kén được tạo ra bởi tầm gốc và chuyển gen được thu hoạch và lớp serixin được loại bỏ nhờ việc khuấy kén nhẹ nhàng trong Na_2CO_3 0,05% (khối lượng/thể tích) trong 15 phút ở nhiệt độ 85°C với tỷ lệ kén: dung môi là 1:50 (khối lượng/thể tích)⁴⁰. Tơ đã khử keo được lấy khỏi dung dịch và được rửa hai lần bằng nước nóng ($50-60^\circ\text{C}$) bằng cách khuấy cẩn thận và cũng vẫn tỷ lệ kén: dung môi tương tự. Tơ đã khử keo sau đó được làm đông khô và được cân để ước tính hiệu quả loại bỏ lớp serixin. Tơ đã khử keo được sử dụng cho thủ tục chiết protein tuần tự, bằng sự quay trên bánh xe trộn để bảo đảm sự khuấy không đổi, như sau. Ba mươi mg tơ đã khử keo được xử lý bằng 1ml nước muối đệm phosphat (phosphate buffered saline: PBS; NaCl 137mM, KCl 2,7mM, Na_2PO_4 10mM, KH_2PO_4 1,8mM) trong 16 giờ ở nhiệt độ 4°C . Tơ được phân lập thành các phần không tan và tan nhờ ly tâm, dịch nổi được lấy ra và được giữ ở nhiệt độ -20°C như là phần tan trong PBS, và viên vê nhỏ được đưa chiết tiếp. Viên này được huyền phù hoá lại trong 1ml 2% (khối lượng/thể tích) Độ lệch chuẩnS và được ủ trong 16 giờ ở nhiệt độ trong phòng. Tiếp tục, tơ được phân lập thành các phần không tan và tan nhờ ly tâm, dịch nổi được lấy ra và được giữ ở nhiệt độ -20°C như là phần tan trong Độ lệch chuẩnS, và

viên vè nhỏ được đưa chiết tiếp. Viên này được huyền phù hoá lại trong 1ml LiSCN 9M chứa 2% (theo thể tích) β -mercaptoetanol và được ủ trong 16-48 giờ ở nhiệt độ trong phòng. Sau khi ly tâm, dịch nổi được giữ ở nhiệt độ -20°C như là phần tan trong LiSCN 9M/BME. Viên vè nhỏ sau cùng thu được trong bước này được huyền phù hoá lại trong 1ml LiSCN 16M chứa 5% (theo thể tích) BME và được ủ trong khoảng một giờ ở nhiệt độ trong phòng. Việc này dẫn đến hoà tan hoàn toàn và tạo ra phần chiết sau cùng, nó được giữ như là phần tan trong LiSCN 16M/BME ở nhiệt độ -20°C cho đến khi thử nghiệm thẩm tách miễn dịch được thực hiện.

Phân tích protein tơ

Tuyến tơ của các thử nghiệm bắn phá *ex vivo* cũng như của tằm gốc và chuyển gen chưa được xử lý được đồng nhất hoá trên đá trong chất đệm natri phosphat (Na_2PO_4 30 mM, pH 7,4) chứa 1% (khối lượng/thể tích) Độ lệch chuẩn S và urê 5M, sau đó được làm trong trong 5 phút với tốc độ 13.500 vòng trên phút trên vi máy ly tâm ở nhiệt độ 4°C . Dịch nổi được thu hoạch như các phần chiết tuyến tơ và các phần chiết này, cũng như các phần chiết kén tuần tự được mô tả ở trên được hoà tan 4 lần bằng chất đệm Tris-HCl 10mM/2% Độ lệch chuẩn S/5% BME và các mẫu chứa $\sim 90\mu\text{g}$ tổng protein được trộn 1:1 với chất đệm tải Độ lệch chuẩn S-PAGE, được đun sôi ở nhiệt độ 95°C trong 5 phút, và được đưa lên trên các gel gradien 4-20% (Pierce Protein Products; Rockford, IL). Sau khi phân tách, các protein được chuyển từ các gel đến các màng PVDF (ImmobilonTM; Millipore, Billerica, MA) bằng cách sử dụng tế bào chuyển Bio-Rad, theo chỉ dẫn của nhà sản xuất. Phát hiện miễn dịch được thực hiện bằng cách sử dụng kháng huyết thanh thỏ đa dòng đặc hiệu protein tơ nhện được tạo ra chống lại peptit A2 giống tơ hình roi *Nephila clavipes* (GenScript Corporation, Piscataway, NJ) hoặc kháng thể chuột đơn dòng đặc hiệu EGFP thương mại (Living Colors[®] GFP, Clontech Laboratories, Mountain View, CA) làm các kháng thể sơ cấp. Các kháng thể thứ cấp là IgG-HRP ở dê kháng thỏ (Promega Corporation, Madison, WI) hoặc thể liên hợp IgG H+L HRP ở dê kháng chuột (EMD Chemicals, Gibbstown, NJ), một cách tương ứng. Tất cả các kháng thể được sử dụng với độ pha loãng 1:10.000 trong chất đệm chặn tiêu chuẩn (1x PBST/0,05% sữa khô không béo) và các phản ứng kháng thể - kháng nguyên được nhìn thấy bằng mắt nhờ sự phát quang hoá học bằng cách sử dụng bộ dụng cụ thương mại (ECLTM Western Blotting Detection Reagents; GE healthcare).

Kết quả:

Sau mỗi bước trong quy trình này, các phần tan và không tan được phân lập nhờ ly tâm, phần hoà tan được giữ để thẩm tách miễn dịch, và phần không tan được dùng cho chiết tiếp theo. Dung môi chiết cuối cùng hoà tan hoàn toàn tơ còn lại. Các đối chứng thẩm tách miễn dịch xác định rằng kháng huyết thanh đặc hiệu protein tơ nhện không nhận ra bất kỳ protein nào trong tơ pnd-w1 (Fig.19B, các làn 3-6), nhưng nhận ra protein tơ tằm/nhện A2S8¹⁴ khảm được tạo ra trong *E. coli* (Fig.19B, làn 2). Việc chiết tuần tự kén đã khử keo của tằm chuyển gen bằng cách sử dụng nước muối (Fig.19B, các làn 8 và 13), Độ lệch chuẩnS (Fig.19B, các làn 9 và 14), và LiSCN 8M/ β -mercaptoetanol 2% (Fig.19B, các làn 10 và 15) không giải phóng bất kỳ protein phản ứng miễn dịch phát hiện được nào. Tuy nhiên, việc chiết tiếp theo của tơ còn lại bằng LiSCN 16M/ β -mercaptoetanol 5% giải phóng protein phản ứng miễn dịch có M_r ~106kDa từ nhện 6 còn lại (Fig.19, làn 11) và hai protein phản ứng miễn dịch có M_r s ~130 và ~110kDa từ tơ nhện 6-GFP còn lại (Fig.19, dòng 16). Tất cả các protein này đều lớn hơn dự kiến (78kDa và 106kDa lần lượt đối với nhện 6 và nhện 6-GFP). Sự giải thích có thể đối với các protein khác nhau này bao gồm ‘việc lặp’ phiên mã/dịch mã do bởi bản chất lặp cao của các trình tự tơ nhện, sự di chuyển bất thường của các sản phẩm protein trên Độ lệch chuẩnS-PAGE, và/hoặc cải biến sau dịch mã của protein tơ tằm/nhện khảm. Protein tơ tằm/nhện A2S8₁₄ khảm được tạo ra trong *E. coli* là đối chứng dương tính dùng cho thẩm tách miễn dịch, cũng có M_r lớn hơn (~75kDa) so với dự kiến (60kDa). Các phần chiết bằng LiSCN 16M/ β -mercaptoetanol 5% của kén đã khử keo của cả hai dòng tằm chuyển gen cũng chứa kính phết phản ứng miễn dịch có các M_r từ ~40 đến ~75kDa, phản ánh sự phân hủy có thể xảy ra của các protein tơ tằm/nhện khảm và/hoặc sự kết thúc dịch mã sớm. Không kể kích thước của các sản phẩm chuyển gen hoặc lý do xuất hiện của chúng, kết quả chiết tuần tự chứng minh rõ ràng rằng tằm chuyển gen được đề xuất như được mô tả trong bản mô tả này biểu hiện các protein tơ tằm/nhện khảm được hợp nhất rất ổn định vào trong tơ composit.

Ví dụ 10 – Tính chất cơ học của tơ composit

Các tính chất cơ học của tơ tự nhiên và composit đã được khử keo của tơ composit được tạo ra bởi tằm chuyển gen được mô tả trong bản mô tả này.

Các phương pháp mà nhờ chúng tơ composit được chuẩn bị để thử nghiệm, và việc thử nghiệm được thực hiện như thế nào, được trình bày dưới đây trong phần Phương pháp.

Phương pháp:

Tơ tằm đã khử keo dùng để thử nghiệm cơ học có chiều dài ban đầu (L_0) 19mm. Thử nghiệm tơ đơn được thực hiện trong điều kiện môi trường (nhiệt độ 20-22°C và độ ẩm 19-22%) bằng cách sử dụng hệ thống MTS Synergie 100 (MTS Systems Corporation, Eden Prairie MN) được lắp với cả ô 50N tiêu chuẩn và ô tải 10g được chế tạo thông thường (Transducer Techniques, Temecula CA). Các số liệu cơ học (tải trọng và độ giãn) được ghi lại từ cả hai ô tải bằng phần mềm TestWorks® 4.05 (MTS Systems Corporation, Eden Prairie, MN) với tốc độ biến dạng bằng 5mm/phút và tần số 250MHz cho phép tính các giá trị ứng suất và biến dạng. Các đường cong ứng suất/biến dạng của các bộ số liệu được tập hợp cho từng loại tơ được vẽ bằng cách sử dụng MATLAB (Version 7.1) để xác định độ bền (hoặc năng lượng phá hủy), mô đun Young (độ cứng ban đầu), ứng suất tối đa, và độ kéo dài tối đa (= % biến dạng tối đa).

Kết quả:

Các kết quả đã chứng minh rằng tơ composit đã khử keo chứa các protein tơ tằm/nhện khả năng gắn nhãn hoặc không gắn nhãn EGFP có khả năng kéo dài lớn hơn đáng kể và độ bền và độ cứng được cải thiện một chút so với tơ tự nhiên của tằm pnd-w1 (bảng 3 và Fig.20). Bảng 3: Tính chất cơ học của 12-15 tơ được tạo ra bởi tằm cha và chuyển gen được đo trong điều kiện tương hợp chính xác về nhiệt độ, độ ẩm, và tốc độ thử nghiệm và các giá trị trung bình và các độ lệch tiêu chuẩn được trình bày trong Bảng. Các tính chất cơ học trung bình của tơ nhện tơ kéo (*Nephila clavipes*) được xác định đồng thời trong điều kiện giống hệt được đưa ra để so sánh.

Bảng 3: Tính chất cơ học của tơ tự nhiên và composit đã được khử keo

Tính chất cơ học	Pnd-w1		Nhện 6		Nhện 6-GFP (dòng 1)		Nhện 6-GFP (dòng 4)		Tơ kéo (nhện)
	Trung bình	Độ lệch chuẩn	Trung bình	Độ lệch chuẩn	Trung bình	Độ lệch chuẩn	Trung bình	Độ lệch chuẩn	Trung bình
Ứng suất tối đa	198,0	28,1	315,3	65,8	281,9	57,7	338,4	87,0	744,5

(MPa)									
Độ giãn tối đa (%)	22,0	5,8	31,8	5,2	32,5	4,3	31,1	4,5	30,6
Độ bền MJ/m ³	32,0	10,0	71,7	13,9	68,9	16,2	77,2	29,5	138,7
Môđun Young (MPa)	3705,0	999,6	5266,8	1656,5	4860,9	1269,2	5498,1	1181,2	9267,7

Các tính chất cơ học của 12-15 tơ được tạo ra bởi tầm gốc và tầm chuyển gen được xác định và các giá trị trung bình và các độ lệch tiêu chuẩn được trình bày trong Bảng. Các tính chất cơ học tối ưu của tơ nhện tơ kéo (*Nephila clavipes*) được xác định trong cùng một điều kiện được đưa ra để so sánh.

Vì vậy, các tơ composít này bền hơn tơ tầm tự nhiên. Tính chất cơ học của tơ composít được tạo ra bởi tầm chuyển gen biến đổi hơn nhiều so với tính chất của tơ tự nhiên được tạo ra bởi chủng cha mẹ. Ngoài ra, tơ composít được tạo ra bởi hai dòng nhện 6-GFP khác nhau có khả năng kéo dài tương tự, nhưng độ bền kéo khác nhau. Các thay đổi được quan sát thấy trong tính chất cơ học của tơ composít trong dòng chuyển gen riêng rẽ và thay đổi dòng đối với dòng có thể phản ánh tính không đồng nhất của tơ composít, tính không đồng nhất có thể là do sự khác nhau về tỷ lệ protein tơ tầm/nhện khảm và/hoặc sự định vị của các protein này dọc theo tơ. Có thể thấy bằng chứng về tính không đồng nhất của tơ composít trên Fig.18D. Việc so sánh các đặc tính cơ học tốt nhất được quan sát thấy đối với tơ composít của tầm chuyển gen, tơ tự nhiên của tầm cha mẹ, và tơ nhện tơ kéo đại diện được thể hiện trên Fig.20. Kết quả cho thấy tất cả các tơ composít đều bền hơn tơ tự nhiên của tầm pnd-w1. Hơn nữa, tơ composít của tầm nhện 6-GFP dòng 4 chuyển gen thậm chí bền hơn tơ nhện tơ kéo tự nhiên được thử nghiệm trong cùng một điều kiện. Kết quả này chứng minh rằng việc hợp nhất các protein tơ

tằm/nhện khảm có thể cải thiện đáng kể tính chất cơ học của tơ composít được tạo ra bằng cách sử dụng nền tảng tằm chuyển gen.

Các đặc tính cơ học tốt nhất được đo với tơ tằm tự nhiên (pnd-w1) và nhện (tơ kéo *N. clavipes*) được so sánh với các đặc tính thu được với tơ composít được tạo ra bởi tằm chuyển gen. Tất cả các tơ được thử nghiệm trong cùng một điều kiện. Các giá trị bền nhất là: tằm pnd-w1 (đường màu lam, 43,9MJ/m³); nhện 6 dòng 7 (đường màu cam, 86,3MJ/m³); nhện 6-GFP dòng 1 (đường màu lục sẫm, 98,2MJ/m³), nhện 6-GFP dòng 4 (đường màu lục sáng, 167,2MJ/m³); và tơ kéo *N. clavipes* (đường màu đỏ, 138,7MJ/m³) (xem Bảng 3).

Ví dụ 11 – Tơ composít chứa protein tơ tằm/nhện khảm được hợp nhất ổn định

Tơ nhện có rất nhiều tính năng sử dụng như các vật liệu sinh học cho nhiều ứng dụng khác nhau. Trước đây, các khó khăn lớn đối với việc nuôi nhện to đã làm tê liệt nỗ lực sản xuất tự nhiên. Sự cần thiết phát triển phương pháp công nghệ sinh học có hiệu quả để sản xuất tơ nhện được trình bày trong nền được đưa ra trong phần bản chất kỹ thuật của sáng chế. Trong khi các nền khác đã được mô tả để sử dụng trong sản xuất các protein tơ nhện tái tổ hợp, việc gia công có hiệu quả các protein này thành tơ hữu dụng vẫn còn khó khăn. Yêu cầu sản xuất tơ, không chỉ các protein, đặt tằm như là nền tảng chất lượng đối với ứng dụng công nghệ sinh học cụ thể này.

Tằm chuyển gen được xử lý để tạo ra protein tơ nhện được phân lập bằng cách sử dụng vectơ *piggyBac* mã hoá protein tơ spidroin-1 ampullate lớn *Nephila clavipes* tự nhiên dưới sự kiểm soát phiên mã của trình tự khởi đầu serixin (sericin: Ser1) *Bombyx mori*. Trình tự spidroin được hoà lẫn vào trình tự sau mã hoá peptit fhc đầu tận cùng C. Tằm chuyển gen được phân lập bằng cách sử dụng cấu trúc *piggyBac* này tạo ra kén chứa protein tơ tằm/nhện khảm, nhưng protein này chỉ được tìm thấy trong lớp serixin liên kết yếu. Ngược lại, protein tơ tằm/nhện khảm được tạo ra bởi tằm chuyển gen được bộc lộ trong sáng chế này là thành phần không thể thiếu của tơ composít. Liên kết tương đối yếu của protein tơ tằm/nhện khảm được thiết kế bởi các người khác, có thể, trong các vấn đề khác, phản ánh sự không có miền fhc của tằm đầu tận cùng N. Cách khác, việc sử dụng trình tự khởi đầu Ser1 trong vectơ *piggyBac* có thể, trong các vấn đề khác, không phù hợp với sự gắn tơ chính xác, vì trình tự khởi đầu này hoạt động phiên mã trong tuyến tơ giữa, trong khi đó các trình tự khởi đầu fhc, flc, và fhx kiểm soát lần lượt sự biểu hiện fhc, chuỗi nhẹ fibroin, và các protein hexamerin, hoạt động trong tuyến tơ sau. Việc gắn

protein tơ tằm vào trong tơ được kiểm soát, một phần, nhờ sự điều khiển không gian và thời gian chặt chẽ đối với sự biểu hiện gen tơ. Vì vậy, các vectơ được bọc lộ trong sáng chế này được xử lý bằng trình tự khởi đầu *hfc* để điều khiển việc tích lũy protein tơ tằm/nhện khảm trong cùng một chỗ và trong cùng một thời gian như protein tơ tự nhiên, để tạo điều kiện hợp nhất ổn định protein khảm vào trong tơ composít mới được lắp ráp. Các tác giả khác đã mô tả sự tăng nhẹ về độ đàn hồi và độ bền kéo của tơ từ kén được tạo ra bởi một số tằm chuyển gen. Tuy nhiên, lớp serixin không bị loại trước khi thử nghiệm cơ học, và bước khử keo này là cần thiết trong việc xử lý kén để sản xuất tơ thương mại. Vì vậy, nếu kén đã được xử lý theo cách thông thường, protein tơ nhện/tằm tái tổ hợp có thể bị loại và tơ thu được có thể không có tính chất cơ học được cải thiện.

Tằm chuyển gen tạo ra các protein tơ nhện được báo cáo như thành phần tương đối nhỏ của các nghiên cứu khác tập trung vào việc tái tạo tơ từ các protein tơ hoà tan trong các dung môi hexaflo. Tuy nhiên, nghiên cứu này mô tả hai tằm chuyển gen được tạo ra bằng vectơ *piggyBac* mã hoá các trình tự “giống tơ” tổng hợp cực ngắn của các protein spidroin-1 ampullate lớn *Nephila clavipes* hoặc tơ hình roi. Cả hai peptit giống tơ được gắn trong các miền *hfc* tận cùng *N* và *C*. Thử nghiệm cơ học cho thấy tơ được tạo bởi các tằm chuyển gen này có độ bền kéo lớn hơn (41-73MPa), và không có thay đổi về độ đàn hồi. Các nhà nghiên cứu cũng báo cáo là các thay đổi tương đối nhỏ được quan sát thấy trong tính chất cơ học của tơ composít của chúng phản ánh mức độ hợp nhất protein tái tổ hợp thấp. Cũng có thể là trình tự peptit giống tơ nhện đặc trưng được sử dụng trong các cấu trúc này và/hoặc các kích thước nhỏ của chúng có thể, ít nhất một phần, giải thích cho các thay đổi tương đối nhỏ về tính chất cơ học của tơ composít được tạo ra bởi tằm chuyển gen này.

Tằm chuyển gen và tơ composít này là các nhân tố đầu tiên tạo ra các dòng tằm chuyển gen mà tạo ra tơ composít chứa protein tơ tằm/nhện khảm được hợp nhất ổn định cải thiện đáng kể tính chất cơ học của chúng. Tơ nhện/tơ tằm composít được tạo ra bởi các dòng tằm chuyển gen này thậm chí còn bền hơn tơ nhện tơ kéo tự nhiên. Trong các nhân tố khác, điều này có thể ít nhất một phần do sử dụng trình tự tơ nhện tổng hợp 2,4 kbp A2S8₁₄ mã hoá các môtip đàn hồi (GPGGA)₄ (SEQ ID NO: 6) giống hình roi và môtip tinh thể spidroin-2 ampullate lớn [alanin₈ liên kết] ("alanin₈" được bọc lộ như SEQ ID NO: 5). Protein tơ nhện tổng hợp tương đối lớn này có thể được kéo thành sợi nhờ đùn sau khi được tạo ra trong *E. coli*, cho biết nó duy trì khả năng tự nhiên để gắn vào

trong tơ. Tuy nhiên, protein này có thể được biểu hiện phối hợp với và có thể phải tương tác với các protein fhc, flc, và fhx của tằm nội sinh để được hợp nhất vào trong tơ. Vì vậy, trình tự tơ nhện A2S8₁₄ được gắn trong các miền fhc tận cùng N và C để định hướng quá trình gắn. Cùng với khả năng của trình tự khởi đầu fhc kích thích biểu hiện của chúng trong lân cận không gian và thời gian đối với các protein tơ tằm nội sinh, các đặc điểm này có thể ít nhất một phần giải thích cho khả năng của protein tơ tằm/nhện tham gia trong quá trình lắp ráp tơ composit và góp phần đáng kể vào tính chất cơ học của chúng.

Ví dụ 12 – Cấu trúc vectơ *piggyBac* và phân khuếch đại PCR của các thành phần của vectơ *piggyBac*

Một số mảnh gen được phân lập nhờ các phản ứng chuỗi polymeraza với ADN hệ gen được phân lập từ tuyến tơ của giống *Bombyx mori* P50/Daizo và các đoạn môi đặc hiệu gen được thể hiện trong Bảng 4. Các mảnh này chứa trình tự khởi đầu chính fhc và yếu tố tăng cường trước (MP-UUE), hai phiên bản của trình tự khởi đầu cơ bản fhc (BP) và miền đầu tận cùng N (NTD; exon 1/intron 1/exon 2) có các vị trí giới hạn gắn bên sườn ở vị trí 5' và 3' khác nhau, miền đầu tận cùng C fhc (CTD; trình tự mã hoá 3' và tín hiệu poly A), và EGFP. Trong mỗi trường hợp, các sản phẩm khuếch đại được làm sạch bằng gel, và các mảnh ADN có kích thước mong muốn được cắt ra và được phục hồi. Tiếp theo, các mảnh MP-UUE fhc, CTD fhc, và EGFP được phân lập dòng vào trong pSLfa1180fa, hai mảnh NTD khác nhau được phân lập dòng vào trong pCR4-TOPO (Invitrogen Corporation, Carlsbad, CA), và các thể biến nạp *E. coli* chứa các sản phẩm khuếch đại chính xác được nhận dạng bởi việc lập bản đồ giới hạn và xác định bằng cách giải trình tự. Các mảnh này sau đó được sử dụng để lắp ráp các vectơ *piggyBac* được sử dụng trong nghiên cứu này như sau. Trình tự tơ nhện A2S8₁₄ tổng hợp được cắt khỏi tiền plasmit pBluescript SKII+ bằng BamHI và BspEL, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước CTD trong plasmit trung gian pSL được mô tả ở trên. Bước này tạo ra plasmit được gọi là pSL-nhện 6-CTD. Mảnh NotI/BamHI sau đó được cắt khỏi một plasmit trong số các plasmit trung gian pCR4-TOPO-NTD được mô tả ở trên, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước trình tự nhện 6-CTD trong pSL-nhện 6-CTD để tạo ra pSL-NTD-nhện 6-CTD. Đồng thời, mảnh NotI/XbaI được cắt khỏi plasmit trung gian pCR4-TOPO-NTD khác được mô tả ở trên, được làm sạch bằng gel, được

phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước gen khuếch đại EGFP trong plasmit trung gian pSL-EGFP được mô tả ở trên. Việc này tạo ra plasmit chứa mảnh NTD-EGFP được cắt ra bằng NotI và BamHI và được tách dòng phụ vào trong các vị trí tương ứng trước các trình tự nhện 6-CTD trong pSL-nhện 6-CTD. Mảnh MP-UEE sau đó được cắt bằng SfiI và NotI khỏi plasmit trung gian pSL được mô tả ở trên, được làm sạch bằng gel, được phục hồi, và được tách dòng phụ vào trong các vị trí tương ứng trước các trình tự NTD-nhện 6-CTD và NTD-EGFP- nhện 6-CTD trong hai plasmit pSL trung gian khác nhau được mô tả ở trên. Cuối cùng, các cắt xet MP-UEE-NTD- A2S8₁₄-CTD hoặc MP-UEE- NTD-EGFP-A2S8₁₄-CTD được lắp ráp hoàn chỉnh được cắt bằng AseI và FseI khỏi các plasmit pSL cuối cùng tương ứng và được tách dòng phụ vào trong các vị trí tương ứng của pBAC[3XP3-DsRedaf] (Horn, *et al.* (2002), *Insect Biochem. Mol. Biol.*, 32:1221-1235). Bước tách dòng phụ cuối cùng này tạo ra hai vector *piggyBac* riêng rẽ được dành cho nhện 6 và nhện 6-EGFP để chỉ sự không có hoặc có gen đánh dấu EGFP. Bảng dưới đây cung cấp danh mục một số thành phần chính của các vector *piggyBac* được sử dụng. Bảng 4 bộc lộ lần lượt SEQ ID NO 7-17, theo thứ tự xuất hiện.

Bảng 4						
#	Tên	Trình tự (5' đến 3')	Vị trí giới hạn bổ sung	Khuôn mẫu ADN	Tổ hợp đoạn môi cho PCR	Sản phẩm khuếch đại & kích thước
1	Trình tự khởi đầu chính (SP)	TAAC TCGAGGCTCAAAGC CTCATCCCAATTTGGAG	5' XhoI			Trình tự khởi đầu chính fhc
2	Trình tự khởi đầu chính (ASP)	ATACCGCGGTGCAGAAGA CAAGCCATCGCAACGGTG	3' SacII		1 & 2	-5,000 đến -3,844 (1,157bp)
3	UEE (SP)	ATACCGCGGAAAGATGTT TTGTACGGAAAGTTTGAA	5' SacII		3 & 4	Trình tự tăng cường fhc -1,659 đến -

						1,590 (70bp)
4	UEE (ASP)	TTAGCGGCCGCGGAACCC TAAACATTGTTACGTTAC GTTACTTG	3' NotI	ADN hệ gen <i>B.</i> <i>mori</i>		
5	Trình tự khởi đầu fhc +NTD (SP)	TAAGCGGCCGCGGGAGAA AGCATGAAGTAAGTTCTT TAAATATTACAAAAA	5' NotI		5 & 6 5 & 7 (-) (+)	Các catxet biểu hiện nhện 6 EGFP (-) hoặc (+)
6	Trình tự khởi đầu fhc+ NTD (ASP)	ATAGGATCCACGACTGCA GCACTAGTGCTGCTGAAA TCGC	3' BamHI			Trình tự khởi đầu cơ bản fhc & 5' cds
7	Trình tự khởi đầu fhc+ NTD (ASP cho EGFP)	ATATCTAGAACGACTGCA GCACTAGTGCTGCTGAAA TCGC	3' XbaI			+62,118 đến +63,816 (1,744bp)
8	EGFP (SP)	CAATCTAGACGTGAGCAA GGGCGAGGAGCTGTTCAC C	5' XbaI	ADN plasmid pEGFP- N1	8 & 9	EGFP (720bp)
9	EGFP (ASP)	TAAGGATCCAGCTTGTAC AGCTCGTCCATGCCGAGA G	3' BamHI			
10	CTD fhc (SP)	ATACCCGGGAAGCGTCAG TTACGGAGCTGGCAG	5' XmaI	ADN hệ gen <i>B.</i> <i>mori</i>	10 & 11	Fhc 3' cds & tín hiệu poly- A
11	CTD fhc (ASP)	CAAGCTGACTATAGTATT CTTAGTTGAGAAGGCATA C	3' SalI			+79,021 đến +79,500 (480bp)

Ví dụ 13 – Tách dòng masp

Ví dụ này chứng minh tính hữu dụng theo sáng chế nhờ việc cung cấp cấu trúc di truyền chứa vùng NTD trong plasmit, và cụ thể, plasmit pXLBacII ECFP.

Các dòng dương tính tiềm năng chứa vùng NTD có plasmit pXLBacII ECFP được thể hiện nhờ việc sàng lọc quần thể bằng PCR.

Masp cấu trúc di truyền dùng cho pXLBacII-ECFP NTD CTD maspX16 (10.458bp) (Fig.12A) và pXLBacII-ECFP NTD CTD maspX24 (11.250bp) (Fig.12B) được tạo thành.

Rõ ràng đối với người có hiểu biết trung bình trong lĩnh vực kỹ thuật này là các cải biến và các thay đổi khác nhau có thể được thực hiện đối với sáng chế mà không đi trệch khỏi phạm vi của sáng chế. Vì vậy, sáng chế bao gồm cả các cải biến và các thay đổi theo sáng chế miễn là chúng nằm trong phạm vi của các yêu cầu bảo hộ đính kèm.

DANH MỤC THAM KHẢO

Các tài liệu tham khảo hiện tại được kết hợp cụ thể ở đây bằng cách tham khảo.

1. Berghammer, A., Bucher, G., Maderspacher, F., và Klingler, M. (1999), *Dev. Genes Evol.*, 209: 382-389.
2. Birnboim, H. C., và Doly, J. (1979), *Nucl. Acids Res.*, 7: 1513-1523.
3. Brooks, A. E., Creager, M., và Lewis, R. V. (2005), *Altering the Mechanics of Spider Silk Through Methanol Post-spin Draw*. In "Biomedical Sciences Instrumentation", tập 41, trang 1-6.
4. Cary, L. C., et al. (1989), *Virology*, 172: 156-169.
5. Choudary, P. V., Kamita, S. G., và Maeda, S. (1995), *Expression of foreign genes in Bombyx mori larvae using baculovirus vectors*. In "Baculovirus expression protocols" (C. D. Richardson, Ed.), tập 39, trang 243-264. Humana Press, Clifton, N.J.
6. Colgin, M., và Lewis, R. V. (1998), *Protein Science*, 7: 667-672.
7. Denny, M. W. (1980), *Symp. Soc. Exp. Biol.*, 34: 247-272.
8. Dooling, D. (2005), *Growing your own spare parts: NASA assists ligament replacement research*.
9. Elick, T. A., Bauser, C. A., Principe, N. M., và Fraser, M. J., Jr. (1996), *Genetica*, 97: 127-139.
10. Fahnestock, S. R., và Bedzyk, L. A. (1997), *Appl. Microbiol. Biotechnol.*, 47: 33-39.
11. Fraser, M. J. (2000), *The TTAA-specific family of transposable elements*. In "Insect Transgenesis: Methods và Applications." (A. A. James, và A. H. Handler, Eds.). CRC Press, Orlando.
12. Fraser, M. J., Brusca, J. S., Smith, G. E., và Summers, M. D. (1985), *Virology*, 145: 356-361.
13. Fraser, M. J., Cary, L., Boonvisudhi, K., và Wang, H. G. (1995), *Virology*, 211: 397-407.

14. Fraser, M. J., Smith, G. E., và Summers, M. D. (1983), *J. Virol.*, 47: 287-300.
15. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., và Lewis, R. (2001), *Science*, 291: 2603-2605.
16. Gosline, J. M., Denny, M. W., và DeMont, M. E. (1984), *Nature* 309: 551-552.
17. Handler, A. M., và Gomez, S. P. (1995), *Mol. Gen. Genet.*, 247: 399-408.
18. Handler, A. M., và Gomez, S. P. (1996), *Genetics*, 143: 1339-1347.
19. Handler, A. M., và Harrell, R. A., 2nd (1999), *Insect Mol. Biol.*, 8: 449-457.
20. Handler, A. M., và Harrell, R. A., 2nd (2001), *Insect Biochem. Mol. Biol.*, 31: 199-205.
21. Handler, A. M., McCombs, S. D., Fraser, M. J., và Saul, S. H. (1998), *Proc. Natl. Acad. Sci. U. S. A.*, 95: 7520-7525.
22. Hayashi, C. Y., và Lewis, R. V. (2000), *Science*, 287: 1477-1479.
23. Hayashi, C. Y., Shipley, N. H., và Lewis, R. V. (1999), *Int. J. Biol. Macromol.*, 24: 271-275.
24. Hinman, M. B., và Lewis, R. V. (1992), *J. Biol. Chem.*, 267: 19320- 19324.
25. Holland, C., Terry, A. E., Porter, D., và Vollrath, F. (2006), *Nat. Mater.*, 5: 870-874.
26. Horard, B., Mange, A., Pelissier, B., và Couble, P. (1994), *Insect Mol. Biol.*, 3: 261-265.
27. Horn, C., Jaunich, B., và Wimmer, E. A. (2000), *Dev. Genes Evol.*, 210: 623-629.
28. Huemmerich, D., et al. (2004), *Curr., Biol.* 14: 2070-2074.
29. Imamura, M., et al. (2003), *Genetics*, 165: 1329-1340.
30. Inoue, S., et al. (2005), *Insect Biochem. Mol. Biol.*, 35: 51-59.

31. Inoue, S., et al. (2000), *J. Biol. Chem.*, 275: 40517-40528.
32. Lazaris, A., et al. (2002), *Science* 295: 472-476.
33. Lewis, R. V., et al. (1996), *Prot. Expr. Purif.*, 7: 400-406.
34. Lobo, N., Li, X., và Fraser, M. J., Jr. (1999), *Mol. Gen. Genet.*, 261: 803-810.
35. Maeda, S., et al. (1985), *Nature*, 315: 592-594.
36. Mori, K., et al. (1995). *J. Mol. Biol.* 251: 217-228.
37. O'Brochta, D. A., Gomez, S. P., và Handler, A. M. (1991), *Mol. Gen. Genet.* 225: 387-394.
38. Peloquin, J. J., et al. (2000), *Insect Mol. Biol.*, 9: 323-333.
39. Sambrook, J., Fritsch, E. F., và Maniatis, T. (1989), "Molecular Cloning: A Laboratory Manual." 2nd edition ed. Cold Spring Harbor Press, Cold Spring Harbor, New York.
40. Scheller, J., Guhrs, K. H., Grosse, F., và Conrad, U. (2001), *Nat. Biotechnol.*, 19: 573-577.
41. Southern, E. M. (1975), *J. Mol. Biol.*, 98: 503-517.
42. Takei, F., et al. (1984), *J. Cell Biol.*, 99: 2005-2010.
43. Tamura, T., et al. (2000), *Nat. Biotechnol.*, 18: 81-84.
44. Thibault, S. T., Luu, H. T., Vann, N., và Miller, T. A. (1999), *Insect Mol. Biol.*, 8: 119-123.
45. Thomas, J. L., et al. (2002), *Insect Biochem. Mol. Biol.*, 32, 247-253.
46. Tomita, M., et al. (2003), *Nat. Biotechnol.*, 21: 52-56.
47. Towbin, H., et al. (1979), *Proc. Natl. Acad. of Sci. U.S.A.*, 76: 4350-4354.
48. Urry, D. W. (2002), *Philosophical Transactions of the Royal Society of London B* 357:169-184.
49. Wang, H. G., ADN Fraser, M. J. (1993), *Insect Mol. Biol.*, 1: 109-116.
50. Wang, H. H., Fraser, M. J., và Cary, L. C. (1989), *Gene.*, 81: 97-108.

51. Wong Po Foo, C., et al. (2006), Appl. Phys. A., 82: 223-233.
 52. Wurm, F. M. (2003), Nat. Biotechnol., 21: 34-35.
 53. Xu, M., và Lewis, R. V. (1990), Proc. Natl. Acad. Sci. U.S.A., 87: 7120-7124.
 54. Yamao, M., et al. (1999), Genes. Dev., 13: 511-516.
 55. Yun, et al. (2001), "Altering fibrin heavy chain gene of silkworm *Bombyx mori* by homologous recombination," Shengwu Huaxe yu Shengwu Wuli Xuebao 33(1): 112-116.
 56. GenBank Acc. No. AF226688, Zhou, et al. "*Bombyx mori* fibroin heavy chain Fib-H (fib-H) gene, complete cds.," US Natl. Library of Medicine, Bethesda, MD, USA, 19 tháng 6, 2000.
 57. Zhao, et al. (2001), Acta Biochimica et Biophysica Sinica, 33(1): 112-116.
 58. Zhang, et al. (1999), Acta Biochimica et Biophysica Sinica, 31(2): 119-123.
 59. Zhou, C.Z., et al. (2000), Nucleic Acids Res., 28. (12): 2413-2419.
 60. Tomita, M., et al. (2003), Nat. Biotechnol., 21 (1): 52-56.
 61. Yoshizato, Katsutoshi, "A Proposal for Application of Recombinant Insects (Kumikaetai Konchu Riyo Eno Teigen)", Sanshi Konchuken Shiryo, số 28, trang 93-95.
 62. Toshiki, et al. (2000), Nature Biotechnology, 16: 81-85.
 63. Okano, et al. (2000), Journal of Interferon và Cytokine Research, 20: 1015-1022.
 64. Xiao-Hui, et al. (2000), Acta Pharmacol. Sin., 21 (9): 797-801.
 65. Ishihara, et al. (1999), Biochimica et Biophysica Acta, 1451: 48-58.
 66. T. Tamura, "Construction và utilization of transgenic silkworm using gen nhảy", Fiber Preprints, Japan, tập 56, số 2, 2001, trang 38-41.
- A. Yanai, et al. (2002), Research Journal of Food và Agriculture, 25 (2): 30-33.

67. T. Tamura, et al. (2000), *Agriculture và Horticulture*, 75 (8): 17-24.
68. Yoshizato, Katsutoshi (2001), "A Proposal for Application of Recombinant Insects (Kumikaetai Konchu Riyo Eno Teigen)", *Sanshi Konchuken Shiryo*, 28: 93 - 95.
69. United States Patent 7,674,882 - Kaplan, et al.
70. United States Patent 7,659,112 - Hiramatsu, et al.
71. United States Patent 7,521,228 - Lewis, et al.
72. United States Patent 6,268,169 - Fahnestock.
73. United States Patent 5,994,099 - Lewis.
74. United States Patent 5,989,894 - Lewis.
75. United States Patent 5,756,677 - Lewis
76. United States Patent 5,733,771 - Lewis.
77. Kluge, J.A., Rabotyagova, O., Leisk, G.G. & Kaplan, D.L. (2008), *Trends Biotechnol.*, 26:244-251.
78. Scheibel, T. (2004), *Microb. Cell. Fact.* 3, 14.
79. MacIntosh, A.C., Kearns, V.R., Crawford, A. & Hatton, P.V. (2008), *J. Tiss. Engr. Reg. Med.*, 2:71-80.
80. Gosline, J.M., Guerette, P.A., Ortlepp, C.S. & Savage, K.N. (1999), *J. Exp. Biol.*, 202:3295-3303.
81. Lewis, R.V. (2006), *Chem. Rev.*, 106:3762-3774.
82. Hardy, J.G., L.M., R. & T.R. (2008), S., *Polymer*, 49:4309-4327.
83. Teulé, F., et al. (2007), *J. Mat. Sci.*, 42:8974-8985.
84. Teulé, F., et al. (2009), *Nat. Protoc.* 4:341-355.
85. Fahnestock, S.R. & Irwin, S.L. (1997), *Appl. Microbiol. Biotechnol.*, 47:23-32.
86. Fahnestock, S.R. & Bedzyk, L.A. (1997), *Appl. Microbiol. Biotechnol.*, 47:33-39.

87. Zhang, Y., et al. (2008), *Mol. Biol. Rep.* 35:329-335.
88. Miao, Y., et al. (2006), *Appl. Microbiol. Biotechnol.*, 71:192-199.
89. Kato, T., Kajikawa, M., Maenaka, K. & Park, E.Y. (2010), *Appl. Microbiol. Biotechnol.*, 85:459-470.
90. Royer, C., et al. (2005), *Transgenic Res.*, 14:463-472.
91. Kojima, K., et al. (2007), *Biosci. Biotechnol. Biochem.* 71, 2943-2951.
92. Kurihara, H., Sezutsu, H., Tamura, T. & Yamada, K. (2007), *Biochem. Biophys. Res. Commun.*, 355:976-980.
93. Shimizu, K., et al. (2007), *Insect Biochem. Mol. Biol.*, 37:713-725.
94. Yanagisawa, S., et al. (2007), *Biomacromolecules*, 8:3487-3492.
95. Wen, H., et al. (2010), *Mol. Biol. Rep.*, 37:1815-1821.
96. Zhu, Z., et al. (2010), *J. Biomater. Sci. olyn. Ed.*, 21:395-411.
97. Sehna, F. & Akai, H. (1990), *Int. Insect Morph. Embryol.*, 19:79-132.
98. Horn, C., et al. (2002), *Insect Biochem. Mol. Biol.*, 32:1221-1235.
99. Yamada, H., Nakao, H., Takasu, Y. & Tsubouchi, K. (2001), *Mat. Sci. Engr. C*, 14: 41- 46.
100. United States Patent 5,728,810 Lewis.

YÊU CẦU BẢO HỘ

1. Axit nucleic mã hóa polypeptit tơ nhện khảm, trong đó polypeptit khảm này bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C,
 - (a) miền đầu tận cùng N (NTD) của polypeptit tơ chuỗi nặng fibroin *Bombyx mori* (fhc);
 - (b) 14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L) và môtip bền (S), trong đó môtip đàn hồi (E) là GPGGA (SEQ ID NO: 2) và môtip bền (S) là GGPSGPGSAAAAAAAAA (SEQ ID NO: 3); và
 - (c) miền đầu tận cùng C (CTD) của polypeptit tơ *Bombyx mori* fhc.
2. Axit nucleic theo điểm 1, trong đó axit nucleic này mã hóa polypeptit tơ nhện khảm được gắn bên sườn ở đầu 5' của nó một axit nucleic chứa trình tự khởi đầu chính, yếu tố tăng cường ngược dòng (UEE), và trình tự khởi đầu cơ bản của gen chuỗi nặng *B. mori* fibroin (fhc), và được gắn bên sườn ở đầu 3' của nó một axit nucleic chứa phần kết thúc phiên mã và các điểm polyadenyl hóa của gen chuỗi nặng *B. mori* fibroin (fhc).
3. Axit nucleic chứa các trình tự dưới đây, theo hướng từ 5' đến 3':
 - (a) trình tự bao gồm đoạn lặp đầu tận cùng đầu tiên của gen nhảy;
 - (b) trình tự điều hòa thứ nhất chứa trình tự khởi đầu chính, yếu tố tăng cường ngược dòng (UEE), và trình tự khởi đầu cơ bản của gen chuỗi nặng *B. mori* fibroin (fhc), trong đó trình tự khởi đầu đã nêu được liên kết có điều khiển với
 - (c) trình tự axit nucleic mã hóa polypeptit tơ nhện khảm, trong đó polypeptit khảm này bao gồm, theo hướng từ đầu tận cùng N đến đầu tận cùng C:
 - (i) miền đầu tận cùng N (NTD) của polypeptit tơ nhện chuỗi nặng *B. mori* fibroin (fhc);

- (ii) 14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L), và môtip bền (S);
trong đó môtip đàn hồi (E) là GPGGA (SEQ ID NO: 2);
trong đó môtip bền (S) là GGPSGPGSAAAAAAAAA (SEQ ID NO: 3); và
 - (iii) miền đầu tận cùng C (CTD) của polypeptit tơ nhện *Bombyx mori* fhc;
 - (d) trình tự điều hòa thứ hai chứa phần kết thúc phiên mã và các điểm polyadenyl hóa của gen chuỗi nặng *B. mori* fibroin (fhc); và
 - (e) trình tự chứa đoạn lặp đầu tận cùng thứ hai của gen nhảy;
trong đó ít nhất một trong các trình tự khởi đầu đã nêu hoạt động trong các tế bào hoặc mô *B. mori* được biến nạp;
trong đó ít nhất một trong các đoạn lặp đầu tận cùng đã nêu tạo điều kiện dễ dàng cho sự chuyển vị của các trình tự (b), (c) và (d) vào trong hệ gen của tằm *B. mori* bị biến đổi.
4. Vectơ chứa axit nucleic theo điểm bất kỳ trong số các điểm 1, 2, và 3.
 5. Vectơ theo điểm 4, trong đó vectơ này bao gồm đoạn lặp đầu tận cùng thứ nhất và đoạn lặp đầu tận cùng thứ hai của gen nhảy.
 6. Vectơ theo điểm 5, trong đó đoạn lặp đầu tận cùng thứ nhất và thứ hai của gen nhảy này thu được từ gen nhảy *piggyBac*.
 7. Vectơ được chọn từ nhóm bao gồm:
 - (a) vectơ pXLBacII-ECFP NTD CTD maspX16, bao gồm trình tự được mô tả trong SEQ ID NO: 38; và
 - (b) vectơ pXLBacII-ECFP NTD CTD maspX24, bao gồm trình tự được mô tả trong SEQ ID NO: 39.
 8. Vectơ được chọn từ nhóm bao gồm:

- (a) vector pSL-nhện#4 bao gồm trình tự được mô tả trong SEQ ID NO: 30;
- (b) vector pSL-nhện#4+GFP bao gồm trình tự được mô tả trong SEQ ID NO: 32;
- (c) vector pSL-nhện#6 bao gồm trình tự được mô tả trong SEQ ID NO: 34; và
- (d) vector pSL-nhện#6+GFP bao gồm trình tự được mô tả trong SEQ ID NO: 36.

9. Phương pháp tạo ra tầm chuyển gen *Bombyx mori* biểu hiện một cách ổn định polypeptit tơ nhện khảm thích hợp để gắn vào trong tơ nhện khảm, trong đó phương pháp này bao gồm bước:

- (a) chèn vector chứa axit nucleic mã hóa polypeptit tơ nhện khảm, trong đó axit nucleic này mã hóa polypeptit bao gồm:
 - (i) miền đầu tận cùng N (NTD) của polypeptit tơ *Bombyx mori* fhc;
 - (ii) 14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L) và môtip bền (S); trong đó môtip đàn hồi (E) là GPGGA (SEQ ID NO: 2) và môtip bền (S) là GGPSGPGSAAAAAAAAA (SEQ ID NO: 3); và
 - (iii) miền đầu tận cùng C (CTD) của polypeptit tơ *Bombyx mori* fhc vào trứng *Bombyx mori* đột biến để tạo ra trứng *Bombyx mori* được tiêm;
- (b) để cho trứng nở trong điều kiện ủ thích hợp để tạo ra ấu trùng;
- (c) cho phép ấu trùng trưởng thành trong các điều kiện ủ thích hợp; và
- (d) lựa chọn tầm chuyển gen *Bombyx mori*.

10. Phương pháp theo điểm 9, trong đó axit nucleic còn bao gồm axit nucleic mã hóa polypeptit đánh dấu để tạo điều kiện thuận lợi cho việc sàng lọc hoặc chọn lọc *Bombyx mori* chuyển gen.

11. Phương pháp theo điểm 9, trong đó vector đã nêu được chọn từ nhóm bao gồm:

- (a) vector pSL-nhận#4 bao gồm trình tự được mô tả trong SEQ ID NO: 30;
- (b) vector pSL-nhận#4+GFP bao gồm trình tự được mô tả trong SEQ ID NO: 32;
- (c) vector pSL-nhận#6 bao gồm trình tự được mô tả trong SEQ ID NO: 34; và
- (d) vector pSL-nhận#6+GFP bao gồm trình tự được mô tả trong SEQ ID NO: 36.

12. Phương pháp tạo ra tơ nhện khảm bao gồm:

- (a) chuẩn bị tấm chuyển gen bằng cách sử dụng vector chứa axit nucleic mã hóa polypeptit tơ nhện khảm, trong đó polypeptit khảm này bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C,
 - (i) miền đầu tận cùng N (NTD) của polypeptit tơ *Bombyx mori* fhc,
 - (ii) 14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L) và môtip bền (S);
trong đó môtip đàn hồi (E) là GPGGA (SEQ ID NO: 2) và môtip bền (S) là GGPSGPGSAAAAAAAAA (SEQ ID NO: 3); và
 - (iii) miền đầu tận cùng C (CTD) của polypeptit tơ *Bombyx mori* fhc vào trong trứng *Bombyx mori* đột biến để tạo ra trứng *Bombyx mori* được tiêm;
- (b) để cho tấm chuyển gen tạo ra kén trong các điều kiện sinh lý tự nhiên thích hợp đối với tấm; và

- (c) thu thập và chiết một hoặc nhiều tơ nhện khảm từ kén đã nêu.
13. Phương pháp theo điểm 12, trong đó tầm chuyển gen đã nêu là tầm *Bombyx mori* chuyển gen.
14. Polypeptit tơ nhện khảm bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C:
- (i) miền đầu tận cùng N (NTD) của polypeptit tơ *Bombyx mori* fhc,
 - (ii) 14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L) và môtip bền (S);
trong đó môtip đàn hồi (E) là GPGGA (SEQ ID NO: 2) và môtip bền (S) là GGPSGPGSAAAAAAAAA (SEQ ID NO: 3); và
 - (iii) miền đầu tận cùng C (NTD) của polypeptit tơ *Bombyx mori* fhc.
15. Polypeptit tơ nhện khảm theo điểm 14, trong đó 4 đến 16 bản sao của môtip đàn hồi này được chọn từ nhóm bao gồm:
- (GPGGA)₄, được ký hiệu là A1, như được nêu trong SEQ ID NO: 40;
 - (GPGGA)₈, được ký hiệu là A2, như được nêu trong SEQ ID NO: 41; và
 - (GPGGA)₁₆, được ký hiệu là A4, như được nêu trong SEQ ID NO: 43.
16. Polypeptit tơ nhện khảm theo điểm 15, trong đó ít nhất một trong các đoạn lặp đã nêu của các môtip tơ nhện được chọn từ nhóm bao gồm:
- trình tự chứa một bản sao của SEQ ID NO: 43 và một bản sao của SEQ ID NO: 3, và;
 - trình tự chứa một bản sao của SEQ ID NO: 41 và một bản sao của SEQ ID NO: 3.
17. Polypeptit tơ nhện khảm theo điểm 15, trong đó polypeptit tơ nhện khảm còn bao gồm tác nhân sinh trưởng.
18. Tơ composit bao gồm polypeptit tơ nhện khảm theo điểm bất kỳ trong số các điểm 14 và từ 15 đến 17, trong đó tơ đã nêu có độ bền kéo lớn hơn so với sợi tơ tầm tự nhiên.

19. Tơ composít theo điểm 18, trong đó tơ composít còn bao gồm một hoặc nhiều tác nhân trị liệu.

20. Tơ composít theo điểm 19, trong đó ít nhất một trong các tác nhân trị liệu đã nêu được chọn từ nhóm bao gồm tác nhân chống nhiễm khuẩn, tác nhân hóa trị liệu, tác nhân chống thải loại, tác nhân giảm đau, tác nhân chống viêm, hormon và tác nhân sinh trưởng.

21. Tơ composít theo điểm 20, trong đó ít nhất một trong các tác nhân trị liệu đã nêu là tác nhân sinh trưởng.

22. Tầm chuyển gen chứa axit nucleic theo điểm bất kỳ trong số các điểm 1, 2 và 3.

23. Tầm chuyển gen theo điểm 22, trong đó tầm có khả năng tạo ra polypeptit tơ nhện khảm thích hợp để tạo ra tơ khảm có độ bền kéo lớn hơn độ bền kéo của tơ tầm tự nhiên.

24. Tầm chuyển gen theo điểm 22, trong đó tầm đã nêu là tầm *Bombyx mori*.

25. Tầm chuyển gen theo điểm 24, trong đó 4 đến 16 bản sao của môtip đàn hồi được chọn từ nhóm bao gồm:

(GPGGA)₄, được ký hiệu là A1, như được nêu trong SEQ ID NO: 40;

(GPGGA)₈, được ký hiệu là A2, như được nêu trong SEQ ID NO: 41; và

(GPGGA)₁₆, được ký hiệu là A4, như được nêu trong SEQ ID NO: 43.

26. Tầm chuyển gen theo điểm 24, trong đó ít nhất một trong các đoạn lặp đã nêu của các môtip tơ nhện được chọn từ nhóm bao gồm:

trình tự chứa một bản sao của SEQ ID NO: 43 và một bản sao của SEQ ID NO: 3, và;

trình tự chứa một bản sao của SEQ ID NO: 41 và một bản sao của SEQ ID NO: 3.

27. Phương pháp tạo ra tầm *Bombyx mori* chuyển gen biểu hiện một cách ổn định polypeptit tơ nhện khảm thích hợp để gắn vào trong tơ nhện khảm, phương pháp nêu trên bao gồm bước:

- (a) chèn vector chứa axit nucleic mã hóa polypeptit tơ nhện khảm, trong đó axit nucleic này mã hóa polypeptit bao gồm

miền đầu N tận cùng (NTD) của polypeptit tơ *Bombyx mori* fhc;

một hoặc nhiều đoạn lặp chứa một hoặc nhiều môtip đàn hồi tơ nhện;

và

miền đầu tận cùng C (CTD) của polypeptit tơ *Bombyx mori* fhc

vào thể đột biến trứng *Bombyx mori* để tạo ra trứng *Bombyx mori* được tiêm;

trong đó vector đã nêu được chọn từ nhóm bao gồm:

- (i) vector được ký hiệu là pXLBacII-ECFP NTD CTD maspX16 bao gồm trình tự được mô tả trong SEQ ID NO: 38;

- (ii) vector được ký hiệu là pXLBacII-ECFP NTD CTD maspX24 bao gồm trình tự được mô tả trong SEQ ID NO: 39;

- (b) để cho trứng nở trong điều kiện ủ thích hợp để tạo ra ấu trùng;

- (c) cho phép ấu trùng trưởng thành trong các điều kiện ủ thích hợp; và

- (d) lựa chọn tằm *Bombyx mori* chuyển gen.

28. Tằm chuyển gen được tạo ra bằng phương pháp theo điểm 27.

29. Tằm chuyển gen tạo ra sợi bao gồm polypeptit chứa:

14 đến 42 đoạn lặp của các môtip tơ nhện, mỗi đoạn lặp bao gồm theo hướng từ đầu tận cùng N đến đầu tận cùng C có từ 4 đến 16 bản sao của môtip đàn hồi (E), tác nhân liên kết tùy ý (L), và môtip bền (S);

trong đó môtip đàn hồi (E) là GPGGA như được nêu trong SEQ ID NO: 2, và;

trong đó môtip bền (S) là GGPSGPGSAAAAAAAAA như được nêu trong SEQ ID NO: 3.

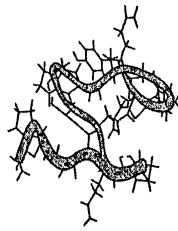


FIG. 4

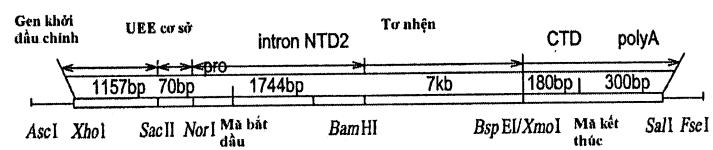


FIG. 5

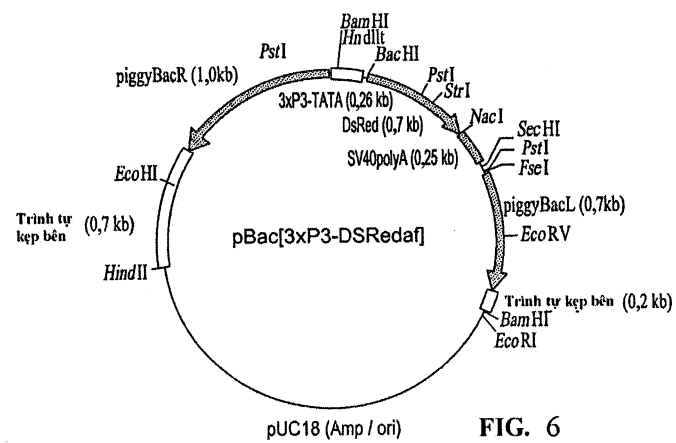
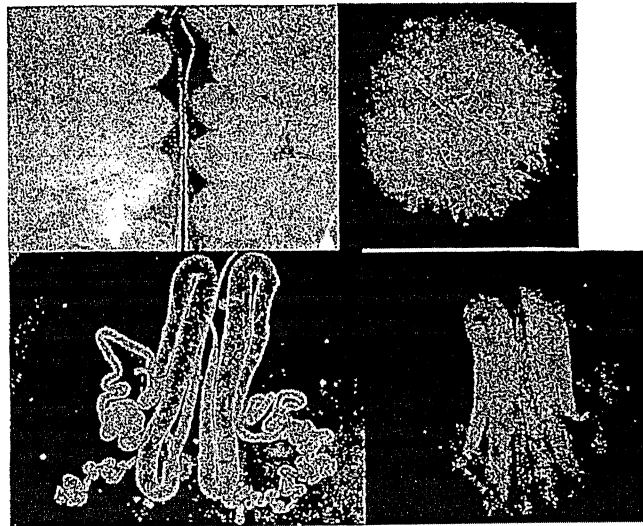
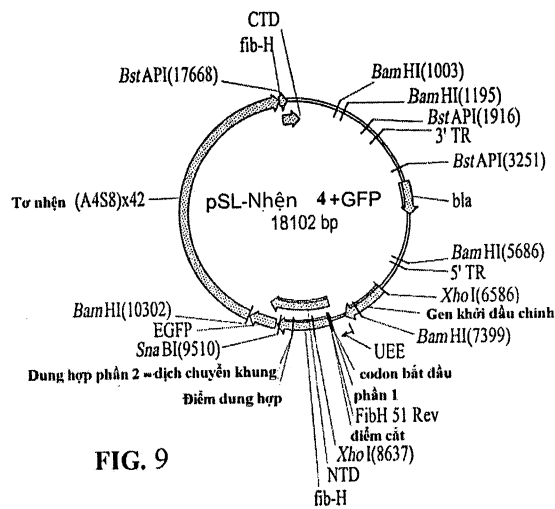
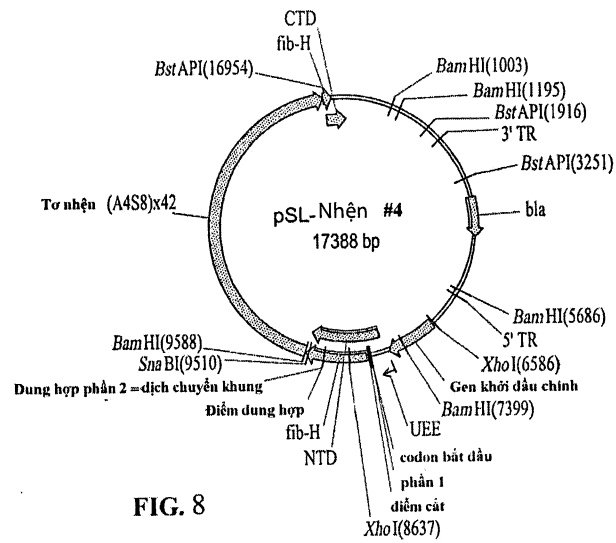


FIG. 6

FIG. 7





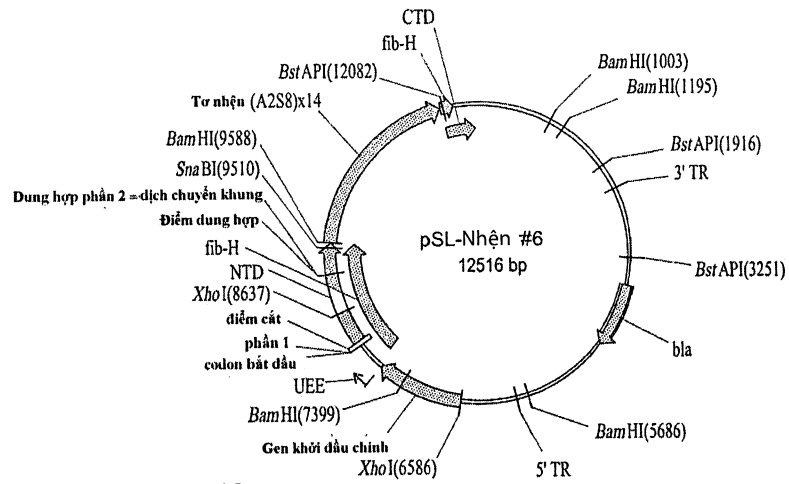


FIG. 10

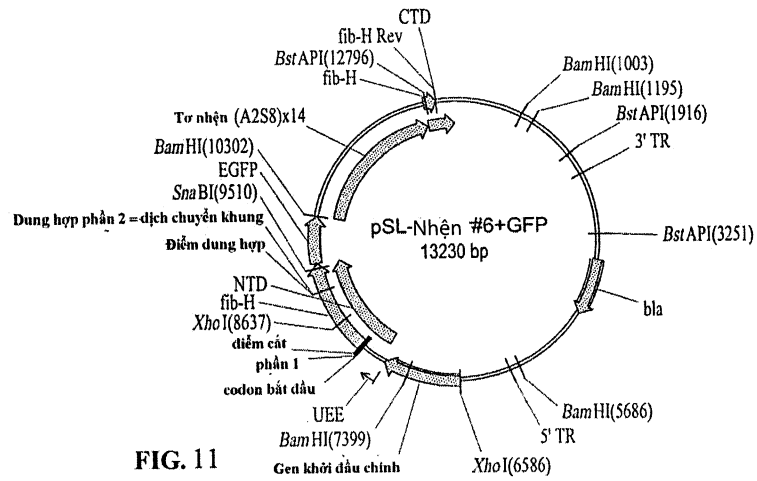


FIG. 11

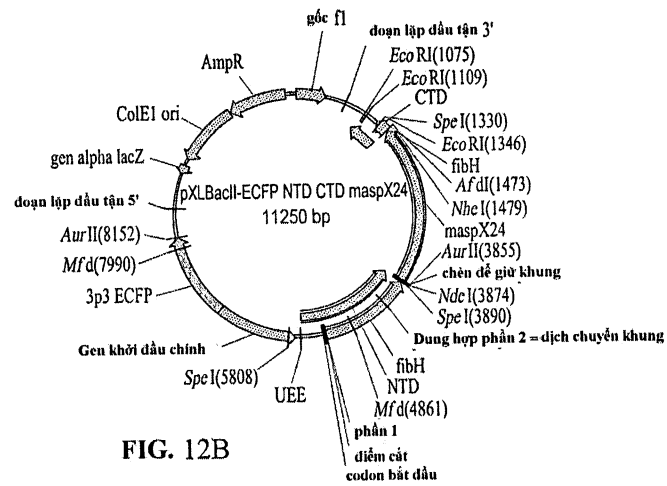
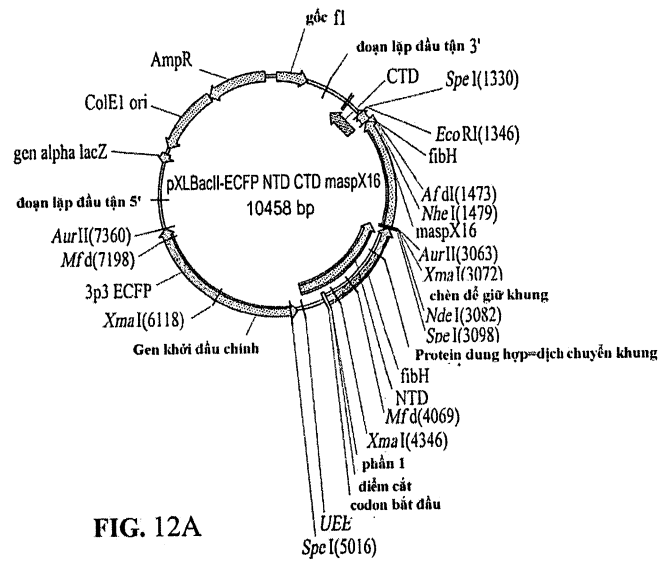


FIG. 13
pSL- Nhện #4

Đặc điểm	Vị trí cặp bazơ *
đoạn lặp đầu tận 3'	nhóm bổ sung (2270..2332)
Bla (Kháng ampicillin)	3633..4493
đoạn lặp đầu tận 5'	nhóm bổ sung (5836..5870)
Gen khởi đầu chính	6586..7751
UEE	7754..7823
Miền đầu tận cùng N (NTD)	7832..9577
Gen Fib - H trong NTD	8152..9575
Tư nhện (A4S8) #42	9597..16901
Miền đầu tận cùng C (CTD)	16908..17387
Gen Fib - H trong CTD	16908..17087

* Khi bố thế được sử dụng trong trường hợp này có nghĩa là trình tự của dấu hiệu là dài AND đối so với trình tự được mô tả ở đây. Nói một cách khác, nó là bố thế ngược của trình tự được thể hiện ở đây.

Số đếm cơ sở gốc bazơ	3766 a	3953 c	5571 g	4098 t		
1	tcgacgtccc	atggccattc	gaattcggcc	ggcctaggcg	cgccgtacgc	gtatcgataa
61	gctttaagat	acattgatga	gtttggacaa	accacaacta	gaatgcagtg	aaaaaatg
121	tttatttggtg	aaatttggtg	tgctattgct	ttatttgtaa	ccattataag	ctgcaataaa
181	caagttaaca	acaacaattg	cattcatttt	atgtttcagg	ttcaggggga	ggtgtgggag
241	gttttttaaa	gcaagtaaaa	cctctacaaa	tgtggtatgg	ctgattatga	tctagagtcg
301	cgcccgctac	aggaacaggt	ggtggcggcc	ctcgggtgcg	tcgtactgct	ccacgatggt
361	gtagtctctcg	ttgtgggagg	tgatgtccag	cttggagtcc	acgtagtagt	agccggggcag
421	ctgcacgggc	ttcttggcca	tgtagatgga	cttgaactcc	accaggtagt	ggccgcccgtc
481	cttcagcttc	agggccttgt	ggatctcgcc	cttcagcacg	ccgtcgcggg	ggtcacggcg
541	ctcgggtggag	gcctcccagc	ccatggtctt	cttctgcatt	acggggccgt	cggaggggaa
601	gttcgcgcga	tgaacttcac	cttgtagatg	aagcagccgt	cctgcaggga	ggagtcctgg
661	gtcacgggtca	ccacgcccgc	gtcctcgaag	ttcatcacgc	gctcccactt	gaagccctcg
721	gggaaggaca	gcttcttgta	gtcggggatg	tcggcggggt	gcttcacgta	caccttggag
781	ccgtactgga	actgggggga	caggatgtcc	caggcggaag	gcagggggcc	gcccttggtc
841	accttcagct	tcacggtgtt	gtggccctcg	tagggggcgg	cctcgcccct	cgcccctcga
901	tctcgaactc	gtggccgttc	acggtgccct	ccatgcgcac	cttgaagcgc	atgaactcct
961	tgatgacgtt	cttggaggag	cgcaccatgg	tggcgaccgg	tggtatcccg	gcccgcggtg
1021	ccgtcgactc	tagcgggtacc	ccgattgttt	agcttgttca	gctgcgcttg	tttatttgct
1081	tagctttcgc	ttagcgacgt	gttcactttg	cttgtttgaa	ttgaattgtc	gctccgtaga
1141	cgaagcgcc	ctattttatc	tccggcggtc	gagggttcga	aatcgataag	cttggatcct
1201	aattgaatta	gctctaattg	aattagtctt	ctaattgaat	tagtctctaa	ttgaattaga
1261	tcccgcggcg	agctcgaatt	aaaccattgt	gggaaccgtg	cgatcaaaca	aacgcgagat
1321	accgggaagt	actgaaaaac	agtcgctcca	ggccagtggt	aacatcgatg	ttttgttttg
1381	acggacccct	tactctcgtc	tcatataaac	cgaagccagc	taagatggta	tacttattat
1441	catcttgtga	tgaggatgct	tctatacagc	aaagtaccgg	taaaccgcaa	atggttatgt
1501	attataatca	aactaaaggc	ggagtggaca	cgtagacca	aatgtgttct	gtgatgacct
1561	gcagtaggaa	gacgaatagg	tggcctatgg	cattattgta	cggaatgata	aacattgcct
1621	gcataaattc	ttttattata	tacagccata	atgtcagtag	caagggagaa	aaggttcaaa
1681	gtcgcacaaa	atttatgaga	aacctttaca	tgagcctgac	gtcatcggtt	atgcgtaagc

```

1741 gtttagaagc tcctactttg aagagatatt tgcgcgataa tatctctaata attttgccaa
1801 atgaagtgcc tggtagatca gatgacagta ctgaagagcc agtaatgaaa aaacgtactt
1861 actgtactta ctgcccctct aaaataaggc gaaaggcaaa tgcattcgtc aaaaaatgca
1921 aaaaagttaa ttgtcgcagag cataaatattg atatgtgcc aagttgtttc tgactgacta
1981 ataagataaa ttgttttcta ttatgtataa gtttaagctaa ttacttattt tataatacaa
2041 catgactgtt tttaaagtac aaaataagtt tatttttgta aaagagagaa tgtttaaaag
2101 ttttgttact ttatagaaga aattttgagt ttttgtttt ttttaataaa taataataca
2161 taaataaatt gtttgttgaa tttattatta gtatgtaagt gtaaatataa taaaacttaa
2221 tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtcaat
2281 tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggtaataata
2341 tagtttctaa tttttttatt attcagcctg ctgctgtgaa tacogtatat ctcaacgctg
2401 tctgtgagat tgcgtattc tagccttttt agtttttcgc tcatcgactt gataattgtc
2461 gacacatttt cgtcgatttg cgttttgatc aaagacttga gcagagacac gtttaatcaac
2521 tgttcaaat gattccatatt aacgatatca acccgatgag tatatggtgc gtaaaataa
2581 ttttttaacc ctcttatact ttgcactctg cgttaatacg cgttcgtgta cagacgtaat
2641 catgttttct ttttttgata aaactcctac tgagtttgac ctcatattag accctcaca
2701 gttgcaaaac gtggcatttt ttaccaatga agaatttaaa gttattttta aaaatttcat
2761 cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaatacg accagttaat
2821 caacgtgtac acagacgcgt cggcaaaaaa cgcgcagccc gacgtgttgg ctaaaattat
2881 taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttctca aaaagttaa
2941 gaccaacaag ttacaggaca ctattaatta tttgattttg cccacttca tttgtggga
3001 tcacaatttt gttatatatt taaaacaaag ctttggcact ggcgtcgtt ttacaacgtc
3061 gtgactggga aaaccttggt gttaccaaac ttaatcgctt tgacgacat cccctttcgt
3121 ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttcccaacag ttgcgcagcc
3181 tgaatggcga atggcgctgt atgcggtatt ttctccttac gcatctgtgc ggtatttca
3241 accgcataat gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc
3301 gacacccgcc aacacccgct gacgcgcctt gacgggcttg tctgctccc gcatccgctt
3361 acagacaagc tgtgaccgtc tccgggagct gcatgtgca gaggttttca cgtcatcac
3421 cgaaacgcgc gagagaaaag ggcctcgtga tacgcctatt tttataggtt aatgtcatga
3481 taataatggt ttcttagacg tcagggtggc cttttcgggg aaatgtgccc ggaaccccta
3541 tttgtttatt ttcttaataa cattcaataa tgtatccgct catgagacaa taacctgat
3601 aaatgcttca ataatttga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgcgc
3661 ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcaaccagaa acgctggtga
3721 aagtaaaaga tgcgtgaagt cagttgggtg cagcagtggt ttacatcgaa ctggatctca
3781 acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt
3841 ttaagtttct gctatgtggc gcggtattat cccgtattga cgcggggcaa gagcaactcg
3901 gtcgcgcgat acactattct cagaatgact tgggttgagta ctccaccagtc acagaaaagc
3961 atctttacga tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata
4021 acactgcggc caacttactt ctgacaaacg tcggaggacc gaaggagcta accgcttttt
4081 tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag
4141 ccataccaaa cgcagagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca
4201 aactattaac tggcgaacta cttactctag cttcccgcca gctcggccct tccggctggc tggttatttg
4261 aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggttatttg
4321 ctgataaatc tggagccggt gagcgtgggt ctcgcgggtt cattgcagca ctggggccag
4381 atggttaagc ctcccgtatc gtagttatct acacgacggg gactcaggca actatggatg
4441 aacgaaatag acagatcgct gagatagggt cctcaactgat taagcattgg taactgtcag
4501 accaagttta ctcatatata ctttagattg atttaaaact tcaattttta tttaaaagga
4561 tctagggtgaa gatccttttt gataatctca tgacaaaat cccttaacgt gagttttcgt
4621 tccactgagc gtcagacccc gtagaaaaga tcaaaaggatc ttcttgagat cctttttttc
4681 tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttggttgc
4741 cggatcaaga gctaccaact ctttttcgga aggtaactgg cttcagcaga gcgcagatac
4801 caaatactgt tcttctagtg tagccgtagt taggccacca cttcaagaac tctgtagcac
4861 cgcctacata cctcgcctct ctaatcctgt taccagtggc tgctgccagt ggcgataagt
4921 cgtgtcttac cgggttgagc tcaagacgat agttaccgga taaggcgag cggctcgggt

```


4981 gaacggggggg ttctgtgcaca cagcccagct tggagcgaaac gacctacacc gaactgagat
 5041 acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaaag gcggacaggt
 5101 atccggtaag cggcagggtc ggaacaggag agcgacagag ggagcttcca ggggaaacg
 5161 cctgatattc ttatagtcct gtcgggtttc gccacctctg acctgagcgt cgatttttgt
 5221 gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg
 5281 ttctctggcct tttgctggcc ttttgcacac atgttcttcc ctgcgttacc ccttgattct
 5341 gtggataacc gttattaccgc ctttgagtga gctgataccg ctgcgcgcag ccgaacgacc
 5401 gagcgacgag agtcagtgag cgaggaagcg gaagagcgcc caatacgcga accgcctctc
 5461 cccgcgcggt ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg
 5521 ggcaagttagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta
 5581 cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca
 5641 ggaaacagct atgacatgat tacgaattcg agctcggtag ccggggatcc tctagagtcg
 5701 acgctcgcgc gacttggttt gccattcttt agcgcgcgct gcgtcacaca gcttgccacc
 5761 aatgtggttt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaaagc
 5821 gcaaatcttt ttaaacctta gaaagatagt ctgcgtaaaa ttgacgcagt cactcttgaa
 5881 atattgtctc ctctttctaa atagcgcgaa tccgtcgtcg tgcatttagg acatctcagt
 5941 cgccgcttgg agctcccgtg aggcgtgctt gtcaatgcgg taagtgtcac tgattttgaa
 6001 ctataacgac cgctgtgagtc aaaaatgcgc atgattatct tttacgtgac ttttaagatt
 6061 taactcatac gataattata ttgttatttc atgttctact taogtgataa cttatttata
 6121 atatatattc ttgttataga tatcgtgact aatatataat aaaaagggta gttctttaga
 6181 cgatgagcat atcctctctg ctcttctgca aagcgatgac gagcttgttg gtgaggattc
 6241 tgacagtgaa atatcagatc acgtaagtga agatgacgta cagagcgata cagaagaagc
 6301 gtttatagat gaggtacatg aagtgacgcc aacgtcaagc ggtagtgaaa tattagacga
 6361 acaaaaatgt attgaacaac caggttcttc atagattctg gttgggtcaac ttcaaaagtc
 6421 accttgccac agaggactat tagaggttaag aataaacatt gttgggtcaac ttcaaaagtc
 6481 acgagcgcta gccagctctc tgcactgaac attgtcagat ccgagatcgg ccggcctagg
 6541 cgcgccaaag ttaagtgca cggcccaagt ggccactagt acttctcgag gctcaaaagc
 6601 tcatcccaat ttggagtcac tcaagacatc cttgatttaag gcagctgcgc atattgacat
 6661 ggacctcggt cgtgctgcga tagacgactg gccgcgcaga ttgaaggcct gtattcaaaa
 6721 tcacggaggt cattttgaat aaactttagt gtcataagaa tctatgtttt gtttaagtta
 6781 ttttgggtata tgaatgggta cataatgaat aaacttgttt caattatttt acattaaaca
 6841 tgtgacagaa tttatgacct gactaggttag gtacaaaacg cctttttgat attagaaaac
 6901 taagtataat agcctacggt cacatctctt tccgtgggtg togttaaagg gcgacttaga
 6961 gaaccaccaa gaacgtagca gaatcctcag agtgtcatac cagcatacag ccacgcgtaa
 7021 ctgctatttta ctggtaatag ggcacattgt aatctcactt aaccatactg tcggggccacc
 7081 atctagccta tttctgccac gaatcaatcg tgagttagtg acatagagaa actattagtt
 7141 gagaagaaaa caagagcact aaagggttga tattgacaaa aatctacttc gccgtcactc
 7201 catagggtta ttgtctctca ttagtccaga acagcagtta cagacgtaag cttttacgca
 7261 caaactacag ggttgctctt tattgtatcg aaaatatggg acctgaataa gggcgatttt
 7321 gacgcgtcct gcccgcccat tcccgcctct cctctcagat ccactaacgg tgctttaggt acctcaagca
 7381 cccaaacacg tcatttcgga tccctcagat ccactaacgg tgctttaggt acctcaagca
 7441 ccggtcatcg ttctcgtcgg acccgtcgtc tgcgacgaag ggctcgacga gcaaaattaac
 7501 cctcagacac agcccactga gtttctcggc ggatcttctc agcgggtcgc gtttccgac
 7561 cgttggtaga ttctgcgaag cacggctctt gctaggattc gtgttagcaa cgtcgtcagg
 7621 tttgagcccc gtgagctcac ttactagtta aggttacgct gaaatagcct ctcaaggctc
 7681 tcagctaggt aggaacaaaa aaaaaaagtc ctgcccctaa caccgttgcg atggcttgct
 7741 ttctgcaccg cggaaagatg tttgtacgg aaagtttgaa taagtgtcta attgcaagta
 7801 acgtaacaat gttttagggt tcggcgcccg cgggagaaaag catgaagtaa gttctttaaa
 7861 tattacaaaa aaattgaacg atattataaa attctttaaa atattaaaag taagaacaat
 7921 aagatcaatt aaatcataat taatcacatt gttcatgac acaatttaat ttacttcata
 7981 cgttgtattg ttatgtttaa taaaaagatt aatttctatg taattgtatc tgtacaatac
 8041 aatgtgtaga tgtttattct atcgaagata aatacgtcaa aactcgaata ttttcagtat
 8101 aaaaagggtc aactttttca aatcagatc agttcgggtc caactctcaa gatgagagtc
 8161 aaaacctttg tgatcttggt ctgcgctctg caggtgagtt aattatttta ctattatttc

8221 agaaggtggc cagacgatat cacggggccac ctgataataa gtggctcgcca aaacgcacag
 8281 atatcgtaaa ttgtgccatt tgatttgtca cgcccggggg ggctacggaa taaactacat
 8341 ttatttattt aaaaaatgaa ccttagatta tgtaacttgt gatttatttg cgtcaaaagt
 8401 aggcaagatg aatctatgta aatacctggg cagacttgca atatacctatt tcaccggtaa
 8461 atcagcattg caatatgcaa tgcattttca acaatatgta aaacaattcg taaagcatca
 8521 ttgaaaaata gacgaaagaa attgcataaa attataaccg cattattaat ttattatgat
 8581 atctattaac aattgctatt gccttttttt cgcaaatatt aatcattttc ataacctoga
 8641 ggttagcattc tgttacattt taatacatgt gtatgtgatt ataacacgag ctgcccactg
 8701 agtttctcgc cagatcttct cagtgggtcg cgttaccgat cacgtgatag attctatgaa
 8761 gcaactgctct tgttagggct agtgttagca aattctttca ggttagctct gagagctcac
 8821 ctaccatcg gagcgtagct ggaataggct accagctaag aggtagggaa aacaaagctc
 8881 gaaacaagct caagtaataa caacataatg tgaccataaa atctcgtggt gtatgagata
 8941 caattatgta ctttcccaca aatgtttaca taattagaat gttgttcaac ttgcttaacg
 9001 cccagctag aacattcaat tattactatt accactacta aggcagtagt tcctaactcg
 9061 ttccagatca gcgctaactt cgattgaatg tgcgaaattt atagctcaat attttagcac
 9121 ttatcgtatt gatttaagaa aaaattgtta acattttggt tcagtatgtc gcttatacaa
 9181 atgcaaacat caatgatttt gatgaggact attttgggag tgatgtcact gtccaaagta
 9241 gtaatacaac agatgaaata attagagatg catctggggc agttatcgaa gaacaaatta
 9301 caactaaaaa aatgcaacgg aaaaataaaa accatggaat acttggaata aatgaaaaaa
 9361 tgatcaagac gttcgttata accacggatt cogacggtaa caggtccatt gttaggaaag
 9421 atgtgctcat gaagacactt tccgatggta ctgttgctca aagtattggt gctgctgatg
 9481 cgggagcata ttctcagagc gggccatacg tatcaaacag tggatacagc actcatcaag
 9541 gatatacgag cgatttcagc actagtgtcg cagtcgttct agacctggat cccccgggtg
 9601 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 9661 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 9721 gcccgctcgg tccagggctcc gctgcagcgg cggctgctcg agcaggtccg ggtggagcag
 9781 gaccaggagg tgctggacct ggtgggtgct gaccaggagg tgctgggtccg ggtggcccgt
 9841 gaccaggagg tgctggacct ggtgggtgct gaccaggagg tgctgggtccg ggtggcccgt
 9901 ctggtccagg ctccgctgca gcggcggtcg ctgcagcagg tccgggtgga gcaggaccag
 9961 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 10021 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtc
 10081 caggctccgc tgcagcggcg gctgctgcag caggtccggg tggagcagga ccaggaggtg
 10141 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 10201 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctc ggtccaggct
 10261 ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
 10321 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 10381 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggtccgctg
 10441 cagcgcggcg tgcctgcagc ggtccgggtg gagcaggacc aggaggtgct ggacctggtg
 10501 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggacctggtg
 10561 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggacctggtg
 10621 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 10681 gaccaggagg tgctggacct ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 10741 gaccaggagg tgctggacct ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 10801 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 10861 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 10921 gaggtgctgg tccgggtgga cbgtctggtc caggctccgc tgcagcggcg gctgctgag
 10981 cagggtccgg tggagcagga ccaggaggtg ctgacctgg tgggtgctgga ccaggaggtg
 11041 ctggtccggg tggagcagga ccaggaggtg ctgacctgg tgggtgctgga ccaggaggtg
 11101 ctggtccggg tggcccgctc ggtccaggtc ccgctgcagc ggcggtgctc gcagcaggtc
 11161 cgggtggagc aggaccagga ggtgctggac ctggtggtg tggaccagga ggtgctggtc
 11221 cgggtggagc aggaccagga ggtgctggac ctggtggtg tggaccagga ggtgctggtc
 11281 cgggtggccc gtcgtggtcca ggtcccgctg cagcggcggc tgctgcagca ggtccgggtg
 11341 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 11401 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg

11461 gccctgtctgg tccaggctcc gctgcagcgg cggctgtctgc agcagggtccg ggtggagcag
 11521 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggagcag
 11581 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggacctg
 11641 ctgggtccagg ctccgctgca cggcggtgctg ctgcagcagg tccgggtgga gcaggaccag
 11701 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 11761 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtggc ccgtctggtc
 11821 cagggtccgc tgagcggcg gctgctgcag cagggtccgg tggagcagga ccaggaggtg
 11881 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 11941 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctt ggtccaggct
 12001 ccgctgcagc ggcgggtgct gcagcaggct cgggtggagc aggaccagga ggtgctggac
 12061 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 12121 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggtccgctg
 12181 cagcggcgcc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggtccaggct
 12241 gctgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggtccaggct
 12301 gctgctggacc aggaggtgct ggtccgggtg gcccgctctg tccagggtcc gctgcagcgg
 12361 cgggtggtgc agcagggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 12421 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 12481 gaccaggagg tgctggtccg ggtggcccgct ctggtccagg ctccgctgca gcggcggtg
 12541 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 12601 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 12661 gaggtgctgg tccgggtgga ccgtctggtc cagggtccgc tgacgcggcg gctgctgca
 12721 caggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 12781 ctggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 12841 ctggtccggg tggcccgctc ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc
 12901 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggac
 12961 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggac
 13021 cgggtggccc gtctggtcca ggtcccgctg cagcgcgccg tgctgcagca ggtccgggtg
 13081 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 13141 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 13201 gcccgctcgg tccaggctcc gctgcagcgg cggctgctgc agcagggtcc ggtggagcag
 13261 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtcc ggtggagcag
 13321 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtcc ggtggacctg
 13381 ctggtccagg ctccgctgca cggcggtgct ctgcagcagg tccgggtgga gcaggaccag
 13441 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtggc ccgtctggtc
 13501 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtggc ccgtctggtc
 13561 cagggtccgc tgagcggcg gctgctgcag cagggtccgg tggagcagga ccaggaggtg
 13621 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 13681 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctt ggtccaggct
 13741 ccgctgcagc ggcgggtgct gcagcaggct cgggtggagc aggaccagga ggtgctggac
 13801 ctggtggtgc tggaccagga ggtgctggac cgggtggagc aggaccagga ggtgctggac
 13861 ctggtggtgc tggaccagga ggtgctggac cgggtggccc gtctggtcca ggtccgctg
 13921 cagcggcgcc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggtccaggct
 13981 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggtccaggct
 14041 gtgctggacc aggaggtgct ggtccgggtg gcccgctctg tccagggtcc gctgcagcgg
 14101 cggctgctgc agcagggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 14161 gaccaggagg tgctgggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 14221 gaccaggagg tgctgggtccg ggtggcccgct ctggtccagg ctccgctgca gcggcggtg
 14281 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 14341 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 14401 gaggtgctgg tccgggtggc ccgtctggtc cagggtccgc tgagcggcg gctgctgca
 14461 caggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 14521 ctggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 14581 ctggtccggg tggcccgctt ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc
 14641 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggac

14701 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 14761 cgggtggccc gtctgggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
 14821 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 14881 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 14941 gccctgtctg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 15001 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggagcag
 15061 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggcccg
 15121 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
 15181 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 15241 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtc
 15301 caggctccgc tgcagcggcg gctgctgcag caggtccggg tggagcagga ccaggaggtg
 15361 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 15421 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctc ggtccaggct
 15481 ccgctgcagc ggcggtgctc gcagcaggtc cgggtggagc aggaccagga ggtgctggac
 15541 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 15601 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggtcccgctg
 15661 cagcggcggc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggacctggtg
 15721 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggacctggtg
 15781 gtgctggacc aggaggtgct ggtccgggtg gcccgtctgg tccaggctcc gctgcagcgg
 15841 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 15901 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 15961 gaccaggagg tgctggtccg ggtggcccgt ctggtccagg ctccgctgca gcggcggctg
 16021 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 16081 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 16141 gaggtgctgg tccgggtgga ccgtctggtc caggctccgc tgcagcggcg gctgctgcag
 16201 caggctccgg tggagcagga ccaggaggtg ctggacctgg tgggtctgga ccaggaggtg
 16261 ctggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtctgga ccaggaggtg
 16321 ctggtccggg tggcccgtct ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc
 16381 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 16441 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 16501 cgggtggccc gtctgggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
 16561 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 16621 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 16681 gccctgtctg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 16741 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggagcag
 16801 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggcccgt
 16861 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggaagc gtcagttacg
 16921 gagctggcag gggatacgga caaggtgcag gaagtgcagc ttcctctgtg tcatctgctt
 16981 catctcgag ttacgactat tctcgtcgta acgtccgcaa aaactgtgga attcctagaa
 17041 gacaactagt tgttaaattc agagcactgc cttgtgtgaa ttgctaattt ttaataataa
 17101 ataacccttg tttcttaatt cgtcctggat acatctatgt ttttttttcc gttataataa
 17161 gagagcattt aagttattgt ttttaattac ttttttttag aaaacagatt tcggattttt
 17221 tgatgcattt ttatttgaat gtactaatat aatcaattaa tcaatgaatt catttattta
 17281 agggataaca ataattccatg aattcacatg cacatttaaa aaaaaactaa attacaatat
 17341 gttcataata aaacaacaag tatgccttct caactaagaa tactatag
 //

FIG. 14

pSL- Nhện #4+GFP

đoạn lặp đầu tận cùng 3'	Vị trí cặp bazơ*
Bla (Kháng ampicillin)	nhóm bổ sung (2270..2332)
đoạn lặp đầu tận cùng 5'	3633..4493
Gen khởi đầu chính	nhóm bổ sung (5836..5870)
UEE	6586..7751
Miền đầu tận cùng N (NTD)	7754..7823
Gen Fib H trong NTD	7832..9577
EGFP	8152..9575
Tơ nhện (A4S8) X42	9585..10298
Miền đầu tận cùng C (CTD)	10311..17615
Gen Fib-H trong CTD	17622..18101
	17622..17801

Đến bazơ Gốc	3937 a	4193 c	5773 g	4199 t		
1	tcgacgtccc	atggccattc	gaattcggcc	ggcctaggcg	cgccgtacgc	gtatcgataa
61	gctttaagat	acattgatga	gtttggacaa	accacaacta	gaatgcagtg	aaaaaaatgc
121	tttattttgtg	aaattttgtga	tgctattgct	ttattttgtaa	ccattataag	ctgcaataaa
181	caagtttaaca	acaacaattg	cattcatttt	atgttttcagg	ttcaggggga	ggtgtgggag
241	gttttttaaa	gcaagtaaaa	cctctacaaa	tgttggtatgg	ctgattatga	tctagagtcg
301	cgcccgctac	aggaacaggt	ggtggcgccc	ctcgggtgcg	tcgtactgct	ccacgatggt
361	gtatgtccctg	ttgtgggagg	tgatgtccag	cttgaggatcc	acgttagtagt	agccgggcag
421	ctgcacgggc	ttcttggcca	tgtagatgga	cttgaactcc	accaggtagt	ggcccgccgc
481	cttcagcttc	agggccttgt	ggatctcgcc	cttcagcacg	ccgtcgcggg	ggtacaggcg
541	ctcgtgtggag	gcctcccagc	ccatggtcct	ctctctgatt	acggggccgt	cggaggggaa
601	gttccgcgga	tgaacttcac	ctttagatag	aagcagccgt	cctgcaggga	ggagctctgg
661	gtcacggta	ccacgcgcgc	gtcctcgaag	ttcatcacgc	gctcccactt	gaagccctcg
721	gggaaggaca	gcttcttgta	gtcggggatg	tcggcggggg	gcttcacgta	caaccttggag
781	ccgtactgga	actgggggga	caggatgttc	caggcgaagg	gcaggggggc	gcccttggtc
841	accttcagct	tcacggtggt	gtggccctcg	taggggcggc	cctcgcccct	cgccccctga
901	tctcgaactc	gtggccgttc	acggtgccct	ccatgcgcac	cttgaagcgc	atgaactcct
961	tgatgacgtt	cttggaggag	cgcaccatgg	tggcgaccgg	tggatcccgg	gcccgcggtg
1021	ccgtcgactc	tagcgttacc	ccgattgttt	agcttgttca	gctgcgcttg	tttatttgc
1081	tagctttcgc	ttagcgacgt	gttcaacttt	cttgtttgaa	ttgaattgtc	gctccgtaga
1141	cgaagcgct	ctatttatac	tccggcggtc	gagggttcga	aatcgataag	cttggatcct
1201	aattgaatta	gctctaattg	aattagtcct	ctaattgaat	tagtctctaa	ttgaattaga
1261	tccccgggag	agctcgaatt	aaaccattgt	gggaaccgtg	cgatcaaaca	aacgcgagat
1321	accgggaagt	actgaaaaac	agtcgctcca	ggccagtggt	aacatcgatg	ttttgtttg
1381	acggacccct	tactctcgct	tcatataaac	cgaagccagc	taagatggta	tacttattat
1441	catctttgta	tgaggatgct	tctatcaacg	aaagtaccgg	taaaccgcaa	atggttatgt
1501	attataatca	aactaaaggc	ggagtggaca	cgctagacca	aatgtgttct	gtgatgacct
1561	gcagtaggaa	gacgaatagg	tggcctatgg	cattattgta	cggaatgata	aacattgcct
1621	gcataaatto	ttttattata	tacagccata	atgtcagtag	caaggggaga	aaggttcaaa
1681	gtcgcaaaaa	atttatgaga	aacctttaca	tgagcctgac	gtcatcgttt	atgctgaagc
1741	gtttagaagc	tcctactttg	aagagatatt	tgccgcgataa	tatctctaatt	attttgccaa
1801	atgaagtgc	tggtacatca	gatgacagta	ctgaagagcc	agtaaatgaaa	aaacgtactt
1861	actgtactta	ctgcccctct	aaaataaggc	gaaagggcaaa	tgcatcgtgc	aaaaaatgca
1921	aaaaagttat	ttgtcgagag	cataatattg	atatgtgcca	aagttgtttc	tgactgacta
1981	ataagtataa	tttgtttcta	ttatgtataa	gttaagctaa	ttacttattt	tataatacaa

```

2041 catgactggt tttaaagtac aaaataagtt ttttttgta aaagagagaa tgtttaaaag
2101 ttttggtact ttatagaaga aattttgagt ttttggtttt ttttaataaa taaataaaaca
2161 taaataaaatt gtttggtgaa tttattatta gtatgtaagt gtaaatataa taaaacttaa
2221 tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtaaat
2281 tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggttaataaa
2341 tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taccgtatat ctcaacgctg
2401 tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcactgactt gatattgtcc
2461 gcacattttt cgtcgatttg cgttttgatc aaagacttga gcagagacac gttaatcaac
2521 ggttcaaatt gatccataatt aacgatatca acccgatgcy tatatggtgc gtaaaatata
2581 ttttttaacc ctcttatact ttgcactctg cgttaatacy cgttcgtgta cagacgtaat
2641 catgttttct tttttggata aaactcctac tgagtttgac ctcatattag accctcacaa
2701 gttgcaaaac gtggcatttt ttaccaatga agaattttaa gttattttaa aaaatttcat
2761 cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaatcg accagttaat
2821 caacgtgtac acagacgctg cggcaaaaaa cgcgcagccc gacgtgttgg ctaaaaattat
2881 taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttcctca aaaagtgtga
2941 gaccaacaag tttacggaca ctattaatta tttgattttg cccacttca tttgtggga
3001 tcacaatttt gttatatatt taaaacaaag ctttggcact ggccgtcgtt ttacaacgtc
3061 gtgactggga aaacctggc gttaccacac ttaatcgctt tgcagcacat ccccttttgc
3121 ccagctggcg taatagcgaa gagggccgca cggatcgccc ttccaacag ttgcgcagcc
3181 tgaatggcga atggcgccgt atgcggtatt ttctccttac gcatctgtgc ggtatttcac
3241 accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc
3301 gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctccc gcatccgctt
3361 acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac
3421 cgaaacgcgc gagacgaaag ggccctcgtg tacgcctatt tttataggtt aatgtcatga
3481 taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccctta
3541 tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat
3601 aaatgcttca ataattattg aaaaggaaaga gtatgagtat tcaacatttc cgtgtcgccc
3661 ttattccctt ttttgcggca ttttgccttc ctgtttttgc tcacccagaa acgctgggtga
3721 aagtaaaaga tgctgaagat cagttgggtg cagcagtggt ttacatcgaa ctggatctca
3781 acagcggtaa gatccttgag agttttcggc ccgaagaacg ttttccaatg atgagcaatt
3841 ttaaaagtct gctatgtggc gcggtattat cccgtattga cggcgggcaa gagcaactcg
3901 gtcgcccgat acactattct cagaatgact tggttgagta ctcaaccgct acagaaaagc
3961 atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata
4021 acactgggc caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt
4081 tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag
4141 ccataccaaa cgcagagcgt gacaccacga tgctgtgagc aatggcaaca acgttgcgca
4201 aactattaac tggcgaaacta cttactctag cttcccgcca acaatttaata gactggatgg
4261 aggcggataa agttgcagga ccacttctgc gtcgggccct tccggctggc tggtttattg
4321 ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag
4381 atggtaagcc ctcccgtatc gtatgtatct acacgacggg gagtcaggca actatggatg
4441 aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag
4501 accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga
4561 tctaggtgaa gatccttttt gataatctca tgaccaaaat ccottaacgt gagttttcgt
4621 tccactgagc gtcagacccc gtgaaaaaga tcaaaaggatc ttcttgagat cctttttttc
4681 tgccgctaatt ctgctgcttg caaacaacaa aaccacgcgt accagcgggt gttgttttg
4741 cggatcaaga gctaccaact ctttttccga aggttaactgg cttcagcaga gcgcagatac
4801 caaatactgt tcttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac
4861 cgcctacata cctcgctctg ctaatcctgt taccagtggc tgctgccagt gccgataagt
4921 cgtgtcttac cgggttgagc tcaagacgat agttaccgga taaggcgagc cgtcgggct
4981 gaacgggggg ttctgtgaca cagcccagct tggagcgaac gacctacacc gaactgagat
5041 acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaaag gcggacaggt
5101 atccggttaag cggcagggtc ggaacaggag agcgcacgag ggaagctcca gggggaacg
5161 cctgatattc ttatagctct gtcgggtttc gccacctctg acttgagcgt cgatttttgt
5221 gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg

```

5281 ttccctggcct tttgctggcc ttttgcac atgttcttcc ctgcgttacc cctgattct
 5341 gtggataacc gtattaccgc ctttgagtga gctgataccg ctgcgcgcag ccgaacgacc
 5401 gagcgacgag agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc
 5461 cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttccga ctggaaagcg
 5521 ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attagggacc ccaggcttta
 5581 cactttatgc ttccggctcg tatgttggtg ggaattgtga gcggataaca atttcacaca
 5641 ggaaacagct atgacatgat tacgaattcg agctcggtag ccggggatcc tctagagtcg
 5701 acgctcgcgc gacttggttt gccattcttt agcgcgcgct gcgtcacaca gcttgccac
 5761 aatgtggttt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaacg
 5821 gcaaatcttt tttaacccca gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa
 5881 atattgctct ctctttctaa atagcgcgaa tccgtcgcgtg tgcatttagg acatctcagt
 5941 cgccgcttgg agctcccgtg aggcgtgctt gtcaatgcgg taagtgtcac tgaatttgaa
 6001 ctataacgac cgctgtagtc aaaatgacgc atgattatct ttacgtgac ttttaagatt
 6061 taactcatag gataattata ttgttatttc atgttctact tacgtgataa cttattatat
 6121 atataatttc ttgttataga tatcgtgact aatatataat aaaatgggta gttctttaga
 6181 cgatgagcat atcctctctg ctcttctgca aagcgatgac gagctgtgtg gtgaggattc
 6241 tgacagtgaa atatcagatc acgtaagtga agatgacgac cagagcgata cagaagaagc
 6301 gtttatagat gaggtacatg aagtgcagcc aacgtcaagc ggtagtgaat tattagacga
 6361 acaaaatggt attgaacaac caggttcttc atagattctg ttagaagcca aagaatcttg
 6421 accttgccac agaggactat tagaggtaag aataaacatt gttggtcaac ttcaaagtcc
 6481 acgagcgcta gccgagtcct tgcactgaac attgtcagat ccgagatcgg ccggcctagg
 6541 cgcgccaagc ttaaggtgca cgcccacgt ggccactagt acttctcgag gctcaaagcc
 6601 tcatcccaat ttggagtcac tcaagacatc cttgattaa gacgtgcgcg atattgacat
 6661 ggacctcgtt cgtgtgcgca tagacgactg gcgcgcgaga ttgaaggcct gtattcaaaa
 6721 tcacggaggt cattttgaat aaactttagt gtcataagaa tctatgtttt gtttaagtca
 6781 ttttggtata tgaatgggta cataatgaat aaactgtttt caattttttt acattaaaca
 6841 tgtgacagaa tttatgacct gactaggtag gtacaaacag cctttttgat attagaaaaa
 6901 taagtataat agcctacggt cacatctctt tccgtgggtg tctgttaaag gcgacttaga
 6961 gaaccaccaa gaacgtagca gaatcctcag agtgtcatac cagcatacag ccactcgctaa
 7021 ctgctattta ctggtaatag ggcacattgt aatctcactt aaccatactg tcggggccacc
 7081 atctagccta tttctgcccac gaatcaatcg tgaagtatgg acatagagaa actattagtt
 7141 gagaagaaaa caagagcact aaagggttga tattgacaaa aatctacttc gccgtcactc
 7201 cataggttta ttgtctctca ttagtccaga acagcagtta cagacgtaag cttttacgca
 7261 caaactacag ggttgctctt tattgtatcg aaaatatggg acctgaataa gggcgatttt
 7321 gacgcgtcct gcccgcccat tcccgatcct acggacagaa tggcaagcag tcgacgtcgc
 7381 cccaaacacg tcatctcgga tcctcagcat ccactaacgg tgcttttagt acctcaagca
 7441 ccggtcatcg ttctcgtcgg acccgtcgtt tgcgacgaag ggctcgacga gcaaatatca
 7501 cctcagacac agcccactga gtttctcgcc ggtatctctc agcgggtcgc gtttcgacg
 7561 cggtggtaga ttctgcgaag cacggctctt gctaggattc gtgttagcaa cgtcgtcagg
 7621 tttgagcccc gtgagctcac ttactagtta aggttacgct gaaatagcct ctcaaaggctc
 7681 tcagctaggt aggaaacaaa aaaaaagtc ctgcccctaa caccgttgcg atggcttgct
 7741 ttctgcaccg cggaagatg tttgtacgg aaagtgtgaa taagtgtcta attgcaagta
 7801 acgtaacaaat gttttagggg tcggcgggcg cgggagaaag catgaagtaa gttctttaa
 7861 tattacaaaa aaattgaacg atattataaa attctttaa atattaaaaa taagaacaat
 7921 aagatcaatt aaatcataat taatcacatt gttcatgac acaatttaatt ttacttcata
 7981 cgttgtattg ttatgtttaa taaaaagatt aatttctatg taattgtatc tgtacaatac
 8041 aatgtgtaga tgtttattct atcgaaagta aatcgtcaa aactcgaaaa ttttcagtat
 8101 aaaaagggtc tgatcttggt ctgcgctctg caggtgagtt aattatttta ctattatttc
 8161 aaaacctttg tgatcttggt ctgcgctctg caggtgagtt aattatttta ctattatttc
 8221 agaagggtggc cagacgatat cacggggcac ctgataataa gtggtcgcca aaacgcacag
 8281 atatcgtaaa ttgtgccatt tgatttgta cgcccggggg ggctacggaa taaactacat
 8341 ttattttatt aaaaaatgaa ccttagatta tgtaacttgt gattttattg cgtcaaaagt
 8401 aggcaagatg aatctatgta aatacctggg cagacttgca atatctatt tcaccggtaa
 8461 atcagcattg caatatgcaa tgcataattca acaatatgta aaacaattcg taaagcatca

8521 ttagaaaaa gacgaaagaa attgcataaa attataaccg cattattaat ttattatgat
 8581 atctattaac aatttgctatt gccttttttt cgcataattat aatcattttc ataacctcga
 8641 ggtagcattc tgttacattt taatacattg gtatgtgatt ataacacgag ctgcccactg
 8701 agtttctcgc cagatcttct cagtgggtcg cgttaccgat cacgtgatag attctatgaa
 8761 gcactgctct tgttagggct agtgtagca aattctttca ggttgagtct gagagctcac
 8821 ctacccatcg gagcgtagct ggaataggct accagctaata aggtagggaa aacaaagctc
 8881 gaaacaagct caagtaataa caacataatg tgaccataaa atctcgtggt gtatgagata
 8941 caattatgta ctttcccaca aatgtttaca taattagaat gttgttcaac ttgcctaacg
 9001 ccccgactag aacattcaat tattactatt accactacta aggcagtatg tcctaactcg
 9061 ttccagatca gcgctaactt cgattgaatg tgcgaaattt atagctcaat attttagcac
 9121 ttatcgtatt gatttaagaa aaaattgtta acattttgtt tcagtatgct gcttatacaa
 9181 atgcaaacat caatgatttt gatgaggact attttgggag tgatgtcact gtccaaagta
 9241 gtaatacaac agatgaaata attagagatg catctggggc agttatcgaa gaacaaatta
 9301 caactaaaaa aatgcaacgg aaaaaataaa accatgggaat acttggaata aatgaaaaaa
 9361 tgatcaagac gttcgttata accacggatt ccgacggtaa cgagtcacatt gtataggaag
 9421 atgtgctcat gaagacactt tccgatggtt ctgttgctca aagttatggt gctgctgatg
 9481 cgggagcata ttctcagagc gggccatacg tatcaaacag tggatacagc actcatcaag
 9541 gatatacagc cgatttctagc actagtgtcg cagtctgtct agacgtgagc aagggcgagg
 9601 agctgttcac cggggtggtg cccatcctgg tcgagctgga cggcgacgta aacggccaca
 9661 agttcagcgt gtccggcgag ggcgaggggc atgcccacta cggcaagctg accctgaagt
 9721 tcactcgcac caccggcaag ctgcccgtgc cctggcccac cctcgtgacc accctgacct
 9781 acggcgtgca gtgcttcagc cgctaccccg accacatgaa gcagcacgac ttcttcaagt
 9841 ccgcatgccc cgaaggctac gtccaggagc gcaacctctt cttcaaggac gacggcaact
 9901 acaagaccgg cgccgaggtg aagttcgagg gcgacacctt ggtgaaccgc atcgagctga
 9961 agggcatcga cttcaaggag gacggcaaca tcctggggca caagctggag tacaactaca
 10021 acagccacaa cgtctatatc atggccgaca agcagaagaa cggcatcaag gtgaacttca
 10081 agatccgcca caacatcgag gacggcagcg tgacgtctcg cgaccactac cagcagaaca
 10141 ccccatcggc cgacggcccc gtgctgctgc ccgacaacca ctacctgagc acccagtcgg
 10201 ccttgagcaa agaccacca gagaaagcgc atcacatggt cctgctggag ttctgtaccg
 10261 ccgcccggat cactctcggc atggacgagc tgtacaagct ggaatccccg ggtggagcag
 10321 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag
 10381 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggccctg
 10441 ctggtcccagg ctccgctgca cggcgggctg ctgacgagg tccgggtgga gcaggaccag
 10501 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 10561 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctgggtc
 10621 caggctccgc tgacgaggcg gctgctgca caggctccgg tgagcaggga ccaggagggtg
 10681 ctggacctgg tgggtgctga ccaggagggt ctggtccggg tgagcaggga ccaggagggtg
 10741 ctggacctgg tgggtgctga ccaggagggt ctggtccggg tggcccgctt ggtcagggtc
 10801 ccgctgcagc ggcggctgct gcagcagggt cgggtggagc aggaccaggga ggtgctggac
 10861 ctggtggtgc tggaccaggga ggtgctggtc cgggtggagc aggaccaggga ggtgctggac
 10921 ctggtggtgc tggaccaggga ggtgctggtc cgggtggagc aggaccaggga ggtgctggac
 10981 cagcggcggc tgctgcagca ggtccgggtg gagcaggacc aggagggtgct ggaacctgggtg
 11041 gtgctggacc aggagggtgct ggtccgggtg gagcaggacc aggagggtgct ggaacctgggtg
 11101 gtgctggacc aggagggtgct ggtccgggtg gcccgtctg tccagggtcc gctgcagcgg
 11161 cggctgctgc agcagggtcc ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 11221 gaccaggagg tgctgggtcc ggtggcctgt ctggtccagg ctccgctgca gcggcggtc
 11281 gaccaggagg tgctgggtcc ggtggcctgt ctggtccagg ctccgctgca gcggcggtc
 11341 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 11401 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 11461 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 11521 cagggtccgg tgagcaggga ccaggagggt ctggacctgg tggtgctgga ccaggagggtg
 11581 ctggtccggg tgagcaggga ccaggagggt ctggacctgg tggtgctgga ccaggagggtg
 11641 ctggtccggg tgagcaggga ccaggagggt ctggtgctgg tggtgctgga ccaggagggtg
 11701 cgggtggagc aggaccaggga ggtgctggac ctggtggtgc tggaccaggga ggtgctggac

11761 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 11821 cgggtggccc gtctggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
 11881 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 11941 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 12001 gcccgtctgg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 12061 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggccgtg
 12121 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggccgtg
 12181 ctggtccagg ctccgctgca gcggcgctg ctgcagcagg tccgggtgga gcaggaccag
 12241 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 12301 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtc
 12361 cagggtccgc tgccagcggc gctgctgcag caggtccggg tggagcagga ccaggaggtg
 12421 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 12481 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 12541 ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
 12601 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 12661 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 12721 cagcggcggc tggtagcagca ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 12781 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 12841 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 12901 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 12961 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 13021 gaccaggagg tgctggtccg ggtggcctgt ctggtccagg ctccgctgca gcggcgctg
 13081 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 13141 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 13201 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag
 13261 cagggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
 13321 ctggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
 13381 ctggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
 13441 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggac
 13501 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggac
 13561 cgggtggccc gtctggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
 13621 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 13681 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
 13741 gcccgtctgg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 13801 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggacct ggtggccgtg
 13861 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggacct ggtggccgtg
 13921 ctggtccagg ctccgctgca gcggcgctg ctgcagcagg tccgggtgga gcaggaccag
 13981 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 14041 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 14101 cagggtccgc tgccagcggc gctgctgcag caggtccggg tggagcagga ccaggaggtg
 14161 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 14221 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 14281 ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
 14341 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 14401 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 14461 cagcggcggc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 14521 gtgctggacc aggaggtgct ggtccgggtg gcccgtctg tccaggctcc gctgcagcgg
 14581 gtgctggacc aggaggtgct ggtccgggtg gcccgtctg tccaggctcc gctgcagcgg
 14641 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 14701 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 14761 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 14821 ctgcagcagg tccgggtgga gcaggaccag gaggtgctg acctggtggt gctggaccag
 14881 gaggtgctg tccgggtgga gcaggaccag gaggtgctg acctggtggt gctggaccag
 14941 gaggtgctg tccgggtgga gcaggaccag gaggtgctg acctggtggt gctggaccag

15001 cagggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtctgga ccaggaggtg
 15061 ctgggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtctgga ccaggaggtg
 15121 ctgggtccggg tggcccgctc ggtccaggct ccgctgcagc ggcggctgct gcagcaggctc
 15181 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 15241 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 15301 cgggtggccc gtctggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
 15361 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 15421 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 15481 gcccgctctg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 15541 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag
 15601 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggcccg
 15661 ctgggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
 15721 gagggtgctg acctggtggt gctggaccag gagggtgctg tccgggtgga gcaggaccag
 15781 gagggtgctg acctggtggt gctggaccag gagggtgctg tccgggtgga gcaggaccag
 15841 caggctccgc tgcagcggcg gctgctgcag cagggtccgg tggagcagga ccaggaggtg
 15901 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
 15961 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctc ggtccaggct
 16021 ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
 16081 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggtccgctg
 16141 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggtccgctg
 16201 cagcggcggc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 16261 gtgctggacc aggaggtgct ggtccgggtg gagcaggacc aggaggtgct ggtccgggtg
 16321 gtgctggacc aggaggtgct ggtccgggtg gcccgtctg tccaggctcc gctgcagcgg
 16381 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 16441 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 16501 gaccaggagg tgctggtccg ggtggcccgt ctggtccagg ctccgctgca gccggcggctg
 16561 ctgcagcagg tccgggtgga gcaggaccag gagggtgctg acctggtggt gctggaccag
 16621 gagggtgctg tccgggtgga gcaggaccag gagggtgctg acctggtggt gctggaccag
 16681 gagggtgctg tccgggtgga gcaggaccag gagggtgctg acctggtggt gctggaccag
 16741 cagggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 16801 ctggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
 16861 ctggtccggg tggcccgctc ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc
 16921 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 16981 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
 17041 cgggtggccc gtctggtcca ggtccgctg cagcggcggc aggaggtgct ggtccgggtg
 17101 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 17161 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
 17221 gcccgctctg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
 17281 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag
 17341 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggcccg
 17401 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
 17461 gagggtgctg acctggtggt gctggaccag gagggtgctg tccgggtgga gcaggaccag
 17521 gagggtgctg acctggtggt gctggaccag gagggtgctg tccgggtgga gcaggaccag
 17581 caggctccgc tgcagcggcg gctgctgcag cagggtccgg aagcgtcagt tacggagctg
 17641 gcaggggata cggacaaggt gcaggaggtg cagcttcctc tggattcctc agaagacaac
 17701 gcagttacga ctattctcgt cgtaacgtcc gcaaaaaactg tggattcctc agaagacaac
 17761 tagttgttaa attcagagca ctgccttgtg tgaattgcta atttttaata taaaaataac
 17821 ctgttttctt acttcgtcct ggatacatct atgttttttt ttctgttaat aaatgagagc
 17881 atttaagtta ttgtttttta ttactttttt tttagaaaaa gatttcggat tttttgtatg
 17941 catttttatt gaatgtacta atataatcaa ttaataaatg aattcattta ttttaaggat
 18001 aacaataatc catgaattca catgcacatt taaaacaaaa cttaattaca ataggttcat
 18061 ataaaaacaa caagtatgcc ttctcaacta agaatactat ag

//

FIG. 15
pSL-Nhện #6

Đặc điểm	Vị trí cặp bazơ *
đoạn lặp đầu tận cùng 3'	nhóm bổ sung (2270..2332)
Bla (Kháng ampicilin)	3633..4493
đoạn lặp đầu tận cùng 5'	nhóm bổ sung (5836..5870)
Gen khởi đầu chính	6586..7751
UEE	7754..7823
Miền đầu tận cùng N (NTD)	7832..9577
Gen Fib-H trong NTD	8152..9575
Tơ nhện (A4S8) X14	9597..12029
Miền đầu tận cùng C (CTD)	12036..12515
Gen Fib-H trong CTD	12036..12215

* Khi bố thể được sử dụng trong trường hợp này có nghĩa là trình tự của dấu hiệu là dài AND đối so với trình tự được mô tả ở đây. Nói một cách khác, nó là bố thể ngược của trình tự được thể hiện ở đây.

Số đếm cơ sở gốc bazơ	3150 a	2749 c	3331 g	3286 t		
1	tcgacgtccc	atggccattc	gaattcggcc	ggcctaggcg	cgcgctacgc	gtatcgataa
61	gcttttaagat	acattgatga	gtttggacaa	accacaaacta	gaatgcagtg	aaaaaaatgc
121	tttattttgtg	aaattttgtg	tgctatttgc	ttattttgtaa	ccattataag	ctgcaataaa
181	caagttaaca	acaacaattg	cattcatttt	atgttttcagg	ttcaggggga	ggtgtgggag
241	gtttttttaa	gcaagtaaaa	cctctacaaa	tgtggtatgg	ctgattatga	tctagagtcg
301	cggccgctac	aggaacaggt	ggtggcggcc	ctcgggtgcg	tcgtactgct	ccacgatggt
361	gtagtcctcg	ttgtgggagg	tgatgtccag	cttggagtcc	acgtagtagt	agccgggacg
421	ctgcacgggc	ttcttggcca	tgtagatgga	cttgaactcc	accaggtagt	ggccgcgcgc
481	cttcagcttc	agggccttgt	ggatctcgcc	cttcagcacg	ccgtcgcggg	ggtacagggc
541	ctcgggtggg	gcctcccagc	ccatggtcct	cttctgcatt	acggggccgt	cggaggggaa
601	gttccgcccga	tgaacttcac	ctttagatgt	aagcagccgt	cctgcaggga	ggagctctgg
661	gtcacgggtca	ccacgcgcgc	gtcctcgaag	ttcatcacgc	gctcccactt	gaagccctcg
721	gggaaggaca	gcttcttgta	gtcggggatg	tcggcggggg	gcttcacgta	caccttggag
781	ccgtactgga	actgggggga	caggatgtcc	caggcgaagg	gcagggggcc	gcccttggtc
841	accttcagct	tcacggtgtt	gtggccctcg	tagggggcgc	cctcgccctc	cgccctcga
901	tctcgaaactc	gtggccgctc	acggtgccct	ccatgcgcac	cttgaagcgc	atgaactcct
961	tgatgacgtt	cttggaggag	cgcaccatgg	tggcgaccgg	tgatcccg	gcccgcggtg
1021	ccgtcgactc	tagcgggtacc	ccgattgttt	agcttgttca	gctgcgcttg	tttatttgc
1081	tagcttttgc	ttagcgacgt	gttcactttg	cttgtttgaa	ttgaattgtc	gctccgtaga
1141	cgaagcgctc	ctattttatac	tccggcggtc	gaggggttcga	aatcgataag	cttggatcct
1201	aattgaatta	gctctaattg	aattagtctt	ctaattgaat	tagtctctaa	ttgaattaga
1261	tccccggggg	agctcgaatt	aaaccattgt	gggaaccgtg	cgatcaaaac	aacgcgagat
1321	accgggaagt	actgaaaaac	agtcgctcca	ggccagtggg	aacatcgatg	ttttgttttg
1381	acggacccct	tactctcgtc	tcataataaac	cgaagccagc	taagatggta	tacttattat
1441	catcttgtga	tgaggatgct	tctatcaacg	aaagtaccgg	taaaaccgaa	atggttatgt
1501	attataatca	aactaaaggc	ggagtggaca	cgctagacca	aatgtgttct	gtgatgacct
1561	gcagtaggaa	gacgaatagg	tggcctatgg	cattattgta	cggaaatgata	aacattgcct
1621	gcataaattc	ttttattata	tacagccata	atgtcagtag	caagggagaa	aaggttcaaa
1681	gtcgcaaaaa	atttatgaga	aacctttaca	tgagcctgac	gtcatcgttt	atgctgtaagc
1741	gtttagaagc	tcctactttg	aagagatatt	tgcgcgataa	tatctcta	atgttgccaa
1801	atgaagtgcc	tggtacatca	gatgacagta	ctgaagagcc	agtaatgaaa	aaacgtactt

```

1861 actgtactta ctgccccctt aaaataaggg gaaaggcaaa tgcacgtgc aaaaaatgca
1921 aaaaagttaa ttgtcagag cataatattg atatgtgcc aagttgttc tgactgacta
1981 ataagtataa ttgttttcta ttatgtataa gttaagctaa ttacttattt tataatacaa
2041 catgactggt tttaaagta ctaataagtt ttttttgta aaagagagaa tgtttaaaag
2101 ttttgttact ttatagaaga aattttgagt ttttgttttt ttttaataaa taataaaca
2161 taaataaatt gtttgttgaa tttattatta gtatgtaagt gtaaatataa taaaacttaa
2221 tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtcaat
2281 tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggttaataaa
2341 tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taccgtatat ctcaacgctg
2401 tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcatcgactt gatattgtcc
2461 gacacatttt cgtcgtattg cgttttgatc aaagacttga gcagagacac gttaatcaac
2521 tgttcaaatt gatccatatt aacgatatca acccgatcgc tatatggtgc gtaaaatata
2581 ttttttaacc ctcttatact ttgcactctg cgttaatacg cgttcgtgta cagacgtaat
2641 catgttttct ttttttgata aaactcctac tgagtttgac ctcatattag accctcaca
2701 gttgcaaaac gtggcatttt ttaccaatga agaatttaaa gttattttaa aaaatttcat
2761 cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaacg accagttaat
2821 caacgtgtac acagacgctg cggcaaaaaa cagcagcccc gacgtgttgg ctaaaattat
2881 taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttctca aaaagttaa
2941 gaccaacaag tttacggaca ctattaatta tttgattttg cccacttca ttttggga
3001 tcacaatttt gttatatatt taaaacaaag ctttggcact ggccgtcgtt ttacaacgct
3061 gtgactggga aaacctgtgc gttacccaac ttaatgcctt tgcagcacat cccctttcg
3121 ccagctggcg taatagcgaa gaggcgccga cggatcgccc ttccaacag ttgcgcagcc
3181 tgaatggcga atggcgctg atggcggtatt ttctccttac gcatctgtgc ggtatttcac
3241 accgcatatg gtgcactctc agtacaatct gctctgatgc cgcgtgttg tctgctcccg gcatcggctt
3301 gacacccgac aacacccgct gacgcgcctt gacgggcttg tctgctcccg gcatcggctt
3361 acagacaagc tgtgaccgct tccgggagct gcatgtgtca gaggttttca cgtcatcac
3421 cgaacgcgcg gagacgaag gcctcgtgta taccgctatt tttatagggt aatgtcatga
3481 taataatggt ttcttagacg tcaggttgga ctttccggg aaatgtgcgc ggaaccccta
3541 tttgtttatt ttcttaataa cattcaataa tgtatccgct catgagacaa taacctgat
3601 aaatgcttca ataattattg aaaaggaaaga gtatgagtat tcaacatttc cgtgtgcccc
3661 ttattccctt ttttgcggca ttttgccttc ctgtttttgc taccacgaa acgctggtga
3721 aagtaaaaga tgcgtgaagt cagttgggtg cagcagtggt ttacatcgaa ctggatctca
3781 acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt
3841 ttaaagtctt gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg
3901 gtcgccgcat acactattct cagaatgact tgggttagta ctcaccagtc acagaaaagc
3961 atcttacgga tggcatgaca gtaagagaat tatgcagtcg tgccataacc atgagtata
4021 acactgcccg caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt
4081 tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag
4141 ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca
4201 aactattaac tggcgaaacta cttactctag cttcccggca acaattaata gactggatgg
4261 aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg
4321 ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag
4381 atggtaagcc ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg
4441 aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag
4501 accaagttaa ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga
4561 tctaggtgaa gatccttttt gataatctca tgaccaaaat cccttaacgt gacttttctg
4621 tccactgagc gtcagacccc gtgaaaaaga tcaaaaggatc ttcttgagat ccttttttct
4681 tgcgcgtaat ctgctgcttg ctttttcoga aggttaactg cttcagcaga gcgcagatag
4741 cggatcaaga gctaccaact ctttttcoga aggttaactg cttcagcaga gcgcagatag
4801 caaatactgt tcttctagtg tagccgtagt tagggcacca cttcaagaac tctgtagcac
4861 cgcctacata cctcgtctct ctaatcctgt taccagtggc tgctgccagt ggccgataag
4921 cgtgtcttac cgggttggac tcaagacgat agttaccgga taaggcgag cggtcgggct
4981 gaacgggggg ttcgtgcaca cagcccgact tggagcgaac gacctacac gaactgagat
5041 acctacagcg tgagctatga gaaagcgcca cgcttccgca agggagaaag gcggacaggt

```

5101 atccggtgaag cggcagggtc ggaacaggag agcgacagag ggagcttcca gggggaacg
 5161 cctgatatac ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt
 5221 gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacg
 5281 ttccctggcct ttgtctggcc ttttgctcac atgttcttct ctgcgttatc ccctgattct
 5341 gtggataacc gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc
 5401 gagcgagcg agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc
 5461 cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttccoga ctggaaagcg
 5521 ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta
 5581 cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca
 5641 ggaacagcgt atgacatgat tacgaattcg agctcggtag ccggggatcc tctagagtcg
 5701 acgctcgcgc gacttggttt gccattcttt agcgcgcgtc gcgtcacaca gcttggccac
 5761 aatgtgggtt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaacg
 5821 gcaaatcttt ttaaaccta gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa
 5881 atattgctct ctctttctaa atagcgcgaa tccgtcgtcg tgcaattagg acatctcagt
 5941 cgccgcttgg agctcccgtg aggcgtgctt gtcaatgcgg taagtgtcac tgaatttgaa
 6001 ctataacgac cgcgtgagtc aaaaatgacgc atgattatct ttacgtgac ttttaagatt
 6061 taactcatac gataattata ttgttatttc atgttctact tacgtgataa cttatttatat
 6121 atatatcttc ttgttataga tatcgtgact aatatataat aaaaagggtg gttctttaga
 6181 cgatgagcat atcctctctg ctcttctgca aagcgatgac gagcttggg gtgaggattc
 6241 tgacagtgaa atacagatc acgtaagtga agatgacgtc cagagcgata cagaagaagc
 6301 gtttatagat gaggtacatg aagtgcagcc aacgtcaagc ggtagtgaaa tattagacga
 6361 acaaaaatgtt attgaacaac caggttcttc atagattctg ttagaagcca aagaatcttg
 6421 accttgccac agaggactat tagaggtgaag aataaacatt gttggtaaac ttcaaaagtc
 6481 acgagcgcta gccaggtctc tgcactgaac attgtcagat ccgagatcgg ccggcctagg
 6541 cgcgcgaagc ttaaggtgca cggccacgt ggcactagt acttctcgag gctcaaaagc
 6601 tcatcccaat ttggagtcac tcaagacatc cttgattaa ggcagctgccc atattgacat
 6661 ggacctcgtt cgtgctgcga tagacgactg gccgcgcaga ttgaaggcct gtattcaaaa
 6721 tcacggaggt cattttgaat aaactttagt gtcataagaa tctatgtttt gttaaagtca
 6781 ttttggtata tgaattggtt cataatgaat aaacttgttt caattatttt acattaaaca
 6841 tgtgacagaa tttatgacct gactaggtag gtacaaacag cctttttgat attagaaaaa
 6901 taagtaaaat agcctacggt cacatctctt tccgtgggtg tctgttaaag gcgacttaga
 6961 gaaccaccaa gaacgtagca gaatcctcag agtgtcatac cagcatacag ccatcgctaa
 7021 ctgctattta ctggtaatag ggcacattgt aatctcactt aaccatactg tccggccacc
 7081 atctagccta tttctgccac gaatcaatcg tgagtgtatg acatagagaa actattagtt
 7141 gagaagaaaa caagagcact aaagggttga tattgacaaa aatctacttc gccgtcactc
 7201 cataggttta ttgtctctca ttagtccaga acagcagtta cagacgtaa gcttttacga
 7261 caaactacag ggttgcctct tttgtatcgc aaaatatggg acctgaataa gggcgatttt
 7321 gacgcgtcct gcccgcccat tcccgatcct acggacagaa tggcaagcag tcgacgtcgc
 7381 cccaaacacg tcatttcgga tccctacgat ccactaacgg tgccttaggt acctcaagca
 7441 ccggtcatcg ttctcgtcgg acccgtcgtt tgcgacgaag ggctcgacga gcaaatcaac
 7501 cctcagacac agcccactga gtttctcgcc ggtatctctc agcgggtcgc gtttccgact
 7561 cggtggtaga ttctgcgaag cacggctctt gctaggattc gtgttagcaa cgtcgtcagg
 7621 tttgagcccc gtgagctcac ttactagtta aggttacgct gaaatagcct ctcaaggctc
 7681 tcagctaggt aggaacaaa aaaaaaagtc ctgccttaa caccgttgcg atggcttgct
 7741 ttctgcaccg cggaagatg ttttgtacgg aaagtgtgaa taagtgccta attgcaagta
 7801 acgtaacaaat gttttagggt tcggcgcccg cgggagaaa catgaagtaa gttctttaa
 7861 tattacaaaa aaattgaacg atattataaa attcttttaa atattaaaag taagaacaa
 7921 aagatcaatt aaatcataat taatcacatt gttcatgac acaatttaac ttacttcata
 7981 cggtgtattg ttatgtttaa taataagatt aatttctatg taattgtatc tgtacatac
 8041 aatgtgtaga tgtttattct atcgaaagta aatacgtcaa aactcgaaaa ttttcagtat
 8101 aaaaagggtc aactttttca aatcagcatc agttcgggtc caactctcaa gatgagagtc
 8161 aaaacctttg tgatcttggt ctgcgctctg cagggtgagtt aattatttta ctattatttc
 8221 agaaggtggc cagacgatat cacgggccac ctgataataa gtggtcgcca aaacgcacag
 8281 atatcgtaaa ttgtgccatt tgatttgtca cgcccggggg ggctacggaa taaactacat

8341 ttattttattt aaaaaatgaa ccttagatta tgtaacttgt gattttatttg cgtcaaaagt
8401 agggcaagatg aatctatgta aatacctggg cagacttgca atatcctatt tcaccggtaa
8461 atcagcattg caatatgcaa tgcataattca acaatatgta aaacaattcg taaagcatca
8521 ttagaaaaata gacgaaagaa attgcataaa attataaccg cattattaat ttattatgat
8581 atctattaac aattgctatt gccttttttt cgcaaatat aatcattttc ataacctcga
8641 ggtagcattc tgttacattt taatacattg gtatgtgatt ataacacgag ctgcccactg
8701 agtttctcgc cagatcttct cagtgggtcg cgttaccgat cagtgatag attctatgaa
8761 gcactgctct tgtagggct agtggttagca aattccttca ggttgagtct gagagctcac
8821 ctacccatcg gagcgtagct ggaatagct accagcta at aggtagggaa aacaaagctc
8881 gaaacaagct caagtaataa caacataatg tgaccataaa atctcgtggt gtatgagata
8941 caattatgta ctttcccaca aatgtttaca taattagaat gttgttcaac ttgcctaacg
9001 ccccgactag aacattcaat tattactatt accactacta aggcagtatg tcctaactcg
9061 ttccagatca gcgctaactt cgattgaatg tgcgaaattt atagctcaat attttagcac
9121 ttatcgtatt gatttaagaa aaaattgtta acattttgtt tcagtatgct gcttatacaa
9181 atgcaaacat caatgatttt gatgaggact attttgggag tgatgtcact gtccaaagta
9241 gtaatacaac agatgaaata attagagatg catctggggc agttatcgaa gaacaaatta
9301 caactaaaaa aatgcaacgg aaaaataaaa accatggaat acttggaata aatgaaaaaa
9361 tgatcaagac gttcgttata accacggatt ccgacggtaa cagtgccatt gttagggaag
9421 atgtgctcat gaagacactt tccgatggta ctgttgctca aagttatgtt gctgctgatg
9481 cgggagcata ttctcagagc gggccatcac tatcaaacag tggatacagc actcatcaag
9541 gatatacagc cgatttcacg actagtctg cagtctgtct agacctggaat cccccgggtg
9601 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
9661 gagcaggacc aggaggtgct ggacctgggt gtgctggacc aggaggtgct ggtccgggtg
9721 gcccgctctg tccaggtccc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
9781 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag
9841 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggcccg
9901 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
9961 gaggtgctg acctggtggt gctggaccag gaggtgctg tccgggtgga gcaggaccag
10021 gaggtgctg acctggtggt gctggaccag gaggtgctg tccgggtgga ccgtctggtc
10081 caggctccgc tgcagcggcg gctgctgcag caggtccggg tggagcagga ccaggaggtg
10141 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
10201 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggacctgct ggtccaggct
10261 ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
10321 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
10381 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
10441 cagcggcggc tgctgcagca ggtccgggtg gagcaggacc aggaggtgct ggcacctggtg
10501 gtgctggacc aggaggtgct ggtccgggtg gccctgctg tccaggctcc gctgcagcgg
10561 gtgctggacc aggaggtgct ggtggagcag gaccaggagg tgctggacct ggtggtgctg
10621 cggctgctgc agcaggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
10681 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
10741 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
10801 ctgcagcagg tccgggtgga gcaggaccag gaggtgctg acctggtggt gctggaccag
10861 gaggtgctg tccgggtgga gcaggaccag gaggtgctg acctggtggt gctggaccag
10921 gaggtgctg tccgggtgga ccgtctggtc caggctccgc tgcagcggcg gctgctgca
10981 caggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
11041 ctggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
11101 ctggtccggg tggagcagga ccaggaggtg ctggacctg tgggtgctgga ccaggaggtg
11161 cgggtggagc aggaccagga ggtgctggac ctggtggtg tggaccagga ggtgctggac
11221 cgggtggagc aggaccagga ggtgctggac ctggtggtg tggaccagga ggtgctggac
11281 cgggtggccc gctggttcca ggtccgctg cagcggcggc tgctgcagca ggtccgggtg
11341 gagcaggacc aggaggtgct ggaacctggtg gtgctggacc aggaggtgct ggtccgggtg
11401 gagcaggacc aggaggtgct ggaacctggtg gtgctggacc aggaggtgct ggtccgggtg
11461 gcccgctctg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
11521 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag

```

11581 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggcccg
11641 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
11701 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
11761 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtc
11821 caggctccgc tgcagcggcg gctgctgcag caggctccgg tggagcagga ccaggaggtg
11881 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
11941 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgctt ggtccaggct
12001 ccgctgcagc ggcggctgct gcagcaggtc cgggaagcgt cagttacgga gctggcaggg
12061 gatacggaca aggtgcagga agtgcagctt cctctgtgtc atctgcttca tctcgcagtt
12121 acgactatc tgcctgtaac gtccgcaaaa actgtggaat tccagaaga caactagtgt
12181 ttaaattcag agcactgcct tgtgtgaatt gctaattttt aatataaat aacccttgtt
12241 tcttacttcg tcctggatac atctatgttt ttttttcgt taataaatga gagcatttaa
12301 gttattgttt ttaattactt ttttttagaa aacagatttc ggattttttg tatgcatttt
12361 atttgaatgt actaatataa tcaattaatc aatgaattca tttatttaag ggataacaat
12421 aatccatgaa ttcacatgca catttaaac aaaactaat tacaataggt tcataataaa
12481 acaacaagta tgccttctca actaagaata ctatag
//

```

FIG. 16
pSL- Nhện #6+GFP

Đặc điểm	Vị trí cặp bazơ *
Bla (Kháng ampicillin)	3633..4493
đoạn lặp đầu tận cùng 3'	nhóm bổ sung (2270..2332)
đoạn lặp đầu tận cùng 5'	nhóm bổ sung (5836..5870)
Gen khởi đầu chính	6586..7751
UEE	7754..7823
Miền đầu tận cùng N (NTD)	7832..9577
Gen Fib - H trong NTD	8152..9575
EGFP	9585..10298
Tơ nhện (A2S8) X14	10311..12743
Miền đầu tận cùng C (CTD)	12750..13229
Gen Fib - H trong CTD	12750..12929

* Khi bộ thể được sử dụng trong trường hợp này có nghĩa là trình tự của dấu hiệu là dài AND đối so với trình tự được mô tả ở đây. Nói một cách khác, nó là bộ thể ngược của trình tự được thể hiện ở đây.

Số đếm cơ sở gốc bazơ	3321 a	2989 c	3533 g	3387 t		
1	tcgacgtccc	atggccattc	gaattcggcc	ggcctaggcg	cgccgtacgc	gtatcgataa
61	gctttaagat	acattgatga	gtttggacaa	accacaacta	gaatgcagtg	aaaaaaatgc
121	tttatttgtg	aaatttgtga	tgctattgct	ttatttgtaa	ccattataag	ctgcaataaa
181	caagttaaca	acaacaattg	cattcatttt	atgtttcagg	ttcaggggga	ggtgtgggag
241	gttttttaaa	gcaagtaaaa	octctacaaa	tgtggtatgg	ctgattatga	tctagagtcg
301	cgcccgctac	aggaacaggt	ggtggcggcc	ctcgggtgcg	tcgtactgct	ccacgatggg
361	gtagtccctc	ttgtgggagg	tgatgtccag	cttggagtcg	acgtagtagt	agccgggcag
421	ctgcacgggc	ttcttggcca	tgtagatgga	cttgaactcc	accaggtagt	ggcccgccgc
481	cttcagcttc	agggccttgt	ggatctcgcc	cttcagcagc	ccgtcgccgg	ggtacaggcg
541	ctcgggtggg	gcctcccagc	ccatgggtctt	cttcctgcatt	acggggccgt	cggaggggaa
601	gttccgcccga	tgaacttcac	cttgtagatg	aagcagccgt	cctgcaggga	ggagtccctgg
661	gtcacgggtca	ccacgccgcc	gtcctcgaa	ttcatcacgc	gctcccactt	gaagccctcg
721	gggaaggaca	gcttcttgta	gtcggggatg	tcggcggggt	gcttcacgta	caccttggag
781	ccgtactgga	actgggggga	caggatgtcc	caggcgaaag	gcagggggcc	gcccttggtc
841	accttcagct	tcacgggtgt	gtggccctcg	tagggcggcg	cctcgccctt	cgccctctga
901	tctcgaactc	gtggccggtt	acgggtgccct	ccatgcgcac	cttgaagcgc	atgaactcct
961	tgatgacgtt	cttggaggag	cgcaccatgg	tggcgaccgg	tggatcccgg	gcccgcggtg
1021	ccgtcgactc	tagcgggtacc	ccgattgttt	agcttgttca	gctgcgcttg	tttatttgct
1081	tagctttcgc	ttagcgacgt	gttcactttg	cttgtttgaa	ttgaattgtc	gctccgtaga
1141	cgaagcgcct	ctatttatac	tccggcggtc	gaggggttcga	aatcgataag	cttggatcct
1201	aattgaatta	gctctaattg	aattagtctt	ctaattgaat	tagtctctaa	ttgaattaga
1261	tccccgggcy	agctcgaatt	aaaccattgt	gggaaccgtg	cgatcaaaca	aacgcgagat
1321	accgggaagt	actgaaaaac	agtcgctcca	ggccagtggg	aacatcgatg	ttttgttttg
1381	acgggacccct	tactctcgtc	tcatataaac	cgaagccagc	taagatggta	tacttattat
1441	catcttgtag	tgaggatgct	tctatcaacg	aaagtaccgg	taaaccgcaa	atggttatgt
1501	attataatca	aactaaaggc	ggagtggaca	cgctagacca	aatgtgttct	gtgatgacct
1561	gcagtaggaa	gacgaatagg	tggcctatgg	cattattgta	cggaatgata	aacattgcct
1621	gcataaatcc	ttttattata	tacagccata	atgtcagtag	caaggggagaa	aaggttcaaa


```

1681 gtcgcaaaaa atttatgaga aacctttaca tgagcctgac gtcacgtgtt atgcgtaagc
1741 gtttagaagc tcctactttg aagagatatt tgcgcgataa tatctccta attttgccaa
1801 atgaagtggc tggtagatca gatgacagta ctgaagagcc agtaatgaaa aaacgtactt
1861 actgtactta ctgccccctc aaaataaggg gaaaggcaaa tgcatcgtgc aaaaaatgca
1921 aaaaagttat ttgtcgagag cataatatgg atatgtgcc aagttgttct tgactgacta
1981 ataagtataa tttgtttcta ttatgtataa gttaagctaa ttacttatt tataatacaa
2041 catgactgtt tttaaagtac aaaataagtt tatttttgta aaagagagaa tgtttaaag
2101 tttgtgtact ttatagaaga aattttgagt tttgttttt ttttaataaa taaaataaca
2161 taaataaat gtttgttgaa tttattatta gtatgtaagt gtaaatataa taaaacttaa
2221 tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtcaat
2281 tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggtaataaa
2341 tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taaccgtat ctcaacgctg
2401 tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcacgtactt gatattgtcc
2461 gacacatttt cgtcgtattg cgttttgatc aaagacttga gcagagacac gttaatcaac
2521 tgttcaaat gatccatatt aacgatatca acccgatgag tatatgtgc gtaaaatata
2581 ttttttaacc ctcttatact ttgcactctg cgttaataac cgttcgtgta cagacgtaat
2641 catgttttct tttttggata aaactcctac tgagtttgac ctcatattag accctcaaa
2701 gttgcaaac gtggcatttt ttaccaatga agaatttaa gttattttaa aaaatttcat
2761 cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaacg accagttaat
2821 caacgtgtac acagacgcgt cggcaaaaa cacgcagccc gacgtgttg ctaaaattat
2881 taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttcctca aaaagttgaa
2941 gaccaacaag tttacggaca ctattaatta tttgattttg ccccaactca ttttgggga
3001 tcacaatttt gttatatttt taaaacaaag ctttggcact ggccgtcgtt ttacaacgtc
3061 gtgactggga aaacccctggc gttaccacaac ttaatcgctc tgccagacat cccctttcg
3121 ccagctggcg taatagcgaa gagggccgca ccgacgcgcc ttcccaacag ttgcgcagcc
3181 tgaatggcga atggcgccgt atgcggtatt ttctccttac gcactcgtgc ggtatttcac
3241 accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc
3301 gacacccgcc aacacccgct gacgcgcctc gacgggcttg tctgtcccc gcactcgtt
3361 acagacaagc tgtgaccgct tccgggagct gcactgtgca gaggttttca cgtcatcac
3421 cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt tttataggtt aatgtcatga
3481 taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc ggaaccccta
3541 tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat
3601 aaatgcttca ataatttga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc
3661 ttattccctt ttttgcggca ttttgcttcc ctgtttttgc tcaccagaaa acgctggtga
3721 aagtaaaaga tgctgaagat cagttgggtg cagcagtggt ttacatcgaa ctggatctca
3781 acagcggtaa gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt
3841 ttaaagtctt gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg
3901 gtcgcccgc atactattct cagaatgact tgggttagta ctccaccgtc acagaaaagc
3961 atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata
4021 acactgccc caacttactt ctgacaacga tcggagagacc gaaggagcta accgcttttt
4081 tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag
4141 ccataccaaa cgacgagcgt gacaccacga tgccgtgtagc aatggcaaca acgttgcgca
4201 aactattaac tggcgaacta cttactctag cttcccggca acaattaata gactggatgg
4261 aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg
4321 ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag
4381 atggtaagcc ctcccgtatc gtagttatct acacgacggg gagtacggca actatggatg
4441 aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag
4501 accaagttta ctcatatata ctttagattg attttaaact tcatttttaa tttaaaagga
4561 tctaggtgaa gatccttttt gataatctca tgacaaaat cccttaacgt gagttttcgt
4621 tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat ccttttttct
4681 tgcgcgtaat ctgctgcttg caaacaacaa aaccacgctc accagcgggt gtttgtttgc
4741 cggatcaaga gctaccaact ctttttccga aggttaactg cttcagcaga gcgcagatag
4801 caaatactgt tcttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac
4861 cgcctacata cctcgtctct ctaatcctgt taccagtggt tgctgccagt ggcgataagt

```

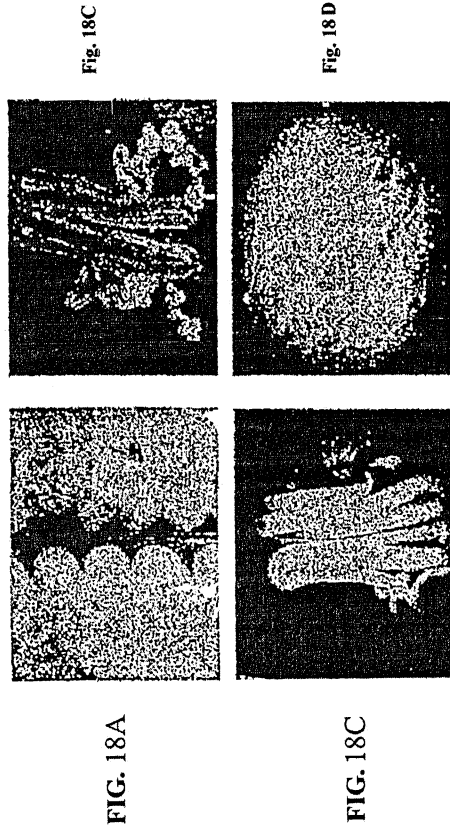
4921 cgtgtctttac cgggttggac tcaagacgat agttaccgga taaggcgag cggtcgggct
 4981 gaacggggggg ttctgtgcaca cagcccagct tggagcgaa gacctacacc gaactgagat
 5041 acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt
 5101 atccggtaag cggcagggtc ggaacaggag agcgacagag ggagcttcca ggggaaacg
 5161 cctgatattct ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt
 5221 gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg
 5281 ttcttgccct ttgtggtggt ttttgctcac atgttctttc ctgcttctc ccctgattct
 5341 gtggataaac gtattaccgc ctttgagtga gctgataccg ctgcccgcag ccgaacgacc
 5401 gagcgcgagc agtcagtgag cgagggaagc gaagagcgcc caatacgcaa accgcctctc
 5461 cccgcgcggt ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg
 5521 ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta
 5581 cactttatgc ttccggctcg tatgttgtgt ggaattgtga gcggataaca atttcacaca
 5641 ggaaacagct atgacatgat tacgaattcg agctcggtac ccggggatcc tctagagtcg
 5701 acgctcgcgc gacttggttt gccattcttt agcgcgcgtc gcgtcacaca gcttggccac
 5761 aatgtgggtt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaaag
 5821 gcaaatcttt tttaaccccta gaaagatagt ctgctgtaaa ttgacgcagt cattcttgaa
 5881 atattgtctc ctctttctaa atagcgcgaa tccgtcgtg tgcaatttag acatctcagt
 5941 cgcgcgcttg agctcccggt aggcgtgctt gtcaatcgcg taagtgtcac tgattttgaa
 6001 ctataacgac cgcgtgagtc aaaaatgacgc atgattatct tttaacgtac ttttaagatt
 6061 taactcatac gataaattata ttgttatttc atgttctact tacgtgataa cttattatat
 6121 atatatcttc ttgttataga tatcgtgact aatatataat aaaaagggtg gttctttaga
 6181 cgatgagcat atcctctctg ctcttctgca aagcgatgac gagcttggtg gtgaggattc
 6241 tgacagtgaa atatcagatc acgtaagtga agatgacgtc cagagcgata cagaagaagc
 6301 gtttatagat gaggtacatg aagtgcagcc aacgtcaagc ggtagtgaat tattagacga
 6361 acaaaatggt attgaacaac cagggttctc atagattctg ttagaagcca aagaatcttg
 6421 accttgccac agaggactat tagaggtaag aataaacatt gttggtcaac tcaaaagtc
 6481 acgagcgcta gccgagtcgc tgcactgaac attgtcagat ccgagatcgg ccggcctagg
 6541 cgcgccaagc ttaaggtgca cggcccacgt ggccactagt acttctcgag gctcaaaagc
 6601 tcatcccaat ttggagtcac tcaagacatc ctgtgattaag gcagctgccc atattgacat
 6661 gacccctggt cgtgtgcga tagacgactg gccgcgcaga ttgaagcgct gtattcaaaa
 6721 tcacggaggt cattttgaat aaactttagt gtcataagaa tctatgtttt gttaaagtca
 6781 ttttggtata tgaatggtta cataatgaat aaacttggtt caattatttt acattaaaaca
 6841 tgtgacagaa tttatgacct gactaggtag gtacaaacag cctttttgat attagaaaac
 6901 taagtaaaat agcctacggt cacatctctt tccgtgggtg tcgttaaagg gcgacttaga
 6961 gaaccaccaa gaacgtagca gaatcctcag agtgtcatac cagcatacag ccactcgtaa
 7021 ctgctattta ctggtaatag ggcacattgt aatctcactt aaccatactg tccggccacc
 7081 atctagccta tttctgcccac gaatcaatcg tgagtgtggt acatagagaa actattagtt
 7141 gagaagaaaa caagagcact aaaggtttga tattgacaaa aatctacttc gccgtcactc
 7201 cataggttta ttgtctctca ttagtccaga acagcagtta cagacgtaag cttttacgca
 7261 caaactacag ggttgctctt tcccgatcct acggacagaa tggcaagcag tcgacgtcgc
 7321 gagcgctcct gcccgcccat tcccgatcct acgtaacggt tgcttttagt acctcaagca
 7381 cccaaacacg tcatttccga tcctcagcat ccactaacgg tgcttttagt acctcaagca
 7441 cccggtcatc ttctcgtcgg acccgctcgt tgcgacgaag ggctcgacga gcaaataaac
 7501 cctcagacac agcccactga gtttctcgcg ggtatctctc agcgggtcgc gtttccgac
 7561 cgttggtaga ttctgcgaag cacggctctt gctaggattc gtgttagcaa cgtcgtcagg
 7621 tttgagcccc gtgagctcac ttactagtta aggttacgct gaaatagcct ctcaaggctc
 7681 tcagctaggt aggaacaaa aaaaaagtc ctgcccttaa caccgttgcg atggcttgct
 7741 ttctgcaccg cggaaagatg tttgttacgg aaagtgtgaa taagtgtcta attgcaagta
 7801 acgtaacaat gttttagggg tcggcgccgc cgggagaaag catgaagtaa gttctttaa
 7861 tattacaaaa aaattgaacg atattataaa attctttaa atatttaa ttaactcata
 7921 aagatcaatt aaatcataat taatcacatt gttcatgac acaattta ttaactcata
 7981 cgttggtatt ttatgtttaa taaaagatt aatttctat taattgtatc tgtacaatac
 8041 aatgtgtaga tgtttattct atcgaagata aatacgtcaa aactcgaaaa ttttcagtat
 8101 aaaaagggtc aactttttca aatcagcatc agttcgggtt caactctcaa gatgagagtc

8161 aaaacctttg tgatcttgtg ctgcgctctg caggtgagtt aattatttta ctattatttc
 8221 agaaggtggc cagacgatat cacgggccac ctgataataa gtggtcgcca aaacgcacag
 8281 atatcgtaaa ttgtgccatt tgatttgtca cgcccgggg ggctacggaa taaactacat
 8341 ttttttattt aaaaaatgaa ccttagatta tgtaacttgt gatttatttg cgtcaaaagt
 8401 aggcgaagtg aatctatgta aatacctggg cagacttgca atacctatt tcaccggtaa
 8461 atcagcattg caatatgcaa tgcatttca acaatatgta aaacaattcg taaagcatca
 8521 ttagaaaata gacgaaagaa attgcataaa attataaccg cattattaat ttattatgat
 8581 atctattaac aattgctatt gccttttttt cgcaaatat aatcattttc ataaccctga
 8641 ggtagcattc tgttacattt taatacattg gtatgtgatt ataacacgag ctgcccactg
 8701 agtttctcgc cagatcttct cagtgggtcg cgttaccgat cacgtgtag attctatgaa
 8761 gactgctct ttttagggct agtgttagca aattctttca ggttagtct gagagctcac
 8821 ctaccatcg gagcgtagct ggaataggct accagctaag aggtaggaa aacaaagctc
 8881 gaaacaagct caagtaataa caacataatg tgaccataaa atctcgtggt gtatgagata
 8941 caattatgta ctttcccaca aatgtttaca taattagaat gttgttcaac ttgcctaacc
 9001 cccagctag aacattcaat tattactatt accactacta aggcagtatg tcttaactcg
 9061 ttccagatca gcgctaactt cgattgaatg tgcgaaattt atagctcaat attttagcac
 9121 ttatcgtatt gatttaagaa aaaattgtta acattttgtt tcagtatgct gcttatacaa
 9181 atgcaaacat caatgatatt gatgaggact attttgggag tgaatgcact gtccaaagta
 9241 gtaatacaac agatgaaata attagagatg catctggggc agttatcgaa gaacaaatta
 9301 caactaaaaa aatgcaacgg aaaaaataaa accatggaaat acttggaata aatgaaaaaa
 9361 tgatcaagac gttcgttata accacggatt ccgacggtaa cgagtccatt gtagaggaa
 9421 atgtgctcat gaagacactt tccgatggta ctgttgctca aagtatggt gctgctgatg
 9481 cgggagcata ttctcagagc gggccatacg tatcaaacag tggatacagc actcatcaag
 9541 gatatacgag cgatttccagc actagtctgt cagtcgttct agacgtgagc aagggcagag
 9601 agctgttcac cggggtgggt cccatcctgg tgcagctgga cggcgacgta aacggccaca
 9661 agttcagcgt gtccggtcga ggcgagggcg atgccaccta cggcaagctg accctgaagt
 9721 tcatctgcac caccggcaag ctgcccgtgc cctggcccac cctcgtgacc accctgaact
 9781 acggcgtgca gtgcttcagc cgtaccctcg accacatgaa gcagcagcag ttcctcaagt
 9841 ccgccatgcc cgaaggctac gtccaggagc gcaccatctt cttcaaggac gacggcaact
 9901 acaagaccgc cgccgaggtg aagttcagag gcgacacccct ggtgaaccgc atcagagctga
 9961 agggcatcga cttcaaggag gacggcaaca tcctggggca caagctggag tacaactaca
 10021 acagccacaa cgtctatatac atggccgaca agcagaagaa cggcatcaag gtgaacttca
 10081 agatccgcca caacatcgag gacggcagcg tgcagctcgc cgaccactac cagcagaaca
 10141 ccccatcgg cgacggcccc gtgctgctgc ccgacaacca ctacctgagc acccagtcgg
 10201 ccctgagcaa agaccccaac gagaagcggc atcacatggt cctgctggag ttcgtgaccg
 10261 ccgcccggat cactctcgcc atggacgagc tgtacaagct ggatcccccg ggtggagcag
 10321 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggagcag
 10381 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggtccg ggtggccctg
 10441 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
 10501 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
 10561 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtg
 10621 caggctccgc tgcagcggcg gctgctgcag caggtccggg tggagcagga ccaggagggtg
 10681 ctggacctgg tgggtgctgga ccaggagggt ctggtccggg tggagcagga ccaggagggtg
 10741 ctggacctgg tgggtgctgga ccaggagggt ctggtccggg tggcccgctt ggtccaggct
 10801 ccgctgcagc ggcggctgct gcagcagggt cgggtggagc aggaccagga ggtgctggac
 10861 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
 10921 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggctccgctg
 10981 cagcggcgcc tgctgcagca ggtccgggtg gagcaggacc aggagggtgct ggacctggtg
 11041 gtgctggacc aggagggtgct ggtccgggtg gagcaggacc aggagggtgct ggacctggtg
 11101 gtgctggacc aggagggtgct ggtccgggtg gcccgctcgg tccaggctcc gctgcagcgg
 11161 cggctgctgc agcagggtcc ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 11221 gaccaggagg tgctggtccg ggtggagcag gaccaggagg tgctggacct ggtggtgctg
 11281 gaccaggagg tgctggtccg ggtggcccgct ctggtccagg ctcgctgca gcggcggtg
 11341 ctgcagcagg tccgggtgga gcaggaccag gaggtgctgg acctggtggt gctggaccag

```

11401 gaggtgctgg tccgggtgga gcaggaccag gaggtgctgg acctgggtgg gctggaccag
11461 gaggtgctgg tccgggtggc ccgtctggtc caggctccgc tgcagcggcg gctgctgcag
11521 caggctccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
11581 ctggtccggg tggagcagga ccaggaggtg ctggacctgg tgggtgctgga ccaggaggtg
11641 ctggtccggg tggcccgtct ggtccaggct ccgtgcagc ggcgggtgct gcagcaggtc
11701 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
11761 cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc
11821 cgggtggccc gtctgggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg
11881 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
11941 gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg
12001 gcccgtctgg tccaggctcc gctgcagcgg cggctgctgc agcaggtccg ggtggagcag
12061 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggctcc ggtggagcag
12121 gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctggctcc ggtggcccgt
12181 ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag
12241 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag
12301 gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga ccgtctggtc
12361 caggctccgc tgcagcggcg gctgctgcag caggctccgg tggagcagga ccaggaggtg
12421 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg
12481 ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgtct ggtccaggct
12541 ccgtgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac
12601 ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac
12661 ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctgggtcca ggtccgctg
12721 cagcggcggc tgctgcagca ggtccgggaa gcgtcagtta cggagctggc aggggatacg
12781 gacaagggtg aggaagtga gcttcctctg tgcctctgc ttcctctgc agttacgact
12841 attctcgtcg taacgtccgc aaaaactgtg gaattcctag aagacaacta gttgttaaat
12901 tcagagcact gccttgtgtg aattgctaatt ttttaataata aataaacctt tgtttcttac
12961 ttcgtcctgg atacatctat gttttttttt tcgttaataa atgagagcat ttaagttatt
13021 gtttttaatt actttttttt agaaaacaga tttcggattt tttgtatgca ttttatttga
13081 atgtactaat ataataaatt aatcaatgaa ttcatttatt taaggataaa caataatcca
13141 tgaattcaca tgcacattta aaacaaaact aaattacaat aggttcatat aaaaaaaca
13201 agtatgcctt ctcaactaag aatactatag
//

```

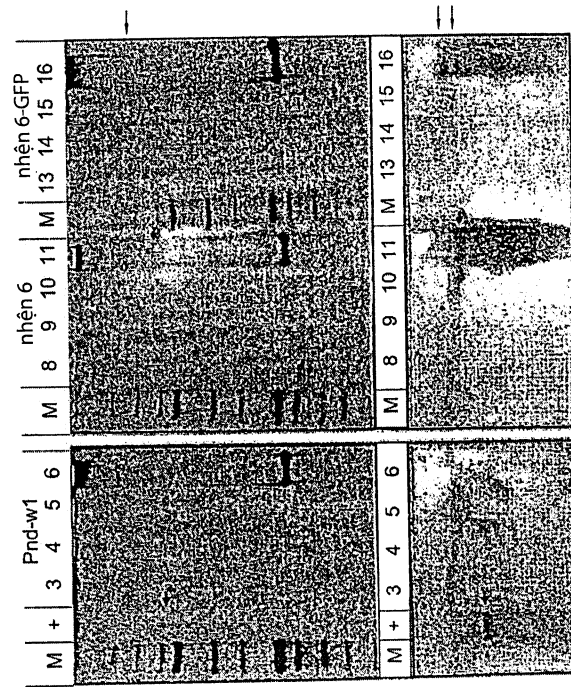


Fig.19

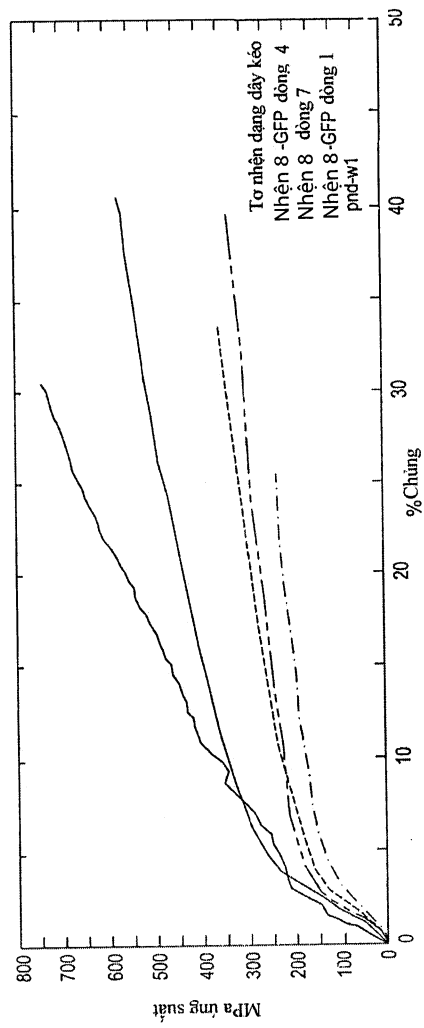


Fig 20

FIG. 21
pXL BAC II ECFP NTD CTD Maspl X16

Đặc điểm	Vị trí cặp bazơ *
Gốc F1	3...459
đoạn lặp đầu tận cùng 3'	746...808
Miền đầu tận cùng C (CTD)	nhóm bổ sung (992...1471)
Gen Fib - H trên CTD	nhóm bổ sung (1292...1471)
Maspl X16	nhóm bổ sung (1478...3066)
Miền đầu tận cùng N (NTD)	nhóm bổ sung (3086...4831)
Gen Fib - H trên NTD	nhóm bổ sung (3088...4511)
UEE	nhóm bổ sung (4840...4909)
Gen khởi đầu chính	nhóm bổ sung (4912...6077)
ColE1 ori	nhóm bổ sung (8529...9469)
Alpha LacZ	nhóm bổ sung (8313...8435)
Kháng ampicillin	nhóm bổ sung (9472...10329)
đoạn lặp đầu tận cùng 5'	7784...7818
3xP3 ECFP	6090...7373

* khi bố thế được sử dụng trong trường hợp này có nghĩa là trình tự của dấu hiệu là dài AND đối so với trình tự được mô tả ở đây. Nói một cách khác, nó là bố thế ngược của trình tự được thể hiện ở đây.

Số đếm cơ sở	2764 a	2759 c	2414 g	2521 t
gốc bazơ				
1	ctaaattgta	agcggttaata	ttttgttaaa	attcgcgtta
61	atatttttaac	caataggccg	aaatcggcaa	aatcccttat
121	gatagggttg	agtgtgtgtc	cagtttgtaa	caagagtcca
181	caacgtcaaa	gggcgaaaaa	ccgtctatca	gggcgatggc
241	ctaatacagt	tttttggggt	cgaggtgccg	taaagcacta
301	ccccgatatt	agagcttgac	ggggaaaagcc	ggcgaacgtg
361	agcgaagga	gcgggcgcta	ggcgctggc	aagtgtagcg
421	cacaccgcc	gcgcttaatg	cgccgctaca	gggcgctgcc
481	caactgttg	gaaggcgat	cggtgcgggc	ctcttcgcta
541	gggatgtgct	gcaaggcgat	taagttgggt	aacgccagg
601	taaaaacgacg	gccagtgagc	gcgcctcggt	cattcacgtt
661	gcagactcgc	ggtgcaaatg	tgttttacag	cgtgatggag
721	gctgcagaac	acgcagctag	attaacccta	gaaagataat
781	agataatcat	gcgtaaaatt	gacgcattgtg	ttttatcggg
841	attaatttga	atagatatata	agttttatta	tatttacact
901	aacaaacaat	ttatttatgt	ttattttatt	attaaaaaaa
961	tctataaagt	aacaaaactt	ttatcgaatt	gtatagtatt
1021	tggtgttttt	atatgaacct	attgtaattt	agttttgttt
1081	tggattattg	ttatccctta	aataaatgaa	ttcattgatt
1141	tcaaataaaa	tgcatacaaa	aaatccgaaa	tctgttttct
1201	aataacttaa	atgctctcat	ttattaacga	aaaaaaaac
1261	gtaagaaaca	aggggttattt	tatattaata	attagcaatt
1321	attttaacaac	tagttgtctt	ctaggaattc	cacagttttt
1381	agtcgtaact	gcgagatgaa	gcagatgaca	cagaggaagc
1441	cgtatccct	gccagctccg	taactgacgc	tcttaaggct
1501	accggcgccg	cctgcagcag	ccgcagcgcc	gccagcacct
1561	gcctgcaccc	tgagagccta	gccaccgta	gccaccttga
1621	cgagcgcg	ccagcacctt	ggccaccag	accaccacg

1681 cccaccgtag ccaccttgac cggcgccgccc tgcagcagcc gcagcgccgc cagcaccttg
 1741 gccacccaga ccaccacggc ctgcaccctg agagcctagc ccaccgtagc cacccttgacc
 1801 ggcgcgcctt gcagcagccg cagcgccgccc agcaccttgg ccaccagac caccacggcc
 1861 tgcaccctga gagcctagcc caccgtagcc acccttgacc ggcgcgcctt cagcagccgc
 1921 agcggcgcca gcaccttggc caccagacc accacggcct gcaccttag agcctagccc
 1981 accgtagcca ccttgaccgg cgcgccttgc agcagccgca gcggcgccag caccctggcc
 2041 accagagacca ccacggcctg caccctgaga gcctagccca ccgtagccac cttgaccggc
 2101 gccgcctgca gcagcgcagc cggcgccagc acccttggcca ccagaccac cagcgccctg
 2161 accctgagag cctagcccac cgtagccacc ttgaccggcg ccgctgtag cagccgcagc
 2221 ggcgcagca ccttggccac ccagaccacc acggcctgca cctgtagagc ctgaccaccc
 2281 gttagccact tgaccggcgc cgcctgtagc agcgcagcg gcgccagcac cttggccacc
 2341 cagaccacca cggcctgcac cctgagagcc tagccaccg tagccacctt gaccggcgcc
 2401 gctgtagca gccgcagcgg cgcgcagacc ttggccaccc agaccaccac ggctgaccc
 2461 ctgagagcct agcccaccgt agccaccttgc accggcgccg cctgtagcag ccgagcggc
 2521 gccagcacct tggccaccca gaccaccacg gcctgacccc tgagagccta gccaccgta
 2581 gccaccttga cggcgccgccc ctgcagcagc cgcagcgccg ccagcacctt ggccaccag
 2641 accaccacgg cctgaccctt gagagcctag ccaccgtag ccaccttgac cggcgccgccc
 2701 tgcagcagcc gcagcgccgc cagcaccttg gccaccaga ccaccagggc ctgacccttg
 2761 agagcctagc ccaccgtagc cacccttgacc ggcgcgcctt gcagcagccg cagcgccgccc
 2821 agcaccttgg ccaccagac caccacggcc tgcacctga gagcctagcc caccgtagcc
 2881 accctgacgg gcgcgcgcct cagcagccgc agcgccgcca gcaccttggc caccagacc
 2941 accacggcct gcaccttag agcctagccc accgtagcca ccttgaccgg cgcgccttgc
 3001 agcagccgca gcggcgccag caccctggcc accagacca ccacggcctg caccctgaga
 3061 gcctagggcg cccggggccc atatgagcac tgcagcacta gtgctgaat cgctgtata
 3121 tccttgatga gtgctgtatc cactgtttga tacgtatggc ccgctctgag aatatgctcc
 3181 cgcacacgca gcaacataac tttgagcaac agtaccatcg gaaagtgtct tcatgagcac
 3241 atcttctctt acaatggact cgttaccgtc ggaatccgtg gttataacga acgtcttgat
 3301 cattttttca ttttttccaa gtattccatg gtttttattt ttccgttgca ttttttagt
 3361 tgtaatttgt tcttcgataa ctgcccaga tgcattctta attatttcat ctgtgtatt
 3421 actactttgg acagtacat cactcccaaa atagtctcca tcaaatcat tgatgtttgc
 3481 atttgtataa gcgacatact gaaacaaaat gttaacaatt ttttcttaaa tcaatacgat
 3541 aagtgtctaa atattgagct ataaatttgc cacattcaat cgaagttagc gctgatctgg
 3601 aacgagttag gacatactgc cttagtagtg gtaatagtaa taattgaatg tctagctgg
 3661 ggcgttaggc aagttgaaca acatttctat tatgtaaaca tttgtgggaa agtacaataa
 3721 tgatctctat acaccacgag attttatggt gtagcctatt ccagctacgc tccgatgggt
 3781 tcgagccttg ttttccctac ctattagctg gtagcctatt ccagccttaa caagagcagt
 3841 aggtgagctc tcagactcaa cctgaaagaa tttgctaaca ctgaccctaa caagagcagt
 3901 gcttcataga atctatcacg tgatcggtaa cgcgaccac tgagaagatc tggcgagaaa
 3961 ctcagtgggc agctcgtgtt ataatcacat accaatgtat taaatgttaa cagaatgcta
 4021 cctcgaggtt atgaaaatga ttataatttg cgaaaaaag gcaatagcaa ttgttaatg
 4081 atatcataat aaattaataa tgcggttata attttatgca atttcttctg tctattttct
 4141 aatgatgctt tacgaattgt tttacataat gttgaatatg cattgcataat tgcaatgctg
 4201 atttaccggt gaaataggat attgcaagtc tgcccaggta tttacataga ttcatcttgc
 4261 ctacttttga cgcaataaaa tcacaagtta cataatctaa ggttcatttt ttaataaaa
 4321 aaatgtagtt tattccgtag cccccccggc cgtgacaaat caaatggcac aatttacgat
 4381 atctgtgcgt tttggcgacc acttattatc aggtggcccg tgatctgctc tggccacctt
 4441 ctgaaataat agtaaaataa ttaactcacc tgcagagcgc agcacaaagt cacaaggtt
 4501 ttgactctca Tcttgagagt tggaaaccgaa ctgatgctga tttgaaaaag ttgaaccttt
 4561 ttatactgaa aattttcgag ttttgacgta tttactttcg atagaataaa catctacaca
 4621 ttgtattgta cagatacaat tacatagaaa ttaacttttt tatttaacat aacaatacaa
 4681 cgtatgaagt aaattaaatt gtgatcatga acaatgtgat taattatgat ttaattgatc
 4741 ttattgttct tacttttaatt attttaaaga attttataat atcgttcaat tttttgttaa
 4801 tatttaaaaga acttacttca tgctttctcc cgcggccgccc gaaccctaaa acattgttac
 4861 gttacttgca attaagcact tattcaaaact ttccgtacaa aacatctttc cgcggtgtag

4921 aagacaagcc atcgcaacgg tghtaagggc aggaacttttt tttttgttcc ctacctagct
 4981 gagagccttg agaggctatt tcagcgtaac cttactagt aagttagctc acggggctca
 5041 aacctgacga cgttgctaac acgaatccta gcaagagccg tgcttcgcag aatctaccac
 5101 cggatcggaa acgcgacccg ctgagaagat ccggcgagaa actcagtggt ctgtgtctga
 5161 gggtaatttt gctcgtcgag cccttcgtcg caagcgacgg gtccgacgag aacgatgacc
 5221 ggtgcttgag gtacctaaag caccgttagt ggatcgtgag gatccgaat gacgtgtttg
 5281 gggcgacgtc gactgcttgc cattctgtcc gtaggatcgg gaatggcgcg gcaggacgcg
 5341 tcaaaatcgc cttatttcag gtcccatatt ttcgatataa taaagagcaa cctgtagtt
 5401 tgtgcgtaaa agcttacgtc tgtaactgct gttctggact aatgagagac aataaaccta
 5461 tggagtgcgc gcgaagtaga tttttgtcaa tatcaaacct ttagtgctct tgttttcttc
 5521 tcaactaata gtttctctat gtccatcact cagcattgat tcgtggcaga aataggctag
 5581 atgggtggccc gacagtatgg ttaagtgaag ttacaatgtg ccctattacc agtaaatagc
 5641 agcttagcgat ggctgtatgc tggatgaca ctctgaggat tctgctacgt tcttggtggt
 5701 tctctaagtc gccctttaac gacacccacg gaaagagatg tgaccgtagg ctattttact
 5761 tagttttcta atatcaaaaa ggctgtttgt acctacctag tcaggctata aattctgtca
 5821 catgtttta atgaaataat tgaacaaggt ttattcatta tgtaaccatt catataccaa
 5881 aatgaactta acaaaacata gattcttatg acactaaagt ttattcaaaa tgacctccgt
 5941 gattttgaa atcaggccttc aatctgcgcg gccagtcgtc tatcgacgca cgaacgaggt
 6001 ccatgtcaat atcggcagct gcccttaaca aggatgtctt gactgactcc aaattgggat
 6061 gaggtcttga gcctcgacct agttcttagt ttcccacaa gtgtaattcg agctcggcg
 6121 gggatctaat tcaatttaga actaattcaa tttagagctaa ttcaattagg atccaaagct
 6181 atcgatttcg aaccctcgac cgccggagta taaatagagg cgcttcgtct acggagcgac
 6241 aattcaattc aaacaagcaa agtgaacacg tcgctaagcg aaagctaagc aaataaacaa
 6301 gcgcagctga acaagctaaa caatcggggt accgctagag tcgacggtac gatccaccgg
 6361 tcgcccacct ggtgagcaag ggcgaggagc tgttcaccgg ggtggtgccc atcctggtcg
 6421 agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cgcgagggcg gaggcgatg
 6481 ccacctacgg caagctgacc ctgaagtcca tctgcaccac cgccaagctg cccgtgccct
 6541 ggcccaccct cgtgaccacc ctgacctggg gcgtgcagtg cttcagcccg taccgccacc
 6601 acatgaagca gcacgacttc ttcaagtcgg ccactgccga aggtactatc caggagcgca
 6661 ccactcttct caaggacgac ggcaactaca agaccccgcg cgaggtgaag ttcgagggcg
 6721 acaccctggt gaaccgcatc gagctgaagg gcactcgact caaggaggac ggcaacatcc
 6781 tgggggacaa gctggagtag aactacatca gccacaacgt ctatatcacc gccgacaagc
 6841 agaagaacgg catcaaggcc aacttcaaga tccgccacaa catcgaggac ggcagcgtgc
 6901 agctcgccga ccaactaccg cagaacaccc ccactcgcca cgcccccggt ctgctgcccg
 6961 acaaccacta cctgagcacc cagtccgccc tgagcaaaag ccccaacgag aagcgcgac
 7021 acatggtcct gctggagttc gtgaccgccc ccgggatcac tctcgccatg gacgagctgt
 7081 acaagttaa agcgccgagac tctagatcat aatcagccat accacatttg tagaggtttt
 7141 acttgcttta aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat
 7201 tgtgtgtgtt aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac
 7261 aaatttcaca aataaagcat ttttttact gcattctagt tgtggtttgt ccaaactcat
 7321 caatgtatct taaagcttat cgatacgcgt acggcgcgcc tagggccggc gatactagag
 7381 cggccgccac cgcggtgag ctccagcttt tgttcccttt agtgaggggt aattagatct
 7441 taatacgact cactataggg cgaattgggt accgggcccc cctcgaggt cgacggtatc
 7501 gataagcttg atatctataa caagaaaata tatatataat aagttatcac gtaagttaga
 7561 catgaaataa caatataatt atcgtatgag ttaaatctta aaagtcacgt ttaccgcat
 7621 catgcgtcat tttgactcac gcggtcggtta tagttcaaaa tcagtgacac ttaccgcat
 7681 gacaagcagc cctcacggga gctccaagcg gcgactgaga tgtcctaat gcacagcgac
 7741 ggattcgcgc tatttagaaa gagagagcaa tatttcaaga atgcatgcgt caattttacg
 7801 cagactatct ttctagggtt aatctagctg catcaggatc atatcgtcgg gtcttttttc
 7861 cggctcagtc atcgcccaag ctggcgctat ctgggcatcg gggagggaag agccccgtcc
 7921 ttttcccgcg aggttgaagc ggcattgaaa gagtttgccg aggatgactg ctgctgcatt
 7981 gacgttgagc gaaaacgcac gtttaccatg atgattcggg aaggtgtggc catgcacgac
 8041 tttacggtg aactgttcgt tcaggccacc tgggatacca gttcgtcgcg gcttttccgg
 8101 acacagttcc ggaatggtcag cccgaagcgc atcagcaaac cgaacaatac cggcgacagc

```

8161 cggaactgcc gtgccggtgt gcagattaat gacagcgggtg cggcgctggg atattacgtc
8221 agcgaggacg ggtatcctgg ctggatgccg cagaaatgga catggatacc cgtgagtta
8281 ccggcggggc ggcgttggcg taatcatggt catagctgtt tctgtgtga aattgttatc
8341 cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tggggtgcct
8401 aatgagttag ctaactcaca ttaattgogt tgcgctcact gcccgtttc cagtcgggaa
8461 acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggg ggtttgcgta
8521 ttgggctcct ttcgcttccc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc
8581 gagcgggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg
8641 caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt
8701 tgctggcggt tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa
8761 gtcagaggtg gcgaaacccg acaggactat aaagatacca ggcggtttccc cctggaagct
8821 cctcgtgctg ctctcctgtt ccgaccctgc cgtttaccgg atacctgtcc gcctttctcc
8881 cttcgggaag cgtggcgctt tctcatagct cagcgttagt gtatctcagt tccgtgtagg
8941 tcgttctcgt caagctgggc tgtgtgcacg aacccccgt tcagcccagc cgtgcgctc
9001 tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg cactggcag
9061 cagccactgg taacaggatt agcagagcga ggtatgtagg cgggtgtaca gatttcttga
9121 agtggtggcc taactacggc taaactagaa ggacagtatt tgggtatctgc gctctgctga
9181 agccagttac cttcggaaaa agagttggtg gctcttgatc cggcaaacaa accaccgtg
9241 gtagcgttgg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggtctcaaag
9301 aagatccttt gatctttctt acggggtctg acgctcagtg gaacgaaaaa tcacgttaag
9361 ggatttttgt catgagatta tcaaaaagga tcttcaccta gatcctttta aattzaaaat
9421 gaagttttaa atcaatctaa agtatatatg agtaaaacttg gtctgacagt taccaaatgct
9481 taatcagtga ggcacctatc tcagcgatct gtctatttct tcatccata gttgctgac
9541 tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa
9601 tgataccgog agaccacgc tcaccggctc cagattttatc agcaataaac cagccagccg
9661 gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt
9721 gttgcccggg agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca
9781 ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcatcc agctccggtt
9841 cccaacgac aaggcgagtt acatgatccc ccatgttggt caaaaaagcg gttagctcct
9901 tcggtcctcc gatcgttggtc agaagtaagt tggccgcagt gttatcactc atggttatgg
9961 cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg
10021 agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg
10081 cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgtct atcattggaa
10141 aacgttcttc gggcgaaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt
10201 aaccactcgt tgcacccaac tgatcttcag catcttttac tttcaccagc gtttctgggt
10261 gagcaaaaaa aggaaggcaa aatgccgcaa aaaagggaa aagggcgaca cggaaatgtt
10321 gaatactcat actcttctct tttcaatatt attgaagcat ttatcagggt tattgtctca
10381 tgagcggata catatttgaa tgtatttaga aaaataaaca aataggggtt ccgcgcacat
10441 ttccccgaaa agtgccac

```

//

FIG. 22
pXL Bac II ECFP NTD CTD Masp1 X24

Đặc điểm	Vị trí cặp bazơ *
Gốc fl	3...459
đoạn lặp đầu tận cùng 3'	746...808
Miền đầu tận cùng C (CTD)	nhóm bổ sung (992...1471)
Fig-H gene on CTD	nhóm bổ sung (1292...1471)
Masp1 X24	nhóm bổ sung (1478...3858)
Miền đầu tận cùng N (NTD)	nhóm bổ sung (3878...5623)
Gen Fib-H trên NTD	nhóm bổ sung (3880...5303)
UEE	nhóm bổ sung (5632...5701)
Gen khởi đầu chính	nhóm bổ sung (5704...6869)
3xP3 ECFP	6882...8165
đoạn lặp đầu tận 5'	8576...8610
Alpha LacZ	nhóm bổ sung (9105...9227)
đoạn lặp đầu tận cùng 5'	nhóm bổ sung (9321...10261)
Kháng ampicillin	nhóm bổ sung (10264...11121)

* khi bố thế được sử dụng trong trường hợp này có nghĩa là trình tự của dấu hiệu là dài AND đối so với trình tự được mô tả ở đây. Nói một cách khác, nó là bố thế ngược của trình tự được thể hiện ở đây.

Số đếm cơ sở	2908 a	3127 c	2622 g	2593 t			
gốc bazơ	1	ctaaattgta	agcggttaata	ttttgttaaa	attcgcgtta	aatttttgtt	aaatcagctc
61	atttttttaac	caataggccg	aaatcggtcaa	aatcccttat	aaatcaaaag	aatagaccga	acgtggactc
121	gatagggttg	agtgtttgtc	cagtttgtaa	caagagtcca	ctattaaaga	acgtggactc	
181	caacgtcaaa	ggcgaaaaa	cgctctatca	ggcgatggc	ccactacgtg	aaccatcacc	
241	ctaatacaagt	tttttggggt	cgaggtgccc	taaagcacta	aatcggaacc	ctaaagggag	
301	cccccgattt	agagcttgac	ggggaaaagcc	ggcgaacgtg	gcgagaaaag	aaggyaagaa	
361	agcgaaaagga	gcgggcgcta	ggcgctggc	aagtgtagcg	gtcacgctgc	gcgtaaccac	
421	cacacccgcc	gcgcttaatg	cgccgcgtaca	ggcgctgcc	cattcgccat	tcaggtctgc	
481	caactgttgg	gaagggcgat	cggtgcgggc	ctcttcgcta	ttacgccagc	tgccgaaaag	
541	gggatgtgct	gcaagggcat	taagttgggt	aacgccaggg	ttttcccgat	cacgacgttg	
601	taaaacgacg	gccagtgagc	gcgcctcggt	cattcacgtt	tttgaacccg	tgaggagcgg	
661	gcagactcgc	ggtgcaaatg	tgttttacag	cgtgatggag	cagatgaaga	tgctcgacac	
721	gctgcagaac	acgcagctag	attaacccta	gaaagataat	catattgtga	cgtagcttaa	
781	agataatcat	gcgtaaaatt	gacgcattgt	ttttatcggt	ctgtatatcg	aggttttatt	
841	attaatttga	atagatatta	agttttatta	tatttacact	tacatactaa	taataaatcc	
901	aacaaacaat	ttatttatgt	ttattttatt	attaaaaaaa	aacaaaaact	caaaatttct	
961	tctataaagt	aacaaaactt	ttatcgattt	gtatagtatt	cttagttgag	aaggcatact	
1021	tggtgttttt	atatgaacct	attgtaattt	agttttgttt	taaatgtgca	tgtgaattca	
1081	tggtattattg	ttatccctta	aataaatgaa	ttcattgatt	aattgattat	attagtagat	
1141	tcaataaaaa	tgcatacaaa	aaatccgaaa	tctgttttct	aaaaaaaagt	aattaaaaac	
1201	aataacttaa	atgctctcat	ttattaacga	aaaaaaaac	atagatgtat	ccaggacgaa	
1261	gtaagaaaca	aggggttatt	tattattaaa	attagcaatt	cacacaaggc	agtgctctga	
1321	atttaacaac	tagttgtcct	ctaggaattc	cacagttttt	gcggacgtta	cgacgagaat	
1381	agtcgtaact	gcgagatgaa	gcagatgaca	cagaggaagc	tgactttcct	gcacctttgc	
1441	cgtatccctt	gccagctccg	taactgacgc	tcttaaggct	agcccaccgt	agccaccttg	
1501	accggcgccg	cctgcagcag	ccgcagcggc	gccagcacct	tgccacccca	gaccaccacg	
1561	gcctgcaccc	tgagagccta	gccacaccga	gccaccttga	ccggcgccgc	ctgcagcagc	
1621	cgcgagcgcg	ccagcacctt	ggccacccag	accaccacgg	cctgcacctt	gagagcctag	

```

1681 cccaccgtag ccaccttgac cggcgccgccc tgcagcagcc gcagcggcgc cagcaccttg
1741 gccacccaga ccaccacggc ctgcaccctg agagcctagc ccaccgtagc caccttgacc
1801 ggcgcgcgct gcagcagccg cagcggcgcc agcaccttgg ccaccagac caccacggcc
1861 tgcaccctga gagcctagcc caccgtagcc accctgaccg gcgcgcgctg cagcagccgc
1921 agcggcgcca gcaccttggc caccagacc accacggcct gcacctgag agcctagccc
1981 accgtagcca ccttgaccgg cgccgcctgc agcagccgca gcggcgccag caccttggcc
2041 acccagacca ccacggcctg caccctgaga gcctagccca ccgtagccac cttgaccggc
2101 gccgcctgca gcagcgcgag cggcgccagc acccttggcca ccagaccac caccgctgc
2161 accctgagag cctagccacc cgtagccacc ttgaccggcg ccgcctgag cagccgcagc
2221 ggcgcagca ccttggccac ccagaccacc acggcctgca ccctgagag ctagccacc
2281 gtaccaccc tgcgcggcg cgctgcagc agccgcagcg gcgccagcac cttggccacc
2341 cagaccacca cgcctgcac cctgagagcc tagccaccg tagccacctt gaccggcgcc
2401 gcctgcagca gccgcagcgg cgcagcacc ttggccacc agaccaccac ggccctgcac
2461 ctgagagcct agccacccgt agccacccgt accggcgcc cctgcagcag ccgcagcgcc
2521 gccagcacc ttggccacca gaccaccacg gcctgcaccc tgagagccta gccaccgta
2581 gccacccctga ccggcgccc ctgcagcag cgcagcggcg ccagcacctt ggccaccag
2641 accaccacgg cctgcacctt gagagcctag cccaccgtag ccaccttgac cggcgccgccc
2701 tgcagcagcc gcagcggcg cagcaccttg gccaccaga ccaccagcg ctagccacctg
2761 agagccctagc ccaccgtagc caccctgacc ggcgcgcctt gcagcagccg cagcggcgcc
2821 agcaccttgg ccaccagac caccacggcc tgcaccctga gagcctagcc caccgtagcc
2881 acccttagcgg gcgcgcgctg cagcagcggc agcggcgcca gcaccttggc caccagacc
2941 accacggcct gcaccttgag agcctagccc accgtagcca ccttgaccgg cgccgctgc
3001 agcagcggca gcggcgccag cacccttgcc acccagacca ccacggcctg caccctgaga
3061 gcctagccca ccgtagccac cttgaccggc gccgcctgca gcagcgcgag cggcgccagc
3121 acccttgcca cccagaccac caccgctgc accctgagag cctagccacc cgtagccacc
3181 ttgaccggcg ccgcctgag cagcgcgagc ggcgcagca ccttggccac ccagaccacc
3241 accggcctgca ccctgagag ctagccacc gtaccacctt gaccggcgcc cgccctgagc
3301 agccgcagcg gcgcagcac cttggccacc cagaccacca cggcctgcac cctgagagcc
3361 tagccaccgg tagccacctt gaccggcgcc gcctgcagca gccgcagcgg cgcagcacc
3421 ttggccacc agaccaccac ggccctgcac ctgagagcct agccaccgt agccacctt
3481 accggcgccg cctgcagcag ccgcagcggc gccagcacc ttggccacca gaccaccacg
3541 gcctgcaccc tgagagccta gccaccgta gccaccttga ccggcgccgc ctgcagcagc
3601 cgcagcggcg ccagcacctt ggccaccacg accaccacgg cctgcacctt gagagcctag
3661 cccaccgtag ccaccttgac cggcgccgccc tgcagcagcc gcagcggcg cagcaccttg
3721 gccaccaga ccaccacggc ctgcaccctg agagcctagc ccaccgtagc caccttgacc
3781 ggcgcgcgct gcagcagccg cagcggcgcc agcaccttgg ccaccagac caccacggcc
3841 tgcaccctga gagcctaggg cgccggggcc acatagtagc actgcagcac tagtgctgaa
3901 atcgctcgta tatccttgat gaggctgta tccactgttt gatacgtatg gcccgctctg
3961 agaataatg cccgcatcag cagcaacata actttgagca acagtacctt cggaaagtgt
4021 cttcatgagc acatcttcc ctacaatgga ctcggtaccg tcggaatccg tgggtataac
4081 gaacgtcttg atcatttttt cattttttcc aagtattcca tgggttttat ttttccgttg
4141 cattttttta gttgtaattt gttcttcgat aactgcccc gatgcatttc taattatttc
4201 atctgttgta ttaactactt ggacagtgc atcaactcca aaatagtcct catcaaaatc
4261 attgatgttt gcaatttgat aagcgacata ctgaaacaaa atgttaacaa tttttcttta
4321 aatcaatacg ataagtgtta aaatattgag ctataaattt cgcacattca atcgaagtta
4381 gcgctgatct ggaacgagtt aggcataact gccttagtag tggtaatagt aataattgaa
4441 tgttctagct gggcgcttag gcaagttgaa caacattcta attatgtaaa catttggtgg
4501 aaagtacata attgtatctc atacaccacg agattttatg gtcacattat gttgttatta
4561 cttgagcttg tttcgagcct tgttttccct accatttagc tggtagccta ttccagctac
4621 gtcctgatgg gtaggtgagc tctcagactc aacctgaaag aatttgctaa cactagccct
4681 aacaagagca gtgcttcata gaattctatc cgtgatcggt aacgcgaccc actgagaaga
4741 tctggcgaga aactcagtg gcagctcgtg ttataatcac ataccaatgt attaaaatgt
4801 aacagaatgc tacctcgagg ttatgaaat gattataatt tgcgaaaaaa aggcaatagc
4861 aattgttaat agatatcata ataaattaat aatgcggtta taattttatg caatttcttt

```

4921 cgtctatattt ctaatgatgc tttacgaatt gttttacata ttgttgaata tgcattgcat
 4981 attgcaatgc tgatttaccg gtgaaatagg atattgcaag tctgcccagg tattttacata
 5041 gattcatctt gcctactttt gacgcaaata aatcacaagt tacataatct aaggttcatt
 5101 ttttaataaa ataaatgtag tttattccgt agccccccg ggcgtgacaa atcaaatggc
 5161 acaatttacg atatctgtgc gttttggcga ccacttatta tcagggtggc cgtgatatcg
 5221 tctggccacc ttctgaaata atagtataat aattaactca cctgcagagc gcagcacaag
 5281 atcacaaagg ttttgactct catcttgaga gttggaaccg aactgatgct gatttgaata
 5341 agttgaacct ttttatactg aaaattttcg agttttgacg tatttacttt cगतगाता
 5401 aacatctaca cattgtattg tacagataca attacataga aattaatctt tttatttaac
 5461 ataacaatac aacgtatgaa gtaaatataa ttgtgatcat gaacaatgtg attaatattg
 5521 atttaattga tcttattgtt cttactttta atatttttaa gaattttata atatcgttca
 5581 atttttttgt aatattttaa gaacttactt catgctttct cccgcggccg ccgaacccta
 5641 aaacattgtt acgttacttg caattaagca cttattcaaa ctttccgtac aaaaatctt
 5701 tccgcgggtg agaagacaag ccatcgcaac ggtgttaagg gcaggacttt tttttttgtt
 5761 tcctacacag ctgagagcct tgagaggcta tttcagcgta accttaacta gtaagtggagc
 5821 tcacgggggt caaacctgac gacgttgcta acacgaatcc tagcaagagc cgtgcttcgc
 5881 agaattctacc accggatcgg aaacgcgacc cgctgagaag atccggcgag aaactcagtg
 5941 ggctgtgtct gagggttaat ttgctcgtcg agcccttctg cgcaagcgac gggctccgagc
 6001 agaacgatga ccggtgcttg aggtacctaa agcacggtta gtggatcgtg aggatccgaa
 6061 atgacgtggt tggggcgacg tcgactgctt gccattctgt ccgtaggacg ggggaatgggc
 6121 gggcaggagc cgtcaaaatc gcccttattc aggtcccata ttttcgatac aataaagagc
 6181 aacctgttag tttgtgcgta aaagcttacg tctgtaactg ctgtttctgga ctaatgagag
 6241 acaataaacc tatggagtga cggcgaagta gatttttctc aatatcaaac ctttagtgct
 6301 cttgtttttct tctcaactaa tagttttctc atgtccatca ctacagattg attcgtggca
 6361 gaaataggct agatggtggc ccgacagtat ggttaagtga gattacaatg tgccttatta
 6421 ccagttaaata gcagttagcg atggctgtat gctggatga cactctgagg attctgctac
 6481 gttcttgggt gttctctaag tcgcccttta acgacaccca cggaaagaga tgtgaccgta
 6541 ggctattttt cttagttttc taatatcaaa aaggctgttt gtacctacct agtcagggtca
 6601 taaattctgt cacatgttta atgtaaaata attgaaacaa gtttattcat tatgtaacca
 6661 tcatatatac aaaatgaact taacaaaaca tagatttcta tgacactaaa gtttattcaa
 6721 aatgacctcc gtgattttga atacaggcct tcaatctgag cggccagtcg tctatcgag
 6781 caccgaacgag gtccatgtca atatcggcag ctgccttaac caaggatgtc ttgagtgaat
 6841 ccaaatggg atgaggcttt gagcctcgac ctagtcttag tgttccaca atgggttaatt
 6901 cgagctcgcc cggggatcta attcaattag agactaattc aattagagct aattcaatta
 6961 ggatccaagc ttatcgattt cgaaccctcg accgcggag tataaataga ggcgcttcgt
 7021 ctacggagcg acaattcaat tcaacaagc aaagtgaaca cgtcgctaa cgaagactaa
 7081 gcaaaataaac aagcgagct gaacaagcta acaatcggg gtaccgtag agtcgacggt
 7141 acgatccacc ggtcgccacc atggtgagca agggcgagga gctgttcacc ggggtggtgc
 7201 ccatacctggt cgagctggac ggcgacgtaa acggccacaa gttcagcggt tccggcgagg
 7261 ggcaggggcga tgccacctac ggcaagctga cctgaaagt catctgcacc accggcaagc
 7321 tgcccggtgcc ctggccacc ctggtgacca cctgacctg gggcggtcag tgcctcagcc
 7381 gctaccccgca ccacatgaag cagcacgact tcttcaagtc cgccatgccc gaaggctacg
 7441 tccaggagcg caccatcttc ttcaaggacg acggcaacta caagaccgc gccgaggtga
 7501 agttcgaggg cgacaccctg gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg
 7561 acggcaacat cctggggcac aagctggagt acaactacat cagccacaac gtctatatca
 7621 ccgcccagcaa gcagaagaac ggcatacagg ccaacttcaa gatccgccac aacatcgagg
 7681 acggcagcgt gcagctcgcc gaccactacc agcagaacac ccccatcgcc gacggccccc
 7741 tgctgctgcc cgacaaccac tacctgagca cccagtccgc cctgagcaaa gaccccaacg
 7801 agaagcgcg tcacatggtc ctgctggagt tcgtgaccgc cgcggggtac actctcggca
 7861 tggacgagct gtacaagtaa agcggccgag actctagatc ataatcagcc ataccacatt
 7921 tgtagaggtt ttacttgctt taaaaaacct cccacacct cccctgaacc tgaacataa
 7981 aatgaatgca attgtgtgtg ttaactgtt tattgcagct tataatggtt acaataaag
 8041 caatagcatc acaaatttca caaataaagc attttttca ctgcattcta gttgtggtt
 8101 gtccaaactc atcaatgtat cttaaagctt atcgatacgc gtacggcgcg cctagggcgg

```

8161 ccgatactag agcggccgccc accgcggtgg agctccagct tttgttcct ttagtgaggg
8221 ttaattagat cttaatacga ctcactatag ggcgaattgg gtaccgggcc cccctcgag
8281 gtcgacggta tcgataagct tgatatctat aacaagaaaa tatatatata ataagttatc
8341 acgtaagtag aacatgaaat aacaatataa ttatcgtatg agttaaatct taaaagtcac
8401 gtaaaagata atcatgctgc attttgactc acgcggtcgt tatagttcaa aatcagtgac
8461 acctaccgca ttgacaagca cgcctcacgg gagctccaag cggcgactga gatgtcctaa
8521 atgcacagcg acggattcgc gctatttaga aagagagagc aatatttcaa gaatgcattc
8581 gtcaatttta cgcagactat ctttctaggg agctggcgct atctgggcat cggggaggaa
8641 gggctctttt tccggctcag tcacgcgcca agctggcgct atctgggcat cggggaggaa
8701 gaagcccggt ccttttcccg cgaggttgaa gcggcatgga aagagtttgc cgaggtatgc
8761 tgctgctgca ttgacgttga gcgaaaacgc acgtttacca tgatgattcg ggaaggtgtg
8821 gccatgcacg cctttaacgg tgaactgttc gttcaggcca cctgggatac cagtccgtcg
8881 cggcttttcc ggacacagtt ccggatggtc agcccgaaag gcatcagcaa cccgaacaat
8941 accggcgaca gccggaactg ccgtgcccgt gtgcagatta atgacagcgg tcggggcgctg
9001 ggatactacg tcagcgaggga cgggtatcct ggctggatgc cgagaaatg gacatggata
9061 ccccgtaggt tacccggcgg gcgcgcttgg cgtaatcatg gtcatactgt tttcctgtgt
9121 gaaattgtta tccgctcaca attccacaca acatacgagc cgggaagcata aagtgtaaag
9181 cctgggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgtct
9241 tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag
9301 gcgggtttcg tattggggcg tcttccgctt cctcgcctac tgactcgtcg cgctcggtcg
9361 ttcggctgcg gcgagcggtt tcagctcact caaaggcggt aatacggtta tccacagaat
9421 caggggataa cgcaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta
9481 aaaaggccgc gttgctggcg tttttccata ggctccgccc cctgacgag catcacaaaa
9541 atcgacgctc aagtcagagg tggcgaaaac cgacaggact ataaagatac caggcgtttc
9601 cccctggaa gctcctcgtg cgtctcctcg ttcggacctt gccgcttacc ggtacctgtg
9661 ccgcttcttc ccttccggga agcgtggcgc tttctcatag ctcacgctgt aggtatctca
9721 gttcgggtga ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccc
9781 accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cagcacttat
9841 cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta
9901 cagagttctt gaagtgtgg cctaactacg gctacactag aaggacagta tttggtatct
9961 gcgctctgct gaagccagtt accctcggaa aaagagttgg tagctcttga tccggcaaac
10021 aaaccaccgc tggtagcggt ggtttttttg tttgcaagca gcagattacg cgcagaaaaa
10081 aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaaacaaa
10141 actcacgtta agggattttg gtcatgagat tatcaaaaag gatcttcacc tagatccttt
10201 taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaaact tggctcgaca
10261 gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctatct cgtttaccca
10321 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggaggggcta ccatctggcc
10381 ccagtgcctg aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa
10441 accagccagc cggaagggcc gagcgagaaa gtgttcctgc aactttatcc gcttccatcc
10501 agtctattaa ttgttgcggg gaagctagag taagtagttc gccagttaat agtttgcgca
10561 acgttgttgc cattgttaca ggcacgttgc tgcacgctc gtcgtttggt atggttcat
10621 tcagctccgg tttcccaacga tcaaggcgag ttacatgac ccccatgttg tgcaaaaaag
10681 cggttagctc cttcgggtcct ccatcgttgc tcagaagtaa gttggccgca gtgttatcac
10741 tcattggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgtttt
10801 ctgtgactgg tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt
10861 gctcttgccc ggcgtcaata cgggataata cgcgcccaca tagcagaact ttaaaagtgc
10921 tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat
10981 ccagttcgat gtaacccact cgtgcaccca actgatcttc agcatctttt actttcacca
11041 cgtgttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgga
11101 caggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg
11161 gttattgtct catgagcgga tacatatattg aatgtattta gaaaaataaa caaatagggg
11221 ttccgcgcac atttcccga aaagtcccac

```

//

DANH MỤC TRÌNH TỰ

<110> THE UNIVERSITY OF NOTRE DAME

<120> AXIT NUCLEIC MÃ HÓA POLYPEPTIT TỌ NHỆN KHẢM, POLYPEPTIT TỌ NHỆN KHẢM, TỌ COMPOZIT CHỨA POLYPEPTIT TỌ NHỆN KHẢM VÀ PHƯƠNG PHÁP TẠO RA TẦM CHUYỂN GEN

<130> 343414-00063

<140> PCT/US2011/053760

<141> 2011-09-28

<150> 61/387,332

<151> 2010-09-28

<160> 35

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Peptit tổng hợp

<400> 1

Gly	Pro	Gly	Gly	Gln	Gly	Pro	Gly	Gly	Tyr	Gly	Pro	Gly	Gly	Gln	Gly
1				5					10					15	

Pro	Gly	Gly	Tyr
			20

<210> 2

<211> 5

<212> PRT

<213> Chưa biết

<220>

<223> Mô tả trình tự chưa biết: Trình tự motif đàn hồi tơ hình roi kiểu xung

<400> 2

Gly	Pro	Gly	Gly	Ala
1				5

<210> 3

<211> 16

<212> PRT

<213> Trình tự chưa biết

<220>

<223> Mô tả trình tự chưa biết: Trình tự môtip bền của tơ dầy kéo giả định

<400> 3

Gly Gly Pro Ser Gly Pro Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
 1 5 10 15

<210> 4

<211> 40

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polypeptit tổng hợp

<400> 4

Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly
 1 5 10 15

Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro
 20 25 30

Gly Gly Ala Gly Pro Gly Gly Ala
 35 40

<210> 5

<211> 8

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Peptit tổng hợp

<400> 5

Ala Ala Ala Ala Ala Ala Ala Ala
 1 5

<210> 6

<211> 20

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Peptit tổng hợp

<400> 6

Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly Pro Gly Gly Ala Gly

1

5

10

15

Pro Gly Gly Ala
20

<210> 7
<211> 35
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 7
taactcgagg ctcaaagcct catcccaatt tggag 35

<210> 8
<211> 36
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 8
ataccgcggt gcagaagaca agccatcgca acggtg 36

<210> 9
<211> 36
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 9
ataccgcgga aagatgtttt gtacggaaag tttgaa 36

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

<220>
<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 10
ttagcggccg ccgaacccta aaacattgtt acgttacgtt acttg 45

<210> 11
 <211> 51
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 11
 taagcggccg cgggagaaaag catgaagtaa gttctttaaa tattacaaaa a

51

<210> 12
 <211> 40
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 12
 ataggatcca cgactgcagc actagtgctg ctgaaatcgc

40

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

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 13
 atatctagaa cgactgcagc actagtgctg ctgaaatcgc

40

<210> 14
 <211> 37
 <212> ADN
 <213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 14
 caatctagac gtgagcaagg gcgaggagct gttcacc

37

<210> 15
 <211> 37
 <212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 15

taaggatcca gcttgtagacg ctctgtccatg ccgagag

37

<210> 16

<211> 33

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit tổng hợp

<400> 16

ataccggga agcgtagtt acggagctgg cag

33

<210> 17

<211> 37

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Oligonucleotit Tổng hợp

<400> 17

caagctgact atagtattct tagttgagaa ggcatac

37

<210> 18

<211> 33

<212> PRT

<213> Nephila clavipes

<400> 18

Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gly Leu Gly Ser Gln Gly Ala
1 5 10 15

Gly Arg Gly Gly Tyr Gly Gly Gln Gly Ala Gly Ala Ala Ala Ala Ala
20 25 30

Ala

<210> 19

<211> 26

<212> PRT

<213> *Lactrodectus geometricus*

<400> 19

Gly Gly Ala Gly Gln Gly Gly Tyr Gly Gln Gly Gly Gln Gly Gly Ala
1 5 10 15

Gly Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
20 25

<210> 20

<211> 34

<212> PRT

<213> *Argiope trifasciata*

<220>

<221> MOD_RES

<222> (14)..(14)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (31)..(31)

<223> Axit amin bất kỳ

<400> 20

Gly Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Xaa Gln Gly
1 5 10 15

Ala Gly Gln Gly Tyr Gly Ala Gly Ser Gly Gly Gln Gly Gly Xaa Gly
20 25 30

Gln Gly

<210> 21

<211> 40

<212> PRT

<213> *Nephila clavipes*

<400> 21

Gly Pro Gly Gln Gln Gly Pro Gly Gly Tyr Gly Pro Gly Gln Gln Gly
1 5 10 15

Pro Gly Gly Tyr Gly Pro Gly Gln Gln Gly Pro Ser Gly Pro Gly Ser
20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala
35 40

<210> 22
 <211> 29
 <212> PRT
 <213> *Lactrodectus geometricus*

<220>
 <221> MOD_RES
 <222> (11)..(11)
 <223> Axit amin bất kỳ

<400> 22
 Gly Pro Gly Gly Tyr Gly Pro Gly Pro Gly Xaa Gln Gln Gly Tyr Gly
 1 5 10 15

Pro Gly Gly Ser Gly Ala Ala Ala Ala Ala Ala Ala Ala
 20 25

<210> 23
 <211> 32
 <212> PRT
 <213> *Argiope trifasciata*

<400> 23
 Gly Pro Gly Gly Gln Gly Pro Gly Gln Gln Gly Pro Gly Gly Tyr Gly
 1 5 10 15

Pro Ser Gly Pro Gly Gly Ala Ser Ala Ala Ala Ala Ala Ala Ala Ala
 20 25 30

<210> 24
 <211> 4949
 <212> PRT
 <213> *Nephila clavipes*

<400> 24
 Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala
 1 5 10 15

Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln
 20 25 30

Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Gly
 35 40 45

Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly
 50 55 60

Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly
65 70 75 80

Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala
85 90 95

Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala
100 105 110

Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly
115 120 125

Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala
130 135 140

Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly
145 150 155 160

Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr
165 170 175

Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala
180 185 190

Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala
195 200 205

Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly
210 215 220

Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala
225 230 235 240

Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly
245 250 255

Ala Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly
260 265 270

Gly Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala
275 280 285

Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg

26612

290				295				300							
Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala
305					310					315					320
Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly
			325					330						335	
Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly
			340					345						350	
Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly
		355						360						365	
Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala
	370						375				380				
Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr
385						390				395					400
Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala
			405							410					415
Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly
			420							425				430	
Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly
			435				440							445	
Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly
		450					455							460	
Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala
465						470				475					480
Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly
										490				495	
Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala
			500							505				510	
Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly
			515							520				525	

Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala
 530 535 540

Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Gly
 545 550 555 560

Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr Gly Ala
 565 570 575

Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly
 580 585 590

Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala
 595 600 605

Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr Gly
 610 615 620

Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly
 625 630 635 640

Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala
 645 650 655

Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr
 660 665 670

Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala Gly Ala
 675 680 685

Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala
 690 695 700

Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly
 705 710 715 720

Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Gly Ala Gly
 725 730 735

Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly
 740 745 750

Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly
755 760 765

Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala
770 775 780

Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala
785 790 795 800

Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln
805 810 815

Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly
820 825 830

Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly
835 840 845

Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly
850 855 860

Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Ala
865 870 875 880

Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala
885 890 895

Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly
900 905 910

Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala
915 920 925

Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly
930 935 940

Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr
945 950 955 960

Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala
965 970 975

26612

Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala
980 985 990

Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Ala Gly Gly
995 1000 1005

Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala
1010 1015 1020

Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly
1025 1030 1035

Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala
1040 1045 1050

Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala
1055 1060 1065

Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala
1070 1075 1080

Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala
1085 1090 1095

Gly Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr
1100 1105 1110

Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala Gly
1115 1120 1125

Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala Gly Ala Gly Ala
1130 1135 1140

Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly Gly Tyr Gly Gly
1145 1150 1155

Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala
1160 1165 1170

Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala
1175 1180 1185

Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly

26612

1190		1195		1200
Gly Tyr 1205	Gly Gly Gln Gly 1210	Gly Tyr Gly Ala Gly 1215	Ala Gly Ala Gly	
Ala Ala 1220	Ala Ala Ala Gly 1225	Ala Gly Ala Gly Gly 1230	Gly Gly Tyr	
Gly Arg 1235	Gly Ala Gly Ala 1240	Ala Gly Ala Ala 1245	Gly Ala Gly	
Ala Gly 1250	Ala Gly Gly Tyr 1255	Gly Gln Gly Gly Tyr 1260	Gly Ala Gly	
Ala Gly 1265	Ala Gly Ala Ala 1270	Ala Ala Gly Ala 1275	Ala Gly Gly	
Ala Gly 1280	Gly Tyr Gly Arg 1285	Ala Gly Ala Gly 1290	Gly Ala Ala	
Ala Gly 1295	Ala Gly Ala Gly 1300	Gly Gly Tyr Gly 1305	Gln Gly Gly	
Tyr Gly 1310	Ala Gly Ala Gly 1315	Gly Ala Ala Ala 1320	Ala Gly Ala	
Gly Ala 1325	Gly Gly Ala Gly 1330	Tyr Gly Arg Gly 1335	Gly Ala Gly	
Ala Gly 1340	Ala Ala Ala Gly 1345	Gly Ala Gly Ala 1350	Gly Tyr Gly	
Gly Gln 1355	Gly Gly Tyr Gly 1360	Gly Ala Gly Ala 1365	Ala Ala Ala	
Ala Ala 1370	Gly Ala Gly Ala 1375	Gly Ala Gly Gly 1380	Gly Arg Gly	
Ala Gly 1385	Ala Gly Ala Gly 1390	Ala Ala Gly Ala 1395	Ala Gly Ala	
Gly Gly 1400	Tyr Gly Gly Gln 1405	Gly Tyr Gly Ala 1410	Ala Gly Ala	

Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly
1415						1420					1425			
Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala
1430						1435					1440			
Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala
1445						1450					1455			
Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly
1460						1465					1470			
Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala
1475						1480					1485			
Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly
1490						1495					1500			
Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly
1505						1510					1515			
Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala
1520						1525					1530			
Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr
1535						1540					1545			
Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala
1550						1555					1560			
Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg
1565						1570					1575			
Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly
1580						1585					1590			
Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly
1595						1600					1605			
Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly
1610						1615					1620			

Gly Tyr	Gly Arg	Gly Ala	Gly	Ala Gly	Ala Gly	Ala	Ala Ala Gly
1625			1630			1635	
Ala Gly	Ala Gly	Ala Gly	Gly	Tyr Gly	Gly Gln	Gly	Gly Tyr Gly
1640			1645			1650	
Ala Gly	Ala Gly	Ala Gly	Ala	Ala Ala	Ala Ala	Gly	Ala Gly Ala
1655			1660			1665	
Gly Gly	Ala Gly	Gly Tyr	Gly	Arg Gly	Ala Gly	Ala	Gly Ala Gly
1670			1675			1680	
Ala Ala	Ala Gly	Ala Gly	Ala	Gly Ala	Gly Gly	Tyr	Gly Gly Gln
1685			1690			1695	
Gly Gly	Tyr Gly	Ala Gly	Ala	Gly Ala	Gly Ala	Ala	Ala Ala Ala
1700			1705			1710	
Gly Ala	Gly Ala	Gly Gly	Ala	Gly Gly	Tyr Gly	Arg	Gly Ala Gly
1715			1720			1725	
Ala Gly	Ala Gly	Ala Ala	Ala	Gly Ala	Gly Ala	Gly	Ala Gly Gly
1730			1735			1740	
Tyr Gly	Gly Gln	Gly Gly	Tyr	Gly Ala	Gly Ala	Gly	Ala Gly Ala
1745			1750			1755	
Ala Ala	Ala Ala	Gly Ala	Gly	Ala Gly	Gly Ala	Gly	Gly Tyr Gly
1760			1765			1770	
Arg Gly	Ala Gly	Ala Gly	Ala	Gly Ala	Ala Ala	Gly	Ala Gly Ala
1775			1780			1785	
Gly Ala	Gly Gly	Tyr Gly	Gly	Gly	Gln Gly	Gly Tyr	Ala Gly Ala
1790			1795			1800	
Gly Ala	Gly Ala	Ala Ala	Ala	Ala	Ala Gly	Ala Gly	Gly Gly Ala
1805			1810			1815	
Gly Gly	Tyr Gly	Arg Gly	Ala	Gly Ala	Gly Ala	Gly	Ala Ala Ala
1820			1825			1830	

Gly Ala 1835	Gly Ala Gly Ala 1840	Gly Tyr Gly Gly Gln 1845	Gly Gly Tyr
Gly Ala 1850	Gly Ala Gly Ala 1855	Ala Ala Ala Ala Ala 1860	Gly Ala Gly
Ala Gly 1865	Gly Ala Gly Gly Tyr 1870	Gly Arg Gly Ala Gly 1875	Ala Gly Ala
Gly Ala 1880	Ala Ala Gly Ala 1885	Ala Gly Ala Gly Gly 1890	Tyr Gly Gly
Gln Gly 1895	Gly Tyr Gly Ala 1900	Ala Gly Ala Gly Ala 1905	Ala Ala Ala
Ala Gly 1910	Ala Gly Ala Gly 1915	Ala Gly Gly Tyr Gly 1920	Arg Gly Ala
Gly Ala 1925	Gly Ala Gly Ala 1930	Ala Gly Ala Gly Ala 1935	Gly Ala Gly
Gly Tyr 1940	Gly Gly Gln Gly 1945	Tyr Gly Ala Gly Ala 1950	Gly Ala Gly
Ala Ala 1955	Ala Ala Ala Gly 1960	Gly Ala Gly Gly Ala 1965	Gly Gly Tyr
Gly Arg 1970	Gly Ala Gly Ala 1975	Ala Gly Ala Ala Ala 1980	Gly Ala Gly
Ala Gly 1985	Ala Gly Gly Tyr 1990	Gly Gln Gly Gly Tyr 1995	Gly Ala Gly
Ala Gly 2000	Ala Gly Ala Ala 2005	Ala Ala Gly Ala Gly 2010	Ala Gly Gly
Ala Gly 2015	Gly Tyr Gly Arg 2020	Ala Gly Ala Gly Ala 2025	Gly Ala Ala
Ala Gly 2030	Ala Gly Ala Gly 2035	Gly Gly Tyr Gly Gly 2040	Gln Gly Gly
Tyr Gly	Ala Gly Ala Gly Ala	Gly Ala Ala Ala Ala	Ala Gly Ala

26612

2045	2050	2055
Gly Ala 2060	Gly Gly Ala Gly Gly 2065	Tyr Gly Arg Gly Ala 2070
Ala Gly 2075	Ala Ala Ala Gly Ala 2080	Gly Gly Tyr Gly 2085
Gly Gln 2090	Gly Gly Tyr Gly Ala 2095	Gly Ala Ala Ala 2100
Ala Ala 2105	Gly Ala Gly Ala Gly 2110	Tyr Gly Arg Gly 2115
Ala Gly 2120	Ala Gly Ala Gly Ala 2125	Ala Gly Ala 2130
Gly Gly 2135	Tyr Gly Gly Gln Gly 2140	Gly Tyr Gly Ala Gly 2145
Gly Ala 2150	Ala Ala Ala Ala Gly 2155	Ala Gly Ala Gly 2160
Tyr Gly 2165	Arg Gly Ala Gly Ala 2170	Gly Ala Gly Ala 2175
Gly Ala 2180	Gly Ala Gly Gly Tyr 2185	Gly Gly Gln Gly 2190
Gly Ala 2195	Gly Ala Gly Ala Ala 2200	Ala Gly Ala Gly 2205
Gly Ala 2210	Gly Gly Tyr Gly Arg 2215	Gly Ala Gly Ala 2220
Ala Ala 2225	Gly Ala Gly Ala Gly 2230	Gly Gly Tyr Gly 2235
Gly Tyr 2240	Gly Ala Gly Ala Gly 2245	Ala Ala Ala Gly 2250
Ala Gly 2255	Ala Gly Gly Ala Gly 2260	Gly Tyr Gly Arg Gly 2265

Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr
2270						2275					2280			
Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala
2285						2290					2295			
Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg
2300						2305					2310			
Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly
2315						2320					2325			
Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly
2330						2335					2340			
Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly
2345						2350					2355			
Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly
2360						2365					2370			
Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly
2375						2380					2385			
Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala
2390						2395					2400			
Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly
2405						2410					2415			
Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln
2420						2425					2430			
Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala
2435						2440					2445			
Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly
2450						2455					2460			
Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly
2465						2470					2475			

Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala
2480						2485					2490			
Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly
2495						2500					2505			
Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala
2510						2515					2520			
Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala
2525						2530					2535			
Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala
2540						2545					2550			
Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala
2555						2560					2565			
Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr
2570						2575					2580			
Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly
2585						2590					2595			
Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala
2600						2605					2610			
Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly
2615						2620					2625			
Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala
2630						2635					2640			
Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala
2645						2650					2655			
Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly
2660						2665					2670			
Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly
2675						2680					2685			

26612

Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr
2690						2695					2700			
Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly
2705						2710					2715			
Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly
2720						2725					2730			
Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly
2735						2740					2745			
Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala
2750						2755					2760			
Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly
2765						2770					2775			
Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala
2780						2785					2790			
Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly
2795						2800					2805			
Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Tyr	Gly
2810						2815					2820			
Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala
2825						2830					2835			
Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly	Tyr	Gly	Arg	Gly
2840						2845					2850			
Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Ala
2855						2860					2865			
Gly	Gly	Tyr	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Ala	Gly	Ala	Gly	Ala
2870						2875					2880			
Gly	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Gly	Ala	Gly	Gly	Ala	Gly	Gly
2885						2890					2895			
Tyr	Gly	Arg	Gly	Ala	Gly	Ala	Gly	Ala	Gly	Ala	Ala	Ala	Gly	Ala

26612

2900	2905	2910
Gly Ala 2915	Gly Ala Gly Gly Tyr 2920	Gly Gly Gln Gly Gly Tyr Gly Ala 2925
Gly Ala 2930	Gly Ala Gly Ala Ala 2935	Ala Ala Ala Gly Ala Gly Ala Gly 2940
Gly Ala 2945	Gly Gly Tyr Gly Arg 2950	Gly Ala Gly Ala Gly Ala Gly Ala 2955
Ala Ala 2960	Gly Ala Gly Ala Gly 2965	Ala Gly Gly Tyr Gly Gly Gln Gly 2970
Gly Tyr 2975	Gly Ala Gly Ala Gly 2980	Ala Gly Ala Ala Ala Ala Ala Ala Gly 2985
Ala Gly 2990	Ala Gly Gly Ala Gly 2995	Gly Tyr Gly Arg Gly Ala Gly Ala 3000
Gly Ala 3005	Gly Ala Ala Ala Gly 3010	Ala Gly Ala Gly Ala Gly Gly Tyr 3015
Gly Gly 3020	Gln Gly Gly Tyr Gly 3025	Ala Gly Ala Gly Ala Gly Ala Gly Ala Ala 3030
Ala Ala 3035	Ala Gly Ala Gly Ala 3040	Gly Gly Ala Gly Gly Tyr Gly Arg 3045
Gly Ala 3050	Gly Ala Gly Ala Gly 3055	Ala Ala Ala Gly Ala Gly Ala Gly 3060
Ala Gly 3065	Gly Tyr Gly Gly Gln 3070	Gly Gly Tyr Gly Ala Gly Ala Gly 3075
Ala Gly 3080	Ala Ala Ala Ala Ala 3085	Gly Ala Gly Ala Gly Gly Ala Gly 3090
Gly Tyr 3095	Gly Arg Gly Ala Gly 3100	Ala Gly Ala Gly Ala Ala Ala Gly 3105
Ala Gly 3110	Ala Gly Ala Gly Gly 3115	Tyr Gly Gly Gln Gly Gly Tyr Gly 3120

Ala Gly 3125	Ala Gly	Ala Gly	Ala Gly	Ala 3130	Ala Ala Ala Ala	Gly 3135	Ala Gly Ala
Gly Gly 3140	Ala Gly	Gly Tyr	Gly 3145	Arg Gly	Ala Gly	Ala 3150	Gly Ala Gly
Ala Ala 3155	Ala Gly	Ala Gly	Ala 3160	Gly Ala	Gly Gly	Tyr 3165	Gly Gly Gln
Gly Gly 3170	Tyr Gly	Ala Gly	Ala 3175	Gly Ala	Gly Ala	Ala 3180	Ala Ala Ala
Gly Ala 3185	Gly Ala	Gly Gly	Ala 3190	Gly Gly	Tyr Gly	Arg 3195	Gly Ala Gly
Ala Gly 3200	Ala Gly	Ala Ala	Ala 3205	Gly Ala	Gly Ala	Gly 3210	Ala Gly Gly
Tyr Gly 3215	Gly Gln	Gly Gly	Tyr 3220	Gly Ala	Gly Ala	Gly 3225	Ala Gly Ala
Ala Ala 3230	Ala Ala	Gly Ala	Gly 3235	Ala Gly	Gly Ala	Gly 3240	Gly Tyr Gly
Arg Gly 3245	Ala Gly	Ala Gly	Ala 3250	Gly Ala	Ala Ala	Gly 3255	Ala Gly Ala
Gly Ala 3260	Gly Gly	Tyr Gly	Gly 3265	Gln Gly	Gly Tyr	Gly 3270	Ala Gly Ala
Gly Ala 3275	Gly Ala	Ala Ala	Ala 3280	Ala Gly	Ala Gly	Ala 3285	Gly Gly Ala
Gly Gly 3290	Tyr Gly	Arg Gly	Ala 3295	Gly Ala	Gly Ala	Gly 3300	Ala Ala Ala
Gly Ala 3305	Gly Ala	Gly Ala	Gly 3310	Gly Tyr	Gly Gly	Gln 3315	Gly Gly Tyr
Gly Ala 3320	Gly Ala	Gly Ala	Gly 3325	Ala Ala	Ala Ala	Ala 3330	Gly Ala Gly

Ala Gly	Gly Ala	Gly Gly	Tyr	Gly Arg	Gly Ala	Gly	Ala Gly	Ala
3335			3340				3345	
Gly Ala	Ala Ala	Gly Ala	Gly	Ala Gly	Ala Gly	Gly	Tyr Gly	Gly
3350			3355				3360	
Gln Gly	Gly Tyr	Gly Ala	Gly	Ala Gly	Ala Gly	Ala	Ala Ala	Ala
3365			3370				3375	
Ala Gly	Ala Gly	Ala Gly	Gly	Ala Gly	Gly Tyr	Gly	Arg Gly	Ala
3380			3385				3390	
Gly Ala	Gly Ala	Gly Ala	Ala	Ala Gly	Ala Gly	Ala	Gly Ala	Gly
3395			3400				3405	
Gly Tyr	Gly Gly	Gln Gly	Gly	Tyr Gly	Ala Gly	Ala	Gly Ala	Gly
3410			3415				3420	
Ala Ala	Ala Ala	Ala Gly	Ala	Gly Ala	Gly Gly	Ala	Gly Gly	Tyr
3425			3430				3435	
Gly Arg	Gly Ala	Gly Ala	Gly	Ala Gly	Ala Ala	Ala	Gly Ala	Gly
3440			3445				3450	
Ala Gly	Ala Gly	Gly Tyr	Gly	Gly Gln	Gly Gly	Tyr	Gly Ala	Gly
3455			3460				3465	
Ala Gly	Ala Gly	Ala Ala	Ala	Ala Ala	Gly Ala	Gly	Ala Gly	Gly
3470			3475				3480	
Ala Gly	Gly Tyr	Gly Arg	Gly	Ala Gly	Ala Gly	Ala	Gly Ala	Ala
3485			3490				3495	
Ala Gly	Ala Gly	Ala Gly	Ala	Gly Gly	Tyr Gly	Gly	Gln Gly	Gly
3500			3505				3510	
Tyr Gly	Ala Gly	Ala Gly	Ala	Gly Ala	Ala Ala	Ala	Ala Gly	Ala
3515			3520				3525	
Gly Ala	Gly Gly	Ala Gly	Gly	Tyr Gly	Arg Gly	Ala	Gly Ala	Gly
3530			3535				3540	

26612

Ala Gly 3545	Ala Ala Ala Gly	Ala Gly Ala Gly Ala Gly 3550	Gly Tyr Gly
Gly Gln 3560	Gly Gly Tyr Gly	Ala Gly Ala Gly Ala Gly 3565	Ala Ala Ala 3570
Ala Ala 3575	Gly Ala Gly Ala	Gly Gly Ala Gly Gly Tyr 3580	Gly Arg Gly 3585
Ala Gly 3590	Ala Gly Ala Gly	Ala Ala Gly Ala Gly 3595	Ala Gly Ala 3600
Gly Gly 3605	Tyr Gly Gly Gln	Gly Gly Tyr Gly Ala Gly 3610	Ala Gly Ala 3615
Gly Ala 3620	Ala Ala Ala Ala	Gly Ala Gly Ala Gly Gly 3625	Ala Gly Gly 3630
Tyr Gly 3635	Arg Gly Ala Gly	Ala Gly Ala Gly Ala Ala 3640	Ala Gly Ala 3645
Gly Ala 3650	Gly Ala Gly Gly	Tyr Gly Gly Gln Gly Gly 3655	Tyr Gly Ala 3660
Gly Ala 3665	Gly Ala Gly Ala	Ala Ala Ala Gly Ala 3670	Gly Ala Gly 3675
Gly Ala 3680	Gly Gly Tyr Gly	Arg Gly Ala Gly Ala Gly 3685	Ala Gly Ala 3690
Ala Ala 3695	Gly Ala Gly Ala	Gly Ala Gly Gly Tyr Gly 3700	Gly Gln Gly 3705
Gly Tyr 3710	Gly Ala Gly Ala	Gly Ala Gly Ala Ala Ala 3715	Ala Ala Gly 3720
Ala Gly 3725	Ala Gly Gly Ala	Gly Gly Tyr Gly Arg Gly 3730	Ala Gly Ala 3735
Gly Ala 3740	Gly Ala Ala Ala	Gly Ala Gly Ala Gly Ala 3745	Gly Gly Tyr 3750
Gly Gly	Gln Gly Gly Tyr	Gly Ala Gly Ala Gly Ala	Gly Ala Ala

3755		3760		3765
Ala Ala Ala Gly Ala Gly	Ala Gly Gly Ala Gly	Gly Tyr Gly Arg		
3770	3775	3780		
Gly Ala Gly Ala Gly Ala	Gly Ala Ala Ala Gly	Ala Gly Ala Gly		
3785	3790	3795		
Ala Gly Gly Tyr Gly Gly	Gln Gly Gly Tyr Gly	Ala Gly Ala Gly		
3800	3805	3810		
Ala Gly Ala Ala Ala Ala	Ala Gly Ala Gly Ala	Gly Gly Ala Gly		
3815	3820	3825		
Gly Tyr Gly Arg Gly Ala	Gly Ala Gly Ala Gly	Ala Ala Ala Gly		
3830	3835	3840		
Ala Gly Ala Gly Ala Gly	Gly Tyr Gly Gly Gln	Gly Gly Tyr Gly		
3845	3850	3855		
Ala Gly Ala Gly Ala Gly	Ala Ala Ala Ala Ala	Gly Ala Gly Ala		
3860	3865	3870		
Gly Gly Ala Gly Gly Tyr	Gly Arg Gly Ala Gly	Ala Gly Ala Gly		
3875	3880	3885		
Ala Ala Ala Gly Ala Gly	Ala Gly Ala Gly Gly	Tyr Gly Gly Gln		
3890	3895	3900		
Gly Gly Tyr Gly Ala Gly	Ala Gly Ala Gly Ala	Ala Ala Ala Ala		
3905	3910	3915		
Gly Ala Gly Ala Gly Gly	Ala Gly Gly Tyr Gly	Arg Gly Ala Gly		
3920	3925	3930		
Ala Gly Ala Gly Ala Ala	Ala Gly Ala Gly Ala	Gly Ala Gly Gly		
3935	3940	3945		
Tyr Gly Gly Gln Gly Gly	Tyr Gly Ala Gly Ala	Gly Ala Gly Ala		
3950	3955	3960		
Ala Ala Ala Ala Gly Ala	Gly Ala Gly Gly Ala	Gly Gly Tyr Gly		
3965	3970	3975		

Arg Gly 3980	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Gly Ala	Ala Ala	Ala Ala	Gly Ala	Ala Gly	Ala Gly	Ala Gly
Gly Ala 3995	Gly Gly	Tyr Gly	Gly Gly	Gln Gly	Gly Gly	Tyr Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly
Gly Ala 4010	Gly Ala	Ala Ala	Ala Ala	Ala Ala	Ala Ala	Gly Ala	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Gly Gly	Ala Gly
Gly Gly 4025	Tyr Gly	Arg Gly	Gly Ala	Gly Ala	Gly Ala	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly
Gly Ala 4040	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Gly Tyr	Gly Gly	Gln Gly	Gly Gly	Tyr Gly	Gly Gly	Tyr Gly
Gly Ala 4055	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Ala Ala	Ala Ala	Ala Ala	Ala Ala	Ala Ala	Gly Ala	Gly Ala
Ala Gly 4070	Gly Ala	Gly Gly	Tyr Gly	Gly Arg	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Ala Gly	Ala Gly	Ala Gly	Ala Gly
Gly Ala 4085	Ala Ala	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Ala Gly	Ala Gly	Gly Gly	Gly Gly	Tyr Gly	Gly Gly	Gly Gly
Gln Gly 4100	Gly Tyr	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Ala Gly	Ala Gly	Gly Ala	Gly Ala	Ala Ala	Ala Ala	Ala Ala
Ala Gly 4115	Ala Gly	Ala Gly	Gly Gly	Gly Gly	Gly Gly	Ala Gly	Gly Gly	Tyr Gly	Gly Gly	Arg Gly	Gly Ala	Ala Gly
Gly Ala 4130	Gly Ala	Gly Ala	Ala Ala	Ala Ala	Ala Ala	Ala Gly	Ala Gly	Gly Ala	Gly Ala	Ala Gly	Gly Ala	Gly Ala
Gly Tyr 4145	Gly Gly	Gln Gly	Gly Gly	Gly Gly	Tyr Gly	Ala Gly	Ala Gly	Gly Ala	Gly Ala	Ala Gly	Gly Ala	Gly Ala
Ala Ala 4160	Ala Ala	Ala Gly	Ala Gly	Ala Gly	Gly Ala	Gly Ala	Gly Gly	Ala Gly	Gly Gly	Gly Gly	Tyr Gly	Ala Gly
Gly Arg 4175	Gly Ala	Gly Ala	Gly Ala	Gly Ala	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Ala Gly	Gly Ala	Gly Ala	Gly Ala

Ala Gly 4190	Ala Gly Gly Tyr 4195	Gly Gln Gly Gly Tyr 4200	Gly Ala Gly
Ala Gly 4205	Ala Gly Ala Ala 4210	Ala Ala Gly Ala Gly 4215	Ala Gly Gly
Ala Gly 4220	Gly Tyr Gly Arg 4225	Ala Gly Ala Gly Ala 4230	Gly Ala Ala
Ala Gly 4235	Ala Gly Ala Gly 4240	Gly Gly Tyr Gly Gly 4245	Gln Gly Gly
Tyr Gly 4250	Ala Gly Ala Gly 4255	Gly Ala Ala Ala Ala 4260	Ala Gly Ala
Gly Ala 4265	Gly Gly Ala Gly 4270	Tyr Gly Arg Gly Ala 4275	Gly Ala Gly
Ala Gly 4280	Ala Ala Ala Gly 4285	Gly Ala Gly Ala Gly 4290	Gly Tyr Gly
Gly Gln 4295	Gly Gly Tyr Gly 4300	Gly Ala Gly Ala Gly 4305	Ala Ala Ala
Ala Ala 4310	Gly Ala Gly Ala 4315	Gly Ala Gly Gly Tyr 4320	Gly Arg Gly
Ala Gly 4325	Ala Gly Ala Gly 4330	Ala Ala Gly Ala Gly 4335	Ala Gly Ala
Gly Gly 4340	Tyr Gly Gly Gln 4345	Gly Tyr Gly Ala Gly 4350	Ala Gly Ala
Gly Ala 4355	Ala Ala Ala Ala 4360	Ala Gly Ala Gly Gly 4365	Ala Gly Gly
Tyr Gly 4370	Arg Gly Ala Gly 4375	Gly Ala Gly Ala Ala 4380	Ala Gly Ala
Gly Ala 4385	Gly Ala Gly Gly 4390	Gly Gly Gln Gly Gly 4395	Tyr Gly Ala

Gly Ala 4400	Gly Ala Gly Ala 4405	Ala Ala Ala Gly 4410	Gly Ala Gly
Gly Ala 4415	Gly Gly Tyr Gly 4420	Arg Gly Ala Gly Ala 4425	Gly Ala Gly Ala
Ala Ala 4430	Gly Ala Gly Ala 4435	Gly Ala Gly Tyr Gly 4440	Gly Gln Gly
Gly Tyr 4445	Gly Ala Gly Ala 4450	Ala Gly Ala Ala 4455	Ala Ala Gly
Ala Gly 4460	Ala Gly Gly Ala 4465	Gly Tyr Gly Arg Gly 4470	Ala Gly Ala
Gly Ala 4475	Gly Ala Ala Ala 4480	Ala Gly Ala Gly 4485	Gly Gly Tyr
Gly Gly 4490	Gln Gly Gly Tyr 4495	Gly Ala Gly Ala Gly 4500	Gly Ala Ala
Ala Ala 4505	Ala Gly Ala Gly 4510	Gly Gly Ala Gly 4515	Tyr Gly Arg
Gly Ala 4520	Gly Ala Gly Ala 4525	Ala Ala Ala Gly 4530	Gly Ala Gly
Ala Gly 4535	Gly Tyr Gly Gly 4540	Gln Gly Gly Tyr Gly 4545	Ala Gly Ala Gly
Ala Gly 4550	Ala Ala Ala Ala 4555	Ala Gly Ala Gly Ala 4560	Gly Ala Gly
Gly Tyr 4565	Gly Arg Gly Ala 4570	Ala Gly Ala Gly 4575	Ala Ala Gly
Ala Gly 4580	Ala Gly Ala Gly 4585	Gly Tyr Gly Gly Gln 4590	Gly Tyr Gly
Ala Gly 4595	Ala Gly Ala Gly 4600	Ala Ala Ala Ala 4605	Ala Gly Ala
Gly Gly	Ala Gly Gly Tyr 4610	Gly Arg Gly Ala Gly 4615	Ala Gly Gly

26612

4610		4615		4620
Ala Ala Ala Gly Ala Gly	Ala Gly Ala Gly Gly Tyr Gly Gly Gln			
4625	4630	4635		
Gly Gly Tyr Gly Ala Gly	Ala Gly Ala Gly Ala Ala Ala Ala Ala			
4640	4645	4650		
Gly Ala Gly Ala Gly Gly	Ala Gly Gly Tyr Gly Arg Gly Ala Gly			
4655	4660	4665		
Ala Gly Ala Gly Ala Ala	Ala Gly Ala Gly Ala Gly Ala Gly Gly			
4670	4675	4680		
Tyr Gly Gly Gln Gly Gly	Tyr Gly Ala Gly Ala Gly Ala Gly Ala			
4685	4690	4695		
Ala Ala Ala Ala Gly Ala	Gly Ala Gly Gly Ala Gly Gly Tyr Gly			
4700	4705	4710		
Arg Gly Ala Gly Ala Gly	Ala Gly Ala Ala Ala Gly Ala Gly Ala			
4715	4720	4725		
Gly Ala Gly Gly Tyr Gly	Gly Gln Gly Gly Tyr Gly Ala Gly Ala			
4730	4735	4740		
Gly Ala Gly Ala Ala Ala	Ala Ala Gly Ala Gly Ala Gly Gly Ala			
4745	4750	4755		
Gly Gly Tyr Gly Arg Gly	Ala Gly Ala Gly Ala Gly Ala Ala Ala			
4760	4765	4770		
Gly Ala Gly Ala Gly Ala	Gly Tyr Gly Gly Gln Gly Gly Tyr			
4775	4780	4785		
Gly Ala Gly Ala Gly Ala	Gly Ala Ala Ala Ala Ala Gly Ala Gly			
4790	4795	4800		
Ala Gly Gly Ala Gly Gly	Tyr Gly Arg Gly Ala Gly Ala Gly Ala			
4805	4810	4815		
Gly Ala Ala Ala Gly Ala	Gly Ala Gly Ala Gly Gly Tyr Gly Gly			
4820	4825	4830		

Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala
4835 4840 4845

Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Arg Gly Ala
4850 4855 4860

Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly Ala Gly Ala Gly
4865 4870 4875

Gly Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly Ala Gly Ala Gly
4880 4885 4890

Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr
4895 4900 4905

Gly Arg Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Gly Ala Gly
4910 4915 4920

Ala Gly Ala Gly Gly Tyr Gly Gly Gln Gly Gly Tyr Gly Ala Gly
4925 4930 4935

Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala
4940 4945

<210> 25
<211> 93
<212> PRT
<213> Argiope trifasciata

<400> 25
Gly Ala Gly Ser Gly Ala Gly Ala Gly Ser Gly Ala Gly Ala Gly Ser
1 5 10 15

Gly Ala Gly Ala Gly Ser Gly Ser Gly Ala Gly Tyr Gly Val Gly Ala
20 25 30

Gly Ser Gly Ala Gly Ala Gly Ser Gly Ala Gly Ala Gly Tyr Gly Ala
35 40 45

Gly Ala Gly Ser Gly Ala Gly Ala Gly Ser Gly Ala Gly Ala Gly Ser
50 55 60

Gly Ala Gly Ser Asp Gly Tyr Gly Arg Gly Phe Gly Ala Gly Ala Gly

26612

65

70

75

80

Ser Gly Ala Gly Ala Gly Ser Gly Ala Gly Tyr Gly Ala
 85 90

<210> 26

<211> 200

<212> PRT

<213> Araneus sp.

<400> 26

Gly Ala Gly Ala Ala Gly Gly Tyr Gly Gly Gly Ala Gly Ala Gly Ala
 1 5 10 15

Gly Gly Ala Gly Gly Tyr Gly Gln Gly Tyr Gly Ala Gly Ala Gly Ala
 20 25 30

Gly Ala Ala Ala Ala Ala Gly Ala Gly Ala Gly Ala Ala Gly Gly Tyr
 35 40 45

Gly Gly Gly Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Gln
 50 55 60

Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Ala Gly Ala
 65 70 75 80

Gly Ala Gly Ala Ala Gly Gly Tyr Gly Gly Gly Ala Gly Ala Gly Ala
 85 90 95

Gly Gly Ala Gly Gly Tyr Gly Gln Gly Tyr Gly Ala Gly Ala Gly Ala
 100 105 110

Gly Ala Ala Ala Ala Gly Ala Gly Ala Gly Ala Ala Gly Gly Tyr
 115 120 125

Gly Gly Gly Ala Gly Ala Gly Ala Gly Gly Ala Gly Gly Tyr Gly Gln
 130 135 140

Gly Tyr Gly Ala Gly Ala Gly Ala Gly Ala Ala Ala Ala Gly Ala
 145 150 155 160

Gly Ala Gly Ala Ala Gly Gly Tyr Gly Gly Gly Ala Gly Ala Gly Ala
 165 170 175

Gly Gly Ala Gly Gly Tyr Gly Gln Gly Tyr Gly Ala Gly Ala Gly Ala
 180 185 190

Gly Ala Ala Ala Ala Ala Gly Ala
 195 200

<210> 27
 <211> 387
 <212> PRT
 <213> Nephila clavipes

<220>
 <221> MOD_RES
 <222> (5)..(5)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (10)..(10)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (15)..(15)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (20)..(20)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (25)..(25)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (30)..(30)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (35)..(35)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (40)..(40)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (45)..(45)

<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (50)..(50)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (55)..(55)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (60)..(60)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (65)..(65)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (70)..(70)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (75)..(75)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (80)..(80)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (85)..(85)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (95)..(95)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (100)..(100)
<223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (10 $\bar{5}$)..(105)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (11 $\bar{0}$)..(110)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (11 $\bar{5}$)..(115)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (12 $\bar{0}$)..(120)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (12 $\bar{5}$)..(125)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (13 $\bar{0}$)..(130)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (13 $\bar{5}$)..(135)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (14 $\bar{0}$)..(140)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (14 $\bar{5}$)..(145)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (15 $\bar{0}$)..(150)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (15 $\bar{5}$)..(155)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES

<222> (160)..(160)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (165)..(165)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (170)..(170)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (175)..(175)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (180)..(180)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (185)..(185)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (190)..(190)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (195)..(195)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (200)..(200)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (205)..(205)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (208)..(208)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (211)..(211)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (214)..(214)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (217)..(217)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (220)..(220)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (223)..(223)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (226)..(226)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (262)..(262)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (267)..(267)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (272)..(272)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (277)..(277)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (282)..(282)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (287)..(287)
 <223> Axit amin bất kỳ

<220>

<221> MOD_RES
 <222> $(29\overline{2})..(292)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(29\overline{7})..(297)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(30\overline{2})..(302)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(30\overline{7})..(307)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(31\overline{2})..(312)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(31\overline{7})..(317)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(32\overline{2})..(322)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(32\overline{7})..(327)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(33\overline{2})..(332)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(33\overline{7})..(337)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(34\overline{2})..(342)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(34\overline{7})..(347)$

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (352)..(352)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (357)..(357)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (362)..(362)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (367)..(367)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (372)..(372)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (377)..(377)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (382)..(382)

<223> Axit amin bất kỳ

<220>

<221> MOD_RES

<222> (387)..(387)

<223> Axit amin bất kỳ

<400> 27

Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly
1				5					10					15	

Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro
			20					25					30		

Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly
		35					40					45			

Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly	Xaa	Gly	Pro	Gly	Gly
	50					55					60				

Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
65 70 75 80

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
85 90 95

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
100 105 110

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly
115 120 125

Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly
130 135 140

Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
145 150 155 160

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
165 170 175

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
180 185 190

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Gly Xaa
195 200 205

Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly
210 215 220

Gly Xaa Thr Ile Ile Glu Asp Leu Asp Ile Thr Ile Asp Gly Ala Asp
225 230 235 240

Gly Pro Ile Thr Ile Ser Glu Glu Leu Thr Ile Ser Gly Ala Gly Gly
245 250 255

Ser Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
260 265 270

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
275 280 285

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
 290 295 300

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly
 305 310 315 320

Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly
 325 330 335

Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
 340 345 350

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
 355 360 365

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
 370 375 380

Gly Gly Xaa
 385

<210> 28
 <211> 329
 <212> PRT
 <213> Nephila sp.

<220>
 <221> MOD_RES
 <222> (5)..(5)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (10)..(10)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (15)..(15)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (20)..(20)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES

<222> (25)..(25)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (30)..(30)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (35)..(35)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (40)..(40)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (45)..(45)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (50)..(50)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (55)..(55)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (60)..(60)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (65)..(65)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (70)..(70)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (75)..(75)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (80)..(80)
 <223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (85)..(85)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (95)..(95)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (100)..(100)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (105)..(105)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (110)..(110)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (115)..(115)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (120)..(120)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (125)..(125)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (130)..(130)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (135)..(135)
<223> Axit amin bất kỳ

<220>

<221> MOD_RES
 <222> $(14\overline{0})..(140)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(14\overline{5})..(145)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(15\overline{0})..(150)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(15\overline{5})..(155)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(16\overline{0})..(160)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(16\overline{5})..(165)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(17\overline{0})..(170)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(17\overline{5})..(175)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(18\overline{0})..(180)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(18\overline{3})..(183)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(18\overline{6})..(186)$
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> $(18\overline{9})..(189)$

```

<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (192)..(192)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (195)..(195)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (198)..(198)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (201)..(201)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (239)..(239)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (244)..(244)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (249)..(249)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (254)..(254)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (259)..(259)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (264)..(264)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (269)..(269)
<223> Axit amin bất kỳ

```

<220>
<221> MOD_RES
<222> (274)..(274)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (279)..(279)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (284)..(284)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (289)..(289)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (294)..(294)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (299)..(299)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (304)..(304)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (309)..(309)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (314)..(314)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (319)..(319)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (324)..(324)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES

<222> (329)..(329)

<223> Axit amin bất kỳ

<400> 28

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
 1 5 10 15

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
 20 25 30

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly
 35 40 45

Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly
 50 55 60

Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
 65 70 75 80

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
 85 90 95

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
 100 105 110

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly
 115 120 125

Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly
 130 135 140

Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
 145 150 155 160

Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
 165 170 175

Pro Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa
 180 185 190

Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Thr Val Ile Glu Asp Leu Asp
 195 200 205

Ile Thr Ile Asp Gly Ala Asp Gly Pro Ile Thr Ile Ser Glu Glu Leu

26612

210	215	220
Thr Ile Gly Gly Ala Gly Ala Gly Gly Ser Gly Pro Gly Gly Xaa Gly		
225	230	235 240
Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro		
	245	250 255
Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly		
	260	265 270
Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly		
	275	280 285
Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa		
	290	295 300
Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly		
305	310	315 320
Pro Gly Gly Xaa Gly Pro Gly Gly Xaa		
	325	

<210> 29
 <211> 125
 <212> PRT
 <213> Argiope trifasciata

<220>
 <221> MOD_RES
 <222> (5)..(5)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (10)..(10)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (15)..(15)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (20)..(20)
 <223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (25)..(25)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (30)..(30)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (52)..(52)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (57)..(57)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (62)..(62)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (67)..(67)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (72)..(72)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (75)..(75)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (78)..(78)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (81)..(81)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES
<222> (84)..(84)
<223> Axit amin bất kỳ

<220>
<221> MOD_RES

<222> (87)..(87)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (90)..(90)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (95)..(95)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (100)..(100)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (105)..(105)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (110)..(110)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (115)..(115)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (120)..(120)
 <223> Axit amin bất kỳ

<220>
 <221> MOD_RES
 <222> (125)..(125)
 <223> Axit amin bất kỳ

<400> 29
 Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
 1 5 10 15

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
 20 25 30

Val Thr Val Asp Val Asp Val Ser Val Gly Gly Ala Pro Gly Gly Gly
 35 40 45

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro

50

55

60

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly
65 70 75 80

Xaa Gly Gly Xaa Gly Gly Xaa Gly Gly Xaa Gly Pro Gly Gly Xaa Gly
85 90 95

Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro
100 105 110

Gly Gly Xaa Gly Pro Gly Gly Xaa Gly Pro Gly Gly Xaa
115 120 125

<210> 30

<211> 17388

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 30

tgcagctccc atggccattc gaattcggcc ggcctaggcg cgccgtacgc gtatcgataa	60
gctttaagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc	120
tttatattgtg aaatttgtga tgctattgct ttattttgtaa ccattataag ctgcaataaa	180
caagttaaca acaacaattg cattcathtt atgtttcagg ttcaggggga ggtgtgggag	240
gttttttaaa gcaagtaaaa cctctacaaa tgtggtatgg ctgattatga tctagagtcg	300
cggccgctac aggaacaggt ggtggcggcc ctcggtgcgc tcgtactgct ccacgatggt	360
gtagtcctcg ttgtgggagg tgatgtccag cttggagtcc acgtagtagt agccgggcag	420
ctgcacgggc ttcttggcca ttagatgga cttgaactcc accaggtagt ggccgccgtc	480
cttcagcttc agggccttgt ggatctcgcc cttcagcacg ccgtcgcggg ggtacaggcg	540
ctcggtggag gcctcccagc ccatggtctt cttctgcatt acggggccgt cggaggggaa	600
gttccgccga tgaacttcac cttgtagatg aagcagccgt cctgcaggga ggagtcctgg	660
gtcacggtca ccacgccgc gtcctcgaag ttcacacgc gctccactt gaagccctcg	720
gggaaggaca gcttcttgta gtcggggatg tcggcggggt gcttcacgta caccttgag	780
ccgtactgga actgggggga caggatgtcc caggcgaagg gcagggggcc gcccttggtc	840

accttcagct tcacgggtgtt gtggccctcg taggggcggc cctcgcccct cgcccctcga	900
tctcgaactc gtggccgttc acggtgccct ccatgcgcac cttgaagcgc atgaactcct	960
tgatgacgtt cttggaggag cgcaccatgg tggcgaccgg tggatcccgg gcccgcggtg	1020
ccgtcgactc tagcgggtacc ccgattgttt agcttggttca gctgcgcttg tttatttgct	1080
tagcttttcgc ttagcgacgt gttcactttg cttgtttgaa ttgaattgtc gctccgtaga	1140
cgaagcgcct ctattttatac tccggcggtc gagggttcga aatcgataag cttggatcct	1200
aattgaatta gctctaattg aattagtctt ctaattgaat tagtctctaa ttgaattaga	1260
tccccgggcg agctcgaatt aaaccattgt gggaaccgtg cgatcaaaca aacgcgagat	1320
accgggaagt actgaaaaac agtcgctcca ggccagtggg aacatcgatg ttttgttttg	1380
acggaccctt tactctcgtc tcatataaac cgaagccagc taagatggta tacttattat	1440
catcttgatga tgaggatgct tctatcaacg aaagtaccgg taaaccgcaa atggttatgt	1500
attataatca aactaaaggc ggagtggaca cgctagacca aatgtgttct gtgatgacct	1560
gcagtaggaa gacgaatagg tggcctatgg cattattgta cggaatgata aacattgcct	1620
gcataaattc ttttattata tacagccata atgtcagtag caagggagaa aagggtcaaa	1680
gtcgcaaaaa atttatgaga aacctttaca tgagcctgac gtcacgtttt atgcgtaagc	1740
gtttagaagc tcctactttg aagagatatt tgcgcgataa tatctctaata attttgccaa	1800
atgaagtgcc tgggtacatca gatgacagta ctgaagagcc agtaatgaaa aaacgtactt	1860
actgtactta ctgcccctct aaaataaggc gaaaggcaaa tgcacgtgac aaaaaatgca	1920
aaaaagttat ttgtcgagag cataatattg atatgtgcca aagttgtttc tgactgacta	1980
ataagtataa tttgtttcta ttatgtataa gttaagctaa ttacttattt tataatacaa	2040
catgactgtt tttaaagtac aaaataagtt tattttttgta aaagagagaa tgtttaaaag	2100
ttttgttact ttatagaaga aattttgagt ttttgttttt ttttaataaa taaataaaca	2160
taaataaatt gtttgttgaa tttattatta gtatgtaagt gtaaatataa taaaacttaa	2220
tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtcaat	2280
tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggtaaataa	2340
tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taccgtatat ctcaacgctg	2400
tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcatcgactt gatattgtcc	2460
gacacatttt cgtcgatttg cgttttgatc aaagacttga gcagagacac gttaatcaac	2520
tgttcaaatt gatccatatt aacgatatca acccgatgcg tatatggtgc gtaaaatata	2580

ttttttaacc	ctcttatact	ttgcactctg	cggttaatacg	cgttcgtgta	cagacgtaat	2640
catgttttct	tttttggata	aaactcctac	tgagtttgac	ctcatattag	accctcacia	2700
gttgcaaaac	gtggcatttt	ttaccaatga	agaatttaaa	gttatttttaa	aaaatttcat	2760
cacagattta	aagaagaacc	aaaaattaaa	ttattttcaac	agtttaaatcg	accagttaat	2820
caacgtgtac	acagacgcgt	cggcaaaaaa	cacgcagccc	gacgtgttgg	ctaaaattat	2880
taaatcaact	tgtgttatag	tcacggattt	gccgtccaac	gtgttcctca	aaaagttgaa	2940
gaccaacaag	tttacggaca	ctattaatta	tttgattttg	ccccacttca	ttttgtggga	3000
tcacaatttt	gttatatttt	taaaacaaag	ctttggcact	ggccgtcggt	ttacaacgtc	3060
gtgactggga	aaaccctggc	gttaccacaac	ttaatcgctt	tgcagcacat	ccccctttcg	3120
ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	3180
tgaatggcga	atggcgcttg	atgcggtatt	ttctccttac	gcattctgtgc	ggtatttcac	3240
accgcatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	agccagcccc	3300
gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgctcccg	gcattccgctt	3360
acagacaagc	tgtgaccgtc	tccgggagct	gcattgtgtca	gaggttttca	ccgtcatcac	3420
cgaaacgcgc	gagacgaaag	ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	3480
taataatggt	ttcttagacg	tcagggtggca	cttttcgggg	aaatgtgcgc	ggaaccccta	3540
tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	3600
aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgcgcc	3660
ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctggtga	3720
aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	3780
acagcggtaa	gattccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	3840
ttaaagttct	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	gagcaactcg	3900
gtcgccgcat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	3960
atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	4020
acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	4080
tgcacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	4140
ccataccaaa	cgacgagcgt	gacaccacga	tgctgttagc	aatggcaaca	acgttgcgca	4200
aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	4260

aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggttttattg	4320
ctgataaatc tggagccggt gagcgtgggt ctgcggtat cattgcagca ctggggccag	4380
atggtaagcc ctcccgatc gtagttatct acacgacggg gagtcaggca actatggatg	4440
aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag	4500
accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga	4560
tctaggtgaa gatccttttt gataatctca tgacaaaaat cccttaacgt gagttttcgt	4620
tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat cctttttttc	4680
tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc	4740
cggatcaaga gctaccaact ctttttccga aggtaactgg cttcagcaga gcgcagatac	4800
caaatactgt tcttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac	4860
cgctacata cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt	4920
cgtgtcttac cgggttgac tcaagacgat agttaccgga taaggcgcag cggtcgggct	4980
gaacgggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat	5040
acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt	5100
atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca gggggaaacg	5160
cctgatatct ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt	5220
gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg	5280
ttcctggcct tttgctggcc ttttgctcac atgttctttc ctgcgttatc ccctgattct	5340
gtggataacc gtattaccgc ctttgagtga gctgataccg ctgccgcag ccgaacgacc	5400
gagcgcagcg agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc	5460
cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg	5520
ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta	5580
cactttatgc ttccggctcg tatgttggtg ggaattgtga gcggataaca atttcacaca	5640
ggaaacagct atgacatgat tacgaattcg agctcggtag ccggggatcc tctagagtcg	5700
acgctcgcgc gacttggttt gccattcttt agcgcgcgtc gcgtcacaca gcttgccac	5760
aatgtggttt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaagc	5820
gcaaactctt ttttaacccta gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa	5880
atattgctct ctctttctaa atagcgcgaa tccgtcgtcg tgcatttagg acatctcagt	5940
cgccgcttgg agctcccgtg aggcgtgctt gtcaatgcgg taagtgtcac tgattttgaa	6000

ctataacgac	cgcgtgagtc	aaaatgacgc	atgattatct	tttacgtgac	ttttaagatt	6060
taactcatac	gataattata	ttgttatttc	atgttctact	tacgtgataa	cttattatat	6120
atatattttc	ttgttataga	tatcgtgact	aatatataat	aaaatgggta	gttctttaga	6180
cgatgagcat	atcctctctg	ctcttctgca	aagcgatgac	gagcttggtg	gtgaggattc	6240
tgacagtga	atatcagatc	acgtaagtga	agatgacgtc	cagagcgata	cagaagaagc	6300
gtttatagat	gaggtacatg	aagtgcagcc	aacgtcaagc	ggtagtgaaa	tattagacga	6360
acaaaatgtt	attgaacaac	caggttcttc	atagattctg	ttagaagcca	aagaatcttg	6420
accttgccac	agaggactat	tagaggtaag	aataaacatt	gttgggtcaac	ttcaaagtcc	6480
acgaggcgta	gccgagtctc	tgactgaac	attgtcagat	ccgagatcgg	ccggcctagg	6540
cgcgccaagc	ttaaggtgca	cggccacgt	ggccactagt	acttctcgag	gctcaaagcc	6600
tcacccaat	ttggagtcac	tcaagacatc	cttgattaag	gcagctgccg	atattgacat	6660
ggacctcgtt	cgtgctgcga	tagacgactg	gccgcgcaga	ttgaaggcct	gtattcaaaa	6720
tcacggaggt	cattttgaat	aaactttagt	gtcataagaa	tctatgtttt	gttaagttca	6780
ttttggtata	tgaatggtta	cataatgaat	aaacttgttt	caattatttt	acattaaaca	6840
tgtgacagaa	tttatgacct	gactaggtag	gtacaaacag	cctttttgat	attagaaaac	6900
taagtaaaat	agcctacggt	cacatctctt	tccgtgggtg	tcgttaaagg	gcgacttaga	6960
gaaccaccaa	gaacgtagca	gaatcctcag	agtgtcatat	cagcatatag	ccatcgctaa	7020
ctgctattta	ctggtaatag	ggcacattgt	aatctcactt	aaccatactg	tcgggccacc	7080
atctagccta	tttctgccac	gaatcaatcg	tgagtgatgg	acatagagaa	actattagtt	7140
gagaagaaaa	caagagcact	aaaggtttga	tattgacaaa	aatctacttc	gccgtcactc	7200
cataggttta	ttgtctctca	ttagtccaga	acagcagtta	cagacgtaag	cttttacgca	7260
caaactacag	ggttgctctt	tattgtatcg	aaaatatggg	acctgaataa	gggcgatttt	7320
gacgcgtcct	gcccgcccat	tcccgatcct	acggacagaa	tggcaagcag	tcgacgtcgc	7380
cccaaacacg	tcatttcgga	tcctcacgat	ccactaacgg	tgcttttaggt	acctcaagca	7440
ccggtcatcg	ttctcgtcgg	acccgtcgtc	tgcgacgaag	ggctcgacga	gcaaattaac	7500
cctcagacac	agcccactga	gtttctcgcc	ggatcttctc	agcgggtcgc	gtttccgata	7560
cgggtggtaga	ttctgcgaag	cacggctctt	gctaggattc	gtgttagcaa	cgtcgtcagg	7620
tttgagcccc	gtgagctcac	ttactagtta	aggttacgct	gaaatagcct	ctcaaggctc	7680

tcagctaggt aggaaacaaa aaaaaaagtc ctgcccttaa caccggttgcg atggcttgtc	7740
ttctgcaccg cggaaagatg ttttgtacgg aaagtttgaa taagtgttta attgcaagta	7800
acgtaacaat gtttttaggg tccggcggccg cgggagaaaag catgaagtaa gttcttttaa	7860
tattacaaaa aaattgaacg atattataaa attcttttaa atattaaaag taagaacaat	7920
aagatcaatt aaatcataat taatcacatt gttcatgac acaattttaat ttacttcata	7980
cgttgtattg ttatgtttaa taaaaagatt aatttctatg taattgtatc tgtacaatac	8040
aatgtgtaga tgtttattct atcgaaagta aatacgtcaa aactcgaaaa ttttcagtat	8100
aaaaagggtc aactttttca aatcagcatc agttcgggtc caactctcaa gatgagagtc	8160
aaaacctttg tgatcttgtg ctgcgctctg caggtgagtt aattatttta ctattatttc	8220
agaagggtggc cagacgatat cacggggcac ctgataataa gtggtcgcca aaacgcacag	8280
ataticgtaaa ttgtgccatt tgatttgtca cggccggggg ggctacggaa taaactacat	8340
ttatttattt aaaaaatgaa ccttagatta tgtaacttgt gatttatttg cgtcaaaagt	8400
aggcaagatg aatctatgta aatacctggg cagacttgca atatcctatt tcaccggtaa	8460
atcagcattg caatatgcaa tgcatattca acaatatgta aaacaattcg taaagcatca	8520
ttagaaaata gacgaaagaa attgcataaa attataaccg cattattaat ttattatgat	8580
atctattaac aattgctatt gccttttttt cgcaaattat aatcattttc ataacctcga	8640
ggtagcattc tgttacattt taatacattg gtatgtgatt ataacacgag ctgccactg	8700
agtttctcgc cagatcttct cagtgggtcg cgttaccgat cacgtgatag attctatgaa	8760
gcactgctct tgttagggct agtgtttagca aattctttca ggttgagtct gagagctcac	8820
ctacccatcg gagcgtagct ggaataggct accagctaata aggtagggaa aacaagctc	8880
gaaacaagct caagtaataa caacataatg tgaccataaa atctcgtggg gtatgagata	8940
caattatgta ctttcccaca aatgtttaca taattagaat gttgttcaac ttgcctaacg	9000
cccagctag aacattcaat tattactatt accactacta aggcagtatg tcctaactcg	9060
ttccagatca gcgctaactt cgattgaatg tgcgaaattt atagctcaat atttttagcac	9120
ttatcgtatt gatttaagaa aaaattgtta acattttgtt tcagtatgtc gcttatacaa	9180
atgcaaacat caatgatatt gatgaggact attttgggag tgatgtcact gtccaaagta	9240
gtaatacaac agatgaaata attagagatg catctggggc agttatcgaa gaacaaatta	9300
caactaaaaa aatgcaacgg aaaaataaaa accatggaat acttgaaaaa aatgaaaaaa	9360
tgatcaagac gttcgttata accacggatt ccgacggtaa cgagtccatt gtagaggaag	9420

atgtgctcat	gaagacactt	tccgatggta	ctgttgctca	aagttatggt	gctgctgatg	9480
cgggagcata	ttctcagagc	gggccatacg	tatcaaacag	tggatacagc	actcatcaag	9540
gatatacagag	cgatttcagc	actagtgtctg	cagtcgttct	agacctggat	ccccgggtg	9600
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gttccgggtg	9660
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gttccgggtg	9720
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggtccg	ggtggagcag	9780
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	9840
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggcccgt	9900
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	9960
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	10020
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	10080
caggctccgc	tgcagcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	10140
ctggacctgg	tggtgctgga	ccaggaggtg	ctggtccggg	tggagcagga	ccaggaggtg	10200
ctggacctgg	tggtgctgga	ccaggaggtg	ctggtccggg	tggcccgtct	gttccaggct	10260
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	10320
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	10380
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	10440
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	10500
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	10560
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	10620
cggctgctgc	agcaggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	10680
gaccaggagg	tgctggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	10740
gaccaggagg	tgctggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	10800
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	10860
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	10920
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	10980
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11040
ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11100

ctggtccggg	tggcccgctct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	11160
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11220
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11280
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	gtcccggtg	11340
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gtcccggtg	11400
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gtcccggtg	11460
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcagggtccg	ggtggagcag	11520
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggagcag	11580
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggcccgt	11640
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	11700
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	11760
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	11820
caggctccgc	tgacgcggcg	gctgctgcag	caggctccggg	tggagcagga	ccaggaggtg	11880
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggtccggg	tggagcagga	ccaggaggtg	11940
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggtccggg	tggcccgctct	ggtccaggct	12000
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	12060
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	12120
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	12180
cagcggcggc	tgctgcagca	gtcccggtg	gagcaggacc	aggaggtgct	ggacctggtg	12240
gtgctggacc	aggaggtgct	gtcccggtg	gagcaggacc	aggaggtgct	ggacctggtg	12300
gtgctggacc	aggaggtgct	gtcccggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	12360
cggctgctgc	agcagggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	12420
gaccaggagg	tgctgggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	12480
gaccaggagg	tgctgggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	12540
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	12600
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	12660
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgacgcggcg	gctgctgcag	12720
caggctccggg	tggagcagga	ccaggaggtg	ctggacctgg	tgggtgctgga	ccaggaggtg	12780
ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tgggtgctgga	ccaggaggtg	12840

ctggtccggg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	12900
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	12960
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	13020
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	13080
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	13140
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	13200
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggctccg	ggtggagcag	13260
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggctccg	ggtggagcag	13320
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggctccg	ggtggcccgt	13380
ctggtccagg	ctccgctgca	gcggcggtg	ctgcagcagg	tccgggtgga	gcaggaccag	13440
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	13500
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	13560
caggctccgc	tgacgcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	13620
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggctccgg	tggagcagga	ccaggaggtg	13680
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggctccgg	tggcccgtct	ggtccaggct	13740
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	13800
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	13860
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	13920
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	13980
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	14040
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	14100
cggctgctgc	agcaggctccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	14160
gaccaggagg	tgctggctccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	14220
gaccaggagg	tgctggctccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggtg	14280
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	14340
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	14400
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgacgcggcg	gctgctgcag	14460
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tgggtgctgga	ccaggaggtg	14520

ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tgggtgctgga	ccaggaggtg	14580
ctggtccggg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	14640
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	14700
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	14760
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	14820
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	14880
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	14940
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggtccg	ggtggagcag	15000
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	15060
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggcccgt	15120
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	15180
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	15240
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	15300
caggctccgc	tgcagcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	15360
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggtccggg	tggagcagga	ccaggaggtg	15420
ctggacctgg	tgggtgctgga	ccaggaggtg	ctggtccggg	tggcccgtct	ggtccaggct	15480
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	15540
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	15600
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	15660
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	15720
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	15780
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	15840
cggctgctgc	agcaggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	15900
gaccaggagg	tgctggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	15960
gaccaggagg	tgctggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	16020
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	16080
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	16140
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	16200
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tgggtgctgga	ccaggaggtg	16260

ctggtccggg tggagcagga ccaggaggtg ctggacctgg tggtgctgga ccaggaggtg	16320
ctggtccggg tggcccgtct ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc	16380
cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc	16440
cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc	16500
cgggtggccc gtctggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg	16560
gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg	16620
gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg	16680
gcccgtctgg tccaggctcc gctgcagcgg cggctgctgc agcagggtccg ggtggagcag	16740
gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggagcag	16800
gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggcccgt	16860
ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggaagc gtcagttacg	16920
gagctggcag gggatacggg caaggtgcag gaagtgcagc ttcctctgtg tcatctgctt	16980
catctcgcag ttacgactat tctcgtcgta acgtccgcaa aaactgtgga attcctagaa	17040
gacaactagt tgttaaattc agagcactgc cttgtgtgaa ttgctaattt ttaataataa	17100
ataacccttg tttcttactt cgtcctggat acatctatgt tttttttttc gttaataaat	17160
gagagcattt aagttattgt ttttaattac ttttttttag aaaacagatt tcggattttt	17220
tgtatgcatt ttatttgaat gtactaatat aatcaattaa tcaatgaatt catttattta	17280
agggataaca ataatccatg aattcacatg cacatttaaa acaaaactaa attacaatag	17340
gttcatataa aaacaacaag tatgccttct caactaagaa tactatag	17388

<210> 31

<211> 18102

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 31

tcgacgtccc atggccattc gaattcggcc ggcctaggcg cgccgtacgc gtatcgataa	60
gctttaagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc	120
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa	180
caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga ggtgtgggag	240

gttttttaaa gcaagtaaaa cctctacaaa tgtggtatgg ctgattatga tctagagtcg	300
cggccgctac aggaacaggt ggtggcggcc ctcggtgcgc tcgtactgct ccacgatggt	360
gtagtcctcg ttgtgggagg tgatgtccag cttggagtcc acgtagtagt agccggggcag	420
ctgcacgggc ttcttgcca tgtagatgga cttgaactcc accaggtagt ggccgccgtc	480
cttcagcttc agggccttgt ggatctcgcc cttcagcacg ccgtcgcggg ggtacaggcg	540
ctcggtggag gcctccagc ccatggtctt cttctgcatt acggggccgt cggaggggaa	600
gttccgccga tgaacttcac cttgtagatg aagcagccgt cctgcaggga ggagtcctgg	660
gtcacggtca ccacgccgcc gtcctcgaag ttcatacgc gctccactt gaagccctcg	720
gggaaggaca gcttcttgta gtcgggatg tcggcggggg gcttcacgta caccttgag	780
ccgtactgga actgggggga caggatgtcc caggcgaagg gcagggggcc gcccttggtc	840
accttcagct tcacggtgtt gtggccctcg taggggcggc cctcgccct cgccctcga	900
tctcgaactc gtggccgttc acggtgccct ccatgcgcac cttgaagcgc atgaactcct	960
tgatgacgtt cttggaggag cgcaccatgg tggcgaccgg tggatcccg gcccgcggt	1020
ccgtcgactc tagcgggtacc ccgattgttt agcttgttca gctgcgcttg tttatttgc	1080
tagctttcgc ttagcgacgt gttcactttg cttgtttgaa ttgaattgtc gctccgtaga	1140
cgaagcgcct ctatttatac tccggcggtc gaggggttcga aatcgataag cttggatcct	1200
aattgaatta gctctaattg aattagtctt ctaattgaat tagtctctaa ttgaattaga	1260
tccccgggcg agctcgaatt aaaccattgt gggaaccgtg cgatcaaaca aacgcgagat	1320
accgggaagt actgaaaaac agtcgctcca ggccagtggg aacatcgatg ttttgtttg	1380
acggaccctt tactctcgtc tcatataaac cgaagccagc taagatggta tacttattat	1440
catcttgtga tgaggatgct tctatcaacg aaagtaccgg taaaccgcaa atggttatgt	1500
attataatca aactaaaggc ggagtggaca cgctagacca aatgtgttct gtgatgacct	1560
gcagtaggaa gacgaatagg tggcctatgg cattattgta cggaatgata aacattgcct	1620
gcataaattc ttttattata tacagccata atgtcagtag caaggagaa aaggttcaaa	1680
gtcgcaaaaa atttatgaga aacctttaca tgagcctgac gtcacgttt atgcgtaagc	1740
gtttagaagc tcctactttg aagagatatt tgccgcgataa tatctctaata attttgccaa	1800
atgaagtgcc tggatcatca gatgacagta ctgaagagcc agtaatgaaa aaacgtactt	1860
actgtactta ctgcccctct aaaataaggc gaaaggcaaa tgcacgtgc aaaaaatgca	1920

aaaaagttat ttgtcgagag cataatattg atatgtgcca aagttgtttc tgactgacta	1980
ataagtataa tttgtttcta ttatgtataa gttaagctaa ttacttattt tataatacaa	2040
catgactggt tttaaagtac aaaataagtt tatttttgta aaagagagaa tgtttaaaag	2100
ttttgttact ttatagaaga aattttgagt ttttgttttt ttttaataaa taaataaaca	2160
taaataaatt gtttggtgaa tttattatta gtatgtaagt gtaaataataa taaaacttaa	2220
tatctattca aattaataaa taaacctcga tatacagacc gataaaacac atgcgtcaat	2280
tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggtaaataa	2340
tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taccgtatat ctcaacgctg	2400
tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcatcgactt gatattgtcc	2460
gacacatttt cgtcgatttg cgttttgatc aaagacttga gcagagacac gttaatcaac	2520
tgttcaaatt gatccatatt aacgatatca acccgatgcg tatatgggtgc gtaaaatata	2580
ttttttaacc ctcttatact ttgcactctg cgttaatacg cgttcgtgta cagacgtaat	2640
catgttttct tttttggata aaactcctac tgagtttgac ctcatattag accctcacia	2700
gttgcaaaac gtggcatttt ttaccaatga agaatttaaa gttatttttaaaaatttcat	2760
cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaatcg accagttaat	2820
caacgtgtac acagacgcgt cggcaaaaaa cacgcagccc gacgtggttg ctaaaattat	2880
taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttcctca aaaagttgaa	2940
gaccaacaag tttacggaca ctattaatta tttgattttg cccacttca ttttggtgga	3000
tcacaatttt gttatatattt taaaacaaag ctttggcact ggccgctggt ttacaacgtc	3060
gtgactggga aaaccctggc gttaccaaac ttaatcgct tgcagacat cccctttcg	3120
ccagctggcg taatagcgaa gaggccgca ccgatcgccc ttcccaacag ttgcgcagcc	3180
tgaatggcga atggcgctg atgcggtatt ttctccttac gcatctgtgc ggtatttcac	3240
accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc	3300
gacacccgcc aacaccgct gacgcgcctt gacgggcttg tctgctcccg gcatccgctt	3360
acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac	3420
cgaaacgcgc gagacgaaag ggctcgtga tacgcctatt tttatagggt aatgtcatga	3480
taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc ggaacccta	3540
tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat	3600
aaatgcttca ataattattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc	3660

ttattccctt	ttttgoggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctggtga	3720
aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	3780
acagcggtaa	gaccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	3840
ttaaagttct	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	gagcaactcg	3900
gtcgccgcat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	3960
atcttacgga	tggtcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	4020
acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	4080
tgcaacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	4140
ccataccaaa	cgacgagcgt	gacaccacga	tgctgttagc	aatggcaaca	acgttgcgca	4200
aactattaac	tggtgaacta	cttactctag	cttcccgga	acaattaata	gactggatgg	4260
aggcggataa	agttgcagga	ccacttctgc	gctcgccct	tccggctggc	tggtttattg	4320
ctgataaatc	tggtgacggg	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	4380
atggtgaagc	ctcccgatc	gtagtattct	acacgacggg	gagtcaggca	actatggatg	4440
aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	4500
accaagttta	ctcatatata	ctttagattg	atttaaaact	tcatTTTTaa	tttaaaagga	4560
tctaggtgaa	gacctttttt	gataatctca	tgaccaaaat	cccttaacgt	gagttttcgt	4620
tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	cctttttttc	4680
tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcgggtg	gtttgtttgc	4740
cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	4800
caaatactgt	tcttctagt	tagccgtagt	taggccacca	cttcaagaac	tctgtagcac	4860
cgcctacata	cctcgctctg	ctaactctgt	taccagtggc	tgctgccagt	ggcgataagt	4920
cgtgtcttac	cgggttggtg	tcaagacgat	agttaccgga	taaggcgag	cggtcgggct	4980
gaacgggggg	ttcgtgcaca	cagcccagct	tggtgagcga	gacctacacc	gaactgagat	5040
acctacagcg	tgagctatga	gaaagcgcca	cgcttcccga	aggagagaa	gcggacaggt	5100
atccggtaag	cggcagggtc	ggaacaggag	agcgcacgag	ggagcttcca	gggggaaacg	5160
cctgatatct	ttatagtcct	gtcgggtttc	gccacctctg	acttgagcgt	cgatttttgt	5220
gatgctcgtc	acggggggcg	gagcctatgg	aaaaacgcca	gcaacgcggc	ctttttacgg	5280
ttcctggcct	tttgctggcc	ttttgctcac	atgttctttc	ctgcgttatc	ccctgattct	5340

gtggataacc	gtattaccgc	ctttgagtga	gctgataccg	ctcgccgcag	ccgaacgacc	5400
gagcgcagcg	agtcagtgag	cgaggaagcg	gaagagcgcc	caatacgcaa	accgcctctc	5460
cccgcgcggt	ggccgattca	ttaatgcagc	tggcacgaca	ggtttcccca	ctggaaagcg	5520
ggcagtgagc	gcaacgcaat	taatgtgagt	tagctcactc	attaggcacc	ccaggcttta	5580
cactttatgc	ttccggctcg	tatgttggtg	ggaattgtga	gcggataaca	atttcacaca	5640
ggaaacagct	atgacatgat	tacgaattcg	agctcggtag	ccgggggatcc	tctagagtcg	5700
acgctcgcgc	gacttggttt	gccattcttt	agcgcgcgtc	gcgtcacaca	gcttggccac	5760
aatgttggttt	ttgtcaaacg	aagattctat	gacgtgttta	aagtttaggt	cgagtaaagc	5820
gcaaattcttt	tttaacccta	gaaagatagt	ctgcgtaaaa	ttgacgcatg	cattcttgaa	5880
atattgctct	ctctttctaa	atagcgcgaa	tccgtcgcgtg	tgcathtagg	acatctcagt	5940
cgccgcttgg	agctcccgtg	aggcgtgctt	gtcaatgcgg	taagtgtcac	tgattttgaa	6000
ctataacgac	cgcgtgagtc	aaaatgacgc	atgattatct	tttacgtgac	ttttaagatt	6060
taactcatac	gataattata	ttgttatttc	atgttctact	tacgtgataa	cttattatat	6120
atataattttc	ttgttataga	tatcgtgact	aatatataat	aaaatgggta	gttctttaga	6180
cgatgagcat	atcctctctg	ctcttctgca	aagcgatgac	gagcttggtg	gtgaggattc	6240
tgacagtgaa	atatcagatc	acgtaagtga	agatgacgtc	cagagcgata	cagaagaagc	6300
gtttatagat	gaggtacatg	aagtgcagcc	aacgtcaagc	ggtagtgaag	tattagacga	6360
acaaaatggt	attgaacaac	caggttcttc	atagattctg	ttagaagcca	aagaatcttg	6420
accttgccac	agaggactat	tagaggtaag	aataaacatt	gttgggtcaac	ttcaaagtcc	6480
acgaggcgta	gccgagtctc	tgcaactgaac	attgtcagat	ccgagatcgg	ccggcctagg	6540
cgcgcccaagc	ttaaggtgca	cggcccacgt	ggccactagt	acttctcgag	gctcaaagcc	6600
tcatcccaat	ttggagtcac	tcaagacatc	cttgattaag	gcagctgccg	atattgacat	6660
ggacctcggt	cgtgctgcga	tagacgactg	gccgcgcaga	ttgaaggcct	gtattcaaaa	6720
tcacggaggt	cattttgaat	aaactttagt	gtcataagaa	tctatgtttt	gttaagttca	6780
ttttggtata	tgaatggtta	cataatgaat	aaacttggtt	caattatttt	acattaaaca	6840
tgtgacagaa	tttatgacct	gactaggtag	gtacaaacag	cctttttgat	attagaaaac	6900
taagtaaaat	agcctacggt	cacatctctt	tccgtgggtg	tcgttaaagg	gcgacttaga	6960
gaaccaccaa	gaacgtagca	gaatcctcag	agtgtcatac	cagcatacag	ccatcgctaa	7020
ctgctattta	ctggtaatag	ggcacattgt	aatctcactt	aaccatactg	tcgggccacc	7080

atctagccta	tttctgccac	gaatcaatcg	tgagtgatgg	acatagagaa	actattagtt	7140
gagaagaaaa	caagagcact	aaaggtttga	tattgacaaa	aatctacttc	gccgtcactc	7200
cataggttta	ttgtctctca	ttagtccaga	acagcagtta	cagacgtaag	cttttacgca	7260
caaaactacag	ggttgctctt	tattgtatcg	aaaatatggg	acctgaataa	gggcgatttt	7320
gacgcgtcct	gcccgcccat	tcccgatcct	acggacagaa	tggaagcag	tcgacgtcgc	7380
cccaaacacg	tcatttcgga	tcctcacgat	ccactaacgg	tgcttttaggt	acctcaagca	7440
ccggtcatcg	ttctcgtcgg	acccgctcgt	tgcgacgaag	ggctcgacga	gcaaattaac	7500
cctcagacac	agcccactga	gtttctcgcc	ggatcttctc	agcgggtcgc	gtttccgatac	7560
cgggtggtaga	ttctgcgaag	cacggctctt	gctaggattc	gtgttagcaa	cgctcgtcagg	7620
tttgagcccc	gtgagctcac	ttactagtta	aggttacgct	gaaatagcct	ctcaaggctc	7680
tcagctaggt	aggaaacaaa	aaaaaaagtc	ctgcccttaa	caccgttgcg	atggcttgtc	7740
ttctgcaccg	cggaaagatg	ttttgtacgg	aaagtttgaa	taagtgttta	attgcaagta	7800
acgtaacaat	gttttagggg	tcggcggccg	cgggagaaa	catgaagtaa	gttctttaaa	7860
tattacaaaa	aaattgaacg	atattataaa	attcttttaa	atattaaaag	taagaacaat	7920
aagatcaatt	aatcataat	taatcacatt	gttcatgatac	acaatttaat	ttacttcata	7980
cgttggtattg	ttatgtttaa	taaaaagatt	aatttctatg	taattgtatc	tgtacaatac	8040
aatgtgtaga	tgtttattct	atcgaaagta	aatacgtcaa	aactcgaaaa	ttttcagtat	8100
aaaaagggtc	aactttttca	aatcagcatac	agttcgggtc	caactctcaa	gatgagagtc	8160
aaaacctttg	tgatcttggtg	ctgcgctctg	cagggtgagtt	aattattttta	ctattatttc	8220
agaagggtggc	cagacgatat	cacgggccac	ctgataataa	gtggtcgcca	aaacgcacag	8280
atatcgtaaa	ttgtgccatt	tgatttgtca	cggccggggg	ggctacggaa	taaactacat	8340
ttattttattt	aaaaaatgaa	ccttagatta	tgtaacttgt	gatttatttg	cgtcaaaagt	8400
aggcaagatg	aatctatgta	aatacctggg	cagacttgca	atatacctatt	tcaccggtaa	8460
atcagcattg	caatatgcaa	tgcatattca	acaatatgta	aaacaattcg	taaagcatca	8520
ttagaaaata	gacgaaagaa	attgcataaa	attataaccg	cattattaat	ttattatgat	8580
atctattaac	aattgctatt	gccttttttt	cgcaaattat	aatcattttc	ataacctcga	8640
ggtagcattc	tgttacattt	taatacattg	gtatgtgatt	ataacacgag	ctgccactg	8700
agtttctcgc	cagatcttct	cagtgggtcg	cgttaccgat	cacgtgatag	attctatgaa	8760

gcactgctct	tgtagggct	agtgttagca	aattctttca	ggttgagtct	gagagctcac	8820
ctacccatcg	gagcgtagct	ggaataggct	accagcta	aggtagggaa	aacaaagctc	8880
gaaacaagct	caagtaataa	caacataatg	tgaccataaa	atctcgtggg	gtatgagata	8940
caattatgta	ctttccca	aatgtttaca	taattagaat	gttggttcaac	ttgcctaacg	9000
ccccagctag	aacattcaat	tattactatt	accactacta	aggcagtatg	tcctaactcg	9060
ttccagatca	gcgctaactt	cgattgaatg	tgcgaaattt	atagctcaat	attttagcac	9120
ttatcgtatt	gatttaagaa	aaaattgtta	acattttgtt	tcagtatgtc	gcttatacaa	9180
atgcaaaca	caatgat	gatgaggact	attttgggag	tgatgtcact	gtccaaagta	9240
gtaatacaac	agatgaaata	attagagatg	catctggggc	agttatcgaa	gaacaaatta	9300
caactaaaa	aatgcaacgg	aaaaataaaa	accatggaat	acttgga	aatgaaaaa	9360
tgatcaagac	gttcgttata	accacggatt	ccgacggtaa	cgagtcatt	gtagaggaag	9420
atgtgctcat	gaagacactt	tccgatggta	ctgttgctca	aagttatgtt	gctgctgatg	9480
cgggagcata	ttctcagagc	gggccatacg	tatcaaacag	tggtacagc	actcatcaag	9540
gatatacgag	cgatttcagc	actagtgtg	cagtcgttct	agacgtgagc	aagggcgagg	9600
agctgttcac	cggggtggg	cccatcctgg	tcgagctgga	cggcgacgta	aacggccaca	9660
agttcagcgt	gtccggcgag	ggcgagggcg	atgccaccta	cggcaagctg	accctgaagt	9720
tcattctgcac	caccggcaag	ctgcccgtgc	cctggcccac	cctcgtgacc	accctgacct	9780
acggcgtgca	gtgcttcagc	cgctaccccg	accacatgaa	gcagcacgac	ttcttcaagt	9840
ccgccatgcc	cgaaggctac	gtccaggagc	gcaccatctt	cttcaaggac	gacggcaact	9900
acaagacccg	cgccgaggtg	aagttcgagg	gcgacaccct	ggtgaaccgc	atcgagctga	9960
agggcatcga	cttcaaggag	gacggcaaca	tcctggggca	caagctggag	tacaactaca	10020
acagccacaa	cgtctatatc	atggccgaca	agcagaagaa	cggcatcaag	gtgaacttca	10080
agatccgcca	caacatcgag	gacggcagcg	tgagctcgc	cgaccactac	cagcagaaca	10140
cccccatcgg	cgacggcccc	gtgctgctgc	ccgacaacca	ctacctgagc	accagtc	10200
cctgagcaa	agacccaac	gagaagcgcg	atcacatggt	cctgctggag	ttcgtgaccg	10260
ccgccgggat	cactctcggc	atggacgagc	tgtacaagct	ggatcccccg	ggtggagcag	10320
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	10380
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggcccg	10440
ctggtccagg	ctccgctgca	gcggcggtg	ctgcagcagg	tccgggtgga	gcaggaccag	10500

gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	10560
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	10620
caggctccgc	tgcagcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	10680
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggagcagga	ccaggaggtg	10740
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggcccgtct	ggtccaggct	10800
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	10860
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	10920
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	10980
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	11040
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	11100
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	11160
cggctgctgc	agcaggctcc	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	11220
gaccaggagg	tgctggctcc	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	11280
gaccaggagg	tgctggctcc	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	11340
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	11400
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	11460
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	11520
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	11580
ctggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	11640
ctggctccgg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	11700
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11760
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11820
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	11880
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	11940
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	12000
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggctcc	ggtggagcag	12060
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggctcc	ggtggagcag	12120
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggctcc	ggtggcccgt	12180

ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	12240
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	12300
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	12360
caggctccgc	tgacgcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	12420
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggagcagga	ccaggaggtg	12480
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggcccgtct	ggtccaggct	12540
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	12600
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	12660
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	12720
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	12780
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	12840
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	12900
cggctgctgc	agcaggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	12960
gaccaggagg	tgctgggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	13020
gaccaggagg	tgctgggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	13080
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	13140
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	13200
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgacgcggcg	gctgctgcag	13260
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	13320
ctggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	13380
ctggctccgg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	13440
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	13500
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	13560
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	13620
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	13680
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	13740
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggtccg	ggtggagcag	13800
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggagcag	13860
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggcccgt	13920

ctggtccagg	ctccgctgca	gcggcggtg	ctgcagcagg	tccgggtgga	gcaggaccag	13980
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	14040
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	14100
caggctccgc	tgcagcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	14160
ctggacctgg	tggctgctgga	ccaggaggtg	ctggtccggg	tggagcagga	ccaggaggtg	14220
ctggacctgg	tggctgctgga	ccaggaggtg	ctggtccggg	tggcccgtct	gttccaggct	14280
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	gtgctggac	14340
ctggtggtgc	tggaccagga	gtgctggtc	cgggtggagc	aggaccagga	gtgctggac	14400
ctggtggtgc	tggaccagga	gtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	14460
cagcggcggc	tgctgcagca	gttccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	14520
gtgctggacc	aggaggtgct	gttccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	14580
gtgctggacc	aggaggtgct	gttccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	14640
cggctgctgc	agcagggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	14700
gaccaggagg	tgctggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	14760
gaccaggagg	tgctggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggtg	14820
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	14880
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	14940
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	15000
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	15060
ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	15120
ctggtccggg	tggcccgtct	gttccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	15180
cgggtggagc	aggaccagga	gtgctggac	ctggtggtgc	tggaccagga	gtgctggtc	15240
cgggtggagc	aggaccagga	gtgctggac	ctggtggtgc	tggaccagga	gtgctggtc	15300
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	gttccgggtg	15360
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gttccgggtg	15420
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gttccgggtg	15480
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcagggtccg	ggtggagcag	15540
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	15600

gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggcccgt	15660
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	15720
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	15780
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	15840
caggctccgc	tgacgcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	15900
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggagcagga	ccaggaggtg	15960
ctggacctgg	tggctgctgga	ccaggaggtg	ctggctccgg	tggcccgtct	ggtccaggct	16020
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	16080
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	16140
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	16200
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	16260
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	16320
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	16380
cggctgctgc	agcaggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	16440
gaccaggagg	tgctggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	16500
gaccaggagg	tgctggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	16560
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	16620
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	16680
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgacgcggcg	gctgctgcag	16740
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	16800
ctggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggctgctgga	ccaggaggtg	16860
ctggctccgg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	16920
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	16980
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	17040
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	17100
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	17160
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	17220
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggtccg	ggtggagcag	17280
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	17340

```

gaccaggagg tgctggacct ggtgggtgctg gaccaggagg tgctgggtccg ggtggcccg 17400
ctggtccagg ctccgctgca gcggcggtg ctgcagcagg tccgggtgga gcaggaccag 17460
gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag 17520
gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtggc ccgtctggtc 17580
caggctccgc tgcagcggcg gctgctgcag cagggtccggg aagcgtcagt tacggagctg 17640
gcaggggata cggacaaggt gcaggaagtg cagcttcctc tgtgtcatct gcttcatctc 17700
gcagttacga ctattctcgt cgtaacgtcc gcaaaaactg tggaattcct agaagacaac 17760
tagttgttaa attcagagca ctgccttgctg tgaattgcta atttttaata taaaataacc 17820
cttgtttctt acttcgtcct ggatacatct atgttttttt tttcgtaaat aaatgagagc 17880
atttaagtta ttgtttttta ttactttttt ttagaaaaca gatttcggat tttttgatg 17940
cattttatth gaatgtacta atataatcaa ttaatcaatg aattcattta ttttaaggat 18000
aacaataatc catgaattca catgcacatt taaaacaaaa ctaaattaca ataggttcat 18060
ataaaaacaa caagtatgcc ttctcaacta agaatactat ag 18102

```

<210> 32

<211> 12516

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 32

```

tcgacgtccc atggccattc gaattcggcc ggccctaggcg cgccgtacgc gtatcgataa 60
gctttaagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 120
tttatttgctg aaatttgctg tgctattgct ttatttgtaa ccattataag ctgcaataaa 180
caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga ggtgtgggag 240
gttttttaaa gcaagtaaaa cctctacaaa tgtgggtatg ctgattatga tctagagtcg 300
cgcccgctac aggaacaggt ggtggcggcc ctcggtgcgc tcgtactgct ccacgatggt 360
gtagtcctcg ttgtgggagg tgatgtccag cttggagtcc acgtagtagt agccgggcag 420
ctgcacgggc ttcttgcca ttagatgga cttgaactcc accaggtagt ggccgccgtc 480
cttcagcttc agggccttgt ggatctcgcc cttcagcacg ccgtcgcggg ggtacaggcg 540
ctcggtggag gcctcccagc ccatggtctt cttctgcatt acggggccgt cgaggggaa 600

```

gttccgccga tgaacttcac cttgtagatg aagcagccgt cctgcagggg	ggagtccttg	660
gtcacgggtca ccacgccgcc gtcctcgaag ttcatacacgc gctcccactt	gaagccctcg	720
gggaaggaca gcttcttgta gtcgggggatg tcggcgggggt gcttcacgta	caccttgga	780
ccgtactgga actgggggga caggatgtcc caggcgaagg gcagggggcc	gcccttggtc	840
accttcagct tcacgggtgtt gtggccctcg taggggcggc cctcgcccct	cgcccctcga	900
tctcgaactc gtggccgttc acggtgccct ccatgcgcac cttgaagcgc	atgaactcct	960
tgatgacgtt cttggaggag cgcacatggt tggcgaccgg tggatcccg	gcccgcggt	1020
ccgtcgactc tagcgggtacc ccgattgttt agcttgttca gctgcgcttg	tttatttgct	1080
tagctttcgc ttagcgacgt gttcactttg cttgtttgaa ttgaattgtc	gctccgtaga	1140
cgaagcgcct ctatttatac tccggcggtc gaggggttca aatcgataag	cttggtacct	1200
aattgaatta gctctaattg aattagtctt ctaattgaat tagtctctaa	ttgaattaga	1260
tccccgggcg agctcgaatt aaaccattgt gggaaccgtg cgatcaaaca	aacgcgagat	1320
accgggaagt actgaaaaac agtcgctcca ggccagtggg aacatcgatg	ttttgttttg	1380
acggacccct tactctcgtc tcatataaac cgaagccagc taagatggta	tacttattat	1440
catcttgtga tgaggatgct tctatcaacg aaagtaccgg taaaccgcaa	atggttatgt	1500
attataatca aactaaaggc ggagtggaca cgctagacca aatgtgttct	gtgatgacct	1560
gcagtaggaa gacgaatagg tggcctatgg cattattgta cggaatgata	aacattgcct	1620
gcataaattc ttttattata tacagccata atgtcagtag caaggagaa	aaggttcaaa	1680
gtcgcaaaaa atttatgaga aacctttaca tgagcctgac gtcacgtttt	atgcgtaagc	1740
gtttagaagc tcctactttg aagagatatt tgcgcgataa tatctcta	atatttgcaa	1800
atgaagtgcc tggatcatca gatgacagta ctgaagagcc agtaatgaa	aaacgtactt	1860
actgtactta ctgcccctct aaaataaggc gaaaggcaaa tgcacgtgc	aaaaaatgca	1920
aaaaagttat ttgtcgagag cataatattg atatgtgcca aagttgtttc	tgactgacta	1980
ataagtataa tttgtttcta ttatgtataa gttaagctaa ttacttattt	tataatacaa	2040
catgactggt tttaaagtac aaaataagtt tatttttgta aaagagagaa	tgtttaaaag	2100
ttttgttact ttatagaaga aattttgagt ttttgttttt ttttaataaa	taaataaaca	2160
taaataaatt gtttggtgaa tttattatta gtatgtaagt gtaaatataa	taaaacttaa	2220
tatctattca aattaataaa taaacctcga tatacagacc gataaaacac	atgcgtcaat	2280

tttacgcatg attatcttta acgtacgtca caatatgatt atctttctag ggtaaataa	2340
tagtttctaa tttttttatt attcagcctg ctgtcgtgaa taccgtatat ctcaacgctg	2400
tctgtgagat tgtcgtattc tagccttttt agtttttcgc tcatcgactt gatattgtcc	2460
gacacatttt cgtcgtattg cgttttgatc aaagacttga gcagagacac gttaatcaac	2520
tgttcaaatt gatccatatt aacgatatca acccgatgcg tatatgggtgc gtaaaatata	2580
ttttttaacc ctcttatact ttgcactctg cgttaatacg cgttcgtgta cagacgtaat	2640
catgttttct tttttggata aaactcctac tgagtttgac ctcatattag accctcacia	2700
gttgcaaaac gtggcatttt ttaccaatga agaatttaaa gttatttttaaaaatttcat	2760
cacagattta aagaagaacc aaaaattaaa ttatttcaac agtttaatcg accagttaat	2820
caacgtgtac acagacgcgt cggcaaaaaa cacgcagccc gacgtgttgg ctaaaattat	2880
taaatcaact tgtgttatag tcacggattt gccgtccaac gtgttcctca aaaagttgaa	2940
gaccaacaag tttacggaca ctattaatta tttgattttg cccacttca ttttgtggga	3000
tcacaatttt gttatatttt taaaacaaag ctttggcact ggccgtcgtt ttacaacgtc	3060
gtgactggga aaaccctggc gttaccaac ttaatcgctt tgcagcacat cccctttcg	3120
ccagctggcg taatagegaa gaggccgca ccgacgccc ttccaacag ttgcgcagcc	3180
tgaatggcga atggcgctg atgcggtatt ttctccttac gcatctgtgc ggtatttcac	3240
accgcatatg gtgcactctc agtacaatct gctctgatgc cgcatagtta agccagcccc	3300
gacacccgcc aacaccgct gacgcgcctt gacgggcttg tctgctcccg gcatccgctt	3360
acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac	3420
cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt tttatagggt aatgtcatga	3480
taataatggt ttcttagacg tcaggtggca cttttcgggg aaatgtgcgc ggaacccta	3540
tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa taaccctgat	3600
aaatgcttca ataatttga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc	3660
ttattccctt ttttgoggca ttttgccttc ctgtttttgc tcaccagaa acgctggtga	3720
aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca	3780
acagcggtaa gatccttgag agttttcgcc ccgaagaacg tttccaatg atgagcactt	3840
ttaaagtctt gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg	3900
gtcgccgcat aactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc	3960
atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata	4020

acactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt	4080
tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag	4140
ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca	4200
aactattaac tggcgaacta cttactctag cttcccggca acaattaata gactggatgg	4260
aggcggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg	4320
ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag	4380
atggtaagcc ctcccgatc gtagttatct acacgacggg gagtcaggca actatggatg	4440
aacgaaatag acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag	4500
accaagttta ctcatatata ctttagattg atttaaaact tcatTTTTta tttaaaagga	4560
tctaggtgaa gatccttttt gataatctca tgacccaaaat cccttaacgt gagttttcgt	4620
tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat ctttttttc	4680
tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc	4740
cggatcaaga gctaccaact ctttttccga aggtaactgg cttcagcaga gcgcagatac	4800
caaatactgt tcttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac	4860
cgcctacata cctcgtctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt	4920
cgtgtcttac cgggttgac tcaagacgat agttaccgga taaggcgag cggtcgggct	4980
gaacgggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat	5040
acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt	5100
atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca ggggaaacg	5160
cctgatatct ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt	5220
gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg	5280
ttcctggcct tttgctggcc ttttctcac atgttctttc ctgcgttatc ccctgattct	5340
gtggataacc gtattaccgc ctttgagtga gctgataccg ctcgccgcag ccgaacgacc	5400
gagcgcagcg agtcagtgag cgaggaagcg gaagagcgcc caatacgcaa accgcctctc	5460
cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg	5520
ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta	5580
cactttatgc ttccggctcg tatgttggtg ggaattgtga gcggataaca atttcacaca	5640
ggaaacagct atgacatgat tacgaattcg agctcggtag ccgggggatcc tctagagtcg	5700

acgctcgcgc gacttggttt gccattcttt agcgcgcgctc gcgtcacaca gcttggccac	5760
aatgtgggttt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaagc	5820
gcaaattcttt ttttaacccta gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa	5880
atattgctct ctcttttctaa atagcgcgaa tccgtcgctg tgcatttagg acatctcagt	5940
cgccgcttgg agctcccgtg aggcggtgctt gtcaatgcgg taagtgtcac tgattttgaa	6000
ctataacgac cgcgtagctc aaaatgacgc atgattatct tttacgtgac ttttaagatt	6060
taactcatac gataattata ttgttatttc atgttctact tacgtgataa cttattatat	6120
atatatttttc ttgttataga tatcgtgact aatatataat aaaatgggta gttctttaga	6180
cgatgagcat atcctctctg ctcttctgca aagcgatgac gagcttggtg gtgaggattc	6240
tgacagtgaa atatcagatc acgtaagtga agatgacgtc cagagcgata cagaagaagc	6300
gtttatagat gaggtacatg aagtgcagcc aacgtcaagc ggtagtgaaa tattagacga	6360
acaaaatggt attgaacaac caggttcttc atagattctg ttagaagcca aagaatcttg	6420
accttgccac agaggactat tagaggtaag aataaacatt gttggtcaac ttcaaagtcc	6480
acgaggcgta gccgagtctc tgcactgaac attgtcagat ccgagatcgg ccggcctagg	6540
cgcgccaaagc ttaaggtgca cggccacgt ggccactagt acttctcgag gctcaaagcc	6600
tcatcccaat ttggagtcac tcaagacatc cttgattaag gcagctgccg atattgacat	6660
ggacctcggt cgtgctgca tagacgactg gccgcgcaga ttgaaggcct gtattcaaaa	6720
tcacggaggt catTTTTgaat aaactttagt gtcataagaa tctatgtttt gttaagttca	6780
TTTTggTata tgaatggTta cataatgaat aaacttgTtt caattatttt acattaaaca	6840
tgtgacagaa tttatgacct gactaggtag gtacaaacag cttttttgat attagaaaac	6900
taagtaaaat agcctacggt cacatctctt tccgtgggtg tcgttaaagg gcgacttaga	6960
gaaccaccaa gaacgtagca gaatcctcag agtgtcatac cagcatacag ccatcgctaa	7020
ctgctatttta ctggtaatag ggcacattgt aatctcactt aaccatactg tcgggccacc	7080
atctagccta tttctgccac gaatcaatcg tgagtgatgg acatagagaa actattagtt	7140
gagaagaaaa caagagcact aaaggtttga tattgacaaa aatctacttc gccgtcactc	7200
cataggttta ttgtctctca ttagtccaga acagcagtta cagacgtaag cttttacgca	7260
caaaactacag ggttgctctt tattgtatcg aaaatatggg acctgaataa gggcgatttt	7320
gacgcgtcct gcccgcccat tcccgatcct acggacagaa tggcaagcag tcgacgtcgc	7380
cccaaacacg tcatttcgga tcctcacgat ccactaacgg tgcttttaggt acctcaagca	7440

ccggtcatcg	ttctcgtcgg	acccgtcgct	tgcgacgaag	ggctcgacga	gcaaattaac	7500
cctcagacac	agcccactga	gtttctcgcc	ggatcttctc	agcgggctgc	gtttccgac	7560
cgggtgtaga	ttctgcgaag	cacggctctt	gctaggattc	gtgttagcaa	cgctcgtcagg	7620
tttgagcccc	gtgagctcac	ttactagtta	aggttacgct	gaaatagcct	ctcaaggctc	7680
tcagctaggt	aggaaacaaa	aaaaaaagtc	ctgcccttaa	caccgttgcg	atggcttgtc	7740
ttctgcaccg	cggaaagatg	ttttgtacgg	aaagtttgaa	taagtgccta	attgcaagta	7800
acgtaacaat	gttttagggg	tcggcgggcg	cgggagaaa	catgaagtaa	gttctttaaa	7860
tattacaaaa	aaattgaacg	atattataaa	attctttaaa	atattaaaag	taagaacaat	7920
aagatcaatt	aatcataat	taatcacatt	gttcatgac	acaatttaat	ttacttcata	7980
cgttggtattg	ttatgttaaa	taaaaagatt	aatttctatg	taattgtatc	tgtacaatac	8040
aatgtgtaga	tgtttattct	atcgaaagta	aatacgtcaa	aactcgaaaa	ttttcagtat	8100
aaaaagggtc	aactttttca	aatcagcatc	agttcgggtc	caactctcaa	gatgagagtc	8160
aaaacctttg	tgatcttggtg	ctgcgctctg	cagggtgagtt	aattatttta	ctattatttc	8220
agaagggtggc	cagacgatat	cacggggccac	ctgataataa	gtggctcgcca	aaacgcacag	8280
atatcgtaaa	ttgtgccatt	tgatttgta	cgcccggggg	ggctacggaa	taaactacat	8340
ttatttatatt	aaaaaatgaa	ccttagatta	tgtaacttgt	gatttatattg	cgtcaaaagt	8400
aggcaagatg	aatctatgta	aatacctggg	cagacttgca	atatacctatt	tcaccggtaa	8460
atcagcattg	caatatgcaa	tgcatattca	acaatatgta	aaacaattcg	taaagcatca	8520
ttagaaaata	gacgaaagaa	attgcataaa	attataaccg	cattattaat	ttattatgat	8580
atctattaac	aattgctatt	gccttttttt	cgcaaattat	aatcattttc	ataacctcga	8640
ggtagcattc	tgttacattt	taatacattg	gtatgtgatt	ataacacgag	ctgcccactg	8700
agtttctcgc	cagatcttct	cagtgggtcg	cgttaccgat	cacgtgatag	attctatgaa	8760
gcactgctct	tgtaggggct	agtgttagca	aattctttca	ggttgagtct	gagagctcac	8820
ctacccatcg	gagcgtagct	ggaataggct	accagcta	aggtagggaa	aacaaagctc	8880
gaaacaagct	caagtaataa	caacataatg	tgaccataaa	atctcgtggg	gtatgagata	8940
caattatgta	ctttcccaca	aatgtttaca	taattagaat	gttgttcaac	ttgcctaacg	9000
ccccagctag	aacattcaat	tattactatt	accactacta	aggcagtatg	tcctaactcg	9060
ttccagatca	gcgctaactt	cgattgaatg	tgcgaaattt	atagctcaat	atttttagcac	9120

ttatcgtatt	gatttaagaa	aaaattgtta	acattttgtt	tcagtatgtc	gcttatacaa	9180
atgcaaacat	caatgatatt	gatgaggact	attttgggag	tgatgtcact	gtccaaagta	9240
gtaatacaac	agatgaaata	attagagatg	catctggggc	agttatcgaa	gaacaaatta	9300
caactaaaaa	aatgcaacgg	aaaaataaaa	accatggaat	acttggaana	aatgaaaaaa	9360
tgatcaagac	gttcggtata	accacggatt	ccgacggtaa	cgagtccatt	gtagaggaag	9420
atgtgctcat	gaagacactt	tccgatggta	ctgttgctca	aagttatgtt	gctgctgatg	9480
cgggagcata	ttctcagagc	gggccatacg	tatcaaacag	tggatacagc	actcatcaag	9540
gatatacgag	cgatttcagc	actagtgtct	cagtcgttct	agacctggat	cccccggtg	9600
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gtcccggtg	9660
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	gtcccggtg	9720
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcagggtccg	gtggagcag	9780
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	gtggagcag	9840
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	gtggccccgt	9900
ctgggtccagg	ctccgctgca	gcggcggtcg	ctgcagcagg	tccgggtgga	gcaggaccag	9960
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	10020
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	10080
caggctccgc	tgacgcggcg	gctgctgcag	cagggtccggg	tggagcagga	ccaggaggtg	10140
ctggacctgg	tgggtgctgga	ccaggaggtg	ctgggtccggg	tggagcagga	ccaggaggtg	10200
ctggacctgg	tgggtgctgga	ccaggaggtg	ctgggtccggg	tggccccgtct	gtccaggct	10260
ccgctgcagc	ggcggtgct	gcagcaggtc	cgggtggagc	aggaccagga	gtgctggac	10320
ctgggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	gtgctggac	10380
ctgggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	10440
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	10500
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	10560
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	10620
cggctgctgc	agcagggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	10680
gaccaggagg	tgctgggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	10740
gaccaggagg	tgctgggtccg	ggtggccccgt	ctgggtccagg	ctccgctgca	gcggcggtcg	10800
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	10860

gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	10920
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	10980
caggctccggg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11040
ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11100
ctggtccggg	tggcccgtct	ggtccaggct	ccgctgcagc	ggcggctgct	gcagcaggtc	11160
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11220
cgggtggagc	aggaccagga	ggtgctggac	ctggtggtgc	tggaccagga	ggtgctggtc	11280
cgggtggccc	gtctggtcca	ggctccgctg	cagcggcggc	tgctgcagca	ggtccgggtg	11340
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	11400
gagcaggacc	aggaggtgct	ggacctggtg	gtgctggacc	aggaggtgct	ggtccgggtg	11460
gcccgtctgg	tccaggctcc	gctgcagcgg	cggctgctgc	agcaggctccg	ggtggagcag	11520
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggagcag	11580
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctggtccg	ggtggcccgt	11640
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	11700
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	11760
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	11820
caggctccgc	tgcagcggcg	gctgctgcag	caggctccggg	tggagcagga	ccaggaggtg	11880
ctggacctgg	tggtgctgga	ccaggaggtg	ctggctccggg	tggagcagga	ccaggaggtg	11940
ctggacctgg	tggtgctgga	ccaggaggtg	ctggctccggg	tggcccgtct	ggtccaggct	12000
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggaagcgt	cagttacgga	gctggcaggg	12060
gatacggaca	aggtgcagga	agtgcagctt	cctctgtgtc	atctgcttca	tctcgcagtt	12120
acgactattc	tcgtcgtaac	gtccgcaaaa	actgtggaat	tcctagaaga	caactagttg	12180
ttaaattcag	agcactgcct	tgtgtgaatt	gctaattttt	aatataaaat	aacccttggt	12240
tcttacttcg	tcctggatac	atctatgttt	tttttttcgt	taataaatga	gagcatttaa	12300
gttattgttt	ttaattactt	tttttttagaa	aacagatttc	ggattttttg	tatgcatttt	12360
atttgaatgt	actaatataa	tcaattaatc	aatgaattca	tttattttaag	ggataacaat	12420
aatccatgaa	ttcacatgca	catttaaaac	aaaactaaat	tacaataggt	tcatataaaa	12480
acaacaagta	tgcccttctca	actaagaata	ctatag			12516

<210> 33
 <211> 13230
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 33
 tcgacgtccc atggccattc gaattcggcc ggcctaggcg cgccgtacgc gtatcgataa 60
 gctttaagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 120
 tttattttgtg aaattttgtga tgctattgct ttattttgtaa ccattataag ctgcaataaa 180
 caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga ggtgtgggag 240
 gtttttttaa gcaagtaaaa cctctacaaa tgtgggtatg ctgattatga tctagagtcg 300
 cgcccgctac aggaacaggt ggtggcggcc ctcggtgcmc tcgtactgct ccacgatggt 360
 gtagtcctcg ttgtgggagg tgatgtccag cttggagtcc acgtagtagt agccgggcag 420
 ctgcacgggc ttcttggcca ttagatgga cttgaactcc accaggtagt ggccgccgctc 480
 cttcagcttc agggccttgt ggatctcgcc cttcagcacg ccgtcgcggg ggtacaggcg 540
 ctcggtggag gcctcccagc ccatggtctt cttctgcatt acggggccgt cgaggggaa 600
 gttccgccga tgaacttcac cttgtagatg aagcagccgt cctgcaggga ggagtcctgg 660
 gtcacggtca ccacgcgcc gtcctcgaag ttcacacgc gtcacctt gaagccctcg 720
 gggaaggaca gcttcttgta gtcggggatg tcggcggggt gcttcacgta caccttggag 780
 ccgtactgga actgggggga caggatgtcc caggcgaagg gcagggggcc gcccttggtc 840
 accttcagct tcacggtgtt gtggccctcg tagggcggc cctcgccct cgccctcgga 900
 tctcgaactc gtggccgttc acggtgccct ccatgcgcac cttgaagcgc atgaactcct 960
 tgatgacgtt cttggaggag cgcacatgg tggcgaccgg tggatcccg gcccgcggtg 1020
 ccgtcgactc tagcgttacc ccgattgttt agcttgttca gctgcgcttg tttatttgct 1080
 tagctttcgc ttagcgacgt gttcactttg cttgtttgaa ttgaattgtc gctccgtaga 1140
 cgaagcgcct ctatttatac tccggcggtc gagggttcga aatcgataag cttggatcct 1200
 aattgaatta gctctaattg aattagtctt ctaattgaat tagtctctaa ttgaattaga 1260
 tccccgggag agctcgaatt aaaccattgt gggaaccgtg cgatcaaaca aacgcgagat 1320
 accgggaagt actgaaaaac agtcgctcca ggccagtggg aacatcgatg ttttgttttg 1380

acggaccctt	tactctcgtc	tcatataaac	cgaagccagc	taagatggta	tacttattat	1440
catcttgtga	tgaggatgct	tctatcaacg	aaagtaccgg	taaaccgcaa	atggttatgt	1500
attataatca	aactaaaggc	ggagtggaca	cgctagacca	aatgtgttct	gtgatgacct	1560
gcagtaggaa	gacgaatagg	tggcctatgg	cattattgta	cggaatgata	aacattgcct	1620
gcataaattc	ttttattata	tacagccata	atgtcagtag	caagggagaa	aaggttcaaa	1680
gtcgcaaaaa	atztatgaga	aacctttaca	tgagcctgac	gtcatcgttt	atgcgtaagc	1740
gtttagaagc	tcctactttg	aagagatatt	tgcgcgataa	tatctctaata	attttgccaa	1800
atgaagtgcc	tggtacatca	gatgacagta	ctgaagagcc	agtaatgaaa	aaacgtactt	1860
actgtactta	ctgcccctct	aaaataaggc	gaaaggcaaa	tgcatcggtc	aaaaaatgca	1920
aaaaagttat	ttgtcgagag	cataatattg	atatgtgcca	aagttgtttc	tgactgacta	1980
ataagtataa	tttgtttcta	ttatgtataa	gttaagctaa	ttacttattt	tataatacaa	2040
catgactgtt	tttaaagtac	aaaataagtt	tatttttgta	aaagagagaa	tgtttaaaag	2100
ttttgttact	ttatagaaga	aattttgagt	ttttgttttt	ttttaataaa	taaataaaca	2160
taaataaatt	gtttgttgaa	tttattatta	gtatgtaagt	gtaaataataa	taaaacttaa	2220
tatctattca	aattaataaa	taaacctcga	tatacagacc	gataaaacac	atgcgtcaat	2280
tttacgcatg	attatcttta	acgtacgtca	caatatgatt	atctttctag	ggttaaataa	2340
tagtttctaa	tttttttatt	attcagcctg	ctgtcgtgaa	taccgtatat	ctcaacgctg	2400
tctgtgagat	tgtcgtattc	tagccttttt	agtttttcgc	tcatcgactt	gatattgtcc	2460
gacacatttt	cgtcgatttg	cgttttgatc	aaagacttga	gcagagacac	gttaatcaac	2520
tgttcaaatt	gatccatatt	aacgatatca	acccgatgcg	tatatgggtc	gtaaaatata	2580
ttttttaacc	ctcttatact	ttgcactctg	cgtttaatacg	cgttcgtgta	cagacgtaat	2640
catgttttct	tttttgata	aaactcctac	tgagtttgac	ctcatattag	accctcacia	2700
gttgcaaaac	gtggcatttt	ttaccaatga	agaatttaaa	gttattttaa	aaaatttcac	2760
cacagattta	aagaagaacc	aaaaattaaa	ttatttcaac	agtttaatcg	accagttaat	2820
caacgtgtac	acagacgcgt	cggcaaaaaa	cacgcagccc	gacgtggttg	ctaaaattat	2880
taaatcaact	tgtgttatag	tcacggattt	gccgtccaac	gtgttcctca	aaaagttgaa	2940
gaccaacaag	tttacggaca	ctattaatta	tttgattttg	ccccacttca	ttttgtggga	3000
tcacaatttt	gttatatttt	taaaacaaag	ctttggcact	ggccgtcggt	ttacaacgtc	3060
gtgactggga	aaaccctggc	gttaccacaac	ttaatcgctt	tgacgcacat	ccccctttcg	3120

ccagctggcg	taatagcgaa	gaggcccgca	ccgatcgccc	ttcccaacag	ttgcgcagcc	3180
tgaatggcga	atggcgccctg	atgcggtatt	ttctccttac	gcatctgtgc	ggtattttcac	3240
accgcatatg	gtgcactctc	agtacaatct	gctctgatgc	cgcatagtta	agccagcccc	3300
gacacccgcc	aacacccgct	gacgcgccct	gacgggcttg	tctgctcccg	gcatccgctt	3360
acagacaagc	tgtgaccgtc	tccgggagct	gcatgtgtca	gaggttttca	ccgtcatcac	3420
cgaaacgcgc	gagacgaaag	ggcctcgtga	tacgcctatt	tttatagggt	aatgtcatga	3480
taataatggt	ttcttagacg	tcaggtggca	cttttcgggg	aaatgtgcgc	ggaacccta	3540
tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	taaccctgat	3600
aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	cgtgtcgccc	3660
ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	acgctggtga	3720
aagtaaaaga	tgctgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	ctggatctca	3780
acagcggtaa	gacccctgag	agttttcgcc	ccgaagaacg	ttttccaatg	atgagcactt	3840
ttaaagtctt	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	gagcaactcg	3900
gtcgccgcat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	acagaaaagc	3960
atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	atgagtgata	4020
acactgcggc	caacttactt	ctgacaacga	tcggaggacc	gaaggagcta	accgcttttt	4080
tgcaacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	ctgaatgaag	4140
ccataccaaa	cgacgagcgt	gacaccacga	tgccctgtagc	aatggcaaca	acgttgcgca	4200
aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	gactggatgg	4260
aggcggataa	agttgcagga	ccacttctgc	gctcggccct	tccggctggc	tggttttattg	4320
ctgataaatc	tggagccggt	gagcgtgggt	ctcgcggtat	cattgcagca	ctggggccag	4380
atggtaagcc	ctcccgatatc	gtagttatct	acacgacggg	gagtcaggca	actatggatg	4440
aacgaaatag	acagatcgct	gagataggtg	cctcactgat	taagcattgg	taactgtcag	4500
accaagttta	ctcatatata	ctttagattg	atttaaaaact	tcatttttta	tttaaaagga	4560
tctaggtgaa	gacccctttt	gataatctca	tgacccaaaat	cccttaacgt	gagttttcgt	4620
tccactgagc	gtcagacccc	gtagaaaaga	tcaaaggatc	ttcttgagat	cctttttttc	4680
tgcgcgtaat	ctgctgcttg	caaacaaaaa	aaccaccgct	accagcgggtg	gtttgtttgc	4740
cggatcaaga	gctaccaact	ctttttccga	aggtaactgg	cttcagcaga	gcgcagatac	4800

caaatactgt tcttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac	4860
cgctacata cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt	4920
cgtgtcttac cgggttggac tcaagacgat agttaccgga taaggcgag cggtcgggct	4980
gaacgggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat	5040
acctacagcg tgagctatga gaaagcgcca cgcttcccga agggagaaag gcggacaggt	5100
atccggtaa cggcagggtc ggaacaggag agcgcacgag ggagcttcca gggggaaacg	5160
cctgatatct ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt	5220
gatgctcgtc acggggggcg gagcctatgg aaaaacgcca gcaacgcggc ctttttacgg	5280
ttcctggcct tttgctggcc ttttgtcac atgttctttc ctgcgttatc ccctgattct	5340
gtggataacc gtattaccgc ctttgagtga gctgataccg ctgcgcgag ccgaacgacc	5400
gagcgcagcg agtcagtga cgaggaagcg gaagagcgcc caatacgcaa accgcctctc	5460
cccgcgcgtt ggccgattca ttaatgcagc tggcacgaca ggtttcccga ctggaaagcg	5520
ggcagtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc ccaggcttta	5580
cactttatgc ttccggctcg tatgttggtg ggaattgtga gcggataaca atttcacaca	5640
ggaaacagct atgacatgat tacgaattcg agctcggtag ccggggatcc tctagagtcg	5700
acgctcgcgc gacttggttt gccattcttt agcgcgcgtc gcgtcacaca gcttggccac	5760
aatgtgggtt ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaagc	5820
gcaaattctt tttaacccta gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa	5880
atattgctct ctctttctaa atagcgcgaa tccgtcgcgtg tgcatttagg acatctcagt	5940
cgccgcttgg agctcccgtag aggcgtgctt gtcaatgcgg taagtgtcac tgattttgaa	6000
ctataacgac cgcgtagtc aaaatgacgc atgattatct ttacgtgac ttttaagatt	6060
taactcatac gataattata ttgttatttc atgttctact tacgtgataa cttattatat	6120
atatattttc ttgttataga tatcgtgact aatatataat aaaatgggta gttctttaga	6180
cgatgagcat atcctctctg ctcttctgca aagcgatgac gagcttggtg gtgaggattc	6240
tgacagtga atatcagatc acgtaagtga agatgacgtc cagagcgata cagaagaagc	6300
gtttatagat gaggtacatg aagtcagcc aacgtcaagc gtagtgaaa tattagacga	6360
acaaaatggt attgaacaac caggttcttc atagattctg ttagaagcca aagaatcttg	6420
accttgccac agaggactat tagaggtaag aataaacatt gttggtcaac ttcaaagtcc	6480
acgaggcgta gccgagtctc tgcactgaac attgtcagat ccgagatcgg ccggcctagg	6540

cgcgccaagc	ttaaggtgca	cggcccacgt	ggccactagt	acttctcgag	gctcaaagcc	6600
tcattcccaat	ttggagtcac	tcaagacatc	cttgattaag	gcagctgccg	atattgacat	6660
ggacctcggt	cgtgctgcga	tagacgactg	gccgcgcaga	ttgaaggcct	gtattcaaaa	6720
tcacggaggt	cattttgaat	aaactttagt	gtcataagaa	tctatgtttt	gttaagttca	6780
ttttgggtata	tgaatgggta	cataatgaat	aaacttgttt	caattatttt	acattaaaca	6840
tgtgacagaa	tttatgacct	gactaggtag	gtacaaacag	cctttttgat	attagaaaac	6900
taagtaaaat	agcctacggt	cacatctctt	tccgtgggtg	tcgttaaagg	gcgacttaga	6960
gaaccaccaa	gaacgtagca	gaatcctcag	agtgtcatac	cagcatacag	ccatcgctaa	7020
ctgctattta	ctggtaatag	ggcacattgt	aatctcactt	aaccatactg	tcgggccacc	7080
atctagccta	tttctgccac	gaatcaatcg	tgagtgatgg	acatagagaa	actattagtt	7140
gagaagaaaa	caagagcact	aaaggtttga	tattgacaaa	aatctacttc	gccgtcactc	7200
cataggttta	ttgtctctca	ttagtccaga	acagcagtta	cagacgtaag	cttttacgca	7260
caaaactacag	ggttgctctt	tattgtatcg	aaaatatggg	acctgaataa	gggcgatttt	7320
gacgcgtcct	gcccgcocat	tcccgatcct	acggacagaa	tggcaagcag	tcgacgtcgc	7380
cccaaacacg	tcatttcgga	tcctcacgat	ccactaacgg	tgcttttaggt	acctcaagca	7440
cgggtcatcg	ttctcgtcgg	acccgtcgct	tgcgacgaag	ggctcgacga	gcaaattaac	7500
cctcagacac	agcccactga	gtttctcgcc	ggatcttctc	agcgggtcgc	gtttccgatc	7560
cggtggtaga	ttctgcgaag	cacggctctt	gctaggattc	gtgttagcaa	cgtcgtcagg	7620
tttgagcccc	gtgagctcac	ttactagtta	aggttacgct	gaaatagcct	ctcaaggctc	7680
tcagctaggt	aggaaacaaa	aaaaaaagtc	ctgcccttaa	caccgttgcg	atggcttgtc	7740
ttctgcaccg	cggaaagatg	ttttgtacgg	aaagtttgaa	taagtgttta	attgcaagta	7800
acgtaacaat	gttttagggg	tcggcggccg	cgggagaaa	catgaagtaa	gttctttaaa	7860
tattacaaaa	aaattgaacg	atattataaa	attcttttaa	atattaaaag	taagaacaat	7920
aagatcaatt	aatcataat	taatcacatt	gttcatgatc	acaatttaat	ttacttcata	7980
cgttggtattg	ttatgttaaa	taaaaagatt	aatttctatg	taattgtatc	tgtacaatac	8040
aatgtgtaga	tgtttattct	atcgaaagta	aatacgtcaa	aactcgaaaa	ttttcagtat	8100
aaaaagggtc	aactttttca	aatcagcatc	agttcgggtc	caactctcaa	gatgagagtc	8160
aaaacctttg	tgatcttgtg	ctgcgctctg	caggtgagtt	aattatttta	ctattatttc	8220

agaaggtggc	cagacgatat	cacggggccac	ctgataataa	gtggtcgccca	aaacgcacag	8280
atatcgtaaa	ttgtgccatt	tgatttgtca	cgcccggggg	ggctacggaa	taaactacat	8340
ttattttattt	aaaaaatgaa	ccttagatta	tgtaacttgt	gattttatttg	cgtaaaaagt	8400
aggcaagatg	aatctatgta	aatacctggg	cagacttgca	atatacctatt	tcaccggtaa	8460
atcagcattg	caatatgcaa	tgcatattca	acaatatgta	aaacaattcg	taaagcatca	8520
ttagaaaata	gacgaaagaa	attgcataaa	attataaccg	cattattaat	ttattatgat	8580
atctattaac	aattgctatt	gccttttttt	cgcaaattat	aatcattttc	ataacctcga	8640
ggtagcattc	tgttacattt	taatacattg	gtatgtgatt	ataacacgag	ctgcccactg	8700
agtttctcgc	cagatcttct	cagtgggtcg	cgttaccgat	cacgtgatag	attctatgaa	8760
gcactgctct	tgtagggct	agtgttagca	aattctttca	ggttgagtct	gagagctcac	8820
ctacccatcg	gagcgtagct	ggaataggct	accagctaata	aggtagggaa	aacaaagctc	8880
gaaacaagct	caagtaataa	caacataatg	tgaccataaa	atctcgtggg	gtatgagata	8940
caattatgta	ctttcccaca	aatgtttaca	taattagaat	gttgttcaac	ttgcctaacg	9000
ccccagctag	aacattcaat	tattactatt	accactacta	aggcagtatg	tcctaactcg	9060
ttccagatca	gcgctaactt	cgattgaatg	tgcgaaattt	atagctcaat	attttagcac	9120
ttatcgtatt	gatttaagaa	aaaattgtta	acattttgtt	tcagtatgtc	gcttatacaa	9180
atgcaaacat	caatgatattt	gatgaggact	attttgggag	tgatgtcact	gtccaaagta	9240
gtaatacaac	agatgaaata	attagagatg	catctggggc	agttatcgaa	gaacaaatta	9300
caactaaaaa	aatgcaacgg	aaaaataaaa	accatggaat	acttgaaaaa	aatgaaaaaa	9360
tgatcaagac	gttcgttata	accacggatt	ccgacggtaa	cgagtcatt	gtagaggaag	9420
atgtgctcat	gaagacactt	tccgatggta	ctgttgctca	aagttatgtt	gctgctgatg	9480
cgggagcata	ttctcagagc	gggccatacg	tatcaaacag	tggatacagc	actcatcaag	9540
gatatacgag	cgatttcagc	actagtgtg	cagtcgttct	agacgtgagc	aagggcgagg	9600
agctgttcac	cggggtgggtg	cccatcctgg	tcgagctgga	cggcgacgta	aacggccaca	9660
agttcagcgt	gtccggcgag	ggcgagggcg	atgccaccta	cggcaagctg	accctgaagt	9720
tcattctgcac	caccggcaag	ctgcccgtgc	cctggcccac	cctcgtgacc	accctgacct	9780
acggcgtgca	gtgcttcagc	cgctaccccg	accacatgaa	gcagcacgac	ttcttcaagt	9840
ccgccatgcc	cgaaggctac	gtccaggagc	gcaccatctt	cttcaaggac	gacggcaact	9900
acaagaccgg	cgccgaggtg	aagttcgagg	gcgacaccct	ggtgaaccgc	atcgagctga	9960

agggcatcga	cttcaaggag	gacggcaaca	tcttggggca	caagctggag	tacaactaca	10020
acagccacaa	cgtctatatc	atggccgaca	agcagaagaa	cggcatcaag	gtgaacttca	10080
agatccgcca	caacatcgag	gacggcagcg	tgcagctcgc	cgaccactac	cagcagaaca	10140
cccccatcgg	cgacggcccc	gtgctgctgc	ccgacaacca	ctacctgagc	accaggtccg	10200
ccctgagcaa	agaccccaac	gagaagcgcg	atcacatggc	cctgctggag	ttcgtgaccg	10260
ccgccgggat	cactctcggc	atggacgagc	tgtacaagct	ggatcccccg	ggtggagcag	10320
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggagcag	10380
gaccaggagg	tgctggacct	ggtggtgctg	gaccaggagg	tgctgggtccg	ggtggcccgt	10440
ctggtccagg	ctccgctgca	gcggcggctg	ctgcagcagg	tccgggtgga	gcaggaccag	10500
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtgga	gcaggaccag	10560
gaggtgctgg	acctggtggt	gctggaccag	gaggtgctgg	tccgggtggc	ccgtctggtc	10620
caggctccgc	tgcagcggcg	gctgctgcag	caggctccgg	tggagcagga	ccaggaggtg	10680
ctggacctgg	tggtgctgga	ccaggaggtg	ctggctccgg	tggagcagga	ccaggaggtg	10740
ctggacctgg	tggtgctgga	ccaggaggtg	ctggctccgg	tggcccgtct	ggtccaggct	10800
ccgctgcagc	ggcggctgct	gcagcaggtc	cgggtggagc	aggaccagga	ggtgctggac	10860
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggagc	aggaccagga	ggtgctggac	10920
ctggtggtgc	tggaccagga	ggtgctggtc	cgggtggccc	gtctggtcca	ggctccgctg	10980
cagcggcggc	tgctgcagca	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	11040
gtgctggacc	aggaggtgct	ggtccgggtg	gagcaggacc	aggaggtgct	ggacctggtg	11100
gtgctggacc	aggaggtgct	ggtccgggtg	gcccgtctgg	tccaggctcc	gctgcagcgg	11160
cggctgctgc	agcagggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	11220
gaccaggagg	tgctgggtccg	ggtggagcag	gaccaggagg	tgctggacct	ggtggtgctg	11280
gaccaggagg	tgctgggtccg	ggtggcccgt	ctggtccagg	ctccgctgca	gcggcggctg	11340
ctgcagcagg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	11400
gaggtgctgg	tccgggtgga	gcaggaccag	gaggtgctgg	acctggtggt	gctggaccag	11460
gaggtgctgg	tccgggtggc	ccgtctggtc	caggctccgc	tgcagcggcg	gctgctgcag	11520
caggctccgg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11580
ctggtccggg	tggagcagga	ccaggaggtg	ctggacctgg	tggtgctgga	ccaggaggtg	11640

ctggtccggg tggcccgtct ggtccaggct ccgctgcagc ggcggctgct gcagcaggtc	11700
cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc	11760
cgggtggagc aggaccagga ggtgctggac ctggtggtgc tggaccagga ggtgctggtc	11820
cgggtggccc gtctggtcca ggctccgctg cagcggcggc tgctgcagca ggtccgggtg	11880
gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg	11940
gagcaggacc aggaggtgct ggacctggtg gtgctggacc aggaggtgct ggtccgggtg	12000
gcccgctctgg tccaggctcc gctgcagcgg cggctgctgc agcagggtccg ggtggagcag	12060
gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggagcag	12120
gaccaggagg tgctggacct ggtggtgctg gaccaggagg tgctgggtccg ggtggcccgt	12180
ctggtccagg ctccgctgca gcggcggctg ctgcagcagg tccgggtgga gcaggaccag	12240
gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtgga gcaggaccag	12300
gaggtgctgg acctggtggt gctggaccag gaggtgctgg tccgggtggc ccgtctggtc	12360
caggctccgc tgcagcggcg gctgctgcag cagggtccggg tggagcagga ccaggaggtg	12420
ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggagcagga ccaggaggtg	12480
ctggacctgg tgggtgctgga ccaggaggtg ctggtccggg tggcccgtct ggtccaggct	12540
ccgctgcagc ggcggctgct gcagcaggtc cgggtggagc aggaccagga ggtgctggac	12600
ctggtggtgc tggaccagga ggtgctggtc cgggtggagc aggaccagga ggtgctggac	12660
ctggtggtgc tggaccagga ggtgctggtc cgggtggccc gtctggtcca ggctccgctg	12720
cagcggcggc tgctgcagca ggtccgggaa gcgtcagtta cggagctggc aggggatacg	12780
gacaagggtgc aggaagtgca gcttcctctg tgtcatctgc ttcattctgc agttacgact	12840
attctcgtcg taacgtccgc aaaaactgtg gaattcctag aagacaacta gttgttaa	12900
tcagagcact gccttggtg aattgcta ttttaataata aaataaccct tgtttcttac	12960
ttcgtcctgg atacatctat gttttttttt tcgttaataa atgagagcat ttaagttatt	13020
gtttttaatt actttttttt agaaaacaga tttcggattt tttgtatgca ttttatttga	13080
atgtactaat ataataatt aatcaatgaa ttcatttatt taagggataa caataatcca	13140
tgaattcaca tgcacattta aaacaaaact aaattacaat aggttcatat aaaaacaaca	13200
agtatgcctt ctcaactaag aatactatag	13230

<210> 34

<211> 10458

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 34

ctaaattgta agcgtaata ttttgtaaa attcggtta aatttttggt aaatcagctc	60
attttttaac caataggccg aaatcgcaa aatcccttat aaatcaaaag aatagaccga	120
gatagggttg agtggtgttc cagtttgga caagagtcca ctattaaaga acgtggactc	180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc	240
ctaatcaagt tttttgggggt cgaggtgccg taaagcacta aatcggaacc ctaaaggag	300
ccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa	360
agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac	420
cacacccgcc gcgcttaatg cgccgtaca gggcgctcc cattcgccat tcaggctgcg	480
caactgttgg gaagggcgat cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg	540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg	600
taaaacgacg gccagtgagc gcgcctcggt cattcacgtt tttgaacccg tggaggacgg	660
gcagactcgc ggtgcaaatg tgttttacag cgtgatggag cagatgaaga tgctcgacac	720
gctgcagaac acgcagctag attaacccta gaaagataat catattgtga cgtacgttaa	780
agataatcat gcgtaaaatt gacgcatgtg ttttatcgggt ctgtatatcg aggtttat	840
attaatttga atagatatta agttttatta tatttacact tacatactaa taataaat	900
aacaaacaat ttatttatgt ttatttattt attaaaaaa aacaaaaact caaaatttct	960
tctataaagt aacaaaaact ttatcgaatt gtatagtatt cttagttgag aaggcatact	1020
tggtgttttt atatgaacct attgtaattt agttttgttt taaatgtgca tgtgaattca	1080
tggattattg ttatccctta aataaatgaa ttcattgatt aattgattat attagtacat	1140
tcaaataaaa tgcatacaaa aaatccgaaa tctgttttct aaaaaaagt aattaaac	1200
aataacttaa atgctctcat ttattaacga aaaaaaaaa atagatgtat ccaggacgaa	1260
gtaagaaaca agggttat	1320
ttatataaaa attagcaatt cacacaaggc agtgctctga	1380
atttaacaac tagttgtctt ctaggaattc cacagttttt gcggacgtta cgacgagaat	1440
agtcgtaact gcgagatgaa gcagatgaca cagaggaagc tgcacttcct gcaccttgct	1500
cgtatcccct gccagctccg taactgacgc tcttaaggct agcccaccgt agccaccttg	

accggcgccg cctgcagcag ccgcagcggc gccagcacct tggccacca gaccaccacg	1560
gcctgcaccc tgagagccta gcccaccgta gccaccttga ccggcgccgc ctgcagcagc	1620
cgcagcggcg ccagcacctt ggccaccag accaccacgg cctgcaccct gagagcctag	1680
cccaccgtag ccaccttgac cggcgcggcc tgcagcagcc gcagcggcg cagcaccttg	1740
gccaccaga ccaccacggc ctgcaccctg agagcctagc ccaccgtagc caccttgacc	1800
ggcgccgcct gcagcagccg cagcggcgcc agcaccttgg ccaccagac caccacggcc	1860
tgcaccctga gagcctagcc caccgtagcc accttgaccg gcgccgcctg cagcagccgc	1920
agcggcgcca gcaccttggc caccagacc accacggcct gcaccctgag agcctagccc	1980
accgtagcca ccttgaccgg cgccgcctgc agcagccgca gcggcgccag caccttggcc	2040
accagacca ccacggcctg caccctgaga gcctagcca ccgtagccac cttgaccggc	2100
gccgcctgca gcagccgcag cggcgccagc accttggcca ccagaccac cacggcctgc	2160
accctgagag cctagcccac cgtagccacc ttgaccggcg ccgcctgcag cagccgcagc	2220
ggcgccagca ccttggccac ccagaccacc acggcctgca ccctgagagc ctagcccacc	2280
gtagccacct tgaccggcg cgctgcagc agccgcagcg gcgccagcac cttggccacc	2340
cagaccacca cggcctgcac cctgagagcc tagcccaccg tagccacctt gaccggcgcc	2400
gcctgcagca gccgcagcgg cgccagcacc ttggccaccc agaccaccac ggccctgcacc	2460
ctgagagcct agcccaccgt agccaccttg accggcgccg cctgcagcag ccgcagcggc	2520
gccagcacct tggccacca gaccaccacg gcctgcaccc tgagagccta gcccaccgta	2580
gccaccttga ccggcgccgc ctgcagcagc cgcagcggcg ccagcacctt ggccaccag	2640
accaccacgg cctgcaccct gagagcctag cccaccgtag ccaccttgac cggcgcggcc	2700
tgcagcagcc gcagcggcg cagcaccttg gccaccaga ccaccacggc ctgcaccctg	2760
agagcctagc ccaccgtagc caccttgacc ggcgccgcct gcagcagccg cagcggcgcc	2820
agcaccttgg ccaccagac caccacggcc tgcaccctga gagcctagcc caccgtagcc	2880
accttgaccg gcgccgcctg cagcagccgc agcggcgcca gcaccttggc caccagacc	2940
accacggcct gcaccctgag agcctagccc accgtagcca ccttgaccgg cgccgcctgc	3000
agcagccgca gcggcgccag caccttggcc accagacca ccacggcctg caccctgaga	3060
gcctaggccg cccgggcccac atatgacgac tgcagcacta gtgctgaaat cgctcgata	3120
tccttgatga gtgctgtatc cactgtttga tacgtatggc ccgctctgag aatatgctcc	3180

cgcatcagca gcaacataac tttgagcaac agtaccatcg gaaagtgtct tcatgagcac	3240
atcttcctct acaatggact cgttaccgtc ggaatccgtg gttataacga acgtcttgat	3300
cattttttca ttttttccaa gtattccatg gtttttattt ttccgttgca tttttttagt	3360
tgtaatgtgt tcttcgataa ctgccccaga tgcattctcta attatttcat ctgttgatt	3420
actactttgg acagtgacat cactcccaaa atagtcctca tcaaaatcat tgatgtttgc	3480
atgtgtataa gcgacatact gaaacaaaat gttaacaatt ttttcttaaa tcaatacgat	3540
aagtgcataa atattgagct ataaatttcg cacattcaat cgaagttagc gctgatctgg	3600
aacgagttag gacatactgc cttagtagtg gtaatagtaa taattgaatg ttctagctgg	3660
ggcgttaggc aagttgaaca acattctaatt tatgtaaaca tttgtgggaa agtacataat	3720
tgtatctcat acaccacgag attttatggg cacattatgt tgttattact tgagcttggt	3780
tcgagctttg ttttccctac ctattagctg gtagcctatt ccagctacgc tccgatgggt	3840
aggtgagctc tcagactcaa cctgaaagaa tttgctaaca ctagccctaa caagagcagt	3900
gcttcataga atctatcacg tgatcggtaa cgcgaccac tgagaagatc tggcgagaaa	3960
ctcagtgggc agctcgtgtt ataatcacat accaatgtat taaaatgtaa cagaatgcta	4020
cctcgagggt atgaaaatga ttataatttg cgaaaaaag gcaatagcaa ttgttaatag	4080
atatcataat aaattaataa tgcggttata attttatgca atttctttcg tctattttct	4140
aatgatgctt tacgaattgt ttacatatt gttgaatatg cattgcatat tgcaatgctg	4200
atttaccggt gaaataggat attgcaagtc tgcccaggta ttacataga ttcatcttgc	4260
ctacttttga cgcaaataaa tcacaagtta cataatctaa ggttcatttt ttaaataaat	4320
aaatgtagtt tattccgtag ccccccggt cgtagacaaat caaatggcac aatttacgat	4380
atctgtgctt tttggcgacc acttattatc aggtggcccg tgatatcgtc tggccacctt	4440
ctgaaataat agtaaaataa ttaactcacc tgcagagcgc agcacaagat cacaaagggt	4500
ttgactctca tcttgagagt tggaaccgaa ctgatgctga tttgaaaaag ttgaaccttt	4560
ttatactgaa aattttcgag ttttgacgta ttacttttcg atagaataaa catctacaca	4620
ttgtattgta cagatacaat tacatagaaa ttaatctttt tatttaacat aacaatacaa	4680
cgtatgaagt aaattaaatt gtgatcatga acaatgtgat taattatgat ttaattgatc	4740
ttattgttct tacttttaat attttaaaga attttataat atcgttcaat ttttttgtaa	4800
tatttaaaga acttacttca tgctttctcc cgcggccgcc gaaccctaaa acattgttac	4860
gttacttgca attaagcact tattcaaact ttccgtacaa aacatctttc cgcggtgcag	4920

aagacaagcc atcgcaacgg tgttaagggc aggacttttt tttttgtttc ctacctagct	4980
gagagccttg agaggctatt tcagcgtaac cttaactagt aagtgagctc acggggctca	5040
aacctgacga cgttgctaac acgaatccta gcaagagccg tgcttcgcag aatctaccac	5100
cggatcggaa acgcgacccg ctgagaagat ccggcgagaa actcagtggg ctgtgtctga	5160
gggttaattt gctcgtcgag cccttcgtcg caagcgacgg gtccgacgag aacgatgacc	5220
ggtgcttgag gtacctaaag caccgttagt ggatcgtgag gatccgaaat gacgtgtttg	5280
gggcgacgtc gactgcttgc cattctgtcc gtaggatcgg gaatgggcgg gcaggacgcg	5340
tcaaaatcgc ccttattcag gtcccatatt ttcgatacaa taaagagcaa ccctgtagtt	5400
tgtgcgtaaa agcttacgtc tgtaactgct gttctggact aatgagagac aataaaccta	5460
tggagtgacg gcgaagtaga tttttgtcaa tatcaaacct ttagtgctct tgttttcttc	5520
tcaactaata gtttctctat gtccatcact cacgattgat tcgtggcaga aataggctag	5580
atggtggccc gacagtatgg ttaagtgaga ttacaatgtg ccctattacc agtaaatagc	5640
agttagcgat ggctgtatgc tggatgaca ctctgaggat tctgctacgt tcttggtggt	5700
tctctaagtc gccctttaac gacaccacg gaaagagatg tgaccgtagg ctattttact	5760
tagttttcta atatcaaaaa ggctgtttgt acctacctag tcaggtcata aattctgtca	5820
catgtttaat gtaaaataat tgaaacaagt ttattcatta tgtaaccatt catataccaa	5880
aatgaactta acaaaacata gattcttatg aactaaagt ttattcaaaa tgacctccgt	5940
gattttgaat acaggccttc aatctgcgcg gccagtcgtc tatcgcagca cgaacgaggt	6000
ccatgtcaat atcggcagct gccttaatca aggatgtctt gagtgactcc aaattgggat	6060
gaggctttga gcctcgacct agttctagtg ttcccacaat ggttaattcg agctcgcccg	6120
gggatctaata tcaattagag actaattcaa ttagagctaa ttcaattagg atccaagctt	6180
atcgatttcg aaccctcgac cgccggagta taaatagagg cgcttcgtct acggagcgac	6240
aattcaattc aaacaagcaa agtgaacacg tcgctaagcg aaagctaagc aaataaacia	6300
gcgcagctga acaagctaaa caatcggggg accgctagag tcgacggtac gatccaccgg	6360
tcgccaccat ggtgagcaag ggcgaggagc tgttcaccgg ggtggtgcc atcctggtcg	6420
agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggcgagggc gagggcgatg	6480
ccacctacgg caagctgacc ctgaagttca tctgcaccac cggcaagctg cccgtgccct	6540
ggcccaccct cgtgaccacc ctgacctggg gcgtgcagtg cttcagccgc taccgccacc	6600

acatgaagca	gcacgacttc	ttcaagtccg	ccatgcccga	aggctacgtc	caggagcgca	6660
ccatcttctt	caaggacgac	ggcaactaca	agacccgcgc	cgaggtgaag	ttcgagggcg	6720
acaccctggt	gaaccgcatc	gagctgaagg	gcatcgactt	caaggaggac	ggcaacatcc	6780
tggggcacia	gctggagtag	aactacatca	gccacaacgt	ctatatcacc	gccgacaagc	6840
agaagaacgg	catcaaggcc	aacttcaaga	tccgccacaa	catcgaggac	ggcagcgtgc	6900
agctcgccga	ccactaccag	cagaacaccc	ccatcggcga	cggccccgtg	ctgctgcccg	6960
acaaccacta	cctgagcacc	cagtccgccc	tgagcaaaga	ccccaacgag	aagcgcgac	7020
acatggtcct	gctggagttc	gtgaccgccc	ccgggatcac	tctcggcatt	gacgagctgt	7080
acaagtaaag	cggccgcgac	tctagatcat	aatcagccat	accacatttg	tagaggtttt	7140
acttgcttta	aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	7200
tgttggtgtt	aacttgctta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	7260
aaatttcaca	aataaagcat	tttttccact	gcattctagt	tgtggtttgt	ccaaactcat	7320
caatgtatct	taaagcttat	cgatacgctg	acggcgcgcc	taggccggcc	gatactagag	7380
cggccgccac	cgcggtggag	ctccagcttt	tgttcccttt	agtgagggtt	aattagatct	7440
taatacgact	cactataggg	cgaattgggt	accgggcccc	ccctcgaggt	cgacgggtatc	7500
gataagcttg	atatctataa	caagaaaata	tatatataat	aagttatcac	gtaagtagaa	7560
catgaaataa	caatataatt	atcgtatgag	ttaaattctta	aaagtcacgt	aaaagataat	7620
catgcgtcat	tttgactcac	gcggtcgtta	tagttcaaaa	tcagtgacac	ttaccgcatt	7680
gacaagcacg	cctcacggga	gctccaagcg	gcgactgaga	tgtcctaaat	gcacagcgac	7740
ggattcgcgc	tatttagaaa	gagagagcaa	tatttcaaga	atgcatgcgt	caattttacg	7800
cagactatct	ttctaggggt	aatctagctg	catcaggatc	atatcgtcgg	gtcttttttc	7860
cggctcagtc	atcgcccaag	ctggcgctat	ctgggcatcg	gggaggaaga	agcccggtgcc	7920
ttttcccgcg	aggttgaagc	ggcatggaaa	gagtttgccg	aggatgactg	ctgctgcatt	7980
gacgttgagc	gaaaacgcac	gtttaccatg	atgattcggg	aaggtgtggc	catgcacgcc	8040
tttaacggtg	aactgttcgt	tcaggccacc	tgggatacca	gttcgtcgcg	gcttttccgg	8100
acacagttcc	ggatgggtcag	cccgaagcgc	atcagcaacc	cgaacaatac	cggcgacagc	8160
cggaaactgcc	gtgccgggtgt	gcagattaat	gacagcgggtg	cggcgctggg	atattacgtc	8220
agcgaggacg	ggtatcctgg	ctggatgccg	cagaaatgga	catggatacc	ccgtgagtta	8280
cccggcgggc	gcgcttgggc	taatcatggt	catagctgtt	tcctgtgtga	aattgttatc	8340

cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc tgggggtgcct	8400
aatgagtgag ctaactcaca ttaattgCGt tgcgctcact gcccgctttc cagtcgggaa	8460
acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta	8520
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcggt cggtgcggc	8580
gagcggatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg	8640
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt	8700
tgctggcggt tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa	8760
gtcagagggt gcgaaacccg acaggactat aaagatacca ggcgtttccc cctggaagct	8820
ccctcgTgcg ctctcctggt ccgaccctgc cgcttaccg atacctgtcc gcctttctcc	8880
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg	8940
tcgttcgctc caagctgggc tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct	9000
tatccggtaa ctatcgtctt gagtccaacc cggtaagaca cgacttatcg ccaactggcag	9060
cagccactgg taacaggatt agcagagcga ggtatgtagg cggtgctaca gagttcttga	9120
agtgggtggc taactacggc tacactagaa ggacagtatt tggatatctgc gctctgctga	9180
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaaca accaccgctg	9240
gtagcgggtg tttttttggt tgcaagcagc agattacgcg cagaaaaaaaa ggatctcaag	9300
aagatccttt gatcttttct acgggggtctg acgctcagtg gaacgaaaac tcacgttaag	9360
ggattttggT catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaaat	9420
gaagtttta atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct	9480
taatcagtga ggcacctatc tcagcgatct gtctatctcg ttcattcata gttgcctgac	9540
tccccgtcgt gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa	9600
tgataccgcg agaccacgc tcaccggctc cagatttatc agcaataaac cagccagccg	9660
gaagggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt	9720
gttgccggga agctagagta agtagttcgc cagttaatag tttgcgcaac gttgttgcca	9780
ttgctacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt	9840
cccaacgatc aaggcgagtt acatgatccc ccatgttgTg caaaaaagcg gttagctcct	9900
tcggtcctcc gatcgttgTc agaagtaagt tggccgcagt gttatcactc atggttatgg	9960
cagcactgca taattctctt actgtcatgc catccgtaag atgcttttct gtgactggTg	10020

```

agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg 10080
cgtcaatacg ggataatacc gcgccacata gcagaacttt aaaagtgctc atcattggaa 10140
aacgttcttc ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt 10200
aaccactcg tgcaccaaac tgatcttcag catcttttac tttcaccagc gtttctgggt 10260
gagcaaaaac aggaaggcaa aatgccgcaa aaaaggggaat aagggcgaca cggaatggt 10320
gaatactcat actcttcctt tttcaatatt attgaagcat ttatcagggt tattgtctca 10380
tgagcggata catatttgaa tgtatttaga aaaataaaca aatagggggt ccgcgcacat 10440
ttccccgaaa agtgccac 10458

```

<210> 35

<211> 11250

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Mô tả trình tự nhân tạo: Polynucleotit tổng hợp

<400> 35

```

ctaaattgta agcgttaata ttttgttaaa attcgcgtta aatttttgtt aaatcagctc 60
attttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga 120
gataggggtg agtggtgttc cagtttgga caagagtcca ctattaaaga acgtggactc 180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc 240
ctaatcaagt tttttggggg cgaggtgccg taaagcacta aatcggaacc ctaaaggag 300
ccccgatatt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa 360
agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac 420
cacacccgcc gcgcttaatg cgccgctaca gggcgcgctc cattcgccat tcaggctgcg 480
caactgttgg gaagggcgat cggcgcgggc ctcttcgcta ttacgccagc tggcgaaagg 540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg 600
taaaacgacg gccagtgagc gcgcctcggt cattcacgtt tttgaacccg tggaggacgg 660
gcagactcgc ggtgcaaatt tgttttacag cgtgatggag cagatgaaga tgctcgacac 720
gctgcagaac acgcagctag attaacccta gaaagataat catattgtga cgtacgttaa 780
agataatcat gcgtaaaatt gacgcatgtg ttttatcggg ctgtatatcg aggtttat 840
attaatttga atagatatta agttttatta tatttacact tacatactaa taataaat 900

```

aacaaacaat ttatttatgt ttatttattt attaaaaaaa aacaaaaaact caaaattttct	960
tctataaagt aacaaaaactt ttatcgaatt gtatagtatt cttagttgag aaggcatact	1020
tgttggtttt atatgaacct attgtaattt agttttgttt taaatgtgca tgtgaattca	1080
tggattattg ttatccctta aataaatgaa ttcattgatt aattgattat attagtacat	1140
tcaaataaaa tgcatacaaa aaatccgaaa tctgttttct aaaaaaaagt aattaaaaac	1200
aataacttaa atgctctcat ttattaacga aaaaaaaaac atagatgtat ccaggacgaa	1260
gtaagaaaca agggttattt tatattaaaa attagcaatt cacacaaggc agtgctctga	1320
atttaacaac tagttgtctt ctaggaattc cacagttttt gcggacgtta cgacgagaat	1380
agtcgtaact gcgagatgaa gcagatgaca cagaggaagc tgcacttcct gcaccttgct	1440
cgtatcccct gccagctccg taactgacgc tcttaaggct agcccaccgt agccaccttg	1500
accggcgccg cctgcagcag ccgcagcggc gccagcacct tggccaccca gaccaccacg	1560
gcctgcaccc tgagagccta gcccaccgta gccaccttga ccggcgccgc ctgcagcagc	1620
cgcagcggcg ccagcacctt ggccaccag accaccacgg cctgcaccct gagagcctag	1680
cccaccgtag ccaccttgac cggcgccgcc tgcagcagcc gcagcggcg cagcaccttg	1740
gccaccacga ccaccacggc ctgcaccctg agagcctagc ccaccgtagc caccttgacc	1800
ggcgccgcct gcagcagccg cagcggcgcc agcaccttg ccaccagac caccacggcc	1860
tgcaccctga gagcctagcc caccgtagcc accttgaccg gcgccgcctg cagcagccgc	1920
agcggcgcca gcaccttggc caccagacc accacggcct gcaccctgag agcctagccc	1980
accgtagcca ccttgaccgg cgccgcctgc agcagccgca gcggcgccag caccttggcc	2040
accagacca ccacggcctg caccctgaga gcctagccca ccgtagccac cttgaccggc	2100
gccgcctgca gcagccgcag cggcgccagc accttggcca ccagaccac cacggcctgc	2160
accctgagag cctagcccac cgtagccacc ttgaccggcg ccgcctgcag cagccgcagc	2220
ggcgccagca ccttggccac ccagaccacc acggcctgca ccctgagagc ctagcccacc	2280
gtagccacct tgaccggcg cgctgcagc agccgcagcg gcgccagcac cttggccacc	2340
cagaccacca cggcctgcac cctgagagcc tagcccaccg tagccacctt gaccggcgcc	2400
gcctgcagca gccgcagcgg cgccagcacc ttggccaccc agaccaccac ggctgcacc	2460
ctgagagcct agcccaccgt agccaccttg accggcgccg cctgcagcag ccgcagcggc	2520
gccagcacct tggccaccca gaccaccacg gcctgcaccc tgagagccta gcccaccgta	2580
gccaccttga ccggcgccgc ctgcagcagc cgcagcggcg ccagcacctt ggccaccag	2640

accaccacgg	cctgcaccct	gagagcctag	cccaccgtag	ccaccttgac	cggcgccgcc	2700
tgcagcagcc	gcagcggcgc	cagcaccttg	gccacccaga	ccaccacggc	ctgcaccctg	2760
agagcctagc	ccaccgtagc	caccttgacc	ggcgccgcct	gcagcagccg	cagcggcgcc	2820
agcaccttgg	ccaccagac	caccacggcc	tgcaccctga	gagcctagcc	caccgtagcc	2880
accttgaccg	gcgccgcctg	cagcagccgc	agcggcgcca	gcaccttggc	caccagacc	2940
accacggcct	gcaccctgag	agcctagccc	accgtagcca	ccttgaccgg	cgccgcctgc	3000
agcagccgca	gcggcgccag	caccttggcc	accagacca	ccacggcctg	caccctgaga	3060
gcctagccca	ccgtagccac	cttgaccggc	gccgcctgca	gcagccgcag	cggcgccagc	3120
accttggcca	cccagaccac	cacggcctgc	accctgagag	cctagcccac	cgtagccacc	3180
ttgaccggcg	ccgcctgcag	cagccgcagc	ggcgccagca	ccttggccac	ccagaccacc	3240
acggcctgca	ccctgagagc	ctagcccacc	gtagccacct	tgaccggcgc	cgctgcagc	3300
agccgcagcg	gcgccagcac	cttggccacc	cagaccacca	cggcctgcac	cctgagagcc	3360
tagcccaccg	tagccacctt	gaccggcgcc	gcctgcagca	gccgcagcgg	cgccagcacc	3420
ttggccaccc	agaccaccac	ggcctgcacc	ctgagagcct	agcccaccgt	agccaccttg	3480
accggcgccg	cctgcagcag	ccgcagcggc	gccagcacct	tggccaccca	gaccaccacg	3540
gcctgcaccc	tgagagccta	gccaccgta	gccaccttga	ccggcgccgc	ctgcagcagc	3600
cgcagcggcg	ccagcacctt	ggccaccag	accaccacgg	cctgcaccct	gagagcctag	3660
cccaccgtag	ccaccttgac	cggcgccgcc	tgcagcagcc	gcagcggcgc	cagcaccttg	3720
gccacccaga	ccaccacggc	ctgcaccctg	agagcctagc	ccaccgtagc	caccttgacc	3780
ggcgccgcct	gcagcagccg	cagcggcgcc	agcaccttgg	ccaccagac	caccacggcc	3840
tgcaccctga	gagcctaggc	cggccgggcc	acatatgacg	actgcagcac	tagtgctgaa	3900
atcgctcgta	tatccttgat	gagtgtgtga	tccactgttt	gatacgtatg	gcccgtcttg	3960
agaatatgct	ccgcatacag	cagcaacata	actttgagca	acagtaccat	cggaaagtgt	4020
cttcatgagc	acatcttcct	ctacaatgga	ctcgttaccg	tcggaatccg	tggttataac	4080
gaacgtcttg	atcatttttt	cattttttcc	aagtattcca	tggtttttat	ttttccgttg	4140
cattttttta	gttgtaattt	gttcttcgat	aactgcccc	gatgcatctc	taattatttc	4200
atctgttgta	ttactacttt	ggacagtgac	atcactccca	aaatagtcct	catcaaaatc	4260
attgatgttt	gcatttgtat	aagcgacata	ctgaaacaaa	atgttaacaa	ttttttctta	4320

aatcaatacg	ataagtgc	ataaattgag	ctataaattt	cgacacattca	atcgaagtta	4380
gcgctgatct	ggaacgagtt	aggacatact	gccttagtag	tggtaatagt	aataattgaa	4440
tgttctagct	ggggcgttag	gcaagttgaa	caacattcta	attatgtaaa	catttgtggg	4500
aaagtacata	attgtatctc	atacaccacg	agatttttatg	gtcacattat	gttgttatta	4560
cttgagcttg	tttcgagctt	tgttttccct	acctattagc	tggtagccta	ttccagctac	4620
gctccgatgg	gtaggtgagc	tctcagactc	aacctgaaag	aatttgctaa	cactagccct	4680
aacaagagca	gtgcttcata	gaatctatca	cgtgatcgg	aacgcgaccc	actgagaaga	4740
tctggcgaga	aactcagtgg	gcagctcgtg	ttataatcac	ataccaatgt	attaaaatgt	4800
aacagaatgc	tacctcgagg	ttatgaaaat	gattataatt	tgcgaaaaaa	aggcaatagc	4860
aattgttaat	agatatcata	ataaattaat	aatgcgggta	taatttttatg	caattttctt	4920
cgtctatttt	ctaatgatgc	tttacgaatt	gttttacata	ttgttgaata	tgcatcgc	4980
attgcaatgc	tgattttaccg	gtgaaatagg	atattgcaag	tctgcccagg	tatttacata	5040
gattcatctt	gcctactttt	gacgcaaata	aatcacaagt	tacataatct	aaggttcatt	5100
ttttaataaa	ataaatgtag	tttattccgt	agcccccccg	ggcgtgacaa	atcaaaggc	5160
acaatttacg	atatctgtgc	gttttggcga	ccacttatta	tcagggtggcc	cgtgatatcg	5220
tctggccacc	ttctgaaata	atagtaaaat	aattaactca	cctgcagagc	gcagcacaag	5280
atcacaagg	ttttgactct	catcttgaga	gttggaaaccg	aactgatgct	gatttgaaaa	5340
agttgaacct	ttttatactg	aaaattttcg	agttttgacg	tattttacttt	cgatagaata	5400
aacatctaca	cattgtattg	tacagataca	attacataga	aattaatctt	tttatttaac	5460
ataacaatac	aacgtatgaa	gtaaattaaa	ttgtgatcat	gaacaatgtg	attaattatg	5520
atttaattga	tcttattggt	cttactttta	atattttaaa	gaattttata	atatcgttca	5580
atTTTTTgt	aatattttaa	gaacttactt	catgctttct	cccgcggccg	ccgaacccta	5640
aaacattggt	acgttacttg	caattaagca	cttattcaaa	ctttccgtac	aaaacatctt	5700
tccgcggtgc	agaagacaag	ccatcgcaac	ggtgttaagg	gcaggacttt	TTTTTgtt	5760
tcctacctag	ctgagagcct	tgagaggcta	tttcagcgta	accttaacta	gtaagtgagc	5820
tcacggggct	caaacctgac	gacgttgcta	acacgaatcc	tagcaagagc	cgtgcttcgc	5880
agaatctacc	accggatcgg	aaacgcgacc	cgtcgagaag	atccggcgag	aaactcagtg	5940
ggctgtgtct	gagggttaat	ttgctcgtcg	agcccttcgt	cgcaagcgac	gggtccgacg	6000
agaacgatga	ccggtgcttg	aggtacctaa	agcaccgtta	gtggatcgtg	aggatccgaa	6060

atgacgtggt	tggggcgacg	tcgactgctt	gccattctgt	ccgtaggata	gggaatgggc	6120
gggcaggacg	cgtcaaaatc	gcccttattc	aggtcccata	ttttcgatac	aataaagagc	6180
aaccctgtag	tttgtgcgta	aaagcttacg	tctgtaactg	ctgttctgga	ctaatagagag	6240
acaataaacc	tatggagtga	cggcgaagta	gattttttgtc	aatatcaaac	ctttagtgtc	6300
cttgtttttct	tctcaactaa	tagtttctct	atgtccatca	ctcacgattg	attcgtggca	6360
gaaataggct	agatgggtggc	ccgacagtat	ggttaagtga	gattacaatg	tgccctatta	6420
ccagtaaata	gcagtttagcg	atggctgtat	gctggtatga	cactctgagg	attctgctac	6480
gttcttggtg	gttctctaag	tcgcccttta	acgacacca	cggaaagaga	tgtgaccgta	6540
ggctattttta	cttagttttc	taatatcaaa	aaggctgttt	gtacctacct	agtcagggtca	6600
taaattctgt	cacatgttta	atgtaaaata	attgaaacaa	gtttattcat	tatgtaacca	6660
ttcatatacc	aaaatgaact	taacaaaaca	tagattctta	tgacactaaa	gtttattcaa	6720
aatgacctcc	gtgattttga	atacaggcct	tcaatctgcg	cggccagtcg	tctatcgag	6780
cacgaacgag	gtccatgtca	atatcggcag	ctgccttaat	caaggatgtc	ttgagtgtc	6840
ccaaattggg	atgaggcttt	gagcctcgac	ctagttctag	tgttcccaca	atggttaatt	6900
cgagctcgcc	cggggatcta	attcaattag	agactaattc	aattagagct	aattcaatta	6960
ggatccaagc	ttatcgattt	cgaaccctcg	accgccggag	tataaataga	ggcgcttcgt	7020
ctacggagcg	acaattcaat	tcaaacaagc	aaagtgaaca	cgtcgctaag	cgaaagctaa	7080
gcaaataaac	aagcgcagct	gaacaagcta	aacaatcggg	gtaccgctag	agtcgacggt	7140
acgatccacc	ggtcgccacc	atggtgagca	agggcgagga	gctgttcacc	ggggtggtgc	7200
ccatcctggt	cgagctggac	ggcgacgtaa	acggccacaa	gttcagcgtg	tccggcgagg	7260
gcgagggcga	tgccacctac	ggcaagctga	ccctgaagtt	catctgcacc	accggcaagc	7320
tgcccgtgcc	ctggcccacc	ctcgtgacca	ccctgacctg	ggcgctgcag	tgcttcagcc	7380
gctaccccga	ccacatgaag	cagcacgact	tcttcaagtc	cgccatgccc	gaaggctacg	7440
tccaggagcg	caccatcttc	ttcaaggacg	acggcaacta	caagaccgcg	gccgaggtga	7500
agttcgaggg	cgacaccctg	gtgaaccgca	tcgagctgaa	gggcatcgac	ttcaaggagg	7560
acggcaacat	cctggggcac	aagctggagt	acaactacat	cagccacaac	gtctatatca	7620
ccgccgacaa	gcagaagaac	ggcatcaagg	ccaacttcaa	gatccgccac	aacatcgagg	7680
acggcagcgt	gcagctcgcc	gaccactacc	agcagaacac	ccccatcggc	gacggccccg	7740

tgctgctgcc cgacaaccac tacctgagca cccagtccgc cctgagcaaa gaccccaacg	7800
agaagcgcgga tcacatgggtc ctgctggagt tcgtgaccgc cgccgggatc actctcggga	7860
tggacgagct gtacaagtaa agcggccgcg actctagatc ataatcagcc ataccacatt	7920
tgtagagggtt ttacttgctt taaaaaacct cccacacctc cccctgaacc tgaacataa	7980
aatgaatgca attggtgttg ttaacttggt tattgcagct tataatgggtt acaaataaag	8040
caatagcatc acaaatttca caaataaagc atttttttca ctgcattcta gttgtggttt	8100
gtccaaactc atcaatgtat cttaaagctt atcgatacgc gtacggcgcg cctaggccgg	8160
ccgatactag agcggccgcc accgcggtgg agctccagct tttgttccct ttagtgaggg	8220
ttaattagat cttaatacga ctactatag ggcgaattgg gtaccgggcc cccctcgag	8280
gtcgacggta tcgataagct tgatatctat aacaagaaaa tatatatata ataagttatc	8340
acgtaagtag aacatgaaat aacaataaa ttatcgtatg agttaaatct taaaagtcac	8400
gtaaaagata atcatgcgtc attttgactc acgcggtcgt tatagttcaa aatcagtgac	8460
acttaccgca ttgacaagca cgcctcacgg gagctccaag cggcgactga gatgtcctaa	8520
atgcacagcg acggattcgc gctatttaga aagagagagc aatatttcaa gaatgcatgc	8580
gtcaatttta cgcagactat ctttctaggg ttaatctagc tgcatacagga tcatatcgtc	8640
gggtcttttt tccggctcag tcatcgcca agctggcgct atctgggcat cggggaggaa	8700
gaagcccgtg ctttttcccg cgaggttgaa gcggcatgga aagagtttgc cgaggatgac	8760
tgctgctgca ttgacgttga gcgaaaacgc acgtttacca tgatgattcg ggaagggtgtg	8820
gccatgcacg cttttaacgg tgaactgttc gttcaggcca cctgggatac cagttcgtcg	8880
cggcttttcc ggacacagtt ccggtggtc agcccgaagc gcatacagca cccgaacaat	8940
accggcgaca gccggaactg ccgtgccggt gtgcagatta atgacagcgg tcgggcgctg	9000
ggatattacg tcagcgagga cgggtatcct ggctggatgc cgcagaaatg gacatggata	9060
ccccgtgagt taccggcgcg gcgcgcttgg cgtaatcatg gtcatagctg tttcctgtgt	9120
gaaattgtta tccgctcaca attccacaca acatacgagc cggaagcata aagtgtaaag	9180
cctgggggtgc ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgtt	9240
tccagtcggg aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag	9300
gcggtttgcg tattgggcgc tcttccgctt cctcgctcac tgactcgctg cgctcggtcg	9360
ttcggctgcg gcgagcggta tcagctcact caaaggcggg aatacggtta tccacagaat	9420
caggggataa cgcaggaaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta	9480

aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa	9540
atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc	9600
cccctggaag ctccctcgtg cgctctcctg ttccgaccct gccgcttacc ggatacctgt	9660
ccgcctttct cccttcggga agcgtggcgc tttctcatag ctacgctgt aggtatctca	9720
gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg	9780
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggtaaga cacgacttat	9840
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta	9900
cagagttctt gaagtgggtg cctaactacg gctacactag aaggacagta tttggtatct	9960
gcgctctgct gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac	10020
aaaccaccgc tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa	10080
aaggatctca agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaacgaaa	10140
actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt	10200
taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact tgggtctgaca	10260
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca	10320
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggtta ccatctggcc	10380
ccagtgcctg aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa	10440
accagccagc cggaagggcc gagcgcagaa gtggctcctg aactttatcc gcctccatcc	10500
agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat agtttgcgca	10560
acgttggtgc cattgctaca ggcatcgtg tgtcacgctc gtcgtttggg atggcttcat	10620
tcagctccgg ttcccaacga tcaaggcgag ttacatgata ccccatgttg tgcaaaaaag	10680
cggttagctc cttcggtcct ccgatcgttg tcagaagtaa gttggccgca gtgttatcac	10740
tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt	10800
ctgtgactgg tgagtactca accaagtcatt tctgagaata gtgtatgcgg cgaccgagtt	10860
gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc	10920
tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat	10980
ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt actttcacca	11040
gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca	11100
cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc atttatcagg	11160

26612

gttattgtct catgagcggg tacatatttg aatgtattta gaaaaataaa caaatagggg	11220
ttccgcgcac atttccccga aaagtgccac	11250