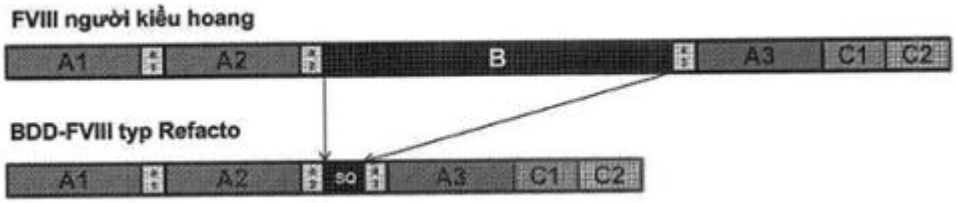
(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ
(11) **1-0041851**
(51)[7] **C07K 14/755**; C12N 15/86; A61K 48/00 (13) **B**

(21) 1-2018-02468 (22) 11/11/2016
(86) PCT/US2016/061684 11/11/2016 (87) WO 2017/083762 18/05/2017
(30) 62/255,317 13/11/2015 US
(45) 25/12/2024 441 (43) 25/03/2019 372A
(73) Takeda Pharmaceutical Company Limited (JP)
1-1, Doshomachi 4-chome, Chuo-ku, Osaka-shi, Osaka, Japan
(72) FALKNER, Falko-Günter (DE); HORLING, Franziska (DE); LENGLER, Johannes (DE); ROTTENSTEINER, Hanspeter (AT); SCHEIFLINGER, Friedrich (AT).
(74) Công ty TNHH Tầm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) POLYNUCLEOTIT ĐƯỢC BIẾN ĐỔI CODON MÃ HÓA CÁC BIẾN THỂ FVIII TÁI TỔ HỢP CÓ SỰ BIỂU HIỆN TĂNG CHO LIỆU PHÁP GEN ĐỂ ĐIỀU TRỊ BỆNH HEMOPHILIA A, VECTƠ VÀ HẠT VIRUT LIÊN HỢP ADENO CHỨA POLYNUCLEOTIT NÀY, QUY TRÌNH SẢN SINH HẠT VIRUT LIÊN HỢP ADENO, TẾ BÀO CHỦ ĐƯỢC GÂY NHIỄM BẰNG HẠT VIRUT NÀY, VÀ QUY TRÌNH TẢI NẠP TẾ BÀO CHỦ

(57) Sáng chế đề xuất, ngoài các khía cạnh khác, polynucleotit được biến đổi codon mã hóa các biến thể Yếu tố VIII để biểu hiện trong các tế bào động vật có vú. Trong một số phương án, sáng chế cũng đề xuất các vectơ và hạt virut liên hợp adeno chứa polynucleotit này, quy trình sản sinh hạt virut liên hợp adeno (AAV) này. Ngoài ra, sáng chế cũng đề xuất tế bào chủ được gây nhiễm bằng hạt virut liên hợp adeno (AAV) nói trên và quy trình tải nạp tế bào chủ.



Lĩnh vực kỹ thuật kỹ thuật được đề cập

Sáng chế đề cập đến, ngoài các khía cạnh khác, polynucleotit được biến đổi codon mã hóa các biến thể Yếu tố VIII để biểu hiện trong các tế bào động vật có vú. Trong một số phương án, sáng chế cũng đề xuất các vectơ và hạt virut liên hợp adeno chứa polynucleotit này, quy trình sản sinh hạt virut liên hợp adeno (AAV) này. Ngoài ra, sáng chế cũng đề xuất tế bào chủ được gây nhiễm bằng hạt virut liên hợp adeno (AAV) nói trên và quy trình tải nạp tế bào chủ.

Tình trạng kỹ thuật của sáng chế

Quá trình đông máu trải qua con đường sinh động học và phức tạp bao gồm các phản ứng hóa sinh phụ thuộc lẫn nhau, còn được gọi là chuỗi đông máu. Yếu tố đông máu VIII (FVIII) là thành phần chính trong chuỗi này. Yếu tố VIII được huy động đến vị trí chảy máu, và tạo ra phức hợp Xase với Yếu tố IX hoạt hóa (FIXa) và Yếu tố X (FX). Phức hợp Xase hoạt hóa FX, yếu tố này lại hoạt hóa prothrombin thành thrombin, sau đó thrombin hoạt hóa các thành phần khác trong chuỗi đông máu để tạo ra cục máu đông ổn định (xem lại trong Saenko et al., Trends Cardiovasc. Med., 9:185-192 (1999); Lenting et al., Blood, 92:3983-3996 (1998)).

Hemophilia A là rối loạn chảy máu liên quan đến nhiễm sắc thể X bẩm sinh được đặc trưng bởi sự thiếu hoạt tính của Yếu tố VIII. Hoạt tính Yếu tố VIII bị giảm sẽ ức chế vòng phản hồi dương tính trong chuỗi đông máu. Điều này gây ra sự đông máu không hoàn toàn, mà biểu hiện thành các giai đoạn chảy máu với thời gian kéo dài, vết thâm tím lan rộng, chảy máu miệng và mũi tự phát, cứng khớp và đau mãn tính, và có thể chảy máu trong và thiếu máu trong các trường hợp nghiêm trọng (Zhang et al., Clinic. Rev. Allerg. Immunol., 37:114-124 (2009)).

Thông thường, bệnh hemophilia A được điều trị bằng liệu pháp thay thế Yếu tố VIII,

bao gồm sử dụng protein Yếu tố VIII (ví dụ, Yếu tố VIII được sản xuất tái tổ hợp hoặc thu được từ huyết tương) cho đối tượng mắc bệnh hemophilia A. Yếu tố VIII được sử dụng với mục đích phòng bệnh để ngăn ngừa hoặc làm giảm tần số của các giai đoạn chảy máu, đáp ứng lại giai đoạn chảy máu cấp tính, và/hoặc xung quanh thời gian phẫu thuật để kiểm soát sự chảy máu trong quá trình phẫu thuật. Tuy nhiên, liệu pháp thay thế Yếu tố VIII có một số tác dụng không mong muốn.

Thứ nhất, liệu pháp thay thế Yếu tố VIII được sử dụng để điều trị hoặc kiểm soát bệnh hemophilia A, nhưng không chữa được sự thiếu hụt Yếu tố VIII cơ bản. Do đó, các đối tượng mắc bệnh hemophilia A cần liệu pháp thay thế Yếu tố VIII suốt đời. Việc điều trị liên tục là tốn kém và đòi hỏi đối tượng phải duy trì sự tuân thủ nghiêm ngặt, vì việc bỏ qua chỉ vài liều phòng bệnh có thể gây ra các hậu quả nghiêm trọng cho đối tượng mắc bệnh hemophilia A nặng.

Thứ hai, do Yếu tố VIII có thời gian bán thải tương đối ngắn trên in vivo, liệu pháp thay thế Yếu tố VIII phòng bệnh thông thường đòi hỏi phải sử dụng mỗi hai ngày một lần hoặc ba ngày một lần. Điều này tạo ra gánh nặng cho đối tượng để duy trì sự tuân thủ trong suốt cuộc đời. Mặc dù các thuốc Yếu tố VIII "tác dụng kéo dài" thế hệ ba có thể làm giảm tần số sử dụng, liệu pháp thay thế Yếu tố VIII phòng bệnh với các thuốc này vẫn đòi hỏi phải sử dụng hàng tháng, hàng tuần hoặc thường xuyên hơn một cách liên tục. Ví dụ, việc điều trị phòng ngừa bằng ELOCTATE™ [Yếu tố chống chảy máu (tái tổ hợp), protein dung hợp Fc] đòi hỏi phải sử dụng mỗi ba đến năm ngày một lần (Thông tin kê đơn ELOCTATE™, Biogen Idec Inc., (2015)). Ngoài ra, tác dụng lâu dài của các chế phẩm sinh học được biến đổi hóa học (ví dụ, polypeptit được pegyl hóa) vẫn chưa được hiểu đầy đủ.

Thứ ba, từ 15% đến 30% đối tượng nhận được liệu pháp thay thế Yếu tố VIII tạo ra các kháng thể ức chế kháng Yếu tố VIII, làm cho liệu pháp này trở nên không hiệu quả. Liệu pháp bỏ qua Yếu tố VIII (ví dụ, sử dụng dịch cô đặc phức hợp prothrombin được sản sinh tái tổ hợp hoặc thu được từ huyết tương) có thể được sử dụng để điều trị bệnh hemophilia ở đối tượng đã sinh ra kháng thể ức chế. Tuy nhiên, liệu pháp bỏ qua Yếu tố VIII có hiệu quả thấp hơn liệu pháp thay thế Yếu tố VIII (Mannucci P.M., J Thromb Haemost., 1(7):1349-55 (2003)) và có thể kèm theo tăng nguy cơ về biến chứng tim mạch (Luu và Ewenstein, Haemophilia, 10 Suppl. 2:10-16 (2004)).

Liệu pháp gen dòng tế bào sinh dưỡng có triển vọng lớn trong việc điều trị bệnh hemophilia A vì nó sửa chữa được tình trạng hoạt tính Yếu tố VIII chức năng biểu hiện thấp cơ bản (ví dụ, do các đột biến sai nghĩa hoặc vô nghĩa), hơn là chỉ cung cấp liều một lần của hoạt tính Yếu tố VIII cho đối tượng này. Do sự khác biệt về cơ chế tác dụng này, so với liệu pháp thay thế Yếu tố VIII, việc sử dụng một lần vectơ liệu pháp gen Yếu tố VIII có thể cung cấp cho đối tượng Yếu tố VIII trong vài năm, làm giảm chi phí điều trị và giảm nhu cầu phải tuân thủ kiên trì liên tục.

Liệu pháp gen Yếu tố đông máu IX (FIX) đã được sử dụng hiệu quả để điều trị cho các đối tượng mắc bệnh hemophilia B, tình trạng đông máu liên quan được đặc trưng bởi hoạt tính Yếu tố IX giảm (Manno C.S., et al., Nat Med., 12(3):342-47 (2006)). Tuy nhiên, liệu pháp gen Yếu tố VIII thể hiện một số thách thức khác biệt. Ví dụ, polypeptit Yếu tố VIII kiểu hoang, có độ dài đầy đủ (2351 axit amin; số truy cập UniProt P00451) lớn hơn năm lần so với polypeptit Yếu tố IX kiểu hoang, có độ dài đầy đủ (461 axit amin; số truy cập UniProt P00740). Do đó, trình tự mã hóa của Yếu tố VIII kiểu hoang là 7053 cặp bazơ, là quá lớn để đóng gói được trong các vectơ liệu pháp gen AAV thông thường. Ngoài ra, sự biểu hiện tái tổ hợp của các biến thể bị mất vùng chức năng B của Yếu tố VIII (BDD-FVIII) đã được báo cáo là yếu. Do đó, một số nhóm đã nỗ lực thay đổi xu hướng sử dụng codon của các cấu trúc BDD-FVIII, với thành công hạn chế.

Bản chất kỹ thuật của sáng chế

Do đó, có nhu cầu về các biến thể Yếu tố VIII mà có các trình tự mã hóa được đóng gói hiệu quả hơn vào, và được phân phối qua, các vectơ liệu pháp gen. Cũng có nhu cầu về các axit nucleic tổng hợp, được biến đổi codon, mà biểu hiện Yếu tố VIII hiệu quả hơn. Các biến thể Yếu tố VIII và axit nucleic được biến đổi codon như vậy cho phép cải thiện việc điều trị các thiếu hụt Yếu tố VIII (ví dụ, bệnh hemophilia A). Các thiếu hụt trên đây và các vấn đề khác liên quan đến việc điều trị các thiếu hụt Yếu tố VIII (ví dụ, bệnh hemophilia A) được giảm đi hoặc loại bỏ nhờ các biến thể Yếu tố VIII được biến đổi codon đã được mô tả này.

Theo một số phương án, sáng chế đề xuất các axit nucleic mã hóa các biến thể Yếu tố VIII mà có sự đồng nhất trình tự cao với các trình tự được biến đổi codon đã được mô tả của chuỗi nặng (ví dụ, CS01-HC-NA, CS04-HC-NA, hoặc CS23-HC-NA) và chuỗi nhẹ

Yếu tố VIII (CS01-LC-NA, CS04-LC-NA, hoặc CS23-LC-NA). Trong một số phương án, các axit nucleic này bao gồm thêm trình tự mã hóa trình tự cầu nối mà thay thế vùng chức năng B Yếu tố VIII tự nhiên (ví dụ, các trình tự cầu nối chứa vị trí cắt furin), giữa các trình tự mã hóa chuỗi nặng và chuỗi nhẹ Yếu tố VIII.

Theo một khía cạnh, sáng chế đề xuất polynucleotit bao gồm trình tự nucleotit mã hóa polypeptit Yếu tố VIII. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất ít nhất là 95% với CS04-HC-NA (SEQ ID NO: 3). Chuỗi nhẹ của polypeptit Yếu tố FVIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất ít nhất là 95% với CS04-LC-NA (SEQ ID NO: 4). Cầu nối polypeptit này chứa vị trí cắt furin.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit này được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO04 (SEQ ID NO: 6).

Theo một khía cạnh, sáng chế đề xuất polynucleotit bao gồm trình tự nucleotit mã hóa polypeptit Yếu tố VIII. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất ít nhất là 95% với CS01-HC-NA (SEQ ID NO: 24). Chuỗi nhẹ của polypeptit Yếu tố FVIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất ít nhất là 95% với CS01-LC-NA (SEQ ID NO: 25). Cầu nối polypeptit này chứa vị trí cắt furin.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit này được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO01 (SEQ ID NO: 5).

Theo một khía cạnh, sáng chế đề xuất polynucleotit bao gồm trình tự nucleotit mã hóa polypeptit Yếu tố VIII. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất ít nhất là 95% với CS23-HC-NA (SEQ ID NO: 22). Chuỗi nhẹ của polypeptit Yếu tố FVIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất ít nhất là 95% với CS23-LC-

NA (SEQ ID NO: 23). Cầu nối polypeptit này chứa vị trí cắt furin.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit này được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO23 (SEQ ID NO: 7).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 96% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 96% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 97% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 97% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 98% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 98% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 99% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 99% với trình tự chuỗi

nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 99,5% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 99,5% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII có sự đồng nhất ít nhất là 99,9% với trình tự chuỗi nặng tương ứng (ví dụ, CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), hoặc CS23-HC-NA (SEQ ID NO: 22)), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII có sự đồng nhất ít nhất là 99,9% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), hoặc CS23-LC-NA (SEQ ID NO: 23)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII là CS04-HC-NA (SEQ ID NO: 3), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII là CS04-LC-NA (SEQ ID NO: 4).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII là CS01-HC-NA (SEQ ID NO: 24), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII là CS01-LC-NA (SEQ ID NO: 25).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit thứ nhất mã hóa chuỗi nặng của polypeptit Yếu tố VIII là CS23-HC-NA (SEQ ID NO: 22), và trình tự nucleotit thứ hai mã hóa chuỗi nhẹ của polypeptit Yếu tố FVIII là CS23-LC-NA (SEQ ID NO: 23).

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự

đồng nhất ít nhất là 95% với CS04-FL-NA, trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS01-FL-NA, trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS23-FL-NA, trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), hoặc CS23-FL-NA (SEQ ID NO: 20)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS04-FL-NA (SEQ ID NO: 1).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS01-FL-NA (SEQ ID NO: 13).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS23-FL-NA (SEQ ID NO: 20).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 95% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 96% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 97% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 98% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 99% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 99,5% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin có sự đồng nhất ít nhất là 99,9% với CS04-FL-AA (SEQ ID NO: 2).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này mã hóa polypeptit Yếu tố VIII chứa trình tự axit amin của CS04-FL-AA (SEQ ID NO: 2).

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS04-SC1-NA (SEQ ID NO: 9), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS04-SC2-NA (SEQ ID NO: 11), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS01-SC1-NA (SEQ ID NO: 26), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS01-SC2-NA (SEQ ID NO: 27), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS23-SC1-NA (SEQ ID NO: 28), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Theo một khía cạnh, sáng chế đề xuất polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 95% với CS23-SC2-NA (SEQ ID NO: 29), trong đó polynucleotit này mã hóa polypeptit Yếu tố VIII chuỗi đơn.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với trình tự polynucleotit có độ dài đầy đủ tương ứng (ví dụ, CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), hoặc CS23-SC2-NA (SEQ ID NO: 29)).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS04-SC1-NA (SEQ ID NO: 9).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS04-SC2-NA (SEQ ID NO: 11).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS01-SC1-NA (SEQ ID NO: 26).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS01-SC2-NA (SEQ ID NO: 27).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS23-SC1-NA (SEQ ID NO: 28).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này là CS23-SC2-NA (SEQ ID NO: 29).

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA,

CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-

SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit

này có sự đồng nhất ít nhất là 98% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-

FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA,

CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với trình tự được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA,

CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, trình tự nucleotit này được chọn từ nhóm bao gồm CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-

SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, và CS23m234-SC2-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm polypeptit glycosyl hóa nằm giữa hai axit amin liên tiếp.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit được mã hóa bao gồm peptit glycosyl hóa với trình tự axit amin có sự đồng nhất ít nhất là 92% với peptit glycosyl hóa được chọn từ nhóm bao gồm NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA và NGV-AA.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit được mã hóa bao gồm peptit glycosyl hóa với trình tự axit amin được chọn từ nhóm bao gồm NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA và NGV-AA.

Trong một phương án của các polynucleotit được mô tả trên đây, peptit glycosyl hóa được mã hóa bởi polynucleotit có trình tự nucleotit có sự đồng nhất ít nhất là 95% với trình tự được chọn từ nhóm bao gồm NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA, NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA và NGV-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, peptit glycosyl hóa này được mã hóa bởi polynucleotit có trình tự nucleotit được chọn từ một trong số NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA,

NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA và NGV-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với trình tự được chọn từ nhóm bao gồm BDLNG1-NA, BDLNG3-NA, BDLNG5-NA, BDLNG6-NA, BDLNG9-NA, BDLNG10-NA, BDLNG16-NA, BDLNG17-NA, BDLNG18-NA, BDLNG19-NA, BDLNG20-NA và BDLNG21-NA.

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm đột biến thế axit amin F328S (SPI, F309S SPE), so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P (SPI; I86V, A108S, G132K, M147T, và L152P, SPE, tương ứng), so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và b) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19). Trong một số phương án (ví dụ, trong đó phân tử FVIII được mã hóa bao gồm một phần của vùng đầu cùng N của vùng chức năng B kiểu hoang), polypeptit Yếu tố VIII được mã hóa cũng bao gồm đột biến mất đoạn các axit amin SF760-761, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến thế axit amin F328S (SPI, F309S SPE), so với FVIII-FL-AA (SEQ ID NO: 19), và b) đột biến thế axit amin C1918G và C1922G (SPI; C1899G và C1903 SPE, tương ứng), so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến thế axit amin F328S (SPI, F309S SPE), so với FVIII-FL-AA (SEQ ID NO: 19), và b) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P (SPI; I86V, A108S, G132K, M147T, và L152P, SPE, tương ứng), so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến thế axit amin F328S, so với FVIII-FL-AA (SEQ ID NO: 19), b) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và c) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19). Trong một số phương án (ví dụ, trong đó phân tử FVIII được mã hóa bao gồm một phần của vùng đầu cùng N của vùng chức năng B kiểu hoang), polypeptit Yếu tố VIII được mã hóa cũng bao gồm đột biến mất đoạn axit amin SF760-761, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, so với FVIII-FL-AA (SEQ ID NO: 19), b) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và c) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19). Trong một số phương án (ví dụ, trong đó phân tử FVIII được mã hóa bao gồm một phần của vùng đầu cùng N của vùng chức năng B kiểu hoang), polypeptit Yếu tố VIII được mã hóa cũng bao gồm đột biến mất đoạn các axit amin SF760-761, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến thế axit amin F328S, so với FVIII-FL-AA (SEQ ID NO: 19), b) các đột biến thế axit amin C1918G và C1922G, so với FVIII-FL-AA (SEQ ID NO: 19), và c) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) đột biến thế axit amin F328S, so với FVIII-FL-AA (SEQ ID NO: 19), b) các đột biến thế axit amin C1918G và C1922G, so với FVIII-FL-AA (SEQ ID NO: 19), c) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và c) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19). Trong một số phương án (ví dụ, trong đó phân tử FVIII được mã hóa bao gồm một phần của vùng đầu cùng N của vùng chức năng B kiểu hoang), polypeptit Yếu tố VIII được mã hóa cũng bao gồm đột biến mất đoạn các axit amin SF760-761, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polypeptit Yếu tố VIII được mã hóa bao gồm a) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, so với FVIII-FL-AA (SEQ ID NO: 19), b) đột biến thế axit amin F328S, so với FVIII-FL-AA (SEQ ID NO: 19), c) các đột biến thế axit amin C1918G và C1922G, so với FVIII-FL-AA (SEQ ID NO: 19), d) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và e) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19). Trong một số phương án (ví dụ, trong đó phân tử FVIII được mã hóa bao gồm một phần của vùng đầu cùng N của vùng chức năng B kiểu hoang), polypeptit Yếu tố VIII được mã hóa cũng bao gồm đột biến mất đoạn các axit amin SF760-761, so với FVIII-FL-AA (SEQ ID NO: 19).

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này cũng bao gồm trình tự khởi động được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này cũng bao gồm trình tự tăng cường được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này cũng bao gồm yếu tố polyadenyl hóa được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

Trong một phương án của các polynucleotit được mô tả trên đây, polynucleotit này cũng bao gồm intron được liên kết chức năng với trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

Trong một phương án của các polynucleotit được mô tả trên đây, intron này nằm giữa trình tự khởi động và vị trí bắt đầu dịch mã (ví dụ, ATG mã hóa thứ nhất) của trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

Theo một khía cạnh khác, sáng chế đề xuất vectơ liệu pháp gen động vật có vú bao gồm polynucleotit như được mô tả trên đây.

Trong một phương án của vectơ liệu pháp gen động vật có vú được mô tả trên đây, vectơ liệu pháp gen động vật có vú là vectơ virut liên hợp adeno (AAV).

Trong một phương án của vectơ liệu pháp gen động vật có vú được mô tả trên đây, vectơ AAV là vectơ AAV-8.

Theo một khía cạnh khác, sáng chế đề xuất phương pháp điều trị bệnh hemophilia A bao gồm sử dụng, cho bệnh nhân cần được điều trị, vectơ liệu pháp gen động vật có vú như được mô tả trên đây.

Theo một khía cạnh khác, sáng chế đề xuất vectơ liệu pháp gen động vật có vú như được mô tả trên đây để điều trị bệnh hemophilia A.

Theo một khía cạnh khác, sáng chế đề xuất việc sử dụng vectơ liệu pháp gen động vật có vú như được mô tả trên đây để sản xuất thuốc để điều trị bệnh hemophilia A.

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII bao gồm chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-HC-AAm23. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-LC-AAm23. Cầu nối polypeptit của polypeptit Yếu tố VIII bao gồm vị trí cắt furin. Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và iii) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19).

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII bao gồm chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-HC-AAm123. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-LC-AAm123. Cầu nối polypeptit của polypeptit Yếu tố VIII bao gồm vị trí cắt furin. Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), iii) đột biến gắn xen axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19), và iv) đột biến thế axit amin F328S.

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII bao gồm chuỗi

nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-HC-AAm234. Chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 95% với trình tự CS01-LC-AAm234. Cầu nối polypeptit của polypeptit Yếu tố VIII bao gồm vị trí cắt furin. Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), iii) đột biến gắn xen axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19), và iv) các đột biến thế axit amin F328S/C1918G/C1922G.

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 96% với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất ít nhất là 96% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 97% với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất ít nhất là 97% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 98% với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất ít nhất là 98% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 99% với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất ít nhất là 99% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của

polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 99,5% với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất ít nhất là 99,5% với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, chuỗi nặng của polypeptit Yếu tố VIII có trình tự đồng nhất với trình tự chuỗi nặng tương ứng (ví dụ, CS01-HC-AAm23, CS01-HC-AAm123, hoặc CS01-HC-AAm234), và chuỗi nhẹ của polypeptit Yếu tố FVIII có trình tự đồng nhất với trình tự chuỗi nhẹ tương ứng (ví dụ, CS01-LC-AAm23, CS01-LC-AAm123, hoặc CS01-LC-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit có sự đồng nhất ít nhất là 95% với BDL-SQ-AA (SEQ ID NO: 30).

Trong một phương án của các polypeptit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit có trình tự axit amin là BDL-SQ-AA (SEQ ID NO: 30).

Trong một phương án của các polynucleotit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit được mã hóa bao gồm peptit glycosyl hóa với trình tự axit amin có sự đồng nhất ít nhất là 92% với peptit glycosyl hóa được chọn từ nhóm bao gồm NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA và NGV-AA.

Trong một phương án của các polypeptit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit bao gồm peptit glycosyl hóa được chọn từ nhóm bao gồm NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA và NGV-AA.

Trong một phương án của các polypeptit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit có trình tự axit amin có sự đồng nhất ít nhất là 95% với trình tự được chọn từ nhóm bao gồm BDLNG1-AA, BDLNG3-AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-AA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA và BDLNG21-AA.

Trong một phương án của các polypeptit Yếu tố VIII được mô tả trên đây, cầu nối polypeptit có trình tự axit amin được chọn từ nhóm bao gồm BDLNG1-AA, BDLNG3-

AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-NA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA và BDLNG21-AA.

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII có trình tự axit amin đồng nhất ít nhất là 95% với CS40-FL-AAm23 (SEQ ID NO: 104). Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), và iii) đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19).

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII có trình tự axit amin đồng nhất ít nhất là 95% với CS40-FL-AAm123. Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), iii) đột biến gắn xen axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19), và iv) đột biến thế axit amin F328S.

Theo một khía cạnh khác, sáng chế đề xuất polypeptit Yếu tố VIII có trình tự axit amin đồng nhất ít nhất là 95% với CS40-FL-AAm234. Polypeptit Yếu tố VIII bao gồm i) các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, ii) đột biến mất đoạn axit amin AIEPR755-759, so với FVIII-FL-AA (SEQ ID NO: 19), iii) đột biến gắn xen axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19), và iv) các đột biến thế axit amin F328S/C1918G/C1922G.

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 96% với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 97% với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 98% với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 99% với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất ít nhất là 99,5% với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Trong một phương án của các polypeptit Yếu tố VIII đã được mô tả, polypeptit Yếu tố VIII có trình tự đồng nhất với trình tự có độ dài đầy đủ tương ứng (ví dụ, CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, hoặc CS40-FL-AAm234).

Mô tả vắn tắt các hình vẽ

Fig. 1 thể hiện sơ đồ minh họa các cấu trúc protein Yếu tố VIII người typ ReFacto và kiểu hoang.

Fig. 2A và Fig. 2B thể hiện trình tự nucleotit được biến đổi codon CS04 (SEQ ID NO: 1) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS04-FL-NA" đối với trình tự mã hóa có độ dài đầy đủ).

Fig. 3 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 2) được mã hóa bởi trình tự nucleotit được biến đổi codon CS04 theo một số phương án ("CS04-FL-AA" đối với trình tự axit amin có độ dài đầy đủ).

Fig. 4 thể hiện một phần của trình tự nucleotit được biến đổi codon CS04 (SEQ ID NO: 3) mã hóa chuỗi nặng của biến thể Yếu tố VIII theo một số phương án ("CS04-HC-NA").

Fig. 5 thể hiện một phần của trình tự nucleotit được biến đổi codon CS04 (SEQ ID NO: 4) mã hóa chuỗi nhẹ của biến thể Yếu tố VIII theo một số phương án ("CS04-LC-NA").

Fig. 6 thể hiện các trình tự mã hóa được lấy làm ví dụ (SEQ ID NO 5-7 và 36-48, tương ứng, theo thứ tự xuất hiện) cho các cầu nối được thế vùng chức năng B theo một số phương án. BDLO01 (SEQ ID NO: 5), BDLO04 (SEQ ID NO: 6), và BDLO23 (SEQ ID NO: 7) lần lượt là các phần tương ứng của các trình tự nucleotit được biến đổi codon CS01,

CS04, và CS23 mà mã hóa cầu nối được thế vùng chức năng B, tương ứng.

Fig. 7A, Fig. 7B, và Fig. 7C thể hiện trình tự vectơ AAV (SEQ ID NO: 8) chứa trình tự nucleotit được biến đổi codon CS04 theo một số phương án ("CS04-AV-NA").

Fig. 8A và Fig. 8B thể hiện trình tự nucleotit được biến đổi codon CS01m1 (SEQ ID NO: 49) mã hóa biến thể Yếu tố VIII có đột biến thế axit amin F328S theo một số phương án ("CS01m1-FL-NA").

Fig. 9A và Fig. 9B thể hiện trình tự nucleotit được biến đổi codon CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE) (SEQ ID NO: 9) mã hóa biến thể Yếu tố VIII chuỗi đơn theo một số phương án ("CS04-SC1-NA").

Fig. 10 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 10) được mã hóa bởi trình tự nucleotit được biến đổi codon CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE), CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE), và CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) theo một số phương án (lần lượt là "CS01-SC1-AA," "CS04-SC1-AA", và "CS23-SC1-AA").

Fig. 11A và Fig. 11B thể hiện trình tự nucleotit được biến đổi codon CS04Δ(772-1667) (SPI; CS04Δ (753-1648), SPE) (SEQ ID NO: 11) mã hóa biến thể Yếu tố VIII chuỗi đơn theo một số phương án ("CS04-SC2-NA").

Fig. 12 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 12) được mã hóa bởi trình tự nucleotit được biến đổi codon CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE), CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE), và CS23Δ(772-1667) (SPI; CS23Δ(753-1648), SPE) theo một số phương án (lần lượt là "CS01-SC2-AA," "CS04-SC2-AA", và "CS23-SC2-AA").

Fig. 13A và Fig. 13B thể hiện các trình tự axit amin và nucleotit cho các peptit glycosyl hóa điển hình được gắn xen vào cầu nối được thế vùng chức năng B theo một số phương án. "NG1" hoặc NG1-AA" là mã cho trình tự axit amin, được thể hiện ở dòng trên cùng. "NG1-NA" là mã cho trình tự axit nucleic, được thể hiện ở dòng dưới cùng cho mỗi nhóm. Fig. 13A và Fig. 13B mô tả các trình tự axit amin là SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, và các trình tự nucleotit lần lượt là SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, theo thứ tự xuất hiện.

Fig. 14 thể hiện kết quả của dự đoán in silico về sự N-glycosyl hóa in vivo của vùng chức năng B Yếu tố VIII kiểu hoang. Hình vẽ này mô tả lần lượt SEQ ID NO 76 và 76-82, theo thứ tự xuất hiện.

Fig. 15 thể hiện các kết quả của dự đoán in silico về sự N-glycosyl hóa in vivo của cầu nối peptit V3. Hình vẽ này mô tả lần lượt SEQ ID NO 83 và 83-89, theo thứ tự xuất hiện.

Fig. 16A và Fig. 16B thể hiện trình tự nucleotit được biến đổi codon CS01 (SEQ ID NO: 13) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS01-FL-NA").

Fig. 17A và Fig. 17B thể hiện trình tự nucleotit được biến đổi codon CS08 (SEQ ID NO: 14) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS08-FL-NA").

Fig. 18A và Fig. 18B thể hiện trình tự nucleotit được biến đổi codon CS10 (SEQ ID NO: 15) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS10-FL-NA").

Fig. 19A và Fig. 19B thể hiện trình tự nucleotit được biến đổi codon CS11 (SEQ ID NO: 16) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS11-FL-NA").

Fig. 20A và Fig. 20B thể hiện trình tự mã hóa ReFacto kiểu hoang CS40 (SEQ ID NO: 17), theo một số phương án ("CS40-FL-NA").

Fig. 21A và Fig. 21B thể hiện trình tự nucleotit được biến đổi codon CH25 (SEQ ID NO: 18) mã hóa biến thể Yếu tố VIII theo một số phương án ("CH25-FL-NA").

Fig. 22 thể hiện trình tự axit amin Yếu tố VIII người kiểu hoang (SEQ ID NO: 19), theo một số phương án ("FVIII-FL-AA").

Fig. 23 minh họa sơ đồ tách dòng các cấu trúc pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, và pCh25, bằng cách gắn xen các trình tự ADN BDD-FVIII typ Refacto tổng hợp vào khung vectơ pCh-BB01 qua các vị trí giới hạn AscI và NotI.

Fig. 24 thể hiện tính nguyên vẹn của các chế phẩm bộ gen vectơ AAV, như được phân tích bằng điện di gel agaroza. Làn 1, trình tự đánh dấu ADN; làn 2, vCS40; làn 3, vCS01; làn 4, vCS04. Các vectơ AAV đều có các bộ gen có cùng kích thước, di chuyển ở xấp xỉ 5 kb (mũi tên, bên phải). Thang bên trái thể hiện kích thước của các đoạn ADN theo kilobazơ (kb).

Fig. 25 thể hiện phân tích protein của các chế phẩm vectơ AAV bởi PAGE và nhuộm bạc. Làn 1, gen đánh dấu protein (M); làn 2, vCS40, làn 3, vCS01; và làn 4, vCS04. Các cấu trúc này đều có cùng các capsit AAV8 chứa VP1, VP2, và VP3 (các mũi tên bên phải). Thang bên trái thể hiện kích thước của trình tự đánh dấu protein theo kilodalton (kDa).

Fig. 26A và Fig. 26B thể hiện trình tự nucleotit được biến đổi codon CS23 (SEQ ID NO: 20) mã hóa biến thể Yếu tố VIII theo một số phương án ("CS23-FL-NA").

Fig. 27 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 21) được mã hóa bởi trình tự nucleotit được biến đổi codon CS23 theo một số phương án ("CS23-FL-AA").

Fig. 28 thể hiện một phần của trình tự nucleotit được biến đổi codon CS23 (SEQ ID NO: 22) mã hóa chuỗi nặng của biến thể Yếu tố VIII theo một số phương án ("CS23-HC-NA").

Fig. 29 thể hiện một phần của trình tự nucleotit được biến đổi codon CS23 (SEQ ID NO: 23) mã hóa chuỗi nhẹ của biến thể Yếu tố VIII theo một số phương án ("CS23-LC-NA").

Fig. 30A và Fig. 30B thể hiện trình tự nucleotit được biến đổi codon CS01m13 (SEQ ID NO: 90) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m1 (F328S) và m3 theo một số phương án ("CS01-FL-NA-m13").

Fig. 31A và Fig. 31B thể hiện trình tự nucleotit được biến đổi codon CS01m23 (SEQ ID NO: 91) mã hóa biến thể Yếu tố VIII với các tập hợp đột biến m2 và m3 theo một số phương án ("CS01-FL-NA-m23").

Fig. 32A và Fig. 32B thể hiện trình tự nucleotit được biến đổi codon CS01m3 (SEQ ID NO: 92) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m3 theo một số phương án ("CS01-FL-NA-m3").

Fig. 33A và Fig. 33B thể hiện trình tự nucleotit được biến đổi codon CS01m2 (SEQ ID NO: 93) mã hóa biến thể Yếu tố VIII với tập hợp đột biến m2 (các đột biến thế axit amin I105V/A127S/G151K/M166T/L171P (SPI)) theo một số phương án ("CS01-FL-NA-m2").

Fig. 34A và Fig. 34B thể hiện trình tự nucleotit được biến đổi codon CS04m2 (SEQ ID NO: 94) mã hóa biến thể Yếu tố VIII với các đột biến m2 (các đột biến thế axit amin I105V/A127S/G151K/M166T/L171P (SPI)) theo một số phương án ("CS01-FL-NA-m2").

Fig. 35A và Fig. 35B thể hiện trình tự nucleotit được biến đổi codon CS04m3 (SEQ ID NO: 95) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m3 theo một số phương án ("CS04-FL-NA-m3").

Fig. 36A và Fig. 36B thể hiện trình tự nucleotit được biến đổi codon CS04m23 (SEQ ID NO: 96) mã hóa biến thể Yếu tố VIII với tập hợp đột biến m2 (I105V/A127S/G151K/M166T/L171P (SPI)) và các đột biến thế axit amin m3 theo một số phương án ("CS04-FL-NA-m23").

Fig. 37A và Fig. 37B thể hiện trình tự nucleotit được biến đổi codon CS04m1 (SEQ ID NO: 97) mã hóa biến thể Yếu tố VIII với đột biến thế axit amin m1 (F328S) theo một số phương án ("CS04-FL-NA-m1").

Fig. 38A và Fig. 38B thể hiện trình tự nucleotit được biến đổi codon CS04m13 (SEQ ID NO: 98) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m1 và m3 theo một số phương án ("CS04-FL-NA-m13").

Fig. 39A và Fig. 39B thể hiện trình tự nucleotit được biến đổi codon CS23m13 (SEQ ID NO: 99) mã hóa biến thể Yếu tố VIII có đột biến thế axit amin m1 và m3 theo một số phương án ("CS23m13-FL-NA").

Fig. 40A và Fig. 40B thể hiện trình tự nucleotit được biến đổi codon CS23m3 (SEQ ID NO: 100) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m3 theo một số phương án ("CS23-FL-NA-m3").

Fig. 41A và Fig. 41B thể hiện trình tự nucleotit được biến đổi codon CS23m2 (SEQ ID NO: 101) mã hóa biến thể Yếu tố VIII với tập hợp đột biến m2 (các đột biến thế axit amin I105V/A127S/G151K/M166T/L171P) theo một số phương án ("CS23-FL-NA-m2").

Fig. 42A và Fig. 42B thể hiện trình tự nucleotit được biến đổi codon CS23m1 (SEQ ID NO: 102) mã hóa biến thể Yếu tố VIII với đột biến thế axit amin m1 (F328S) theo một số phương án ("CS23-FL-NA-m1").

Fig. 43A và Fig. 43B thể hiện trình tự nucleotit được biến đổi codon CS23m23 (SEQ ID NO: 103) mã hóa biến thể Yếu tố VIII với tập hợp đột biến m2 (I105V/A127S/G151K/M166T/L171P) và các đột biến thế axit amin m3 theo một số phương án ("CS23-FL-NA-m23").

Fig. 44 mô tả việc tách dòng các cấu trúc pCS, được thực hiện bằng cách gắn xen BDD-FVIII typ Refacto tổng hợp mang nhiều đột biến khác nhau (xem bảng kèm theo) vào khung vectơ pCh-BB01 qua các vị trí giới hạn AscI và NotI.

Fig. 45 mô tả phân tích protein của các chế phẩm vectơ AAV bởi PAGE và nhuộm bạc. Làn 1, gen đánh dấu protein (M); làn 2, vCS01, làn 3, vCS17; làn 4, vCS19; làn 5, vCS20; làn 6, vCS40; làn 7, vCS04; làn 8, vCS17; làn 9, cấu trúc vCS24. Các cấu trúc này đều có cùng các capsit AAV8 chứa VP1, VP2 và VP3 (mũi tên bên phải). Thang bên trái thể hiện kích thước của gen đánh dấu protein theo kilo Dalton (kDa).

Fig. 46 mô tả sự nguyên vẹn của các chế phẩm bộ gen vectơ AAV được phân tích bằng điện di gel agaroza. Làn 1, gen đánh dấu ADN (M); làn 2, vCS04, làn 3, vCS17; làn 4, vCS20; làn 5, vCS24; làn 6, vCS16; làn 7, cấu trúc vCS40. Tải lượng vectơ là 1.5E10 vg mỗi làn. Các vectơ AAV có các bộ gen có cùng kích thước, di chuyển ở xấp xỉ 5 kb (mũi tên, bên phải). Thang bên trái thể hiện kích thước của các đoạn ADN theo kilobazơ (kb).

Fig. 47 thể hiện một phần của trình tự nucleotit được biến đổi codon CS01 (SEQ ID NO: 24) mã hóa chuỗi nặng của biến thể Yếu tố VIII theo một số phương án ("CS01-HC-NA").

Fig. 48 thể hiện một phần của trình tự nucleotit được biến đổi codon CS01 (SEQ ID NO: 25) mã hóa chuỗi nhẹ của biến thể Yếu tố VIII theo một số phương án ("CS01-LC-NA").

Fig. 49A và Fig. 49B thể hiện trình tự nucleotit được biến đổi codon CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE) (SEQ ID NO: 26) mã hóa biến thể Yếu tố VIII chuỗi đơn theo một số phương án ("CS01-SC1-NA").

Fig. 50A và Fig. 50B thể hiện trình tự nucleotit được biến đổi codon CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE) (SEQ ID NO: 27) mã hóa biến thể Yếu tố VIII chuỗi

đơn theo một số phương án ("CS01-SC2-NA").

Fig. 51A và Fig. 51B thể hiện trình tự nucleotit được biến đổi codon CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) (SEQ ID NO: 28) mã hóa biến thể Yếu tố VIII chuỗi đơn theo một số phương án ("CS23-SC1-NA").

Fig. 52A và Fig. 52B thể hiện trình tự nucleotit được biến đổi codon CS23Δ(772-1667) (SPI; CS23Δ (753-1648), SPE) (SEQ ID NO: 29) mã hóa biến thể Yếu tố VIII chuỗi đơn theo một số phương án ("CS23-SC2-NA").

Fig. 53 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 104) được mã hóa bởi trình tự nucleotit được biến đổi codon CS01m23 theo một số phương án ("CS01m23-FL-AA").

Fig. 54 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 105) được mã hóa bởi trình tự nucleotit được biến đổi codon CS04m3 theo một số phương án ("CS01m23-FL-AA").

Fig. 55 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 106) được mã hóa bởi trình tự nucleotit được biến đổi codon CS01m12 theo một số phương án ("CS01m12-FL-AA").

Fig. 56 thể hiện trình tự axit amin biến thể Yếu tố VIII (SEQ ID NO: 107) được mã hóa bởi trình tự nucleotit được biến đổi codon CS04m12 theo một số phương án ("CS04m12-FL-AA").

Fig. 57A và Fig. 57B thể hiện trình tự nucleotit được biến đổi codon CS01m12 (SEQ ID NO: 108) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m1 (F328S) và m2 theo một số phương án ("CS01-FL-NAm12").

Fig. 58A và Fig. 58B thể hiện trình tự nucleotit được biến đổi codon CS04m12 (SEQ ID NO: 109) mã hóa biến thể Yếu tố VIII với các đột biến thế axit amin m1 (F328S) và m2 theo một số phương án ("CS04-FL-NAm12").

Mô tả chi tiết sáng chế

Giới thiệu

Liệu pháp gen dựa trên AAV có tiềm năng lớn đối với việc điều trị cho người mắc bệnh máu khó đông. Đối với bệnh hemophilia B, số liệu lâm sàng đầu tiên gợi ý là các mức FIX khoảng 10% có thể được duy trì ở ít nhất là một số bệnh nhân trong hơn 1 năm. Tuy nhiên, đối với bệnh hemophilia A, việc đạt được các mức biểu hiện có tác dụng điều trị từ 5-10% với các vectơ AAV vẫn là một thách thức do nhiều nguyên nhân khác nhau. Thứ nhất, trình tự mã hóa Yếu tố VIII là quá lớn đối với các vectơ dựa trên AAV thông thường. Thứ hai, các cấu trúc Yếu tố VIII bị cắt hoặc bị mất vùng chức năng B được thao tác di truyền có mức biểu hiện kém trên in vivo, ngay cả khi được tối ưu hóa codon. Thứ ba, các cấu trúc biến thể Yếu tố VIII bị cắt hoặc bị mất vùng chức năng B này có thời gian bán thải ngắn in vivo, làm trầm trọng hơn các tác động của sự biểu hiện kém. Thứ tư, ngay cả khi được biểu hiện, FVIII không được tiết đủ từ các tế bào, như là các yếu tố đông máu khác, như Yếu tố IX.

Ngoài ra, các thách thức này không thể được giải quyết bằng cách sử dụng đơn giản các liều cao hơn của cấu trúc liệu pháp gen này. Theo hiểu biết hiện nay, liều vectơ của vectơ liệu pháp gen dựa trên AAV nên được gia tăng trên $2x10^{12}$ vg/kg thể trọng. Đó là vì ở các liều cao này, đáp ứng miễn dịch tế bào T được khởi động, đáp ứng này phá vỡ các tế bào đã tải nạp và kết quả là, sự biểu hiện gen chuyển bị giảm hoặc thậm chí là bị triệt tiêu. Do đó, các chiến lược để cải thiện sự biểu hiện của FVIII là cần thiết để làm cho liệu pháp gen FVIII trở thành lựa chọn điều trị sống còn cho bệnh nhân mắc bệnh hemophilia A.

Sáng chế đề cập đến việc phát hiện ra các trình tự mã hóa biến thể Yếu tố VIII được biến đổi codon mà giải quyết được các vấn đề này và các vấn đề khác liên quan đến liệu pháp gen Yếu tố VIII. Ví dụ, các polynucleotit được mô tả ở đây gây ra sự biểu hiện được cải thiện đáng kể ở các tế bào động vật có vú, và biểu hiện khả năng đóng gói virion được cải thiện do các tương tác đóng gói được ổn định. Trong một số phương án, các ưu điểm này được nhận ra bằng cách sử dụng các trình tự mã hóa cho chuỗi nặng và chuỗi nhẹ của Yếu tố VIII có sự đồng nhất trình tự cao với các cấu trúc CS01, CS04, và CS23 được biến đổi codon (ví dụ, có sự đồng nhất trình tự cao với một trong số các trình tự mã hóa chuỗi nặng CS01-HC, CS04-HC, và CS23-HC và sự đồng nhất trình tự cao với một trong số các trình tự mã hóa chuỗi nhẹ CS01-LC, CS04-LC, và CS23-LC).

Trong một số phương án, các phân tử Yếu tố VIII được mã hóa bởi các polynucleotit được mô tả ở đây được làm ngắn đi bằng cách cắt, làm mất đoạn, hoặc thay thế vùng chức

năng B kiểu hoang. Do đó, các polynucleotit này là thích hợp hơn để biểu hiện Yếu tố VIII bằng các vectơ liệu pháp gen thông thường, mà biểu hiện không hiệu quả các polypeptit lớn hơn, như Yếu tố VIII kiểu hoang.

Thuận lợi là, được thể hiện ở đây là các trình tự mã hóa biến thể Yếu tố VIII được biến đổi codon CS01, CS04, và CS23 gây ra sự biểu hiện tốt hơn của cấu trúc Yếu tố VIII bị mất vùng chức năng B in vivo. Ví dụ, đã chứng minh được trong Ví dụ 2 và Ví dụ 4 là việc sử dụng trong tĩnh mạch các vectơ liệu pháp gen dựa trên AAV có trình tự mã hóa CS01 (SEQ ID NO: 13), CS04 (SEQ ID NO: 1), và CS23 (SEQ ID NO: 20) gây ra sự tăng biểu hiện Yếu tố VIII 18 lần, 74 lần và 30 lần, so với cấu trúc CS40 tương ứng được mã hóa với trình tự polynucleotit kiểu hoang (SEQ ID NO: 17), ở chuột bất hoạt Yếu tố VIII (Bảng 4 và Bảng 7).

Ngoài ra, cũng được chỉ ra ở đây là các trình tự mã hóa biến thể Yếu tố VIII được biến đổi codon CS01 và CS04 gây ra sự đóng gói virion và sản xuất virut tốt hơn. Ví dụ, đã chứng minh được trong Ví dụ 1 là các cấu trúc vectơ AAV chứa các cấu trúc CS01 và CS04 tạo ra năng suất virut cao hơn từ 5 đến 7 lần, so với cấu trúc CS40 tương ứng được mã hóa bằng trình tự polynucleotit kiểu hoang, khi được phân lập từ cùng một lượng pelet tế bào.

Thuận lợi là, các tác giả sáng chế còn phát hiện ra là hoạt tính Yếu tố VIII cải thiện được tạo ra từ các trình tự được biến đổi codon CS01, CS04, và CS23 có thể được tăng cường thêm bằng cách đưa các đột biến vào trình tự polypeptit Yếu tố VIII cơ bản. Ví dụ, như được chứng minh trong Ví dụ 4, các đột biến F328S, X5, và X1, một mình và kết hợp với một chất khác, làm tăng thêm hoạt tính FVIII khi được biểu hiện in vivo trong nền được biến đổi codon CS01 hoặc CS04 từ 2 đến 7 lần, so với các cấu trúc được biến đổi codon, kiểu hoang (Bảng 7). Đáng chú ý hơn là, các trình tự được biến đổi codon này, mã hóa các đột biến Yếu tố VIII đột biến, gây ra sự tăng đến 246 lần so với cấu trúc CS40 tương ứng được mã hóa bằng trình tự polynucleotit kiểu hoang (Bảng 7).

Các định nghĩa

Như được sử dụng ở đây, các thuật ngữ sau đây có ý nghĩa như được mô tả cho chúng, trừ khi có quy định khác.

Như được sử dụng ở đây, các thuật ngữ "Yếu tố VIII" và "FVIII" được sử dụng thay thế cho nhau, và dùng để chỉ protein bất kỳ có hoạt tính Yếu tố VIII (ví dụ, FVIII hoạt hóa, thường được gọi là FVIIIa) hoặc tiền protein (ví dụ, tiền protein có trình tự ức chế (pro-protein) hoặc tiền protein có cả trình tự tín hiệu và trình tự ức chế (pre-pro-protein)) của protein có hoạt tính Yếu tố VIII, cụ thể là hoạt tính đồng nhất tố Yếu tố IXa. Trong một ví dụ cụ thể, polypeptit Yếu tố VIII dùng để chỉ polypeptit có các trình tự có sự đồng nhất trình tự cao (ví dụ, ít nhất là 70%, 75%, 80%, 85%, 90%, 95%, 99%, hoặc cao hơn) với chuỗi nặng và nhẹ của polypeptit Yếu tố VIII kiểu hoang. Trong một số phương án, vùng chức năng B của polypeptit Yếu tố VIII bị mất đoạn, bị cắt hoặc bị thay thế bằng polypeptit cầu nối để làm giảm kích thước của polynucleotit mã hóa polypeptit Yếu tố VIII. Trong một phương án điển hình, các axit amin 20-1457 của SEQ ID NO: 2 tạo thành polypeptit Yếu tố VIII.

Các ví dụ không giới hạn về các polypeptit Yếu tố VIII kiểu hoang bao gồm tiền Yếu tố VIII có cả trình tự tín hiệu và trình tự ức chế của người (ví dụ, số truy cập Ngân hàng Gen AAA52485, CAA25619, AAA58466, AAA52484, AAA52420, AAV85964, BAF82636, BAG36452, CAI41660, CAI41666, CAI41672, CAI43241, CAO03404, EAW72645, AAH22513, AAH64380, AAH98389, AAI11968, AAI11970, hoặc AAB61261), tiền Yếu tố VIII có trình tự ức chế tương ứng (pro-Factor), và các biến thể tự nhiên của nó, tiền Yếu tố VIII có cả trình tự tín hiệu và trình tự ức chế ở lợn (ví dụ, số truy cập UniProt F1RZ36 hoặc K7GSZ5), tiền Yếu tố VIII có trình tự ức chế tương ứng, và các biến thể tự nhiên của chúng; tiền Yếu tố VIII có cả trình tự tín hiệu và trình tự ức chế của chuột nhắt (ví dụ, số truy cập Ngân hàng Gen AAA37385, CAM15581, CAM26492, hoặc EDL29229), tiền Yếu tố VIII có trình tự ức chế tương ứng, và các biến thể tự nhiên của chúng, tiền Yếu tố VIII có cả trình tự tín hiệu và trình tự ức chế của chuột cống (ví dụ, số truy cập Ngân hàng Gen là AAQ21580), tiền Yếu tố VIII có trình tự ức chế tương ứng, và các biến thể tự nhiên của chúng; tiền Yếu tố VIII có cả trình tự ức chế và trình tự tín hiệu của chuột cống; và các chất tương tự Yếu tố VIII ở động vật có vú khác (ví dụ, khỉ, linh trưởng, chuột hang, chuột lang nhà, v.v...).

Như được sử dụng ở đây, polypeptit Yếu tố VIII bao gồm các biến thể tự nhiên và các cấu trúc nhân tạo có hoạt tính đồng nhân tố Yếu tố IX. Như được sử dụng trong phần mô tả này, Yếu tố VIII bao gồm các biến thể tự nhiên bất kỳ, các trình tự thay thế, các chất

đồng đẳng, hoặc các protein đột biến mà vẫn giữ được một số hoạt tính đồng nhân tố Yếu tố IX cơ bản (ví dụ, ít nhất là 5%, 10%, 25%, 50%, 75%, hoặc nhiều hơn hoạt tính kiểu hoang tương ứng). Các ví dụ về các biến đổi axit amin Yếu tố VIII (so với FVIII-FL-AA (SEQ ID NO: 19)) được tìm thấy trong quần thể người bao gồm, không giới hạn ở, S19R, R22T, Y24C, Y25C, L26P/R, E30V, W33G, Y35C/H, G41C, R48C/K, K67E/N, L69P, E72K, D75E/V/Y, P83R, G89D/V, G92A/V, A97P, E98K, V99D, D101G/H/V, V104D, K108T, M110V, A111T/V, H113R/Y, L117F/R, G121S, E129V, G130R, E132D, Y133C, D135G/Y, T137A/I, S138R, E141K, D145H, V147D, Y155H, V159A, N163K, G164D/V, P165S, C172W, S176P, S179P, V181E/M, K185T, D186G/N/Y, S189L, L191F, G193R, L195P, C198G, S202N/R, F214V, L217H, A219D/T, V220G, D222V, E223K, G224W, T252I, V253F, N254I, G255V, L261P, P262L, G263S, G266F, C267Y, W274C, H275L, G278R, G280D, E284K, V285G, E291G/K, T294I, F295L, V297A, N299I, R301C/H/L, A303E/P, I307S, S308L, F312S, T314A/I, A315V, G323E, L326P, L327P/V, C329F, I331V, M339T, E340K, V345A/L, C348R/S/Y, Y365C, R391C/H/P, S392L/P, A394S, W401G, I405F/S, E409G, W412G/R, K427I, L431F/S, R437P/W, I438F, G439D/S/V, Y442C, K444R, Y450D/N, T454I, F455C, G466E, P470L/R/T, G474E/R/V, E475K, G477V, D478N, T479R, F484C, A488G, R490G, Y492C/H, Y492H, I494T, P496R, G498R, R503H, G513S/V, I522Y, K529E, W532G, P540T, T541S, D544N, R546W, R550C/G/H, S553P, S554C/G, V556D, R560T, D561G/H/Y, I567T, P569R, S577F, V578A, D579A/H, N583S, Q584H/K/R, I585R/T, M586V, D588G/Y, L594Q, S596P, N601D/K, R602G, S603I/R, W604C, Y605H/S, N609I, R612C, N631K/S, M633I, S635N, N637D/I/S, Y639C, L644V, L650F, V653A/M, L659P, A663V, Q664P, F677L, M681I, V682F, Y683C/N, T686R, F698L, M699T/V, M701I, G705V, G710W, N713I, R717L/W, G720D/S, M721I/L, A723T, L725Q, V727F, E739K, Y742C, R795G, P947R, V1012L, E1057K, H1066Y, D1260E, K1289Q, Q1336K, N1460K, L1481P, A1610S, I1698T, Y1699C/F, E1701K, Q1705H, R1708C/H, T1714S, R1715G, A1720V, E1723K, D1727V, Y1728C, R1740G, K1751Q, F1762L, R1768H, G1769R, L1771P, L1775F/V, L1777P, G1779E/R, P1780L, I1782R, D1788H, M1791T, A1798P, S1799H, R1800C/G/H, P1801A, Y1802C, S1803Y, F1804S, L1808F, M1842I, P1844S, T1845P, E1848G, A1853T/V, S1858C, K1864E, D1865N/Y, H1867P/R, G1869D/V, G1872E, P1873R, L1875P, V1876L, C1877R/Y, L1882P, R1888I, E1894G, I1901F, E1904D/K, S1907C/R, W1908L, Y1909C, A1939T/V, N1941D/S, G1942A, M1945V, L1951F, R1960L/Q,

L1963P, S1965I, M1966I/V, G1967D, S1968R, N1971T, H1973L, G1979V, H1980P/Y, F1982I, R1985Q, L1994P, Y1998C, G2000A, T2004R, M2007I, G2013R, W2015C, R2016P/W, E2018G, G2022D, G2028R, S2030N, V2035A, Y2036C, N2038S, 2040Y, G2045E/V, I2051S, I2056N, A2058P, W2065R, P2067L, A2070V, S2082N, S2088F, D2093G/Y, H2101D, T2105N, Q2106E/P/R, G2107S, R2109C, I2117F/S, Q2119R, F2120C/L, Y2124C, R2135P, S2138Y, T2141N, M2143V, F2145C, N2148S, N2157D, P2162L, R2169C/H, P2172L/Q/R, T2173A/I, H2174D, R2178C/H/L, R2182C/H/P, M2183R/V, L2185S/W, S2192I, C2193G, P2196R, G2198V, E2200D, I2204T, I2209N, A2211P, A2220P, P2224L, R2228G/L/P/Q, L2229F, V2242M, W2248C/S, V2251A/E, M2257V, T2264A, Q2265R, F2279C/I, I2281T, D2286G, W2290L, G2304V, D2307A, P2319L/S, R2323C/G/H/L, R2326G/L/P/Q, Q2330P, W2332R, I2336F, R2339T, G2344C/D/S, và C2345S/Y. Các protein Yếu tố VIII cũng bao gồm các polypeptit chứa các biến đổi sau dịch mã.

Nói chung, các polynucleotit mã hóa Yếu tố VIII mã hóa polypeptit chuỗi đơn không hoạt động (ví dụ, tiền protein có cả trình tự tín hiệu và trình tự ức chế (pre-pro-protein)) mà trải qua quá trình xử lý sau dịch mã để tạo ra protein Yếu tố VIII hoạt động (ví dụ, FVIIIa). Ví dụ, theo Fig. 1, tiền protein có cả trình tự tín hiệu và trình tự ức chế của Yếu tố VIII người kiểu hoang đầu tiên được cắt để giải phóng peptit tín hiệu được mã hóa (không được thể hiện), tạo ra tiền protein chuỗi đơn đầu tiên (được chỉ ra là "FVIII người kiểu hoang). Tiền protein này sau đó được cắt giữa vùng chức năng B và A3 để tạo ra polypeptit thứ nhất bao gồm chuỗi nặng Yếu tố VIII (ví dụ, các vùng chức năng A1 và A2) và vùng chức năng B, và polypeptit thứ hai bao gồm chuỗi nhẹ Yếu tố VIII (ví dụ, bao gồm các vùng chức năng A3, C1, và C3). Polypeptit thứ nhất được cắt thêm để loại bỏ vùng chức năng B, và để tách các vùng chức năng A1 và A2, mà vẫn liên kết với chuỗi nhẹ Yếu tố VIII trong protein Yếu tố VIIIa thành thục. Để xem lại về quy trình thành thục Yếu tố VIII, xem Graw et al., Nat Rev Genet., 6(6):488-501 (2005), nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn toàn bộ với tất cả các mục đích.

Tuy nhiên, trong một số phương án, polypeptit Yếu tố VIII là polypeptit Yếu tố VIII chuỗi đơn. Polypeptit Yếu tố VIII chuỗi đơn được thao tác di truyền để loại bỏ các vị trí cắt tự nhiên, và tùy ý loại bỏ, cắt, hoặc thay thế vùng chức năng B của Yếu tố VIII. Do đó, chúng không được thành thục bằng cách cắt (khác với cắt peptit tín hiệu tùy ý và/hoặc

peptit dẫn đầu), và hoạt động như là một chuỗi đơn. Các ví dụ không giới hạn về polypeptit Yếu tố VIII chuỗi đơn được mô tả trong Zollner et al. (Thromb Res, 134(1):125-31 (2014)) và Donath et al. (Biochem J., 312(1):49-55 (1995)), phần mô tả của tài liệu này được đưa vào đây bằng cách viện dẫn toàn bộ cho tất cả các mục đích.

Như được sử dụng ở đây, thuật ngữ "chuỗi nặng Yếu tố VIII", hoặc đơn giản "chuỗi nặng", dùng để chỉ sự kết hợp của các vùng chức năng A1 và A2 của polypeptit Yếu tố VIII. Trong một phương án điển hình, các axit amin 20-759 của CS04-FL-AA (SEQ ID NO: 2) tạo thành chuỗi nặng Yếu tố VIII.

Như được sử dụng ở đây, thuật ngữ "chuỗi nhẹ Yếu tố VIII", hoặc đơn giản là "chuỗi nhẹ" dùng để chỉ sự kết hợp của các vùng chức năng A3, C1, và C2 của polypeptit Yếu tố VIII. Trong một phương án điển hình, các axit amin 774-1457 của CS04-FL-AA (SEQ ID NO: 2) tạo thành chuỗi nhẹ Yếu tố VIII. Trong một số phương án, chuỗi nhẹ Yếu tố VIII không bao gồm peptit axit a3, peptit này được giải phóng trong quá trình thành thục in vivo.

Nói chung, các chuỗi nặng và nhẹ Yếu tố VIII được biểu hiện dưới dạng chuỗi polypeptit đơn, ví dụ, cùng với vùng chức năng B tùy ý hoặc cầu nối được thay thế vùng chức năng B. Tuy nhiên, trong một số phương án, chuỗi nặng Yếu tố VIII và chuỗi nhẹ Yếu tố VIII được biểu hiện dưới dạng các chuỗi polypeptit riêng biệt (ví dụ, được đồng biểu hiện), và được kết hợp lại để tạo ra protein Yếu tố VIII (ví dụ, trên in vivo hoặc in vitro).

Như được sử dụng ở đây, thuật ngữ "cầu nối được thế vùng chức năng B" và "cầu nối Yếu tố VIII" được sử dụng thay thế lẫn nhau, và dùng để chỉ các dạng bị cắt của vùng chức năng B Yếu tố VIII kiểu hoang (ví dụ, các axit amin 760-1667 của FVIII-FL-AA (SEQ ID NO: 19)) hoặc các peptit được thao tác di truyền để thay thế vùng chức năng B của polypeptit Yếu tố VIII. Như được sử dụng ở đây, cầu nối Yếu tố VIII được đặt giữa đầu cùng C của chuỗi nặng Yếu tố VIII và đầu cùng N của chuỗi nhẹ Yếu tố VIII trong polypeptit biến thể Yếu tố VIII theo một số phương án. Ví dụ không giới hạn về các cầu nối được thế vùng chức năng B được mô tả trong Patent Mỹ số 4,868,112, 5,112,950, 5,171,844, 5,543,502, 5,595,886, 5,610,278, 5,789,203, 5,972,885, 6,048,720, 6,060,447, 6,114,148, 6,228,620, 6,316,226, 6,346,513, 6,458,563, 6,924,365, 7,041,635, và

7,943,374; Công bố đơn yêu cầu cấp Patent Mỹ số 2013/024960, 2015/0071883, và 2015/0158930; và Công bố đơn PCT số WO 2014/064277 và WO 2014/127215, phần mô tả của tài liệu này được đưa vào đây bằng cách viện dẫn, toàn bộ, cho tất cả các mục đích.

Trừ khi có quy định khác, việc đánh số các axit amin Yếu tố VIII dùng để chỉ axit amin tương ứng trong trình tự Yếu tố VIII người, kiểu hoang, có độ dài đầy đủ (FVIII-FL-AA), được thể hiện bằng SEQ ID NO: 19 trong Fig. 22. Do đó, khi đề cập đến đột biến thế axit amin trong protein biến thể Yếu tố VIII được mô tả ở đây, số của axit amin được trích dẫn dùng để chỉ axit amin tương tự (ví dụ, tương đương về cấu trúc hoặc chức năng) và/hoặc tương đồng (ví dụ, được bảo tồn trong quá trình tiến hóa trong trình tự axit amin sơ cấp) trong trình tự Yếu tố VIII kiểu hoang, có độ dài đầy đủ. Ví dụ, đột biến thế axit amin T2105N dùng để chỉ đột biến thế T thành N ở vị trí 2105 của trình tự Yếu tố VIII người, kiểu hoang, có độ dài đầy đủ (FVIII-FL-AA; SEQ ID NO: 19), đột biến thế T thành N ở vị trí 1211 của protein biến thể Yếu tố VIII được mã hóa bởi CS04 (CS04-FL-AA; SEQ ID NO: 2), và đột biến thế T thành N ở vị trí 1212 của biến thể Yếu tố VIII được mã hóa bởi CS04m3 (CS04m3-FL-AA; SEQ ID NO: 105).

Như được mô tả ở đây, hệ thống đánh số axit amin Yếu tố VIII là phụ thuộc vào việc peptit tín hiệu Yếu tố VIII (ví dụ, axit amin 1-19 của trình tự Yếu tố VIII người kiểu hoang, có độ dài đầy đủ) có được bao gồm hay không. Khi peptit tín hiệu được bao gồm, việc đánh số được gọi là "bao gồm peptit tín hiệu" hoặc "SPI". Khi peptit tín hiệu không được bao gồm, việc đánh số được gọi là "không bao gồm peptit tín hiệu" hoặc "SPE." Ví dụ, F328S là sự đánh số SPI cho cùng một axit amin với F309S, theo cách đánh số SPE. Trừ khi có quy định khác, toàn bộ việc đánh số axit amin dùng để chỉ axit amin tương ứng trong trình tự Yếu tố VIII người kiểu hoang, có độ dài đầy đủ (FVIII-FL-AA), được thể hiện bằng SEQ ID NO: 19 trong Fig. 22.

Như được mô tả ở đây, các polynucleotit được biến đổi codon gây ra sự biểu hiện tăng của Yếu tố VIII chuyển gen trên in vivo (ví dụ, khi được sử dụng như một phần của vectơ liệu pháp gen), so với mức độ biểu hiện Yếu tố VIII được gây ra bởi cấu trúc Yếu tố VIII được mã hóa tự nhiên (ví dụ, polynucleotit mã hóa cùng một cấu trúc Yếu tố VIII sử dụng các codon người kiểu hoang). Như được sử dụng ở đây, thuật ngữ "sự biểu hiện tăng" dùng để chỉ mức hoạt tính Yếu tố VIII chuyển gen được tăng cường trong máu của động vật sử dụng polynucleotit được biến đổi codon mã hóa Yếu tố VIII, so với mức hoạt tính

Yếu tố VIII chuyển gen trong máu của động vật sử dụng cấu trúc Yếu tố VIII được mã hóa tự nhiên. Các mức hoạt tính này có thể đo được bằng cách sử dụng hoạt tính Yếu tố VIII bất kỳ đã biết trong lĩnh vực này. Thử nghiệm điển hình để xác định hoạt tính Yếu tố VIII là thử nghiệm Technochrome FVIII (Technoclone, Vienna, Áo).

Trong một số phương án, sự biểu hiện tăng dùng để chỉ hoạt tính Yếu tố VIII chuyển gen trong máu của động vật được sử dụng polynucleotit Yếu tố VIII được biến đổi codon cao hơn ít nhất là 25% so với mức hoạt tính Yếu tố VIII chuyển gen trong máu của động vật được sử dụng polynucleotit Yếu tố VIII được mã hóa tự nhiên. Trong một số phương án, sự biểu hiện tăng dùng để chỉ hoạt tính Yếu tố VIII chuyển gen trong máu của động vật được sử dụng polynucleotit Yếu tố VIII được biến đổi codon cao hơn ít nhất là 50%, cao hơn ít nhất là 75%, cao hơn ít nhất là 100%, cao hơn ít nhất là 3 lần, cao hơn ít nhất là 4 lần, cao hơn ít nhất là 5 lần, cao hơn ít nhất là 6 lần, cao hơn ít nhất là 7 lần, cao hơn ít nhất là 8 lần, cao hơn ít nhất là 9 lần, cao hơn ít nhất là 10 lần, cao hơn ít nhất là 15 lần, cao hơn ít nhất là 20 lần, cao hơn ít nhất là 25 lần, cao hơn ít nhất là 30 lần, cao hơn ít nhất là 40 lần, cao hơn ít nhất là 50 lần, cao hơn ít nhất là 60 lần, cao hơn ít nhất là 70 lần, cao hơn ít nhất là 80 lần, cao hơn ít nhất là 90 lần, cao hơn ít nhất là 100 lần, cao hơn ít nhất là 125 lần, cao hơn ít nhất là 150 lần, cao hơn ít nhất là 175 lần, cao hơn ít nhất là 200 lần, cao hơn ít nhất là 225 lần, hoặc cao hơn ít nhất là 250 lần so với mức hoạt tính Yếu tố VIII chuyển gen trong máu của động vật được sử dụng polynucleotit Yếu tố VIII được mã hóa tự nhiên.

Như được sử dụng ở đây, các polynucleotit được biến đổi codon gây ra sự sản xuất vectơ tăng, so với mức sản xuất vectơ được gây ra bởi cấu trúc Yếu tố VIII được mã hóa tự nhiên (ví dụ, polynucleotit mã hóa cùng một cấu trúc Yếu tố VIII bằng cách sử dụng các codon người kiểu hoang). Như được sử dụng ở đây, thuật ngữ "sự sản xuất virut tăng" dùng để chỉ năng suất vectơ được gia tăng trong môi trường nuôi cấy tế bào (ví dụ, hiệu giá trên mỗi lit môi trường nuôi cấy) được gây nhiễm bằng polynucleotit được biến đổi codon mã hóa Yếu tố VIII, so với năng suất vectơ trong môi trường nuôi cấy tế bào được gây nhiễm bằng cấu trúc Yếu tố VIII được mã hóa tự nhiên. Năng suất vectơ có thể đo được bằng cách sử dụng thử nghiệm hiệu giá vectơ bất kỳ đã biết trong lĩnh vực này. Thử nghiệm điển hình để xác định năng suất vectơ (ví dụ, của vectơ AAV) là qPCR gắn đích các trình tự lặp đầu cùng đảo AAV2 (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)).

Trong một số phương án, sự sản xuất virut tăng dùng để chỉ năng suất vectơ được biến đổi codon cao hơn ít nhất là 25% so với năng suất của cấu trúc Yếu tố VIII được mã hóa tự nhiên trong cùng một loại môi trường nuôi cấy. Trong một số phương án, sự sản xuất vectơ tăng dùng để chỉ năng suất vectơ được biến đổi codon cao hơn ít nhất là 50%, cao hơn ít nhất là 75%, cao hơn ít nhất là 100%, cao hơn ít nhất là 3 lần, cao hơn ít nhất là 4 lần, cao hơn ít nhất là 5 lần, cao hơn ít nhất là 6 lần, cao hơn ít nhất là 7 lần, cao hơn ít nhất là 8 lần, cao hơn ít nhất là 9 lần, cao hơn ít nhất là 10 lần, cao hơn ít nhất là 15 lần, hoặc cao hơn ít nhất là 20 lần so với năng suất của cấu trúc Yếu tố VIII được mã hóa tự nhiên trong cùng một loại môi trường nuôi cấy.

Như được sử dụng ở đây, thuật ngữ “bệnh hemophilia" dùng để chỉ nhóm tình trạng bệnh được đặc trưng bởi sự đông máu hoặc tạo cục máu đông giảm. Bệnh hemophilia có thể dùng để chỉ bệnh hemophilia Typ A, Typ B, hoặc Typ C, hoặc kết hợp của tất cả ba loại bệnh này. Bệnh hemophilia Typ A (hemophilia A) bị gây ra bởi sự giảm hoặc mất hoạt tính Yếu tố VIII (FVIII) và là phổ biến nhất trong số các typ phụ của bệnh hemophilia. Bệnh hemophilia Typ B (hemophilia B) bị gây ra bởi sự mất hoặc giảm chức năng đông máu của yếu tố IX (FIX). Bệnh hemophilia Typ C (hemophilia C) là hậu quả của sự mất hoặc giảm hoạt tính chống đông máu của yếu tố XI (FXI). Bệnh hemophilia A và B là các bệnh liên quan đến nhiễm sắc thể X, trong khi bệnh hemophilia C là bệnh liên quan đến nhiễm sắc thể thường. Các phương pháp điều trị thông thường cho bệnh hemophilia bao gồm cả việc sử dụng để phòng ngừa và theo nhu cầu các yếu tố đông máu, như FVIII, FIX, bao gồm Bebulin®–VH, và FXI, cũng như FEIBA-VH, desmopressin, và các dịch truyền huyết tương.

Như được sử dụng ở đây, thuật ngữ "liệu pháp gen FVIII" bao gồm phương pháp điều trị bất kỳ cung cấp axit nucleic mã hóa Yếu tố VIII cho bệnh nhân để làm dịu, làm giảm, hoặc phòng ngừa sự tái phát một hoặc nhiều triệu chứng (ví dụ, các yếu tố lâm sàng) bị gây ra bởi bệnh hemophilia. Thuật ngữ này bao gồm việc sử dụng hợp chất, thuốc, quy trình, hoặc phác đồ bất kỳ chứa axit nucleic mã hóa phân tử Yếu tố VIII, bao gồm dạng được cải biến bất kỳ của Yếu tố VIII (ví dụ, biến thể Yếu tố VIII), để duy trì hoặc cải thiện sức khỏe của cá thể mắc bệnh hemophilia. Người có hiểu biết trung bình trong lĩnh vực này sẽ hiểu rõ là liệu trình của liệu pháp FVIII hoặc liều của dược chất FVIII có thể được thay đổi, ví dụ, dựa vào các kết quả thu được theo sáng chế.

Như được sử dụng ở đây, thuật ngữ "liệu pháp đi tắt" bao gồm phương pháp điều trị bất kỳ cung cấp chất, hợp chất đông máu hoặc yếu tố đông máu không phải Yếu tố VIII cho bệnh nhân để làm dịu, làm giảm, hoặc ngăn ngừa sự tái phát của một hoặc nhiều triệu chứng (ví dụ, các yếu tố lâm sàng) bị gây ra bởi bệnh hemophilia. Các hợp chất và các yếu tố đông máu không phải Yếu tố VIII bao gồm, nhưng không giới hạn ở, phức hợp kháng chất ức chế Yếu tố VIII bắc cầu (FEIBA), yếu tố VII hoạt hóa (FVIIa) tái tổ hợp, phức hợp prothrombin đậm đặc, và phức hợp prothrombin đậm đặc hoạt hóa. Các hợp chất và các yếu tố đông máu không phải Yếu tố VIII có thể là tái tổ hợp hoặc thu được từ huyết tương. Người có hiểu biết trung bình trong lĩnh vực này sẽ hiểu rõ là liệu trình của liệu pháp đi tắt hoặc liều của liệu pháp đi tắt có thể được thay đổi, ví dụ dựa vào các kết quả thu được theo sáng chế.

Như được sử dụng ở đây, "liệu pháp kết hợp" bao gồm sử dụng axit nucleic mã hóa phân tử Yếu tố VIII và dược chất điều trị bệnh hemophilia A thông thường bao gồm phương pháp điều trị bất kỳ cung cấp cả axit nucleic mã hóa phân tử Yếu tố VIII và phân tử Yếu tố VIII và/hoặc chất làm đông máu không phải Yếu tố VIII (ví dụ, dược chất đi tắt) cho bệnh nhân để làm dịu, làm giảm, hoặc phòng ngừa sự tái phát một hoặc nhiều triệu chứng (ví dụ, các yếu tố lâm sàng) bị gây ra bởi bệnh hemophilia. Thuật ngữ này bao gồm việc sử dụng hợp chất, thuốc, quy trình, hoặc phác đồ bất kỳ chứa axit nucleic mã hóa phân tử Yếu tố VIII, bao gồm dạng được cải biến bất kỳ của Yếu tố VIII, mà là hữu ích để duy trì hoặc cải thiện sức khỏe của cá thể mắc bệnh hemophilia và bao gồm bất kỳ trong số các dược chất được mô tả ở đây.

Thuật ngữ "liều hoặc lượng có tác dụng điều trị" hoặc "liều hoặc lượng đủ để điều trị" hoặc "liều hoặc lượng đủ hoặc hữu hiệu" dùng để chỉ liều gây ra các tác dụng điều trị mà nó được sử dụng cho. Ví dụ, lượng có tác dụng điều trị của thuốc hữu hiệu để điều trị bệnh hemophilia có thể là lượng có khả năng phòng ngừa hoặc làm giảm một hoặc nhiều triệu chứng bị gây ra bởi bệnh hemophilia. Liều chính xác sẽ phụ thuộc vào mục đích của việc điều trị, và sẽ được đánh giá bởi người có hiểu biết trung bình trong lĩnh vực này bằng cách sử dụng các kỹ thuật đã biết (*xem, ví dụ*, Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); và *Remington: The Science and Practice of Pharmacy*, xuất bản lần thứ 20, 2003, Gennaro, Ed., Lippincott, Williams

& Wilkins).

Như được sử dụng ở đây, thuật ngữ "gen" dùng để chỉ một đoạn của phân tử ADN mã hóa cho một chuỗi polypeptit (ví dụ, vùng mã hóa). Trong một số phương án, gen được định vị bởi các vùng ngay trước, sau, và/hoặc xen vào giữa vùng mã hóa mà tham gia vào việc sản sinh ra chuỗi polypeptit (ví dụ, các yếu tố điều hòa như trình tự khởi động, trình tự tăng cường, trình tự polyadenyl hóa, vùng không được dịch mã - 5', vùng không được dịch mã -3', hoặc intron).

Như được sử dụng ở đây, thuật ngữ "các yếu tố điều hòa" dùng để chỉ các trình tự nucleotit, như trình tự khởi động, trình tự tăng cường, trình tự kết thúc, trình tự polyadenyl hóa, intron, v.v... mà gây ra sự biểu hiện của trình tự mã hóa trong tế bào.

Như được sử dụng ở đây, thuật ngữ "trình tự khởi động" dùng để chỉ trình tự nucleotit mà hỗ trợ kiểm soát sự biểu hiện của trình tự mã hóa. Nói chung, các trình tự khởi động được đặt ở đầu 5' của vị trí khởi động dịch mã của gen. Tuy nhiên, trong một số phương án, trình tự khởi động có thể được đặt trong trình tự intron, hoặc đầu 3' của trình tự mã hóa. Trong một số phương án, trình tự khởi động hữu ích cho vectơ liệu pháp gen là thu được từ gen tự nhiên của protein đích (ví dụ, trình tự khởi động Yếu tố VIII). Trong một số phương án, trình tự khởi động hữu ích cho vectơ liệu pháp gen là đặc hiệu để biểu hiện trong một tế bào hoặc mô cụ thể của sinh vật đích (ví dụ, trình tự khởi động đặc hiệu gan). Vẫn trong các phương án khác, một trong nhiều trình tự khởi động đã được mô tả rõ ràng được sử dụng trong vectơ liệu pháp gen đã mô tả ở đây. Ví dụ không giới hạn về các trình tự khởi động đã được mô tả rõ ràng bao gồm trình tự khởi động sớm CMV, trình tự khởi động β-actin, và trình tự khởi động protein liên kết metyl CpG 2 (MeCP2). Trong một số phương án, trình tự khởi động này là trình tự khởi động cơ định, mà điều khiển sự biểu hiện gần như hằng định của protein đích. Trong các phương án khác, trình tự khởi động này là trình tự khởi động cảm ứng, mà điều khiển sự biểu hiện của protein đích đáp ứng với một kích thích cụ thể (ví dụ, tiếp xúc với chất hoặc phương pháp điều trị cụ thể). Để xem lại về việc thiết kế các trình tự khởi động cho liệu pháp gen qua trung gian AAV, xem Gray et al. (Human Gene Therapy 22:1143-53 (2011)), nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn toàn bộ cho tất cả các mục đích.

Như được sử dụng ở đây, thuật ngữ "vectơ" dùng để chỉ phương tiện bất kỳ được sử

dụng để chuyển axit nucleic (ví dụ, mã hóa cấu trúc liệu pháp gen Yếu tố VIII) vào tế bào chủ. Trong một số phương án, vectơ bao gồm đơn vị sao chép, có chức năng sao chép phương tiện này, cùng với axit nucleic đích. Ví dụ không giới hạn về các vectơ hữu ích cho liệu pháp gen bao gồm plasmit, thể thực khuẩn, cosmid, nhiễm sắc thể nhân tạo, và các virut, mà hoạt động như là các đơn vị sao chép tự trị trên in vivo. Trong một số phương án, vectơ là phương tiện virut để đưa axit nucleic đích vào (ví dụ, polynucleotit được biến đổi codon mã hóa biến thể Yếu tố VIII). Nhiều virut nhân chuẩn đã được biến đổi hữu ích cho liệu pháp gen đã được biết trong lĩnh vực này. Ví dụ, các virut liên hợp adeno (AAV) là đặc biệt thích hợp để sử dụng trong liệu pháp gen người vì con người là vật chủ tự nhiên đối với virut, các virut tự nhiên chưa được biết là góp phần vào các bệnh bất kỳ, và các virut này gây ra đáp ứng miễn dịch nhẹ.

Như được sử dụng ở đây, thuật ngữ "đảo CpG" dùng để chỉ vùng bên trong polynucleotit có mật độ các dinucleotit CpG tăng có ý nghĩa thống kê. Như được sử dụng ở đây, vùng của polynucleotit (ví dụ, polynucleotit mã hóa protein Yếu tố VIII được biến đổi codon) là đảo CpG nếu, qua một cửa sổ 200 cặp bazơ: (i) vùng này có hàm lượng GC lớn hơn 50%, và (ii) tỉ lệ của các dinucleotit CpG quan sát được trên các dinucleotit CpG được dự đoán ít nhất là 0,6, như được xác định bằng hệ thức:

$$\frac{N[CpG]*N[\text{độ dài của cửa sổ}]}{N[C]*N[G]} \geq 0{,}6.$$

Để có thông tin thêm về các phương pháp phát hiện đảo CpG, xem Gardiner-Garden M. et al., J Mol Biol., 196(2):261-82 (1987), nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn, toàn bộ, cho tất cả các mục đích.

Như được sử dụng ở đây, thuật ngữ "axit nucleic" dùng để chỉ các deoxyribonucleotit hoặc ribonucleotit và các polyme của chúng ở dạng chuỗi đơn hoặc chuỗi kép, và các phần bổ sung của chúng. Thuật ngữ này bao gồm các axit nucleic chứa các chất tương tự nucleotit đã biết hoặc các gốc khung đã biến đổi hoặc các cầu nối, mà là tổng hợp, xuất hiện tự nhiên, và xuất hiện không tự nhiên, mà có các đặc điểm liên kết tương tự với axit nucleic tham chiếu, và được chuyển hóa theo cách tương tự với các nucleotit tham chiếu. Ví dụ về các chất tương tự này bao gồm, không giới hạn ở, phosphorothioat, phosphoramidat, metyl phosphonat, metyl phosphonat bất đối, 2-O-metyl ribonucleotit, và các peptit-axit nucleic (PNA).

Thuật ngữ "axit amin" dùng để chỉ các axit amin xuất hiện tự nhiên và không tự nhiên, bao gồm các chất tương tự axit amin và các chất giả axit amin mà hoạt động theo cách tương tự với các axit amin xuất hiện tự nhiên. Các axit amin xuất hiện tự nhiên bao gồm các các axit amin được mã hóa bởi mã di truyền, cũng như các axit amin được biến đổi sau đó, ví dụ, hydroxyprolin, y-carboxyglutamat, và O-phosphoserin. Các axit amin xuất hiện tự nhiên có thể bao gồm, ví dụ, các axit amin D và L. Các axit amin được sử dụng ở đây cũng có thể bao gồm các axit amin không tự nhiên. Các chất tương tự axit amin dùng để chỉ các hợp chất có cấu trúc hóa học cơ bản giống như axit amin xuất hiện tự nhiên, nghĩa là, nguyên tử cacbon bất kỳ được liên kết với nguyên tử hydro, nhóm carboxyl, nhóm amino, và nhóm R, ví dụ, homoserin, norleuxin, methionin sulfoxit, hoặc methionin metyl sulfoni. Các chất tương tự này có các nhóm R được biến đổi (ví dụ, norleuxin) hoặc các khung peptit được biến đổi, nhưng vẫn giữ được cấu trúc hóa học cơ bản giống như axit amin xuất hiện tự nhiên. Các chất giả axit amin dùng để chỉ các hợp chất hóa học có cấu trúc khác với cấu trúc hóa học chung của axit amin, nhưng hoạt động theo cách tương tự với axit amin xuất hiện tự nhiên. Các axit amin có thể được gọi ở đây bằng các ký hiệu ba ký tự thường được biết của chúng hoặc bằng các ký hiệu 1 ký tự được đề nghị bởi Hội đồng danh pháp hóa sinh IUPAC-IUB. Tương tự nucleotit có thể được gọi bằng các mã một ký tự đã được chấp nhận phổ biến của chúng.

Các trình tự nucleotit mã hóa các cấu trúc Yếu tố VIII đột biến ở đây có thể là đồng nhất với trình tự mã hóa được đề xuất ở đây hoặc có thể là trình tự mã hóa khác, trình tự mà, do sự dư thừa hoặc thoái hóa mã di truyền, mã hóa các polypeptit giống như các trình tự mã hóa được đề xuất ở đây. Một người có hiểu biết trung bình trong lĩnh vực này sẽ nhận ra là mỗi codon trong axit nucleic (ngoại trừ AUG, mà thường là codon duy nhất cho methionin, và TGG, mà thường là codon duy nhất cho tryptophan) có thể được biến đổi để tạo ra phân tử tương tự về mặt chức năng. Do đó, mỗi biến đổi của axit nucleic mà mã hóa cùng một polypeptit có liên quan mật thiết với mỗi trình tự đã mô tả về sản phẩm biểu hiện, nhưng không phải về cấu trúc liệu pháp gen thực tế.

Đối với các trình tự axit amin, người có hiểu biết trung bình trong lĩnh vực này sẽ nhận ra là từng đột biến thế, đột biến mất đoạn hoặc đột biến bổ sung vào trình tự axit nucleic hoặc trình tự peptit mà làm thay đổi, bổ sung thêm hoặc làm mất một axit amin duy nhất hoặc một tỉ lệ phần trăm nhỏ các axit amin trong trình tự được mã hóa là "biến thể

được biến đổi bảo tồn" trong đó biến đổi này gây ra sự thay thế một axit amin bằng một axit amin tương tự về hóa học. Các bảng đột biến thế bảo tồn cung cấp các axit amin tương tự về mặt chức năng đã được biết rõ trong lĩnh vực này. Các biến thể đã được cải biến bảo tồn này, ngoài ra và không loại trừ các biến thể đa hình, là các dạng tương đương giữa các loài, và các alen theo sáng chế.

Các đột biến thế axit amin bảo tồn tạo ra các axit amin tương tự về mặt chức năng đã được biết rõ trong lĩnh vực này. Tùy theo chức năng của một axit amin cụ thể, ví dụ, các axit amin xúc tác, cấu trúc hoặc quan trọng về mặt không gian, các nhóm axit amin khác nhau có thể được xem là các đột biến thế bảo tồn cho nhau. Bảng 1 cung cấp các nhóm axit amin được xem là các đột biến thế bảo tồn dựa trên điện tích và tính phân cực của axit amin, tính kỵ nước của axit amin, sự bộc lộ bề mặt/bản chất cấu trúc của axit amin, và xu hướng cấu trúc bậc hai của axit amin này.

Bảng 1. Các nhóm đột biến thế axit amin bảo tồn dựa trên chức năng của gốc này trong protein.

Đặc điểm quan trọng	Các nhóm bảo tồn
Điện tích/Tính phân cực	1. H, R, và K 2. D và E 3. C, T, S, G, N, Q, và Y 4. A, P, M, L, I, V, F, và W
Tính kỵ nước	1. D, E, N, Q, R, và K 2. C, S, T, P, G, H, và Y 3. A, M, I, L, V, F, và W
Sự bộc lộ cấu trúc/bề mặt	1. D, E, N, Q, R, và K 2. C, S, T, P, A, G, W, và Y 3. M, I, L, V, và F
Xu hướng cấu trúc bậc hai	1. A, E, Q, H, K, M, L, và R 2. C, T, I, V, F, Y, và W 3. S, G, P, D, và N
Bảo tồn tiến hóa	1. D và E 2. H, K, và R 3. N và Q 4. S và T 5. L, I, và V 6. F, Y, và W 7. A và G 8. M và C

Các thuật ngữ "đồng nhất" hoặc tỉ lệ phần trăm "đồng nhất", trong ngữ cảnh của hai

hoặc nhiều axit nucleic hoặc trình tự peptit, dùng để chỉ hai hoặc nhiều trình tự hoặc tiểu trình tự mà giống nhau hoặc có tỉ lệ phần trăm xác định các gốc axit amin hoặc nucleotit là giống nhau (nghĩa là, đồng nhất khoảng 60%, tốt hơn, nếu là 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, hoặc đồng nhất cao hơn trên một vùng xác định, khi được so sánh và sắp xếp thẳng hàng để có sự giống nhau tối đa trên cửa sổ so sánh hoặc vùng được chỉ định) như được đo bằng cách sử dụng các thuật toán so sánh trình tự BLAST hoặc BLAST 2.0 với các tham số mặc định được mô tả dưới đây, hoặc bằng cách sắp xếp thẳng hàng thủ công và kiểm tra bằng mắt thường.

Như đã biết trong lĩnh vực này, nhiều chương trình khác nhau có thể được sử dụng để xác định một protein (hoặc axit nucleic như được thảo luận dưới đây) có sự đồng nhất hoặc tương đồng về trình tự với trình tự đã biết hay không. Sự đồng nhất và/hoặc tương đồng trình tự được xác định bằng cách sử dụng các kỹ thuật chuẩn đã biết trong lĩnh vực này, bao gồm, nhưng không giới hạn ở thuật toán đồng nhất trình tự khu trú của Smith & Waterman, Adv. Appl. Math., 2:482 (1981), bằng thuật toán sắp xếp thẳng hàng đồng nhất trình tự của Needleman & Wunsch, J. Mol. Biol., 48:443 (1970), bằng cách tra cứu phương pháp tương tự của Pearson & Lipman, Proc. Natl. Acad. Sci. U.S.A., 85:2444 (1988), bằng cách thực hiện các thuật toán này trên máy vi tính (GAP, BESTFIT, FASTA, và TFASTA trong Gói phần mềm Wisconsin Genetics, Genetics Computer Group, 575 Science Drive, Madison, WI), chương trình giải trình tự Best Fit được mô tả bởi Devereux et al., Nucl. Acid Res., 12:387-395 (1984), tốt hơn, nếu sử dụng các cài đặt mặc định hoặc bằng cách kiểm tra. Tốt hơn, nếu tỉ lệ phần trăm đồng nhất được tính bằng FastDB dựa vào các tham số sau đây: điểm phạt bắt cặp nhầm là 1; điểm phạt khoảng trống 1; điểm phạt kích thước khoảng trống là 0,33; và điểm phạt kết hợp là 30, "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc, tất cả các tài liệu này được đưa vào đây bằng cách viện dẫn.

Ví dụ về thuật toán hữu ích là PILEUP. PILEUP tạo ra sự sắp xếp thẳng hàng nhiều trình tự từ nhóm các trình tự liên quan bằng cách sử dụng các sắp xếp thẳng hàng theo cặp, lũy tiến. Cũng có thể lập một cây thể hiện mối quan hệ tạo nhóm được sử dụng để tạo ra sự sắp xếp thẳng hàng. PILEUP sử dụng kỹ thuật đơn giản hóa của phương pháp sắp xếp thẳng hàng lũy tiến của Feng & Doolittle, J. Mol. Evol. 35:351-360 (1987); phương pháp

này là tương tự với phương pháp được mô tả bởi Higgins & Sharp CABIOS 5:151-153 (1989), đều được đưa vào đây bằng cách viện dẫn. Các tham số PILEUP hữu ích bao gồm khối lượng khoảng trống mặc định là 3,00, khối lượng độ dài khoảng trống mặc định là 0,10, và các khoảng trống đầu cùng được đo.

Một ví dụ khác về thuật toán hữu ích là thuật toán BLAST, được mô tả trong: Altschul et al., J. Mol. Biol. 215, 403-410, (1990); Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997); và Karlin et al., Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787 (1993), đều được đưa vào đây bằng cách viện dẫn. Chương trình BLAST đặc biệt hữu ích là chương trình WU-BLAST-2 mà thu được từ Altschul et al., Methods in Enzymology, 266:460-480 (1996); http://blast.wustl/edu/blast/ README.html]. WU-BLAST-2 sử dụng một số tham số tra cứu, hầu hết trong số đó được thiết lập đến các giá trị mặc định. Các tham số điều chỉnh được được thiết lập với các giá trị sau đây: khoảng xếp chồng =1, tỉ lệ xếp chồng = 0,125, ngưỡng từ (T) = 11. Các tham số HSP S và HSP S2 là các giá trị động học và được thiết lập bởi chính chương trình này phụ thuộc vào thành phần của trình tự cụ thể và thành phần của cơ sở dữ liệu cụ thể mà dựa vào đó trình tự cần quan tâm được tra cứu; tuy nhiên, các giá trị này có thể được điều chỉnh để làm tăng độ nhạy.

Thuật toán hữu ích khác là BLAST có khoảng trống, như được báo cáo bởi Altschul et al., Nucl. Acids Res., 25:3389-3402, được đưa vào đây bằng cách viện dẫn. BLAST có khoảng trống sử dụng các điểm số thay thế BLOSUM-62; tham số ngưỡng T được đặt là 9; phương pháp hai hit (hit: chuỗi con có điểm số lớn hơn một giá trị ngưỡng T) để khởi động sự mở rộng không tạo khoảng trống; độ dài khoảng trống điện tích của chi phí k a là 10+k; Xu được đặt ở 16, và Xg được đặt ở 40 cho giai đoạn tra cứu cơ sở dữ liệu và ở 67 cho giai đoạn tính toán kết quả của thuật toán. Các bắt cặp có khoảng trống được bắt đầu bằng điểm số tương ứng với ~22 bit.

Giá trị % đồng nhất trình tự axit amin được xác định bằng số lượng các gốc bắt cặp đồng nhất chia cho tổng số gốc của trình tự "dài hơn" trong vùng được bắt cặp. Trình tự "dài hơn" là trình tự có hầu hết các gốc thực tế nằm trong vùng được bắt cặp (các khoảng trống được đưa vào bằng WU-Blast-2 để tối đa hóa điểm số bắt cặp được bỏ qua). Theo cách tương tự, "tỉ lệ phần trăm (%) đồng nhất trình tự axit nucleic" đối với trình tự mã hóa của các polypeptit được xác định được định nghĩa là tỉ lệ phần trăm các gốc nucleotit trong trình tự ứng viên giống với các gốc nucleotit trong trình tự mã hóa của protein chu kỳ tế

bào. Phương pháp được ưu tiên hơn sử dụng môđun BLASTN của WU-BLAST-2 được thiết lập đến các tham số mặc định, với khoảng xếp chồng và tỉ lệ xếp chồng được đặt lần lượt ở 1 và 0,125.

Sự bắt cặp có thể bao gồm việc đưa các khoảng trống vào các trình tự cần được sắp xếp thẳng hàng. Ngoài ra, đối với các trình tự chứa nhiều hoặc ít axit amin hơn so với protein được mã hóa bởi trình tự trong Fig. 2 (SEQ ID NO:1), được hiểu là, trong một phương án, tỉ lệ phần trăm đồng nhất trình tự sẽ được xác định dựa vào số lượng các axit amin hoặc nucleotit đồng nhất so với tổng số axit amin hoặc nucleotit. Do đó, ví dụ, sự đồng nhất trình tự của các trình tự ngắn hơn so với các trình tự được thể hiện trong Fig. 2 (SEQ ID NO:1), như được thảo luận dưới đây, sẽ được xác định bằng cách sử dụng số lượng nucleotit trong trình tự ngắn hơn, trong một phương án. Trong phép tính tỉ lệ phần trăm đồng nhất, khối lượng tương đối không được gán cho nhiều biểu hiện khác nhau của sự biến đổi trình tự, như sự gắn xen, mất đoạn, thay thế, v.v...

Trong một phương án, chỉ có sự đồng nhất được tính điểm dương (+1) và tất cả các dạng của biến đổi trình tự bao gồm các khoảng trống được gán giá trị "0", điều này tránh được nhu cầu về thang đo hoặc các tham số như được mô tả dưới đây cho các phép tính tương đồng trình tự. Tỉ lệ phần trăm đồng nhất trình tự có thể được tính toán, ví dụ, bằng cách chia số lượng các gốc bắt cặp giống nhau cho tổng số gốc của trình tự "ngắn hơn" trong vùng được bắt cặp và nhân với 100. Trình tự "dài hơn" là trình tự có hầu hết các gốc thực tế nằm trong vùng được bắt cặp.

Thuật ngữ "biến thể alen" dùng để chỉ các dạng đa hình của gen ở locut di truyền cụ thể, cũng như các cADN có nguồn gốc từ các bản sao mARN của các gen này, và các polypeptit được mã hóa bởi chúng. Thuật ngữ "codon động vật có vú được ưu tiên hơn" dùng để chỉ tập hợp con các codon trong tập hợp các codon mã hóa axit amin mà được sử dụng thường xuyên nhất trong các protein được biểu hiện trong tế bào động vật có vú như được chọn từ danh mục sau đây: Gly (GGC, GGG); Glu (GAG); Asp (GAC); Val (GTG, GTC); Ala (GCC, GCT); Ser (AGC, TCC); Lys (AAG); Asn (AAC); Met (ATG); Ile (ATC); Thr (ACC); Trp (TGG); Cys (TGC); Tyr (TAT, TAC); Leu (CTG); Phe (TTC); Arg (CGC, AGG, AGA); Gln (CAG); His (CAC); và Pro (CCC).

Như được sử dụng ở đây, thuật ngữ được biến đổi codon dùng để chỉ trình tự

polynucleotit mã hóa polypeptit (ví dụ, protein biến thể Yếu tố VIII), trong đó ít nhất một codon của polynucleotit tự nhiên mã hóa polypeptit này đã được thay đổi để cải thiện đặc tính của trình tự polynucleotit này. Trong một số phương án, đặc tính được cải thiện này thúc đẩy sự phiên mã tăng của mARN mã hóa cho polypeptit này, gia tăng độ ổn định của mARN này (ví dụ, thời gian bán thải của mARN được cải thiện), gia tăng sự dịch mã của polypeptit này, và/hoặc gia tăng sự gói polynucleotit vào trong vectơ. Ví dụ không giới hạn về các biến đổi có thể được sử dụng để đạt được các đặc tính được cải thiện bao gồm thay đổi việc sử dụng và/hoặc phân bố của các codon cho các axit amin cụ thể, điều chỉnh hàm lượng GC toàn bộ và/hoặc khu trú, loại bỏ các trình tự giàu AT, loại bỏ các yếu tố trình tự lặp lại, điều chỉnh hàm lượng dinucleotit CpG toàn bộ và/hoặc khu trú, loại bỏ các yếu tố điều hòa chưa biết (ví dụ, các yếu tố hộp TATA và hộp CCAAT), loại bỏ vị trí nối intron/exon, cải thiện các trình tự điều hòa (ví dụ, đưa trình tự liên ứng Kozak vào), và loại bỏ các yếu tố trình tự có khả năng tạo ra cấu trúc bậc hai (ví dụ, vòng thân) trong mARN đã được phiên mã.

Như được thảo luận ở đây, có nhiều danh pháp khác nhau để chỉ các thành phần của phần mô tả ở đây. "Số CS" (ví dụ, "CS04", "CS01", "CS23", v.v...) dùng để chỉ các polynucleotit được biến đổi codon mã hóa polypeptit FVIII và/hoặc các polypeptit được mã hóa, bao gồm các biến thể. Ví dụ, CS01-FL dùng để chỉ trình tự polynucleotit hoặc trình tự axit amin CS01 được biến đổi codon có độ dài đầy đủ (đôi khi được gọi ở đây là "CS01-FL-AA" đối với trình tự axit amin và "CS01-FL-NA" đối với trình tự axit nucleic) được mã hóa bởi trình tự polynucleotit CS01. Tương tự, "CS01-LC" dùng để chỉ trình tự axit nucleic được biến đổi codon ("CS01-LC-NA") mã hóa chuỗi nhẹ của polypeptit FVIII hoặc trình tự axit amin (đôi khi còn được gọi ở đây là "CS01-LC-AA") của chuỗi nhẹ FVIII được mã hóa bởi trình tự polynucleotit CS01 này. Tương tự, CS01-HC, CS01-HC-AA và CS01-HC-NA là giống nhau đối với chuỗi nặng FVIII. Như được hiểu rõ bởi người có hiểu biết trung bình trong lĩnh vực này, đối với các cấu trúc như CS01, CS04, CS23, v.v..., mà chỉ được biến đổi codon (ví dụ, chúng không chứa thêm đột biến thế axit amin so với Refacto), các trình tự axit amin sẽ là giống nhau, vì các trình tự axit amin không bị biến đổi bởi sự tối ưu hóa codon. Do đó, các cấu trúc trình tự của phần mô tả này bao gồm, nhưng không giới hạn ở, CS01-FL-NA, CS01-FL-AA, CS01-LC-NA, CS01-LC-AA, CS01-HC-AA, CS01-HC-NA, CS04-FL-NA, CS04-FL-AA, CS04-LC-NA, CS04-LC-AA, CS04-HC-AA, CS04-HC-NA, CS23-FL-NA, CS23-FL-AA, CS23-LC-NA, CS23-

LC-AA, CS23-HC-AA và CS23-HC-NA.

Danh pháp này cũng được áp dụng cho các peptit glycosyl hóa như được thể hiện trong Fig. 13, sao cho "NGA1-AA" dùng để chỉ trình tự axit amin và NGA1-NA dùng để chỉ trình tự axit nucleic.

Phần mô tả này cũng bao gồm các biến thể Yếu tố VIII mới khác, như được mô tả dưới đây, với danh pháp thích hợp.

Các biến thể Yếu tố VIII được biến đổi codon

Trong một số phương án, sáng chế đề xuất các polynucleotit được biến đổi codon mã hóa các biến thể Yếu tố VIII. Các polynucleotit được biến đổi codon này gây ra sự biểu hiện được cải thiện rõ rệt của Yếu tố VIII khi được sử dụng trong cấu trúc liệu pháp gen dựa trên AAV. Polynucleotit được biến đổi codon này cũng thể hiện tác dụng gói virion AAV được cải thiện, so với các cấu trúc được tối ưu hóa codon thông thường. Như được thể hiện trong Ví dụ 2 và Ví dụ 4, tác giả sáng chế đã đạt được các lợi ích này nhờ việc tìm ra ba polynucleotit được biến đổi codon (CS01-FL-NA, CS04-FL-NA, và CS23-FL-NA) mã hóa polypeptit Yếu tố VIII có các chuỗi nặng và chuỗi nhẹ Yếu tố VIII kiểu hoang của người, và cầu nối được thế vùng chức năng B, 14 axit amin, ngắn (cầu nối "SQ") chứa vị trí cắt furin để tạo thuận lợi cho sự thành thục của protein FVIIIa hoạt động trên in vivo. Như đã chứng minh thêm trong Ví dụ 4, việc đưa các tổ hợp khác nhau của các đột biến axit amin F328S, X5, và X1 vào phân tử Yếu tố VIII được mã hóa làm tăng thêm sự biểu hiện in vivo của hoạt tính Yếu tố VIII.

Trong một phương án, polynucleotit được biến đổi codon được đề xuất ở đây có trình tự nucleotit có sự đồng nhất trình tự cao với ít nhất là các trình tự trong CS01, CS04, hoặc CS23 (SEQ ID NO 13, 1, và 20, tương ứng) mã hóa chuỗi nặng Yếu tố VIII và chuỗi nhẹ Yếu tố VIII. Như đã biết trong lĩnh vực này, vùng chức năng B của Yếu tố VIII là không cần thiết cho hoạt tính in vivo. Do đó, trong một số phương án, các polynucleotit được biến đổi codon được đề xuất ở đây thiếu hoàn toàn vùng chức năng B Yếu tố VIII. Trong một số phương án, vùng chức năng B Yếu tố VIII tự nhiên được thay thế bằng cầu nối axit amin ngắn chứa vị trí cắt furin, ví dụ, cầu nối "SQ" chứa các axit amin 760-773 của các cấu trúc CS01, CS04, hoặc CS23 (SEQ ID NOS 2, 2, và 21, tương ứng). Cầu nối "SQ" còn được gọi là BDLO04, (-AA cho trình tự axit amin và -NA cho trình tự nucleotit

được thể hiện trong Fig. 6).

Trong một phương án, chuỗi nặng và chuỗi nhẹ Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon là chuỗi nặng và chuỗi nhẹ Yếu tố VIII của người, tương ứng. Trong các phương án khác, chuỗi nặng và chuỗi nhẹ Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon là các trình tự chuỗi nặng và chuỗi nhẹ từ một động vật có vú khác (ví dụ, Yếu tố VIII của lợn). Vẫn trong các phương án khác, chuỗi nặng và chuỗi nhẹ Yếu tố VIII là các chuỗi nặng và chuỗi nhẹ khảm (ví dụ, sự kết hợp của trình tự người và trình tự động vật có vú thứ hai). Vẫn trong các phương án khác, chuỗi nặng và chuỗi nhẹ Yếu tố VIII là dạng được làm tương thích với người của các chuỗi nặng và chuỗi nhẹ từ một động vật có vú khác, ví dụ, các trình tự chuỗi nặng và chuỗi nhẹ từ một động vật có vú khác mà trong đó các gốc người được thay thế ở các vị trí được chọn để làm giảm khả năng gây miễn dịch của peptit tạo thành khi được sử dụng cho người.

Hàm lượng GC của gen người thay đổi trong khoảng rộng, từ nhỏ hơn 25% đến lớn hơn 90%. Tuy nhiên, nói chung, các gen người có hàm lượng GC cao hơn được biểu hiện ở mức độ cao hơn. Ví dụ, Kudla et al. (PLoS Biol., 4(6):80 (2006)) chứng minh là sự gia tăng hàm lượng GC của gen làm tăng sự biểu hiện của polypeptit được mã hóa, chủ yếu do làm tăng phiên mã và đem lại mức trạng thái ổn định cao hơn cho bản sao mARN. Nói chung, hàm lượng GC mong muốn của cấu trúc gen được tối ưu hóa codon là bằng hoặc lớn hơn 60%. Tuy nhiên, các bộ gen AAV tự nhiên có hàm lượng GC xấp xỉ 56%.

Do đó, trong một số phương án, polynucleotit được biến đổi codon được đề xuất ở đây có hàm lượng CG phù hợp chặt chẽ hơn với hàm lượng GC của các virion AAV tự nhiên (ví dụ, xấp xỉ 56% GC), mà nhỏ hơn so với các hàm lượng CG được ưu tiên hơn của các polynucleotit được tối ưu hóa codon thông thường để biểu hiện trong tế bào động vật có vú (ví dụ, tại hoặc trên 60% GC). Như được nêu vắn tắt trong Ví dụ 1, CS04-FL-NA (SEQ ID NO: 1), mà có hàm lượng GC là khoảng 56%, có tác dụng gói virion cải thiện so với các trình tự mã hóa được biến đổi codon tương tự có hàm lượng GC cao hơn.

Do đó, trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là nhỏ hơn 60%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là nhỏ hơn 59%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit

được biến đổi codon mã hóa polypeptit Yếu tố VIII là nhỏ hơn 58%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là nhỏ hơn 57%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là không lớn hơn 56%.

Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 54% đến 59%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 55% đến 59%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 56% đến 59%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 54% đến 58%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 55% đến 58%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 56% đến 58%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 54% đến 57%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 55% đến 57%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 56% đến 57%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 54% đến 56%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là từ 55% đến 56%.

Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là 56±0,5%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là 56±0,4%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là 56±0,3%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là 56±0,2%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là 56±0,1%. Trong một số phương án, hàm lượng GC toàn phần của polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII là

56%.

Các đột biến thế axit amin Yếu tố VIII

Để làm tăng thêm hiệu quả của sự biểu hiện dựa trên vectơ AAV của các cấu trúc Yếu tố VIII được mô tả ở đây, các đột biến thế axit amin đã được biết là cải thiện sự tiết, làm tăng hoạt tính đặc hiệu, và/hoặc tăng cường độ ổn định của Yếu tố VIII được đưa vào thêm, trong một số phương án. Nhiều biến thể tiềm năng đã được xác định là làm tăng mức hoạt tính FVIII trong huyết tương ở liều vectơ đã cho. Các biến thể này bao gồm các biến thể có peptit tín hiệu hiệu quả hơn, các đột biến thế axit amin ngăn cản các tương tác BiP, các đột biến thế axit amin tương tự với các gen trực giao Yếu tố VIII được tiết hiệu quả hơn (ví dụ, Yếu tố VIII của lợn), các biến thể Yếu tố VIII chuỗi đơn, và các đột biến thế axit amin làm ổn định Yếu tố VIII và/hoặc làm giảm sự phân ly tiểu đơn vị.

Đột biến của các gốc A108, R121, và L2302 (SPE), nằm ở bề mặt phân cách giữa các vùng chức năng A1 và C2, làm tăng độ ổn định của Yếu tố VIII. Ví dụ, đột biến thế axit amin A108I đưa vào gốc kỵ nước mà lấp đầy không gian giữa các vùng chức năng tốt hơn, làm ổn định tương tác. Tương tự, đột biến thế axit amin kép R121C/L2302C (SPE) đưa vào liên kết disulfua nối các vùng chức năng A1-C2, làm ổn định thêm tương tác. Cùng với nhau, cả ba đột biến thế axit amin làm tăng độ ổn định nhiệt của Yếu tố VIII từ 3 đến 4 lần. Để xem lại, xem Wakabayashi et al., J Biol Chem. 286(29):25748-55 (2011) và Wakabayashi et al., Thromb Haemost. 10(3):492-95 (2012). Do đó, trong một số phương án, polypeptit Yếu tố VIII được mã hóa bao gồm các đột biến thế axit amin A108I và/hoặc R121C/L2302C.

Đột biến E113 (SPE), nằm trong vùng chức năng liên kết canxi của Yếu tố VIII, làm tăng hoạt tính đông máu FVIII đặc hiệu. Ví dụ, E113A có thể làm tăng sự tạo thành FXase nhờ ái lực FVIII đối với Yếu tố IXa được tăng cường. Cụ thể là, đột biến thế axit amin E113A làm tăng hoạt tính đông máu FVIII đặc hiệu thêm hai lần và làm tăng ái lực đối với Yếu tố IXa thêm bốn lần (Biochemistry, 41:8485 (2002); J. Biol. Chem., 279:12677 (2004); và Biochemistry, 44:10298 (2005)). Do đó, trong một số phương án, polypeptit Yếu tố VIII được mã hóa bao gồm đột biến thế axit amin E113A.

Đột biến thế một hoặc nhiều gốc axit amin xung quanh vị trí cắt APC Yếu tố VIII (các gốc 331-341 (SPE)) làm giảm sự bất hoạt Yếu tố VIIIa bởi protein C được hoạt hóa,

mà không ảnh hưởng đến hoạt tính FVIII. Ví dụ, các đột biến thế axit amin PQL333-335VDQ (SPE) làm giảm sự bất hoạt Yếu tố VIII 16 lần. Tương tự, các đột biến thế axit amin MKN336-339GNQ làm giảm sự bất hoạt Yếu tố VIII 9 lần. Khi được kết hợp, hai đột biến thế axit amin bộ ba (ví dụ, PQLRMKN333-339VDQRGNQ) (SEQ ID NO 34 và 35, tương ứng) làm giảm sự bất hoạt Yếu tố VIII 100 lần (J. Biol. Chem., 282:20264 (2007). Do đó, trong một số phương án, polypeptit Yếu tố VIII được mã hóa bao gồm các đột biến thế axit amin PQL333-335VDQ và/hoặc MKN337-339GNQ (SPE).

Các đột biến nằm trong bề mặt phân cách vùng chức năng A2 cũng làm tăng độ ổn định của Yếu tố VIII. Cụ thể là, việc gây đột biến các gốc tích điện trong mặt phân cách vùng chức năng A1-A2 và A2-A3 làm tăng độ ổn định và khả năng duy trì tiểu đơn vị A2 trong Yếu tố VIIIa. Ví dụ, đột biến D519, E665, và E1984 thành V hoặc A tạo ra độ ổn định tăng đến 2 lần trong Yếu tố VIII và độ ổn định tăng đến 5 lần trong Yếu tố VIIIa. Cụ thể là, các đột biến thế axit amin D519A/E665V gây ra sự tăng độ ổn định 3 lần; các đột biến thế axit amin D519V/E665V gây ra sự tăng độ ổn định 2 lần, làm giảm sự phân ly A2 8 lần, và làm tăng khả năng tạo thành thrombin từ 2 đến 4 lần; các đột biến thế axit amin D519V/E1984A gây ra sự tăng độ ổn định 2 lần; và các đột biến thế axit amin D519V/E665V/E1984A gây ra sự tăng độ ổn định 2 lần (Blood 112:2761-69 (2008); J. Thromb. Haemost., 7:438-44 (2009)). Do đó, trong một số phương án, polypeptit Yếu tố VIII được mã hóa bao gồm một hoặc nhiều trong số các đột biến thế axit amin D519A/V, E665A/V, và E1984A/V.

Đặc biệt thích hợp theo sáng chế là nhiều đột biến đặc hiệu mà có thể được bao gồm riêng biệt hoặc kết hợp với các biến thể khác được mô tả ở đây. Các biến thể này được mã hóa dưới dạng các tập hợp như dưới đây: "m1" dùng để chỉ một thay đổi axit amin duy nhất, "m2" là tập hợp của 5 biến thể axit amin, "m3" là sự kết hợp của đột biến mất đoạn 7 axit amin và đột biến gắn xen sáu axit amin mà kéo dài mối nối giữa cầu nối polypeptit và chuỗi nặng, "m4" là sự kết hợp của đột biến duy nhất m1 và đột biến kép m5, và "m5" là tập hợp của hai đột biến loại bỏ xystein. Các đột biến này được mô tả dưới đây. Các đột biến này có thể được bao gồm trong cấu trúc cụ thể bất kỳ, một mình hoặc kết hợp với các biến thể khác, và do đó chúng được mã hóa. Ví dụ, "m23" là sự kết hợp của các biến thể m2 và m3 lên giàn cụ thể, như được phác họa ở đây; do đó "CS01m23-FL-NA" hoặc "CS01-FL-NAm23" dùng để chỉ trình tự polynucleotit được biến đổi codon CS01 với các

nucleotit mã hóa các đột biến m2 và m3 được bao gồm, và "CS01m23-FL-AA" hoặc "CS01-FL-AAm23" dùng để chỉ trình tự axit amin. Vì CS01 được biến đổi codon nhưng không làm thay đổi trình tự axit amin của Refacto, các biến đổi này có thể được cho là các đột biến trên mức axit amin so với trình tự axit amin Refacto CS01-FL-AA (SEQ ID NO: 2).

Trong nhiều phương án, các polypeptit theo sáng chế được tạo ra với biến thể "m1" được bao gồm. Các đột biến trong phiến β kỵ nước 11 axit amin trong vùng chức năng A1, tương tác với BiP, làm tăng sự tiết Yếu tố VIII. Ví dụ, đột biến thế axit amin F328S (SPI, F309S SPE) trong túi này làm tăng sự tiết Yếu tố VIII 3 lần. Biến thể F328S được gọi ở đây là đột biến "m1" và nằm trong chuỗi nặng. Như được mô tả ở đây, việc đánh số biến thể có thể được thực hiện có bao gồm peptit tín hiệu, "bao gồm peptit tín hiệu", hoặc "SPI", hoặc bắt đầu từ trình tự protein cuối được xử lý, "không bao gồm peptit tín hiệu", hoặc "SPE". Do đó, bằng cách sử dụng cách đánh số SPI, đột biến F328S là giống với đột biến F309 SPE. Nói chung, phần mô tả này sử dụng cách đánh số SPI, nhưng như được hiểu rõ bởi người có hiểu biết trung bình trong lĩnh vực này, cả hai hệ thống đánh số đều dẫn đến (các) đột biến giống nhau.

Do đó, được bao gồm trong sáng chế là các polypeptit chứa đột biến m1, bao gồm CS01-FL-AAm1, CS01-HC-AAm1, CS04-FL-AAm1, CS04-HC-AAm1 CS23-FL-AAm1, CS23-HC-AAm1, CS40-FL-AAm1 và CS40-HC-AAm1 (tất cả đều mã hóa các trình tự protein giống nhau tương ứng).

Ngoài ra, được bao gồm trong sáng chế không chỉ là các trình tự polypeptit chứa đột biến m1, mà cả các trình tự polynucleotit được biến đổi codon mã hóa các protein có đột biến m1, như CS01-FL-NAm1, CS01-HC-NAm1, CS04-FL-NAm1, CS04-HC-NAm1, CS23-FL-NAm1, CS23-HC-NA-m1, CS40-FL-NAm1 và CS40-HC-NAm1.

Trong nhiều phương án, các polypeptit theo sáng chế được tạo ra với tập hợp biến thể "m2" được bao gồm, đó là các đột biến I105V/A127S/G151K/M166T/L171P (đánh số SPI; (đánh số SPE lần lượt là V86I/S108A/K132G/T147M/P152L). Tập hợp đột biến m2 dựa trên thực tế là sự thay thế các axit amin 82-176 của lợn cho các axit amin người tương ứng trong cấu trúc liệu pháp gen bị mất vùng chức năng B làm tăng hoạt tính Yếu tố VIII khi được biểu hiện trong các tế bào HEK293 (W. Xiao, communication). *Id.* Đột biến

ngược từng axit amin lợn vào cấu trúc BDD-FVIII người xác định được năm axit amin trong vùng chức năng A1 góp phần vào hiện tượng này: I105V, A127S, G151K, M166T, và L171P (SPI). Việc đưa tổ hợp các đột biến này vào cấu trúc người lặp lại hoạt tính cải thiện của đột biến thế lớn hơn ở lợn. *Id.* Do đó, trong một số phương án, các polypeptit Yếu tố VIII được mã hóa chứa một hoặc nhiều đột biến thế axit amin được chọn từ I105V, A127S, G151K, M166T, và L171P, với toàn bộ tập hợp 5 axit amin, m2, hữu ích trong nhiều phương án. Như đối với đột biến m1, các biến thể m2 nằm trong chuỗi nặng, và do đó sáng chế bao gồm các polypeptit chứa đột biến m2, bao gồm CS01-FL-AAm2, CS01-HC-AAm2, CS04-FL-AAm2, CS04-HC-AAm2, CS23-FL-AAm2, CS23-HC-AAm2, CS40-FL-AAm2 và CS40-HC-AAm2 (tất cả đều mã hóa các trình tự protein giống nhau tương ứng).

Ngoài ra, được bao gồm trong sáng chế không chỉ là các trình tự polypeptit chứa đột biến m2, mà cả các trình tự polynucleotit được biến đổi codon mã hóa các protein có đột biến m2, như CS01-FL-NAm2, CS01-HC-NAm2, CS04-FL-NAm2, CS04-HC-NAm2, CS23-FL-NAm2, CS23-HC-NA-m2, CS40-FL-NAm2 và CS40-HC-NAm2.

Trong các phương án khác, các polypeptit và các polynucleotit theo sáng chế bao gồm các đột biến m3. m3 là đột biến thay thế bảy axit amin cho sáu axit amin qua bề mặt phân cách HC-vùng chức năng B mà đưa vị trí glycosyl hóa bổ sung đến gần với bề mặt phân cách này. Do đó, trong một số phương án, m3 là đột biến làm mất các axit amin AIEPRSF755-761 và gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19) (ví dụ, AIEPRSF755-761TTYVNRSL) ("TTYVNRSL" được mô tả là SEQ ID NO: 33). Các gốc AIEPR755-759, so với SEQ ID NO: 19, nằm ở đầu cùng của chuỗi nặng, trong khi các gốc S760 và F761 nằm trong vùng chức năng B. Trong một số phương án, trong đó vùng chức năng B của FVIII bị mất, bị cắt, hoặc bị thay thế, các gốc S760 và F761 không thể có mặt trong trình tự axit amin cơ bản đang được gây đột biến. Do đó, trong một số phương án, m3 là đột biến mất đoạn của các axit amin AIEPR755-759 và đột biến gắn xen các axit amin TTYVNRSL (SEQ ID NO: 33) sau N754, so với FVIII-FL-AA (SEQ ID NO: 19) (ví dụ, AIEPR755-759TTYVNRSL ("TTYVNRSL" được mô tả là SEQ ID NO: 33).

Các biến thể m3 nằm trong đoạn nối giữa chuỗi nặng và vùng chức năng B, và do đó sáng chế bao gồm các polypeptit chứa đột biến m3, bao gồm CS01-FL-AAm3, CS01-

HC-AAm3, CS04-FL-AAm3, CS04-HC-AAm3, CS23-FL-AAm3, CS23-HC-AAm3, CS40-FL-AAm3 và CS40-HC-AAm3 (tất cả đều mã hóa các trình tự protein giống nhau tương ứng).

Ngoài ra, được bao gồm trong sáng chế không chỉ là các trình tự polypeptit chứa đột biến m3, mà cả các trình tự polynucleotit được biến đổi codon mã hóa các protein có đột biến m3, như CS01-FL-NAm3, CS01-HC-NAm3, CS04-FL-NAm3, CS04-HC-NAm3, CS23-FL-NAm3, CS23-HC-NA-m3, CS40-FL-NAm3 và CS40-HC-NAm3.

Trong các phương án khác, các polypeptit và polynucleotit theo sáng chế bao gồm các đột biến m4. Việc loại bỏ liên kết disulfua C1899-C1903 trong Yếu tố VIII cũng làm tăng sự tiết. Ngoài ra, tác dụng làm tăng tiết Yếu tố VIII là cộng thêm đối với tổ hợp của các đột biến thế axit amin F328S (SPI, F309S SPE) và C1918G/C1922G (Miao et al., Blood, 103:3412-19 (2004); Selvaraj et al., J. Thromb. Haemost., 10:107-15 (2012)). Do đó, trong một số phương án, polypeptit Yếu tố VIII được mã hóa chứa các đột biến m4, mà là các đột biến thế axit amin F328S (SPI, F309S SPE) và C1918G/C1922G (SPI). Vì biến thể F328S nằm trong chuỗi nặng và hai biến thể xystein nằm trong chuỗi nhẹ, các trình tự polypeptit chứa đột biến m4 là CS01-FL-AAm4, CS01-HC-AAm4, CS01-LC-AAm4, CS04-FL-AAm4, CS04-HC-AAm4, CS04-LC-AAm4, CS23-FL-AAm4, CS23-HC-AAm4 và CS23-LC-AAm4.

Ngoài ra, được bao gồm trong sáng chế không chỉ là các trình tự polypeptit chứa đột biến m4, mà còn bao gồm cả các trình tự polynucleotit được biến đổi codon mã hóa các protein có đột biến m4, như CS01-FL-NAm4, CS01-HC-NAm4, CS01-LC-NAm4, CS04-FL-NAm4, CS04-HC-NAm4, CS04-LC-NAm4, CS23-FL-NAm4, CS23-HC-NAm4, CS23-LC-NAm4, CS40-FL-NA-m4, CS40-HC-NA-m4 và CS40-LC-NA-m4.

Trong các phương án khác, các polypeptit và polynucleotit theo sáng chế bao gồm các đột biến m5. Như trên đây, việc loại bỏ liên kết disulfua C1899-C1903 trong Yếu tố VIII cũng làm tăng sự tiết. Các đột biến thế axit amin C1918G/C1922G (SPI), nằm trong chuỗi nhẹ, được gọi ở đây là tập hợp đột biến m5.

Các biến thể m5 nằm trong chuỗi nhẹ, và do đó sáng chế bao gồm các polypeptit chứa đột biến m5, bao gồm CS01-FL-AAm5, CS01-LC-AAm5, CS04-FL-AAm5, CS04-LC-AAm5, CS23-FL-AAm5, CS23-LC-AAm5, CS40-FL-AAm5 và CS40-LC-AAm5 (tất

cả đều mã hóa các trình tự protein giống nhau tương ứng).

Ngoài ra, được bao gồm trong sáng chế không chỉ là các trình tự polypeptit chứa đột biến m5, mà còn bao gồm cả các trình tự polynucleotit được biến đổi codon mã hóa các protein có đột biến m5, như CS01-FL-NAm5, CS01-LC-NAm5, CS04-FL-NAm5, CS04-LC-NAm5, CS23-FL-NA-m5, CS23-LC-NA-m5, CS40-FL-NA-m5 và CS40-LC-NA-m5.

Ngoài các cấu trúc đặc hiệu (cả axit amin và axit nucleic) mà bao gồm m1, m2, m3, m4 và m5 đơn lẻ, các tổ hợp của các tập hợp đột biến này có thể được tạo ra như đã đề cập ở đây. Như được chú thích ở đây, các tổ hợp được đề cập là "m12", là tổ hợp của các tập hợp m1 và m2, hoặc "m123" là tổ hợp của các tập hợp m1, m2 và m3. Do đó, được bao gồm trong sáng chế là các tổ hợp kép bao gồm m12, m13, m14, m15, m23, m24, m25, m34, m35 và m45. Cũng được bao gồm là các tổ hợp bộ ba, m123, m124, m125, m234, m235 và m345. Được bao gồm thêm là các tổ hợp bộ bốn, m1234, m1235, m1345 và tổ hợp m12345.

Được đặc biệt quan tâm trong một số phương án là các tập hợp đột biến sau đây: m1, m2, m3 và m4, m23, m123, và m234.

Các cầu nối được thay thế vùng chức năng B Yếu tố VIII

Trong một số phương án, cầu nối giữa chuỗi nặng và chuỗi nhẹ FVIII (ví dụ, vùng chức năng B trong Yếu tố VIII kiểu hoang) được biến đổi thêm. Do các ràng buộc về kích thước của khả năng gói AAV, các biến thể bị mất, bị cắt vùng chức năng B và/hoặc các biến thể được thế cầu nối nên cải thiện tác dụng của cấu trúc liệu pháp gen FVIII. Cầu nối được thế vùng chức năng B được sử dụng phổ biến nhất là của SQ FVIII, mà chỉ giữ lại 14 axit amin của vùng chức năng B làm trình tự cầu nối. Một biến thể khác của yếu tố VIII của lợn ("OBI-1", được mô tả trong Patent Mỹ số 6,458,563) được biểu hiện tốt trong các tế bào CHO, và có cầu nối dài hơn một chút là 24 axit amin. Trong một số phương án, các cấu trúc Yếu tố VIII được mã hóa bởi các polynucleotit được biến đổi codon đã mô tả ở đây bao gồm trình tự cầu nối vùng chức năng B typ SQ. Trong một số phương án, các cấu trúc Yếu tố VIII được mã hóa bởi các polynucleotit được biến đổi codon đã mô tả ở đây bao gồm trình tự cầu nối vùng chức năng B typ OBI-1.

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao

gồm cầu nối vùng chức năng B typ SQ, bao gồm các axit amin 760-762/1657-1667 của vùng chức năng B Yếu tố VIII người kiểu hoang (FVIII-FL-AA; SEQ ID NO: 19) (Sandberg et al. Thromb. Haemost. 85:93 (2001)). Trong một số phương án, cầu nối vùng chức năng B typ SQ có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ SQ có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ SQ. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao gồm cầu nối vùng chức năng B typ Greengene, bao gồm các axit amin 760/1582-1667 của vùng chức năng B Yếu tố VIII người kiểu hoang (FVIII-FL-AA; SEQ ID NO: 19) (Oh et al., Biotechnol. Prog., 17:1999 (2001)). Trong một số phương án, cầu nối vùng chức năng B typ Greengene có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ Greengene có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ Greengene. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao gồm cầu nối vùng chức năng B typ SQ kéo dài (SFSQNPPVLKRHQR; BDL-SQ-AA; SEQ ID NO: 30), bao gồm các axit amin 760-769/1657-1667 của vùng chức năng B Yếu tố VIII người kiểu hoang (FVIII-FL-AA; SEQ ID NO: 19) (Thim et al., Haemophilia, 16:349 (2010)). Trong một số phương án, cầu nối vùng chức năng B typ SQ kéo dài có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ SQ kéo dài có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ SQ kéo dài. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao

gồm cầu nối vùng chức năng B typ OBI-1 lợn, bao gồm các axit amin SFAQNSRPPSASAPKPPVLRRHQR (SEQ ID NO: 31) từ vùng chức năng B Yếu tố VIII lợn kiểu hoang (Toschi et al., Curr. Opin. Mol. Ther. 12:517 (2010)). Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của lợn có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của lợn có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ OBI-1 của lợn. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao gồm cầu nối vùng chức năng B typ OBI-1 của người, bao gồm các axit amin 760-772/1655-1667 của vùng chức năng B Yếu tố VIII người kiểu hoang (FVIII-FL-AA; SEQ ID NO: 19). Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của người có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của người có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ OBI-1 của người. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa đã mô tả ở đây bao gồm cầu nối vùng chức năng B typ O8, bao gồm các axit amin SFSQNSRHQAYRYRRG (SEQ ID NO: 32) từ vùng chức năng B Yếu tố VIII lợn kiểu hoang (Toschi et al., Curr. Opin. Mol. Ther. 12:517 (2010)). Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của lợn có một đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, cầu nối vùng chức năng B typ OBI-1 của lợn có hai đột biến thế axit amin so với trình tự kiểu hoang tương ứng. Trong một số phương án, peptit glycosyl hóa được gắn xen vào cầu nối vùng chức năng B typ OBI-1 của lợn. Trong một số phương án, peptit glycosyl hóa được chọn từ các peptit được thể hiện trong Fig. 13 (SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện).

Việc loại bỏ vùng chức năng B khỏi các cấu trúc Yếu tố VIII không ảnh hưởng đến

hoạt tính của enzym đã hoạt hóa (ví dụ, FVIIIa), có thể là vì vùng chức năng B được loại bỏ trong quá trình hoạt hóa. Tuy nhiên, vùng chức năng B của Yếu tố VIII chứa một số gốc được biến đổi sau dịch mã, ví dụ, bằng cách glycosyl hóa liên kết N hoặc O. Phân tích in silico (Prediction of N-glycosylation sites in human proteins, R. Gupta, E. Jung và S. Brunak, in preparation (2004)) vùng chức năng B Yếu tố VIII kiểu hoang dự đoán là ít nhất bốn trong số các vị trí này được glycosyl hóa in vivo (Fig. 14). Được cho là các biến đổi trong vùng chức năng B này góp phần vào việc điều hòa sau dịch mã và/hoặc thời gian bán thải của Yếu tố VIII trên in vivo.

Mặc dù vùng chức năng B Yếu tố VIII vắng mặt trong protein Yếu tố VIIIa thành thục, sự glycosyl hóa trong vùng chức năng B của phân tử tiền chất Yếu tố VIII có thể làm tăng thời gian bán thải tuần hoàn của protein này trước khi hoạt hóa. Do đó, trong một số phương án, cầu nối polypeptit của các cấu trúc Yếu tố VIII được mã hóa đã mô tả ở đây bao gồm một hoặc nhiều trình tự glycosyl hóa, để cho phép glycosyl hóa trên in vivo. Trong một số phương án, cầu nối polypeptit bao gồm ít nhất một trình tự glycosyl hóa liên ứng (ví dụ, trình tự liên ứng glycosyl hóa liên kết N hoặc O). Trong một số phương án, cầu nối polypeptit bao gồm ít nhất là hai trình tự glycosyl hóa liên ứng. Trong một số phương án, cầu nối polypeptit bao gồm ít nhất là ba trình tự glycosyl hóa liên ứng. Trong một số phương án, cầu nối polypeptit bao gồm ít nhất là bốn trình tự glycosyl hóa liên ứng. Trong một số phương án, cầu nối polypeptit bao gồm ít nhất là năm trình tự glycosyl hóa liên ứng. Trong một số phương án, cầu nối polypeptit bao gồm ít nhất là 6, 7, 8, 9, 10, hoặc nhiều trình tự glycosyl hóa liên ứng.

Trong một số phương án, cầu nối polypeptit chứa ít nhất một trình tự glycosyl hóa liên kết N N-X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T. Trong một số phương án, cầu nối polypeptit chứa ít nhất là hai trình tự glycosyl hóa liên kết N N-X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T. Trong một số phương án, cầu nối polypeptit chứa ít nhất là ba trình tự glycosyl hóa liên kết N N-X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T. Trong một số phương án, cầu nối polypeptit chứa ít nhất là bốn trình tự glycosyl hóa liên kết N N-X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T. Trong một số phương án, cầu nối polypeptit chứa ít nhất là năm trình tự glycosyl hóa liên kết N N-X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T. Trong một số phương án, cầu nối polypeptit chứa ít nhất là 6, 7, 8, 9, 10, hoặc nhiều trình tự glycosyl hóa liên kết N N-

X-S/T, trong đó X là axit amin bất kỳ khác P, S, hoặc T.

Trong một số phương án, cầu nối polypeptit này chứa peptit glycosyl hóa có sự đồng nhất trình tự cao với bất kỳ một trong số SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong Fig. 13A-13B. Trong một số phương án, polypeptit glycosyl hóa có sự đồng nhất ít nhất là 92% với bất kỳ một trong số SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá hai đột biến thế axit amin so với bất kỳ một trong số SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá một đột biến thế axit amin so với bất kỳ trong số SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có trình tự axit amin được chọn từ bất kỳ trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B.

Trong một số phương án, peptit glycosyl hóa có sự đồng nhất ít nhất là 92% với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 90% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có sự đồng nhất ít nhất là 92% với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 95% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có sự đồng nhất ít nhất là 92% với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 98% với trình tự nucleotit tương

ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B.

Trong một số phương án, peptit glycosyl hóa có không quá hai đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 90% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá hai đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 95% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá hai đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 98% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B.

Trong một số phương án, peptit glycosyl hóa có không quá một đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 90% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá một đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 95% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng,

theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có không quá một đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 98% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B.

Trong một số phương án, peptit glycosyl hóa có trình tự được chọn từ bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 90% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có trình tự được chọn từ bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 95% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa có trình tự được chọn từ bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B và được mã hóa bởi trình tự polynucleotit có sự đồng nhất ít nhất là 98% với trình tự nucleotit tương ứng được chọn từ các SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B.

Trong một số phương án, polypeptit Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon đã mô tả ở đây có cầu nối được thế vùng chức năng B, trong đó peptit glycosyl hóa được gắn xen vào trình tự cầu nối SQ (các axit amin 760-773 của CS04-FL-AA; SEQ ID NO: 2). Trong một phương án cụ thể, peptit glycosyl hóa được chọn từ bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B, peptit glycosyl hóa có sự đồng nhất ít nhất là 92% với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59,

61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B, peptit glycosyl hóa có không quá hai đột biến thế axit amin so với bất kỳ một trong số SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B, và peptit glycosyl hóa có không quá một đột biến thế axit amin so với bất kỳ một trong số các SEQ ID NO 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, và 75, tương ứng, theo thứ tự xuất hiện, như được thể hiện trong các Fig. 13A-13B. Trong một số phương án, peptit glycosyl hóa được gắn xen vào peptit SQ giữa các gốc N768 và P769 (so với CS04-FL-AA; SEQ ID NO: 2).

Trong một số phương án, cầu nối polypeptit của cấu trúc Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất trình tự cao với bất kỳ một trong số các trình tự được thể hiện trong Fig. 6 (SEQ ID NO 5-7 và 36-48, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 96% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 97% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 98% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 99% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 99,5% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 99,9% với bất kỳ một trong các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, và 74, tương ứng, theo thứ tự xuất hiện). Trong một số phương án, trình tự nucleotit thứ ba là đồng nhất với trình tự bất kỳ trong số các trình tự được thể hiện trong Fig. 13 (SEQ ID NO 50, 52, 54, 56, 58, 60, 62, 64, 66, 68,

70, 72, và 74, tương ứng, theo thứ tự xuất hiện).

Polynucleotit được biến đổi codon mã hóa biến thể Yếu tố VIII có cầu nối cắt được

Các polynucleotit được biến đổi codon CS04

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII có cầu nối có thể cắt được trên in vivo. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS04-HC-NA (SEQ ID NO: 3), mà là một phần của CS04-FL-NA (SEQ ID NO: 1) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS04-LC-NA (SEQ ID NO: 4), mà là một phần của CS04-FL-NA (SEQ ID NO: 1) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit bao gồm vị trí cắt furin, cho phép thành thục trên in vivo (ví dụ, sau khi biểu hiện trên in vivo hoặc sử dụng tiền polypeptit).

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4).

Trong một số phương án, cầu nối polypeptit của cấu trúc Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất trình tự cao với BDLO04 (SEQ ID NO: 6), mã hóa cầu nối 14 axit amin tương ứng với các axit amin 760-773 của CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 96% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 97% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 98% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba là đồng nhất với BDLO04 (SEQ ID NO: 6).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS04-FL-NA (SEQ ID NO: 1). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS04-FL-NA (SEQ ID NO: 1).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS04-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit

amin này đồng nhất với CS04-FL-AA (SEQ ID NO: 2).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04, có sự tương đồng trình tự cao với CS04-FL-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m125.

Các polynucleotit được biến đổi codon CS01

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây

bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII có cầu nối có thể cắt được trên in vivo. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS01-HC-NA (SEQ ID NO: 24), mà là một phần của CS01-FL-NA (SEQ ID NO: 13) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS01-LC-NA (SEQ ID NO: 25), mà là một phần của CS01-FL-NA (SEQ ID NO: 13) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit bao gồm vị trí cắt furin, cho phép thành thục trên in vivo (ví dụ, sau khi biểu hiện trên in vivo hoặc sử dụng tiền polypeptit).

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25).

Trong một số phương án, cầu nối polypeptit của cấu trúc Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất trình tự cao với BDLO04 (SEQ ID NO: 6), mã hóa cầu nối 14 axit amin tương ứng với các axit amin 760-773 của CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có

sự đồng nhất ít nhất là 96% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 97% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 98% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba là đồng nhất với BDLO04 (SEQ ID NO: 6).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS01-FL-NA (SEQ ID NO: 13). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS01-FL-NA (SEQ ID NO: 13).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS01-FL-AA (SEQ ID NO: 2). Trong một số phương án, trình tự axit amin này đồng nhất với CS01-FL-AA (SEQ ID NO: 2).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01, có sự tương đồng trình tự cao với CS01-FL-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m125.

Các polynucleotit được biến đổi codon CS23

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII có cầu nối có thể cắt được trên in vivo. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS23-HC-NA (SEQ ID NO: 22), mà là một phần của CS23-FL-NA (SEQ ID NO: 20) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu

tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS23-LC-NA (SEQ ID NO: 23), mà là một phần của CS23-FL-NA (SEQ ID NO: 20) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit bao gồm vị trí cắt furin, cho phép thành thục trên in vivo (ví dụ, sau khi biểu hiện trên in vivo hoặc sử dụng tiền polypeptit).

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23).

Trong một số phương án, cầu nối polypeptit của cấu trúc Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất trình tự cao với BDLO04 (SEQ ID NO: 6), mã hóa cầu nối 14 axit amin tương ứng với các axit amin 760-773 của CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 96% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 97% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 98% với BDLO04 (SEQ ID NO: 6). Trong một số phương án, trình tự nucleotit thứ ba là đồng nhất với BDLO04 (SEQ ID NO: 6).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có

sự đồng nhất trình tự cao với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS23-FL-NA (SEQ ID NO: 20). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS23-FL-NA (SEQ ID NO: 20).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS23-FL-AA (SEQ ID NO: 21). Trong một số phương án, trình tự axit amin này đồng nhất với CS23-FL-AA (SEQ ID NO: 21).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23, có sự tương đồng trình tự cao với CS23-FL-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi

polynucleotit CS23 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m125.

Polynucleotit được biến đổi codon mã hóa protein Yếu tố VIII chuỗi đơn

Các cấu trúc Yếu tố VIII, trong đó vị trí cắt furin nằm ở đầu cùng C của vùng chức năng B được loại bỏ, vẫn giữ được hoạt tính như là polypeptit chuỗi đơn, mặc dù sự thành thục bình thường của phân tử Yếu tố VIII không thể xảy ra (Leyte et al. (1991)). Tương tự, cấu trúc Yếu tố VIII bị mất vùng chức năng B có vị trí furin bị suy giảm (chứa đột biến thế axit amin R1664H) hoạt động sinh học mạnh hơn so với cấu trúc Yếu tố VIII tương ứng có vị trí cắt furin kiểu hoang (Siner et al. (2013)). Do đó, trong một số phương án, polynucleotit được biến đổi codon được đề xuất ở đây bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII chuỗi đơn. Polypeptit Yếu tố VIII chuỗi đơn bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Cầu nối polypeptit này không bao gồm vị trí cắt furin.

Các polynucleotit được biến đổi codon CS04 chuỗi đơn

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây

bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII chuỗi đơn. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit tùy ý nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS04-HC-NA (SEQ ID NO: 3), mà là một phần của CS04-FL-NA (SEQ ID NO: 1) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS04-LC-NA (SEQ ID NO: 4), mà là một phần của CS04-FL-NA (SEQ ID NO: 1) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit tùy ý này không bao gồm vị trí cắt furin.

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS04-HC-NA và CS04-LC-NA (SEQ ID NO 3 và 4).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự

nucleotit này có sự đồng nhất ít nhất là 98% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS04-SC1-NA (SEQ ID NO: 9). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS04-SC1-NA (SEQ ID NO: 9).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS04-SC2-NA (SEQ ID NO: 11). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS04-SC2-NA (SEQ ID NO: 11).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS04-SC1-AA (SEQ ID NO: 10; Yếu tố VIIIΔ của người (760-1667) (SPI; *Hs*FVIIIΔ(741-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS04-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này đồng nhất với CS04-SC1-AA (SEQ ID NO: 10).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04-SC1, có sự tương đồng trình tự cao với CS04-SC1-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS04-SC2-AA (SEQ ID NO: 12; Yếu tố VIIIΔ của người (772-1667) (SPI; *Hs*FVIIIΔ(753-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS04-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này đồng nhất với CS04-SC2-AA (SEQ ID NO: 12).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS04-SC2, có sự tương đồng trình tự cao với CS04-SC2-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII

được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS04 chứa các đột biến thế axit amin m125.

Các polynucleotit được biến đổi codon CS01 chuỗi đơn

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII chuỗi đơn. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit tùy ý nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS01-HC-NA (SEQ ID NO: 24), mà là một phần của CS01-FL-NA (SEQ ID NO: 13) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS01-LC-NA (SEQ ID NO: 25), mà là một phần của CS01-FL-NA (SEQ ID NO: 13) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit tùy ý này không bao gồm vị trí cắt furin.

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số

phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS01-HC-NA và CS01-LC-NA (SEQ ID NO 24 và 25).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS01-SC1-NA (SEQ ID NO: 26). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS01-SC1-NA (SEQ ID NO: 26).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất

là 99,5% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS01-SC2-NA (SEQ ID NO: 27). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS01-SC2-NA (SEQ ID NO: 27).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS01-SC1-AA (SEQ ID NO: 10; Yếu tố VIIIΔ của người (760-1667) (SPI; *Hs*FVIIIΔ(741-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS01-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này đồng nhất với CS01-SC1-AA (SEQ ID NO: 10).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01-SC1, có sự tương đồng trình tự cao với CS01-SC1-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS01-SC2-AA (SEQ ID NO: 12; Yếu tố VIIIΔ của người (772-1667) (SPI; *Hs*FVIIIΔ(753-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin

này có sự đồng nhất ít nhất là 99,9% với CS01-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này đồng nhất với CS01-SC2-AA (SEQ ID NO: 12).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS01-SC2, có sự tương đồng trình tự cao với CS01-SC2-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS01 chứa các đột biến thế axit amin m125.

Các polynucleotit được biến đổi codon CS23 chuỗi đơn

Trong một phương án, các polynucleotit được biến đổi codon được đề xuất ở đây bao gồm trình tự nucleotit mã hóa polypeptit biến thể Yếu tố VIII chuỗi đơn. Polypeptit Yếu tố VIII bao gồm chuỗi nhẹ Yếu tố VIII, chuỗi nặng Yếu tố VIII, và cầu nối polypeptit tùy ý nối đầu cùng C của chuỗi nặng với đầu cùng N của chuỗi nhẹ. Chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất trình tự cao với CS23-HC-NA (SEQ ID NO: 22), mà là một phần của CS23-FL-NA (SEQ ID NO: 20) mã hóa chuỗi nặng Yếu tố VIII. Chuỗi nhẹ của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất trình tự cao với CS23-LC-NA (SEQ ID NO: 23), mà là một phần của CS23-FL-NA (SEQ ID NO: 20) mã hóa chuỗi nhẹ Yếu tố VIII. Cầu nối polypeptit tùy ý này không bao gồm vị trí cắt furin.

Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 95% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 96% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 97% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 98% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,5% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, các trình tự nucleotit thứ nhất và thứ hai có sự đồng nhất trình tự ít nhất là 99,9% lần lượt với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23). Trong một số phương án, trình tự nucleotit thứ nhất và thứ hai lần lượt là đồng nhất với CS23-HC-NA và CS23-LC-NA (SEQ ID NO 22 và 23).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng

nhất ít nhất là 97% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS23-SC1-NA (SEQ ID NO: 28). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS23-SC1-NA (SEQ ID NO: 28).

Trong một số phương án, polynucleotit được biến đổi codon có trình tự nucleotit có sự đồng nhất trình tự cao với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 95% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 96% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 97% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 98% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,5% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này có sự đồng nhất ít nhất là 99,9% với CS23-SC2-NA (SEQ ID NO: 29). Trong một số phương án, trình tự nucleotit này là đồng nhất với CS23-SC2-NA (SEQ ID NO: 29).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS23-SC1-AA (SEQ ID NO: 10; Yếu tố VIIIΔ của người (760-1667) (SPI; CS04Δ(741-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS23-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS23-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS23-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS23-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS23-SC1-AA (SEQ ID NO: 10). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS23-SC1-AA (SEQ ID NO: 10). Trong một số

phương án, trình tự axit amin này đồng nhất với CS23-SC1-AA (SEQ ID NO: 10).

Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23-SC1, có sự tương đồng trình tự cao với CS23-SC1-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS23-SC2-AA (SEQ ID NO: 12; Yếu tố VIIIΔ của người (772-1667) (SPI; *Hs*FVIIIΔ(753-1648), SPE)). Trong một số phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit được biến đổi codon có trình tự axit amin có sự đồng nhất trình tự cao với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 97% với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 98% với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99% với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,5% với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này có sự đồng nhất ít nhất là 99,9% với CS23-SC2-AA (SEQ ID NO: 12). Trong một số phương án, trình tự axit amin này đồng nhất với CS23-SC2-AA (SEQ ID NO: 12).

Trong một số phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS23-SC2, có sự tương đồng trình tự cao với CS23-SC2-AA (ví dụ, đồng nhất ít nhất là 95%, 96%, 97%, 98%, 99%, 99,5%, hoặc 99,9%), chứa một hoặc nhiều đột biến thế axit amin được chọn từ m1, m2, m3, m4, và m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m1. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m2. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m3. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m4. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa đột biến thế axit amin m5.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit

CS23 chứa các đột biến thế axit amin m12. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m13. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m23. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m24. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m25. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m34. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m35.

Trong một phương án, biến thể Yếu tố VIII chuỗi đơn được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m123. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m234. Trong một phương án, biến thể Yếu tố VIII được mã hóa bởi polynucleotit CS23 chứa các đột biến thế axit amin m125.

Vectơ biểu hiện Yếu tố VIII

Trong một số phương án, các polynucleotit được biến đổi codon được mô tả ở đây được tích hợp vào các vectơ biểu hiện. Ví dụ không giới hạn về các vectơ biểu hiện bao gồm vectơ virut (ví dụ, các vectơ thích hợp cho liệu pháp gen), vectơ plasmit, vectơ thể thực khuẩn, cosmid, phagemid, nhiễm sắc thể nhân tạo, và tương tự.

Ví dụ không giới hạn về các vectơ virut bao gồm: retrovirut, ví dụ, virut bệnh bạch cầu Moloney ở chuột (MMLV), virut sarcoma Harvey ở chuột, virut gây ung thư vú ở chuột, và virut Rous sarcoma; adenovirut, virut liên hợp adeno; virut typ SV40; polyomavirut; virut Epstein-Barr; virut papilloma; virut herpes; virut vaccinia; và virut bại liệt.

Trong một số phương án, các polynucleotit được biến đổi codon được mô tả ở đây được tích hợp vào vectơ liệu pháp gen. Trong một số phương án, vectơ liệu pháp gen là retrovirut, và cụ thể là retrovirut không có khả năng sao chép. Quy trình sản xuất các retrovirut không có khả năng sao chép đã được biết trong lĩnh vực này. Để xem lại, xem Kriegler, M., Gene Transfer and Expression, A Laboratory Manual, W.H. Freeman Co., New York (1990) và Murry, E. J., Methods in Molecular Biology, Vol. 7, Humana Press,

Inc., Cliffton, N.J. (1991)).

Trong một phương án, vectơ liệu pháp gen là vectơ liệu pháp gen dựa trên virut liên hợp adeno (AAV). Các hệ AAV đã được mô tả trước đây và đã được biết rõ trong lĩnh vực này (Kelleher and Vos, *Biotechniques*, 17(6):1110-17 (1994); Cotten et al., *Proc Natl Acad Sci USA*, 89(13):6094-98 (1992); Curiel, *Nat Immun*, 13(2-3):141-64 (1994); Muzyczka, *Curr Top Microbiol Immunol*, 158:97-129 (1992); và Asokan A, et al., Mol. Ther., 20(4):699-708 (2012), mỗi tài liệu này được đưa vào đây bằng cách viện dẫn toàn bộ cho tất cả các mục đích). Các chi tiết liên quan đến việc tạo ra và sử dụng các vectơ rAAV đã được mô tả, ví dụ, trong Patent Mỹ số 5,139,941 và 4,797,368, mỗi tài liệu này được đưa vào đây bằng cách viện dẫn toàn bộ cho tất cả các mục đích. Trong một phương án cụ thể, vectơ AAV là vectơ AAV-8.

Trong một số phương án, các polynucleotit được biến đổi codon được mô tả ở đây được tích hợp vào vectơ biểu hiện retrovirut. Các hệ này đã được mô tả trước đây, và đã được biết rõ trong lĩnh vực này (Mann *et al.*, *Cell*, 33:153-159, 1983; Nicolas và Rubinstein, In: Vectors: A survey of molecular cloning vectors and their uses, Rodriguez and Denhardt, eds., Stoneham: Butterworth, pp. 494-513, 1988; Temin, In: Gene Transfer, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986). Trong một phương án cụ thể, vectơ retrovirut là vectơ lentivirut (xem, ví dụ, Naldini *et al.*, *Science*, 272(5259):263-267, 1996; Zufferey *et al.*, *Nat Biotechnol*, 15(9):871-875, 1997; Blomer *et al.*, *J Virol.*, 71(9):6641-6649, 1997; Patent Mỹ số 6,013,516 và 5,994,136).

Nhiều loại vectơ có thể được sử dụng để biểu hiện polypeptit Yếu tố VIII từ polypeptit được biến đổi codon trong môi trường nuôi cấy tế bào, bao gồm các vectơ biểu hiện trong tế bào nhân chuẩn và tiền nhân. Trong một số phương án, vectơ plasmit được dự kiến sử dụng để biểu hiện polypeptit Yếu tố VIII trong môi trường nuôi cấy tế bào. Nói chung, các vectơ plasmit chứa đơn vị sao chép và các trình tự điều khiển thu được từ loài tương thích với tế bào chủ được sử dụng liên quan đến các vật chủ này. Vectơ này có thể mang vị trí sao chép, cũng như các trình tự đánh dấu mà cung cấp khả năng lựa chọn kiểu hình trong các tế bào được biến nạp. Plasmit này sẽ bao gồm polynucleotit được biến đổi codon mã hóa polypeptit Yếu tố VIII, được liên kết chức năng với một hoặc nhiều trình tự điều khiển, ví dụ, trình tự khởi động.

Ví dụ không giới hạn về các vectơ để biểu hiện trong tế bào tiền nhân bao gồm các plasmit như pRSET, pET, pBAD, v.v..., trong đó các trình tự khởi động được sử dụng trong các vectơ biểu hiện trong tế bào tiền nhân bao gồm lac, trc, trp, recA, araBAD, v.v... Ví dụ về các vectơ để biểu hiện trong tế bào nhân chuẩn bao gồm: (i) để biểu hiện trong nấm men, các vectơ như pAO, pPIC, pYES, pMET, sử dụng các trình tự khởi động như AOX1, GAP, GAL1, AUG1, v.v...; (ii) để biểu hiện trong tế bào côn trùng, các vectơ như pMT, pAc5, pIB, pMIB, pBAC, v.v..., sử dụng các trình tự khởi động như PH, p10, MT, Ac5, OpIE2, gp64, polh, v.v...., và (iii) để biểu hiện trong tế bào động vật có vú, các vectơ như pSVL, pCMV, pRc/RSV, pcDNA3, pBPV, v.v..., và các vectơ thu được từ hệ virut như virut vaccinia, virut liên hợp adeno, virut herpes, retrovirut, v.v..., sử dụng các trình tự khởi động như CMV, SV40, EF-1, UbC, RSV, ADV, BPV, và β-actin.

Ví dụ thực hiện sáng chế

Ví dụ 1 – Cấu trúc của trình tự biểu hiện biến thể Yếu tố VIII được biến đổi codon

Có hai trở ngại cần phải vượt qua để tạo ra trình tự mã hóa Yếu tố VIII hữu ích cho liệu pháp gen để điều trị bệnh hemophilia A. Thứ nhất, do các hạn chế về kích thước bộ gen của các vectơ mang liệu pháp gen thông thường (ví dụ, các virion AAV), polypeptit Yếu tố VIII được mã hóa phải được làm ngắn đi đáng kể. Thứ hai, trình tự mã hóa phải được biến đổi để: (i) ổn định các tương tác đóng gói trong vectơ vận chuyển, (ii) ổn định chất trung gian mARN, và (iii) cải thiện độ mạnh của quá trình phiên mã/dịch mã của mARN.

Để đạt được mục tiêu thứ nhất, các tác giả sáng chế bắt đầu với cấu trúc biến thể Yếu tố VIII bị mất vùng chức năng B, được gọi là "FVIII-BDD-SQ." Trong cấu trúc này, vùng chức năng B được thay thế bằng trình tự mười bốn axit amin được gọi là trình tự "SQ". FVIII-BDD-SQ tái tổ hợp được bán với tên thương mại là REFACTO®, và đã được chỉ ra là hiệu quả trong việc quản lý bệnh hemophilia A. Tuy nhiên, trình tự mã hóa tự nhiên cho FVIII-BDD-SQ, mà bao gồm các trình tự axit nucleic người kiểu hoang cho chuỗi nặng và chuỗi nhẹ Yếu tố VIII, được biểu hiện không hiệu quả trong các vectơ liệu pháp gen.

Để khắc phục sự biểu hiện kém của FVIII-BDD-SQ tự nhiên, thuật toán tối ưu hóa codon được mô tả trong Fath et al. (PLoS ONE, 6:e17596 (2011)), được biến đổi như được

mô tả trong Ward et al. (Blood, 117:798 (2011)) và trong McIntosh et al. (Blood, 121, 3335-3344 (2013)), được áp dụng cho trình tự FVIII-BDD-SQ để tạo ra trình tự mã hóa trung gian thứ nhất CS04a. Tuy nhiên, tác giả sáng chế nhận ra là trình tự CS04a được tạo ra bằng cách sử dụng thuật toán biến đổi có thể được cải thiện bằng cách thay đổi thêm trình tự này. Theo đó, tác giả sáng chế lại đưa vào các dinucleotit CpG, lại đưa vào codon CGC cho arginin, thay đổi sự phân bố codon leuxin và serin, lại đưa vào các cặp codon được bảo tồn cao, và loại bỏ hộp TATA, hộp CCAAT không xác định, và các yếu tố vị trí nối, trong khi vẫn tránh được các đảo CpG và sự trình diện quá mức cục bộ của các dải giàu AT và giàu GC.

Thứ nhất, thuật toán được biến đổi này thay thế một cách hệ thống các codon chứa các CpG-dinucleotit (ví dụ, các codon arginin) bằng các codon dinucleotit không phải CpG, và làm giảm/tránh các CpG-dinucleotit được tạo ra bởi các codon lân cận. Việc tránh hoàn toàn các CpG dinucleotit này thường được thực hiện để ngăn ngừa tính miễn dịch được cảm ứng bởi TLR sau khi tiêm trong cơ vacxin ADN. Tuy nhiên, việc này lại hạn chế khả năng tối ưu hóa codon. Ví dụ, thuật toán được biến đổi này loại trừ việc sử dụng tập hợp các codon CGX arginin hoàn chỉnh. Điều này phá vỡ quá trình mã hóa của các gen cho sự biểu hiện trong tế bào người, vì CGC là codon arginin được sử dụng thường xuyên nhất trong các gen người được biểu hiện cao. Ngoài ra, việc tránh tạo ra các CpG bởi các codon lân cận làm hạn chế thêm khả năng tối ưu hóa (ví dụ, giới hạn số lượng các cặp codon có thể được sử dụng cùng nhau).

Vì tính miễn dịch được cảm ứng bởi TLR không được mong đợi trở thành một vấn đề liên quan đến liệu pháp gen dựa trên AAV, hướng đến gan, các codon bao gồm các CpG, và các codon lân cận tạo ra các CpG, được đưa lại vào trình tự mã hóa trung gian CS04a, tốt hơn là, vào trình tự mã hóa chuỗi nhẹ Yếu tố VIII (ví dụ, ở đầu 3' của trình tự mã hóa FVIII-BDD-SQ). Điều này cho phép sử dụng thường xuyên hơn các codon người được ưu tiên, cụ thể là các codon cho arginin. Tuy nhiên, phải cẩn thận để tránh tạo ra các đảo CpG, là các vùng của trình tự mã hóa có tần số các vị trí CpG cao. Điều này trái ngược với tài liệu của Krinner et al. (Nucleic Acids Res., 42(6):3551-64 (2014)), tài liệu này gợi ý là các vùng chức năng CpG xuôi dòng của vị trí khởi đầu phiên mã tăng cường mức độ biểu hiện gen cao.

Thứ hai, thuật toán biến đổi này sử dụng một số codon riêng, như CTG cho leuxin,

GTG cho valin, và CAG cho glutamin. Tuy nhiên, điều này vi phạm nguyên tắc sử dụng codon cân bằng, ví dụ, như được đề xuất trong Haas et al. (Current Biology, 6(3):315-24 (1996)). Do việc sử dụng quá mức các codon được ưu tiên bởi thuật toán biến đổi này, các codon leuxin luân phiên được đưa lại vào vị trí được cho phép bởi các nguyên tắc khác được áp dụng cho sự thay đổi codon (ví dụ, tần số CpG và hàm lượng GC).

Thứ ba, thuật toán được biến đổi này thay thế các cặp codon bất kể việc chúng được bảo tồn như thế nào trong tự nhiên, khi một số tiêu chuẩn (ví dụ, sự có mặt của CG-dinucleotit) được đáp ứng. Do các đặc điểm có lợi có thể được bảo tồn bởi sự tiến hóa, các cặp codon được bảo tồn nhiều nhất mà được thay thế bởi thuật toán này và các cặp codon được ưu tiên được bảo tồn nhiều nhất, ví dụ, như được mô tả trong Tats et al. (BMC Genomics 9:463 (2008)), được phân tích và được điều chỉnh tại vị trí được cho phép bởi các nguyên tắc khác được áp dụng cho sự biến đổi codon (ví dụ, tần số CpG và hàm lượng GC).

Thứ tư, các codon serin được sử dụng trong trình tự mã hóa trung gian cũng được thao tác di truyền lại. Cụ thể là, các codon AGC, TCC, và TCT serin được đưa vào trình tự mã hóa biến đổi với tần số cao hơn, để bắt cặp tốt hơn toàn bộ cho việc sử dụng codon người (Haas et al., ở trên).

Thứ năm, các yếu tố hộp TATA, hộp CCAAT, và các vị trí nối intron/exon được sàng lọc và được loại bỏ khỏi trình tự mã hóa biến đổi. Khi biến đổi trình tự mã hóa, phải cẩn thận để tránh sự trình diện quá mức cục bộ các dải giàu AT hoặc giàu GC.

Cuối cùng, ngoài việc tối ưu hóa việc sử dụng codon trong trình tự mã hóa, các yêu cầu về cấu trúc của virion AAV cơ bản được xem xét khi tinh chế thêm trình tự mã hóa trung gian CS04a. Các vectơ AAV (ví dụ, phần axit nucleic của virion AAV) được đóng gói dưới dạng các phân tử ADN chuỗi đơn vào trong các capsit của chúng (để xem lại, xem, Daya và Berns, Clin. Microbiol Rev., 21(4):583-93 (2008)). Hàm lượng GC của vectơ này do đó có thể ảnh hưởng đến việc đóng gói của bộ gen, và do đó, ảnh hưởng đến năng suất vectơ trong quá trình sản xuất. Tương tự nhiều thuật toán, thuật toán biến đổi được sử dụng ở đây tạo ra trình tự gen được tối ưu hóa với hàm lượng GC ít nhất là 60% (*xem*, Fath et al., PLoS One, 6(3):e17596 (2011) (đính chính trong: PLoS One, (6)3 (2011)). Tuy nhiên, protein capsit AAV8 được mã hóa bởi trình tự nucleotit có hàm lượng GC thấp hơn

là khoảng 56%. Do đó, để giả trình tự mã hóa protein capsit AAV8 tự nhiên tốt hơn, hàm lượng GC của trình tự mã hóa trung gian CS04a được giảm xuống 56%.

Trình tự mã hóa CS04 tạo thành, được thể hiện trong Fig. 2, có hàm lượng GC toàn phần là 56%. Hàm lượng CpG-dinucleotit của trình tự này là trung bình. Tuy nhiên, các CpG dinucleotit chủ yếu xuất hiện trong phần xuôi dòng của trình tự mã hóa, ví dụ, phần mã hóa chuỗi nhẹ Yếu tố VIII. Trình tự CS04 có sự đồng nhất trình tự nucleotit là 79,77% với trình tự mã hóa tương ứng trong Yếu tố VIII kiểu hoang (số truy cập Ngân hàng gen M14113).

Với mục đích so sánh, một số cấu trúc ReFacto, được tối ưu hóa codon khác được tạo ra. CS01 được tạo ra bằng cách áp dụng thuật toán tối ưu hóa codon của Fath et al., như được biến đổi bởi Ward et al., như được thực hiện cho CS04. Tuy nhiên, khác với CS04, cấu trúc CS01 không chứa các đảo CpG bất kỳ. Cấu trúc CS08 ReFacto được tối ưu hóa codon như được mô tả trong Radcliff P.M. et al., Gene Therapy, 15:289-97 (2008), nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn, toàn bộ, cho tất cả các mục đích. Cấu trúc ReFacto được tối ưu hóa codon CS10 thu được từ Eurofins Genomics (Ebersberg, Đức). Cấu trúc ReFacto được tối ưu hóa codon CS11 thu được từ Integrated DNA Technologies, Inc. (Coralville, USA). Cấu trúc ReFacto được tối ưu hóa codon CH25 thu được từ các dịch vụ GeneArt của ThermoFischer Scientific (Regensburg, Đức). Cấu trúc CS40 ReFacto chứa trình tự mã hóa Yếu tố VIII kiểu hoang. Thuật toán được sử dụng để tạo ra CS23 là dựa trên công cụ JCAT (www.jcat.de), công cụ trực tuyến để tối ưu hóa codon (Grote et al., 2005; Nucl. Acids Res. W526-31). Trình tự này được biến đổi thêm để phản ánh nhiều hơn việc sử dụng codon của siêu họ albumin (Mirsafian et al. 2014). Sc. Word Journal 2014, ID 639682). Các tỉ lệ đồng nhất trình tự chung giữa mỗi trình tự mã hóa ReFacto được thể hiện trong Bảng 2, dưới đây.

Bảng 2 – Ma trận tỉ lệ phần trăm đồng nhất đối với các cấu trúc Yếu tố VIII được biến đổi codon.

	CS01	CS04	CS08	CS10	CS11	CS40	CH25	CS23
CS01	100%							
CS04	93,0%	100%						
CS08	80,7%	82,2%	100%					

CS10	79,1%	79,4%	78,4%	100%				
CS11	78,3%	78,3%	78,1%	77,5%	100%			
CS40	79,6%	79,8%	76,7%	77,6%	75,4%	100%		
CH25	81,3%	85,1%	85,0%	79,9%	79,4%	75,8%	100%	
CS23	84,3%	89,2%	85,1%	80,3%	79,9	76,5%	93,2%	100%

Các plasmit của mỗi cấu trúc được tạo ra bằng cách tách dòng các đoạn ADN tổng hợp khác nhau vào cùng một plasmit khung vectơ (pCh-BB01). Quy trình tổng hợp ADN của các đoạn BDD-FVIII typ Refacto với các vị trí giới hạn enzym AscI và NotI bên sườn được thực hiện bởi ThermoFischer Scientific (Regensburg, Đức). Khung vectơ này chứa hai đoạn lặp đầu cùng đảo (ITR) thu được từ AAV2 bên sườn mà bao gồm trình tự khởi động/trình tự tăng cường thu được từ gen transthyretin chuột đặc hiệu gan, vị trí giới hạn enzym AscI và NotI để gắn xen BDD-FVIII typ Refacto tương ứng và vị trí polyA tổng hợp. Sau khi nối khung vectơ đã tạo ra và gắn xen qua các vị trí AscI và NotI, các plasmit tạo thành được khuếch đại theo thang milligam. Các trình tự BDD-FVIII typ Refacto của các cấu trúc này được kiểm tra bằng cách giải trình tự trực tiếp (Microsynth, Balgach, Thụy Sĩ). Việc tách dòng tạo ra bảy cấu trúc plasmit khác nhau được đặt tên là pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, và pCh25 (Fig. 23). Các cấu trúc này có cùng một khung vectơ và mã hóa cùng một protein FVIII bị mất vùng chức năng B (BDD-FVIII typ Refacto), nhưng khác nhau ở trình tự mã hóa FVIII của chúng.

Các vectơ dựa trên AAV8 được tạo ra bằng phương pháp chuyển nhiễm ba plasmit, như được mô tả trong Grieger JC, et al. (Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector, Mol Ther., Oct 6. (2015) doi: 10.1038/mt.2015.187. [Epub ahead of print]), nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn, toàn bộ, cho tất cả các mục đích. Các tế bào huyền phù HEK293 được sử dụng để chuyển nhiễm plasmit bằng cách sử dụng plasmit vectơ FVIII tương ứng, plasmit pXX6-80 trợ giúp (mang gen trợ giúp adenovirut), và plasmit đóng gói pGSK2/8 (đóng góp các gen rep2 và cap8). Để phân lập các cấu trúc AAV8, các pelet tế bào của một lit môi trường nuôi cấy được xử lý bằng cách sử dụng các gradien iodixanol, như được mô tả trong Grieger et al. (2015, ở trên). Quy trình này tạo ra các chế phẩm vectơ được gọi là vCS01, vCS04, vCS08, vCS10, vCS11, và vCH25. Các vectơ được định lượng bằng qPCR bằng cách sử dụng quy trình qPCR thông

dụng gắn đích các đoạn lặp đầu cùng đảo AAV2 (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)). Plasmit vectơ đối chứng mang các đoạn lặp đầu cùng đảo AAV2 đóng vai trò để tạo ra đường cong chuẩn. Cấu trúc vCS04 tạo thành được thể hiện bằng SEQ ID NO: 8 trong các Fig. 7A-7C.

Tính nguyên vẹn của các bộ gen vectơ được phân tích bằng điện di gen AAV agaroza. Điện di được thực hiện như đã mô tả trong Fagone et al., Human Gene Therapy Methods 23:1-7 (2012). Nói ngắn gọn, các chế phẩm vectơ AAV được ủ ở 75°C trong 10 phút khi có mặt 0,5% SDS và sau đó được làm nguội xuống nhiệt độ phòng. Xấp xỉ 1,5E10 bộ gen vectơ (vg) được nạp cho mỗi làn trên gel agaroza 1% 1xTAE và được điện di trong 60 phút ở độ dài gel 7 V/cm. Sau đó gel này được nhuộm trong dung dịch 2x GelRed (Biotium Cat# 41003) và được chụp ảnh bằng ChemiDocTMMP (Biorad). Các kết quả được thể hiện trong Fig. 24 chứng minh là các vectơ virut vCS01, vCS04, và vCS40 có bộ gen có cùng kích thước, được chỉ ra bởi dải riêng biệt trong khoảng 5kb (Fig. 24, làn 2-4). Mặc dù kích thước vectơ xấp xỉ 5,2 kb, bộ gen này là một dải đồng nhất xác nhận sự đóng gói chính xác bộ gen quá cỡ một chút (so với bộ gen kiểu hoang AAV dài 4,7 kb). Tất cả các chế phẩm vectơ vCS khác đều thể hiện cùng một kích thước bộ gen (số liệu không được thể hiện).

Để xác nhận mẫu protein capsit được mong đợi, SDS PAGE sau đó là nhuộm bạc được thực hiện với các vectơ vCS01, vCS04, và vCS40 (Fig. 25). Như được thể hiện trong hình vẽ này, quy trình tinh chế xuôi dòng tạo ra nguyên liệu được tinh chế cao thể hiện mẫu protein được mong đợi VP1, VP2 và VP3 (Fig. 25, làn 2-4). Mẫu giống như vậy được quan sát thấy với tất cả các chế phẩm virut khác (không được thể hiện). Quy trình SDS-PAGE của các chế phẩm AAV được thực hiện theo các quy trình chuẩn. Mỗi làn chứa 1E10 vg cấu trúc virut tương ứng, và được tách riêng trên gel Bis-Tris 4-12% (NuPAGE® Novex, Life Technologies) theo hướng dẫn của nhà sản xuất. Việc nhuộm bạc được thực hiện với kit SilverQuestTM (Novex, Life Technologies) theo hướng dẫn của nhà sản xuất.

Ngạc nhiên là, các vectơ AAV vCS01 và vCS04 có sự đóng gói virion cao hơn, được đo bằng năng suất cao hơn trong quá trình sản xuất virut AAV, so với cấu trúc mã hóa kiểu hoang vCS40 và các cấu trúc được tối ưu hóa codon còn lại. Như được thể hiện trong Bảng 3, các vectơ vCS01 và vCS04 sao chép tốt hơn đáng kể so với vCS40, tạo ra sự tăng năng suất 5-7 lần về hiệu giá AAV.

Bảng 3 – Năng suất cho mỗi lit môi trường nuôi cấy tế bào thu được với các cấu trúc vectơ AAV vCS01, vCS04, và vCD40, như được tinh chế từ các pelet tế bào.

Cấu trúc	Nồng độ vectơ [vg/ml] x10E12	Năng suất [vg /lit] x10E12	Số lần tăng so với dạng kiểu hoang
vCS40	2,0	11,0	-
vCS01	9,2	51,4	4,7
vCS04 – Mẫu 1	17,6	79,2	7,2
vCS04 – Mẫu 2	15,9	58,8	5,4

Ví dụ 2 – Biểu hiện in vivo các trình tự biểu hiện biến thể Yếu tố VIII được biến đổi codon

Để kiểm tra hiệu lực sinh học của các trình tự biến thể Yếu tố VIII được biến đổi codon, các cấu trúc FVIII typ ReFacto được mô tả trong Ví dụ 1 được sử dụng cho chuột thiếu Yếu tố VIII. Nói ngắn gọn, các thử nghiệm được thực hiện ở chuột bị bất hoạt (ko) C57Bl/6 FVIII (với 6-8 động vật mỗi nhóm) bằng cách tiêm tĩnh mạch đuôi các bộ gen vectơ 4E12 (vg) trên một kilogam thể trọng của chuột. Máu được lấy 14 ngày sau khi tiêm bằng cách chích sau hố mắt và huyết tương được điều chế và làm đông lạnh bằng cách sử dụng các quy trình chuẩn. Các mức biểu hiện ở ngày 14 được lựa chọn vì tác động của các kháng thể ức chế ở thời điểm này là nhỏ nhất, các kháng thể này được quan sát thấy ở một số động vật thuộc mô hình chuột ở các thời điểm muộn hơn. Hoạt tính FVIII trong huyết tương chuột được xác định bằng cách sử dụng thử nghiệm Technochrome FVIII được thực hiện, chỉ với các biến đổi nhỏ, như được gợi ý bởi nhà sản xuất (Technoclone, Vienna, Áo). Đối với thử nghiệm này, các mẫu huyết tương được pha loãng thích hợp và được trộn với chất phản ứng thử nghiệm, chứa thrombin, yếu tố IX hoạt hóa (FIXa), phospholipit, yếu tố X và canxi. Sau khi hoạt hóa FVIII bởi thrombin, phức hợp với FIXa, phospholipit và canxi được tạo thành. Phức hợp này hoạt hóa FX thành FX đã hoạt hóa (FXa), chất này lại tách para-nitroanilit (pNA) từ chất nền tạo màu. Các đặc điểm động học của sự tạo thành pNA được đo ở 405 nm. Tốc độ này tỷ lệ thuận với nồng độ FVIII trong mẫu này. Các nồng độ FVIII được đọc từ cuvet tham chiếu và kết quả được đưa ra trong IU FVIII/millilit.

Các kết quả này, được trình bày trong Bảng 4 dưới đây, chứng minh là các trình tự được biến đổi codon được thiết kế bằng cách sử dụng các thuật toán thương mại (CS10, CS11, và CH25) chỉ tạo ra sự tăng vừa phải BDD-Yếu tố VIII (3-4 lần) so với cấu trúc

BDD-Yếu tố VIII kiểu hoang (CS40). Tương tự, cấu trúc BDD-Yếu tố VIII được biến đổi codon được tạo ra như được mô tả trong Radcliffe et al. (CS08), chỉ làm tăng 3-4 lần sự biểu hiện BDD-FVIII. Kết quả này phù hợp với các kết quả được báo cáo trong Radcliff et al. Ngạc nhiên là, các cấu trúc CS01, CS04, và CS23 tạo ra sự biểu hiện BDD-FVIII cao hơn nhiều trong các thử nghiệm hiệu lực sinh học in-vivo (tăng lần lượt 18, 74, và 30 lần).

Bảng 4 – Sự biểu hiện FVIII trong huyết tương của chuột bất hoạt FVIII được cảm ứng bởi các cấu trúc vectơ AAV khác nhau.

Cấu trúc	Thuật toán codon	Sự biểu hiện FVIII trung bình ở ngày 14 [IU/ml]	Độ lệch chuẩn	Số lượng chuột	Số lần tăng so với dạng kiểu hoang
vCS40	Người kiểu hoang	0,03	0,03	12	-
vCS01	Của tác giả sáng chế	0,55	0,28	22	18,3
vCS04	Của tác giả sáng chế	2,21	1,20	55	73,7
vCS08	Radcliffe et al.	0,11	0,01	6	3,6
vCS10	Eurofins	0,09	0,01	7	3,0
vCS11	IDT	0,08	0,02	8	2,7
vCH25	GeneArt	0,13	0,12	18	4,3
vCS23	Của tác giả sáng chế	0,91	0,32	5	30,3

Ví dụ 3 – Thiết kế peptit glycosyl hóa cho cầu nối được thế vùng chức năng B

Các tác giả khác đã chỉ ra rằng việc đưa peptit nhỏ ("peptit V3") chứa sáu vị trí glycosyl hóa liên kết N giả định từ vùng chức năng B Yếu tố VIII kiểu hoang, vào cấu trúc liệu pháp gen bị mất vùng chức năng B, làm tăng mức nồng độ Yếu tố VIII trong huyết tương của chuột (McIntosh et al., Blood 121(17):3335-44 (2013)). Tuy nhiên, để duy trì kích thước nhỏ của cầu nối được thế vùng chức năng B, các vị trí glycosyl hóa được lấy ra trong trường hợp vùng chức năng B kiểu hoang. Dự đoán in silico (Gupta et al., ở trên) về cầu nối chứa peptit V3 gợi ý là chỉ hai trong số các vị trí glycosyl hóa này trong peptit V3 sẽ được biến đổi in vivo (Fig. 15).

Do đó, tác giả sáng chế nỗ lực phát hiện các peptit glycosyl hóa khác mà củng cố mức glycosyl hóa cao hơn trên in vivo, phù hợp với sự glycosyl hóa kiểu hoang chặt chẽ hơn so với peptit V3. Các tác giả sáng chế thiết kế và thử nghiệm một số peptit glycosyl hóa khác, in silico. Một số trong số các peptit này, được thể hiện trong Fig. 13A-13B, được dự đoán là có mức độ glycosyl hóa bằng hoặc lớn hơn trên in vivo so với peptit V3, khi nằm giữa các axit amin N768 và P769 của cầu nối được thế vùng chức năng B trong SEQ ID NO:2. Các kết quả của dự đoán in silico được thể hiện trong Bảng 5, dưới đây. Bảng 5 cũng báo cáo các kết quả của thử nghiệm biểu hiện được thực hiện cho một số cấu trúc mã hóa protein Yếu tố VIII typ ReFacto với peptit glycosyl hóa được đưa vào cầu nối được thế vùng chức năng B, trong nền được tối ưu hóa codon CS01.

Bảng 5 – Dự đoán mức độ N-glycosyl hóa trong các peptit cầu nối được thế vùng chức năng B và hiệu quả của các cấu trúc vectơ AAV trên in vivo.

Trình tự	Số lượng vị trí N-glycosyl hóa được dự đoán	Biểu hiện ở ngày 28 [IU/ml]	SD	Số lượng chuột [n]	Số lần tăng biểu hiện
vCS01	0	0,74	0,52	5	21
vNG1/CS01	4	Không xác định	-	-	-
vNG4/CS01	3	1,93	0,57	6	55
vNG5/CS01	2	Không xác định	-	-	-
vNG6/CS01	1	0,80	0,67	5	23
vNG9/CS01	1	Không xác định	-	-	-
vNG10/CS01	2	2,66	0,52	6	76
vNG16/CS01	2	1,59	0,57	6	45
vNG17/CS01	2	Không xác định	-	-	-
vNG18/CS01	2	Không xác định	-	-	-
vNG19/CS01	2	0,88	0,25	5	25
vNG20/CS01	2	Không xác định	-	-	-
vNG21/CS01	2	Không xác định	-	-	-

vCS40	0	0,035	0,030	12	1

Các vectơ AAV chứa các biến thể NG được tạo ra như đã mô tả trong Ví dụ 1 và được thử nghiệm ở chuột bất hoạt FVIII như được mô tả trong Ví dụ 2. Tất cả các vectơ virut (ngoại trừ vectơ đối chứng vCS40) được thể hiện trong **Bảng 5** là dựa trên thuật toán được sử dụng trong vCS01. Tập hợp các cấu trúc tương tự sử dụng thuật toán vCS04 cũng được tạo ra (chuỗi vNG/CS04) và được thử nghiệm trong mô hình chuột. Các kết quả được so sánh với mức độ biểu hiện đạt được với cấu trúc vCS40 kiểu hoang. Mức độ biểu hiện ở ngày 28 được chọn trong ví dụ này, vì các mức độ biểu hiện của phần lớn cấu trúc đạt tới mức cao nhất ở thời điểm này. Ba vectơ AAV đạt được các mức độ biểu hiện FVIII cao hơn 40 lần bao gồm vNG4/CS01, vNG10/CS01 và vNG16/CS01 (**Bảng 5**). Các cấu trúc tương đương vNG4/CS04, vNG10/CS04 và vNG16/CS04 được dự đoán thể hiện mức độ biểu hiện còn cao hơn vì chúng dựa trên thuật toán vCS04 cao hơn.

Ngạc nhiên là, các vectơ AAV của chuỗi vNG/CS01 có sự đóng gói virion cao hơn, được đo bằng năng suất cao hơn trong quy trình sản xuất virut AAV, so với cấu trúc mã hóa kiểu hoang vCS40. Như được thể hiện trong **Bảng 6**, các vectơ dựa trên vNG/CS01 sao chép tốt hơn đáng kể so với vCS40, làm tăng năng suất xấp xỉ 3 lần trong hiệu giá AAV.

Bảng 6 – Năng suất trên mỗi lit môi trường nuôi cấy tế bào thu được với các cấu trúc vectơ AAV được tinh chế từ pelet tế bào.

Trình tự	Nồng độ vectơ [vg/ml] x10^{12}	Năng suất [vg/lit] x10^{12}	Số lần tăng so với dạng kiểu hoang
vCS01	9,17	51,35	4,7
vNG1/CS01	2,13	17,04	1,5
vNG4/CS01	5,74	33,01	3,0
vNG5/CS01	6,91	27,29	2,5
vNG6/CS01	7,01	40,66	3,7
vNG9/CS01	6,39	29,39	2,7
vNG10/CS01	8,57	37,71	3,4
vNG16/CS01	5,3	28,36	2,6
vNG17/CS01	4,24	32,22	2,9
vNG18/CS01	6,11	37,88	3,4
vNG19/CS01	9,42	39,56	3,6
vNG20/CS01	4,09	30,27	2,8

vNG21/CS01	Không xác định	-	-
vCS40	2,03	11	1,0

Ví dụ 4 – Xây dựng cấu trúc BDD-FVIII đột biến

Nhiều cấu trúc BDD-FVIII typ Refacto được gây đột biến khác nhau, mang các đột biến axit amin trong chuỗi nặng Yếu tố VIII và/hoặc cầu nối được thế vùng chức năng B, được tách dòng và sàng lọc. Các vectơ tương đương, được gọi ở đây là chuỗi "vCS" của các vectơ, mã hóa các biến thể BDD-FVIII trong nền được biến đổi codon CS01, CS04, và CS23. Phương pháp được sử dụng để tạo ra các nền CS01 và CS04 được mô tả trong Ví dụ 1. Phương pháp được sử dụng để tạo ra CS23 là dựa trên công cụ JCAT (www.jcat.de), công cụ trực tuyến để tối ưu hóa codon (Grote et al., 2005; Nucl. Acids Res. W526-31). Trình tự này được biến đổi thêm để phản ánh tốt hơn việc sử dụng codon của siêu họ albumin (Mirsafian et al. Word Journal, ID 639682 (2014)), nội dung này được đưa vào đây bằng cách viện dẫn, toàn bộ, cho tất cả các mục đích.

Các tổ hợp của ba loại đột biến này được đưa vào các trình tự FVIII của chuỗi cấu trúc vCS. Sự thay đổi axit amin thứ nhất được đưa vào trình tự FVIII là đột biến X1 (TTYVNRSL (SEQ ID NO: 33); X. Xiao), đưa vị trí glycosyl hóa bổ sung vào gần cầu nối được thế vùng chức năng B. Đột biến X1 còn được gọi ở đây là đột biến "m3". Sự thay đổi axit amin thứ hai được tạo ra trong trình tự FVIII bao gồm đột biến F328S (SPI, F309S SPE), sự thay đổi axit amin được biết là cải thiện sự tiết FVIII (Swaaroop, J. Biol. Chem., 272:24121–24 (1997)). Đột biến này còn được gọi ở đây là đột biến "m1". Sự thay đổi thứ ba là đột biến được gọi là X5, là tổ hợp của năm thay đổi axit amin trong vùng chức năng A1 của chuỗi nặng mà cải thiện hoạt tính đặc hiệu và sự tiết BDD-FVIII (Cao et al., 2014; ASGCT abstract #460; chi tiết về các đột biến này được mô tả trong bài thuyết trình miệng). Đột biến X5 còn được gọi ở đây là đột biến "m2". Tiếp theo, các tổ hợp của X1 và F328S (SPI, F309S SPE) được tạo ra, sau đó là các tổ hợp của X1 và X5, còn được gọi là "X6", và các tổ hợp khác của X5 và F328S (SPI, F309S SPE) được tạo ra (Bảng 7).

Tổng hợp gen và tách dòng các plasmit vectơ. Các plasmit được tạo ra bằng cách tách dòng nhiều đoạn ADN tổng hợp khác nhau vào cùng một plasmit khung vectơ (pCh-BB01). Quá trình tổng hợp ADN của các đoạn BDD-FVIII typ Refacto có các vị trí giới hạn enzym *Asc*I và *Not*I bên sườn được thực hiện bởi ThermoFischer Scientific (Regensburg, Đức). Khung vectơ chứa hai đoạn lặp đầu cùng đảo (ITR) thu được từ AAV2

bên sườn mà bao gồm trình tự khởi động/trình tự tăng cường thu được từ gen transthyretin chuột đặc hiệu gan, các vị trí giới hạn enzym *Asc*I và *Not*I để gắn xen BDD-FVIII typ Refacto tương ứng, và vị trí polyA tổng hợp. Sau khi nối khung vectơ được tạo ra và các đoạn gắn xen qua vị trí *Asc*I và *Not*I, các plasmit tạo thành được khuếch đại theo quy mô miligam. Các trình tự BDD-FVIII typ Refacto của các cấu trúc này được kiểm tra bằng cách giải trình tự trực tiếp (Microsynth, Balgach, Thụy Sĩ). Việc tách dòng tạo ra nhiều cấu trúc plasmit khác nhau, như được thể hiện trong Fig. 44.

Các chế phẩm vectơ quy mô nhỏ và định lượng bằng PCR định lượng (qPCR). Các vectơ dựa trên AAV8 được tạo ra bằng phương pháp chuyển nhiễm ba plasmit như được mô tả trong Grieger et al. (2015, ở trên). Các tế bào huyền phù HEK293 được sử dụng để chuyển nhiễm plasmit bằng cách sử dụng plasmit vectơ FVIII tương ứng, plasmit pXX6X80 trợ giúp (mang gen trợ giúp adenovirut), và plasmit đóng gói pGSK2/8 (đóng góp các gen rep2 và cap8). Trong quy trình xuôi dòng, pelet tế bào của một lit môi trường nuôi cấy được xử lý bằng cách sử dụng các gradien iodixanol như được mô tả trên đây. Quy trình này tạo ra các chế phẩm vectơ như được chỉ ra trong Bảng 8. Các vectơ được định lượng bằng qPCR bằng cách sử dụng quy trình qPCR toàn cầu gắn đích các đoạn lặp đầu cùng đảo AAV2 (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)). Plasmit vectơ được định lượng chính xác mang các đoạn lặp đầu cùng đảo AAV2 đóng vai trò tạo đường cong chuẩn.

Mô tả vectơ AAV. Tính nguyên vẹn của các bộ gen vectơ được phân tích bằng điện di gen agaroza AAV. Kỹ thuật điện di được thực hiện tương tự như đã mô tả trong Fagone et al. (Human Gene Therapy Methods, 23:1-7 (2012)). Các chế phẩm vectơ AAV được ủ ở 75°C trong 10 phút khi có mặt 0,5% SDS và sau đó được làm nguội xuống nhiệt độ phòng. Xấp xỉ 1,5E10 bộ gen vectơ (vg) được nạp cho mỗi làn trên gel agaroza 1% 1xTAE và được điện di trong 60 phút ở 7 V/cm độ dài gel. Sau đó gel được nhuộm trong dung dịch 2x GelRed (Biotium Cat# 41003) và được chụp ảnh bằng ChemiDoc™MP (Biorad). Kết quả của việc chọn lọc vectơ được thể hiện trong Fig. 45. Các vectơ virut vCS04 (đối chứng), vCS17, vCS20, vCS24, vCS16 và vCS40 (đối chứng) thể hiện toàn bộ bộ gen có cùng kích thước với dải riêng biệt trong khoảng 5kb (Fig. 45, các làn 2-7; mũi tên bên phải). Mặc dù kích thước vectơ xấp xỉ 5,2 kb, bộ gen này là dải đồng nhất xác nhận việc đóng gói chính xác bộ gen quá cỡ một chút (so với bộ gen kiểu hoang AAV dài 4,7 kb).

Để xác nhận độ tinh khiết của vectơ này và mẫu protein capsit được dự đoán, SDS PAGE sau đó là nhuộm bạc được thực hiện với các vectơ, như được thể hiện trong Fig. 46. Như được thể hiện trong hình vẽ này, quy trình tinh chế xuôi dòng tạo ra nguyên liệu tinh khiết cao thể hiện mẫu protein mong đợi là VP1, VP2 và VP3 (Fig. 46 làn 2-9; mũi tên bên phải). Quy trình SDS-PAGE của các chế phẩm AAV được thực hiện theo các quy trình chuẩn. Lượng 1E10 vg mỗi làn được tách riêng trên gel Bis-Tris 4-12% (NuPAGE® Novex, Life Technologies) theo hướng dẫn của nhà sản xuất. Kỹ thuật nhuộm bạc được thực hiện bằng kit SilverQuest™ (Novex, Life Technologies) theo hướng dẫn của nhà sản xuất.

Sàng lọc hiệu lực sinh học in-vivo các vectơ. Các cấu trúc BDD-FVIII typ Refacto khác nhau được sàng lọc ở chuột. Các thử nghiệm được thực hiện ở chuột bị bất hoạt (ko) C57B1/6 FVIII (với 6-8 động vật mỗi nhóm) bằng cách tiêm tĩnh mạch đuôi các bộ gen vectơ 4E12 (vg) trên một kilogam thể trọng của chuột. Máu được lấy 14 ngày sau khi tiêm bằng cách chích sau hố mắt và huyết tương được điều chế và làm đông lạnh bằng cách sử dụng các quy trình chuẩn. Hoạt tính FVIII trong huyết tương chuột được xác định bằng thử nghiệm tạo màu từ Technoclone với các biến đổi nhỏ (Technochrome FVIII, Technoclone, Vienna, Áo). Nói ngắn gọn, mẫu huyết tương được pha loãng thích hợp và được trộn với các chất phản ứng thử nghiệm, chứa thrombin, yếu tố IX được hoạt hóa (FIXa), phospholipit, yếu tố X và canxi. Sau khi hoạt hóa FVIII bởi thrombin, phức hợp với FIXa, phospholipit và canxi được tạo thành. Phức hợp này hoạt hóa FX thành FX đã hoạt hóa (FXa), chất này lại tách para-nitroanilit (pNA) từ chất nền tạo màu. Các đặc điểm động học của sự tạo thành pNA được đo ở 405 nm. Tốc độ này tỷ lệ thuận với nồng độ FVIII trong mẫu này. Các nồng độ FVIII được đọc từ cuvet tham chiếu và kết quả được đưa ra với đơn vị IU FVIII/millilit.

Kết quả của thử nghiệm hiệu lực sinh học ở chuột (số liệu biểu hiện vào ngày 14 của FVIII theo đơn vị quốc tế trên mỗi mililit [IU/ml] trong huyết tương chuột và số lần tăng biểu hiện so với đối chứng vCS40 kiểu hoang) được thể hiện trong Bảng 7. Tất cả các vectơ AAV vCS19, vCS26 và vCS32 đều chứa vị trí glycosyl hóa X1 trong nền được biến đổi codon lần lượt là CS01, CS04, và CS23. Như được thấy trong Bảng 7, các mức biểu hiện cao bất ngờ đạt được, so với cấu trúc kiểu hoang vCS40 (mức được xác định là 1). vCS26, ví dụ, biểu hiện mức độ cao hơn 202 lần so với vectơ vCS40 kiểu hoang. Một cấu

trúc đối chứng khác cho chuỗi vectơ X1, vCH111, mà chứa đột biến X1 trong trường hợp codon Geneart, thể hiện sự tăng biểu hiện vừa phải hơn (12 lần).

Tất cả các vectơ vCS16, vCS28, và vCS34 đều chứa đột biến F328S (SPI, F309S SPE) làm tăng sự tiết trong nền được biến đổi codon lần lượt là CS01, CS04, và CS23. Như được quan sát thấy trong Bảng 7, mức độ biểu hiện cao (cao hơn 45-93 lần so với đối chứng vCS40 kiểu hoang) đạt được với vCS16 và vCS28.

Các vectơ vCS20, vCS24, và vCS33 chứa đột biến X5 trong các nền được biến đổi codon lần lượt là CS01, CS04, và CS23. Biến thể hiệu quả nhất trong chuỗi X5 là vCS20, đạt được các mức >3 đơn vị/ml sau ngày 14 và tăng 121 lần so với đối chứng vCS40 kiểu hoang.

Các vectơ vCS17, vCS29, và vCS31 chứa tổ hợp của các đột biến X1 và F328S (SPI, F309S SPE) trong nền được biến đổi codon lần lượt là CS01, CS04, và CS23 (Bảng 6). Các cấu trúc vCS17 và vCS29 đạt được các mức biểu hiện rất cao trong các nghiên cứu ở chuột (tăng từ 115 đến 246 lần so với đối chứng vCS40). Đáng chú ý, trong mô hình chuột FVIII KO được sử dụng, phần lớn chuột được điều trị bằng cấu trúc vCS17 không phát triển các kháng thể trung hòa theo thời gian, được chứng minh bằng các mức tăng dần của FVIII ở các thời điểm muộn hơn (ví dụ, ngày 28 và ngày 42; số liệu không được thể hiện). Đây là phát hiện bất ngờ, vì trong một số cấu trúc khác, các mức biểu hiện bắt đầu giảm theo thời gian do sự tạo thành các kháng thể trung hòa. Nền CS01 kết hợp với các đột biến làm tăng sự tiết F328S (SPI, F309S SPE) và X1 gây ra sự cảm ứng khả năng tạo miễn dịch thấp.

Các vectơ vCS18, vCS27, và vCS35 chứa tổ hợp của các đột biến X1 và X5 trong các nền được biến đổi codon lần lượt là CS01, CS04, và CS23. Tổ hợp của hai đột biến này cũng là rất hiệu quả. Sự tăng 145 lần so với đối chứng vCS40 có thể đạt được với, ví dụ, vCS18 (Bảng 7).

Các vectơ vCS48 và vCS49 chứa tổ hợp của các đột biến X5 và F328S (SPI, F309S SPE) trong các nền được biến đổi codon lần lượt là CS01 và CS04. Tổ hợp của hai đột biến này cũng là rất hiệu quả. Một trong các sự tăng lớn nhất trong số tất cả các đột biến, sự tăng 239 lần so với đối chứng vCS40, có thể đạt được với vCS49 xác nhận giá trị đặc biệt của các tổ hợp bao gồm đột biến F328S (SPI, F309S SPE).

Quan sát bất ngờ hơn là vectơ AAV đột biến phát triển tốt hơn đáng kể so với cấu trúc vCS40 mang các codon BDD-FVIII kiểu hoang. Việc tối ưu hóa trình tự gây ra sự tăng năng suất vài lần trong quá trình sản xuất vectơ. Trong một số cấu trúc biểu hiện tốt nhất (ví dụ, vCS29, vCS17, vCS20, và vCS26) sự tăng năng suất do biến đổi codon và/hoặc trình tự đột biến là cao hơn xấp xỉ 3-5 lần, so với vectơ kiểu hoang (Bảng 8).

Sự biểu hiện của BDD-FVIII trong huyết tương của chuột bị bất hoạt FVIII được cảm ứng bởi các cấu trúc vectơ AAV khác nhau được thể hiện trong Bảng 7. Các cấu trúc này có cùng một khung vectơ, tuy nhiên, mã hóa các typ FVIII đột biến khác nhau, bao gồm các nền tối ưu hóa codon khác nhau. Các mức biểu hiện vào ngày 14 được chọn vì ở thời điểm này sự ảnh hưởng của các kháng thể ức chế là tối thiểu, các kháng thể ức chế này thường được quan sát thấy ở một số động vật trong mô hình chuột ở các thời điểm muộn hơn. N.d., không xác định.

Bảng 7 – Số liệu về hiệu lực sinh học in vivo của các cấu trúc vCS.

#	Vectơ	Thuật toán, các đột biến	Sự biểu hiện ở ngày 14 [IU/ml]	SD	Số lượng chuột [n]	Số lần tăng biểu hiện
1	vCS19	CS01/X1	2,34	1,10	13	78
2	vCS26	CS04/X1	6,07	2,72	12	202
3	vCS32	CS23/X1	Không xác định	-	-	-
4	vCS16	CS01,F328S	1,35	0,88	6	45
5	vCS28	CS04/F328S	2,78	0,92	7	93
6	vCS34	CS23/F328S	Không xác định	-	-	-
7	vCS20	CS01/X5	3,62	1,96	21	121
8	vCS24	CS04/X5	0,79	0,89	18	26
9	vCS33	CS23/X5	Không xác định	-	-	Không xác định
10	vCS17	CS01, X1, F328S	3,44	1,92	20	115
11	vCS29	CS04, X1, F328S	7,39	2,64	9	246
12	vCS31	CS23, X1, F328S	Không xác định			Không xác định
13	vCS18	CS01, X1+X5 (X6)	4,34	2,50	6	145
14	vCS27	CS04, X1+X5 (X6)	8,03	3,97-	6-	268-
15	vCS35	CS23, X1+X5 (X6)	Không xác định	-	-	-

19	vCS48	CS01, X5, F328S	2,54	0,72	8	85
20	vCS49	CS04, X5, F328S	7,17	1,30	7	239
	đối chứng					
16	vCS40	Người kiểu hoang	0,03	0,03	12	1
17	vCh25	Geneart	0,13	0,12	18	4
18	vCh111	Geneart +X1	0,37	0,21	17	12

Bảng 8 – Năng suất trên mỗi lit môi trường nuôi cấy tế bào (hiệu quả đóng gói) thu được với các cấu trúc vectơ AAV khác nhau. Các vectơ này được tinh chế ngoài các pelet tế bào; n.d., không xác định.

	cấu trúc	Thuật toán, các đột biến	Nồng độ vectơ [vg/ml] $x10^{12}$	Năng suất [vg /lit] $x10^{12}$	Số lần tăng so với trình tự kiểu hoang
1	vCS19	CS01/X1	9,71	36	3,22
2	vCS26	CS04/X1	5,93	32	2,87
3	vCS32	CS23/X1	Không xác định	Không xác định	Không xác định
4	vCS16	CS01/F328S	6,51	29	2,56
5	vCS28	CS04/F328S	5,85	32	2,88
6	vCS34	CS23/F328S	Không xác định	Không xác định	Không xác định
7	vCS20	CS01/X5	9,90	50	4,48
8	vCS24	CS04/X5	3,00	16	1,46
9	vCS33	CS23/X5	Không xác định	Không xác định	Không xác định
10	vCS17	CS01, X1, F328S	8,94	37	3,34
11	vCS29	CS04, X1, F328S	7,42	53	4,72
12	vCS31	CS23, X1, F328S	Không xác định	Không xác định	Không xác định
13	vCS18	CS01, X1+X5 (X6)	21,20	53	4,75
14	vCS27	CS04, X1+X5 (X6)	4,15	19	1,67
15	vCS35	CS23, X1+X5 (X6)	Không xác định	Không xác định	Không xác định
16	vCS48	CS01, X5, F328S	7,14	42,1	3,77
17	vCS49	CS04, X5, F328S	8,27	37,2	3,33
18	vCS40	Người kiểu hoang	2,03	11	1,00

Được hiểu là các ví dụ và các phương án được mô tả ở đây chỉ nhằm mục đích minh họa và nhiều biến đổi hoặc thay đổi khác nhau của chúng sẽ được gợi ý cho người có kiến thức trung bình trong lĩnh vực này và được bao gồm trong tinh thần và phạm vi của đơn này và phạm vi của yêu cầu bảo hộ kèm theo. Tất cả các công bố, patent, và các đơn yêu cầu cấp patent được trích dẫn ở đây được đưa vào đây bằng cách viện dẫn toàn bộ vì tất cả các mục đích.

Viện dẫn chéo đến các đơn liên quan

Đơn này yêu cầu hưởng quyền ưu tiên từ đơn yêu cầu cấp Patent tạm thời Mỹ số 62/255,317, nộp ngày 13 tháng 11 năm 2015, nội dung của tài liệu này được đưa vào đây bằng cách viện dẫn, toàn bộ và cho tất cả các mục đích

Danh mục trình tự

Đơn này chứa Danh mục trình tự đã được nộp bằng hình thức điện tử ở dạng thức ASCII và được đưa toàn bộ vào đây bằng cách viện dẫn. Bản sao ASCII này, được tạo ra vào ngày 9 tháng 11 năm 2016, có tên là 008073_5107_WO_Sequence_Listing.txt và có kích thước 353.479 byt.

YÊU CẦU BẢO HỘ

1. Polynucleotit chứa trình tự nucleotit mã hóa polypeptit Yếu tố VIII, polypeptit Yếu tố VIII này chứa chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng này với đầu cùng N của chuỗi nhẹ này,

trong đó, chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất ít nhất là 99% với toàn bộ độ dài của SEQ ID NO: 3;

trong đó chuỗi nhẹ của polypeptit Yếu tố FVIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất ít nhất là 99% với toàn bộ độ dài của SEQ ID NO: 4; và

trong đó cầu nối polypeptit chứa vị trí cắt furin và peptit glycosyl hóa có trình tự axit amin SEQ ID NO: 55.

2. Polynucleotit theo điểm 1, trong đó cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với toàn bộ độ dài của SEQ ID NO: 38.

3. Polynucleotit theo điểm 1, trong đó cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba SEQ ID NO: 38.

4. Polynucleotit theo điểm 1, trong đó:

trình tự nucleotit thứ nhất là SEQ ID NO: 3; và

trình tự nucleotit thứ hai là SEQ ID NO: 4.

5. Polynucleotit theo điểm 4, trong đó cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba SEQ ID NO: 38.

6. Polynucleotit theo điểm 1, trong đó polynucleotit này chứa thêm trình tự khởi động được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

7. Polynucleotit theo điểm 6, trong đó trình tự khởi động này là trình tự khởi động đặc hiệu gan ngược dòng trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

8. Polynucleotit theo điểm 7, trong đó polynucleotit này chứa thêm trình tự intron nằm giữa trình tự khởi động đặc hiệu gan và trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

9. Vectơ virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 1.

10. Hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 1.

11. Tế bào chủ được gây nhiễm bằng hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 1.

12. Quy trình in vitro sản sinh hạt virut liên hợp adeno (AAV) bao gồm bước đưa polynucleotit theo điểm 1 vào tế bào chủ động vật có vú, trong đó polynucleotit này có khả năng sao chép trong tế bào chủ động vật có vú.

13. Quy trình in vitro tải nạp tế bào chủ bao gồm bước cho tế bào chủ tiếp xúc với hạt virut liên hợp adeno (AAV) theo điểm 10.

14. Polynucleotit chứa trình tự nucleotit có sự đồng nhất ít nhất là 99% với toàn bộ độ dài của SEQ ID NO: 1, trong đó polynucleotit này mã hoá polypeptit Yếu tố VIII có các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, so với SEQ ID NO: 19.

15. Polynucleotit theo điểm 14, trong đó:

polypeptit Yếu tố VIII chứa chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng này với đầu cùng N của chuỗi nhẹ này; và

cầu nối polypeptit chứa peptit glycosyl hoá có trình tự axit amin SEQ ID NO: 55 được chèn giữa chuỗi nặng và chuỗi nhẹ.

16. Polynucleotit theo điểm 14, trong đó polynucleotit này chứa trình tự SEQ ID NO: 94.

17. Polynucleotit theo điểm 14, trong đó polynucleotit này chứa thêm trình tự khởi động được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

18. Polynucleotit theo điểm 17, trong đó trình tự khởi động là trình tự khởi động đặc hiệu gan ngược dòng trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

19. Polynucleotit theo điểm 18, trong đó polynucleotit này chứa thêm trình tự intron

nằm giữa trình tự khởi động đặc hiệu gan và trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

20. Vectơ virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 14.

21. Hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 14.

22. Tế bào chủ được gây nhiễm bằng hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 14.

23. Quy trình in vitro sản sinh hạt virut liên hợp adeno (AAV) bao gồm bước đưa polynucleotit theo điểm 14 vào tế bào chủ động vật có vú, trong đó polynucleotit này có khả năng sao chép trong tế bào chủ động vật có vú.

24. Quy trình in vitro tải nạp tế bào chủ bao gồm bước cho tế bào chủ tiếp xúc với hạt virut liên hợp adeno (AAV) theo điểm 21.

25. Polynucleotit chứa trình tự nucleotit mã hóa polypeptit Yếu tố VIII, polypeptit Yếu tố VIII này chứa chuỗi nhẹ, chuỗi nặng, và cầu nối polypeptit nối đầu cùng C của chuỗi nặng này với đầu cùng N của chuỗi nhẹ này, polypeptit Yếu tố VIII này có các đột biến thế axit amin I105V, A127S, G151K, M166T, và L171P, so với SEQ ID NO: 19,
trong đó chuỗi nặng của polypeptit Yếu tố VIII được mã hóa bởi trình tự nucleotit thứ nhất có sự đồng nhất ít nhất là 99% với toàn bộ độ dài của SEQ ID NO: 3;
trong đó chuỗi nhẹ của polypeptit Yếu tố FVIII được mã hóa bởi trình tự nucleotit thứ hai có sự đồng nhất ít nhất là 99% với toàn bộ độ dài của SEQ ID NO: 4); và
trong đó cầu nối polypeptit chứa vị trí cắt furin và peptit glycosyl hoá có trình tự axit amin SEQ ID NO: 55.

26. Polynucleotit theo điểm 25, trong đó cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba có sự đồng nhất ít nhất là 95% với toàn bộ độ dài của SEQ ID NO: 38.

27. Polynucleotit theo điểm 25, trong đó cầu nối polypeptit được mã hóa bởi trình tự nucleotit thứ ba SEQ ID NO: 38.

28. Polynucleotit theo điểm 25, trong đó polynucleotit này chứa thêm trình tự khởi động được liên kết chức năng với polynucleotit mã hóa polypeptit Yếu tố VIII.

29. Polynucleotit theo điểm 28, trong đó trình tự khởi động là trình tự khởi động đặc hiệu gan ngược dòng trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

30. Polynucleotit theo điểm 29, trong đó polynucleotit này chứa thêm trình tự intron nằm giữa trình tự khởi động đặc hiệu gan và trình tự nucleotit mã hóa polypeptit Yếu tố VIII.

31. Vectơ virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 25.

32. Hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 25.

33. Tế bào chủ được gây nhiễm bằng hạt virut liên hợp adeno (AAV) chứa polynucleotit theo điểm 25.

34. Quy trình in vitro sản sinh hạt virut liên hợp adeno (AAV) bao gồm bước đưa polynucleotit theo điểm 25 vào tế bào chủ động vật có vú, trong đó polynucleotit này có khả năng sao chép trong tế bào chủ động vật có vú.

35. Quy trình in vitro tải nạp tế bào chủ bao gồm bước cho tế bào chủ tiếp xúc với hạt virut liên hợp adeno (AAV) theo điểm 32.

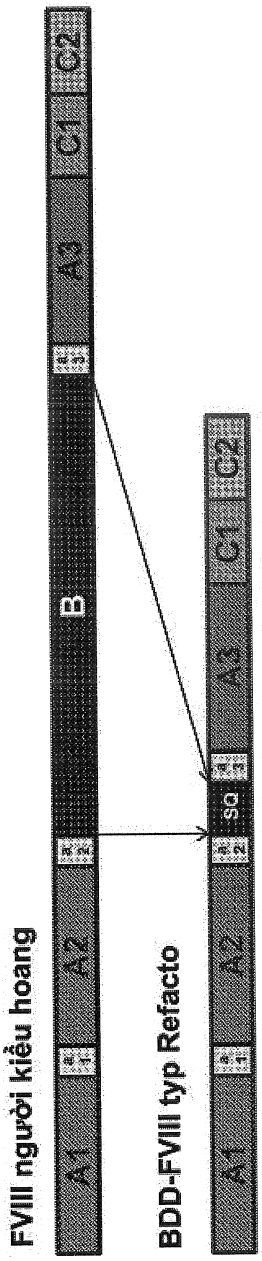

FVIII người kiểu hoang
A1
a1
a3
A2
a2
B
A3
C1
C2
BDD-FVIII typ Refacto
A1
a1
a3
A2
a2
SQ
A3
C1
C2

Fig. 1

CS04-FL-NA

atgcagattgagctgagcacctgcttcttcctgtgcctgctgaggttctgcttctctgccaccagga
gatactacctgggggctgtggagctttcttgggactacatgcagtctgacctgggggagctgcctgt
ggatgccaggttcccacccagagtgcccaaatccttcccattcaacacctctgtggtctacaagaag
accctctttgtggagttcactgaccacctgttcaacattgccaaacccaggccaccctggatgggac
tcctgggacccaccattcaggctgaggtgtatgacactgtggtcatcaccctcaagaacatggcctc
ccaccctgtgagcctgcatgctgtgggggtcagctactggaaggcctctgaggggctgagtatgat
gaccagacctcccagagggagaaggaggatgacaaagtgttccctgggggcagccacacctatgtgt
ggcaggtcctcaaggagaatggccccatggcctctgacccactctgcctgacctactcctacctttc
tcatgtggacctggtcaaggacctcaactctggactgattggggccctgctggtgtgcagggagggc
tccctggccaaagagaagacccagaccctgcacaagttcattctcctgtttgctgtctttgatgagg
gcaagagctggcactctgaaaccaagaactccctgatgcaggacagggatgctgcctctgccagggc
ctggcccaagatgcacactgtgaatggctatgtgaacaggagcctgcctggactcattggctgccac
aggaaatctgtctactggcatgtgattggcatggggacaacccctgaggtgcactccatttcctgg
agggccacaccttcctggtcaggaaccacagacaggccagcctggagatcagccccatcaccttcct
cactgcccagaccctgctgatggacctcggacagttcctgctgttctgccacatcagctcccaccag
catgatggcatggaggcctatgtcaaggtggacagctgccctgaggagccacagctcaggatgaaga
acaatgaggaggctgaggactatgatgatgacctgactgactctgagatggatgtggtccgctttga
tgatgacaacagcccatccttcattcagatcaggtctgtggccaagaaacaccccaagacctgggtg
cactacattgctgctgaggaggaggactgggactatgcccactggtcctggcccctgatgacagga
gctacaagagccagtacctcaacaatggcccacagaggattggacgcaagtacaagaaagtcaggtt
catggcctacactgatgaaaccttcaagaccagggaggccattcagcatgagtctggcatcctgggc
ccactcctgtatgggaggtgggggacaccctgctcatcatcttcaagaaccaggcctccaggccct
acaacatctacccacatggcatcactgatgtcaggcccctgtacagccgcaggctgccaaaggggt
gaaacacctcaaggacttccccattctgcctgggagatcttcaagtacaagtggactgtcactgtg
gaggatggaccaaccaaatctgaccccaggtgcctcaccagatactactccagctttgtgaacatgg
agagggacctggcctctggcctgattggcccactgctcatctgctacaaggagtctgtggaccagag
gggaaaccagatcatgtctgacaagaggaatgtgattctgttctctgtctttgatgagaacaggagc
tggtacctgactgagaacattcagcgcttcctgcccaaccctgctggggtgcagctggaggaccctg
agttccaggccagcaacatcatgcactccatcaatggctatgtgtttgacagcctccagctttctgt
ctgcctgcatgaggtggcctactggtacattctttctattggggcccagactgacttcctttctgtc
ttcttctctggctacaccttcaaacacaagatggtgtatgaggacaccctgaccctcttcccattct
ctggggagactgtgttcatgagcatggagaaccctggcctgtggattctgggatgccacaactctga
cttccgcaacaggggcatgactgccctgctcaaagtctcctcctgtgacaagaacactggggactac
tatgaggacagctatgaggacatctctgcctacctgctcagcaagaacaatgccattgagcccagga
gcttcagccagaatccacctgtcctgaaacgccaccagagggagatcaccaggaccaccctccagtc
tgaccaggaggagattgactatgatgacaccatttctgtggagatgaagaaagaggactttgacatc
tatgacgaggacgagaaccagagcccaaggagcttccagaagaagaccaggcactacttcattgctg
ctgtggagcgcctgtgggactatggcatgagctccagccccatgtcctcaggaacagggcccagtc
tggctctgtgccacagttcaagaaagtggtcttccaagagttcactgatggcagcttcacccagccc
ctgtacagaggggagctgaatgagcacctgggactcctgggcccatacatcagggctgaggtggagg
acaacatcatggtgaccttccgcaaccaggcctccaggccctacagcttctacagctccctcatcag
ctatgagGaggaccagaggcaggggctgagccacgcaagaactttgtgaaacccaatgaaaccaag
acctacttctggaaagtccagcaccacatggcccccaccaaggatgagtttgactgcaaggcctggg

(còn tiếp)

Fig. 2A

cctacttctctgatgtggacctggagaaggatgtgcactctggcctgattggcccactcctggtctg
ccacaccaacaccctgaaccctgcccatggaaggcaagtgactgtgcaggagtttgccctcttcttc
accatctttgatgaaaccaagagctggtacttcactgagaacatggagcgcaactgcagggccccat
gcaacattcagatggaggaccccaccttcaaagagaactaccgcttccatgccatcaatggctacat
catggacaccctgcctgggcttgtcatggcccaggaccagaggatcaggtggtacctgctttctatg
ggctccaatgagaacattcactccatccacttctctgggcatgtcttcactgtgcgcaagaaggagg
agtacaagatggccctgtacaacctctaccctggggtctttgagactgtggagatgctgccctccaa
agctggcatctggagggtggagtgcctcattggggagcacctgcatgctggcatgagcaccctgttc
ctggtctacagcaacaagtgccagacccccctgggaatggcctctggccacatcagggacttccaga
tcactgcctctggccagtatggccagtgggccccaagctggccaggctccactactctggatccat
caatgcctggagcaccaaggagccattcagctggatcaaagtggacctgctggcccccatgatcatc
catggcatcaagacccaggggccaggcagaagttctccagcctgtacatcagccagttcatcatca
tgtacagcctggatggcaagaaatggcagacctacagaggcaactccactggaacactcatggtctt
ctttggcaatgtggacagctctggcatcaagcacaacatcttcaacccccccaatcatcgccagatac
atcaggctgcaccccacccactacagcatccgcagcaccctcaggatggagctgatgggctgtgacc
tgaactcctgcagcatgcccctgggcatggagagcaaggccatttctgatgcccagatcactgcctc
cagctacttcaccaacatgtttgccacctggagcccaagcaaggccaggctgcacctccagggaagg
agcaatgcctggaggccccaggtcaacaacccaaaggagtggctgcaggtggacttccagaagacca
tgaaggtcactggggtgaccacccaggggtcaagagcctgctcaccagcatgtatgtgaaggagtt
cctgatcagctccagccaggatggccaccagtggaccctcttcttccagaatggcaaggtcaaggtg
ttccagggcaaccaggacagcttcacCcctgtggtgaacagcctggacccccccctcctgaccagat
acctgaggattcacccccagagctgggtccaccagattgccctgaggatggaggtcctgggatgtga
ggcccaggacctgtactga (SEQ ID NO:1)

Fig. 2B

CS04-FL-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDAR'FPPRVPKSFPFNTSVVYK
KTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEY
DDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCRE
GSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGC
HRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSH
QHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTW
VHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGIL
GPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVT
VEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENR
SWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLS
VFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGD
YYEDSYEDISAYLLSKNNAIEPRSFSQNPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFD
IYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQ
PLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNET
KTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALF
FTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLS
MGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTL
FLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI
IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIAR
YIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQG
RSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVK
VFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY (SEQ ID NO:2)

Fig. 3

CS04-HC-NA

gcc
accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg
ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac
acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt
gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat
gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg
ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg
gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc
aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat
gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag
ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc
tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat
gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg
acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac
agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg
gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag
gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat
gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc
tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc
cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc
aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc
attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg
ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact
gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc
cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca
accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg
gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag
aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag
aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg
cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg
tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct
attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag
atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc
atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc
atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac
agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagg
(SEQ ID NO:3)

Fig. 4

CS04-LC-NA

g agatcaccag gaccaccctc
cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag
gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc
aggcactact tcattgctgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc
catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggtcttc
caagagttca ctgatggcag cttcacccag cccctgtaca gaggggagct gaatgagcac
ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgaccttc
cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac
cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac
ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg
gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc
ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag
tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttcac tgagaacatg
gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac
taccgcttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggcc
caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc
atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg
tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc
tggagggtgg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttcctg
gtctacagca acaagtgcca gacccccctg ggaatggcct ctggccacat cagggacttc
cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac
tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg
ctggcccccа tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc
ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac
agaggcaact ccactggaac actcatggtc ttctttggca atgtggacag ctctggcatc
aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca ccccacccac
tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctcctgcagc
atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac
ttcaccaaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccagggaagg
agcaatgcct ggaggcccca ggtcaacaac ccaaaggagt ggctgcaggt ggacttccag
aagaccatga aggtcactgg ggtgaccacc caggggtca agagcctgct caccagcatg
tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc
cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac
agcctggacc ccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac
cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta c
(SEQ ID NO:4)

Fig. 5

BDLO01 - agc ttcttctcaga atccacctgt cctgaagaga caccagaga (SEQ ID NO:5)

BDLO04 - agc ttcagccaga atccacctgt cctgaaacgc caccagagg (SEQ ID NO:6)

BDLO23 - agc ttcagccaga acccccccgt gctgaagagg caccagagg (SEQ ID NO:7)

BDLNG1 - agcttcagccagaatGTGAGCAACAATGTGAGCAACAATGCCACCAATAATGCTACCAACccacctgtcctgaaacgccaccagagg (SEQ ID NO:36)

BDLNG4 - agcttcagccagaatGTGAGCAACAATGCCACCAACAATGTGAGCAACccacctgtcctgaaacgccaccagagg (SEQ ID NO:37)

BDLNG5 - agcttcagccagaatGTGAGCAATAATGCCACCAACccacctgtcctgaaacgccaccagagg (SEQ ID NO:38)

BDLNG6 - agcttcagccagaatGTGAGCAATAATccacctgtcctgaaacgccaccagagg (SEQ ID NO:39)

BDLNG9 - agcttcagccagaatAGGAGCCTGccacctgtcctgaaacgccaccagagg (SEQ ID NO:40)

BDLNG10 - agcttcagccagaatGCCACTAATGTGTCTAACAACTCTGCTACCTCTGCTGTGAGCccacctgtcctgaaacgccaccagagg
(SEQ ID NO:41)

BDLNG16 - agcttcagccagaatGCCACCAACTATGTGAACAGGAGCCTGccacctgtcctgaaacgccaccagagg
(SEQ ID NO:42)

BDLNG17 - agcttcagccagaatGCCACCAACTATGTGAACAGGAGCCTGTCTGCCACCTCTGCTGACTCTGCTGTGAGCCAGAATccacctgtcctgaaacgccaccag
agg(SEQ ID NO:43)

BDLNG18 - agcttcagccagaatGTGAGCAACAATGTGAGCAATGCTGTGTCTGCTGTCTGCTccacctgtcctgaaacgccaccagagg (SEQ ID NO:44)

BDLNG19 - agcttcagccagaatATCACTGTGGCCTCTGCCACCTCTAACATCACTGTGGCCTCTGCTGACCccacctgtcctgaaacgccaccagagg
(SEQ ID NO:45)

BDLNG20 - agcttcagccagaatATCACTGTGACCAACATCACTGTGACTGCCccacctgtcctgaaacgccaccagagg (SEQ ID NO:46)

BDLNG21 - agcttcagccagaatCAGACTGTGACCAACATCACTGTGACTGCCccacctgtcctgaaacgccaccagagg (SEQ ID NO:47)

BDLNGV - agcttcagccagaatGCCACTAATGTGTCTAACAACAGCAACACCAGCAACAGCAATGTGTCTccacctgtcctgaaacgccaccagagg
(SEQ ID NO:48)

Fig. 6

CS04-AV-NA

```
1     tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
61    cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
121   ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
181   accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc
241   attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat
301   tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt
361   tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cctcgagatt taaatgacgt
421   tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca aaggtcgccc
481   gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcaga gagggagtgg
541   ccaactccat cactaggggt tcctgagttt aaacttcgtc gacgattcga gcttgggctg
601   caggtcgagg gcactgggag gatgttgagt aagatggaaa actactgatg acccttgcag
661   agacagagta ttaggacatg tttgaacagg ggccgggcga tcagcaggta gctctagagg
721   atccccgtct gtctgcacat ttcgtagagc gagtgttccg atactctaat ctccctaggc
781   aaggttcata tttgtgtagg ttacttattc tccttttgtt gactaagtca ataatcagaa
841   tcagcaggtt tggagtcagc ttggcaggga tcagcagcct gggttggaag gaggggtat
901   aaaagccccc tcaccaggag aagccgtcac acagactagg cgcgccaccg ccaccatgca
961   gattgagctg agcacctgct tcttcctgtg cctgctgagg ttctgcttct ctgccaccag
1021  gagatactac ctgggggctg tggagctttc ttgggactac atgcagtctg acctgggggа
1081  gctgcctgtg gatgccaggt tcccacccag agtgcccaaa tccttcccat tcaacctc
1141  tgtggtctac aagaagaccc tctttgtgga gttcactgac cacctgttca acattgccaa
1201  acccaggcca ccctggatgg gactcctggg acccaccatt caggctgagg tgtatgacac
1261  tgtggtcatc accctcaaga acatggcctc ccaccctgtg agcctgcatg ctgtgggggt
1321  cagctactgg aaggcctctg aggggctga gtatgatgac cagacctccc agagggagaa
1381  ggaggatgac aaagtgttcc ctgggggcag ccacacctat gtgtggcagg tcctcaagga
1441  gaatggcccc atggcctctg acccactctg cctgacctac tcctaccttt ctcatgtgga
1501  cctggtcaag gacctcaact ctggactgat tggggccctg ctggtgtgca gggagggctc
1561  cctggccaaa gagaagaccc agaccctgca caagttcatt ctcctgtttg ctgtctttga
1621  tgagggcaag agctggcact ctgaaaccaa gaactccctg atgcaggaca gggatgctgc
1681  ctctgccagg gcctggccca agatgcacac tgtgaatggc tatgtgaaca ggagcctgcc
1741  tggactcatt ggctgccaca ggaaatctgt ctactggcat gtgattggca tggggacaac
1801  ccctgaggtg cactccattt tcctggaggg ccacaccttc ctggtcagga accacagaca
1861  ggccagcctg gagatcagcc ccatcacctt cctcactgcc cagaccctgc tgatggacct
1921  cggacagttc ctgctgttct gccacatcag ctcccaccag catgatggca tggaggccta
1981  tgtcaaggtg gacagctgcc ctgaggagcc acagctcagg atgaagaaca atgaggaggc
2041  tgaggactat gatgatgacc tgactgactc tgagatggat gtggtccgct ttgatgatga
2101  caacagccca tccttcattc agatcaggtc tgtggccaag aaacacccca agacctgggt
2161  gcactacatt gctgctgagg aggaggactg ggactatgcc ccactggtcc tggccctga
2221  tgacaggagc tacaagagcc agtacctcaa caatggccca cagaggattg gacgcaagta
2281  caagaaagtc aggttcatgg cctacactga tgaaaccttc aagaccaggg aggccattca
2341  gcatgagtct ggcatcctgg gcccactcct gtatggggag gtgggggaca ccctgctcat
2401  catcttcaag aaccaggcct ccaggcccta caacatctac ccacatggca tcactgatgt
2461  caggcccctg tacagccgca ggctgccaaa gggggtgaaa cacctcaagg acttccccat
```

Fig. 7A

```
2521 tctgcctggg gagatcttca agtacaagtg gactgtcact gtggaggatg gaccaaccaa
2581 atctgacccc aggtgcctca ccagatacta ctccagcttt gtgaacatgg agagggacct
2641 ggcctctggc ctgattggcc cactgctcat ctgctacaag gagtctgtgg accagagggg
2701 aaaccagatc atgtctgaca agaggaatgt gattctgttc tctgtctttg atgagaacag
2761 gagctggtac ctgactgaga acattcagcg cttcctgccc aaccctgctg gggtgcagct
2821 ggaggaccct gagttccagg ccagcaacat catgcactcc atcaatggct atgtgtttga
2881 cagcctccag ctttctgtct gcctgcatga ggtggcctac tggtacattc tttctattgg
2941 ggcccagact gacttccttt ctgtcttctt ctctggctac accttcaaac acaagatggt
3001 gtatgaggac accctgaccc tcttcccatt ctctggggag actgtgttca tgagcatgga
3061 gaaccctggc ctgtggattc tgggatgcca caactctgac ttccgcaaca ggggcatgac
3121 tgccctgctc aaagtctcct cctgtgacaa gaacactggg gactactatg aggacagcta
3181 tgaggacatc tctgcctacc tgctcagcaa gaacaatgcc attgagccca ggagcttcag
3241 ccagaatcca cctgtcctga aacgccacca gagggagatc accaggacca ccctccagtc
3301 tgaccaggag gagattgact atgatgacac catttctgtg gagatgaaga aagaggactt
3361 tgacatctat gacgaggacg agaaccagag cccaaggagc ttccagaaga agaccaggca
3421 ctacttcatt gctgctgtgg agcgcctgtg ggactatggc atgagctcca gccccatgt
3481 cctcaggaac agggcccagt ctggctctgt gccacagttc aagaaagtgg tcttccaaga
3541 gttcactgat ggcagcttca cccagccctt gtacagaggg gagctgaatg agcacctggg
3601 actcctgggc ccatacatca gggctgaggt ggaggacaac atcatggtga ccttccgcaa
3661 ccaggcctcc aggccctaca gcttctacag ctccctcatc agctatgagg aggaccagag
3721 gcaggggggt gagccacgca agaactttgt gaaacccaat gaaaccaaga cctacttctg
3781 gaaagtccag caccacatgg cccccaccaa ggatgagttt gactgcaagg cctgggccta
3841 cttctctgat gtggacctgg agaaggatgt gcactctggc ctgattggcc cactcctggt
3901 ctgccacacc aacaccctga accctgccca tggaaggcaa gtgactgtgc aggagtttgc
3961 cctcttcttc accatctttg atgaaaccaa gagctggtac ttcactgaga acatggagcg
4021 caactgcagg gccccatgca acattcagat ggaggacccc accttcaaag agaactaccg
4081 cttccatgcc atcaatggct acatcatgga caccctgcct gggcttgtca tggcccagga
4141 ccagaggatc aggtggtacc tgctttctat gggctccaat gagaacattc actccatcca
4201 cttctctggg catgtcttca ctgtgcgcaa gaaggaggag tacaagatgg ccctgtacaa
4261 cctctaccct ggggtctttg agactgtgga gatgctgccc tccaaagctg gcatctggag
4321 ggtggagtgc ctcattgggg agcacctgca tgctggcatg agcaccctgt tcctggtcta
4381 cagcaacaag tgccagaccc ccctgggaat ggcctctggc cacatcaggg acttccagat
4441 cactgcctct ggccagtatg gccagtgggc ccccaagctg gccaggctcc actactctgg
4501 atccatcaat gcctggagca ccaaggagcc attcagctgg atcaaagtgg acctgctggc
4561 ccccatgatc atccatggca tcaagaccca gggggccagg cagaagttct ccagcctgta
4621 catcagccag ttcatcatca tgtacagcct ggatggcaag aaatggcaga cctacagagg
4681 caactccact ggaacactca tggtcttctt tggcaatgtg gacagctctg gcatcaagca
4741 caacatcttc aacccccaa tcatcgccag atacatcagg ctgcacccca cccactacag
4801 catccgcagc accctcagga tggagctgat gggctgtgac ctgaactcct gcagcatgcc
4861 cctgggcatg gagagcaagg ccatttctga tgcccagatc actgcctcca gctacttcac
4921 caacatgttt gccacctgga gcccaagcaa ggccaggctg cacctccagg gaaggagcaa
4981 tgcctggagg ccccaggtca acaacccaaa ggagtggctg caggtggact tccagaagac
```

Fig. 7B

```
5041 catgaaggtc actggggtga ccacccaggg ggtcaagagc ctgctcacca gcatgtatgt
5101 gaaggagttc ctgatcagct ccagccagga tggccaccag tggaccctct tcttccagaa
5161 tggcaaggtc aaggtgttcc agggcaacca ggacagcttc acccctgtgg tgaacagcct
5221 ggaccccccc ctcctgacca gatacctgag gattcacccc cagagctggg tccaccagat
5281 tgccctgagg atggaggtcc tgggatgtga ggcccaggac ctgtactgat gacgagcggc
5341 cgctcttagt agcagtatcg ataataaaag atctttattt tcattagatc tgtgtgttgg
5401 ttttttgtgt gttaattaag ctcgcgaagg aacccctagt gatggagttg gccactccct
5461 ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaaa ggtcgcccga cgcccgggct
5521 ttgcccgggc ggcctcagtg agcgagcgag cgcgcagaga gggagtggcc aagacgattt
5581 aaatgacaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc
5641 tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat
5701 gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc
5761 tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg
5821 ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag
5881 cggtatcagc tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag
5941 gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc
6001 tggcgttttt ccataggctc cgcccccctg acgagcatca caaaaatcga cgctcaagtc
6061 agaggtggcg aaacccgaca ggactataaa gataccaggc gtttcccccc ggaagctccc
6121 tcgtgcgctc tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt
6181 cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg
6241 ttcgctccaa gctgggctgt gtgcacgaac ccccgttca gcccgaccgc tgcgccttat
6301 ccggtaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag
6361 ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt
6421 ggtggcctaa ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc
6481 cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta
6541 gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag
6601 atcctttgat cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga
6661 ttttggtcat gagattatca aaaaggatct tcacctagat cctttttaaat taaaaatgaa
6721 gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa
6781 tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc
6841 ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga
6901 taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa
6961 gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt
7021 gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg
7081 ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc
7141 aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg
7201 gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag
7261 cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt
7321 actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt
7381 caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac
7441 gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac
7501 ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag
7561 caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa
7621 tactcatact cttcctttt caatattatt gaagcattta tcagggttat tgtctcatga
7681 gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc
7741 cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa
7801 ataggcgtat cacgaggccc tttcgtc  (SEQ ID NO:8)
```

Fig. 7C

CS01m1-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGGA
GATACTACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCTGT
GGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAGAAG
ACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATGGGAC
TCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATGGCATC
CCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAGTATGAT
GACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACCTATGTGT
GGCAAGTCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCCTACCTTTC
TCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGCAGGGAAGGA
TCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTCTTTGATGAGG
GCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCCTCTGCCAGGGC
ATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTCATTGGCTGCCAC
AGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCACTCCATTTTCCTGG
AGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCTCCCATCACCTTCCT
CACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTCCTGCCACATCTCTTCCCACCAG
CATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCACAGCTCAGGATGAAGA
ACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCAGATTTGA
TGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAACACCCCAAGACATGGGTG
CACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTCCTGGCCCCTGATGACAGGA
GCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGAAAGTACAAGAAAGTCAGATT
CATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAGCATGAGTCTGGCATTCTGGGA
CCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTCAAGAACCAGGCCTCCAGGCCCT
ACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTACAGCAGGAGACTGCCAAAAGGGGT
GAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTCAAGTACAAGTGGACTGTCACTGTG
GAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGATACTACTCCTCTTTTGTGAACATGG
AGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATCTGCTACAAGGAGTCTGTGGACCAGAG
AGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTGTTCTCTGTCTTTGATGAGAACAGATCA
TGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAACCCTGCTGGGGTGCAACTGGAAGACCCTG
AGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGCTATGTGTTTGACTCTCTCCAGCTTTCTGT
CTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCTATTGGGGCACAAACTGACTTCCTTTCTGTC
TTCTTCTCTGGATACACCTTCAAGCACAAGATGGTGTATGAGGACACCCTGACACTCTTCCCATTCT
CTGGGGAAACTGTGTTCATGAGCATGGAGAACCCTGGACTGTGGATTCTGGGATGCCACAACTCTGA
CTTCAGAAACAGGGGAATGACTGCACTGCTCAAAGTCTCCTCCTGTGACAAGAACACTGGGGACTAC
TATGAGGACTCTTATGAGGACATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTGAGCCCAGAA
GCTTCTCTCAGAATCCACCTGTCCTGAAGAGACACCAGAGAGAGATCACCAGGACAACCCTCCAGTC
TGACCAGGAAGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATC
TATGATGAGGACGAGAACCAGTCTCCAAGATCATTCCAGAAGAAGACAAGACACTACTTCATTGCTG
CTGTGGAAAGACTGTGGGACTATGGCATGTCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTC
TGGCTCTGTGCCACAGTTCAAGAAAGTGGTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCC
CTGTACAGAGGGGAACTGAATGAGCACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAG
ACAACATCATGGTGACATTCAGAAACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAG
CTATGAGGAAGACCAGAGACAAGGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAG
ACCTACTTCTGGAAAGTCCAGCACCACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGG

(Còn tiếp)

Fig. 8A

CATACTTCTCTGATGTGGACCTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTG
CCACACCAACACCCTGAACCCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTC
ACCATCTTTGATGAAACCAAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCAT
GCAACATTCAGATGGAAGACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACAT
CATGGACACCCTGCCTGGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATG
GGATCCAATGAGAACATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGG
AATACAAGATGGCCCTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAA
AGCTGGCATCTGGAGGGTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTC
CTGGTCTACAGCAACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGA
TCACTGCCTCTGGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCAT
CAATGCATGGTCAACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATT
CATGGCATCAAGACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCA
TGTACTCTCTGGATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTT
CTTTGGCAATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATAC
ATCAGGCTGCACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACC
TGAACTCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATC
CTCTTACTTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGA
AGCAATGCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAA
TGAAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTT
CCTGATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGAT
ACCTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGA
GGCACAAGACCTGTACTGA (SEQ ID NO:49)

Fig. 8B

CS04Δ(760-1667) - CS04-SC1-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATAC
TACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCTGTGGATGCCAGG
TTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAGAAGACCCTCTTTGTGGAG
TTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATGGGACTCCTGGGACCCACCATTCAG
GCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATGGCCTCCCACCCTGTGAGCCTGCATGCTGTG
GGGGTCAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCTCCCAGAGGGAGAAGGAGGAT
GACAAAGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCT
GACCCACTCTGCCTGACCTACTCCTACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATT
GGGGCCCTGCTGGTGTGCAGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTC
CTGTTTGCTGTCTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGAT
GCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCACTCCATT
TTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGCCCCATCACCTTC
CTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCACATCAGCTCCCACCAGCAT
GATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCACAGCTCAGGATGAAGAACAATGAG
GAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCCGCTTTGATGATGACAACAGC
CCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAACACCCCAAGACCTGGGTGCACTACATTGCTGCTGAG
GAGGAGGACTGGGACTATGCCCCACTGGTCCTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAAC
AATGGCCCCACAGAGGATTGGACGCAAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAG
ACCAGGGAGGCCATTCAGCATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTG
CTCATCATCTTCAAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCC
CTGTACAGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGATACTAC
TCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATCTGCTACAAGGAG
TCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTGTTCTCTGTCTTTGATGAG
AACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAACCCTGCTGGGGTGCAGCTGGAGGAC
CCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGCTATGTGTTTGACAGCCTCCAGCTTTCTGTC
TGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCTATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTC
TCTGGCTACACCTTCAAACACAAGATGGTGTATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACT
GTGTTCATGAGCATGGAGAACCCTGGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGC
ATGACTGCCCTGCTCAAAGTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGAC
ATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTGAGCCCAGGGAGATCACCAGGACCACCCTCCAGTCT
GACCAGGAGGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGAC
GAGGACGAGAACCAGAGCCCAAGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGC
CTGTGGGACTATGGCATGAGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAG
TTCAAGAAAGTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAAT
GAGCACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC
CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAG
CCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCACATGGCCCCC

(còn tiếp)

Fig. 9A

ACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCT
GGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCTGCCCATGGAAGGCAAGTGACTGTG
CAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGCGC
AACTGCAGGGCCCCATGCAACATTCAGATGGAGGACCCCACCTTCAAAGAGAACTACCGCTTCCATGCCATC
AATGGCTACATCATGGACACCCTGCCTGGGCTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTT
TCTATGGGCTCCAATGAGAACATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAG
GAGTACAAGATGGCCCTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCT
GGCATCTGGAGGGTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTAC
AGCAACAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGCACCAAG
GAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCCAGGGGGCC
AGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGCAAGAAATGGCAG
ACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCAC
AACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTGCACCCCACCCACTACAGCATCCGCAGCACC
CTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATT
TCTGATGCCCAGATCACTGCCTCCAGCTACTTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGG
CTGCACCTCCAGGGAAGGAGCAATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGAC
TTCCAGAAGACCATGAAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTG
AAGGAGTTCCTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAG
GTGTTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGTGAGGCCCAG
GACCTGTACTGA (SEQ ID NO:9)

Fig. 9B

CS04Δ(760-1667) - CS04-SC1-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKK
TLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYD
DQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREG
SLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCH
RKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQ
HDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWV
HYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILG
PLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTV
EDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENRS
WYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLSV
FFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDY
YEDSYEDISAYLLSKNNAIEPREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQK
KTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLG
PYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTK
DEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTEN
MERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGH
VFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMA
SGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSS
LYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTL
RMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEW
LQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNS
LDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY (SEQ ID NO:10)

Fig. 10

CS04Δ(772-1667) - CS04-SC2-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATTCCAGACACCCCAGCACC
AGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATTTCT
GTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGGAGCTTC

(còn tiếp)

Fig. 11A

CAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATGAGCTCC
AGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTGGTCTTC
CAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCACCTGGGA
CTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAACCAGGCC
TCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAG
CCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCACATG
GCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAG
GATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCTGCCCAT
GGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAGAGCTGG
TACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGACCCCACC
TTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGGCTTGTC
ATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACATTCACTCC
ATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTGTACAAC
CTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGGGTGGAG
TGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGCAACAAGTGC
CAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTAT
GGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGCACCAAG
GAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCCAG
GGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGC
AAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAATGTGGAC
AGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTGCACCCC
ACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCCTGCAGC
ATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGCTACTTCACC
AACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAATGCCTGG
AGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTCACT
GGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTGATCAGC
TCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTTCCAGGGC
AACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGATACCTGAGG
ATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGTGAGGCCCAG
GACCTGTACTGA (SEQ ID NO:11)

Fig. 11B

CS04Δ(772-1667) - CS04-SC2-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFVEF
TDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDK
VFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFA
VFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEG
HTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDY
DDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRI
GRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLP
KGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQ
IMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWY
ILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSC
DKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNSRHPSTREITRTTLQSDQEEIDYDDTISVEMKKEDFDIY
DEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELN
EHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPT
KDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNC
RAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYK
MALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQ
WAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGN
STGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQIT
ASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISS
SQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY
(SEQ ID NO:12)

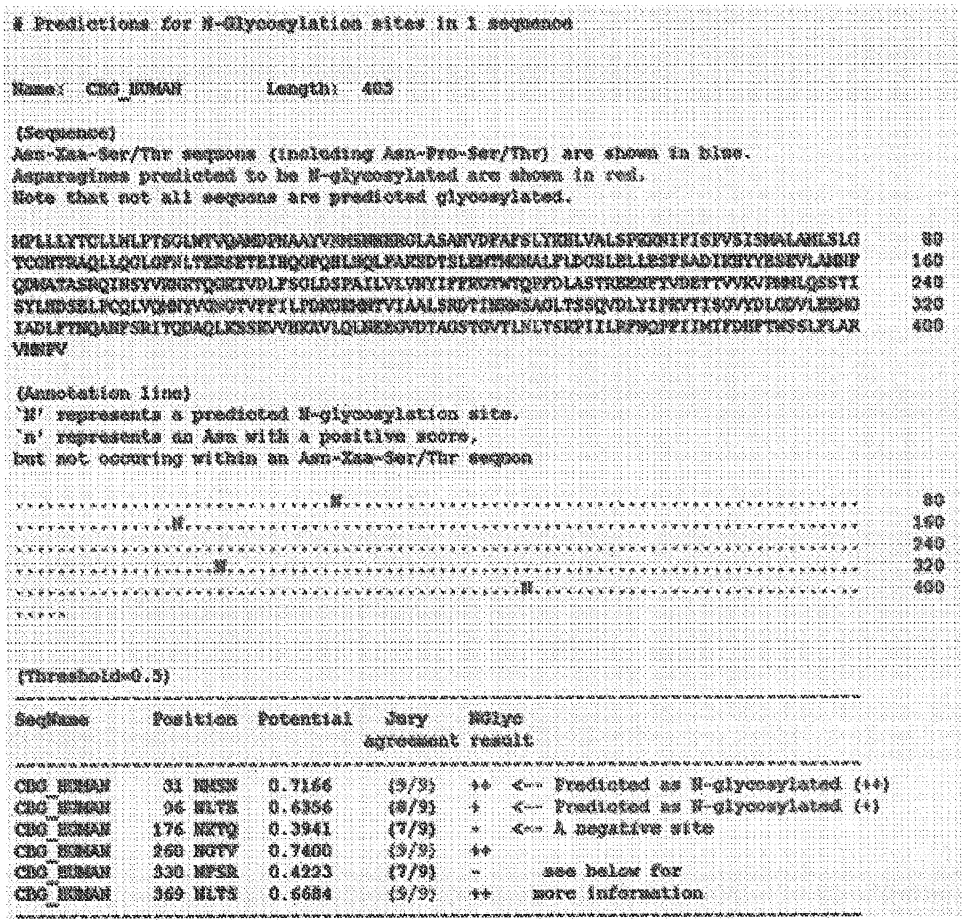
Fig. 12

```
NG1:   V  S  N  N  V  S  N  N  A  T  N  N  A  T  N  (SEQ ID NO:51)
       GTG AGC AAC AAT GTG AGC AAC AAT GCC ACC AAT AAT GCT ACC AAC (SEQ ID NO:50)

NG4:   V  S  N  N  A  T  N  N  V  S  N  (SEQ ID NO:53)
       GTG AGC AAC AAT GCC ACC AAC AAT GTG AGC AAC (SEQ ID NO:52)

NG5:   V  S  N  N  A  T  N  (SEQ ID NO:55)
       GTG AGC AAT AAT GCC ACC AAC (SEQ ID NO:54)

NG6:   V  S  N  N  (SEQ ID NO:57)
       GTG AGC AAT AAT (SEQ ID NO:56)

NG9:   R  S  L  (SEQ ID NO:59)
       AGG AGC CTG (SEQ ID NO:58)

NG10:  A  T  N  V  S  N  N  S  A  T  S  A  D  S  A  V  S  (SEQ ID NO:61)
       GCC ACT AAT GTG TCT AAC AAC TCT GCT ACC TCT GCT GAC TCT GCT GTG AGC (SEQ ID NO:60)

NG16:  A  T  N  Y  V  N  R  S  L  (SEQ ID NO:63)
       GCC ACC AAC TAT GTG AAC AGG AGC CTG (SEQ ID NO:62)
```

Fig. 13A

```
NG17:  A  T  N  Y  V  N  R  S  L  S  A  T  S  A  D  S  A  V  S  Q  N  (SEQ ID NO:65)
       GCC ACC AAC TAT GTG AAC AGG AGC CTG TCT GCC ACC TCT GCT GAC TCT GCT GTG AGC CAG AAT (SEQ ID NO:64)

NG18:  V  S  N  N  V  S  N  A  V  S  A  V  S  A  (SEQ ID NO:67)
       GTG AGC AAC AAT GTG AGC AAT GCT GTG TCT GCT GTG TCT GCT (SEQ ID NO:66)

NG19:  I  T  V  A  S  A  T  S  N  I  T  V  A  S  A  D  (SEQ ID NO:69)
       ATC ACT GTG GCC TCT GCC ACC TCT AAC ATC ACT GTG GCC TCT GCT GAC (SEQ ID NO:68)

NG20:  I  T  V  T  N  I  T  V  T  A  (SEQ ID NO:71)
       ATC ACT GTG ACC AAC ATC ACT GTG ACT GCC (SEQ ID NO:70)

NG21:  Q  T  V  T  N  I  T  V  T  A  (SEQ ID NO:73)
       CAG ACT GTG ACC AAC ATC ACT GTG ACT GCC (SEQ ID NO:72)

NGV:   A  T  N  V  S  N  N  S  N  T  S  N  D  S  N  V  S  (SEQ ID NO:75)
       GCC ACT AAT GTG TCT AAC AAC AGC AAC ACC AGC AAT GAC AGC AAT GTG TCT (SEQ ID NO:74)
```

Fig. 13B



```
# Predictions for N-Glycosylation sites in 1 sequence

Name:  CBG_HUMAN        Length:  403

(Sequence)
Asn-Xaa-Ser/Thr sequons (including Asn-Pro-Ser/Thr) are shown in blue.
Asparagines predicted to be N-glycosylated are shown in red.
Note that not all sequons are predicted glycosylated.

MFLLLYTCLLHLPTSCLNTVQAMDTNAAIVDFSNFMIQLASANVDPAPSLYRHLVALSPERNIFISPVSISMALAMLSLG      80
TCGHTRAQLLQGLGFNLTERSETEIHQGFQHLHQLFAKSDTSLEMTMGNALFLDGSLELLESFSADIKHYYESEVLAMNF     160
QDWATASRQINSYVRNKTQGRIVDLFSGLDSPAILVLVNYIFFKGTWTQPFDLASTREENFTVDETTVVKVPMMLQSSTI     240
SYLHDSELPCQLVQMNYVGNGTVFFILPDKGKMKTVIAALSRDTINRWSAGLTSSQVDLYIPKVTISGVYDLKDVLEDMG     320
IADLFTNQANFSRITQDAQLKSSKVVHKAVLQLSEDGVDTAGSTGVTLNLTSKPIILRFNQPFIIMIFDHFTWSSLFLAR     400
VMNPV

(Annotation line)
'N' represents a predicted N-glycosylation site.
'n' represents an Asn with a positive score,
but not occuring within an Asn-Xaa-Ser/Thr sequon

.............................N.............................................................      80
.....................N.....................................................................     160
.....................................................................................     240
.....................N.....................................................................     320
.....................................................N.....................................     400
.....

(Threshold=0.5)
-------------------------------------------------------------------------------
SeqName      Position  Potential   Jury    NGlyc
                                  agreement result
-------------------------------------------------------------------------------
CBG_HUMAN      31 NMSN   0.7166    (9/9)   ++  <-- Predicted as N-glycosylated (++)
CBG_HUMAN      96 NLTE   0.6356    (8/9)   +   <-- Predicted as N-glycosylated (+)
CBG_HUMAN     176 NKTQ   0.3941    (7/9)   -   <-- A negative site
CBG_HUMAN     260 NGTV   0.7400    (9/9)   ++
CBG_HUMAN     330 NFSR   0.4223    (7/9)   -      see below for
CBG_HUMAN     369 NLTS   0.6684    (9/9)   ++    more information
-------------------------------------------------------------------------------
```

Fig. 14

```
Name: Sequence          Length: 41
LSKNNAIEPRSFSQNATNVSNNSNTSNDSNVSPPVLKRHQR
. . . . . N..N. . . . . .
(Threshold=0.5)
-------------------------------------------------------------------
SeqName      Position  Potential    Jury     N-Glyc
                                  agreement result
-------------------------------------------------------------------
Sequence       15 NATN   0.6224     (8/9)    +
Sequence       18 NVSN   0.6453     (9/9)    ++
Sequence       21 NNSN   0.4158     (8/9)    -
Sequence       24 NTSN   0.4158     (8/9)    -
Sequence       27 NDSN   0.3619     (8/9)    -
Sequence       30 NVSP   0.1149     (9/9)    ---
-------------------------------------------------------------------
```

Fig. 15

CS01-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGAAGCTTCTCTCAGAATCCACCTGTCCTGAAGAGA
CACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGA
TCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATG
TCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG
GTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAGCAC

(còn tiếp)

Fig. 16A

CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
TCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTG
CACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAACTCC
TGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCTTAC
TTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTG
ATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTC
CAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGATAC
CTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGAG
GCACAAGACCTGTACTGA (SEQ ID NO:13)

Fig. 16B

CS08-FL-NA

ATGCAGATCGAACTGAGCACTTGCTTCTTCCTGTGTCTCCTGCGCTTTTGCTTCTCCGCCACAAGG
AGATACTATCTCGGTGCCGTGGAGCTCAGCTGGGACTACATGCAGAGCGACTTGGGTGAACTGCCT
GTGGACGCCAGGTTTCCACCCCGCGTGCCCAAGAGTTTCCCGTTCAACACCAGTGTCGTGTACAAG
AAAACCCTCTTCGTGGAATTCACCGACCACCTGTTCAACATCGCCAAACCGCGCCCTCCCTGGATG
GGGCTGCTCGGCCCGACGATCCAGGCTGAGGTCTATGACACGGTGGTGATTACCCTCAAGAACATG
GCTAGCCACCCGGTGAGCCTGCACGCCGTGGGCGTGTCCTATTGGAAAGCGTCCGAGGGTGCGGAG
TACGATGACCAGACTTCACAGCGGGAGAAGGAAGACGACAAAGTGTTCCCCGGGGGTTCCCACACC
TATGTCTGGCAGGTCCTGAAGGAGAATGGTCCTATGGCCTCCGACCCATTGTGCCTCACCTACTCT
TACCTAAGCCATGTGGATCTCGTCAAGGACCTGAACTCGGGGCTGATCGGCGCCCTGCTCGTGTGC
CGGGAGGGCTCACTGGCCAAGGAGAAGACCCAAACTCTGCACAAGTTCATCCTGCTGTTCGCGGTA
TTCGACGAGGGGAAGTCCTGGCACTCCGAGACCAAGAACAGCCTGATGCAGGACCGCGACGCAGCC
TCGGCCCGTGCGTGGCCAAAGATGCACACCGTGAACGGCTACGTTAACAGGAGCCTACCCGGCCTG
ATCGGCTGCCACCGCAAATCGGTCTACTGGCATGTGATCGGAATGGGCACAACGCCCGAGGTCCAC
AGTATCTTCCTCGAGGGCCACACTTTCCTGGTCCGGAATCACCGCCAGGCCAGCCTGGAGATCAGC
CCCATAACCTTTCTGACGGCGCAGACCTTACTCATGGATCTCGGCCAGTTCCTCCTGTTCTGCCAC
ATTTCGTCCCACCAGCACGATGGGATGGAAGCATATGTGAAAGTGGACTCCTGCCCCGAGGAACCC
CAGCTTAGGATGAAGAACAATGAGGAGGCCGAGGACTACGACGATGACCTTACCGATTCAGAAATG
GACGTAGTACGCTTTGACGACGACAACTCTCCATCCTTCATACAGATTCGCTCCGTCGCCAAGAAG
CACCCTAAGACTTGGGTGCACTACATCGCGGCCGAGGAGGAGGACTGGGATTATGCTCCCCTGGTG
CTGGCCCCCGACGACCGCAGCTACAAGAGCCAGTACCTGAATAACGGGCCCCAGCGCATCGGCCGG
AAGTACAAGAAAGTGCGGTTCATGGCTTACACGGACGAGACCTTCAAGACCCGGGAGGCTATCCAG
CATGAGAGCGGCATCTTGGGGCCCCTCCTGTACGGCGAAGTTGGAGACACACTGCTGATCATCTTC
AAGAACCAGGCGAGCAGGCCCTACAACATCTACCCCCACGGCATTACCGATGTCCGGCCGTTGTAC
AGCCGACGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTTCCGATCCTGCCGGGCGAGATCTTC
AAGTACAAGTGGACTGTGACCGTGGAGGATGGGCCGACCAAGAGCGATCCGCGCTGCCTGACCCGT
TACTACTCCAGCTTTGTCAATATGGAGCGCGACCTCGCTAGCGGCTTGATTGGCCCTCTGCTGATC
TGCTACAAGGAGTCCGTGGACCAGAGGGGGAATCAGATCATGAGTGACAAGAGGAACGTGATCCTG
TTCTCCGTGTTCGACGAAAACCGCAGCTGGTATCTCACCGAGAATATCCAGCGCTTCCTGCCCAAC
CCGGCCGGTGTGCAGCTGGAGGACCCCGAGTTTCAGGCCAGCAACATCATGCATTCTATCAACGGA
TATGTGTTTGATTCCCTGCAGCTCTCAGTGTGTCTGCACGAGGTCGCCTACTGGTATATCCTCAGC
ATTGGGGCACAGACCGACTTCCTGAGCGTGTTCTTCTCCGGGTATACCTTCAAGCACAAGATGGTG
TACGAGGATACCCTGACCCTGTTCCCCTTTAGCGGCGAAACCGTGTTTATGTCTATGGAGAACCCC
GGGCTCTGGATCCTTGGCTGCCATAACTCCGACTTCCGCAACCGCGGAATGACCGCGCTCCTGAAA
GTGTCGAGTTGTGACAAGAACACCGGCGACTATTACGAGGACAGTTACGAGGACATCTCTGCGTAC
CTCCTTAGCAAGAATAACGCCATCGAGCCAAGATCCTTCAGCCAGAACCCCCCAGTGCTGAAGAGG
CATCAGCGGGAGATCACCCGCACGACCCTGCAGTCGGATCAGGAGGAGATTGATTACGACGACACG
ATCAGTGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAAGATGAAAACCAGTCCCCTCGG
TCCTTCCAAAAGAAGACCCGGCACTACTTCATCGCCGCTGTGGAACGCCTGTGGGACTATGGAATG

(còn tiếp)

Fig. 17A

TCTTCTAGCCCTCACGTTTTGAGGAACCGCGCCCAGTCGGGCAGCGTGCCCCAGTTCAAGAAAGTG
GTGTTCCAGGAGTTCACCGACGGCTCCTTCACCCAGCCACTTTACCGGGGCGAGCTCAATGAACAT
CTGGGCCTGCTGGGACCCTACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACATTCCGGAAT
CAGGCCAGCAGACCATACAGTTTCTACAGTTCACTCATCTCCTACGAGGAGGACCAGCGCCAGGGG
GCTGAACCCCGTAAGAACTTCGTGAAGCCAAACGAAACAAAGACCTACTTCTGGAAGGTCCAGCAC
CACATGGCACCTACCAAGGACGAGTTCGATTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTG
GAGAAAGATGTGCACAGCGGCCTGATTGGCCCTCTGCTGGTGTGTCACACGAACACACTCAACCCT
GCACACGGGCGGCAGGTCACTGTGCAGGAATTCGCCCTGTTCTTTACCATCTTTGATGAGACGAAG
TCCTGGTATTTCACCGAAAACATGGAGAGGAACTGCCGCGCACCCTGCAACATCCAGATGGAAGAT
CCGACATTCAAGGAGAACTACCGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGC
CTCGTGATGGCCCAAGACCAGCGTATCCGCTGGTATCTGCTGTCGATGGGCTCCAACGAGAACATC
CATAGTATCCACTTCAGCGGGCATGTCTTCACGGTGAGGAAAAAGGAGGAGTACAAGATGGCACTG
TACAACCTCTATCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCTCCAAGGCCGGCATCTGGAGA
GTGGAATGCCTGATCGGCGAGCACCTCCACGCTGGGATGTCCACGCTGTTCCTCGTTTACAGCAAT
AAGTGCCAGACCCCTCTGGGCATGGCGAGCGGCCACATCCGCGACTTCCAGATTACAGCCAGCGGC
CAGTACGGTCAGTGGGCTCCAAAGCTGGCCCGTCTGCACTACTCCGGATCCATCAACGCCTGGTCC
ACCAAGGAACCGTTCTCCTGGATCAAAGTAGACCTGCTAGCCCCCATGATCATTCACGGCATCAAG
ACACAAGGCGCCCGACAGAAGTTCTCGAGCCTCTATATCTCCCAGTTCATCATCATGTATAGCCTG
GACGGAAAGAAGTGGCAGACTTACCGCGGAAACTCGACAGGGACCCTGATGGTATTCTTCGGTAAC
GTGGACAGCTCCGGAATCAAGCACAACATCTTCAACCCACCCATTATCGCCCGCTACATCCGCCTG
CACCCCACTCACTATAGCATTAGGTCCACCCTGCGAATGGAGCTCATGGGCTGTGACCTGAACAGC
TGTAGCATGCCCCTCGGCATGGAGTCTAAGGCGATCTCCGACGCACAGATAACGGCATCATCCTAC
TTTACCAACATGTTCGCTACCTGGTCCCCCTCCAAGGCCCGACTCCACCTGCAAGGGAGATCCAAC
GCCTGGCGGCCACAGGTCAACAATCCCAAGGAGTGGCTGCAAGTGGACTTTCAGAAAACTATGAAA
GTCACCGGAGTGACCACACAGGGAGTGAAGTCTCTGCTGACCAGCATGTACGTGAAGGAGTTCCTC
ATCTCCAGTTCGCAGGATGGCCACCAGTGGACGTTGTTCTTCCAAAACGGTAAAGTCAAAGTCTTC
CAAGGGAACCAGGACAGCTTTACACCCGTCGTGAACTCCCTGGACCCCCCGCTTCTCACTAGATAC
CTCCGCATCCACCCTCAGAGCTGGGTGCACCAGATTGCCCTGCGCATGGAGGTTCTGGGGTGTGAA
GCCCAGGACCTGTACTAA (SEQ ID NO:14)

Fig. 17B

CS10-FL-NA

ATGCAGATTGAGCTCTCCACCTGCTTCTTTCTCTGCCTTCTTCGCTTCTGCTTTTCTGCCACACGC
AGGTACTATTTGGGAGCAGTGGAACTGAGCTGGGATTACATGCAGAGTGACCTTGGTGAACTTCCT
GTGGACGCTCGTTTTCCACCTAGAGTTCCCAAGTCCTTCCCCTTCAACACCTCAGTGGTCTACAAG
AAAACGCTGTTTGTGGAGTTCACTGACCACCTCTTCAACATTGCCAAACCAAGACCCCCTTGGATG
GGATTGCTGGGACCCACAATACAAGCAGAAGTCTACGACACGGTGGTGATTACCCTGAAGAACATG
GCGTCACACCCTGTTTCACTTCACGCTGTTGGGGTCAGTTATTGGAAAGCCTCAGAGGGTGCGGAA
TACGATGATCAAACCAGCCAGAGGGAGAAGGAAGATGACAAGGTCTTTCCTGGGGGTAGCCATACC
TATGTTTGGCAGGTGCTGAAAGAGAATGGGCCTATGGCCTCTGATCCCTTGTGCCTCACATACTCT
TACCTGAGTCACGTCGACCTGGTGAAAGACCTGAATAGCGGTCTGATTGGTGCACTGCTTGTTTGT
AGAGAGGGGAGTTTGGCCAAGGAGAAAACTCAGACTCTCCACAAGTTTATCCTCCTGTTTGCTGTG
TTCGACGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAGGACAGAGATGCTGCA
TCTGCAAGGGCTTGGCCAAAAATGCACACAGTGAACGGCTATGTGAATCGATCACTGCCAGGACTG
ATAGGCTGTCATCGCAAGTCAGTGTATTGGCACGTTATCGGGATGGGAACAACTCCAGAAGTGCAC
AGCATCTTCCTTGAGGGCCACACTTTCCTGGTTCGGAATCATAGACAGGCCAGCCTTGAGATCAGC
CCAATCACCTTTCTGACTGCCCAAACCTTGCTGATGGATCTGGGACAGTTCCTCCTGTTTTGTCAC
ATCTCCTCCCACCAACATGACGGGATGGAGGCTTATGTGAAGGTCGATAGCTGTCCGGAGGAACCA
CAACTGAGGATGAAGAACAACGAAGAGGCAGAGGACTATGACGACGATCTGACTGACAGTGAAATG
GACGTGGTTCGGTTCGACGATGACAATTCTCCTTCATTTATCCAGATCCGTTCCGTGGCCAAGAAG
CACCCCAAGACTTGGGTTCATTACATCGCTGCTGAGGAGGAGGATTGGGACTACGCGCCCTTGGTG
TTGGCCCCAGACGATCGCTCATACAAGAGCCAGTACCTTAACAATGGTCCACAAAGGATCGGCCGG
AAGTACAAGAAGGTTAGATTTATGGCTTATACCGACGAGACTTTTAAAACTAGGGAAGCAATTCAG
CATGAAAGTGGCATTCTTGGACCCCTGCTGTATGGCGAGGTTGGCGACACCCTGCTGATTATCTTT
AAGAACCAGGCAAGCCGGCCCTACAACATCTACCCGCACGGCATAACCGATGTACGACCCCTGTAC
AGTCGCAGACTTCCTAAAGGGGTGAAACACCTGAAGGACTTCCCAATTCTGCCCGGGGAGATCTTC
AAGTATAAATGGACCGTGACGGTTGAGGATGGTCCCACAAAGTCCGATCCGAGATGCCTTACCCGA
TATTATTCCAGCTTCGTGAACATGGAAAGGGACCTGGCCAGCGGGCTGATTGGCCCACTGCTGATT
TGTTACAAGGAGTCTGTCGATCAAAGAGGAAACCAAATAATGAGCGACAAACGTAACGTCATCCTG
TTCAGCGTCTTTGATGAGAATAGAAGCTGGTACCTCACAGAAAATATTCAGCGGTTTCTGCCTAAC
CCCGCAGGCGTCCAGCTGGAAGATCCCGAGTTCCAAGCCTCAAACATCATGCATAGCATCAACGGA
TACGTATTCGATAGCCTGCAGCTGTCCGTCTGTCTCCATGAAGTGGCATATTGGTACATCCTGAGT
ATCGGGGCGCAGACCGACTTCCTGAGCGTGTTCTTTTCTGGATACACGTTCAAACACAAAATGGTC
TATGAAGATACCCTGACTCTGTTTCCATTCTCAGGAGAGACAGTCTTTATGAGTATGGAAAATCCT
GGACTGTGGATCCTGGGCTGTCACAATTCTGATTTTCGGAACAGAGGCATGACAGCCCTGCTTAAA
GTGAGCTCATGCGACAAGAACACCGGTGATTACTACGAAGATAGCTATGAGGACATCAGTGCGTAT
TTGCTCTCCAAGAACAACGCTATCGAGCCACGGTCTTTCAGTCAGAATCCTCCCGTTCTGAAGCGG
CATCAGCGCGAAATAACACGCACAACCCTTCAGTCAGACCAAGAGGAAATCGACTACGATGATACT
ATCTCTGTGGAGATGAAGAAGGAGGATTTCGACATTTACGACGAGGACGAGAATCAGTCCCCAAGG
AGCTTTCAGAAGAAAACAAGACACTATTTCATTGCCGCCGTGGAGCGACTGTGGGACTACGGCATG

(còn tiếp)

Fig. 18A

TCTAGCTCTCCGCATGTACTTAGAAATAGGGCACAAAGCGGATCCGTGCCTCAGTTTAAGAAAGTT
GTCTTTCAGGAGTTTACAGATGGCTCCTTCACCCAGCCCTTGTATCGCGGGGAACTCAATGAACAC
CTGGGCCTCCTGGGTCCTTATATTAGGGCCGAAGTCGAGGACAATATCATGGTGACCTTTAGGAAC
CAGGCATCTAGACCTTACTCTTTCTACTCCTCCCTGATATCCTATGAGGAGGACCAGCGGCAAGGC
GCTGAGCCTCGGAAGAACTTTGTGAAGCCAAATGAAACCAAAACATACTTTTGGAAAGTTCAGCAC
CACATGGCTCCCACGAAGGACGAATTTGACTGTAAAGCCTGGGCCTACTTCTCAGATGTAGATCTC
GAGAAAGACGTGCACTCAGGGCTCATTGGTCCCCTCCTGGTCTGTCATACTAATACCCTCAATCCA
GCACACGGACGTCAGGTAACCGTCCAGGAATTTGCCCTGTTCTTTACCATTTTCGATGAGACTAAA
TCCTGGTACTTTACCGAAAACATGGAGAGGAATTGCAGAGCCCCATGCAACATCCAGATGGAGGAC
CCTACCTTCAAAGAGAACTATCGCTTCCATGCCATTAACGGTTACATTATGGATACTCTCCCAGGA
CTTGTGATGGCACAGGATCAGCGGATAAGATGGTATCTGTTGAGCATGGGCTCCAACGAGAATATT
CACAGCATCCATTTCTCCGGTCACGTGTTTACAGTGAGAAAGAAAGAAGAGTACAAGATGGCTCTG
TATAATCTCTATCCAGGCGTATTCGAAACGGTGGAGATGTTGCCTAGCAAGGCCGGCATTTGGCGA
GTAGAATGCCTTATCGGGGAACATCTGCATGCCGGAATGAGCACGCTCTTCCTGGTGTATAGTAAC
AAGTGCCAGACTCCGCTGGGCATGGCATCTGGCCATATACGGGACTTTCAGATTACGGCTAGCGGG
CAGTATGGGCAGTGGGCACCCAAACTTGCGCGACTGCACTATTCAGGCTCTATCAATGCATGGTCC
ACCAAGGAACCCTTCTCTTGGATTAAGGTGGACCTTTTGGCGCCCATGATAATCCATGGGATCAAA
ACCCAGGGCGCTCGTCAGAAATTCTCATCACTCTACATCTCTCAGTTCATAATAATGTATTCACTG
GATGGGAAGAAATGGCAGACTTACAGAGGAAACAGCACCGGGACGCTGATGGTGTTCTTTGGCAAC
GTGGACAGCAGCGGCATCAAACACAACATCTTCAATCCTCCCATTATTGCCCGTTATATTAGACTG
CATCCCACTCACTACTCTATACGCAGCACACTTAGGATGGAGCTCATGGGATGCGACCTGAACAGT
TGTAGTATGCCCTTGGGGATGGAGTCCAAAGCTATAAGCGACGCACAAATTACAGCTAGCTCTTAC
TTTACGAATATGTTCGCCACGTGGAGCCCAAGCAAAGCCCGGCTGCATTTGCAGGGTCGGAGTAAT
GCTTGGCGCCCACAGGTGAATAACCCTAAGGAATGGTTGCAAGTAGATTTCCAGAAAACTATGAAG
GTAACCGGCGTCACTACACAGGGAGTCAAGTCCCTCTTGACCTCTATGTACGTCAAGGAGTTCCTG
ATTAGCAGCAGTCAGGATGGGCACCAATGGACACTGTTCTTCCAGAATGGGAAAGTTAAAGTATTT
CAGGGTAACCAGGACTCCTTTACACCTGTGGTGAATAGCCTCGACCCACCCCTGCTGACACGATAC
CTCCGCATCCACCCTCAGTCTTGGGTGCATCAAATTGCCCTGCGAATGGAGGTGTTGGGATGCGAA
GCTCAGGACCTCTACTGA (SEQ ID NO:15)

Fig. 18B

CS11-FL-NA

ATGCAGATCGAACTCTCTACTTGCTTCTTCCTGTGCCTTCTGAGGTTCTGCTTCTCTGCCACTCGC
CGATATTACCTCGGGGCCGTGGAGTTGAGTTGGGACTACATGCAATCAGATCTGGGCGAACTCCCT
GTGGATGCCCGATTCCCACCGCGCGTGCCCAAGTCTTTCCCATTTAATACTTCTGTGGTGTACAAG
AAGACATTGTTTGTGGAGTTTACCGATCACCTGTTCAACATCGCCAAACCGCGGCCCCCATGGATG
GGTCTGCTTGGGCCCACCATTCAAGCGGAGGTCTATGATACAGTGGTGATAACGCTTAAGAACATG
GCGAGCCACCCAGTGTCTCTGCATGCCGTTGGTGTATCATATTGGAAGGCCAGCGAAGGAGCGGAG
TACGATGACCAGACCTCTCAGAGAGAGAAGGAAGACGATAAGGTTTTTCCTGGCGGAAGTCATACA
TATGTATGGCAGGTCCTGAAAGAGAATGGGCCGATGGCTTCTGACCCCCTTTGTCTTACCTATAGT
TATCTGAGCCACGTGGACCTGGTCAAGGACCTCAACAGTGGTCTGATTGGGGCTCTGCTTGTTTGT
AGAGAGGGTAGCTTGGCTAAGGAGAAAACCCAAACACTCCATAAGTTCATTTTGCTGTTCGCGGTG
TTCGACGAGGGAAAGAGTTGGCACAGCGAAACAAAGAATTCACTGATGCAAGACAGGGACGCCGCT
TCCGCAAGGGCTTGGCCTAAGATGCATACGGTGAATGGGTATGTGAACCGGAGCCTCCCGGGGCTG
ATCGGGTGCCATCGCAAGTCTGTTTACTGGCACGTCATTGGAATGGGGACAACGCCAGAGGTACAT
AGTATATTTCTTGAAGGCCACACGTTCCTCGTACGGAACCACCGACAGGCTTCCCTGGAGATAAGC
CCCATTACCTTTCTGACCGCTCAGACTCTGCTGATGGACCTTGGCCAGTTTCTCCTGTTCTGCCAT
ATTAGCAGCCACCAGCACGACGGTATGGAAGCATACGTGAAAGTCGATAGCTGTCCTGAGGAGCCT
CAGCTCAGAATGAAGAACAACGAGGAGGCCGAAGACTATGACGATGACCTTACAGATTCCGAGATG
GACGTGGTGCGCTTTGACGACGATAACAGTCCTAGTTTCATTCAAATCAGATCCGTAGCCAAAAAG
CATCCAAAGACATGGGTGCATTACATTGCAGCCGAAGAGGAGGATTGGGATTATGCGCCCCTTGTT
CTGGCTCCAGATGACAGGAGCTATAAGTCCCAGTACTTGAACAACGGGCCACAGCGAATCGGTAGA
AAATATAAGAAGGTAAGATTCATGGCCTACACTGACGAAACATTTAAAACCAGGGAAGCTATCCAA
CACGAATCTGGAATTCTCGGCCCTCTGCTCTACGGTGAGGTGGGGGACACCTTGCTGATCATTTTC
AAAAATCAGGCATCCAGGCCTTACAACATATACCCCCATGGCATCACCGATGTCCGCCCGCTGTAT
TCCAGAAGACTCCCCAAGGGAGTGAAACATCTGAAAGATTTTCCCATCCTGCCGGGCGAGATCTTT
AAATACAAATGGACTGTGACTGTAGAGGACGGGCCTACAAAATCAGACCCACGGTGCCTGACAAGG
TATTACAGTAGCTTCGTCAACATGGAACGCGACCTCGCCAGCGGACTCATTGGCCCACTGTTGATC
TGTTACAAAGAGTCAGTGGATCAGAGGGGAAATCAGATCATGAGCGATAAGAGAAACGTTATCCTG
TTTAGTGTCTTCGACGAGAACCGGTCTTGGTACCTTACTGAGAACATCCAGAGGTTCCTGCCGAAT
CCGGCTGGCGTTCAGCTCGAGGACCCAGAGTTCCAGGCCAGTAATATAATGCACTCAATCAACGGT
TATGTGTTCGATAGCCTGCAGCTGAGCGTCTGCCTCCACGAGGTAGCCTATTGGTACATATTGTCC
ATCGGGGCTCAGACCGATTTTCTGTCCGTGTTCTTTAGCGGGTATACCTTTAAACATAAAATGGTC
TATGAAGACACCCTGACCCTGTTCCCATTCTCCGGTGAGACTGTGTTCATGTCCATGGAGAACCCA
GGGCTGTGGATCCTGGGGTGTCACAATAGTGACTTTAGGAATCGGGGAATGACGGCACTGCTGAAG
GTGAGTTCTTGCGATAAAAATACAGGAGATTACTATGAGGATAGTTACGAGGATATCAGTGCCTAT
CTGCTTTCAAAAAACAACGCAATTGAGCCCCGGTCTTTCTCACAAAACCCCCCGGTGCTGAAGCGC
CACCAGCGCGAAATTACCCGGACAACCTTGCAGTCCGACCAGGAGGAAATCGATTATGACGATACT
ATCAGTGTAGAAATGAAAAAGGAGGATTTTGATATTTACGACGAAGACGAGAACCAGTCTCCGCGA

(còn tiếp)

Fig. 19A

AGTTTTCAGAAGAAAACGCGACACTACTTTATAGCTGCCGTGGAACGACTCTGGGATTATGGCATG
TCCTCCAGCCCTCATGTCCTTAGGAATCGAGCGCAGAGTGGCTCTGTGCCTCAGTTCAAAAAGGTT
GTGTTCCAGGAATTCACCGACGGCTCATTTACCCAGCCGCTGTACAGAGGCGAACTCAACGAACAC
CTTGGGCTGCTTGGGCCATATATTCGAGCAGAGGTGGAAGATAATATCATGGTAACCTTTAGAAAC
CAGGCGTCAAGACCCTATTCCTTCTACAGTTCTCTGATCAGCTACGAGGAGGACCAAAGACAGGGA
GCTGAACCCAGGAAGAACTTTGTGAAACCTAATGAGACCAAGACCTACTTCTGGAAGGTCCAGCAC
CATATGGCCCCAACTAAAGATGAATTCGATTGCAAGGCCTGGGCTTATTTCAGCGACGTGGATCTC
GAAAAGGATGTGCACAGCGGGTTGATCGGACCGCTTTTGGTGTGCCACACAAATACCCTCAATCCT
GCCCACGGGCGGCAGGTCACAGTTCAAGAGTTTGCACTCTTCTTTACAATATTTGACGAGACAAAG
TCATGGTATTTTACAGAGAATATGGAGAGAAATTGTCGCGCACCTTGCAACATTCAGATGGAGGAC
CCCACATTTAAGGAGAATTACAGATTTCATGCTATCAATGGGTACATTATGGATACTCTGCCTGGT
CTGGTCATGGCCCAGGATCAGCGCATAAGGTGGTACTTGCTGAGCATGGGATCTAATGAGAATATA
CACAGCATTCACTTCAGTGGCCACGTTTTTACTGTTAGAAAGAAGGAGGAGTACAAAATGGCGCTC
TACAACCTTTACCCGGGTGTGTTTGAGACAGTGGAGATGCTGCCAAGCAAGGCAGGCATCTGGAGG
GTTGAGTGTCTTATTGGGGAGCATCTGCATGCTGGAATGTCCACCCTCTTTCTTGTGTACAGCAAT
AAGTGCCAGACACCGCTTGGCATGGCCAGCGGCCACATTAGGGACTTTCAGATAACTGCCAGTGGA
CAGTACGGCCAGTGGGCTCCCAAGCTTGCAAGACTCCACTACTCCGGAAGCATAAACGCATGGAGC
ACCAAGGAACCCTTCTCTTGGATTAAGGTGGACCTGCTGGCGCCAATGATCATTCACGGCATAAAA
ACCCAAGGGGCACGACAGAAATTTTCATCTTTGTATATTAGTCAGTTTATCATCATGTACAGCTTG
GATGGAAAGAAGTGGCAGACGTACAGGGGCAATTCTACAGGAACACTTATGGTGTTTTTTGGGAAT
GTCGATTCCAGCGGGATCAAACATAACATCTTCAATCCTCCTATTATCGCCCGATATATCCGCCTG
CACCCTACGCATTACTCCATCAGGTCCACATTGAGAATGGAACTGATGGGGTGCGACCTGAATAGT
TGTAGTATGCCACTGGGCATGGAGTCTAAAGCCATCAGCGATGCACAGATCACTGCCAGCTCTTAC
TTCACCAACATGTTTGCAACTTGGTCCCCCTCTAAAGCTCGCCTGCATCTGCAGGGACGCTCAAAT
GCATGGCGACCACAGGTGAACAATCCAAAAGAGTGGCTCCAGGTCGACTTTCAGAAGACAATGAAG
GTAACAGGAGTGACAACCCAGGGTGTAAAAAGCCTCCTTACGAGTATGTACGTTAAGGAGTTTCTG
ATTTCTAGCTCCCAGGACGGACACCAGTGGACTCTGTTCTTCCAGAACGGCAAAGTGAAGGTATTT
CAGGGAAACCAGGATTCTTTTACCCCGGTAGTGAATAGCCTGGATCCACCGTTGCTGACCCGCTAT
CTGAGAATTCATCCACAATCCTGGGTGCATCAGATTGCCCTCCGGATGGAAGTGCTCGGCTGTGAA
GCTCAGGATCTGTATTAG (SEQ ID NO:16)

Fig. 19B

CS40-FL-NA

ATGCAAATAGAGCTCTCCACCTGCTTCTTTCTGTGCCTTTTGCGATTCTGCTTTAGTGCCACCAGA
AGATACTACCTGGGTGCAGTGGAACTGTCATGGGACTATATGCAAAGTGATCTCGGTGAGCTGCCT
GTGGACGCAAGATTTCCTCCTAGAGTGCCAAAATCTTTTCCATTCAACACCTCAGTCGTGTACAAA
AAGACTCTGTTTGTAGAATTCACGGATCACCTTTTCAACATCGCTAAGCCAAGGCCACCCTGGATG
GGTCTGCTAGGTCCTACCATCCAGGCTGAGGTTTATGATACAGTGGTCATTACACTTAAGAACATG
GCTTCCCATCCTGTCAGTCTTCATGCTGTTGGTGTATCCTACTGGAAAGCTTCTGAGGGAGCTGAA
TATGATGATCAGACCAGTCAAAGGGAGAAAGAAGATGATAAAGTCTTCCCTGGTGGAAGCCATACA
TATGTCTGGCAGGTCCTGAAAGAGAATGGTCCAATGGCCTCTGACCCACTGTGCCTTACCTACTCA
TATCTTTCTCATGTGGACCTGGTAAAAGACTTGAATTCAGGCCTCATTGGAGCCCTACTAGTATGT
AGAGAAGGGAGTCTGGCCAAGGAAAAGACACAGACCTTGCACAAATTTATACTACTTTTTGCTGTA
TTTGATGAAGGGAAAAGTTGGCACTCAGAAACAAAGAACTCCTTGATGCAGGATAGGGATGCTGCA
TCTGCTCGGGCCTGGCCTAAAATGCACACAGTCAATGGTTATGTAAACAGGTCTCTGCCAGGTCTG
ATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCACTCCTGAAGTGCAC
TCAATATTCCTCGAAGGTCACACATTTCTTGTGAGGAACCATCGCCAGGCGTCCTTGGAAATCTCG
CCAATAACTTTCCTTACTGCTCAAACACTCTTGATGGACCTTGGACAGTTTCTACTGTTTTGTCAT
ATCTCTTCCCACCAACATGATGGCATGGAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCC
CAACTACGAATGAAAAATAATGAAGAAGCGGAAGACTATGATGATGATCTTACTGATTCTGAAATG
GATGTGGTCAGGTTTGATGATGACAACTCTCCTTCCTTTATCCAAATTCGCTCAGTTGCCAAGAAG
CATCCTAAAACTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTC
CTCGCCCCCGATGACAGAAGTTATAAAAGTCAATATTTGAACAATGGCCCTCAGCGGATTGGTAGG
AAGTACAAAAAAGTCCGATTTATGGCATACACAGATGAAACCTTTAAGACTCGTGAAGCTATTCAG
CATGAATCAGGAATCTTGGGACCTTTACTTTATGGGGAAGTTGGAGACACACTGTTGATTATATTT
AAGAATCAAGCAAGCAGACCATATAACATCTACCCTCACGGAATCACTGATGTCCGTCCTTTGTAT
TCAAGGAGATTACCAAAAGGTGTAAAACATTTGAAGGATTTTCCAATTCTGCCAGGAGAAATATTC
AAATATAAATGGACAGTGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGGTGCCTGACCCGC
TATTACTCTAGTTTCGTTAATATGGAGAGAGATCTAGCTTCAGGACTCATTGGCCCTCTCCTCATC
TGCTACAAAGAATCTGTAGATCAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCATCCTG
TTTTCTGTATTTGATGAGAACCGAAGCTGGTACCTCACAGAGAATATACAACGCTTTCTCCCCAAT
CCAGCTGGAGTGCAGCTTGAGGATCCAGAGTTCCAAGCCTCCAACATCATGCACAGCATCAATGGC
TATGTTTTTGATAGTTTGCAGTTGTCAGTTTGTTTGCATGAGGTGGCATACTGGTACATTCTAAGC
ATTGGAGCACAGACTGACTTCCTTTCTGTCTTCTTCTCTGGATATACCTTCAAACACAAAATGGTC
TATGAAGACACACTCACCCTATTCCCATTCTCAGGAGAAACTGTCTTCATGTCGATGGAAAACCCA
GGTCTATGGATTCTGGGGTGCCACAACTCAGACTTTCGGAACAGAGGCATGACCGCCTTACTGAAG
GTTTCTAGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGATATTTCAGCATAC
TTGCTGAGTAAAAACAATGCCATTGAACCAAGAAGCTTCTCCCAGAATCCACCAGTCTTGAAACGC
CATCAACGGGAAATAACTCGTACTACTCTTCAGTCAGATCAAGAGGAAATTGACTATGATGATACC
ATATCAGTTGAAATGAAGAAGGAAGATTTTGACATTTATGATGAGGATGAAAATCAGAGCCCCCGC
AGCTTTCAAAAGAAAACACGACACTATTTTATTGCTGCAGTGGAGAGGCTCTGGGATTATGGGATG

(còn tiếp)

Fig. 20A

AGTAGCTCCCCACATGTTCTAAGAAACAGGGCTCAGAGTGGCAGTGTCCCTCAGTTCAAGAAAGTT
GTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATACCGTGGAGAACTAAATGAACAT
TTGGGACTCCTGGGGCCATATATAAGAGCAGAAGTTGAAGATAATATCATGGTAACTTTCAGAAAT
CAGGCCTCTCGTCCCTATTCCTTCTATTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGA
GCAGAACCTAGAAAAAACTTTGTCAAGCCTAATGAAACCAAAACTTACTTTTGGAAAGTGCAACAT
CATATGGCACCCACTAAAGATGAGTTTGACTGCAAAGCCTGGGCTTATTTCTCTGATGTTGACCTG
GAAAAAGATGTGCACTCAGGCCTGATTGGACCCCTTCTGGTCTGCCACACTAACACACTGAACCCT
GCTCATGGGAGACAAGTGACAGTACAGGAATTTGCTCTGTTTTTCACCATCTTTGATGAGACCAAA
AGCTGGTACTTCACTGAAAATATGGAAAGAAACTGCAGGGCTCCCTGCAATATCCAGATGGAAGAT
CCCACTTTTAAAGAGAATTATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGC
TTAGTAATGGCTCAGGATCAAAGGATTCGATGGTATCTGCTCAGCATGGGCAGCAATGAAAACATC
CATTCTATTCATTTCAGTGGACATGTGTTCACTGTACGAAAAAAAAGAGGAGTATAAAATGGCACTG
TACAATCTCTATCCAGGTGTTTTTGAGACAGTGGAAATGTTACCATCCAAAGCTGGAATTTGGCGG
GTGGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTTCTGGTGTACAGCAAT
AAGTGTCAGACTCCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGA
CAATATGGACAGTGGGCCCCAAAGCTGGCCAGACTTCATTATTCCGGATCAATCAATGCCTGGAGC
ACCAAGGAGCCCTTTTCTTGGATCAAGGTGGATCTGTTGGCACCAATGATTATTCACGGCATCAAG
ACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTCTT
GATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCTTTGGCAAT
GTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTCGATACATCCGTTTG
CACCCAACTCATTATAGCATTCGCAGCACTCTTCGCATGGAGTTGATGGGCTGTGATTTAAATAGT
TGCAGCATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCACAGATTACTGCTTCATCCTAC
TTTACCAATATGTTTGCCACCTGGTCTCCTTCAAAAGCTCGACTTCACCTCCAAGGGAGGAGTAAT
GCCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACAGGAGTAACTACTCAGGGAGTAAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCTC
ATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTTCAGAATGGCAAAGTAAAGGTTTTT
CAGGGAAATCAAGACTCCTTCACACCTGTGGTGAACTCTCTAGACCCACCGTTACTGACTCGCTAC
CTTCGAATTCACCCCCAGAGTTGGGTGCACCAGATTGCCCTGAGGATGGAGGTTCTGGGCTGCGAG
GCACAGGACCTCTACTGA (SEQ ID NO:17)

Fig. 20B

CH25-FL-NA

ATGCAGATCGAGCTGTCCACATGCTTTTTTCTGTGCCTGCTGCGGTTCTGCTTCAGCGCCACCCGG
CGGTACTACCTGGGCGCCGTGGAGCTGTCCTGGGACTACATGCAGAGCGACCTGGGCGAGCTGCCC
GTGGACGCCCGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCAGCGTGGTGTACAAG
AAAACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCTCCGAGGGCGCCGAG
TACGACGACCAGACCAGCCAGCGGGAGAAAGAGGACGACAAAGTCTTTCCTGGCGGCAGCCACACC
TACGTGTGGCAGGTCCTGAAAGAAAACGGCCCCATGGCCTCCGACCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACAGCGGGCTGATTGGGGCCCTGCTGGTCTGC
CGGGAGGGCAGCCTGGCCAAAGAGAAAACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGACGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACCGGGACGCCGCC
TCTGCCAGAGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGAAGCCTGCCCGGCCTG
ATTGGCTGCCACCGGAAGAGCGTGTACTGGCACGTGATCGGCATGGGCACCACACCCGAGGTGCAC
AGCATCTTTCTGGAAGGGCACACCTTTCTGGTGCGGAACCACCGGCAGGCCAGCCTGGAAATCAGC
CCTATCACCTTCCTGACCGCCCAGACACTGCTGATGGACCTGGGCCAGTTCCTGCTGTTTTGCCAC
ATCAGCTCTCACCAGCACGACGGCATGGAAGCCTACGTGAAGGTGGACTCCTGCCCCGAGGAACCC
CAGCTGCGGATGAAGAACAACGAGGAAGCCGAGGACTACGACGACGACCTGACCGACAGCGAGATG
GACGTGGTGCGGTTCGACGACGACAACAGCCCCAGCTTCATCCAGATCAGAAGCGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAAGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGAAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGCGGATCGGCCGG
AAGTACAAGAAAGTGCGGTTCATGGCCTACACCGACGAGACCTTCAAGACCCGGGAGGCCATCCAG
CACGAGAGCGGCATCCTGGGCCCCCTGCTGTACGGCGAAGTGGGCGACACACTGCTGATCATCTTC
AAGAACCAGGCCAGCCGGCCCTACAACATCTACCCCCACGGCATCACCGACGTGCGGCCCCTGTAC
AGCAGGCGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGACGGCCCCACCAAGAGCGACCCCAGATGCCTGACCCGG
TACTACAGCAGCTTCGTGAACATGGAACGGGACCTGGCCTCCGGGCTGATCGGACCTCTGCTGATC
TGCTACAAAGAAAGCGTGGACCAGCGGGGCAACCAGATCATGAGCGACAAGCGGAACGTGATCCTG
TTCAGCGTGTTCGATGAGAACCGGTCCTGGTATCTGACCGAGAACATCCAGCGGTTTCTGCCCAAC
CCTGCCGGGGTGCAGCTGGAAGATCCCGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TACGTGTTCGACAGCCTGCAGCTGTCCGTGTGTCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGAGCGTGTTCTTCAGCGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCTTTCAGCGGCGAGACCGTGTTCATGAGCATGGAAAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCCGGAACCGGGGCATGACCGCCCTGCTGAAG
GTGTCCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGATATCAGCGCCTAC
CTGCTGTCCAAGAACAACGCCATCGAGCCCAGAAGCTTCAGCCAGAACCCCCCTGTGCTGAAGCGG
CACCAGAGAGAGATCACCCGGACCACCCTGCAGTCCGACCAGGAAGAGATCGATTACGACGACACC

(còn tiếp)

Fig. 21A

ATCAGCGTGGAGATGAAAAAAGAAGATTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCCGG
TCCTTCCAGAAGAAAACCCGGCACTACTTTATCGCCGCCGTGGAGCGGCTGTGGGACTACGGCATG
AGCAGCAGCCCCCACGTGCTGCGGAACCGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAAGTG
GTGTTCCAGGAATTCACCGACGGCAGCTTCACCCAGCCCCTGTACCGGGGCGAGCTGAACGAGCAC
CTGGGGCTGCTGGGGCCCTACATCAGGGCCGAAGTGGAGGACAACATCATGGTGACCTTCCGGAAT
CAGGCCAGCAGACCCTACTCCTTCTACAGCAGCCTGATCAGCTACGAAGAGGACCAGCGGCAGGGC
GCTGAACCCCGGAAGAACTTCGTGAAGCCCAATGAGACCAAGACCTACTTCTGGAAAGTGCAGCAC
CACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCAGCGACGTGGATCTG
GAAAAGGACGTGCACTCTGGACTGATTGGCCCTCTGCTGGTGTGCCACACCAACACCCTGAACCCC
GCCCACGGCCGGCAGGTGACCGTGCAGGAATTCGCCCTGTTCTTCACCATCTTCGACGAGACCAAG
TCCTGGTACTTCACCGAGAATATGGAACGGAACTGCAGAGCCCCCTGCAACATCCAGATGGAAGAT
CCTACCTTCAAAGAGAACTACCGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCTGGC
CTGGTGATGGCCCAGGACCAGAGGATCCGGTGGTATCTGCTGTCCATGGGCAGCAACGAGAATATC
CACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAAGAAGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGGCGG
GTGGAGTGTCTGATCGGCGAGCACCTGCATGCCGGGATGAGCACCCTGTTTCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCCGGGACTTCCAGATCACCGCCTCCGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCCGGCTGCACTACAGCGGCAGCATCAACGCCTGGTCC
ACCAAAGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCTATGATCATCCACGGCATTAAG
ACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GACGGCAAGAAGTGGCAGACCTACCGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGCAAC
GTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCCGGTACATCCGGCTG
CACCCCACCCACTACAGCATCAGATCCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAACTCC
TGCAGCATGCCTCTGGGCATGGAAAGCAAGGCCATCAGCGACGCCCAGATCACAGCCAGCAGCTAC
TTCACCAACATGTTCGCCACCTGGTCCCCCTCCAAGGCCAGGCTGCACCTGCAGGGCCGGTCCAAC
GCCTGGCGGCCTCAGGTGAACAACCCCAAAGAATGGCTGCAGGTGGACTTTCAGAAAACCATGAAG
GTGACCGGCGTGACCACCCAGGGCGTGAAAAGCCTGCTGACCAGCATGTACGTGAAAGAGTTTCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTTCAGAACGGCAAGGTGAAAGTGTTC
CAGGGCAACCAGGACTCCTTCACCCCCGTGGTGAACTCCCTGGACCCCCCCCTGCTGACCCGCTAC
CTGCGGATCCACCCCCAGTCTTGGGTGCACCAGATCGCCCTGAGGATGGAAGTGCTGGGATGTGAG
GCCCAGGATCTGTACTGA (SEQ ID NO:18)

Fig. 21B

FVIII-FL-AA

mqielstcff lcllrfcfsa trryylgave lswdymqsdl gelpvdarfp prvpksfpfn
tsvvykktlf veftdhlfni akprppwmgl lgptiqaevy dtvvitlknm ashpvslhav
gvsywkaseg aeyddqtsqr ekeddkvfpg gshtyvwqvl kengpmasdp lcltysylsh
vdlvkdlnsg ligallvcre gslakektqt lhkfillfav fdegkswhse tknslmqdrd
aasarawpkm htvngyvnrs lpgligchrk svywhvigmg ttpevhsifl eghtflvrnh
rqasleispi tfltaqtllm dlgqfllfch isshqhdgme ayvkvdscpe epqlrmknne
eaedydddlt dsemdvvrfd ddnspsfiqi rsvakkhpkt wvhyiaaeee dwdyaplvla
pddrsyksqy lnngpqrigr kykkvrfmay tdetfktrea iqhesgilgp llygevgdtl
liifknqasr pyniyphgit dvrplysrrl pkgvkhlkdf pilpgeifky kwtvtvedgp
tksdprcltr yyssfvnmer dlasgligpl licykesvdq rgnqimsdkr nvilfsvfde
nrswylteni qrflpnpagv qledpefqas nimhsingyv fdslqlsvcl hevaywyils
igaqtdflsv ffsgytfkhk mvyedtltlf pfsgetvfms menpglwilg chnsdfrnrg
mtallkvssc dkntgdyyed syedisayll sknnaieprs fsqnsrhpst rqkqfnatti
pendiektdp wfahrtpmpk iqnvsssdll mllrqsptph glslsdlqea kyetfsddps
pgaidsnnsl semthfrpql hhsgdmvftp esglqlrlne klgttaatel kkldfkvsst
snnlistips dnlaaqtdnt sslgppsmpv hydsqldttl fgkkssplte sggplslsee
nndskllesg lmnsqesswg knvsstesgr lfkgkrahgp alltkdnalf kvsislltkn
ktsnnsatnr kthidgpsll ienspsvwqn ilesdtefkk vtplihdrml mdknatalrl
nhmsnkttss knmemvqqkk egpippdaqn pdmsffkmlf lpesarwiqr thgknslnsg
qgpspkqlvs lgpeksvegq nflseknkvv vgkgeftkdv glkemvfpss rnlfltnldn
lhennthnqe kkiqeeiekk etliqenvvl pqihtvtgtk nfmknlflls trqnvegsyd
gayapvlqdf rslndstnrt kkhtahfskk geeenleglg nqtkqiveky acttrispnt
sqqnfvtqrs kralkqfrlp leetelekri ivddtstqws knmkhltpst ltqidyneke
kgaitqspls dcltrshsip qanrsplpia kvssfpsirp iyltrvlfqd nsshlpaasy
rkkdsgvqes shflqgakkn nlslailtle mtgdqrevgs lgtsatnsvt ykkventvlp
kpdlpktsgk vellpkvhiy qkdlfptets ngspghldlv egsllqgteg aikwneanrp
gkvpflrvat essaktpskl ldplawdnhy gtqipkeewk sqekspekta fkkkdtilsl
nacesnhaia ainegqnkpe ievtwakqgr terlcsqnpp vlkrhqreit rttlqsdqee
idyddtisve mkkedfdiyd edenqsprsf qkktrhyfia averlwdygm sssphvlrnr
aqsgsvpqfk kvvfqeftdg sftqplyrge lnehlgllgp yiraevedni mvtfrnqasr
pysfysslis yeedqrqgae prknfvkpne tktyfwkvqh hmaptkdefd ckawayfsdv
dlekdvhsgl igpllvchtn tlnpahgrqv tvqefalfft ifdetkswyf tenmerncra
pcniqmedpt fkenyrfhai ngyimdtlpg lvmaqdqrir wyllsmgsne nihsihfsgh
vftvrkkeey kmalynlypg vfetvemlps kagiwrvecl igehlhagms tlflvysnkc
qtplgmasgh irdfqitasg qygqwapkla rlhysgsina wstkepfswi kvdllapmii
hgiktqgarq kfsslyisqf iimysldgkk wqtyrgnstg tlmvffgnvd ssgikhnifn
ppiiaryirl hpthysirst lrmelmgcdl nscsmplgme skaisdaqit assyftnmfa
twspskarlh lqgrsnawrp qvnnpkewlq vdfqktmkvt gvttqgvksl ltsmyvkefl
isssqdghqw tlffqngkvk vfqgnqdsft pvvnsldppl ltrylrihpq swvhqialrm
evlgceaqdl y (SEQ ID NO:19)

Fig. 22

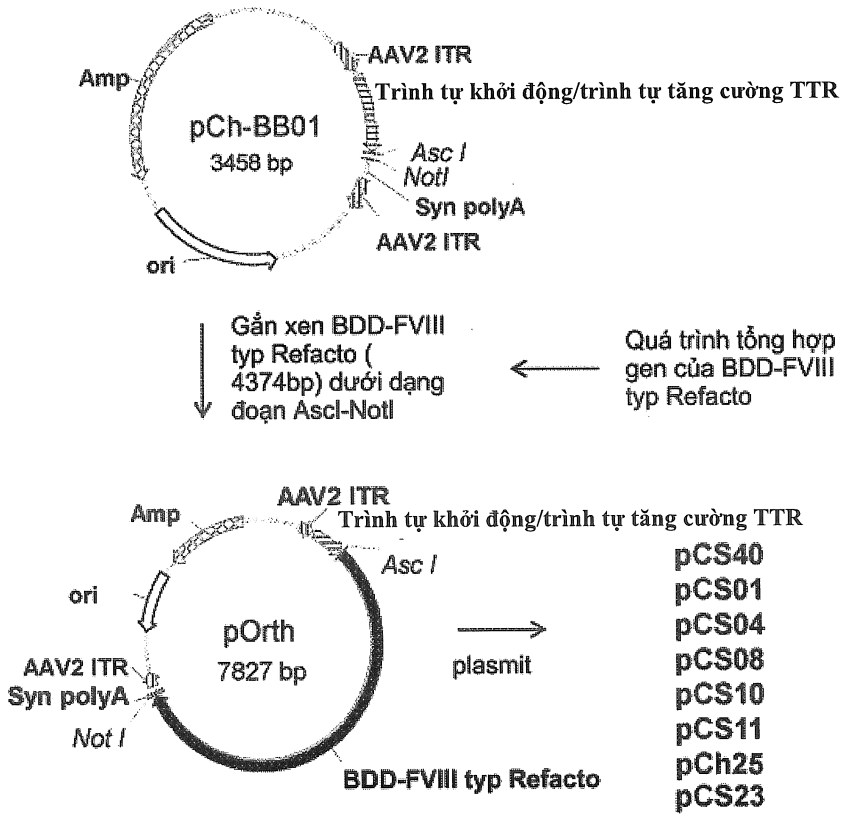

Amp
pCh-BB01
3458 bp
ori
AAV2 ITR
Trình tự khởi động/trình tự tăng cường TTR
Asc I
NotI
Syn polyA
AAV2 ITR
Gắn xen BDD-FVIII typ Refacto (4374bp) dưới dạng đoạn AscI-NotI
Quá trình tổng hợp gen của BDD-FVIII typ Refacto
Amp
AAV2 ITR
Trình tự khởi động/trình tự tăng cường TTR
Asc I
ori
pOrth
7827 bp
AAV2 ITR
Syn polyA
Not I
BDD-FVIII typ Refacto
plasmit
pCS40
pCS01
pCS04
pCS08
pCS10
pCS11
pCh25
pCS23

Fig. 23

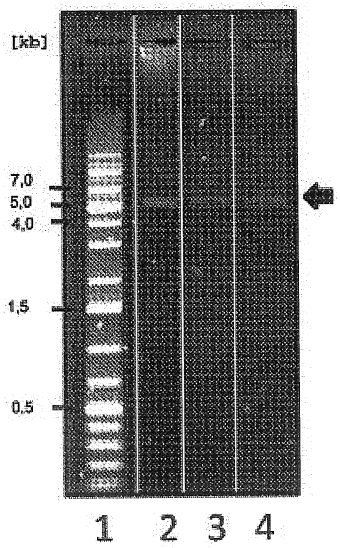


Fig. 24

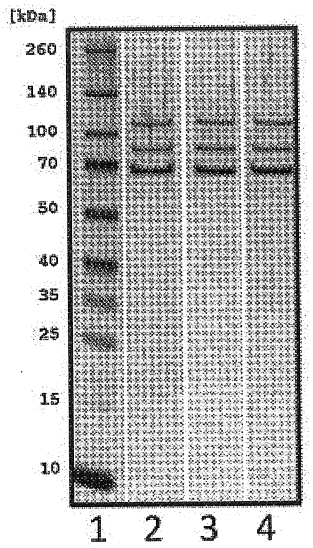


Fig. 25

CS23-FL-NA

atgcagattgagctgagcacctgcttcttcctgtgcctgctgaggttctgcttctctgccaccagg
agatactacctgggcgccgtggagctgagctgggactacatgcagtctgacctgggcgagctgcct
gtggacgccaggttcccccccagagtgcccaagagcttcccttcaacacctcagtggtgtacaag
aagaccctgttcgtggagttcaccgaccacctgttcaacatcgccaagcccaggccccctggatg
ggcctgctgggccccaccatccaggccgaggtgtacgacaccgtggtgatcaccctgaagaacatg
gccagccacccccgtgagcctgcacgccgtgggcgtgagctactggaaggcctctgagggcgccgag
tatgacgaccagaccagccagagggagaaggaggacgacaaggtgttccccggcggcagccacacc
tacgtgtggcaggtgctgaaggagaacggccccatggccagcgacccccctgtgcctgacctacagc
tacctgagccacgtggacctggtgaaggacctgaactctggcctgatcggcgccctgctggtgtgc
agggagggcagcctggccaaggagaagacccagaccctgcacaagttcatcctgctgttcgccgtg
ttcgatgagggcaagagctggcacagcgagaccaagaacagcctgatgcaggacagggatgccgcc
tctgccagggcctggcccaagatgcacaccgtgaacggctacgtgaacaggagcctgccccggcctg
atcggctgccacaggaagtctgtgtactggcacgtgatcggcatgggcaccacccccgaggtgcac
agcatcttcctggagggccacaccttcctggtgaggaaccacaggcaggccagcctggagatcagc
cccatcaccttcctgaccgcccagaccctgctgatggacctgggccagttcctgctgttctgccac
atcagcagccaccagcacgacggcatggaggcctacgtgaaggtggacagctgccccgaggagccc
cagctgaggatgaagaacaacgaggaggccgaggactatgatgatgacctgaccgactctgagatg
gacgtggtgaggtttgatgatgacaacagccccagcttcatccagatcaggtctgtggccaagaag
caccccaagacctgggtgcactacatcgccgccgaggaggaggactgggactacgcccccctggtg
ctggcccccgacgacaggagctacaagagccagtacctgaacaacggcccccagaggatcggcagg
aagtacaagaaggtcagattcatggcctacaccgacgagaccttcaagaccagggaggccatccag
cacgagtctggcatcctgggcccccctgctgtacggcgaggtgggcgacaccctgctgatcatcttc
aagaaccaggccagcaggccctacaacatctacccccacggcatcaccgatgtgaggcccctgtac
agcaggaggctgcccaagggcgtgaagcacctgaaggacttccccatcctgcccggcgagatcttc
aagtacaagtggaccgtgaccgtggaggatggccccaccaagtctgaccccaggtgcctgaccagg
tactacagcagcttcgtgaacatggagagggacctggcctctggcctgatcggcccccctgctgatc
tgctacaaggagagcgtggaccagaggggcaaccagatcatgtctgacaagaggaacgtgatcctg
ttctctgtgttcgatgagaacaggagctggtatctgaccgagaacatccagaggttcctgcccaac
cccgccggcgtgcagctggaggaccccgagttccaggccagcaacatcatgcacagcatcaacggc
tacgtgttcgacagcctgcagctgtctgtgtgcctgcacgaggtggcctactggtacatcctgagc
atcggcgcccagaccgacttcctgtctgtgttcttctctggctacaccttcaagcacaagatggtg
tacgaggacaccctgaccctgttcccctcagcggcgagaccgtgttcatgagcatggagaacccc
ggcctgtggatcctgggctgccacaacagcgacttcaggaacaggggcatgaccgccctgctgaaa
gtcagcagctgcgacaagaacaccggcgactactacgaggacagctacgaggacatcagcgcctac
ctgctgagcaagaacaacgccatcgagcccaggagcttcagccagaaccccccccgtgctgaagagg
caccagagggagatcaccaggaccaccctgcagagcgaccaggaggagatcgactatgatgacacc

(còn tiếp)

Fig. 26A

atcagcgtggagatgaagaaggaggacttcgacatctacgacgaggacgagaaccagagccccagg
agcttccagaagaagaccaggcactacttcatcgccgccgtggagaggctgtgggactatggcatg
agcagcagcccccacgtgctgaggaacagggcccagagcggcagcgtgccccagttcaagaaggtg
gtgttccaggagttcaccgacggcagcttcacccagcccctgtacagaggcgagctgaacgagcac
ctgggcctgctgggcccctacatcagggccgaggtggaggacaacatcatggtgaccttcaggaac
caggccagcaggccctacagcttctacagcagcctgatcagctacgaggaggaccagaggcagggc
gccgagcccaggaagaacttcgtgaagcccaacgagaccaagacctacttctggaaggtgcagcac
cacatggcccccaccaaggacgagttcgactgcaaggcctgggcctacttctctgatgtggacctg
gagaaggacgtgcacagcggcctgatcggcccctgctggtgtgccacaccaacaccctgaacccc
gcccacggcaggcaggtgaccgtgcaggagttcgccctgttcttcaccatcttcgacgagaccaag
agctggtacttcaccgagaacatggagaggaactgcagggcccctgcaacatccagatggaggac
cccaccttcaaggagaactacaggttccacgccatcaacggctacatcatggacaccctgcccggc
ctggtgatggcccaggaccagaggatcaggtggtatctgctgagcatgggcagcaacgagaacatc
cacagcatccacttcagcggccacgtgttcaccgtgaggaagaaggaggagtacaagatggccctg
tacaacctgtaccccggcgtgttcgagaccgtggagatgctgcccagcaaggccggcatctggagg
gtggagtgcctgatcggcgagcacctgcacgccggcatgagcaccctgttcctggtgtacagcaac
aagtgccagacccccctgggcatggccagcggccacatcagggacttccagatcaccgcctctggc
cagtacggccagtgggcccccaagctggccaggctgcactacagcggcagcatcaacgcctggagc
accaaggagcccttcagctggatcaaggtggacctgctggcccccatgatcatccacggcatcaag
acccagggcgccaggcagaagttcagcagcctgtacatcagccagttcatcatcatgtacagcctg
gacggcaagaagtggcagacctacaggggcaacagcaccggcaccctgatggtgttcttcggcaac
gtggacagcagcggcatcaagcacaacatcttcaacccccccatcatcgccaggtacatcaggctg
caccccacccactacagcatcaggagcaccctgcggatggaactgatgggctgcgacctgaacagc
tgcagcatgcccctgggcatggagagcaaggccatctctgacgcccagatcaccgccagcagctac
ttcaccaacatgttcgccacctggagccccagcaaggccaggctgcacctgcagggcaggagcaac
gcctggaggccccaggtgaacaaccccaaggagtggctgcaggtggacttccagaagaccatgaag
gtgaccggcgtgaccacccagggcgtgaagagcctgctgaccagcatgtacgtgaaggagttcctg
atcagcagcagccaggacggccaccagtggaccctgttcttccagaacggcaaagtgaaggtgttc
cagggcaaccaggacagcttcacCcccgtggtgaacagcctggacccccccctgctgaccaggtat
ctgaggatccacccccagagctgggtgcaccagatcgccctgagaatggaagtgctgggatgcgag
gcccaggacctgtactga (SEQ ID NO:20)

Fig. 26B

CS23-FL-AA

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDAR'FPPRVPKSFPFNTSVVYK
KTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEY
DDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCRE
GSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGC
HRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSH
QHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTW
VHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGIL
GPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVT
VEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENR
SWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLS
VFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGD
YYEDSYEDISAYLLSKNNAIEPRSFSQNPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFD
IYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQ
PLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNET
KTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALF
FTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLS
MGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTL
FLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMI
IHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIAR
YIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQG
RSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVK
VFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY (SEQ ID NO:21)

Fig. 27

CS23-HC-NA

gcc
accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg
ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac
acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc
gccaagccca ggccccccctg gatgggcctg ctgggccccca ccatccaggc cgaggtgtac
gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg
ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg
gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg
aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac
gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag
ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg
ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat
gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc
ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc
accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac
aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg
gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag
gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag
gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat
gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc
tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc
cccgacgaca ggagctacaa gagccagtac ctgaacaacg gccccccagag gatcggcagg
aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc
atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg
ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctacccccca cggcatcacc
gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc
cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc
accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg
gacctggcct ctggcctgat cggcccccctg ctgatctgct acaaggagag cgtggaccag
aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag
aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg
cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg
ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc
atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag
atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc
atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc
atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac
agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccagg
(SEQ ID NO:22)

Fig. 28

CS23-LC-NA

g agatcaccag gaccaccctg
cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag
gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc
aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc
cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc
caggagttca ccgacggcag cttcacccag cccctgtaca gaggcgagct gaacgagcac
ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc
aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac
cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac
ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg
gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cggcccctg
ctggtgtgcc acaccaacac cctgaacccc gcccacggca ggcaggtgac cgtgcaggag
ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaacatg
gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac
tacaggttcc acgccatcaa cggctacatc atggacaccc tgcccggcct ggtgatggcc
caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc
atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg
tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc
tggagggtgg agtgcctgat cggcgagcac ctgcacgccg gcatgagcac cctgttcctg
gtgtacagca acaagtgcca gacccccctg ggcatggcca gcggccacat cagggacttc
cagatcaccg cctctggcca gtacggccag tgggcccccа agctggccag gctgcactac
agcggcagca tcaacgcctg gagcaccaag gagcccttca gctggatcaa ggtggacctg
ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc
ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac
aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc
aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccacccac
tacagcatca ggagcaccct gcggatggaa ctgatgggct gcgacctgaa cagctgcagc
atgcccctgg gcatggagag caaggccatc tctgacgccc agatcaccgc cagcagctac
ttcaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg
agcaacgcct ggaggcccca ggtgaacaac cccaaggagt ggctgcaggt ggacttccag
aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg
tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctgttcttc
cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaac
agcctggacc ccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac
cagatcgccc tgagaatgga agtgctggga tgcgaggccc aggacctgta c
(SEQ ID NO:23)

Fig. 29

CS01m13-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTCCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCCTGAAG
AGACACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 30A

GTGGTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAG
CACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGG
TCAACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGG
CTGCACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAAC
TCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCT
TACTTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO:90)

Fig. 30B

CS01m23-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAATCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCACTGCATCTGACCCACCCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCCTGAAG
AGACACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 31A

GTGGTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAG
CACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGG
TCAACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGG
CTGCACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAAC
TCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCT
TACTTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO:91)

Fig. 31B

CS01m3-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGTCTCAGAATCCACCTGTCCTGAAG
AGACACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCA
AGATCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGC
ATGTCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 32A

GTGGTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAG
CACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGA
AACCAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAA
GGGGCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGAC
CTGGAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGTCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAA
GACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGG
TCAACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATC
AAGACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCT
CTGGATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGG
CTGCACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAAC
TCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCT
TACTTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGC
AATGCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATG
AAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTC
CTGATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTG
TTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGA
TACCTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGT
GAGGCACAAGACCTGTACTGA (SEQ ID NO:92)

Fig. 32B

CS01m2-FL-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAATCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCACTGCATCTGACCCACCCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGAAGCTTCTCTCAGAATCCACCTGTCCTGAAGAGA
CACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGA
TCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATG
TCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(còn tiếp)

Fig. 33A

GTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAGCAC
CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
TCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTG
CACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAACTCC
TGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCTTAC
TTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTG
ATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTC
CAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGATAC
CTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGAG
GCACAAGACCTGTACTGA (SEQ ID NO:93)

Fig. 33B

CS04m2-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGTCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGAAGAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCACTGCCTCTGACCCACCCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATCCACCTGTCCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(còn tiếp)

Fig. 34A

GTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC
CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGC
ACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCC
TGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGCTAC
TTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAAT
GCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTTC
CAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGTGAG
GCCCAGGACCTGTACTGA (SEQ ID NO:94)

Fig. 34B

CS04m3-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGAGCCAGAATCCACCTGTCCTGAAA
CGCCACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGC
ATGAGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 35A

GTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGC
AACCAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGAGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAG
GACCCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGG
CTGCACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGC
TACTTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGA
TACCTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGT
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:95)

Fig. 35B

CS04m23-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGTCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGAAGAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCACTGCCTCTGACCCACCCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTTCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGAGCCAGAATCCACCTGTCCTGAAA
CGCCACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGC
ATGAGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 36A

GTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGC
AACCAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGAGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAG
GACCCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGG
CTGCACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGC
TACTTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGA
TACCTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGT
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:96)

Fig. 36B

CS04m1-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTCCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATCCACCTGTCCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(còn tiếp)

Fig. 37A

GTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC
CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGC
ACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCC
TGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGCTAC
TTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAAT
GCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTTC
CAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGTGAG
GCCCAGGACCTGTACTGA (SEQ ID NO:97)

Fig. 37B

CS04m13-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGGCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGGGCAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCATGGCCTCTGACCCACTCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTCCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATACCACCTACGTGAACCGCTCCCTGAGCCAGAATCCACCTGTCCTGAAA
CGCCACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGAC
ACCATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCA
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGC
ATGAGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAA

(còn tiếp)

Fig. 38A

GTGGTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAG
CACCTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGC
AACCAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAG
GGGGCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAG
CACCACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAAC
CCTGCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACC
AAGAGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAG
GACCCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCT
GGGCTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAAC
ATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGG
AGGGTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGC
AACAAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCT
GGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGG
AGCACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATC
AAGACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGC
AATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGG
CTGCACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAAC
TCCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGC
TACTTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGC
AATGCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTC
CTGATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGA
TACCTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGT
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:98)

Fig. 38B

CS23m13-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTCCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACACCACCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCCGTGCTGAAG
AGGCACCAGAGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAG

(còn tiếp)

Fig. 39A

GTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAG
CACCTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAAC
CCCGCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACC
AAGAGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCC
GGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGG
AGGGTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGC
AACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCT
GGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCCACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGC
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:99)

Fig. 39B

CS23m3-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACACCACCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCCGTGCTGAAG
AGGCACCAGAGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAG

(còn tiếp)

Fig. 40A

GTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAG
CACCTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAAC
CCCGCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACC
AAGAGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCC
GGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGG
AGGGTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGC
AACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCT
GGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCCACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGC
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:100)

Fig. 40B

CS23m2-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGGTCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGTCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCAAGAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCACTGCCAGCGACCCCCCCTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACCCCCCCGTGCTGAAGAGG
CACCAGAGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACC
ATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGCATG
AGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAGGTG

(còn tiếp)

Fig. 41A

GTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAGCAC
CTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAAC
CAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGC
GCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCAC
CACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCC
GCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACCAAG
AGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAGGAC
CCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCCGGC
CTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAACATC
CACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGGAGG
GTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGGAGC
ACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAG
ACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGCAAC
GTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCCATCATCGCCAGGTACATCAGGCTG
CACCCCACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAACAGC
TGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGCTAC
TTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAAC
GCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTCCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTGTTC
CAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGCGAG
GCCCAGGACCTGTACTGA (SEQ ID NO:101)

Fig. 41B

CS23m1-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTCCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACCCCCCCCGTGCTGAAGAGG
CACCAGAGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACC
ATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGCATG
AGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAGGTG

(còn tiếp)

Fig. 42A

GTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAGCAC
CTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAAC
CAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGC
GCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCAC
CACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCC
GCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACCAAG
AGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAGGAC
CCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCCGGC
CTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAACATC
CACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGGAGG
GTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGCAAC
AAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGGAGC
ACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAG
ACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGCAAC
GTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGGCTG
CACCCCACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAACAGC
TGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGCTAC
TTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAAC
GCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTCCTG
ATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTGTTC
CAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGCGAG
GCCCAGGACCTGTACTGA (SEQ ID NO:102)

Fig. 42B

CS23m23-FL-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGGTCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGTCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCAAGAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCACTGCCAGCGACCCCCCCTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACACCACCTACGTGAACCGCTCCCTGAGCCAGAACCCCCCCGTGCTGAAG
AGGCACCAGAGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGAC
ACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCC
AGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGC
ATGAGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAG

(còn tiếp)

Fig. 43A

GTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAG
CACCTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGG
AACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAG
GGCGCCGAGCCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAG
CACCACATGGCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGAC
CTGGAGAAGGACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAAC
CCCGCCCACGGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACC
AAGAGCTGGTACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCCACCTTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCC
GGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAAC
ATCCACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCC
CTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGG
AGGGTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGC
AACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCT
GGCCAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATC
AAGACCCAGGGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGC
CTGGACGGCAAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGC
AACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGG
CTGCACCCCACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAAC
AGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGC
TACTTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGC
AACGCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATG
AAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTC
CTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTG
TTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGG
TATCTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGC
GAGGCCCAGGACCTGTACTGA (SEQ ID NO:103)

Fig. 43B

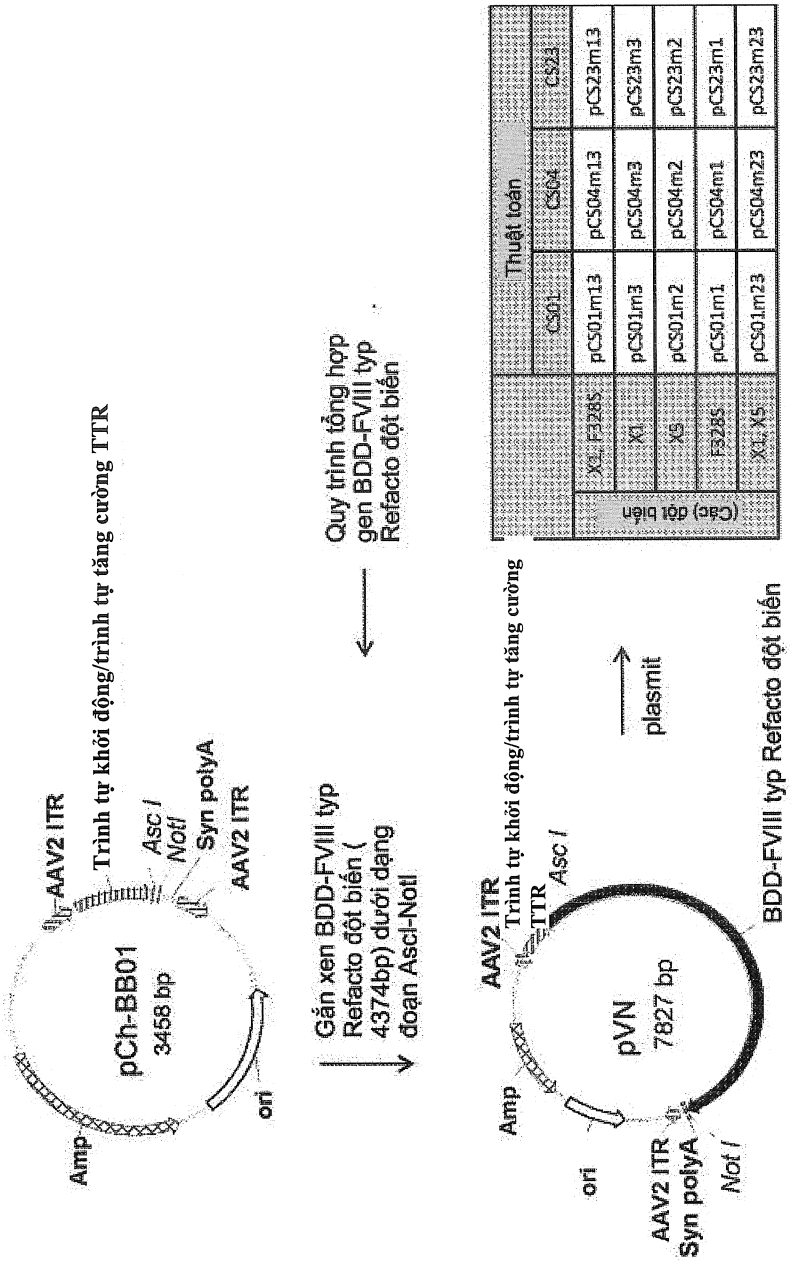

AAV2 ITR
Trình tự khởi động/trình tự tăng cường TTR
Asc I
Not I
Syn polyA
AAV2 ITR
pCh-BB01
3458 bp
Amp
ori
Gắn xen BDD-FVIII typ Refacto đột biến (4374bp) dưới dạng đoạn AscI-NotI
Quy trình tổng hợp gen BDD-FVIII typ Refacto đột biến
Thuật toán
CS01
CS04
CS23
(Các) đột biến
X1; F328S
X1
X9
F328S
X1; X5
pCS01m13
pCS01m3
pCS01m2
pCS01m1
pCS01m23
pCS04m13
pCS04m3
pCS04m2
pCS04m1
pCS04m23
pCS23m13
pCS23m3
pCS23m2
pCS23m1
pCS23m23
plasmit
AAV2 ITR
Trình tự khởi động/trình tự tăng cường
TTR
Asc I
pVN
7827 bp
Amp
ori
AAV2 ITR
Syn polyA
Not I
BDD-FVIII typ Refacto đột biến

Fig. 44

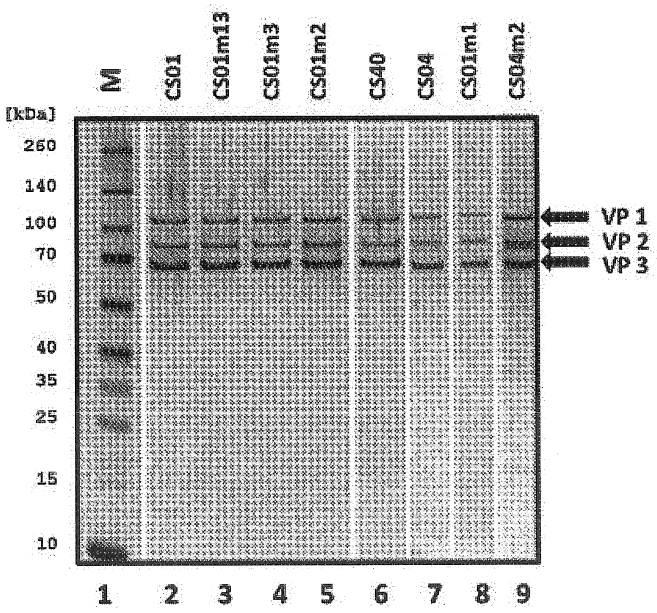

M
CS01
CS01m13
CS01m3
CS01m2
CS40
CS04
CS01m1
CS04m2
[kDa]
260
140
100
70
50
40
35
25
15
10
VP 1
VP 2
VP 3
1 2 3 4 5 6 7 8 9

Fig. 45

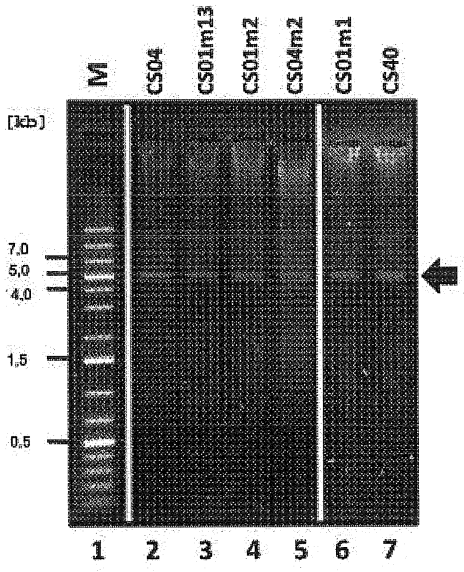

M
CS04
CS01m13
CS01m2
CS04m2
CS01m1
CS40
[kb]
7,0
5,0
4,0
1,5
0,5
1 2 3 4 5 6 7

Fig. 46

CS01-HC-NA

gcc
accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg
ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac
acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt
gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat
gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg
ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga
gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc
aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat
gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa
ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc
tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat
gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca
ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg
acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac
agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg
gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa
gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag
gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat
gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca
tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc
cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga
aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc
attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg
ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact
gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc
cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca
acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga
gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag
agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag
aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctgggtg
caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg
tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct
attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag
atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc
atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga
atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac
tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaga
(SEQ ID NO:24X)

Fig. 47

CS01-LC-NA

g agatcaccag gacaaccctc
cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag
gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca
agacactact tcattgctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc
catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggtcttc
caggagttca ctgatggctc attcacccag cccctgtaca gaggggaact gaatgagcac
ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc
agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac
cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac
ttctggaaag tccagcacca catggcaccc accaaggatg agtttgactg caaggcctgg
gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc
ctggtctgcc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag
tttgccctct tcttcaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg
gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac
tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca
caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc
atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg
tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc
tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttcctg
gtctacagca acaagtgcca gacacccctg ggaatggcct ctggccacat cagggacttc
cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac
tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg
ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct
ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac
agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc
aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca ccccacccac
tactcaatca gatcaaccct caggatggaa ctgatgggat gtgacctgaa ctcctgctca
atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atcctcttac
ttcaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccagggaaga
agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag
aagacaatga aagtcactgg ggtgacaacc caggggtca agtctctgct cacctcaatg
tatgtgaagg agttcctgat ctcttcctca caggatggcc accagtggac actcttcttc
cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac
tcactggacc ccccctcct gacaagatac ctgagaattc acccccagtc ttgggtccac
cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta c
(SEQ ID NO:25)

Fig. 48

CS01Δ(760-1667) - CS01-SC1-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGGAGATAC
TACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCTGTGGATGCCAGG
TTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAGAAGACACTCTTTGTGGAA
TTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATGGGACTCCTGGGACCCACCATTCAG
GCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATGGCATCCCACCCTGTGTCTCTGCATGCTGTG
GGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAGTATGATGACCAGACATCCCAGAGAGAGAAAAGAGGAT
GACAAGGTGTTCCCTGGGGGATCTCACACCTATGTGTGGCAAGTCCTCAAGGAGAATGGACCCATGGCATCT
GACCCACTCTGCCTGACATACTCCTACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATT
GGGGCACTGCTGGTGTGCAGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTC
CTGTTTGCTGTCTTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGAT
GCTGCCTCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCACTCCATT
TTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCTCCCATCACCTTC
CTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCACATCTCTTCCCACCAGCAT
GATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCACAGCTCAGGATGAAGAACAATGAG
GAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATGGATGTGGTCAGATTTGATGATGACAACTCT
CCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAACACCCCAAGACATGGGTGCACTACATTGCTGCTGAG
GAAGAGGACTGGGACTATGCACCACTGGTCCTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAAC
AATGGCCCACAAAGAATTGGAAGAAAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAG
ACAAGAGAAGCCATTCAGCATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTG
CTCATCATCTTCAAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCC
CTGTACAGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGATACTAC
TCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATCTGCTACAAGGAG
TCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTGTTCTCTGTCTTTGATGAG
AACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAACCCTGCTGGGGTGCAACTGGAAGAC
CCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGCTATGTGTTTGACTCTCTCCAGCTTTCTGTC
TGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCTATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTC
TCTGGATACACCTTCAAGCACAAGATGGTGTATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACT
GTGTTCATGAGCATGGAGAACCCTGGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGA
ATGACTGCACTGCTCAAAGTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGAC
ATCTCTGCCTACCTGCTCAGCAAGAACAATGCCATTGAGCCCAGAGAGATCACCAGGACAACCCTCCAGTCT
GACCAGGAAGAGATTGACTATGATGACACCATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGAT
GAGGACGAGAACCAGTCTCCAAGATCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGA
CTGTGGGACTATGGCATGTCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAG
TTCAAGAAAGTGGTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAAT
GAGCACCTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGGGCTGAG
CCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCACATGGCACCC

(còn tiếp)

Fig. 49A

ACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTGGAGAAAGATGTGCACTCT
GGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCTGCACATGGAAGGCAAGTGACTGTG
CAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAGTCATGGTACTTCACTGAGAACATGGAGAGA
AACTGCAGAGCACCATGCAACATTCAGATGGAAGACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATC
AATGGCTACATCATGGACACCCTGCCTGGGCTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTT
TCTATGGGATCCAATGAGAACATTCACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAG
GAATACAAGATGGCCCTGTACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCT
GGCATCTGGAGGGTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTAC
AGCAACAAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCAACCAAG
GAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAGACACAGGGGGCA
AGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTGGATGGCAAGAAGTGGCAG
ACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCAC
AACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTGCACCCCACCCACTACTCAATCAGATCAACC
CTCAGGATGGAACTGATGGGATGTGACCTGAACTCCTGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATT
TCTGATGCCCAGATCACTGCATCCTCTTACTTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGG
CTGCACCTCCAGGGAAGAAGCAATGCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGAC
TTCCAGAAGACAATGAAAGTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTG
AAGGAGTTCCTGATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAG
GTGTTCCAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGATAC
CTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGAGGCACAA
GACCTGTACTGA (SEQ ID NO:26)

Fig. 49B

CS01Δ(772-1667) - CS01-SC2-NA

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCATCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAAGCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGGGATCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCATGGCATCTGACCCACTCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTTCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGAAGCTTCTCTCAGAATTCCAGACACCCCAGCACC
AGGGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACCATTTCT
GTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGATCATTC

(còn tiếp)

Fig. 50A

CAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATGTCTTCC
TCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTGGTCTTC
CAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAGCACCTGGGA
CTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAACCAGGCC
TCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGGGCTGAG
CCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCACCACATG
GCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTGGAGAAA
GATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCTGCACAT
GGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAGTCATGG
TACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGACCCCACC
TTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGGCTTGTC
ATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATTCACTCC
ATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTGTACAAC
CTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGGGTGGAA
TGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGCAACAAGTGC
CAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTAT
GGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCAACCAAG
GAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAGACACAG
GGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTGGATGGC
AAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGCAATGTGGAC
AGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTGCACCCC
ACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAACTCCTGCTCA
ATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCTTACTTCACC
AACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAATGCCTGG
AGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAAGTCACT
GGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTGATCTCT
TCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTCCAGGGC
AACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGATACCTGAGA
ATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGAGGCACAA
GACCTGTACTGA (SEQ ID NO:27)

Fig. 50B

CS23Δ(760-1667) - CS23-SC1-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGATAC
TACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCTGTGGACGCCAGG
TTCCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAGAAGACCCTGTTCGTGGAG
TTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATGGGCCTGCTGGGCCCCACCATCCAG
GCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCCAGCCACCCCGTGAGCCTGCACGCCGTG
GGCGTGAGCTACTGGAAGGCCTCTGAGGGCGCCGAGTATGACGACCAGACCAGCCAGAGGGAGAAGGAGGAC
GACAAGGTGTTCCCCGGCGGCAGCCACACCTACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCCATGGCCAGC
GACCCCCTGTGCCTGACCTACAGCTACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATC
GGCGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTG
CTGTTCGCCGTGTTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGAT
GCCGCCTCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCACAGCATC
TTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGCCCCATCACCTTC
CTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCACATCAGCAGCCACCAGCAC
GACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCCCAGCTGAGGATGAAGAACAACGAG
GAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATGGACGTGGTGAGGTTTGATGATGACAACAGC
CCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCACCCCAAGACCTGGGTGCACTACATCGCCGCCGAG
GAGGAGGACTGGGACTACGCCCCCCTGGTGCTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAAC
AACGGCCCCCAGAGGATCGGCAGGAAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAG
ACCAGGGAGGCCATCCAGCACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTG
CTGATCATCTTCAAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCC
CTGTACAGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGGTACTAC
AGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATCTGCTACAAGGAG
AGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTGTTCTCTGTGTTCGATGAG
AACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAACCCCGCCGGCGTGCAGCTGGAGGAC
CCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGCTACGTGTTCGACAGCCTGCAGCTGTCTGTG
TGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGCATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTC
TCTGGCTACACCTTCAAGCACAAGATGGTGTACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACC
GTGTTCATGAGCATGGAGAACCCCGGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGC
ATGACCGCCCTGCTGAAAGTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGAC
ATCAGCGCCTACCTGCTGAGCAAGAACAACGCCATCGAGCCCAGGGAGATCACCAGGACCACCCTGCAGAGC
GACCAGGAGGAGATCGACTATGATGACACCATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGAC
GAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGG
CTGTGGGACTATGGCATGAGCAGCAGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAG
TTCAAGAAGGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAAC
GAGCACCTGGGCCTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAAC
CAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGCGCCGAG

(còn tiếp)

Fig. 51A

CCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCACCACATGGCCCCC
ACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGACGTGCACAGC
GGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCCGCCCACGGCAGGCAGGTGACCGTG
CAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACCAAGAGCTGGTACTTCACCGAGAACATGGAGAGG
AACTGCAGGGCCCCCTGCAACATCCAGATGGAGGACCCCACCTTCAAGGAGAACTACAGGTTCCACGCCATC
AACGGCTACATCATGGACACCCTGCCCGGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTG
AGCATGGGCAGCAACGAGAACATCCACAGCATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAG
GAGTACAAGATGGCCCTGTACAACCTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCC
GGCATCTGGAGGGTGGAGTGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTAC
AGCAACAAGTGCCAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGC
CAGTACGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGGAGCACCAAG
GAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAGACCCAGGGCGCC
AGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGACGGCAAGAAGTGGCAG
ACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGCAACGTGGACAGCAGCGGCATCAAGCAC
AACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGGCTGCACCCCACCCACTACAGCATCAGGAGCACC
CTGCGGATGGAACTGATGGGCTGCGACCTGAACAGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATC
TCTGACGCCCAGATCACCGCCAGCAGCTACTTCACCAACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGG
CTGCACCTGCAGGGCAGGAGCAACGCCTGGAGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGAC
TTCCAGAAGACCATGAAGGTGACCGGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTG
AAGGAGTTCCTGATCAGCAGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAG
GTGTTCCAGGGCAACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCTGCTGACCAGGTAT
CTGAGGATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGCGAGGCCCAG
GACCTGTACTGA (SEQ ID NO:28)

Fig. 51B

CS23Δ(772-1667) - CS23-SC2-NA

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGCGCCGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGCGAGCTGCCT
GTGGACGCCAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCAGTGGTGTACAAG
AAGACCCTGTTCGTGGAGTTCACCGACCACCTGTTCAACATCGCCAAGCCCAGGCCCCCCTGGATG
GGCCTGCTGGGCCCCACCATCCAGGCCGAGGTGTACGACACCGTGGTGATCACCCTGAAGAACATG
GCCAGCCACCCCGTGAGCCTGCACGCCGTGGGCGTGAGCTACTGGAAGGCCTCTGAGGGCGCCGAG
TATGACGACCAGACCAGCCAGAGGGAGAAGGAGGACGACAAGGTGTTCCCCGGCGGCAGCCACACC
TACGTGTGGCAGGTGCTGAAGGAGAACGGCCCCATGGCCAGCGACCCCCTGTGCCTGACCTACAGC
TACCTGAGCCACGTGGACCTGGTGAAGGACCTGAACTCTGGCCTGATCGGCGCCCTGCTGGTGTGC
AGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACCCTGCACAAGTTCATCCTGCTGTTCGCCGTG
TTCGATGAGGGCAAGAGCTGGCACAGCGAGACCAAGAACAGCCTGATGCAGGACAGGGATGCCGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACCGTGAACGGCTACGTGAACAGGAGCCTGCCCGGCCTG
ATCGGCTGCCACAGGAAGTCTGTGTACTGGCACGTGATCGGCATGGGCACCACCCCCGAGGTGCAC
AGCATCTTCCTGGAGGGCCACACCTTCCTGGTGAGGAACCACAGGCAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTGACCGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCTGTTCTGCCAC
ATCAGCAGCCACCAGCACGACGGCATGGAGGCCTACGTGAAGGTGGACAGCTGCCCCGAGGAGCCC
CAGCTGAGGATGAAGAACAACGAGGAGGCCGAGGACTATGATGATGACCTGACCGACTCTGAGATG
GACGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAG
CACCCCAAGACCTGGGTGCACTACATCGCCGCCGAGGAGGAGGACTGGGACTACGCCCCCCTGGTG
CTGGCCCCCGACGACAGGAGCTACAAGAGCCAGTACCTGAACAACGGCCCCCAGAGGATCGGCAGG
AAGTACAAGAAGGTCAGATTCATGGCCTACACCGACGAGACCTTCAAGACCAGGGAGGCCATCCAG
CACGAGTCTGGCATCCTGGGCCCCCTGCTGTACGGCGAGGTGGGCGACACCCTGCTGATCATCTTC
AAGAACCAGGCCAGCAGGCCCTACAACATCTACCCCCACGGCATCACCGATGTGAGGCCCCTGTAC
AGCAGGAGGCTGCCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCATCCTGCCCGGCGAGATCTTC
AAGTACAAGTGGACCGTGACCGTGGAGGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGG
TACTACAGCAGCTTCGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATCGGCCCCCTGCTGATC
TGCTACAAGGAGAGCGTGGACCAGAGGGGCAACCAGATCATGTCTGACAAGAGGAACGTGATCCTG
TTCTCTGTGTTCGATGAGAACAGGAGCTGGTATCTGACCGAGAACATCCAGAGGTTCCTGCCCAAC
CCCGCCGGCGTGCAGCTGGAGGACCCCGAGTTCCAGGCCAGCAACATCATGCACAGCATCAACGGC
TACGTGTTCGACAGCCTGCAGCTGTCTGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTGAGC
ATCGGCGCCCAGACCGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTTCAAGCACAAGATGGTG
TACGAGGACACCCTGACCCTGTTCCCCTTCAGCGGCGAGACCGTGTTCATGAGCATGGAGAACCCC
GGCCTGTGGATCCTGGGCTGCCACAACAGCGACTTCAGGAACAGGGGCATGACCGCCCTGCTGAAA
GTCAGCAGCTGCGACAAGAACACCGGCGACTACTACGAGGACAGCTACGAGGACATCAGCGCCTAC
CTGCTGAGCAAGAACAACGCCATCGAGCCCAGGAGCTTCAGCCAGAACTCCAGACACCCCAGCACC

(còn tiếp)

Fig. 52A

AGGGAGATCACCAGGACCACCCTGCAGAGCGACCAGGAGGAGATCGACTATGATGACACCATCAGC
GTGGAGATGAAGAAGGAGGACTTCGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTC
CAGAAGAAGACCAGGCACTACTTCATCGCCGCCGTGGAGAGGCTGTGGGACTATGGCATGAGCAGC
AGCCCCCACGTGCTGAGGAACAGGGCCCAGAGCGGCAGCGTGCCCCAGTTCAAGAAGGTGGTGTTC
CAGGAGTTCACCGACGGCAGCTTCACCCAGCCCCTGTACAGAGGCGAGCTGAACGAGCACCTGGGC
CTGCTGGGCCCCTACATCAGGGCCGAGGTGGAGGACAACATCATGGTGACCTTCAGGAACCAGGCC
AGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCTACGAGGAGGACCAGAGGCAGGGCGCCGAG
CCCAGGAAGAACTTCGTGAAGCCCAACGAGACCAAGACCTACTTCTGGAAGGTGCAGCACCACATG
GCCCCCACCAAGGACGAGTTCGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAG
GACGTGCACAGCGGCCTGATCGGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCCGCCCAC
GGCAGGCAGGTGACCGTGCAGGAGTTCGCCCTGTTCTTCACCATCTTCGACGAGACCAAGAGCTGG
TACTTCACCGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAGGACCCCACC
TTCAAGGAGAACTACAGGTTCCACGCCATCAACGGCTACATCATGGACACCCTGCCCGGCCTGGTG
ATGGCCCAGGACCAGAGGATCAGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAACATCCACAGC
ATCCACTTCAGCGGCCACGTGTTCACCGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAAC
CTGTACCCCGGCGTGTTCGAGACCGTGGAGATGCTGCCCAGCAAGGCCGGCATCTGGAGGGTGGAG
TGCCTGATCGGCGAGCACCTGCACGCCGGCATGAGCACCCTGTTCCTGGTGTACAGCAACAAGTGC
CAGACCCCCCTGGGCATGGCCAGCGGCCACATCAGGGACTTCCAGATCACCGCCTCTGGCCAGTAC
GGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACAGCGGCAGCATCAACGCCTGGAGCACCAAG
GAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCACGGCATCAAGACCCAG
GGCGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGACGGC
AAGAAGTGGCAGACCTACAGGGGCAACAGCACCGGCACCCTGATGGTGTTCTTCGGCAACGTGGAC
AGCAGCGGCATCAAGCACAACATCTTCAACCCCCCCATCATCGCCAGGTACATCAGGCTGCACCCC
ACCCACTACAGCATCAGGAGCACCCTGCGGATGGAACTGATGGGCTGCGACCTGAACAGCTGCAGC
ATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGACGCCCAGATCACCGCCAGCAGCTACTTCACC
AACATGTTCGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAACGCCTGG
AGGCCCCAGGTGAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACC
GGCGTGACCACCCAGGGCGTGAAGAGCCTGCTGACCAGCATGTACGTGAAGGAGTTCCTGATCAGC
AGCAGCCAGGACGGCCACCAGTGGACCCTGTTCTTCCAGAACGGCAAAGTGAAGGTGTTCCAGGGC
AACCAGGACAGCTTCACCCCCGTGGTGAACAGCCTGGACCCCCCCCCTGCTGACCAGGTATCTGAGG
ATCCACCCCCAGAGCTGGGTGCACCAGATCGCCCTGAGAATGGAAGTGCTGGGATGCGAGGCCCAG
GACCTGTACTGA (SEQ ID NO:29)

Fig. 52B

CS01m23-FL-AA (SEQ ID NO: 104)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFN
TSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVVTLKNMASHPVSLHAV
GVSYWKSSEGAEYDDQTSQREKEDDKVFPGKSHTYVWQVLKENGPTASDPPCLTYSYLSH
VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD
AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNH
RQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNE
EAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTL
LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP
TKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDE
NRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRG
MTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNTTYVNRSLSQNPPVLKRHQREITRTT
LQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSS
PHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVT
FRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKA
WAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTEN
MERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIH
SIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLF
LVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVD
LLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSG
IKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASS
YFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTS
MYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWV
HQIALRMEVLGCEAQDLY

Fig. 53

CS04m3-FL-AA (SEQ ID NO: 105)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFN
TSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAV
GVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSH
VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD
AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNH
RQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNE
EAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTL
LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP
TKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDE
NRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRG
MTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNTTYVNRSLSQNPPVLKRHQREITRTT
LQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSS
PHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVT
FRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKA
WAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTEN
MERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIH
SIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLF
LVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVD
LLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSG
IKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASS
YFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTS
MYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWV
HQIALRMEVLGCEAQDLY

Fig. 54

CS01-FL-AAm12 (SEQ ID NO: 106)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYK
KTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVVTLKNMASHPVSLHAVGVSYWKSSEGAE
YDDQTSQREKEDDKVFPGKSHTYVWQVLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVC
REGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGL
IGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLSCH
ISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKK
HPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQ
HESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIF
KYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVIL
FSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLK
VSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNPPVLKRHQREITRTTLQSDQEEIDYDDT
ISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKV
VFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQG
AEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNP
AHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPG
LVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWR
VECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWS
TKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGN
VDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSY
FTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFL
ISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCE
AQDLY

Fig. 55

CS04-FL-AAm12 (SEQ ID NO: 107)

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYK
KTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVVTLKNMASHPVSLHAVGVSYWKSSEGAE
YDDQTSQREKEDDKVFPGKSHTYVWQVLKENGPTASDPPCLTYSYLSHVDLVKDLNSGLIGALLVC
REGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGL
IGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLSCH
ISSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKK
HPKTWVHYIAAEEEDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQ
HESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIF
KYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVIL
FSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGMTALLK
VSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNPPVLKRHQREITRTTLQSDQEEIDYDDT
ISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKV
VFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQG
AEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNP
AHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPG
LVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWR
VECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWS
TKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGN
VDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSY
FTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFL
ISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCE
AQDLY

Fig. 56

CS01-FL-NAm12 (SEQ ID NO: 108)

ATGCAGATTGAGCTGTCCACCTGCTTCTTTCTGTGCCTGCTGAGATTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGGCTGTGGAACTTTCTTGGGACTACATGCAGTCTGACCTGGGAGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAGTCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACACTCTTTGTGGAATTCACTGACCACCTGTTCAACATTGCAAAACCCAGACCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCATCCCACCCTGTGTCTCTGCATGCTGTGGGAGTCTCATACTGGAAATCCTCTGAAGGGGCTGAG
TATGATGACCAGACATCCCAGAGAGAGAAAGAGGATGACAAGGTGTTCCCTGGGAAGTCTCACACC
TATGTGTGGCAAGTCCTCAAGGAGAATGGACCCACTGCATCTGACCCACCCTGCCTGACATACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCACTGCTGGTGTGC
AGGGAAGGATCCCTGGCCAAGGAGAAAACCCAGACACTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGTCTTGGCACTCTGAAACAAAGAACTCCCTGATGCAAGACAGGGATGCTGCC
TCTGCCAGGGCATGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGATCACTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAAGTGCAC
TCCATTTTCCTGGAGGGACACACCTTCCTGGTCAGGAACCACAGACAAGCCTCTCTGGAGATCTCT
CCCATCACCTTCCTCACTGCACAGACACTGCTGATGGACCTTGGACAGTTCCTGCTGTCCTGCCAC
ATCTCTTCCCACCAGCATGATGGCATGGAAGCCTATGTCAAGGTGGACTCATGCCCTGAGGAACCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCAGATTTGATGATGACAACTCTCCATCCTTCATTCAGATCAGGTCTGTGGCAAAGAAA
CACCCCAAGACATGGGTGCACTACATTGCTGCTGAGGAAGAGGACTGGGACTATGCACCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGTCTCAGTACCTCAACAATGGCCCACAAAGAATTGGAAGA
AAGTACAAGAAAGTCAGATTCATGGCCTACACTGATGAAACCTTCAAGACAAGAGAAGCCATTCAG
CATGAGTCTGGCATTCTGGGACCACTCCTGTATGGGGAAGTGGGAGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCAGGAGACTGCCAAAAGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGAGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACAAAGTCTGACCCCAGGTGCCTCACCAGA
TACTACTCCTCTTTTGTGAACATGGAGAGAGACCTGGCATCTGGACTGATTGGACCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGAGGCAACCAGATCATGTCTGACAAGAGAAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGATCATGGTACCTGACTGAGAACATTCAGAGATTCCTGCCCAAC
CCTGCTGGGGTGCAACTGGAAGACCCTGAGTTCCAGGCAAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACTCTCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCACAAACTGACTTCCTTTCTGTCTTCTTCTCTGGATACACCTTCAAGCACAAGATGGTG
TATGAGGACACCCTGACACTCTTCCCATTCTCTGGGGAAACTGTGTTCATGAGCATGGAGAACCCT
GGACTGTGGATTCTGGGATGCCACAACTCTGACTTCAGAAACAGGGGAATGACTGCACTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACTCTTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGAAGCTTCTCTCAGAATCCACCTGTCCTGAAGAGA
CACCAGAGAGAGATCACCAGGACAACCCTCCAGTCTGACCAGGAAGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTATGATGAGGACGAGAACCAGTCTCCAAGA
TCATTCCAGAAGAAGACAAGACACTACTTCATTGCTGCTGTGGAAAGACTGTGGGACTATGGCATG
TCTTCCTCTCCCCATGTCCTCAGGAACAGGGCACAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG

(còn tiếp)

Fig. 57A

GTCTTCCAGGAGTTCACTGATGGCTCATTCACCCAGCCCCTGTACAGAGGGGAACTGAATGAGCAC
CTGGGACTCCTGGGACCATACATCAGGGCTGAGGTGGAAGACAACATCATGGTGACATTCAGAAAC
CAGGCCTCCAGGCCCTACAGCTTCTACTCTTCCCTCATCAGCTATGAGGAAGACCAGAGACAAGGG
GCTGAGCCAAGAAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCACCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCATACTTCTCTGATGTGGACCTG
GAGAAAGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCACATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
TCATGGTACTTCACTGAGAACATGGAGAGAAACTGCAGAGCACCATGCAACATTCAGATGGAAGAC
CCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCACAGGACCAGAGAATCAGATGGTACCTGCTTTCTATGGGATCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGAGAAAGAAGGAGGAATACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAATGCCTCATTGGGGAGCACCTGCATGCTGGCATGTCAACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACACCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCACCCAAACTGGCCAGGCTCCACTACTCTGGCTCCATCAATGCATGGTCA
ACCAAGGAGCCATTCTCTTGGATCAAGGTGGACCTGCTGGCACCCATGATCATTCATGGCATCAAG
ACACAGGGGGCAAGACAGAAATTCTCCTCTCTGTACATCTCACAGTTCATCATCATGTACTCTCTG
GATGGCAAGAAGTGGCAGACATACAGAGGCAACTCCACTGGCACCCTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCTCCCATCATTGCCAGATACATCAGGCTG
CACCCCACCCACTACTCAATCAGATCAACCCTCAGGATGGAACTGATGGGATGTGACCTGAACTCC
TGCTCAATGCCCCTGGGAATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCATCCTCTTAC
TTCACCAACATGTTTGCCACCTGGTCACCATCAAAAGCCAGGCTGCACCTCCAGGGAAGAAGCAAT
GCCTGGAGACCCCAGGTCAACAACCCAAAGGAATGGCTGCAAGTGGACTTCCAGAAGACAATGAAA
GTCACTGGGGTGACAACCCAGGGGGTCAAGTCTCTGCTCACCTCAATGTATGTGAAGGAGTTCCTG
ATCTCTTCCTCACAGGATGGCCACCAGTGGACACTCTTCTTCCAGAATGGCAAAGTCAAGGTGTTC
CAGGGCAACCAGGACTCTTTCACACCTGTGGTGAACTCACTGGACCCCCCCCTCCTGACAAGATAC
CTGAGAATTCACCCCCAGTCTTGGGTCCACCAGATTGCCCTGAGAATGGAAGTCCTGGGATGTGAG
GCACAAGACCTGTACTGA

Fig. 57B

CS04-FL-NAm12 (SEQ ID NO: 109)

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGG
AGATACTACCTGGGGGCTGTGGAGCTTTCTTGGGACTACATGCAGTCTGACCTGGGGGAGCTGCCT
GTGGATGCCAGGTTCCCACCCAGAGTGCCCAAATCCTTCCCATTCAACACCTCTGTGGTCTACAAG
AAGACCCTCTTTGTGGAGTTCACTGACCACCTGTTCAACATTGCCAAACCCAGGCCACCCTGGATG
GGACTCCTGGGACCCACCATTCAGGCTGAGGTGTATGACACTGTGGTCGTCACCCTCAAGAACATG
GCCTCCCACCCTGTGAGCCTGCATGCTGTGGGGGTCAGCTACTGGAAGTCCTCTGAGGGGGCTGAG
TATGATGACCAGACCTCCCAGAGGGAGAAGGAGGATGACAAAGTGTTCCCTGGGAAGAGCCACACC
TATGTGTGGCAGGTCCTCAAGGAGAATGGCCCCACTGCCTCTGACCCACCCTGCCTGACCTACTCC
TACCTTTCTCATGTGGACCTGGTCAAGGACCTCAACTCTGGACTGATTGGGGCCCTGCTGGTGTGC
AGGGAGGGCTCCCTGGCCAAAGAGAAGACCCAGACCCTGCACAAGTTCATTCTCCTGTTTGCTGTC
TTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACTCCCTGATGCAGGACAGGGATGCTGCC
TCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGTGAACAGGAGCCTGCCTGGACTC
ATTGGCTGCCACAGGAAATCTGTCTACTGGCATGTGATTGGCATGGGGACAACCCCTGAGGTGCAC
TCCATTTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGACAGGCCAGCCTGGAGATCAGC
CCCATCACCTTCCTCACTGCCCAGACCCTGCTGATGGACCTCGGACAGTTCCTGCTGTCCTGCCAC
ATCAGCTCCCACCAGCATGATGGCATGGAGGCCTATGTCAAGGTGGACAGCTGCCCTGAGGAGCCA
CAGCTCAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGATG
GATGTGGTCCGCTTTGATGATGACAACAGCCCATCCTTCATTCAGATCAGGTCTGTGGCCAAGAAA
CACCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCACTGGTC
CTGGCCCCTGATGACAGGAGCTACAAGAGCCAGTACCTCAACAATGGCCCACAGAGGATTGGACGC
AAGTACAAGAAAGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATTCAG
CATGAGTCTGGCATCCTGGGCCCACTCCTGTATGGGGAGGTGGGGGACACCCTGCTCATCATCTTC
AAGAACCAGGCCTCCAGGCCCTACAACATCTACCCACATGGCATCACTGATGTCAGGCCCCTGTAC
AGCCGCAGGCTGCCAAAGGGGGTGAAACACCTCAAGGACTTCCCCATTCTGCCTGGGGAGATCTTC
AAGTACAAGTGGACTGTCACTGTGGAGGATGGACCAACCAAATCTGACCCCAGGTGCCTCACCAGA
TACTACTCCAGCTTTGTGAACATGGAGAGGGACCTGGCCTCTGGCCTGATTGGCCCACTGCTCATC
TGCTACAAGGAGTCTGTGGACCAGAGGGGAAACCAGATCATGTCTGACAAGAGGAATGTGATTCTG
TTCTCTGTCTTTGATGAGAACAGGAGCTGGTACCTGACTGAGAACATTCAGCGCTTCCTGCCCAAC
CCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGCAACATCATGCACTCCATCAATGGC
TATGTGTTTGACAGCCTCCAGCTTTCTGTCTGCCTGCATGAGGTGGCCTACTGGTACATTCTTTCT
ATTGGGGCCCAGACTGACTTCCTTTCTGTCTTCTTCTCTGGCTACACCTTCAAACACAAGATGGTG
TATGAGGACACCCTGACCCTCTTCCCATTCTCTGGGGAGACTGTGTTCATGAGCATGGAGAACCCT
GGCCTGTGGATTCTGGGATGCCACAACTCTGACTTCCGCAACAGGGGCATGACTGCCCTGCTCAAA
GTCTCCTCCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGCCTAC
CTGCTCAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATCCACCTGTCCTGAAACGC
CACCAGAGGGAGATCACCAGGACCACCCTCCAGTCTGACCAGGAGGAGATTGACTATGATGACACC
ATTTCTGTGGAGATGAAGAAAGAGGACTTTGACATCTATGACGAGGACGAGAACCAGAGCCCAAGG
AGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGCGCCTGTGGGACTATGGCATG
AGCTCCAGCCCCCATGTCCTCAGGAACAGGGCCCAGTCTGGCTCTGTGCCACAGTTCAAGAAAGTG
GTCTTCCAAGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCAC
CTGGGACTCCTGGGCCCATACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCCGCAAC

(còn tiếp)

Fig. 58A

CAGGCCTCCAGGCCCTACAGCTTCTACAGCTCCCTCATCAGCTATGAGGAGGACCAGAGGCAGGGG
GCTGAGCCACGCAAGAACTTTGTGAAACCCAATGAAACCAAGACCTACTTCTGGAAAGTCCAGCAC
CACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTG
GAGAAGGATGTGCACTCTGGCCTGATTGGCCCACTCCTGGTCTGCCACACCAACACCCTGAACCCT
GCCCATGGAAGGCAAGTGACTGTGCAGGAGTTTGCCCTCTTCTTCACCATCTTTGATGAAACCAAG
AGCTGGTACTTCACTGAGAACATGGAGCGCAACTGCAGGGCCCCATGCAACATTCAGATGGAGGAC
CCCACCTTCAAAGAGAACTACCGCTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGG
CTTGTCATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTTTCTATGGGCTCCAATGAGAACATT
CACTCCATCCACTTCTCTGGGCATGTCTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCCCTG
TACAACCTCTACCCTGGGGTCTTTGAGACTGTGGAGATGCTGCCCTCCAAAGCTGGCATCTGGAGG
GTGGAGTGCCTCATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTCTACAGCAAC
AAGTGCCAGACCCCCCTGGGAATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGC
CAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTCCACTACTCTGGATCCATCAATGCCTGGAGC
ACCAAGGAGCCATTCAGCTGGATCAAAGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAG
ACCCAGGGGGCCAGGCAGAAGTTCTCCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTG
GATGGCAAGAAATGGCAGACCTACAGAGGCAACTCCACTGGAACACTCATGGTCTTCTTTGGCAAT
GTGGACAGCTCTGGCATCAAGCACAACATCTTCAACCCCCCAATCATCGCCAGATACATCAGGCTG
CACCCCACCCACTACAGCATCCGCAGCACCCTCAGGATGGAGCTGATGGGCTGTGACCTGAACTCC
TGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATTTCTGATGCCCAGATCACTGCCTCCAGCTAC
TTCACCAACATGTTTGCCACCTGGAGCCCAAGCAAGGCCAGGCTGCACCTCCAGGGAAGGAGCAAT
GCCTGGAGGCCCCAGGTCAACAACCCAAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAG
GTCACTGGGGTGACCACCCAGGGGGTCAAGAGCCTGCTCACCAGCATGTATGTGAAGGAGTTCCTG
ATCAGCTCCAGCCAGGATGGCCACCAGTGGACCCTCTTCTTCCAGAATGGCAAGGTCAAGGTGTTC
CAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTCCTGACCAGATAC
CTGAGGATTCACCCCCAGAGCTGGGTCCACCAGATTGCCCTGAGGATGGAGGTCCTGGGATGTGAG
GCCCAGGACCTGTACTGA

Fig. 58B

DANH MỤC TRÌNH TỰ

<110> Takeda Pharmaceutical Company Limited

<120> POLYNUCLEOTIT ĐƯỢC BIẾN ĐỔI CODON MÃ HÓA CÁC BIẾN THỂ FVIII TÁI TỔ HỢP CÓ SỰ BIỂU HIỆN TĂNG CHO LIỆU PHÁP GEN ĐỂ ĐIỀU TRỊ BỆNH HEMOPHILIA A, VECTƠ VÀ HẠT VIRUT LIÊN HỢP ADENO CHỨA POLYNUCLEOTIT NÀY, QUY TRÌNH SẢN SINH HẠT VIRUT LIÊN HỢP ADENO, TẾ BÀO CHỦ ĐƯỢC GÂY NHIỄM BẰNG HẠT VIRUT NÀY, VÀ QUY TRÌNH TẢI NẠP TẾ BÀO CHỦ

<130> 008073-5107-WO

<140>
<141>

<150> 62/255,317
<151> 13-11-2015

<160> 109

<170> PatentIn phiên bản 3.5

<210> 1
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 1
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaggagc 2280

ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccaccctc 2340

cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400

gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520

catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caagagttca ctgatggcag cttcacccag cccctgtaca gaggggagct gaatgagcac 2640

ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgaccttc 2700

cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760

cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttcac tgagaacatg 3060

gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120

taccgcttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggcc 3180

caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttcctg 3420

gtctacagca acaagtgcca gacccccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540

tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600

ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660

ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720

agaggcaact ccactggaac actcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca ccccacccac 3840

tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctcctgcagc 3900

atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960

ttcaccaaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccagggaagg 4020

agcaatgcct ggaggcccca ggtcaacaac ccaaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtcactgg ggtgaccacc caggggggtca agagcctgct caccagcatg 4140

tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc 4200

cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac 4260

agcctggacc ccccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac 4320

cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta ctga 4374

<210> 2
<211> 1457
<212> PRT

<213> Trình tự nhân tạo

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

<400> 2

```
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1               5                  10                  15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
            20                  25                  30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
        35                  40                  45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
    50                  55                  60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65                  70                  75                  80

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

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220
```

```
Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460
```

```
Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
        515                 520                 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
    530                 535                 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545                 550                 555                 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
                565                 570                 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
            580                 585                 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
        595                 600                 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
    610                 615                 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625                 630                 635                 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
                645                 650                 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
            660                 665                 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
        675                 680                 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
    690                 695                 700
```

```
Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705                 710                 715                 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
                725                 730                 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
            740                 745                 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
        755                 760                 765

Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
    770                 775                 780

Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
785                 790                 795                 800

Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
                805                 810                 815

Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
            820                 825                 830

Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
        835                 840                 845

Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
    850                 855                 860

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
865                 870                 875                 880

Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
                885                 890                 895

Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
            900                 905                 910

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
        915                 920                 925

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
    930                 935                 940

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
```

945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
980 985 990

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
995 1000 1005

Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn
1010 1015 1020

Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys
1025 1030 1035

Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr
1040 1045 1050

Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr
1055 1060 1065

Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe
1070 1075 1080

Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met
1085 1090 1095

Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met
1100 1105 1110

Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly
1115 1120 1125

Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser
1130 1135 1140

Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg
1145 1150 1155

Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro
1160 1165 1170

Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser
1175 1180 1185

```
Thr Lys  Glu Pro Phe Ser Trp  Ile Lys Val Asp Leu  Leu Ala Pro
    1190                1195                1200

Met Ile  Ile His Gly Ile Lys  Thr Gln Gly Ala Arg  Gln Lys Phe
    1205                1210                1215

Ser Ser  Leu Tyr Ile Ser Gln  Phe Ile Ile Met Tyr  Ser Leu Asp
    1220                1225                1230

Gly Lys  Lys Trp Gln Thr Tyr  Arg Gly Asn Ser Thr  Gly Thr Leu
    1235                1240                1245

Met Val  Phe Phe Gly Asn Val  Asp Ser Ser Gly Ile  Lys His Asn
    1250                1255                1260

Ile Phe  Asn Pro Pro Ile Ile  Ala Arg Tyr Ile Arg  Leu His Pro
    1265                1270                1275

Thr His  Tyr Ser Ile Arg Ser  Thr Leu Arg Met Glu  Leu Met Gly
    1280                1285                1290

Cys Asp  Leu Asn Ser Cys Ser  Met Pro Leu Gly Met  Glu Ser Lys
    1295                1300                1305

Ala Ile  Ser Asp Ala Gln Ile  Thr Ala Ser Ser Tyr  Phe Thr Asn
    1310                1315                1320

Met Phe  Ala Thr Trp Ser Pro  Ser Lys Ala Arg Leu  His Leu Gln
    1325                1330                1335

Gly Arg  Ser Asn Ala Trp Arg  Pro Gln Val Asn Asn  Pro Lys Glu
    1340                1345                1350

Trp Leu  Gln Val Asp Phe Gln  Lys Thr Met Lys Val  Thr Gly Val
    1355                1360                1365

Thr Thr  Gln Gly Val Lys Ser  Leu Leu Thr Ser Met  Tyr Val Lys
    1370                1375                1380

Glu Phe  Leu Ile Ser Ser Ser  Gln Asp Gly His Gln  Trp Thr Leu
    1385                1390                1395

Phe Phe  Gln Asn Gly Lys Val  Lys Val Phe Gln Gly  Asn Gln Asp
    1400                1405                1410
```

Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr
1415 1420 1425

Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala
1430 1435 1440

Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 3
<211> 2220
<212> ADN
<213> Trình tự nhân tạo

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

<400> 3
gccaccagga gatactacct gggggctgtg gagctttctt gggactacat gcagtctgac 60

ctgggggagc tgcctgtgga tgccaggttc ccacccagag tgcccaaatc cttcccattc 120

aacacctctg tggtctacaa gaagaccctc tttgtggagt tcactgacca cctgttcaac 180

attgccaaac ccaggccacc ctggatggga ctcctgggac ccaccattca ggctgaggtg 240

tatgacactg tggtcatcac cctcaagaac atggcctccc accctgtgag cctgcatgct 300

gtgggggtca gctactggaa ggcctctgag ggggctgagt atgatgacca gacctcccag 360

agggagaagg aggatgacaa agtgttccct gggggcagcc acacctatgt gtggcaggtc 420

ctcaaggaga atggccccat ggcctctgac ccactctgcc tgacctactc ctacctttct 480

catgtggacc tggtcaagga cctcaactct ggactgattg gggccctgct ggtgtgcagg 540

gagggctccc tggccaaaga gaagacccag accctgcaca agttcattct cctgtttgct 600

gtctttgatg agggcaagag ctggcactct gaaaccaaga actccctgat gcaggacagg 660

gatgctgcct ctgccagggc ctggcccaag atgcacactg tgaatggcta tgtgaacagg 720

agcctgcctg gactcattgg ctgccacagg aaatctgtct actggcatgt gattggcatg 780

gggacaaccc ctgaggtgca ctccattttc ctggagggcc acaccttcct ggtcaggaac 840

cacagacagg ccagcctgga gatcagcccc atcaccttcc tcactgccca gaccctgctg 900

atggacctcg gacagttcct gctgttctgc cacatcagct cccaccagca tgatggcatg 960

gaggcctatg tcaaggtgga cagctgccct gaggagccac agctcaggat gaagaacaat 1020

gaggaggctg aggactatga tgatgacctg actgactctg agatggatgt ggtccgcttt 1080

gatgatgaca acagcccatc cttcattcag atcaggtctg tggccaagaa acaccccaag 1140

acctgggtgc actacattgc tgctgaggag gaggactggg actatgcccc actggtcctg 1200

gcccctgatg acaggagcta caagagccag tacctcaaca atggcccaca gaggattgga 1260

cgcaagtaca agaaagtcag gttcatggcc tacactgatg aaaccttcaa gaccagggag 1320

gccattcagc atgagtctgg catcctgggc ccactcctgt atggggaggt gggggacacc 1380

ctgctcatca tcttcaagaa ccaggcctcc aggccctaca acatctaccc acatggcatc 1440

actgatgtca ggcccctgta cagccgcagg ctgccaaagg gggtgaaaca cctcaaggac 1500

ttccccattc tgcctgggga gatcttcaag tacaagtgga ctgtcactgt ggaggatgga 1560

ccaaccaaat ctgaccccag gtgcctcacc agatactact ccagctttgt gaacatggag 1620

agggacctgg cctctggcct gattggccca ctgctcatct gctacaagga gtctgtggac 1680

cagaggggaa accagatcat gtctgacaag aggaatgtga ttctgttctc tgtctttgat 1740

gagaacagga gctggtacct gactgagaac attcagcgct tcctgcccaa ccctgctggg 1800

gtgcagctgg aggaccctga gttccaggcc agcaacatca tgcactccat caatggctat 1860

gtgtttgaca gcctccagct ttctgtctgc ctgcatgagg tggcctactg gtacattctt 1920

tctattgggg cccagactga cttcctttct gtcttcttct ctggctacac cttcaaacac 1980

aagatggtgt atgaggacac cctgaccctc ttcccattct ctggggagac tgtgttcatg 2040

agcatggaga accctggcct gtggattctg ggatgccaca actctgactt ccgcaacagg 2100

ggcatgactg ccctgctcaa agtctcctcc tgtgacaaga acactgggga ctactatgag 2160

gacagctatg aggacatctc tgcctacctg ctcagcaaga acaatgccat tgagcccagg 2220

<210> 4
<211> 2052
<212> ADN
<213> Trình tự nhân tạo

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

<400> 4
gagatcacca ggaccaccct ccagtctgac caggaggaga ttgactatga tgacaccatt 60

tctgtggaga tgaagaaaga ggactttgac atctatgacg aggacgagaa ccagagccca 120

aggagcttcc agaagaagac caggcactac ttcattgctg ctgtggagcg cctgtgggac 180

tatggcatga gctccagccc ccatgtcctc aggaacaggg cccagtctgg ctctgtgcca 240

cagttcaaga aagtggtctt ccaagagttc actgatggca gcttcaccca gcccctgtac 300

agaggggagc tgaatgagca cctgggactc ctgggcccat acatcagggc tgaggtggag 360

gacaacatca tggtgacctt ccgcaaccag gcctccaggc cctacagctt ctacagctcc 420

ctcatcagct atgaggagga ccagaggcag ggggctgagc cacgcaagaa ctttgtgaaa 480

cccaatgaaa ccaagaccta cttctggaaa gtccagcacc acatggcccc caccaaggat 540

gagtttgact gcaaggcctg ggcctacttc tctgatgtgg acctggagaa ggatgtgcac 600

tctggcctga ttggcccact cctggtctgc cacaccaaca ccctgaaccc tgcccatgga 660

aggcaagtga ctgtgcagga gtttgccctc ttcttcacca tctttgatga aaccaagagc 720

tggtacttca ctgagaacat ggagcgcaac tgcagggccc catgcaacat tcagatggag 780

gaccccacct tcaaagagaa ctaccgcttc catgccatca atggctacat catggacacc 840

ctgcctgggc ttgtcatggc ccaggaccag aggatcaggt ggtacctgct ttctatgggc 900

tccaatgaga acattcactc catccacttc tctgggcatg tcttcactgt gcgcaagaag 960

gaggagtaca agatggccct gtacaacctc taccctgggg tctttgagac tgtggagatg 1020

ctgccctcca aagctggcat ctggagggtg gagtgcctca ttggggagca cctgcatgct 1080

ggcatgagca ccctgttcct ggtctacagc aacaagtgcc agacccccct gggaatggcc 1140

tctggccaca tcagggactt ccagatcact gcctctggcc agtatggcca gtgggccccc 1200

aagctggcca ggctccacta ctctggatcc atcaatgcct ggagcaccaa ggagccattc 1260

agctggatca aagtggacct gctggccccc atgatcatcc atggcatcaa gacccagggg 1320

gccaggcaga agttctccag cctgtacatc agccagttca tcatcatgta cagcctggat 1380

ggcaagaaat ggcagaccta cagaggcaac tccactggaa cactcatggt cttctttggc 1440

aatgtggaca gctctggcat caagcacaac atcttcaacc ccccaatcat cgccagatac 1500

atcaggctgc accccaccca ctacagcatc cgcagcaccc tcaggatgga gctgatgggc 1560

tgtgacctga actcctgcag catgcccctg ggcatggaga gcaaggccat ttctgatgcc 1620

cagatcactg cctccagcta cttcaccaac atgtttgcca cctggagccc aagcaaggcc 1680

aggctgcacc tccagggaag gagcaatgcc tggaggcccc aggtcaacaa cccaaaggag 1740

tggctgcagg tggacttcca gaagaccatg aaggtcactg gggtgaccac ccaggggtc 1800

aagagcctgc tcaccagcat gtatgtgaag gagttcctga tcagctccag ccaggatggc 1860

caccagtgga ccctcttctt ccagaatggc aaggtcaagg tgttccaggg caaccaggac 1920

agcttcaccc ctgtggtgaa cagcctggac ccccccctcc tgaccagata cctgaggatt 1980

caccccaga gctgggtcca ccagattgcc ctgaggatgg aggtcctggg atgtgaggcc 2040

caggacctgt ac 2052

<210> 5
<211> 42
<212> ADN
<213> Trình tự nhân tạo

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

<400> 5
agcttctctc agaatccacc tgtcctgaag agacaccaga ga 42

<210> 6
<211> 42
<212> ADN
<213> Trình tự nhân tạo

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

<400> 6
agcttcagcc agaatccacc tgtcctgaaa cgccaccaga gg 42

<210> 7
<211> 42
<212> ADN
<213> Trình tự nhân tạo

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

<400> 7
agcttcagcc agaacccccc cgtgctgaag aggcaccaga gg 42

<210> 8
<211> 7827
<212> ADN
<213> Trình tự nhân tạo

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

<400> 8
tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60

cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120

ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180

accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240

attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat 300

tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360

tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cctcgagatt taaatgacgt 420

tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca aaggtcgccc 480

gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcaga gagggagtgg 540

ccaactccat cactaggggt tcctgagttt aaacttcgtc gacgattcga gcttgggctg 600

caggtcgagg gcactgggag gatgttgagt aagatggaaa actactgatg acccttgcag 660

agacagagta ttaggacatg tttgaacagg ggccgggcga tcagcaggta gctctagagg 720

atccccgtct gtctgcacat ttcgtagagc gagtgttccg atactctaat ctccctaggc 780

aaggttcata tttgtgtagg ttacttattc tccttttgtt gactaagtca ataatcagaa 840

tcagcaggtt tggagtcagc ttggcaggga tcagcagcct gggttggaag gaggggtat 900

aaaagcccct tcaccaggag aagccgtcac acagactagg cgcgccaccg ccaccatgca 960

gattgagctg agcacctgct tcttcctgtg cctgctgagg ttctgcttct ctgccaccag 1020

gagatactac ctgggggctg tggagctttc ttgggactac atgcagtctg acctggggga 1080

gctgcctgtg gatgccaggt tcccacccag agtgcccaaa tccttcccat tcaacacctc 1140

tgtggtctac aagaagaccc tctttgtgga gttcactgac cacctgttca acattgccaa 1200

acccaggcca ccctggatgg gactcctggg acccaccatt caggctgagg tgtatgacac 1260

tgtggtcatc accctcaaga acatggcctc ccaccctgtg agcctgcatg ctgtgggggt 1320

cagctactgg aaggcctctg agggggctga gtatgatgac cagacctccc agagggagaa 1380

ggaggatgac aaagtgttcc ctgggggcag ccacacctat gtgtggcagg tcctcaagga 1440

gaatggcccc atggcctctg acccactctg cctgacctac tcctaccttt ctcatgtgga 1500

cctggtcaag gacctcaact ctggactgat tggggccctg ctggtgtgca gggagggctc 1560

cctggccaaa gagaagaccc agaccctgca caagttcatt ctcctgtttg ctgtctttga 1620

tgagggcaag agctggcact ctgaaaccaa gaactccctg atgcaggaca gggatgctgc 1680

ctctgccagg gcctggccca agatgcacac tgtgaatggc tatgtgaaca ggagcctgcc 1740

tggactcatt ggctgccaca ggaaatctgt ctactggcat gtgattggca tggggacaac 1800

ccctgaggtg cactccattt tcctggaggg ccacaccttc ctggtcagga accacagaca 1860

ggccagcctg gagatcagcc ccatcacctt cctcactgcc cagaccctgc tgatggacct 1920

cggacagttc ctgctgttct gccacatcag ctcccaccag catgatggca tggaggccta 1980

tgtcaaggtg gacagctgcc ctgaggagcc acagctcagg atgaagaaca atgaggaggc 2040

tgaggactat gatgatgacc tgactgactc tgagatggat gtggtccgct ttgatgatga 2100

caacagccca tccttcattc agatcaggtc tgtggccaag aaacacccca agacctgggt 2160

gcactacatt gctgctgagg aggaggactg ggactatgcc ccactggtcc tggcccctga 2220

tgacaggagc tacaagagcc agtacctcaa caatggccca cagaggattg gacgcaagta 2280

caagaaagtc aggttcatgg cctacactga tgaaaccttc aagaccaggg aggccattca 2340

gcatgagtct ggcatcctgg gcccactcct gtatggggag gtgggggaca ccctgctcat 2400

catcttcaag aaccaggcct ccaggcccta caacatctac ccacatggca tcactgatgt 2460

caggcccctg tacagccgca ggctgccaaa gggggtgaaa cacctcaagg acttccccat 2520

tctgcctggg gagatcttca agtacaagtg gactgtcact gtggaggatg gaccaaccaa 2580

atctgacccc aggtgcctca ccagatacta ctccagcttt gtgaacatgg agagggacct 2640

ggcctctggc ctgattggcc cactgctcat ctgctacaag gagtctgtgg accagagggg 2700

aaaccagatc atgtctgaca agaggaatgt gattctgttc tctgtctttg atgagaacag 2760

gagctggtac ctgactgaga acattcagcg cttcctgccc aaccctgctg gggtgcagct 2820

ggaggaccct gagttccagg ccagcaacat catgcactcc atcaatggct atgtgtttga 2880

cagcctccag ctttctgtct gcctgcatga ggtggcctac tggtacattc tttctattgg 2940

ggcccagact gacttccttt ctgtcttctt ctctggctac accttcaaac acaagatggt 3000

gtatgaggac accctgaccc tcttcccatt ctctggggag actgtgttca tgagcatgga 3060

gaaccctggc ctgtggattc tgggatgcca caactctgac ttccgcaaca ggggcatgac 3120

tgccctgctc aaagtctcct cctgtgacaa gaacactggg gactactatg aggacagcta 3180

tgaggacatc tctgcctacc tgctcagcaa gaacaatgcc attgagccca ggagcttcag 3240

ccagaatcca cctgtcctga aacgccacca gagggagatc accaggacca ccctccagtc 3300

tgaccaggag gagattgact atgatgacac catttctgtg gagatgaaga aagaggactt 3360

tgacatctat gacgaggacg agaaccagag cccaaggagc ttccagaaga agaccaggca 3420

ctacttcatt gctgctgtgg agcgcctgtg ggactatggc atgagctcca gcccccatgt 3480

cctcaggaac agggcccagt ctggctctgt gccacagttc aagaaagtgg tcttccaaga 3540

gttcactgat ggcagcttca cccagcccct gtacagaggg gagctgaatg agcacctggg 3600

actcctgggc ccatacatca gggctgaggt ggaggacaac atcatggtga ccttccgcaa 3660

ccaggcctcc aggccctaca gcttctacag ctccctcatc agctatgagg aggaccagag 3720

gcaggggget gagccacgca agaactttgt gaaacccaat gaaaccaaga cctacttctg 3780

gaaagtccag caccacatgg cccccaccaa ggatgagttt gactgcaagg cctgggccta 3840

cttctctgat gtggacctgg agaaggatgt gcactctggc ctgattggcc cactcctggt 3900

ctgccacacc aacaccctga accctgccca tggaaggcaa gtgactgtgc aggagtttgc 3960

cctcttcttc accatctttg atgaaaccaa gagctggtac ttcactgaga acatggagcg 4020

caactgcagg gccccatgca acattcagat ggaggacccc accttcaaag agaactaccg 4080

cttccatgcc atcaatggct acatcatgga caccctgcct gggcttgtca tggcccagga 4140

ccagaggatc aggtggtacc tgctttctat gggctccaat gagaacattc actccatcca 4200

cttctctggg catgtcttca ctgtgcgcaa gaaggaggag tacaagatgg ccctgtacaa 4260

cctctaccct ggggtctttg agactgtgga gatgctgccc tccaaagctg gcatctggag 4320

ggtggagtgc ctcattgggg agcacctgca tgctggcatg agcaccctgt tcctggtcta 4380

cagcaacaag tgccagaccc ccctgggaat ggcctctggc cacatcaggg acttccagat 4440

cactgcctct ggccagtatg gccagtgggc ccccaagctg gccaggctcc actactctgg 4500

atccatcaat gcctggagca ccaaggagcc attcagctgg atcaaagtgg acctgctggc 4560

ccccatgatc atccatggca tcaagaccca gggggccagg cagaagttct ccagcctgta 4620

catcagccag ttcatcatca tgtacagcct ggatggcaag aaatggcaga cctacagagg 4680

caactccact ggaacactca tggtcttctt tggcaatgtg gacagctctg gcatcaagca 4740

caacatcttc aacccccaa tcatcgccag atacatcagg ctgcacccca cccactacag 4800

catccgcagc accctcagga tggagctgat gggctgtgac ctgaactcct gcagcatgcc 4860

cctgggcatg gagagcaagg ccatttctga tgcccagatc actgcctcca gctacttcac 4920

caacatgttt gccacctgga gcccaagcaa ggccaggctg cacctccagg gaaggagcaa 4980

tgcctggagg ccccaggtca acaacccaaa ggagtggctg caggtggact tccagaagac 5040

catgaaggtc actggggtga ccacccaggg ggtcaagagc ctgctcacca gcatgtatgt 5100

gaaggagttc ctgatcagct ccagccagga tggccaccag tggaccctct tcttccagaa 5160

tggcaaggtc aaggtgttcc agggcaacca ggacagcttc acccctgtgg tgaacagcct 5220

ggaccccccc ctcctgacca gatacctgag gattcacccc cagagctggg tccaccagat 5280

tgccctgagg atggaggtcc tgggatgtga ggcccaggac ctgtactgat gacgagcggc 5340

cgctcttagt agcagtatcg ataataaaag atctttattt tcattagatc tgtgtgttgg 5400

ttttttgtgt gttaattaag ctcgcgaagg aacccctagt gatggagttg gccactccct 5460

ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaaa ggtcgcccga cgcccgggct 5520

ttgcccgggc ggcctcagtg agcgagcgag cgcgcagaga gggagtggcc aagacgattt 5580

aaatgacaag cttggcgtaa tcatggtcat agctgtttcc tgtgtgaaat tgttatccgc 5640

tcacaattcc acacaacata cgagccggaa gcataaagtg taaagcctgg ggtgcctaat 5700

gagtgagcta actcacatta attgcgttgc gctcactgcc cgctttccag tcgggaaacc 5760

tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg gagaggcggt ttgcgtattg 5820

ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag 5880

cggtatcagc tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag 5940

gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc 6000

tggcgttttt ccataggctc cgcccccctg acgagcatca caaaaatcga cgctcaagtc 6060

agaggtggcg aaacccgaca ggactataaa gataccaggc gtttcccccct ggaagctccc 6120

tcgtgcgctc tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt 6180

cgggaagcgt ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg 6240

ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat 6300

ccggtaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag 6360

ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt 6420

ggtggcctaa ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc 6480

cagttacctt cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta 6540

gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag 6600

atcctttgat cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga 6660

ttttggtcat gagattatca aaaaggatct tcacctagat ccttttaaat taaaaatgaa 6720

gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa 6780

tcagtgaggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 6840

ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 6900

taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 6960

gggccgagcg cagaagtggt cctgcaactt tatccgcctc catccagtct attaattgtt 7020

gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg 7080

ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 7140

aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcggtt agctccttcg 7200

gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag 7260

cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt 7320

actcaaccaa gtcattctga gaatagtgta tgcggcgacc gagttgctct tgcccggcgt 7380

caatacggga taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac 7440

gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac 7500

ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 7560

caaaaacagg aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa 7620

tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga 7680

gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc 7740

cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa 7800

ataggcgtat cacgaggccc tttcgtc 7827

<210> 9
<211> 4332
<212> ADN
<213> Trình tự nhân tạo

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

<400> 9
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagggag 2280

atcaccagga ccaccctcca gtctgaccag gaggagattg actatgatga caccatttct 2340

gtggagatga agaaagagga ctttgacatc tatgacgagg acgagaacca gagcccaagg 2400

agcttccaga agaagaccag gcactacttc attgctgctg tggagcgcct gtgggactat 2460

ggcatgagct ccagccccca tgtcctcagg aacagggccc agtctggctc tgtgccacag 2520

ttcaagaaag tggtcttcca agagttcact gatggcagct tcacccagcc cctgtacaga 2580

ggggagctga atgagcacct gggactcctg ggcccataca tcagggctga ggtggaggac 2640

aacatcatgg tgaccttccg caaccaggcc tccaggccct acagcttcta cagctccctc 2700

atcagctatg aggaggacca gaggcagggg gctgagccac gcaagaactt tgtgaaaccc 2760

aatgaaacca agacctactt ctggaaagtc cagcaccaca tggcccccac caaggatgag 2820

tttgactgca aggcctgggc ctacttctct gatgtggacc tggagaagga tgtgcactct 2880

ggcctgattg gcccactcct ggtctgccac accaacaccc tgaaccctgc ccatggaagg 2940

caagtgactg tgcaggagtt tgccctcttc ttcaccatct ttgatgaaac caagagctgg 3000

tacttcactg agaacatgga gcgcaactgc agggccccat gcaacattca gatggaggac 3060

cccaccttca aagagaacta ccgcttccat gccatcaatg gctacatcat ggacaccctg 3120

cctgggcttg tcatggccca ggaccagagg atcaggtggt acctgctttc tatgggctcc 3180

aatgagaaca ttcactccat ccacttctct gggcatgtct tcactgtgcg caagaaggag 3240

gagtacaaga tggccctgta caacctctac cctggggtct ttgagactgt ggagatgctg 3300

ccctccaaag ctggcatctg gagggtggag tgcctcattg gggagcacct gcatgctggc 3360

atgagcaccc tgttcctggt ctacagcaac aagtgccaga ccccccctggg aatggcctct 3420

ggccacatca gggacttcca gatcactgcc tctggccagt atggccagtg ggcccccaag 3480

ctggccaggc tccactactc tggatccatc aatgcctgga gcaccaagga gccattcagc 3540

tggatcaaag tggacctgct ggcccccatg atcatccatg gcatcaagac ccaggggggcc 3600

aggcagaagt tctccagcct gtacatcagc cagttcatca tcatgtacag cctggatggc 3660

aagaaatggc agacctacag aggcaactcc actggaacac tcatggtctt ctttggcaat 3720

gtggacagct ctggcatcaa gcacaacatc ttcaaccccc caatcatcgc cagatacatc 3780

aggctgcacc ccacccacta cagcatccgc agcaccctca ggatggagct gatgggctgt 3840

gacctgaact cctgcagcat gcccctgggc atggagagca aggccatttc tgatgcccag 3900

atcactgcct ccagctactt caccaacatg tttgccacct ggagcccaag caaggccagg 3960

ctgcacctcc agggaaggag caatgcctgg aggccccagg tcaacaaccc aaaggagtgg 4020

ctgcaggtgg acttccagaa gaccatgaag gtcactgggg tgaccaccca gggggtcaag 4080

agcctgctca ccagcatgta tgtgaaggag ttcctgatca gctccagcca ggatggccac 4140

cagtggaccc tcttcttcca gaatggcaag gtcaaggtgt tccagggcaa ccaggacagc 4200

ttcacccctg tggtgaacag cctggacccc cccctcctga ccagatacct gaggattcac 4260

ccccagagct gggtccacca gattgccctg aggatggagg tcctgggatg tgaggcccag 4320

gacctgtact ga 4332

<210> 10
<211> 1443
<212> PRT
<213> Trình tự nhân tạo

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

<400> 10
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val

```
    50                  55                  60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65                  70                  75                  80

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

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300
```

```
Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
        515                 520                 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
    530                 535                 540
```

```
Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545                 550                 555                 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
                565                 570                 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
            580                 585                 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
        595                 600                 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
    610                 615                 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625                 630                 635                 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
                645                 650                 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
            660                 665                 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
        675                 680                 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
    690                 695                 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705                 710                 715                 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
                725                 730                 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
            740                 745                 750

Asn Asn Ala Ile Glu Pro Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser
        755                 760                 765

Asp Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys
    770                 775                 780
```

```
Lys Glu Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg
785                 790                 795                 800

Ser Phe Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg
                805                 810                 815

Leu Trp Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg
            820                 825                 830

Ala Gln Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu
        835                 840                 845

Phe Thr Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn
    850                 855                 860

Glu His Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp
865                 870                 875                 880

Asn Ile Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe
                885                 890                 895

Tyr Ser Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu
            900                 905                 910

Pro Arg Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp
        915                 920                 925

Lys Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys
    930                 935                 940

Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser
945                 950                 955                 960

Gly Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro
                965                 970                 975

Ala His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr
            980                 985                 990

Ile Phe Asp Glu Thr Lys Ser Trp  Tyr Phe Thr Glu Asn  Met Glu Arg
        995                 1000                1005

Asn Cys  Arg Ala Pro Cys Asn  Ile Gln Met Glu Asp  Pro Thr Phe
    1010                1015                1020

Lys Glu  Asn Tyr Arg Phe His  Ala Ile Asn Gly Tyr  Ile Met Asp
```

1025 1030 1035

Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp
1040 1045 1050

Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His
1055 1060 1065

Phe Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys
1070 1075 1080

Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu
1085 1090 1095

Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile
1100 1105 1110

Gly Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr
1115 1120 1125

Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile
1130 1135 1140

Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala
1145 1150 1155

Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp
1160 1165 1170

Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala
1175 1180 1185

Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys
1190 1195 1200

Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu
1205 1210 1215

Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr
1220 1225 1230

Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His
1235 1240 1245

Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His
1250 1255 1260

Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met
1265 1270 1275

Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser
1280 1285 1290

Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr
1295 1300 1305

Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu
1310 1315 1320

Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys
1325 1330 1335

Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly
1340 1345 1350

Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val
1355 1360 1365

Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr
1370 1375 1380

Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
1385 1390 1395

Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
1400 1405 1410

Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
1415 1420 1425

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1430 1435 1440

<210> 11
<211> 4368
<212> ADN
<213> Trình tự nhân tạo

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

<400> 11
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaggagc 2280

ttcagccaga attccagaca ccccagcacc agggagatca ccaggaccac cctccagtct 2340

gaccaggagg agattgacta tgatgacacc atttctgtgg agatgaagaa agaggacttt 2400

gacatctatg acgaggacga gaaccagagc ccaaggagct tccagaagaa gaccaggcac 2460

tacttcattg ctgctgtgga gcgcctgtgg gactatggca tgagctccag cccccatgtc 2520

ctcaggaaca gggcccagtc tggctctgtg ccacagttca agaaagtggt cttccaagag 2580

ttcactgatg gcagcttcac ccagcccctg tacagagggg agctgaatga gcacctggga 2640

ctcctgggcc catacatcag ggctgaggtg gaggacaaca tcatggtgac cttccgcaac 2700

caggcctcca ggccctacag cttctacagc tccctcatca gctatgagga ggaccagagg 2760

caggggctg agccacgcaa gaactttgtg aaacccaatg aaaccaagac ctacttctgg 2820

aaagtccagc accacatggc ccccaccaag gatgagtttg actgcaaggc ctgggcctac 2880

ttctctgatg tggacctgga gaaggatgtg cactctggcc tgattggccc actcctggtc 2940

tgccacacca acaccctgaa ccctgcccat ggaaggcaag tgactgtgca ggagtttgcc 3000

ctcttcttca ccatctttga tgaaaccaag agctggtact tcactgagaa catggagcgc 3060

aactgcaggg ccccatgcaa cattcagatg gaggacccca ccttcaaaga gaactaccgc 3120

ttccatgcca tcaatggcta catcatggac accctgcctg ggcttgtcat ggcccaggac 3180

cagaggatca ggtggtacct gctttctatg ggctccaatg agaacattca ctccatccac 3240

ttctctgggc atgtcttcac tgtgcgcaag aaggaggagt acaagatggc cctgtacaac 3300

ctctaccctg gggtctttga gactgtggag atgctgccct ccaaagctgg catctggagg 3360

gtggagtgcc tcattgggga gcacctgcat gctggcatga gcaccctgtt cctggtctac 3420

agcaacaagt gccagacccc cctgggaatg gcctctggcc acatcaggga cttccagatc 3480

actgcctctg gccagtatgg ccagtgggcc cccaagctgg ccaggctcca ctactctgga 3540

tccatcaatg cctggagcac caaggagcca ttcagctgga tcaaagtgga cctgctggcc 3600

cccatgatca tccatggcat caagacccag ggggccaggc agaagttctc agcctgtac 3660

atcagccagt tcatcatcat gtacagcctg gatggcaaga aatggcagac ctacagaggc 3720

aactccactg gaacactcat ggtcttcttt ggcaatgtgg acagctctgg catcaagcac 3780

aacatcttca acccccccaat catcgccaga tacatcaggc tgcaccccac ccactacagc 3840

atccgcagca ccctcaggat ggagctgatg ggctgtgacc tgaactcctg cagcatgccc 3900

ctgggcatgg agagcaaggc catttctgat gcccagatca ctgcctccag ctacttcacc 3960

aacatgtttg ccacctggag cccaagcaag gccaggctgc acctccaggg aaggagcaat 4020

gcctggaggc cccaggtcaa caacccaaag gagtggctgc aggtggactt ccagaagacc 4080

atgaaggtca ctggggtgac cacccagggg gtcaagagcc tgctcaccag catgtatgtg 4140

aaggagttcc tgatcagctc cagccaggat ggccaccagt ggaccctctt cttccagaat 4200

ggcaaggtca aggtgttcca gggcaaccag gacagcttca cccctgtggt gaacagcctg 4260

gacccccccc tcctgaccag atacctgagg attcaccccc agagctgggt ccaccagatt 4320

gccctgagga tggaggtcct gggatgtgag gcccaggacc tgtactga 4368

<210> 12
<211> 1455
<212> PRT
<213> Trình tự nhân tạo

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

<400> 12

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

```
His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350
```

```
Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
        515                 520                 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
    530                 535                 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545                 550                 555                 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
                565                 570                 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
            580                 585                 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
```

595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro
755 760 765

Ser Thr Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu
770 775 780

Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe
785 790 795 800

Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys
805 810 815

Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr
820 825 830

Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly
835 840 845

```
Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly
    850                 855                 860

Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly
865                 870                 875                 880

Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val
                885                 890                 895

Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu
            900                 905                 910

Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn
        915                 920                 925

Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His
    930                 935                 940

His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr
945                 950                 955                 960

Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly
                965                 970                 975

Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg
            980                 985                 990

Gln Val Thr Val Gln Glu Phe Ala  Leu Phe Phe Thr Ile  Phe Asp Glu
        995                 1000                1005

Thr Lys  Ser Trp Tyr Phe Thr  Glu Asn Met Glu Arg  Asn Cys Arg
    1010                1015                1020

Ala Pro  Cys Asn Ile Gln Met  Glu Asp Pro Thr Phe  Lys Glu Asn
    1025                1030                1035

Tyr Arg  Phe His Ala Ile Asn  Gly Tyr Ile Met Asp  Thr Leu Pro
    1040                1045                1050

Gly Leu  Val Met Ala Gln Asp  Gln Arg Ile Arg Trp  Tyr Leu Leu
    1055                1060                1065

Ser Met  Gly Ser Asn Glu Asn  Ile His Ser Ile His  Phe Ser Gly
    1070                1075                1080
```

```
His Val  Phe Thr Val Arg Lys  Lys Glu Glu Tyr Lys  Met Ala Leu
    1085                1090                1095

Tyr Asn  Leu Tyr Pro Gly Val  Phe Glu Thr Val Glu  Met Leu Pro
    1100                1105                1110

Ser Lys  Ala Gly Ile Trp Arg  Val Glu Cys Leu Ile  Gly Glu His
    1115                1120                1125

Leu His  Ala Gly Met Ser Thr  Leu Phe Leu Val Tyr  Ser Asn Lys
    1130                1135                1140

Cys Gln  Thr Pro Leu Gly Met  Ala Ser Gly His Ile  Arg Asp Phe
    1145                1150                1155

Gln Ile  Thr Ala Ser Gly Gln  Tyr Gly Gln Trp Ala  Pro Lys Leu
    1160                1165                1170

Ala Arg  Leu His Tyr Ser Gly  Ser Ile Asn Ala Trp  Ser Thr Lys
    1175                1180                1185

Glu Pro  Phe Ser Trp Ile Lys  Val Asp Leu Leu Ala  Pro Met Ile
    1190                1195                1200

Ile His  Gly Ile Lys Thr Gln  Gly Ala Arg Gln Lys  Phe Ser Ser
    1205                1210                1215

Leu Tyr  Ile Ser Gln Phe Ile  Ile Met Tyr Ser Leu  Asp Gly Lys
    1220                1225                1230

Lys Trp  Gln Thr Tyr Arg Gly  Asn Ser Thr Gly Thr  Leu Met Val
    1235                1240                1245

Phe Phe  Gly Asn Val Asp Ser  Ser Gly Ile Lys His  Asn Ile Phe
    1250                1255                1260

Asn Pro  Pro Ile Ile Ala Arg  Tyr Ile Arg Leu His  Pro Thr His
    1265                1270                1275

Tyr Ser  Ile Arg Ser Thr Leu  Arg Met Glu Leu Met  Gly Cys Asp
    1280                1285                1290

Leu Asn  Ser Cys Ser Met Pro  Leu Gly Met Glu Ser  Lys Ala Ile
    1295                1300                1305
```

Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe
1310 1315 1320

Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg
1325 1330 1335

Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu
1340 1345 1350

Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr
1355 1360 1365

Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe
1370 1375 1380

Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe
1385 1390 1395

Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe
1400 1405 1410

Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr
1415 1420 1425

Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg
1430 1435 1440

Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 13
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 13
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc 2280

ttctctcaga atccacctgt cctgaagaga caccagagag agatcaccag gacaaccctc 2340

cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag 2400

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca 2460

agacactact tcattgctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc 2520

catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caggagttca ctgatggctc attcacccag cccctgtaca gaggggaact gaatgagcac 2640

ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc 2700

agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac 2760

cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggcaccc accaaggatg agtttgactg caaggcctgg 2880

gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg 3060

gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac 3120

tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca 3180

caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttcctg 3420

gtctacagca acaagtgcca gacacccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540

tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg 3600

ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct 3660

ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720

agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca ccccacccac 3840

tactcaatca gatcaaccct caggatggaa ctgatgggat gtgacctgaa ctcctgctca 3900

atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atcctcttac 3960

ttcaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccagggaaga 4020

agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag 4080

aagacaatga aagtcactgg ggtgacaacc caggggtca agtctctgct cacctcaatg 4140

tatgtgaagg agttcctgat ctcttcctca caggatggcc accagtggac actcttcttc 4200

cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac 4260

tcactggacc ccccctcct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320

cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

<210> 14
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 14
atgcagatcg aactgagcac ttgcttcttc ctgtgtctcc tgcgcttttg cttctccgcc 60

acaaggagat actatctcgg tgccgtggag ctcagctggg actacatgca gagcgacttg 120

ggtgaactgc ctgtggacgc caggtttcca ccccgcgtgc ccaagagttt cccgttcaac 180

accagtgtcg tgtacaagaa aaccctcttc gtggaattca ccgaccacct gttcaacatc 240

gccaaaccgc gccctccctg gatggggctg ctcggcccga cgatccaggc tgaggtctat 300

gacacggtgg tgattaccct caagaacatg gctagccacc cggtgagcct gcacgccgtg 360

ggcgtgtcct attggaaagc gtccgagggt gcggagtacg atgaccagac ttcacagcgg 420

gagaaggaag acgacaaagt gttccccggg ggttcccaca cctatgtctg gcaggtcctg 480

aaggagaatg gtcctatggc ctccgaccca ttgtgcctca cctactctta cctaagccat 540

gtggatctcg tcaaggacct gaactcgggg ctgatcggcg ccctgctcgt gtgccgggag 600

ggctcactgg ccaaggagaa gacccaaact ctgcacaagt tcatcctgct gttcgcggta 660

ttcgacgagg ggaagtcctg gcactccgag accaagaaca gcctgatgca ggaccgcgac 720

gcagcctcgg cccgtgcgtg gccaaagatg cacaccgtga acggctacgt taacaggagc 780

ctacccggcc tgatcggctg ccaccgcaaa tcggtctact ggcatgtgat cggaatgggc 840

acaacgcccg aggtccacag tatcttcctc gagggccaca ctttcctggt ccggaatcac 900

cgccaggcca gcctggagat cagccccata acctttctga cggcgcagac cttactcatg 960

gatctcggcc agttcctcct gttctgccac atttcgtccc accagcacga tgggatggaa 1020

gcatatgtga aagtggactc ctgccccgag gaaccccagc ttaggatgaa gaacaatgag 1080

gaggccgagg actacgacga tgaccttacc gattcagaaa tggacgtagt acgctttgac 1140

gacgacaact ctccatcctt catacagatt cgctccgtcg ccaagaagca ccctaagact 1200

tgggtgcact acatcgcggc cgaggaggag gactgggatt atgctcccct ggtgctggcc 1260

cccgacgacc gcagctacaa gagccagtac ctgaataacg ggccccagcg catcggccgg 1320

aagtacaaga aagtgcggtt catggcttac acggacgaga ccttcaagac ccgggaggct 1380

atccagcatg agagcggcat cttggggccc ctcctgtacg gcgaagttgg agacacactg 1440

ctgatcatct tcaagaacca ggcgagcagg ccctacaaca tctaccccca cggcattacc 1500

gatgtccggc cgttgtacag ccgacggctg cccaagggcg tgaagcacct gaaggacttt 1560

ccgatcctgc cgggcgagat cttcaagtac aagtggactg tgaccgtgga ggatgggccg 1620

accaagagcg atccgcgctg cctgacccgt tactactcca gctttgtcaa tatggagcgc 1680

gacctcgcta gcggcttgat tggccctctg ctgatctgct acaaggagtc cgtggaccag 1740

aggggaatc agatcatgag tgacaagagg aacgtgatcc tgttctccgt gttcgacgaa 1800

aaccgcagct ggtatctcac cgagaatatc cagcgcttcc tgcccaaccc ggccggtgtg 1860

cagctggagg accccgagtt tcaggccagc aacatcatgc attctatcaa cggatatgtg 1920

tttgattccc tgcagctctc agtgtgtctg cacgaggtcg cctactggta tatcctcagc 1980

attggggcac agaccgactt cctgagcgtg ttcttctccg ggtatacctt caagcacaag 2040

atggtgtacg aggataccct gaccctgttc ccctttagcg gcgaaaccgt gtttatgtct 2100

atggagaacc ccgggctctg gatccttggc tgccataact ccgacttccg caaccgcgga 2160

atgaccgcgc tcctgaaagt gtcgagttgt gacaagaaca ccggcgacta ttacgaggac 2220

agttacgagg acatctctgc gtacctcctt agcaagaata acgccatcga gccaagatcc 2280

ttcagccaga acccccccagt gctgaagagg catcagcggg agatcacccg cacgaccctg 2340

cagtcggatc aggaggagat tgattacgac gacacgatca gtgtggagat gaagaaggag 2400

gacttcgaca tctacgacga agatgaaaac cagtcccctc ggtccttcca aaagaagacc 2460

cggcactact tcatcgccgc tgtggaacgc ctgtgggact atggaatgtc ttctagccct 2520

cacgttttga ggaaccgcgc ccagtcgggc agcgtgcccc agttcaagaa agtggtgttc 2580

caggagttca ccgacggctc cttcacccag ccactttacc ggggcgagct caatgaacat 2640

ctgggcctgc tgggacccta catcagggct gaggtggagg acaacatcat ggtgacattc 2700

cggaatcagg ccagcagacc atacagtttc tacagttcac tcatctccta cgaggaggac 2760

cagcgccagg gggctgaacc ccgtaagaac ttcgtgaagc caaacgaaac aaagacctac 2820

ttctggaagg tccagcacca catggcacct accaaggacg agttcgattg caaggcctgg 2880

gcctacttct ccgacgtgga cctggagaaa gatgtgcaca gcggcctgat tggccctctg 2940

ctggtgtgtc acacgaacac actcaaccct gcacacgggc ggcaggtcac tgtgcaggaa 3000

ttcgccctgt tctttaccat ctttgatgag acgaagtcct ggtatttcac cgaaaacatg 3060

gagaggaact gccgcgcacc ctgcaacatc cagatggaag atccgacatt caaggagaac 3120

taccggttcc atgccatcaa tggctacatc atggacaccc tgcctggcct cgtgatggcc 3180

caagaccagc gtatccgctg gtatctgctg tcgatgggct ccaacgagaa catccatagt 3240

atccacttca gcgggcatgt cttcacggtg aggaaaaagg aggagtacaa gatggcactg 3300

tacaacctct atcccggcgt gttcgagacc gtggagatgc tgccctccaa ggccggcatc 3360

tggagagtgg aatgcctgat cggcgagcac ctccacgctg ggatgtccac gctgttcctc 3420

gtttacagca ataagtgcca gacccctctg ggcatggcga gcggccacat ccgcgacttc 3480

cagattacag ccagcggcca gtacggtcag tgggctccaa agctggcccg tctgcactac 3540

tccggatcca tcaacgcctg gtccaccaag gaaccgttct cctggatcaa agtagacctg 3600

ctagccccca tgatcattca cggcatcaag acacaaggcg cccgacagaa gttctcgagc 3660

ctctatatct cccagttcat catcatgtat agcctggacg gaaagaagtg gcagacttac 3720

cgcggaaact cgacagggac cctgatggta ttcttcggta acgtggacag ctccggaatc 3780

aagcacaaca tcttcaaccc acccattatc gcccgctaca tccgcctgca ccccactcac 3840

tatagcatta ggtccaccct gcgaatggag ctcatgggct gtgacctgaa cagctgtagc 3900

atgcccctcg gcatggagtc taaggcgatc tccgacgcac agataacggc atcatcctac 3960

tttaccaaca tgttcgctac ctggtccccc tccaaggccc gactccacct gcaagggaga 4020

tccaacgcct ggcggccaca ggtcaacaat cccaaggagt ggctgcaagt ggactttcag 4080

aaaactatga aagtcaccgg agtgaccaca cagggagtga agtctctgct gaccagcatg 4140

tacgtgaagg agttcctcat ctccagttcg caggatggcc accagtggac gttgttcttc 4200

caaaacggta aagtcaaagt cttccaaggg aaccaggaca gctttacacc cgtcgtgaac 4260

tccctggacc ccccgcttct cactagatac ctccgcatcc accctcagag ctgggtgcac 4320

cagattgccc tgcgcatgga ggttctgggg tgtgaagccc aggacctgta ctaa 4374

<210> 15
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 15
atgcagattg agctctccac ctgcttcttt ctctgccttc ttcgcttctg cttttctgcc 60

acacgcaggt actatttggg agcagtggaa ctgagctggg attacatgca gagtgacctt 120

ggtgaacttc ctgtggacgc tcgttttcca cctagagttc ccaagtcctt ccccttcaac 180

acctcagtgg tctacaagaa aacgctgttt gtggagttca ctgaccacct cttcaacatt 240

gccaaaccaa gacccccttg gatgggattg ctgggaccca caatacaagc agaagtctac 300

gacacggtgg tgattaccct gaagaacatg gcgtcacacc ctgtttcact tcacgctgtt 360

ggggtcagtt attggaaagc ctcagagggt gcggaatacg atgatcaaac cagccagagg 420

gagaaggaag atgacaaggt ctttcctggg ggtagccata cctatgtttg gcaggtgctg 480

aaagagaatg ggcctatggc ctctgatccc ttgtgcctca catactctta cctgagtcac 540

gtcgacctgg tgaaagacct gaatagcggt ctgattggtg cactgcttgt ttgtagagag 600

gggagtttgg ccaaggagaa aactcagact ctccacaagt ttatcctcct gtttgctgtg 660

ttcgacgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca ggacagagat 720

gctgcatctg caagggcttg gccaaaaatg cacacagtga acggctatgt gaatcgatca 780

ctgccaggac tgataggctg tcatcgcaag tcagtgtatt ggcacgttat cgggatggga 840

acaactccag aagtgcacag catcttcctt gagggccaca ctttcctggt tcggaatcat 900

agacaggcca gccttgagat cagcccaatc acctttctga ctgcccaaac cttgctgatg 960

gatctgggac agttcctcct gttttgtcac atctcctccc accaacatga cgggatggag 1020

gcttatgtga aggtcgatag ctgtccggag gaaccacaac tgaggatgaa gaacaacgaa 1080

gaggcagagg actatgacga cgatctgact gacagtgaaa tggacgtggt tcggttcgac 1140

gatgacaatt ctccttcatt tatccagatc cgttccgtgg ccaagaagca ccccaagact 1200

tgggttcatt acatcgctgc tgaggaggag gattgggact acgcgccctt ggtgttggcc 1260

ccagacgatc gctcatacaa gagccagtac cttaacaatg gtccacaaag gatcggccgg 1320

aagtacaaga aggttagatt tatggcttat accgacgaga cttttaaaac tagggaagca 1380

attcagcatg aaagtggcat tcttggaccc ctgctgtatg gcgaggttgg cgacaccctg 1440

ctgattatct ttaagaacca ggcaagccgg ccctacaaca tctacccgca cggcataacc 1500

gatgtacgac ccctgtacag tcgcagactt cctaaagggg tgaaacacct gaaggacttc 1560

ccaattctgc ccggggagat cttcaagtat aaatggaccg tgacggttga ggatggtccc 1620

acaaagtccg atccgagatg ccttacccga tattattcca gcttcgtgaa catggaaagg 1680

gacctggcca gcgggctgat tggcccactg ctgatttgtt acaaggagtc tgtcgatcaa 1740

agaggaaacc aaataatgag cgacaaacgt aacgtcatcc tgttcagcgt ctttgatgag 1800

aatagaagct ggtacctcac agaaaatatt cagcggtttc tgcctaaccc cgcaggcgtc 1860

cagctggaag atcccgagtt ccaagcctca aacatcatgc atagcatcaa cggatacgta 1920

ttcgatagcc tgcagctgtc cgtctgtctc catgaagtgg catattggta catcctgagt 1980

atcggggcgc agaccgactt cctgagcgtg ttcttttctg gatacacgtt caaacacaaa 2040

atggtctatg aagataccct gactctgttt ccattctcag gagagacagt ctttatgagt 2100

atggaaaatc ctggactgtg gatcctgggc tgtcacaatt ctgattttcg gaacagaggc 2160

atgacagccc tgcttaaagt gagctcatgc gacaagaaca ccggtgatta ctacgaagat 2220

agctatgagg acatcagtgc gtatttgctc tccaagaaca acgctatcga gccacggtct 2280

ttcagtcaga atcctcccgt tctgaagcgg catcagcgcg aaataacacg cacaaccctt 2340

cagtcagacc aagaggaaat cgactacgat gatactatct ctgtggagat gaagaaggag 2400

gatttcgaca tttacgacga ggacgagaat cagtccccaa ggagctttca gaagaaaaca 2460

agacactatt tcattgccgc cgtggagcga ctgtgggact acggcatgtc tagctctccg 2520

catgtactta gaaatagggc acaaagcgga tccgtgcctc agtttaagaa agttgtcttt 2580

caggagttta cagatggctc cttcacccag cccttgtatc gcggggaact caatgaacac 2640

ctgggcctcc tgggtcctta tattagggcc gaagtcgagg acaatatcat ggtgaccttt 2700

aggaaccagg catctagacc ttactctttc tactcctccc tgatatccta tgaggaggac 2760

cagcggcaag gcgctgagcc tcggaagaac tttgtgaagc caaatgaaac caaaacatac 2820

ttttggaaag ttcagcacca catggctccc acgaaggacg aatttgactg taaagcctgg 2880

gcctacttct cagatgtaga tctcgagaaa gacgtgcact cagggctcat tggtcccctc 2940

ctggtctgtc atactaatac cctcaatcca gcacacggac gtcaggtaac cgtccaggaa 3000

tttgccctgt tctttaccat tttcgatgag actaaatcct ggtactttac cgaaaacatg 3060

gagaggaatt gcagagcccc atgcaacatc cagatggagg accctacctt caaagagaac 3120

tatcgcttcc atgccattaa cggttacatt atggatactc tcccaggact tgtgatggca 3180

caggatcagc ggataagatg gtatctgttg agcatgggct ccaacgagaa tattcacagc 3240

atccatttct ccggtcacgt gtttacagtg agaaagaaag aagagtacaa gatggctctg 3300

tataatctct atccaggcgt attcgaaacg gtggagatgt tgcctagcaa ggccggcatt 3360

tggcgagtag aatgccttat cggggaacat ctgcatgccg gaatgagcac gctcttcctg 3420

gtgtatagta acaagtgcca gactccgctg ggcatggcat ctggccatat acgggacttt 3480

cagattacgg ctagcgggca gtatgggcag tgggcaccca aacttgcgcg actgcactat 3540

tcaggctcta tcaatgcatg gtccaccaag gaacccttct cttggattaa ggtggacctt 3600

ttggcgccca tgataatcca tgggatcaaa acccagggcg ctcgtcagaa attctcatca 3660

ctctacatct ctcagttcat aataatgtat tcactggatg ggaagaaatg gcagacttac 3720

agaggaaaca gcaccgggac gctgatggtg ttctttggca acgtggacag cagcggcatc 3780

aaacacaaca tcttcaatcc tcccattatt gcccgttata ttagactgca tcccactcac 3840

tactctatac gcagcacact taggatggag ctcatgggat gcgacctgaa cagttgtagt 3900

atgcccttgg ggatggagtc caaagctata agcgacgcac aaattacagc tagctcttac 3960

tttacgaata tgttcgccac gtggagccca agcaaagccc ggctgcattt gcagggtcgg 4020

agtaatgctt ggcgcccaca ggtgaataac cctaaggaat ggttgcaagt agatttccag 4080

aaaactatga aggtaaccgg cgtcactaca cagggagtca agtccctctt gacctctatg 4140

tacgtcaagg agttcctgat tagcagcagt caggatgggc accaatggac actgttcttc 4200

cagaatggga aagttaaagt atttcagggt aaccaggact cctttacacc tgtggtgaat 4260

agcctcgacc caccccctgct gacacgatac ctccgcatcc accctcagtc ttgggtgcat 4320

caaattgccc tgcgaatgga ggtgttggga tgcgaagctc aggacctcta ctga 4374

<210> 16
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 16
atgcagatcg aactctctac ttgcttcttc ctgtgccttc tgaggttctg cttctctgcc 60

actcgccgat attacctcgg ggccgtggag ttgagttggg actacatgca atcagatctg 120

ggcgaactcc ctgtggatgc ccgattccca ccgcgcgtgc ccaagtcttt cccatttaat 180

acttctgtgg tgtacaagaa gacattgttt gtggagttta ccgatcacct gttcaacatc 240

gccaaaccgc ggcccccatg gatgggtctg cttgggccca ccattcaagc ggaggtctat 300

gatacagtgg tgataacgct taagaacatg gcgagccacc cagtgtctct gcatgccgtt 360

ggtgtatcat attggaaggc cagcgaagga gcggagtacg atgaccagac ctctcagaga 420

gagaaggaag acgataaggt ttttcctggc ggaagtcata catatgtatg gcaggtcctg 480

aaagagaatg ggccgatggc ttctgacccc ctttgtctta cctatagtta tctgagccac 540

gtggacctgg tcaaggacct caacagtggt ctgattgggg ctctgcttgt ttgtagagag 600

ggtagcttgg ctaaggagaa aacccaaaca ctccataagt tcattttgct gttcgcggtg 660

ttcgacgagg gaaagagttg gcacagcgaa acaaagaatt cactgatgca agacagggac 720

gccgcttccg caagggcttg gcctaagatg catacggtga atgggtatgt gaaccggagc 780

ctcccggggc tgatcgggtg ccatcgcaag tctgtttact ggcacgtcat tggaatgggg 840

acaacgccag aggtacatag tatatttctt gaaggccaca cgttcctcgt acggaaccac 900

cgacaggctt ccctggagat aagccccatt acctttctga ccgctcagac tctgctgatg 960

gaccttggcc agtttctcct gttctgccat attagcagcc accagcacga cggtatggaa 1020

gcatacgtga aagtcgatag ctgtcctgag gagcctcagc tcagaatgaa gaacaacgag 1080

gaggccgaag actatgacga tgaccttaca gattccgaga tggacgtggt gcgctttgac 1140

gacgataaca gtcctagttt cattcaaatc agatccgtag ccaaaaagca tccaaagaca 1200

tgggtgcatt acattgcagc cgaagaggag gattgggatt atgcgcccct tgttctggct 1260

ccagatgaca ggagctataa gtcccagtac ttgaacaacg ggccacagcg aatcggtaga 1320

aaatataaga aggtaagatt catggcctac actgacgaaa catttaaaac cagggaagct 1380

atccaacacg aatctggaat tctcggccct ctgctctacg gtgaggtggg ggacaccttg 1440

ctgatcattt tcaaaaatca ggcatccagg ccttacaaca tatacccccа tggcatcacc 1500

gatgtccgcc cgctgtattc cagaagactc cccaagggag tgaaacatct gaaagatttt 1560

cccatcctgc cgggcgagat ctttaaatac aaatggactg tgactgtaga ggacgggcct 1620

acaaaatcag acccacggtg cctgacaagg tattacagta gcttcgtcaa catggaacgc 1680

gacctcgcca gcggactcat tggcccactg ttgatctgtt acaaagagtc agtggatcag 1740

aggggaaatc agatcatgag cgataagaga aacgttatcc tgtttagtgt cttcgacgag 1800

aaccggtctt ggtaccttac tgagaacatc cagaggttcc tgccgaatcc ggctggcgtt 1860

cagctcgagg acccagagtt ccaggccagt aatataatgc actcaatcaa cggttatgtg 1920

ttcgatagcc tgcagctgag cgtctgcctc cacgaggtag cctattggta catattgtcc 1980

atcggggctc agaccgattt tctgtccgtg ttctttagcg ggtatacctt taaacataaa 2040

atggtctatg aagacaccct gaccctgttc ccattctccg gtgagactgt gttcatgtcc 2100

atggagaacc cagggctgtg gatcctgggg tgtcacaata gtgactttag gaatcgggga 2160

atgacggcac tgctgaaggt gagttcttgc gataaaaata caggagatta ctatgaggat 2220

agttacgagg atatcagtgc ctatctgctt tcaaaaaaca acgcaattga gccccggtct 2280

ttctcacaaa acccccccggt gctgaagcgc caccagcgcg aaattacccg gacaaccttg 2340

cagtccgacc aggaggaaat cgattatgac gatactatca gtgtagaaat gaaaaaggag 2400

gattttgata tttacgacga agacgagaac cagtctccgc gaagttttca gaagaaaacg 2460

cgacactact ttatagctgc cgtggaacga ctctgggatt atggcatgtc ctccagccct 2520

catgtcctta ggaatcgagc gcagagtggc tctgtgcctc agttcaaaaa ggttgtgttc 2580

caggaattca ccgacggctc atttacccag ccgctgtaca gaggcgaact caacgaacac 2640

cttgggctgc ttgggccata tattcgagca gaggtggaag ataatatcat ggtaaccttt 2700

agaaaccagg cgtcaagacc ctattccttc tacagttctc tgatcagcta cgaggaggac 2760

caaagacagg gagctgaacc caggaagaac tttgtgaaac ctaatgagac caagacctac 2820

ttctggaagg tccagcacca tatggcccca actaaagatg aattcgattg caaggcctgg 2880

gcttatttca gcgacgtgga tctcgaaaag gatgtgcaca gcgggttgat cggaccgctt 2940

ttggtgtgcc acacaaatac cctcaatcct gcccacgggc ggcaggtcac agttcaagag 3000

tttgcactct tctttacaat atttgacgag acaaagtcat ggtattttac agagaatatg 3060

gagagaaatt gtcgcgcacc ttgcaacatt cagatggagg accccacatt taaggagaat 3120

tacagatttc atgctatcaa tgggtacatt atggatactc tgcctggtct ggtcatggcc 3180

caggatcagc gcataaggtg gtacttgctg agcatgggat ctaatgagaa tatacacagc 3240

attcacttca gtggccacgt ttttactgtt agaaagaagg aggagtacaa aatggcgctc 3300

tacaaccttt acccgggtgt gtttgagaca gtggagatgc tgccaagcaa ggcaggcatc 3360

tggagggttg agtgtcttat tggggagcat ctgcatgctg gaatgtccac cctctttctt 3420

gtgtacagca ataagtgcca gacaccgctt ggcatggcca gcggccacat tagggacttt 3480

cagataactg ccagtggaca gtacggccag tgggctccca agcttgcaag actccactac 3540

tccggaagca taaacgcatg gagcaccaag gaacccttct cttggattaa ggtggacctg 3600

ctggcgccaa tgatcattca cggcataaaa acccaagggg cacgacagaa attttcatct 3660

ttgtatatta gtcagtttat catcatgtac agcttggatg gaaagaagtg gcagacgtac 3720

aggggcaatt ctacaggaac acttatggtg ttttttggga atgtcgattc cagcgggatc 3780

aaacataaca tcttcaatcc tcctattatc gcccgatata tccgcctgca ccctacgcat 3840

tactccatca ggtccacatt gagaatggaa ctgatggggt gcgacctgaa tagttgtagt 3900

atgccactgg gcatggagtc taaagccatc agcgatgcac agatcactgc cagctcttac 3960

ttcaccaaca tgtttgcaac ttggtccccc tctaaagctc gcctgcatct gcagggacgc 4020

tcaaatgcat ggcgaccaca ggtgaacaat ccaaaagagt ggctccaggt cgactttcag 4080

aagacaatga aggtaacagg agtgacaacc cagggtgtaa aaagcctcct tacgagtatg 4140

tacgttaagg agtttctgat ttctagctcc caggacggac accagtggac tctgttcttc 4200

cagaacggca aagtgaaggt atttcaggga aaccaggatt cttttacccc ggtagtgaat 4260

agcctggatc caccgttgct gacccgctat ctgagaattc atccacaatc ctgggtgcat 4320

cagattgccc tccggatgga agtgctcggc tgtgaagctc aggatctgta ttag 4374

<210> 17
<211> 4374
<212> ADN

<213> Trình tự nhân tạo

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

<400> 17
atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc 60

accagaagat actacctggg tgcagtggaa ctgtcatggg actatatgca aagtgatctc 120

ggtgagctgc ctgtggacgc aagatttcct cctagagtgc caaaatcttt tccattcaac 180

acctcagtcg tgtacaaaaa gactctgttt gtagaattca cggatcacct tttcaacatc 240

gctaagccaa ggccaccctg gatgggtctg ctaggtccta ccatccaggc tgaggtttat 300

gatacagtgg tcattacact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt 360

ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg 420

gagaaagaag atgataaagt cttccctggt ggaagccata catatgtctg gcaggtcctg 480

aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat 540

gtggacctgg taaaagactt gaattcaggc ctcattggag ccctactagt atgtagagaa 600

gggagtctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttgctgta 660

tttgatgaag ggaaaagttg gcactcagaa acaaagaact ccttgatgca ggatagggat 720

gctgcatctg ctcgggcctg gcctaaaatg cacacagtca atggttatgt aaacaggtct 780

ctgccaggtc tgattggatg ccacaggaaa tcagtctatt ggcatgtgat tggaatgggc 840

accactcctg aagtgcactc aatattcctc gaaggtcaca catttcttgt gaggaaccat 900

cgccaggcgt ccttggaaat ctcgccaata actttcctta ctgctcaaac actcttgatg 960

gaccttggac agtttctact gttttgtcat atctcttccc accaacatga tggcatggaa 1020

gcttatgtca aagtagacag ctgtccagag gaaccccaac tacgaatgaa aaataatgaa 1080

gaagcggaag actatgatga tgatcttact gattctgaaa tggatgtggt caggtttgat 1140

gatgacaact ctccttcctt tatccaaatt cgctcagttg ccaagaagca tcctaaaact 1200

tgggtacatt acattgctgc tgaagaggag gactgggact atgctccctt agtcctcgcc 1260

cccgatgaca gaagttataa aagtcaatat ttgaacaatg gccctcagcg gattggtagg 1320

aagtacaaaa aagtccgatt tatggcatac acagatgaaa cctttaagac tcgtgaagct 1380

attcagcatg aatcaggaat cttgggacct ttactttatg gggaagttgg agacacactg 1440

ttgattatat ttaagaatca agcaagcaga ccatataaca tctaccctca cggaatcact 1500

gatgtccgtc ctttgtattc aaggagatta ccaaaaggtg taaaacattt gaaggatttt 1560

ccaattctgc caggagaaat attcaaatat aaatggacag tgactgtaga agatgggcca 1620

actaaatcag atcctcggtg cctgacccgc tattactcta gtttcgttaa tatggagaga 1680

gatctagctt caggactcat tggccctctc ctcatctgct acaaagaatc tgtagatcaa 1740

agaggaaacc agataatgtc agacaagagg aatgtcatcc tgttttctgt atttgatgag 1800

aaccgaagct ggtacctcac agagaatata caacgctttc tccccaatcc agctggagtg 1860

cagcttgagg atccagagtt ccaagcctcc aacatcatgc acagcatcaa tggctatgtt 1920

tttgatagtt tgcagttgtc agtttgtttg catgaggtgg catactggta cattctaagc 1980

attggagcac agactgactt cctttctgtc ttcttctctg gatatacctt caaacacaaa 2040

atggtctatg aagacacact caccctattc ccattctcag gagaaactgt cttcatgtcg 2100

atggaaaacc caggtctatg gattctgggg tgccacaact cagactttcg gaacagaggc 2160

atgaccgcct tactgaaggt ttctagttgt gacaagaaca ctggtgatta ttacgaggac 2220

agttatgaag atatttcagc atacttgctg agtaaaaaca atgccattga accaagaagc 2280

ttctcccaga atccaccagt cttgaaacgc catcaacggg aaataactcg tactactctt 2340

cagtcagatc aagaggaaat tgactatgat gataccatat cagttgaaat gaagaaggaa 2400

gattttgaca tttatgatga ggatgaaaat cagagccccc gcagctttca aaagaaaaca 2460

cgacactatt ttattgctgc agtggagagg ctctgggatt atgggatgag tagctcccca 2520

catgttctaa gaaacagggc tcagagtggc agtgtccctc agttcaagaa agttgttttc 2580

caggaattta ctgatggctc ctttactcag cccttatacc gtggagaact aaatgaacat 2640

ttgggactcc tggggccata tataagagca gaagttgaag ataatatcat ggtaactttc 2700

agaaatcagg cctctcgtcc ctattccttc tattctagcc ttatttctta tgaggaagat 2760

cagaggcaag gagcagaacc tagaaaaaac tttgtcaagc ctaatgaaac caaaacttac 2820

ttttggaaag tgcaacatca tatggcaccc actaaagatg agtttgactg caaagcctgg 2880

gcttatttct ctgatgttga cctggaaaaa gatgtgcact caggcctgat tggacccctt 2940

ctggtctgcc acactaacac actgaaccct gctcatggga gacaagtgac agtacaggaa 3000

tttgctctgt ttttcaccat ctttgatgag accaaaagct ggtacttcac tgaaaatatg 3060

gaaagaaact gcagggctcc ctgcaatatc cagatggaag atcccacttt taaagagaat 3120

tatcgcttcc atgcaatcaa tggctacata atggatacac tacctggctt agtaatggct 3180

caggatcaaa ggattcgatg gtatctgctc agcatgggca gcaatgaaaa catccattct 3240

attcatttca gtggacatgt gttcactgta cgaaaaaaag aggagtataa aatggcactg 3300

tacaatctct atccaggtgt ttttgagaca gtggaaatgt taccatccaa agctggaatt 3360

tggcgggtgg aatgccttat tggcgagcat ctacatgctg ggatgagcac acttttctg 3420

gtgtacagca ataagtgtca gactcccctg ggaatggctt ctggacacat tagagatttt 3480

cagattacag cttcaggaca atatggacag tgggccccaa agctggccag acttcattat 3540

tccggatcaa tcaatgcctg gagcaccaag gagccctttt cttggatcaa ggtggatctg 3600

ttggcaccaa tgattattca cggcatcaag acccagggtg cccgtcagaa gttctccagc 3660

ctctacatct ctcagtttat catcatgtat agtcttgatg ggaagaagtg gcagacttat 3720

cgaggaaatt ccactggaac cttaatggtc ttctttggca atgtggattc atctgggata 3780

aaacacaata tttttaaccc tccaattatt gctcgataca tccgtttgca cccaactcat 3840

tatagcattc gcagcactct tcgcatggag ttgatgggct gtgatttaaa tagttgcagc 3900

atgccattgg gaatggagag taaagcaata tcagatgcac agattactgc ttcatcctac 3960

tttaccaata tgtttgccac ctggtctcct tcaaaagctc gacttcacct ccaagggagg 4020

agtaatgcct ggagacctca ggtgaataat ccaaaagagt ggctgcaagt ggacttccag 4080

aagacaatga aagtcacagg agtaactact cagggagtaa aatctctgct taccagcatg 4140

tatgtgaagg agttcctcat ctccagcagt caagatggcc atcagtggac tctcttttttt 4200

cagaatggca aagtaaaggt ttttcaggga aatcaagact ccttcacacc tgtggtgaac 4260

tctctagacc caccgttact gactcgctac cttcgaattc acccccagag ttgggtgcac 4320

cagattgccc tgaggatgga ggttctgggc tgcgaggcac aggacctcta ctga 4374

<210> 18
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 18
atgcagatcg agctgtccac atgctttttt ctgtgcctgc tgcggttctg cttcagcgcc 60

acccggcggt actacctggg cgccgtggag ctgtcctggg actacatgca gagcgacctg 120

ggcgagctgc ccgtggacgc ccggttcccc cccagagtgc ccaagagctt ccccttcaac 180

accagcgtgg tgtacaagaa aaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctccgagggc gccgagtacg acgaccagac cagccagcgg 420

gagaaagagg acgacaaagt ctttcctggc ggcagccaca cctacgtgtg gcaggtcctg 480

aaagaaaacg gccccatggc ctccgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaacagcggg ctgattgggg ccctgctggt ctgccgggag 600

ggcagcctgg ccaaagagaa aacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgacgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggaccgggac 720

gccgcctctg ccagagcctg gcccaagatg cacaccgtga acggctacgt gaacagaagc 780

ctgcccggcc tgattggctg ccaccggaag agcgtgtact ggcacgtgat cggcatgggc 840

accacacccg aggtgcacag catctttctg gaagggcaca cctttctggt gcggaaccac 900

cggcaggcca gcctggaaat cagccctatc accttcctga ccgcccagac actgctgatg 960

gacctgggcc agttcctgct gttttgccac atcagctctc accagcacga cggcatggaa 1020

gcctacgtga aggtggactc ctgccccgag gaaccccagc tgcggatgaa gaacaacgag 1080

gaagccgagg actacgacga cgacctgacc gacagcgaga tggacgtggt gcggttcgac 1140

gacgacaaca gccccagctt catccagatc agaagcgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaagag gactgggact acgccccccct ggtgctggcc 1260

cccgacgaca gaagctacaa gagccagtac ctgaacaatg gcccccagcg gatcggccgg 1320

aagtacaaga aagtgcggtt catggcctac accgacgaga ccttcaagac ccgggaggcc 1380

atccagcacg agagcggcat cctgggcccc ctgctgtacg gcgaagtggg cgacacactg 1440

ctgatcatct tcaagaacca ggccagccgg ccctacaaca tctaccccca cggcatcacc 1500

gacgtgcggc ccctgtacag caggcggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggacggcccc 1620

accaagagcg accccagatg cctgacccgg tactacagca gcttcgtgaa catggaacgg 1680

gacctggcct ccgggctgat cggacctctg ctgatctgct acaaagaaag cgtggaccag 1740

cggggcaacc agatcatgag cgacaagcgg aacgtgatcc tgttcagcgt gttcgatgag 1800

aaccggtcct ggtatctgac cgagaacatc cagcggtttc tgcccaaccc tgccggggtg 1860

cagctggaag atcccgagtt ccaggccagc aacatcatgc actccatcaa tggctacgtg 1920

ttcgacagcc tgcagctgtc cgtgtgtctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgagcgtg ttcttcagcg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cctttcagcg gcgagaccgt gttcatgagc 2100

atggaaaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttccg gaaccggggc 2160

atgaccgccc tgctgaaggt gtccagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg atatcagcgc ctacctgctg tccaagaaca acgccatcga gcccagaagc 2280

ttcagccaga acccccctgt gctgaagcgg caccagagag agatcacccg gaccaccctg 2340

cagtccgacc aggaagagat cgattacgac gacaccatca gcgtggagat gaaaaaagaa 2400

gatttcgaca tctacgacga ggacgagaac cagagccccc ggtccttcca gaagaaaacc 2460

cggcactact ttatcgccgc cgtggagcgg ctgtgggact acggcatgag cagcagcccc 2520

cacgtgctgc ggaaccgggc ccagagcggc agcgtgcccc agttcaagaa agtggtgttc 2580

caggaattca ccgacggcag cttcacccag cccctgtacc ggggcgagct gaacgagcac 2640

ctggggctgc tggggcccta catcagggcc gaagtggagg acaacatcat ggtgaccttc 2700

cggaatcagg ccagcagacc ctactccttc tacagcagcc tgatcagcta cgaagaggac 2760

cagcggcagg gcgctgaacc ccggaagaac ttcgtgaagc ccaatgagac caagacctac 2820

ttctggaaag tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880

gcctacttca gcgacgtgga tctggaaaag gacgtgcact ctggactgat tggccctctg 2940

ctggtgtgcc acaccaacac cctgaacccc gcccacggcc ggcaggtgac cgtgcaggaa 3000

ttcgccctgt tcttcaccat cttcgacgag accaagtcct ggtacttcac cgagaatatg 3060

gaacggaact gcagagcccc ctgcaacatc cagatggaag atcctacctt caaagagaac 3120

taccggttcc acgccatcaa cggctacatc atggacaccc tgcctggcct ggtgatggcc 3180

caggaccaga ggatccggtg gtatctgctg tccatgggca gcaacgagaa tatccacagc 3240

atccacttca gcggccacgt gttcaccgtg aggaagaaag aagagtacaa gatggccctg 3300

tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc 3360

tggcgggtgg agtgtctgat cggcgagcac ctgcatgccg ggatgagcac cctgtttctg 3420

gtgtacagca acaagtgcca gacccccctg ggcatggcca gcggccacat ccgggacttc 3480

cagatcaccg cctccggcca gtacggccag tgggccccca agctggcccg gctgcactac 3540

agcggcagca tcaacgcctg gtccaccaaa gagcccttca gctggatcaa ggtggacctg 3600

ctggccccta tgatcatcca cggcattaag acccagggcg ccaggcagaa gttcagcagc 3660

ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720

cggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780

aagcacaaca tcttcaaccc ccccatcatc gcccggtaca tccggctgca ccccacccac 3840

tacagcatca gatccaccct gcggatggaa ctgatgggct gcgacctgaa ctcctgcagc 3900

atgcctctgg gcatggaaag caaggccatc agcgacgccc agatcacagc cagcagctac 3960

ttcaccaaca tgttcgccac ctggtccccc tccaaggcca ggctgcacct gcagggccgg 4020

tccaacgcct ggcggcctca ggtgaacaac cccaaagaat ggctgcaggt ggactttcag 4080

aaaaccatga aggtgaccgg cgtgaccacc cagggcgtga aaagcctgct gaccagcatg 4140

tacgtgaaag agtttctgat cagcagcagc caggacggcc accagtggac cctgttcttt 4200

cagaacggca aggtgaaagt gttccagggc aaccaggact ccttcacccc cgtggtgaac 4260

tccctggacc ccccccctgct gacccgctac ctgcggatcc acccccagtc ttgggtgcac 4320

cagatcgccc tgaggatgga agtgctggga tgtgaggccc aggatctgta ctga 4374

<210> 19
<211> 2351
<212> PRT
<213> Người hiện đại

<400> 19

```
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1               5                  10                  15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
            20                  25                  30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
        35                  40                  45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
    50                  55                  60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65                  70                  75                  80

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

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
```

195 200 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
210 215 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225 230 235 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
245 250 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
260 265 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
275 280 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
290 295 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305 310 315 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
325 330 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
370 375 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385 390 395 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
405 410 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
420 425 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
435 440 445

```
Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
            500                 505                 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
        515                 520                 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
    530                 535                 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545                 550                 555                 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
                565                 570                 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
            580                 585                 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
        595                 600                 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
    610                 615                 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625                 630                 635                 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
                645                 650                 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
            660                 665                 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
        675                 680                 685
```

```
Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
    690                 695                 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705                 710                 715                 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
                725                 730                 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
            740                 745                 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ser Arg His Pro
        755                 760                 765

Ser Thr Arg Gln Lys Gln Phe Asn Ala Thr Thr Ile Pro Glu Asn Asp
    770                 775                 780

Ile Glu Lys Thr Asp Pro Trp Phe Ala His Arg Thr Pro Met Pro Lys
785                 790                 795                 800

Ile Gln Asn Val Ser Ser Ser Asp Leu Leu Met Leu Leu Arg Gln Ser
                805                 810                 815

Pro Thr Pro His Gly Leu Ser Leu Ser Asp Leu Gln Glu Ala Lys Tyr
            820                 825                 830

Glu Thr Phe Ser Asp Asp Pro Ser Pro Gly Ala Ile Asp Ser Asn Asn
        835                 840                 845

Ser Leu Ser Glu Met Thr His Phe Arg Pro Gln Leu His His Ser Gly
    850                 855                 860

Asp Met Val Phe Thr Pro Glu Ser Gly Leu Gln Leu Arg Leu Asn Glu
865                 870                 875                 880

Lys Leu Gly Thr Thr Ala Ala Thr Glu Leu Lys Lys Leu Asp Phe Lys
                885                 890                 895

Val Ser Ser Thr Ser Asn Asn Leu Ile Ser Thr Ile Pro Ser Asp Asn
            900                 905                 910

Leu Ala Ala Gly Thr Asp Asn Thr Ser Ser Leu Gly Pro Pro Ser Met
        915                 920                 925
```

```
Pro Val His Tyr Asp Ser Gln Leu Asp Thr Thr Leu Phe Gly Lys Lys
    930                 935                 940

Ser Ser Pro Leu Thr Glu Ser Gly Gly Pro Leu Ser Leu Ser Glu Glu
945                 950                 955                 960

Asn Asn Asp Ser Lys Leu Leu Glu Ser Gly Leu Met Asn Ser Gln Glu
                965                 970                 975

Ser Ser Trp Gly Lys Asn Val Ser Ser Thr Glu Ser Gly Arg Leu Phe
            980                 985                 990

Lys Gly Lys Arg Ala His Gly Pro  Ala Leu Leu Thr Lys  Asp Asn Ala
        995                 1000                1005

Leu Phe  Lys Val Ser Ile Ser  Leu Leu Lys Thr Asn  Lys Thr Ser
    1010                1015                1020

Asn Asn  Ser Ala Thr Asn Arg  Lys Thr His Ile Asp  Gly Pro Ser
    1025                1030                1035

Leu Leu  Ile Glu Asn Ser Pro  Ser Val Trp Gln Asn  Ile Leu Glu
    1040                1045                1050

Ser Asp  Thr Glu Phe Lys Lys  Val Thr Pro Leu Ile  His Asp Arg
    1055                1060                1065

Met Leu  Met Asp Lys Asn Ala  Thr Ala Leu Arg Leu  Asn His Met
    1070                1075                1080

Ser Asn  Lys Thr Thr Ser Ser  Lys Asn Met Glu Met  Val Gln Gln
    1085                1090                1095

Lys Lys  Glu Gly Pro Ile Pro  Pro Asp Ala Gln Asn  Pro Asp Met
    1100                1105                1110

Ser Phe  Phe Lys Met Leu Phe  Leu Pro Glu Ser Ala  Arg Trp Ile
    1115                1120                1125

Gln Arg  Thr His Gly Lys Asn  Ser Leu Asn Ser Gly  Gln Gly Pro
    1130                1135                1140

Ser Pro  Lys Gln Leu Val Ser  Leu Gly Pro Glu Lys  Ser Val Glu
    1145                1150                1155

Gly Gln  Asn Phe Leu Ser Glu  Lys Asn Lys Val Val  Val Gly Lys
```

1160 1165 1170

Gly Glu Phe Thr Lys Asp Val Gly Leu Lys Glu Met Val Phe Pro
1175 1180 1185

Ser Ser Arg Asn Leu Phe Leu Thr Asn Leu Asp Asn Leu His Glu
1190 1195 1200

Asn Asn Thr His Asn Gln Glu Lys Lys Ile Gln Glu Glu Ile Glu
1205 1210 1215

Lys Lys Glu Thr Leu Ile Gln Glu Asn Val Val Leu Pro Gln Ile
1220 1225 1230

His Thr Val Thr Gly Thr Lys Asn Phe Met Lys Asn Leu Phe Leu
1235 1240 1245

Leu Ser Thr Arg Gln Asn Val Glu Gly Ser Tyr Asp Gly Ala Tyr
1250 1255 1260

Ala Pro Val Leu Gln Asp Phe Arg Ser Leu Asn Asp Ser Thr Asn
1265 1270 1275

Arg Thr Lys Lys His Thr Ala His Phe Ser Lys Lys Gly Glu Glu
1280 1285 1290

Glu Asn Leu Glu Gly Leu Gly Asn Gln Thr Lys Gln Ile Val Glu
1295 1300 1305

Lys Tyr Ala Cys Thr Thr Arg Ile Ser Pro Asn Thr Ser Gln Gln
1310 1315 1320

Asn Phe Val Thr Gln Arg Ser Lys Arg Ala Leu Lys Gln Phe Arg
1325 1330 1335

Leu Pro Leu Glu Glu Thr Glu Leu Glu Lys Arg Ile Ile Val Asp
1340 1345 1350

Asp Thr Ser Thr Gln Trp Ser Lys Asn Met Lys His Leu Thr Pro
1355 1360 1365

Ser Thr Leu Thr Gln Ile Asp Tyr Asn Glu Lys Glu Lys Gly Ala
1370 1375 1380

Ile Thr Gln Ser Pro Leu Ser Asp Cys Leu Thr Arg Ser His Ser
1385 1390 1395

Ile Pro Gln Ala Asn Arg Ser Pro Leu Pro Ile Ala Lys Val Ser
1400 1405 1410

Ser Phe Pro Ser Ile Arg Pro Ile Tyr Leu Thr Arg Val Leu Phe
1415 1420 1425

Gln Asp Asn Ser Ser His Leu Pro Ala Ala Ser Tyr Arg Lys Lys
1430 1435 1440

Asp Ser Gly Val Gln Glu Ser Ser His Phe Leu Gln Gly Ala Lys
1445 1450 1455

Lys Asn Asn Leu Ser Leu Ala Ile Leu Thr Leu Glu Met Thr Gly
1460 1465 1470

Asp Gln Arg Glu Val Gly Ser Leu Gly Thr Ser Ala Thr Asn Ser
1475 1480 1485

Val Thr Tyr Lys Lys Val Glu Asn Thr Val Leu Pro Lys Pro Asp
1490 1495 1500

Leu Pro Lys Thr Ser Gly Lys Val Glu Leu Leu Pro Lys Val His
1505 1510 1515

Ile Tyr Gln Lys Asp Leu Phe Pro Thr Glu Thr Ser Asn Gly Ser
1520 1525 1530

Pro Gly His Leu Asp Leu Val Glu Gly Ser Leu Leu Gln Gly Thr
1535 1540 1545

Glu Gly Ala Ile Lys Trp Asn Glu Ala Asn Arg Pro Gly Lys Val
1550 1555 1560

Pro Phe Leu Arg Val Ala Thr Glu Ser Ser Ala Lys Thr Pro Ser
1565 1570 1575

Lys Leu Leu Asp Pro Leu Ala Trp Asp Asn His Tyr Gly Thr Gln
1580 1585 1590

Ile Pro Lys Glu Glu Trp Lys Ser Gln Glu Lys Ser Pro Glu Lys
1595 1600 1605

Thr Ala Phe Lys Lys Lys Asp Thr Ile Leu Ser Leu Asn Ala Cys
1610 1615 1620

Glu Ser Asn His Ala Ile Ala Ala Ile Asn Glu Gly Gln Asn Lys
1625 1630 1635

Pro Glu Ile Glu Val Thr Trp Ala Lys Gln Gly Arg Thr Glu Arg
1640 1645 1650

Leu Cys Ser Gln Asn Pro Pro Val Leu Lys Arg His Gln Arg Glu
1655 1660 1665

Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile Asp Tyr
1670 1675 1680

Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp Ile
1685 1690 1695

Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys
1700 1705 1710

Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr
1715 1720 1725

Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser
1730 1735 1740

Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
1745 1750 1755

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu
1760 1765 1770

His Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp
1775 1780 1785

Asn Ile Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser
1790 1795 1800

Phe Tyr Ser Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly
1805 1810 1815

Ala Glu Pro Arg Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr
1820 1825 1830

Tyr Phe Trp Lys Val Gln His His Met Ala Pro Thr Lys Asp Glu
1835 1840 1845

Phe Asp Cys Lys Ala Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu
1850 1855 1860

Lys Asp Val His Ser Gly Leu Ile Gly Pro Leu Leu Val Cys His
1865 1870 1875

Thr Asn Thr Leu Asn Pro Ala His Gly Arg Gln Val Thr Val Gln
1880 1885 1890

Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr Lys Ser Trp
1895 1900 1905

Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg Ala Pro Cys Asn
1910 1915 1920

Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg Phe His
1925 1930 1935

Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu Val Met
1940 1945 1950

Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met Gly Ser
1955 1960 1965

Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val Phe Thr
1970 1975 1980

Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn Leu Tyr
1985 1990 1995

Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys Ala Gly
2000 2005 2010

Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His Leu His Ala Gly
2015 2020 2025

Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro
2030 2035 2040

Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile Thr Ala
2045 2050 2055

Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg Leu His
2060 2065 2070

Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro Phe Ser

2075 2080 2085

Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His Gly Ile
2090 2095 2100

Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr Ile Ser
2105 2110 2115

Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp Gln Thr
2120 2125 2130

Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe Gly Asn
2135 2140 2145

Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro Pro Ile
2150 2155 2160

Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg
2165 2170 2175

Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn Ser Cys
2180 2185 2190

Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp Ala Gln
2195 2200 2205

Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr Trp Ser
2210 2215 2220

Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn Ala Trp
2225 2230 2235

Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val Asp Phe
2240 2245 2250

Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly Val Lys
2255 2260 2265

Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile Ser Ser
2270 2275 2280

Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn Gly Lys
2285 2290 2295

Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro Val Val
2300 2305 2310

Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg Ile His
2315 2320 2325

Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu Val Leu
2330 2335 2340

Gly Cys Glu Ala Gln Asp Leu Tyr
2345 2350

<210> 20
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 20
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccaggagc 2280

ttcagccaga accccccccgt gctgaagagg caccagaggg agatcaccag gaccaccctg 2340

cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400

gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460

aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc 2520

cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca gaggcgagct gaacgagcac 2640

ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700

aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac 2760

cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac 2820

ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cggcccctg 2940

ctggtgtgcc acaccaacac cctgaacccc gcccacggca ggcaggtgac cgtgcaggag 3000

ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaacatg 3060

gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120

tacaggttcc acgccatcaa cggctacatc atggacaccc tgcccggcct ggtgatggcc 3180

caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240

atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300

tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc 3360

tggagggtgg agtgcctgat cggcgagcac ctgcacgccg gcatgagcac cctgttcctg 3420

gtgtacagca acaagtgcca gacccccctg ggcatggcca gcggccacat cagggacttc 3480

cagatcaccg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540

agcggcagca tcaacgcctg gagcaccaag gagcccttca gctggatcaa ggtggacctg 3600

ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc 3660

ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720

aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780

aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccacccac 3840

tacagcatca ggagcaccct gcggatggaa ctgatgggct gcgacctgaa cagctgcagc 3900

atgcccctgg gcatggagag caaggccatc tctgacgccc agatcaccgc cagcagctac 3960

ttcaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020

agcaacgcct ggaggcccca ggtgaacaac cccaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg 4140

tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctgttcttc 4200

cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaac 4260

agcctggacc ccccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320

cagatcgccc tgagaatgga agtgctggga tgcgaggccc aggacctgta ctga 4374

<210> 21
<211> 1457
<212> PRT
<213> Trình tự nhân tạo

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

<400> 21
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

```
Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
            20                  25                  30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
        35                  40                  45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
    50                  55                  60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65                  70                  75                  80

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

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255
```

```
Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465                 470                 475                 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
                485                 490                 495
```

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys

740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
755 760 765

Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
770 775 780

Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
785 790 795 800

Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
805 810 815

Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
820 825 830

Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
835 840 845

Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
850 855 860

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
865 870 875 880

Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
885 890 895

Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
900 905 910

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
915 920 925

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
930 935 940

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
980 985 990

```
Gly Arg Gln Val Thr Val Gln Glu  Phe Ala Leu Phe Phe  Thr Ile Phe
       995                1000                1005

Asp Glu  Thr Lys Ser Trp Tyr  Phe Thr Glu Asn Met  Glu Arg Asn
    1010                1015                1020

Cys Arg  Ala Pro Cys Asn Ile  Gln Met Glu Asp Pro  Thr Phe Lys
    1025                1030                1035

Glu Asn  Tyr Arg Phe His Ala  Ile Asn Gly Tyr Ile  Met Asp Thr
    1040                1045                1050

Leu Pro  Gly Leu Val Met Ala  Gln Asp Gln Arg Ile  Arg Trp Tyr
    1055                1060                1065

Leu Leu  Ser Met Gly Ser Asn  Glu Asn Ile His Ser  Ile His Phe
    1070                1075                1080

Ser Gly  His Val Phe Thr Val  Arg Lys Lys Glu Glu  Tyr Lys Met
    1085                1090                1095

Ala Leu  Tyr Asn Leu Tyr Pro  Gly Val Phe Glu Thr  Val Glu Met
    1100                1105                1110

Leu Pro  Ser Lys Ala Gly Ile  Trp Arg Val Glu Cys  Leu Ile Gly
    1115                1120                1125

Glu His  Leu His Ala Gly Met  Ser Thr Leu Phe Leu  Val Tyr Ser
    1130                1135                1140

Asn Lys  Cys Gln Thr Pro Leu  Gly Met Ala Ser Gly  His Ile Arg
    1145                1150                1155

Asp Phe  Gln Ile Thr Ala Ser  Gly Gln Tyr Gly Gln  Trp Ala Pro
    1160                1165                1170

Lys Leu  Ala Arg Leu His Tyr  Ser Gly Ser Ile Asn  Ala Trp Ser
    1175                1180                1185

Thr Lys  Glu Pro Phe Ser Trp  Ile Lys Val Asp Leu  Leu Ala Pro
    1190                1195                1200

Met Ile  Ile His Gly Ile Lys  Thr Gln Gly Ala Arg  Gln Lys Phe
    1205                1210                1215
```

```
Ser Ser  Leu Tyr Ile Ser Gln  Phe Ile Ile Met Tyr  Ser Leu Asp
    1220                 1225                 1230

Gly Lys  Lys Trp Gln Thr Tyr  Arg Gly Asn Ser Thr  Gly Thr Leu
    1235                 1240                 1245

Met Val  Phe Phe Gly Asn Val  Asp Ser Ser Gly Ile  Lys His Asn
    1250                 1255                 1260

Ile Phe  Asn Pro Pro Ile Ile  Ala Arg Tyr Ile Arg  Leu His Pro
    1265                 1270                 1275

Thr His  Tyr Ser Ile Arg Ser  Thr Leu Arg Met Glu  Leu Met Gly
    1280                 1285                 1290

Cys Asp  Leu Asn Ser Cys Ser  Met Pro Leu Gly Met  Glu Ser Lys
    1295                 1300                 1305

Ala Ile  Ser Asp Ala Gln Ile  Thr Ala Ser Ser Tyr  Phe Thr Asn
    1310                 1315                 1320

Met Phe  Ala Thr Trp Ser Pro  Ser Lys Ala Arg Leu  His Leu Gln
    1325                 1330                 1335

Gly Arg  Ser Asn Ala Trp Arg  Pro Gln Val Asn Asn  Pro Lys Glu
    1340                 1345                 1350

Trp Leu  Gln Val Asp Phe Gln  Lys Thr Met Lys Val  Thr Gly Val
    1355                 1360                 1365

Thr Thr  Gln Gly Val Lys Ser  Leu Leu Thr Ser Met  Tyr Val Lys
    1370                 1375                 1380

Glu Phe  Leu Ile Ser Ser Ser  Gln Asp Gly His Gln  Trp Thr Leu
    1385                 1390                 1395

Phe Phe  Gln Asn Gly Lys Val  Lys Val Phe Gln Gly  Asn Gln Asp
    1400                 1405                 1410

Ser Phe  Thr Pro Val Val Asn  Ser Leu Asp Pro Pro  Leu Leu Thr
    1415                 1420                 1425

Arg Tyr  Leu Arg Ile His Pro  Gln Ser Trp Val His  Gln Ile Ala
    1430                 1435                 1440
```

Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 22
<211> 2220
<212> ADN
<213> Trình tự nhân tạo

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

<400> 22
gccaccagga gatactacct gggcgccgtg gagctgagct gggactacat gcagtctgac 60

ctgggcgagc tgcctgtgga cgccaggttc ccccccagag tgcccaagag cttccccttc 120

aacacctcag tggtgtacaa gaagaccctg ttcgtggagt tcaccgacca cctgttcaac 180

atcgccaagc ccaggccccc ctggatgggc ctgctgggcc ccaccatcca ggccgaggtg 240

tacgacaccg tggtgatcac cctgaagaac atggccagcc accccgtgag cctgcacgcc 300

gtgggcgtga gctactggaa ggcctctgag ggcgccgagt atgacgacca gaccagccag 360

agggagaagg aggacgacaa ggtgttcccc ggcggcagcc acacctacgt gtggcaggtg 420

ctgaaggaga acggccccat ggccagcgac cccctgtgcc tgacctacag ctacctgagc 480

cacgtggacc tggtgaagga cctgaactct ggcctgatcg gcgccctgct ggtgtgcagg 540

gagggcagcc tggccaagga gaagacccag accctgcaca agttcatcct gctgttcgcc 600

gtgttcgatg agggcaagag ctggcacagc gagaccaaga acagcctgat gcaggacagg 660

gatgccgcct ctgccagggc ctggcccaag atgcacaccg tgaacggcta cgtgaacagg 720

agcctgcccg gcctgatcgg ctgccacagg aagtctgtgt actggcacgt gatcggcatg 780

ggcaccaccc ccgaggtgca cagcatcttc ctggagggcc acaccttcct ggtgaggaac 840

cacaggcagg ccagcctgga gatcagcccc atcaccttcc tgaccgccca gaccctgctg 900

atggacctgg gccagttcct gctgttctgc cacatcagca gccaccagca cgacggcatg 960

gaggcctacg tgaaggtgga cagctgcccc gaggagcccc agctgaggat gaagaacaac 1020

gaggaggccg aggactatga tgatgacctg accgactctg agatggacgt ggtgaggttt 1080

gatgatgaca acagccccag cttcatccag atcaggtctg tggccaagaa gcaccccaag 1140

acctgggtgc actacatcgc cgccgaggag gaggactggg actacgcccc cctggtgctg 1200

gcccccgacg acaggagcta caagagccag tacctgaaca acggccccca gaggatcggc 1260

aggaagtaca agaaggtcag attcatggcc tacaccgacg agaccttcaa gaccagggag 1320

gccatccagc acgagtctgg catcctgggc cccctgctgt acggcgaggt gggcgacacc 1380

ctgctgatca tcttcaagaa ccaggccagc aggccctaca acatctaccc ccacggcatc 1440

accgatgtga ggcccctgta cagcaggagg ctgcccaagg gcgtgaagca cctgaaggac 1500

ttccccatcc tgcccggcga gatcttcaag tacaagtgga ccgtgaccgt ggaggatggc 1560

cccaccaagt ctgaccccag gtgcctgacc aggtactaca gcagcttcgt gaacatggag 1620

agggacctgg cctctggcct gatcggcccc ctgctgatct gctacaagga gagcgtggac 1680

cagaggggca accagatcat gtctgacaag aggaacgtga tcctgttctc tgtgttcgat 1740

gagaacagga gctggtatct gaccgagaac atccagaggt tcctgcccaa ccccgccggc 1800

gtgcagctgg aggaccccga gttccaggcc agcaacatca tgcacagcat caacggctac 1860

gtgttcgaca gcctgcagct gtctgtgtgc ctgcacgagg tggcctactg gtacatcctg 1920

agcatcggcg cccagaccga cttcctgtct gtgttcttct ctggctacac cttcaagcac 1980

aagatggtgt acgaggacac cctgaccctg ttccccttca gcggcgagac cgtgttcatg 2040

agcatggaga accccggcct gtggatcctg ggctgccaca acagcgactt caggaacagg 2100

ggcatgaccg ccctgctgaa agtcagcagc tgcgacaaga acaccggcga ctactacgag 2160

gacagctacg aggacatcag cgcctacctg ctgagcaaga acaacgccat cgagcccagg 2220

<210> 23
<211> 2052
<212> ADN
<213> Trình tự nhân tạo

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

<400> 23
gagatcacca ggaccaccct gcagagcgac caggaggaga tcgactatga tgacaccatc 60

agcgtggaga tgaagaagga ggacttcgac atctacgacg aggacgagaa ccagagcccc 120

aggagcttcc agaagaagac caggcactac ttcatcgccg ccgtggagag gctgtgggac 180

tatggcatga gcagcagccc ccacgtgctg aggaacaggg cccagagcgg cagcgtgccc 240

cagttcaaga aggtggtgtt ccaggagttc accgacggca gcttcaccca gcccctgtac 300

agaggcgagc tgaacgagca cctgggcctg ctgggccccc acatcagggc cgaggtggag 360

gacaacatca tggtgacctt caggaaccag gccagcaggc cctacagctt ctacagcagc 420

ctgatcagct acgaggagga ccagaggcag ggcgccgagc ccaggaagaa cttcgtgaag 480

cccaacgaga ccaagaccta cttctggaag gtgcagcacc acatggcccc caccaaggac 540

gagttcgact gcaaggcctg ggcctacttc tctgatgtgg acctggagaa ggacgtgcac 600

agcggcctga tcggcccccc gctggtgtgc cacaccaaca ccctgaaccc cgcccacggc 660

aggcaggtga ccgtgcagga gttcgccctg ttcttcacca tcttcgacga gaccaagagc 720

tggtacttca ccgagaacat ggagaggaac tgcagggccc cctgcaacat ccagatggag 780

gaccccacct tcaaggagaa ctacaggttc cacgccatca acggctacat catggacacc 840

ctgcccggcc tggtgatggc ccaggaccag aggatcaggt ggtatctgct gagcatgggc 900

agcaacgaga acatccacag catccacttc agcggccacg tgttcaccgt gaggaagaag 960

gaggagtaca agatggccct gtacaacctg taccccggcg tgttcgagac cgtggagatg 1020

ctgcccagca aggccggcat ctggagggtg gagtgcctga tcggcgagca cctgcacgcc 1080

ggcatgagca ccctgttcct ggtgtacagc aacaagtgcc agacccccct gggcatggcc 1140

agcggccaca tcagggactt ccagatcacc gcctctggcc agtacggcca gtgggccccc 1200

aagctggcca ggctgcacta cagcggcagc atcaacgcct ggagcaccaa ggagcccttc 1260

agctggatca aggtggacct gctggccccc atgatcatcc acggcatcaa gacccagggc 1320

gccaggcaga agttcagcag cctgtacatc agccagttca tcatcatgta cagcctggac 1380

ggcaagaagt ggcagaccta caggggcaac agcaccggca ccctgatggt gttcttcggc 1440

aacgtggaca gcagcggcat caagcacaac atcttcaacc cccccatcat cgccaggtac 1500

atcaggctgc accccaccca ctacagcatc aggagcaccc tgcggatgga actgatgggc 1560

tgcgacctga acagctgcag catgcccctg ggcatggaga gcaaggccat ctctgacgcc 1620

cagatcaccg ccagcagcta cttcaccaac atgttcgcca cctggagccc cagcaaggcc 1680

aggctgcacc tgcagggcag gagcaacgcc tggaggcccc aggtgaacaa ccccaaggag 1740

tggctgcagg tggacttcca gaagaccatg aaggtgaccg gcgtgaccac ccagggcgtg 1800

aagagcctgc tgaccagcat gtacgtgaag gagttcctga tcagcagcag ccaggacggc 1860

caccagtgga ccctgttctt ccagaacggc aaagtgaagg tgttccaggg caaccaggac 1920

agcttcaccc ccgtggtgaa cagcctggac ccccccctgc tgaccaggta tctgaggatc 1980

cacccccaga gctgggtgca ccagatcgcc ctgagaatgg aagtgctggg atgcgaggcc 2040

caggacctgt ac 2052

<210> 24
<211> 2220
<212> ADN
<213> Trình tự nhân tạo

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

<400> 24
gccaccagga gatactacct gggggctgtg gaactttctt gggactacat gcagtctgac 60

ctgggagagc tgcctgtgga tgccaggttc ccacccagag tgcccaagtc cttcccattc 120

aacacctctg tggtctacaa gaagacactc tttgtggaat tcactgacca cctgttcaac 180

attgcaaaac ccagaccacc ctggatggga ctcctgggac ccaccattca ggctgaggtg 240

tatgacactg tggtcatcac cctcaagaac atggcatccc accctgtgtc tctgcatgct 300

gtgggagtct catactggaa agcctctgaa ggggctgagt atgatgacca gacatcccag 360

agagagaaag aggatgacaa ggtgttccct gggggatctc acacctatgt gtggcaagtc 420

ctcaaggaga atggacccat ggcatctgac ccactctgcc tgacatactc ctacctttct 480

catgtggacc tggtcaagga cctcaactct ggactgattg gggcactgct ggtgtgcagg 540

gaaggatccc tggccaagga gaaaacccag acactgcaca agttcattct cctgtttgct 600

gtctttgatg agggcaagtc ttggcactct gaaacaaaga actccctgat gcaagacagg 660

gatgctgcct ctgccagggc atggcccaag atgcacactg tgaatggcta tgtgaacaga 720

tcactgcctg gactcattgg ctgccacagg aaatctgtct actggcatgt gattggcatg 780

gggacaaccc ctgaagtgca ctccattttc ctggagggac acaccttcct ggtcaggaac 840

cacagacaag cctctctgga gatctctccc atcaccttcc tcactgcaca gacactgctg 900

atggaccttg gacagttcct gctgttctgc cacatctctt cccaccagca tgatggcatg 960

gaagcctatg tcaaggtgga ctcatgccct gaggaaccac agctcaggat gaagaacaat 1020

gaggaggctg aggactatga tgatgacctg actgactctg agatggatgt ggtcagattt 1080

gatgatgaca actctccatc cttcattcag atcaggtctg tggcaaagaa acaccccaag 1140

acatgggtgc actacattgc tgctgaggaa gaggactggg actatgcacc actggtcctg 1200

gcccctgatg acaggagcta caagtctcag tacctcaaca atggcccaca aagaattgga 1260

agaaagtaca agaaagtcag attcatggcc tacactgatg aaaccttcaa gacaagagaa 1320

gccattcagc atgagtctgg cattctggga ccactcctgt atggggaagt gggagacacc 1380

ctgctcatca tcttcaagaa ccaggcctcc aggccctaca acatctaccc acatggcatc 1440

actgatgtca ggcccctgta cagcaggaga ctgccaaaag gggtgaaaca cctcaaggac 1500

ttccccattc tgcctggaga gatcttcaag tacaagtgga ctgtcactgt ggaggatgga 1560

ccaacaaagt ctgaccccag gtgcctcacc agatactact cctcttttgt gaacatggag 1620

agagacctgg catctggact gattggacca ctgctcatct gctacaagga gtctgtggac 1680

cagagaggca accagatcat gtctgacaag agaaatgtga ttctgttctc tgtctttgat 1740

gagaacagat catggtacct gactgagaac attcagagat tcctgcccaa ccctgctggg 1800

gtgcaactgg aagaccctga gttccaggca agcaacatca tgcactccat caatggctat 1860

gtgtttgact ctctccagct ttctgtctgc ctgcatgagg tggcctactg gtacattctt 1920

tctattgggg cacaaactga cttccttttct gtcttcttct ctggatacac cttcaagcac 1980

aagatggtgt atgaggacac cctgacactc ttcccattct ctggggaaac tgtgttcatg 2040

agcatggaga accctggact gtggattctg ggatgccaca actctgactt cagaaacagg 2100

ggaatgactg cactgctcaa agtctcctcc tgtgacaaga acactgggga ctactatgag 2160

gactcttatg aggacatctc tgcctacctg ctcagcaaga acaatgccat tgagcccaga 2220

<210> 25
<211> 2052
<212> ADN
<213> Trình tự nhân tạo

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

<400> 25
gagatcacca ggacaaccct ccagtctgac caggaagaga ttgactatga tgacaccatt 60

tctgtggaga tgaagaagga ggactttgac atctatgatg aggacgagaa ccagtctcca 120

agatcattcc agaagaagac aagacactac ttcattgctg ctgtggaaag actgtgggac 180

tatggcatgt cttcctctcc ccatgtcctc aggaacaggg cacagtctgg ctctgtgcca 240

cagttcaaga aagtggtctt ccaggagttc actgatggct cattcaccca gcccctgtac 300

agaggggaac tgaatgagca cctgggactc ctgggaccat acatcagggc tgaggtggaa 360

gacaacatca tggtgacatt cagaaaccag gcctccaggc cctacagctt ctactcttcc 420

ctcatcagct atgaggaaga ccagagacaa ggggctgagc caagaaagaa ctttgtgaaa 480

cccaatgaaa ccaagaccta cttctggaaa gtccagcacc acatggcacc caccaaggat 540

gagtttgact gcaaggcctg ggcatacttc tctgatgtgg acctggagaa agatgtgcac 600

tctggcctga ttggcccact cctggtctgc cacaccaaca ccctgaaccc tgcacatgga 660

aggcaagtga ctgtgcagga gtttgccctc ttcttcacca tctttgatga aaccaagtca 720

tggtacttca ctgagaacat ggagagaaac tgcagagcac catgcaacat tcagatggaa 780

gaccccacct tcaaggagaa ctacaggttc atgccatca atggctacat catggacacc 840

ctgcctgggc ttgtcatggc acaggaccag agaatcagat ggtacctgct ttctatggga 900

tccaatgaga acattcactc catccacttc tctgggcatg tcttcactgt gagaaagaag 960

gaggaataca agatggccct gtacaacctc taccctgggg tctttgagac tgtggagatg 1020

ctgccctcca aagctggcat ctggagggtg gaatgcctca ttggggagca cctgcatgct 1080

ggcatgtcaa ccctgttcct ggtctacagc aacaagtgcc agacacccct gggaatggcc 1140

tctggccaca tcagggactt ccagatcact gcctctggcc agtatggcca gtgggcaccc 1200

aaactggcca ggctccacta ctctggctcc atcaatgcat ggtcaaccaa ggagccattc 1260

tcttggatca aggtggacct gctggcaccc atgatcattc atggcatcaa gacacagggg 1320

gcaagacaga aattctcctc tctgtacatc tcacagttca tcatcatgta ctctctggat 1380

ggcaagaagt ggcagacata cagaggcaac tccactggca ccctcatggt cttctttggc 1440

aatgtggaca gctctggcat caagcacaac atcttcaacc ctcccatcat tgccagatac 1500

atcaggctgc accccaccca ctactcaatc agatcaaccc tcaggatgga actgatggga 1560

tgtgacctga actcctgctc aatgcccctg ggaatggaga gcaaggccat ttctgatgcc 1620

cagatcactg catcctctta cttcaccaac atgtttgcca cctggtcacc atcaaaagcc 1680

aggctgcacc tccagggaag aagcaatgcc tggagacccc aggtcaacaa cccaaaggaa 1740

tggctgcaag tggacttcca gaagacaatg aaagtcactg gggtgacaac ccaggggtc 1800

aagtctctgc tcacctcaat gtatgtgaag gagttcctga tctcttcctc acaggatggc 1860

caccagtgga cactcttctt ccagaatggc aaagtcaagg tgttccaggg caaccaggac 1920

tctttcacac ctgtggtgaa ctcactggac ccccccctcc tgacaagata cctgagaatt 1980

caccccagt cttgggtcca ccagattgcc ctgagaatgg aagtcctggg atgtgaggca 2040

caagacctgt ac 2052

<210> 26
<211> 4332
<212> ADN
<213> Trình tự nhân tạo

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

<400> 26
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagagag 2280

atcaccagga caaccctcca gtctgaccag gaagagattg actatgatga caccatttct 2340

gtggagatga agaaggagga ctttgacatc tatgatgagg acgagaacca gtctccaaga 2400

tcattccaga agaagacaag acactacttc attgctgctg tggaaagact gtgggactat 2460

ggcatgtctt cctctcccca tgtcctcagg aacagggcac agtctggctc tgtgccacag 2520

ttcaagaaag tggtcttcca ggagttcact gatggctcat tcacccagcc cctgtacaga 2580

ggggaactga atgagcacct gggactcctg ggaccataca tcagggctga ggtggaagac 2640

aacatcatgg tgacattcag aaaccaggcc tccaggccct acagcttcta ctcttccctc 2700

atcagctatg aggaagacca gagacaaggg gctgagccaa gaaagaactt tgtgaaaccc 2760

aatgaaacca agacctactt ctggaaagtc cagcaccaca tggcacccac caaggatgag 2820

tttgactgca aggcctgggc atacttctct gatgtggacc tggagaaaga tgtgcactct 2880

ggcctgattg gcccactcct ggtctgccac accaacaccc tgaaccctgc acatggaagg 2940

caagtgactg tgcaggagtt tgccctcttc ttcaccatct ttgatgaaac caagtcatgg 3000

tacttcactg agaacatgga gagaaactgc agagcaccat gcaacattca gatggaagac 3060

cccaccttca aggagaacta caggttccat gccatcaatg gctacatcat ggacaccctg 3120

cctgggcttg tcatggcaca ggaccagaga atcagatggt acctgctttc tatgggatcc 3180

aatgagaaca ttcactccat ccacttctct gggcatgtct tcactgtgag aaagaaggag 3240

gaatacaaga tggccctgta caacctctac cctggggtct ttgagactgt ggagatgctg 3300

ccctccaaag ctggcatctg gagggtggaa tgcctcattg gggagcacct gcatgctggc 3360

atgtcaaccc tgttcctggt ctacagcaac aagtgccaga cacccctggg aatggcctct 3420

ggccacatca gggacttcca gatcactgcc tctggccagt atggccagtg ggcacccaaa 3480

ctggccaggc tccactactc tggctccatc aatgcatggt caaccaagga gccattctct 3540

tggatcaagg tggacctgct ggcacccatg atcattcatg gcatcaagac acaggggca 3600

agacagaaat tctcctctct gtacatctca cagttcatca tcatgtactc tctggatggc 3660

aagaagtggc agacatacag aggcaactcc actggcaccc tcatggtctt ctttggcaat 3720

gtggacagct ctggcatcaa gcacaacatc ttcaaccctc ccatcattgc cagatacatc 3780

aggctgcacc ccacccacta ctcaatcaga tcaaccctca ggatggaact gatgggatgt 3840

gacctgaact cctgctcaat gcccctggga atggagagca aggccatttc tgatgcccag 3900

atcactgcat cctcttactt caccaacatg tttgccacct ggtcaccatc aaaagccagg 3960

ctgcacctcc agggaagaag caatgcctgg agaccccagg tcaacaaccc aaaggaatgg 4020

ctgcaagtgg acttccagaa gacaatgaaa gtcactgggg tgacaaccca gggggtcaag 4080

tctctgctca cctcaatgta tgtgaaggag ttcctgatct cttcctcaca ggatggccac 4140

cagtggacac tcttcttcca gaatggcaaa gtcaaggtgt tccagggcaa ccaggactct 4200

ttcacacctg tggtgaactc actggacccc cccctcctga caagatacct gagaattcac 4260

ccccagtctt gggtccacca gattgccctg agaatggaag tcctgggatg tgaggcacaa 4320

gacctgtact ga 4332

<210> 27
<211> 4368
<212> ADN
<213> Trình tự nhân tạo

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

<400> 27
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc 2280

ttctctcaga attccagaca ccccagcacc agggagatca ccaggacaac cctccagtct 2340

gaccaggaag agattgacta tgatgacacc atttctgtgg agatgaagaa ggaggacttt 2400

gacatctatg atgaggacga gaaccagtct ccaagatcat tccagaagaa gacaagacac 2460

tacttcattg ctgctgtgga aagactgtgg gactatggca tgtcttcctc tccccatgtc 2520

ctcaggaaca gggcacagtc tggctctgtg ccacagttca agaaagtggt cttccaggag 2580

ttcactgatg gctcattcac ccagcccctg tacagagggg aactgaatga gcacctggga 2640

ctcctgggac catacatcag ggctgaggtg gaagacaaca tcatggtgac attcagaaac 2700

caggcctcca ggccctacag cttctactct tccctcatca gctatgagga agaccagaga 2760

caaggggctg agccaagaaa gaactttgtg aaacccaatg aaaccaagac ctacttctgg 2820

aaagtccagc accacatggc acccaccaag gatgagtttg actgcaaggc ctgggcatac 2880

ttctctgatg tggacctgga gaaagatgtg cactctggcc tgattggccc actcctggtc 2940

tgccacacca acaccctgaa ccctgcacat ggaaggcaag tgactgtgca ggagtttgcc 3000

ctcttcttca ccatctttga tgaaaccaag tcatggtact tcactgagaa catggagaga 3060

aactgcagag caccatgcaa cattcagatg gaagacccca ccttcaagga gaactacagg 3120

ttccatgcca tcaatggcta catcatggac accctgcctg ggcttgtcat ggcacaggac 3180

cagagaatca gatggtacct gctttctatg ggatccaatg agaacattca ctccatccac 3240

ttctctgggc atgtcttcac tgtgagaaag aaggaggaat acaagatggc cctgtacaac 3300

ctctaccctg gggtctttga gactgtggag atgctgccct ccaaagctgg catctggagg 3360

gtggaatgcc tcattgggga gcacctgcat gctggcatgt caaccctgtt cctggtctac 3420

agcaacaagt gccagacacc cctgggaatg gcctctggcc acatcaggga cttccagatc 3480

actgcctctg gccagtatgg ccagtgggca cccaaactgg ccaggctcca ctactctggc 3540

tccatcaatg catggtcaac caaggagcca ttctcttgga tcaaggtgga cctgctggca 3600

cccatgatca ttcatggcat caagacacag ggggcaagac agaaattctc ctctctgtac 3660

atctcacagt tcatcatcat gtactctctg gatggcaaga agtggcagac atacagaggc 3720

aactccactg gcaccctcat ggtcttcttt ggcaatgtgg acagctctgg catcaagcac 3780

aacatcttca accctcccat cattgccaga tacatcaggc tgcaccccac ccactactca 3840

atcagatcaa ccctcaggat ggaactgatg ggatgtgacc tgaactcctg ctcaatgccc 3900

ctgggaatgg agagcaaggc catttctgat gcccagatca ctgcatcctc ttacttcacc 3960

aacatgtttg ccacctggtc accatcaaaa gccaggctgc acctccaggg aagaagcaat 4020

gcctggagac cccaggtcaa caacccaaag gaatggctgc aagtggactt ccagaagaca 4080

atgaaagtca ctggggtgac aacccagggg gtcaagtctc tgctcacctc aatgtatgtg 4140

aaggagttcc tgatctcttc ctcacaggat ggccaccagt ggacactctt cttccagaat 4200

ggcaaagtca aggtgttcca gggcaaccag gactctttca cacctgtggt gaactcactg 4260

gacccccccc tcctgacaag atacctgaga attcaccccc agtcttgggt ccaccagatt 4320

gccctgagaa tggaagtcct gggatgtgag gcacaagacc tgtactga 4368

<210> 28
<211> 4332
<212> ADN
<213> Trình tự nhân tạo

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

<400> 28
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccagggag 2280

atcaccagga ccaccctgca gagcgaccag gaggagatcg actatgatga caccatcagc 2340

gtggagatga agaaggagga cttcgacatc tacgacgagg acgagaacca gagccccagg 2400

agcttccaga agaagaccag gcactacttc atcgccgccg tggagaggct gtgggactat 2460

ggcatgagca gcagccccca cgtgctgagg aacagggccc agagcggcag cgtgccccag 2520

ttcaagaagg tggtgttcca ggagttcacc gacggcagct tcacccagcc cctgtacaga 2580

ggcgagctga acgagcacct gggcctgctg ggcccctaca tcagggccga ggtggaggac 2640

aacatcatgg tgaccttcag gaaccaggcc agcaggccct acagcttcta cagcagcctg 2700

atcagctacg aggaggacca gaggcagggc gccgagccca ggaagaactt cgtgaagccc 2760

aacgagacca agacctactt ctggaaggtg cagcaccaca tggcccccac caaggacgag 2820

ttcgactgca aggcctgggc ctacttctct gatgtggacc tggagaagga cgtgcacagc 2880

ggcctgatcg gcccccctgct ggtgtgccac accaacaccc tgaaccccgc ccacggcagg 2940

caggtgaccg tgcaggagtt cgccctgttc ttcaccatct tcgacgagac caagagctgg 3000

tacttcaccg agaacatgga gaggaactgc agggcccccct gcaacatcca gatggaggac 3060

cccaccttca aggagaacta caggttccac gccatcaacg gctacatcat ggacaccctg 3120

cccggcctgg tgatggccca ggaccagagg atcaggtggt atctgctgag catgggcagc 3180

aacgagaaca tccacagcat ccacttcagc ggccacgtgt tcaccgtgag gaagaaggag 3240

gagtacaaga tggccctgta caacctgtac cccggcgtgt tcgagaccgt ggagatgctg 3300

cccagcaagg ccggcatctg gagggtggag tgcctgatcg gcgagcacct gcacgccggc 3360

atgagcaccc tgttcctggt gtacagcaac aagtgccaga ccccccctggg catggccagc 3420

ggccacatca gggacttcca gatcaccgcc tctggccagt acggccagtg ggcccccaag 3480

ctggccaggc tgcactacag cggcagcatc aacgcctgga gcaccaagga gcccttcagc 3540

tggatcaagg tggacctgct ggcccccatg atcatccacg gcatcaagac ccagggcgcc 3600

aggcagaagt tcagcagcct gtacatcagc cagttcatca tcatgtacag cctggacggc 3660

aagaagtggc agacctacag gggcaacagc accggcaccc tgatggtgtt cttcggcaac 3720

gtggacagca gcggcatcaa gcacaacatc ttcaaccccc ccatcatcgc caggtacatc 3780

aggctgcacc ccacccacta cagcatcagg agcaccctgc ggatggaact gatgggctgc 3840

gacctgaaca gctgcagcat gcccctgggc atggagagca aggccatctc tgacgcccag 3900

atcaccgcca gcagctactt caccaacatg ttcgccacct ggagccccag caaggccagg 3960

ctgcacctgc agggcaggag caacgcctgg aggccccagg tgaacaaccc caaggagtgg 4020

ctgcaggtgg acttccagaa gaccatgaag gtgaccggcg tgaccaccca gggcgtgaag 4080

agcctgctga ccagcatgta cgtgaaggag ttcctgatca gcagcagcca ggacggccac 4140

cagtggaccc tgttcttcca gaacggcaaa gtgaaggtgt tccagggcaa ccaggacagc 4200

ttcacccccg tggtgaacag cctggacccc cccctgctga ccaggtatct gaggatccac 4260

ccccagagct gggtgcacca gatcgccctg agaatggaag tgctgggatg cgaggcccag 4320

gacctgtact ga 4332

<210> 29
<211> 4368
<212> ADN
<213> Trình tự nhân tạo

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

<400> 29
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccaggagc 2280

ttcagccaga actccagaca ccccagcacc agggagatca ccaggaccac cctgcagagc 2340

gaccaggagg agatcgacta tgatgacacc atcagcgtgg agatgaagaa ggaggacttc 2400

gacatctacg acgaggacga gaaccagagc cccaggagct tccagaagaa gaccaggcac 2460

tacttcatcg ccgccgtgga gaggctgtgg gactatggca tgagcagcag cccccacgtg 2520

ctgaggaaca gggcccagag cggcagcgtg ccccagttca agaaggtggt gttccaggag 2580

ttcaccgacg gcagcttcac ccagcccctg tacagaggcg agctgaacga gcacctgggc 2640

ctgctgggcc cctacatcag ggccgaggtg gaggacaaca tcatggtgac cttcaggaac 2700

caggccagca ggccctacag cttctacagc agcctgatca gctacgagga ggaccagagg 2760

cagggcgccg agcccaggaa gaacttcgtg aagcccaacg agaccaagac ctacttctgg 2820

aaggtgcagc accacatggc ccccaccaag gacgagttcg actgcaaggc ctgggcctac 2880

ttctctgatg tggacctgga gaaggacgtg cacagcggcc tgatcggccc cctgctggtg 2940

tgccacacca acaccctgaa ccccgcccac ggcaggcagg tgaccgtgca ggagttcgcc 3000

ctgttcttca ccatcttcga cgagaccaag agctggtact tcaccgagaa catggagagg 3060

aactgcaggg ccccctgcaa catccagatg gaggacccca ccttcaagga gaactacagg 3120

ttccacgcca tcaacggcta catcatggac accctgcccg gcctggtgat ggcccaggac 3180

cagaggatca ggtggtatct gctgagcatg ggcagcaacg agaacatcca cagcatccac 3240

ttcagcggcc acgtgttcac cgtgaggaag aaggaggagt acaagatggc cctgtacaac 3300

ctgtaccccg gcgtgttcga gaccgtggag atgctgccca gcaaggccgg catctggagg 3360

gtggagtgcc tgatcggcga gcacctgcac gccggcatga gcaccctgtt cctggtgtac 3420

agcaacaagt gccagacccc cctgggcatg gccagcggcc acatcaggga cttccagatc 3480

accgcctctg gccagtacgg ccagtgggcc cccaagctgg ccaggctgca ctacagcggc 3540

agcatcaacg cctggagcac caaggagccc ttcagctgga tcaaggtgga cctgctggcc 3600

cccatgatca tccacggcat caagacccag ggcgccaggc agaagttcag cagcctgtac 3660

atcagccagt tcatcatcat gtacagcctg gacggcaaga agtggcagac ctacaggggc 3720

aacagcaccg gcaccctgat ggtgttcttc ggcaacgtgg acagcagcgg catcaagcac 3780

aacatcttca acccccccat catcgccagg tacatcaggc tgcaccccac ccactacagc 3840

atcaggagca ccctgcggat ggaactgatg ggctgcgacc tgaacagctg cagcatgccc 3900

ctgggcatgg agagcaaggc catctctgac gcccagatca ccgccagcag ctacttcacc 3960

aacatgttcg ccacctggag ccccagcaag gccaggctgc acctgcaggg caggagcaac 4020

gcctggaggc cccaggtgaa caaccccaag gagtggctgc aggtggactt ccagaagacc 4080

atgaaggtga ccggcgtgac cacccagggc gtgaagagcc tgctgaccag catgtacgtg 4140

aaggagttcc tgatcagcag cagccaggac ggccaccagt ggaccctgtt cttccagaac 4200

ggcaaagtga aggtgttcca gggcaaccag gacagcttca cccccgtggt gaacagcctg 4260

gacccccccc tgctgaccag gtatctgagg atccaccccc agagctgggt gcaccagatc 4320

gccctgagaa tggaagtgct gggatgcgag gcccaggacc tgtactga 4368

<210> 30
<211> 14
<212> PRT
<213> Trình tự nhân tạo

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

<400> 30
Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His Gln Arg
1 5 10

<210> 31
<211> 24
<212> PRT
<213> Sus sp.

<400> 31
Ser Phe Ala Gln Asn Ser Arg Pro Pro Ser Ala Ser Ala Pro Lys Pro
1 5 10 15

Pro Val Leu Arg Arg His Gln Arg
20

<210> 32
<211> 16
<212> PRT
<213> Sus sp.

<400> 32
Ser Phe Ser Gln Asn Ser Arg His Gln Ala Tyr Arg Tyr Arg Arg Gly
1 5 10 15

<210> 33
<211> 8
<212> PRT
<213> Trình tự nhân tạo

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

<400> 33
Thr Thr Tyr Val Asn Arg Ser Leu
1 5

<210> 34
<211> 7
<212> PRT
<213> Người hiện đại

<400> 34
Pro Gln Leu Arg Met Lys Asn
1 5

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

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

<400> 35
Val Asp Gln Arg Gly Asn Gln

1 5

<210> 36
<211> 87
<212> ADN
<213> Trình tự nhân tạo

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

<400> 36
agcttcagcc agaatgtgag caacaatgtg agcaacaatg ccaccaataa tgctaccaac 60

ccacctgtcc tgaaacgcca ccagagg 87

<210> 37
<211> 75
<212> ADN
<213> Trình tự nhân tạo

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

<400> 37
agcttcagcc agaatgtgag caacaatgcc accaacaatg tgagcaaccc acctgtcctg 60

aaacgccacc agagg 75

<210> 38
<211> 63
<212> ADN
<213> Trình tự nhân tạo

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

<400> 38
agcttcagcc agaatgtgag caataatgcc accaacccac ctgtcctgaa acgccaccag 60

agg 63

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

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

<400> 39
agcttcagcc agaatgtgag caataatcca cctgtcctga aacgccacca gagg 54

<210> 40
<211> 51
<212> ADN
<213> Trình tự nhân tạo

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

<400> 40
agcttcagcc agaataggag cctgccacct gtcctgaaac gccaccagag g 51

<210> 41
<211> 93
<212> ADN
<213> Trình tự nhân tạo

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

<400> 41
agcttcagcc agaatgccac taatgtgtct aacaactctg ctacctctgc tgactctgct 60

gtgagcccac ctgtcctgaa acgccaccag agg 93

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

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

<400> 42
agcttcagcc agaatgccac caactatgtg aacaggagcc tgccacctgt cctgaaacgc 60

caccagagg 69

<210> 43
<211> 105
<212> ADN
<213> Trình tự nhân tạo

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

<400> 43
agcttcagcc agaatgccac caactatgtg aacaggagcc tgtctgccac ctctgctgac 60

tctgctgtga gccagaatcc acctgtcctg aaacgccacc agagg 105

<210> 44
<211> 84
<212> ADN

<213> Trình tự nhân tạo

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

<400> 44
agcttcagcc agaatgtgag caacaatgtg agcaatgctg tgtctgctgt gtctgctcca 60

cctgtcctga aacgccacca gagg 84

<210> 45
<211> 90
<212> ADN
<213> Trình tự nhân tạo

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

<400> 45
agcttcagcc agaatatcac tgtggcctct gccacctcta acatcactgt ggcctctgct 60

gacccacctg tcctgaaacg ccaccagagg 90

<210> 46
<211> 72
<212> ADN
<213> Trình tự nhân tạo

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

<400> 46
agcttcagcc agaatatcac tgtgaccaac atcactgtga ctgccccacc tgtcctgaaa 60

cgccaccaga gg 72

<210> 47
<211> 72
<212> ADN
<213> Trình tự nhân tạo

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

<400> 47
agcttcagcc agaatcagac tgtgaccaac atcactgtga ctgccccacc tgtcctgaaa 60

cgccaccaga gg 72

<210> 48
<211> 93
<212> ADN
<213> Trình tự nhân tạo

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

<400> 48
agcttcagcc agaatgccac taatgtgtct aacaacagca acaccagcaa tgacagcaat 60

gtgtctccac ctgtcctgaa acgccaccag agg 93

<210> 49
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 49
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gtcctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc 2280

ttctctcaga atccacctgt cctgaagaga caccagagag agatcaccag gacaaccctc 2340

cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag 2400

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca 2460

agacactact tcattgctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc 2520

catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caggagttca ctgatggctc attcacccag cccctgtaca gaggggaact gaatgagcac 2640

ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc 2700

agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac 2760

cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggcaccc accaaggatg agtttgactg caaggcctgg 2880

gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg 3060

gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac 3120

tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca 3180

caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttcctg 3420

gtctacagca acaagtgcca gacacccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540

tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg 3600

ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct 3660

ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720

agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca ccccacccac 3840

tactcaatca gatcaaccct caggatggaa ctgatgggat gtgacctgaa ctcctgctca 3900

atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atcctcttac 3960

ttcaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccagggaaga 4020

agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag 4080

aagacaatga aagtcactgg ggtgacaacc caggggtca agtctctgct cacctcaatg 4140

tatgtgaagg agttcctgat ctcttcctca caggatggcc accagtggac actcttcttc 4200

cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac 4260

tcactggacc ccccctcct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320

cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

<210> 50
<211> 45
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(45)

<400> 50

gtg agc aac aat gtg agc aac aat gcc acc aat aat gct acc aac 45
Val Ser Asn Asn Val Ser Asn Asn Ala Thr Asn Asn Ala Thr Asn
1 5 10 15

<210> 51
<211> 15
<212> PRT
<213> Trình tự nhân tạo

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

<400> 51
Val Ser Asn Asn Val Ser Asn Asn Ala Thr Asn Asn Ala Thr Asn
1 5 10 15

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

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

<220>
<221> CDS
<222> (1)..(33)

<400> 52
gtg agc aac aat gcc acc aac aat gtg agc aac 33
Val Ser Asn Asn Ala Thr Asn Asn Val Ser Asn
1 5 10

<210> 53
<211> 11
<212> PRT
<213> Trình tự nhân tạo

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

<400> 53
Val Ser Asn Asn Ala Thr Asn Asn Val Ser Asn
1 5 10

<210> 54
<211> 21
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(21)

<400> 54
gtg agc aat aat gcc acc aac 21
Val Ser Asn Asn Ala Thr Asn
1 5

<210> 55
<211> 7
<212> PRT
<213> Trình tự nhân tạo

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

<400> 55
Val Ser Asn Asn Ala Thr Asn
1 5

<210> 56
<211> 12
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(12)

<400> 56
gtg agc aat aat 12
Val Ser Asn Asn
1

<210> 57
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 57
Val Ser Asn Asn
1

<210> 58

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

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

<220>
<221> CDS
<222> (1)..(9)

<400> 58
agg agc ctg 9
Arg Ser Leu
1

<210> 59
<211> 3
<212> PRT
<213> Trình tự nhân tạo

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

<400> 59
Arg Ser Leu
1

<210> 60
<211> 51
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(51)

<400> 60
gcc act aat gtg tct aac aac tct gct acc tct gct gac tct gct gtg 48
Ala Thr Asn Val Ser Asn Asn Ser Ala Thr Ser Ala Asp Ser Ala Val
1 5 10 15

agc 51
Ser

<210> 61
<211> 17
<212> PRT
<213> Trình tự nhân tạo

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

<400> 61

Ala Thr Asn Val Ser Asn Asn Ser Ala Thr Ser Ala Asp Ser Ala Val
1 5 10 15

Ser

<210> 62
<211> 27
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(27)

<400> 62
gcc acc aac tat gtg aac agg agc ctg 27
Ala Thr Asn Tyr Val Asn Arg Ser Leu
1 5

<210> 63
<211> 9
<212> PRT
<213> Trình tự nhân tạo

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

<400> 63
Ala Thr Asn Tyr Val Asn Arg Ser Leu
1 5

<210> 64
<211> 63
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS

<222> (1)..(63)

<400> 64
gcc acc aac tat gtg aac agg agc ctg tct gcc acc tct gct gac tct 48
Ala Thr Asn Tyr Val Asn Arg Ser Leu Ser Ala Thr Ser Ala Asp Ser
1 5 10 15

gct gtg agc cag aat 63
Ala Val Ser Gln Asn
20

<210> 65
<211> 21
<212> PRT
<213> Trình tự nhân tạo

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

<400> 65
Ala Thr Asn Tyr Val Asn Arg Ser Leu Ser Ala Thr Ser Ala Asp Ser
1 5 10 15

Ala Val Ser Gln Asn
20

<210> 66
<211> 42
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(42)

<400> 66
gtg agc aac aat gtg agc aat gct gtg tct gct gtg tct gct 42
Val Ser Asn Asn Val Ser Asn Ala Val Ser Ala Val Ser Ala
1 5 10

<210> 67
<211> 14
<212> PRT
<213> Trình tự nhân tạo

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

<400> 67
Val Ser Asn Asn Val Ser Asn Ala Val Ser Ala Val Ser Ala

1 5 10

<210> 68
<211> 48
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(48)

<400> 68
atc act gtg gcc tct gcc acc tct aac atc act gtg gcc tct gct gac 48
Ile Thr Val Ala Ser Ala Thr Ser Asn Ile Thr Val Ala Ser Ala Asp
1 5 10 15

<210> 69
<211> 16
<212> PRT
<213> Trình tự nhân tạo

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

<400> 69
Ile Thr Val Ala Ser Ala Thr Ser Asn Ile Thr Val Ala Ser Ala Asp
1 5 10 15

<210> 70
<211> 30
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(30)

<400> 70
atc act gtg acc aac atc act gtg act gcc 30
Ile Thr Val Thr Asn Ile Thr Val Thr Ala
1 5 10

<210> 71
<211> 10
<212> PRT
<213> Trình tự nhân tạo

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

<400> 71
Ile Thr Val Thr Asn Ile Thr Val Thr Ala
1 5 10

<210> 72
<211> 30
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(30)

<400> 72
cag act gtg acc aac atc act gtg act gcc 30
Gln Thr Val Thr Asn Ile Thr Val Thr Ala
1 5 10

<210> 73
<211> 10
<212> PRT
<213> Trình tự nhân tạo

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

<400> 73
Gln Thr Val Thr Asn Ile Thr Val Thr Ala
1 5 10

<210> 74
<211> 51
<212> ADN
<213> Trình tự nhân tạo

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

<220>
<221> CDS
<222> (1)..(51)

<400> 74
gcc act aat gtg tct aac aac agc aac acc agc aat gac agc aat gtg 48
Ala Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val

1 5 10 15

tct 51
Ser

<210> 75
<211> 17
<212> PRT
<213> Trình tự nhân tạo

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

<400> 75
Ala Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val
1 5 10 15

Ser

<210> 76
<211> 405
<212> PRT
<213> Người hiện đại

<400> 76
Met Pro Leu Leu Leu Tyr Thr Cys Leu Leu Trp Leu Pro Thr Ser Gly
1 5 10 15

Leu Trp Thr Val Gln Ala Met Asp Pro Asn Ala Ala Tyr Val Asn Met
20 25 30

Ser Asn His His Arg Gly Leu Ala Ser Ala Asn Val Asp Phe Ala Phe
35 40 45

Ser Leu Tyr Lys His Leu Val Ala Leu Ser Pro Lys Lys Asn Ile Phe
50 55 60

Ile Ser Pro Val Ser Ile Ser Met Ala Leu Ala Met Leu Ser Leu Gly
65 70 75 80

Thr Cys Gly His Thr Arg Ala Gln Leu Leu Gln Gly Leu Gly Phe Asn
85 90 95

Leu Thr Glu Arg Ser Glu Thr Glu Ile His Gln Gly Phe Gln His Leu
100 105 110

His Gln Leu Phe Ala Lys Ser Asp Thr Ser Leu Glu Met Thr Met Gly
115 120 125

```
Asn Ala Leu Phe Leu Asp Gly Ser Leu Glu Leu Leu Glu Ser Phe Ser
    130                 135                 140

Ala Asp Ile Lys His Tyr Tyr Glu Ser Glu Val Leu Ala Met Asn Phe
145                 150                 155                 160

Gln Asp Trp Ala Thr Ala Ser Arg Gln Ile Asn Ser Tyr Val Lys Asn
                165                 170                 175

Lys Thr Gln Gly Lys Ile Val Asp Leu Phe Ser Gly Leu Asp Ser Pro
            180                 185                 190

Ala Ile Leu Val Leu Val Asn Tyr Ile Phe Phe Lys Gly Thr Trp Thr
        195                 200                 205

Gln Pro Phe Asp Leu Ala Ser Thr Arg Glu Glu Asn Phe Tyr Val Asp
    210                 215                 220

Glu Thr Thr Val Val Lys Val Pro Met Met Leu Gln Ser Ser Thr Ile
225                 230                 235                 240

Ser Tyr Leu His Asp Ser Glu Leu Pro Cys Gln Leu Val Gln Met Asn
                245                 250                 255

Tyr Val Gly Asn Gly Thr Val Phe Phe Ile Leu Pro Asp Lys Gly Lys
            260                 265                 270

Met Asn Thr Val Ile Ala Ala Leu Ser Arg Asp Thr Ile Asn Arg Trp
        275                 280                 285

Ser Ala Gly Leu Thr Ser Ser Gln Val Asp Leu Tyr Ile Pro Lys Val
    290                 295                 300

Thr Ile Ser Gly Val Tyr Asp Leu Gly Asp Val Leu Glu Glu Met Gly
305                 310                 315                 320

Ile Ala Asp Leu Phe Thr Asn Gln Ala Asn Phe Ser Arg Ile Thr Gln
                325                 330                 335

Asp Ala Gln Leu Lys Ser Ser Lys Val Val His Lys Ala Val Leu Gln
            340                 345                 350

Leu Asn Glu Glu Gly Val Asp Thr Ala Gly Ser Thr Gly Val Thr Leu
        355                 360                 365
```

Asn Leu Thr Ser Lys Pro Ile Ile Leu Arg Phe Asn Gln Pro Phe Ile
370 375 380

Ile Met Ile Phe Asp His Phe Thr Trp Ser Ser Leu Phe Leu Ala Arg
385 390 395 400

Val Met Asn Pro Val
405

<210> 77
<211> 4
<212> PRT
<213> Người hiện đại

<400> 77
Asn Met Ser Asn
1

<210> 78
<211> 4
<212> PRT
<213> Người hiện đại

<400> 78
Asn Leu Thr Glu
1

<210> 79
<211> 4
<212> PRT
<213> Người hiện đại

<400> 79
Asn Lys Thr Gln
1

<210> 80
<211> 4
<212> PRT
<213> Người hiện đại

<400> 80
Asn Gly Thr Val
1

<210> 81
<211> 4
<212> PRT
<213> Người hiện đại

<400> 81
Asn Phe Ser Arg
1

<210> 82
<211> 4
<212> PRT
<213> Người hiện đại

<400> 82
Asn Leu Thr Ser
1

<210> 83
<211> 41
<212> PRT
<213> Trình tự nhân tạo

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

<400> 83
Leu Ser Lys Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Ala
1 5 10 15

Thr Asn Val Ser Asn Asn Ser Asn Thr Ser Asn Asp Ser Asn Val Ser
20 25 30

Pro Pro Val Leu Lys Arg His Gln Arg
35 40

<210> 84
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 84
Asn Ala Thr Asn
1

<210> 85
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 85
Asn Val Ser Asn
1

<210> 86
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 86
Asn Asn Ser Asn
1

<210> 87
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 87
Asn Thr Ser Asn
1

<210> 88
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 88
Asn Asp Ser Asn
1

<210> 89
<211> 4
<212> PRT
<213> Trình tự nhân tạo

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

<400> 89
Asn Val Ser Pro
1

<210> 90
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 90
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gtcctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgtctc agaatccacc tgtcctgaag agacaccaga gagagatcac caggacaacc 2340

ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag 2400

gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag 2460

acaagacact acttcattgc tgctgtggaa agactgtggg actatggcat gtcttcctct 2520

ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaggagt tcactgatgg ctcattcacc cagcccctgt acagagggga actgaatgag 2640

cacctgggac tcctgggacc atacatcagg gctgaggtgg aagacaacat catggtgaca 2700

ttcagaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa 2760

gaccagagac aaggggctga gccaagaaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcacatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaagt catggtactt cactgagaac 3060

atggagagaa actgcagagc accatgcaac attcagatgg aagaccccac cttcaaggag 3120

aactacaggt tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcacaggacc agagaatcag atggtacctg ctttctatgg gatccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgagaaaga aggaggaata caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggaatgcct cattggggag cacctgcatg ctggcatgtc aaccctgttc 3420

ctggtctaca gcaacaagtg ccagacaccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggcac ccaaactggc caggctccac 3540

tactctggct ccatcaatgc atggtcaacc aaggagccat tctcttggat caaggtggac 3600

ctgctggcac ccatgatcat tcatggcatc aagacacagg gggcaagaca gaaattctcc 3660

tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca 3720

tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccctcccatc attgccagat acatcaggct gcaccccacc 3840

cactactcaa tcagatcaac cctcaggatg gaactgatgg gatgtgacct gaactcctgc 3900

tcaatgcccc tgggaatgga gagcaaggcc atttctgatg cccagatcac tgcatcctct 3960

tacttcacca acatgtttgc cacctggtca ccatcaaaag ccaggctgca cctccaggga 4020

agaagcaatg cctggagacc ccaggtcaac aacccaaagg aatggctgca agtggacttc 4080

cagaagacaa tgaaagtcac tggggtgaca acccaggggg tcaagtctct gctcacctca 4140

atgtatgtga aggagttcct gatctcttcc tcacaggatg gccaccagtg gacactcttc 4200

ttccagaatg gcaaagtcaa ggtgttccag ggcaaccagg actctttcac acctgtggtg 4260

aactcactgg acccccccct cctgacaaga tacctgagaa ttcaccccca gtcttgggtc 4320

caccagattg ccctgagaat ggaagtcctg ggatgtgagg cacaagacct gtactga 4377

<210> 91
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 91
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaatc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg aagtctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccactgc atctgaccca ccctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgtctc agaatccacc tgtcctgaag agacaccaga gagagatcac caggacaacc 2340

ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag 2400

gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag 2460

acaagacact acttcattgc tgctgtggaa agactgtggg actatggcat gtcttcctct 2520

ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaggagt tcactgatgg ctcattcacc cagcccctgt acagagggga actgaatgag 2640

cacctgggac tcctgggacc atacatcagg gctgaggtgg aagacaacat catggtgaca 2700

ttcagaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa 2760

gaccagagac aaggggctga gccaagaaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcacatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaagt catggtactt cactgagaac 3060

atggagagaa actgcagagc accatgcaac attcagatgg aagaccccac cttcaaggag 3120

aactacaggt tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcacaggacc agagaatcag atggtacctg ctttctatgg gatccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgagaaaga aggaggaata caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggaatgcct cattggggag cacctgcatg ctggcatgtc aaccctgttc 3420

ctggtctaca gcaacaagtg ccagacaccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggcac ccaaactggc caggctccac 3540

tactctggct ccatcaatgc atggtcaacc aaggagccat tctcttggat caaggtggac 3600

ctgctggcac ccatgatcat tcatggcatc aagacacagg gggcaagaca gaaattctcc 3660

tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca 3720

tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccctcccatc attgccagat acatcaggct gcaccccacc 3840

cactactcaa tcagatcaac cctcaggatg gaactgatgg gatgtgacct gaactcctgc 3900

tcaatgcccc tgggaatgga gagcaaggcc atttctgatg cccagatcac tgcatcctct 3960

tacttcacca acatgtttgc cacctggtca ccatcaaaag ccaggctgca cctccaggga 4020

agaagcaatg cctggagacc ccaggtcaac aacccaaagg aatggctgca agtggacttc 4080

cagaagacaa tgaaagtcac tggggtgaca acccaggggg tcaagtctct gctcacctca 4140

atgtatgtga aggagttcct gatctcttcc tcacaggatg gccaccagtg gacactcttc 4200

ttccagaatg gcaaagtcaa ggtgttccag ggcaaccagg actctttcac acctgtggtg 4260

aactcactgg acccccccct cctgacaaga tacctgagaa ttcaccccca gtcttgggtc 4320

caccagattg ccctgagaat ggaagtcctg ggatgtgagg cacaagacct gtactga 4377

<210> 92
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 92
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaagc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg ggatctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccatggc atctgaccca ctctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgtctc agaatccacc tgtcctgaag agacaccaga gagagatcac caggacaacc 2340

ctccagtctg accaggaaga gattgactat gatgacacca tttctgtgga gatgaagaag 2400

gaggactttg acatctatga tgaggacgag aaccagtctc caagatcatt ccagaagaag 2460

acaagacact acttcattgc tgctgtggaa agactgtggg actatggcat gtcttcctct 2520

ccccatgtcc tcaggaacag ggcacagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaggagt tcactgatgg ctcattcacc cagcccctgt acagagggga actgaatgag 2640

cacctgggac tcctgggacc atacatcagg gctgaggtgg aagacaacat catggtgaca 2700

ttcagaaacc aggcctccag gccctacagc ttctactctt ccctcatcag ctatgaggaa 2760

gaccagagac aaggggctga gccaagaaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggca cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcatact tctctgatgt ggacctggag aaagatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcacatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaagt catggtactt cactgagaac 3060

atggagagaa actgcagagc accatgcaac attcagatgg aagaccccac cttcaaggag 3120

aactacaggt tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcacaggacc agagaatcag atggtacctg ctttctatgg gatccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgagaaaga aggaggaata caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggaatgcct cattggggag cacctgcatg ctggcatgtc aaccctgttc 3420

ctggtctaca gcaacaagtg ccagacaccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggcac ccaaactggc caggctccac 3540

tactctggct ccatcaatgc atggtcaacc aaggagccat tctcttggat caaggtggac 3600

ctgctggcac ccatgatcat tcatggcatc aagacacagg gggcaagaca gaaattctcc 3660

tctctgtaca tctcacagtt catcatcatg tactctctgg atggcaagaa gtggcagaca 3720

tacagaggca actccactgg caccctcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccctcccatc attgccagat acatcaggct gcaccccacc 3840

cactactcaa tcagatcaac cctcaggatg gaactgatgg gatgtgacct gaactcctgc 3900

tcaatgcccc tgggaatgga gagcaaggcc atttctgatg cccagatcac tgcatcctct 3960

tacttcacca acatgtttgc cacctggtca ccatcaaaag ccaggctgca cctccaggga 4020

agaagcaatg cctggagacc ccaggtcaac aacccaaagg aatggctgca agtggacttc 4080

cagaagacaa tgaaagtcac tggggtgaca acccaggggg tcaagtctct gctcacctca 4140

atgtatgtga aggagttcct gatctcttcc tcacaggatg gccaccagtg gacactcttc 4200

ttccagaatg gcaaagtcaa ggtgttccag ggcaaccagg actctttcac acctgtggtg 4260

aactcactgg acccccccct cctgacaaga tacctgagaa ttcaccccca gtcttgggtc 4320

caccagattg ccctgagaat ggaagtcctg ggatgtgagg cacaagacct gtactga 4377

<210> 93
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 93
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaatc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg aagtctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccactgc atctgaccca ccctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gttctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc 2280

ttctctcaga atccacctgt cctgaagaga caccagagag agatcaccag gacaaccctc 2340

cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag 2400

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca 2460

agacactact tcattgctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc 2520

catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caggagttca ctgatggctc attcacccag cccctgtaca gaggggaact gaatgagcac 2640

ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc 2700

agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac 2760

cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggcaccc accaaggatg agtttgactg caaggcctgg 2880

gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg 3060

gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac 3120

tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca 3180

caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttcctg 3420

gtctacagca acaagtgcca gacacccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540

tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg 3600

ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct 3660

ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720

agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca ccccacccac 3840

tactcaatca gatcaaccct caggatggaa ctgatgggat gtgacctgaa ctcctgctca 3900

atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atcctcttac 3960

ttcaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccagggaaga 4020

agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag 4080

aagacaatga aagtcactgg ggtgacaacc caggggggtca agtctctgct cacctcaatg 4140

tatgtgaagg agttcctgat ctcttcctca caggatggcc accagtggac actcttcttc 4200

cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac 4260

tcactggacc ccccccctcct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320

cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

<210> 94
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 94
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccactgc ctctgaccca ccctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaggagc 2280

ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccaccctc 2340

cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400

gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520

catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caagagttca ctgatggcag cttcacccag cccctgtaca gaggggagct gaatgagcac 2640

ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgaccttc 2700

cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760

cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttcac tgagaacatg 3060

gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120

taccgcttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggcc 3180

caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttcctg 3420

gtctacagca acaagtgcca gacccccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540

tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600

ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660

ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720

agaggcaact ccactggaac actcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca ccccacccac 3840

tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctcctgcagc 3900

atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960

ttcaccaaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccagggaagg 4020

agcaatgcct ggaggcccca ggtcaacaac ccaaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtcactgg ggtgaccacc caggggtca agagcctgct caccagcatg 4140

tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc 4200

cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac 4260

agcctggacc ccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac 4320

cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta ctga 4374

<210> 95
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 95
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgagcc agaatccacc tgtcctgaaa cgccaccaga gggagatcac caggaccacc 2340

ctccagtctg accaggagga gattgactat gatgacacca tttctgtgga gatgaagaaa 2400

gaggactttg acatctatga cgaggacgag aaccagagcc caaggagctt ccagaagaag 2460

accaggcact acttcattgc tgctgtggag cgcctgtggg actatggcat gagctccagc 2520

ccccatgtcc tcaggaacag ggcccagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaagagt tcactgatgg cagcttcacc cagcccctgt acagagggga gctgaatgag 2640

cacctgggac tcctgggccc atacatcagg gctgaggtgg aggacaacat catggtgacc 2700

ttccgcaacc aggcctccag gccctacagc ttctacagct ccctcatcag ctatgaggag 2760

gaccagaggc agggggctga gccacgcaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggcc cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcccatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaaga gctggtactt cactgagaac 3060

atggagcgca actgcagggc cccatgcaac attcagatgg aggaccccac cttcaaagag 3120

aactaccgct tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcccaggacc agaggatcag gtggtacctg ctttctatgg gctccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgcgcaaga aggaggagta caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggagtgcct cattggggag cacctgcatg ctggcatgag caccctgttc 3420

ctggtctaca gcaacaagtg ccagaccccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggccc ccaagctggc caggctccac 3540

tactctggat ccatcaatgc ctggagcacc aaggagccat tcagctggat caaagtggac 3600

ctgctggccc ccatgatcat ccatggcatc aagacccagg gggccaggca gaagttctcc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg atggcaagaa atggcagacc 3720

tacagaggca actccactgg aacactcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccccccaatc atcgccagat acatcaggct gcaccccacc 3840

cactacagca tccgcagcac cctcaggatg gagctgatgg gctgtgacct gaactcctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atttctgatg cccagatcac tgcctccagc 3960

tacttcacca acatgtttgc cacctggagc ccaagcaagg ccaggctgca cctccaggga 4020

aggagcaatg cctggaggcc ccaggtcaac aacccaaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtcac tggggtgacc acccaggggg tcaagagcct gctcaccagc 4140

atgtatgtga aggagttcct gatcagctcc agccaggatg gccaccagtg gaccctcttc 4200

ttccagaatg gcaaggtcaa ggtgttccag ggcaaccagg acagcttcac ccctgtggtg 4260

aacagcctgg acccccccct cctgaccaga tacctgagga ttcaccccca gagctgggtc 4320

caccagattg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga 4377

<210> 96
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 96
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccactgc ctctgaccca ccctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gttctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgagcc agaatccacc tgtcctgaaa cgccaccaga gggagatcac caggaccacc 2340

ctccagtctg accaggagga gattgactat gatgacacca tttctgtgga gatgaagaaa 2400

gaggactttg acatctatga cgaggacgag aaccagagcc caaggagctt ccagaagaag 2460

accaggcact acttcattgc tgctgtggag cgcctgtggg actatggcat gagctccagc 2520

ccccatgtcc tcaggaacag ggcccagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaagagt tcactgatgg cagcttcacc cagcccctgt acagagggga gctgaatgag 2640

cacctgggac tcctgggccc atacatcagg gctgaggtgg aggacaacat catggtgacc 2700

ttccgcaacc aggcctccag gccctacagc ttctacagct ccctcatcag ctatgaggag 2760

gaccagaggc aggggctga gccacgcaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggcc cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcccatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaaga gctggtactt cactgagaac 3060

atggagcgca actgcagggc cccatgcaac attcagatgg aggaccccac cttcaaagag 3120

aactaccgct tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcccaggacc agaggatcag gtggtacctg ctttctatgg gctccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgcgcaaga aggaggagta caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggagtgcct cattggggag cacctgcatg ctggcatgag caccctgttc 3420

ctggtctaca gcaacaagtg ccagaccccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggccc ccaagctggc caggctccac 3540

tactctggat ccatcaatgc ctggagcacc aaggagccat tcagctggat caaagtggac 3600

ctgctggccc ccatgatcat ccatggcatc aagacccagg gggccaggca gaagttctcc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg atggcaagaa atggcagacc 3720

tacagaggca actccactgg aacactcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccccccaatc atcgccagat acatcaggct gcaccccacc 3840

cactacagca tccgcagcac cctcaggatg gagctgatgg gctgtgacct gaactcctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atttctgatg cccagatcac tgcctccagc 3960

tacttcacca acatgtttgc cacctggagc ccaagcaagg ccaggctgca cctccaggga 4020

aggagcaatg cctggaggcc ccaggtcaac aacccaaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtcac tggggtgacc acccaggggg tcaagagcct gctcaccagc 4140

atgtatgtga aggagttcct gatcagctcc agccaggatg gccaccagtg gaccctcttc 4200

ttccagaatg gcaaggtcaa ggtgttccag ggcaaccagg acagcttcac ccctgtggtg 4260

aacagcctgg acccccccct cctgaccaga tacctgagga ttcaccccca gagctgggtc 4320

caccagattg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga 4377

<210> 97
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 97
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gtcctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaggagc 2280

ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccaccctc 2340

cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400

gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520

catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caagagttca ctgatggcag cttcacccag cccctgtaca gaggggagct gaatgagcac 2640

ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgaccttc 2700

cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760

cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttcac tgagaacatg 3060

gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120

taccgcttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggcc 3180

caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttcctg 3420

gtctacagca acaagtgcca gacccccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540

tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600

ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660

ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720

agaggcaact ccactggaac actcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca ccccacccac 3840

tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctcctgcagc 3900

atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960

ttcaccaaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccagggaagg 4020

agcaatgcct ggaggcccca ggtcaacaac ccaaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtcactgg ggtgaccacc caggggtca agagcctgct caccagcatg 4140

tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc 4200

cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac 4260

agcctggacc ccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac 4320

cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta ctga 4374

<210> 98
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 98
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcatcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaaggc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg ggcagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccatggc ctctgaccca ctctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gtcctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca ataccaccta cgtgaaccgc 2280

tccctgagcc agaatccacc tgtcctgaaa cgccaccaga gggagatcac caggaccacc 2340

ctccagtctg accaggagga gattgactat gatgacacca tttctgtgga gatgaagaaa 2400

gaggactttg acatctatga cgaggacgag aaccagagcc caaggagctt ccagaagaag 2460

accaggcact acttcattgc tgctgtggag cgcctgtggg actatggcat gagctccagc 2520

ccccatgtcc tcaggaacag ggcccagtct ggctctgtgc cacagttcaa gaaagtggtc 2580

ttccaagagt tcactgatgg cagcttcacc cagcccctgt acagagggga gctgaatgag 2640

cacctgggac tcctgggccc atacatcagg gctgaggtgg aggacaacat catggtgacc 2700

ttccgcaacc aggcctccag gccctacagc ttctacagct ccctcatcag ctatgaggag 2760

gaccagaggc agggggctga gccacgcaag aactttgtga aacccaatga aaccaagacc 2820

tacttctgga aagtccagca ccacatggcc cccaccaagg atgagtttga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggatgtgc actctggcct gattggccca 2940

ctcctggtct gccacaccaa caccctgaac cctgcccatg gaaggcaagt gactgtgcag 3000

gagtttgccc tcttcttcac catctttgat gaaaccaaga gctggtactt cactgagaac 3060

atggagcgca actgcagggc cccatgcaac attcagatgg aggaccccac cttcaaagag 3120

aactaccgct tccatgccat caatggctac atcatggaca ccctgcctgg gcttgtcatg 3180

gcccaggacc agaggatcag gtggtacctg ctttctatgg gctccaatga gaacattcac 3240

tccatccact tctctgggca tgtcttcact gtgcgcaaga aggaggagta caagatggcc 3300

ctgtacaacc tctaccctgg ggtctttgag actgtggaga tgctgccctc caaagctggc 3360

atctggaggg tggagtgcct cattggggag cacctgcatg ctggcatgag caccctgttc 3420

ctggtctaca gcaacaagtg ccagaccccc ctgggaatgg cctctggcca catcagggac 3480

ttccagatca ctgcctctgg ccagtatggc cagtgggccc ccaagctggc caggctccac 3540

tactctggat ccatcaatgc ctggagcacc aaggagccat tcagctggat caaagtggac 3600

ctgctggccc ccatgatcat ccatggcatc aagacccagg gggccaggca gaagttctcc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg atggcaagaa atggcagacc 3720

tacagaggca actccactgg aacactcatg gtcttctttg gcaatgtgga cagctctggc 3780

atcaagcaca acatcttcaa ccccccaatc atcgccagat acatcaggct gcaccccacc 3840

cactacagca tccgcagcac cctcaggatg gagctgatgg gctgtgacct gaactcctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atttctgatg cccagatcac tgcctccagc 3960

tacttcacca acatgtttgc cacctggagc ccaagcaagg ccaggctgca cctccaggga 4020

aggagcaatg cctggaggcc ccaggtcaac aacccaaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtcac tggggtgacc acccaggggg tcaagagcct gctcaccagc 4140

atgtatgtga aggagttcct gatcagctcc agccaggatg gccaccagtg gaccctcttc 4200

ttccagaatg gcaaggtcaa ggtgttccag ggcaaccagg acagcttcac ccctgtggtg 4260

aacagcctgg acccccccct cctgaccaga tacctgagga ttcaccccca gagctgggtc 4320

caccagattg ccctgaggat ggaggtcctg ggatgtgagg cccaggacct gtactga 4377

<210> 99
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 99
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gtcctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acaccaccta cgtgaaccgc 2280

tccctgagcc agaaccccccc cgtgctgaag aggcaccaga gggagatcac caggaccacc 2340

ctgcagagcg accaggagga gatcgactat gatgacacca tcagcgtgga gatgaagaag 2400

gaggacttcg acatctacga cgaggacgag aaccagagcc ccaggagctt ccagaagaag 2460

accaggcact acttcatcgc cgccgtggag aggctgtggg actatggcat gagcagcagc 2520

ccccacgtgc tgaggaacag ggcccagagc ggcagcgtgc cccagttcaa gaaggtggtg 2580

ttccaggagt tcaccgacgg cagcttcacc cagcccctgt acagaggcga gctgaacgag 2640

cacctgggcc tgctgggccc ctacatcagg gccgaggtgg aggacaacat catggtgacc 2700

ttcaggaacc aggccagcag gccctacagc ttctacagca gcctgatcag ctacgaggag 2760

gaccagaggc agggcgccga gcccaggaag aacttcgtga agcccaacga gaccaagacc 2820

tacttctgga aggtgcagca ccacatggcc cccaccaagg acgagttcga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggacgtgc acagcggcct gatcggcccc 2940

ctgctggtgt gccacaccaa caccctgaac cccgcccacg gcaggcaggt gaccgtgcag 3000

gagttcgccc tgttcttcac catcttcgac gagaccaaga gctggtactt caccgagaac 3060

atggagagga actgcagggc cccctgcaac atccagatgg aggaccccac cttcaaggag 3120

aactacaggt tccacgccat caacggctac atcatggaca ccctgcccgg cctggtgatg 3180

gcccaggacc agaggatcag gtggtatctg ctgagcatgg gcagcaacga gaacatccac 3240

agcatccact tcagcggcca cgtgttcacc gtgaggaaga aggaggagta caagatggcc 3300

ctgtacaacc tgtaccccgg cgtgttcgag accgtggaga tgctgcccag caaggccggc 3360

atctggaggg tggagtgcct gatcggcgag cacctgcacg ccggcatgag caccctgttc 3420

ctggtgtaca gcaacaagtg ccagaccccc ctgggcatgg ccagcggcca catcagggac 3480

ttccagatca ccgcctctgg ccagtacggc cagtgggccc ccaagctggc caggctgcac 3540

tacagcggca gcatcaacgc ctggagcacc aaggagccct tcagctggat caaggtggac 3600

ctgctggccc ccatgatcat ccacggcatc aagacccagg gcgccaggca gaagttcagc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg acggcaagaa gtggcagacc 3720

tacaggggca acagcaccgg caccctgatg gtgttcttcg gcaacgtgga cagcagcggc 3780

atcaagcaca acatcttcaa cccccccatc atcgccaggt acatcaggct gcaccccacc 3840

cactacagca tcaggagcac cctgcggatg gaactgatgg gctgcgacct gaacagctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atctctgacg cccagatcac cgccagcagc 3960

tacttcacca acatgttcgc cacctggagc cccagcaagg ccaggctgca cctgcagggc 4020

aggagcaacg cctggaggcc ccaggtgaac aaccccaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtgac cggcgtgacc acccagggcg tgaagagcct gctgaccagc 4140

atgtacgtga aggagttcct gatcagcagc agccaggacg gccaccagtg gaccctgttc 4200

ttccagaacg gcaaagtgaa ggtgttccag ggcaaccagg acagcttcac ccccgtggtg 4260

aacagcctgg acccccccct gctgaccagg tatctgagga tccaccccca gagctgggtg 4320

caccagatcg ccctgagaat ggaagtgctg ggatgcgagg cccaggacct gtactga 4377

<210> 100
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 100
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggccccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggccccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acaccaccta cgtgaaccgc 2280

tccctgagcc agaacccccc cgtgctgaag aggcaccaga gggagatcac caggaccacc 2340

ctgcagagcg accaggagga gatcgactat gatgacacca tcagcgtgga gatgaagaag 2400

gaggacttcg acatctacga cgaggacgag aaccagagcc ccaggagctt ccagaagaag 2460

accaggcact acttcatcgc cgccgtggag aggctgtggg actatggcat gagcagcagc 2520

ccccacgtgc tgaggaacag ggcccagagc ggcagcgtgc cccagttcaa gaaggtggtg 2580

ttccaggagt tcaccgacgg cagcttcacc cagcccctgt acagaggcga gctgaacgag 2640

cacctgggcc tgctgggccc ctacatcagg gccgaggtgg aggacaacat catggtgacc 2700

ttcaggaacc aggccagcag gccctacagc ttctacagca gcctgatcag ctacgaggag 2760

gaccagaggc agggcgccga gcccaggaag aacttcgtga agcccaacga gaccaagacc 2820

tacttctgga aggtgcagca ccacatggcc cccaccaagg acgagttcga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggacgtgc acagcggcct gatcggcccc 2940

ctgctggtgt gccacaccaa caccctgaac cccgcccacg gcaggcaggt gaccgtgcag 3000

gagttcgccc tgttcttcac catcttcgac gagaccaaga gctggtactt caccgagaac 3060

atggagagga actgcagggc cccctgcaac atccagatgg aggaccccac cttcaaggag 3120

aactacaggt tccacgccat caacggctac atcatggaca ccctgcccgg cctggtgatg 3180

gcccaggacc agaggatcag gtggtatctg ctgagcatgg gcagcaacga gaacatccac 3240

agcatccact tcagcggcca cgtgttcacc gtgaggaaga aggaggagta caagatggcc 3300

ctgtacaacc tgtaccccgg cgtgttcgag accgtggaga tgctgcccag caaggccggc 3360

atctggaggg tggagtgcct gatcggcgag cacctgcacg ccggcatgag caccctgttc 3420

ctggtgtaca gcaacaagtg ccagaccccc ctgggcatgg ccagcggcca catcagggac 3480

ttccagatca ccgcctctgg ccagtacggc cagtgggccc ccaagctggc caggctgcac 3540

tacagcggca gcatcaacgc ctggagcacc aaggagccct tcagctggat caaggtggac 3600

ctgctggccc ccatgatcat ccacggcatc aagacccagg gcgccaggca gaagttcagc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg acggcaagaa gtggcagacc 3720

tacaggggca acagcaccgg caccctgatg gtgttcttcg gcaacgtgga cagcagcggc 3780

atcaagcaca acatcttcaa ccccccccatc atcgccaggt acatcaggct gcaccccacc 3840

cactacagca tcaggagcac cctgcggatg gaactgatgg gctgcgacct gaacagctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atctctgacg cccagatcac cgccagcagc 3960

tacttcacca acatgttcgc cacctggagc cccagcaagg ccaggctgca cctgcagggc 4020

aggagcaacg cctggaggcc ccaggtgaac aaccccaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtgac cggcgtgacc acccagggcg tgaagagcct gctgaccagc 4140

atgtacgtga aggagttcct gatcagcagc agccaggacg gccaccagtg gaccctgttc 4200

ttccagaacg gcaaagtgaa ggtgttccag ggcaaccagg acagcttcac ccccgtggtg 4260

aacagcctgg acccccccct gctgaccagg tatctgagga tccaccccca gagctgggtg 4320

caccagatcg ccctgagaat ggaagtgctg ggatgcgagg cccaggacct gtactga 4377

<210> 101
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 101
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tggtcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaagtc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc aagagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccactgc cagcgacccc ccctgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgccccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gccccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccaggagc 2280

ttcagccaga accccccccgt gctgaagagg caccagaggg agatcaccag gaccaccctg 2340

cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400

gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460

aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc 2520

cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca gaggcgagct gaacgagcac 2640

ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700

aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac 2760

cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac 2820

ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cggcccccctg 2940

ctggtgtgcc acaccaacac cctgaacccc gcccacggca ggcaggtgac cgtgcaggag 3000

ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaacatg 3060

gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120

tacaggttcc acgccatcaa cggctacatc atggacaccc tgcccggcct ggtgatggcc 3180

caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240

atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300

tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc 3360

tggagggtgg agtgcctgat cggcgagcac ctgcacgccg gcatgagcac cctgttcctg 3420

gtgtacagca acaagtgcca gacccccctg ggcatggcca gcggccacat cagggacttc 3480

cagatcaccg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540

agcggcagca tcaacgcctg gagcaccaag gagcccttca gctggatcaa ggtggacctg 3600

ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc 3660

ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720

aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780

aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccacccac 3840

tacagcatca ggagcaccct gcggatggaa ctgatgggct gcgacctgaa cagctgcagc 3900

atgcccctgg gcatggagag caaggccatc tctgacgccc agatcaccgc cagcagctac 3960

ttcaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020

agcaacgcct ggaggcccca ggtgaacaac cccaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg 4140

tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctgttcttc 4200

cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaac 4260

agcctggacc ccccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320

cagatcgccc tgagaatgga agtgctggga tgcgaggccc aggacctgta ctga 4374

<210> 102
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 102
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tgatcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaaggc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc ggcagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccatggc cagcgacccc ctgtgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gtcctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acgccatcga gcccaggagc 2280

ttcagccaga acccccccgt gctgaagagg caccagaggg agatcaccag gaccaccctg 2340

cagagcgacc aggaggagat cgactatgat gacaccatca gcgtggagat gaagaaggag 2400

gacttcgaca tctacgacga ggacgagaac cagagcccca ggagcttcca gaagaagacc 2460

aggcactact tcatcgccgc cgtggagagg ctgtgggact atggcatgag cagcagcccc 2520

cacgtgctga ggaacagggc ccagagcggc agcgtgcccc agttcaagaa ggtggtgttc 2580

caggagttca ccgacggcag cttcacccag cccctgtaca gaggcgagct gaacgagcac 2640

ctgggcctgc tgggccccta catcagggcc gaggtggagg acaacatcat ggtgaccttc 2700

aggaaccagg ccagcaggcc ctacagcttc tacagcagcc tgatcagcta cgaggaggac 2760

cagaggcagg gcgccgagcc caggaagaac ttcgtgaagc ccaacgagac caagacctac 2820

ttctggaagg tgcagcacca catggccccc accaaggacg agttcgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gacgtgcaca gcggcctgat cggcccccctg 2940

ctggtgtgcc acaccaacac cctgaacccc gcccacggca ggcaggtgac cgtgcaggag 3000

ttcgccctgt tcttcaccat cttcgacgag accaagagct ggtacttcac cgagaacatg 3060

gagaggaact gcagggcccc ctgcaacatc cagatggagg accccacctt caaggagaac 3120

tacaggttcc acgccatcaa cggctacatc atggacaccc tgcccggcct ggtgatggcc 3180

caggaccaga ggatcaggtg gtatctgctg agcatgggca gcaacgagaa catccacagc 3240

atccacttca gcggccacgt gttcaccgtg aggaagaagg aggagtacaa gatggccctg 3300

tacaacctgt accccggcgt gttcgagacc gtggagatgc tgcccagcaa ggccggcatc 3360

tggagggtgg agtgcctgat cggcgagcac ctgcacgccg gcatgagcac cctgttcctg 3420

gtgtacagca acaagtgcca gacccccctg ggcatggcca gcggccacat cagggacttc 3480

cagatcaccg cctctggcca gtacggccag tgggccccca agctggccag gctgcactac 3540

agcggcagca tcaacgcctg gagcaccaag gagcccttca gctggatcaa ggtggacctg 3600

ctggccccca tgatcatcca cggcatcaag acccagggcg ccaggcagaa gttcagcagc 3660

ctgtacatca gccagttcat catcatgtac agcctggacg gcaagaagtg gcagacctac 3720

aggggcaaca gcaccggcac cctgatggtg ttcttcggca acgtggacag cagcggcatc 3780

aagcacaaca tcttcaaccc ccccatcatc gccaggtaca tcaggctgca ccccacccac 3840

tacagcatca ggagcaccct gcggatggaa ctgatgggct gcgacctgaa cagctgcagc 3900

atgcccctgg gcatggagag caaggccatc tctgacgccc agatcaccgc cagcagctac 3960

ttcaccaaca tgttcgccac ctggagcccc agcaaggcca ggctgcacct gcagggcagg 4020

agcaacgcct ggaggcccca ggtgaacaac cccaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtgaccgg cgtgaccacc cagggcgtga agagcctgct gaccagcatg 4140

tacgtgaagg agttcctgat cagcagcagc caggacggcc accagtggac cctgttcttc 4200

cagaacggca aagtgaaggt gttccagggc aaccaggaca gcttcacccc cgtggtgaac 4260

agcctggacc ccccccctgct gaccaggtat ctgaggatcc acccccagag ctgggtgcac 4320

cagatcgccc tgagaatgga agtgctggga tgcgaggccc aggacctgta ctga 4374

<210> 103
<211> 4377
<212> ADN
<213> Trình tự nhân tạo

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

<400> 103
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg cgccgtggag ctgagctggg actacatgca gtctgacctg 120

ggcgagctgc ctgtggacgc caggttcccc cccagagtgc ccaagagctt ccccttcaac 180

acctcagtgg tgtacaagaa gaccctgttc gtggagttca ccgaccacct gttcaacatc 240

gccaagccca ggcccccctg gatgggcctg ctgggcccca ccatccaggc cgaggtgtac 300

gacaccgtgg tggtcaccct gaagaacatg gccagccacc ccgtgagcct gcacgccgtg 360

ggcgtgagct actggaagtc ctctgagggc gccgagtatg acgaccagac cagccagagg 420

gagaaggagg acgacaaggt gttccccggc aagagccaca cctacgtgtg gcaggtgctg 480

aaggagaacg gccccactgc cagcgacccc ccctgcctga cctacagcta cctgagccac 540

gtggacctgg tgaaggacct gaactctggc ctgatcggcg ccctgctggt gtgcagggag 600

ggcagcctgg ccaaggagaa gacccagacc ctgcacaagt tcatcctgct gttcgccgtg 660

ttcgatgagg gcaagagctg gcacagcgag accaagaaca gcctgatgca ggacagggat 720

gccgcctctg ccagggcctg gcccaagatg cacaccgtga acggctacgt gaacaggagc 780

ctgcccggcc tgatcggctg ccacaggaag tctgtgtact ggcacgtgat cggcatgggc 840

accacccccg aggtgcacag catcttcctg gagggccaca ccttcctggt gaggaaccac 900

aggcaggcca gcctggagat cagccccatc accttcctga ccgcccagac cctgctgatg 960

gacctgggcc agttcctgct gttctgccac atcagcagcc accagcacga cggcatggag 1020

gcctacgtga aggtggacag ctgccccgag gagccccagc tgaggatgaa gaacaacgag 1080

gaggccgagg actatgatga tgacctgacc gactctgaga tggacgtggt gaggtttgat 1140

gatgacaaca gccccagctt catccagatc aggtctgtgg ccaagaagca ccccaagacc 1200

tgggtgcact acatcgccgc cgaggaggag gactgggact acgcccccct ggtgctggcc 1260

cccgacgaca ggagctacaa gagccagtac ctgaacaacg gcccccagag gatcggcagg 1320

aagtacaaga aggtcagatt catggcctac accgacgaga ccttcaagac cagggaggcc 1380

atccagcacg agtctggcat cctgggcccc ctgctgtacg gcgaggtggg cgacaccctg 1440

ctgatcatct tcaagaacca ggccagcagg ccctacaaca tctaccccca cggcatcacc 1500

gatgtgaggc ccctgtacag caggaggctg cccaagggcg tgaagcacct gaaggacttc 1560

cccatcctgc ccggcgagat cttcaagtac aagtggaccg tgaccgtgga ggatggcccc 1620

accaagtctg accccaggtg cctgaccagg tactacagca gcttcgtgaa catggagagg 1680

gacctggcct ctggcctgat cggcccccctg ctgatctgct acaaggagag cgtggaccag 1740

aggggcaacc agatcatgtc tgacaagagg aacgtgatcc tgttctctgt gttcgatgag 1800

aacaggagct ggtatctgac cgagaacatc cagaggttcc tgcccaaccc cgccggcgtg 1860

cagctggagg accccgagtt ccaggccagc aacatcatgc acagcatcaa cggctacgtg 1920

ttcgacagcc tgcagctgtc tgtgtgcctg cacgaggtgg cctactggta catcctgagc 1980

atcggcgccc agaccgactt cctgtctgtg ttcttctctg gctacacctt caagcacaag 2040

atggtgtacg aggacaccct gaccctgttc cccttcagcg gcgagaccgt gttcatgagc 2100

atggagaacc ccggcctgtg gatcctgggc tgccacaaca gcgacttcag gaacaggggc 2160

atgaccgccc tgctgaaagt cagcagctgc gacaagaaca ccggcgacta ctacgaggac 2220

agctacgagg acatcagcgc ctacctgctg agcaagaaca acaccaccta cgtgaaccgc 2280

tccctgagcc agaaccccccc cgtgctgaag aggcaccaga gggagatcac caggaccacc 2340

ctgcagagcg accaggagga gatcgactat gatgacacca tcagcgtgga gatgaagaag 2400

gaggacttcg acatctacga cgaggacgag aaccagagcc ccaggagctt ccagaagaag 2460

accaggcact acttcatcgc cgccgtggag aggctgtggg actatggcat gagcagcagc 2520

ccccacgtgc tgaggaacag ggcccagagc ggcagcgtgc cccagttcaa gaaggtggtg 2580

ttccaggagt tcaccgacgg cagcttcacc cagcccctgt acagaggcga gctgaacgag 2640

cacctgggcc tgctgggccc ctacatcagg gccgaggtgg aggacaacat catggtgacc 2700

ttcaggaacc aggccagcag gccctacagc ttctacagca gcctgatcag ctacgaggag 2760

gaccagaggc agggcgccga gcccaggaag aacttcgtga agcccaacga gaccaagacc 2820

tacttctgga aggtgcagca ccacatggcc cccaccaagg acgagttcga ctgcaaggcc 2880

tgggcctact tctctgatgt ggacctggag aaggacgtgc acagcggcct gatcggcccc 2940

ctgctggtgt gccacaccaa caccctgaac cccgcccacg gcaggcaggt gaccgtgcag 3000

gagttcgccc tgttcttcac catcttcgac gagaccaaga gctggtactt caccgagaac 3060

atggagagga actgcagggc cccctgcaac atccagatgg aggaccccac cttcaaggag 3120

aactacaggt tccacgccat caacggctac atcatggaca ccctgcccgg cctggtgatg 3180

gcccaggacc agaggatcag gtggtatctg ctgagcatgg gcagcaacga gaacatccac 3240

agcatccact tcagcggcca cgtgttcacc gtgaggaaga aggaggagta caagatggcc 3300

ctgtacaacc tgtaccccgg cgtgttcgag accgtggaga tgctgcccag caaggccggc 3360

atctggaggg tggagtgcct gatcggcgag cacctgcacg ccggcatgag caccctgttc 3420

ctggtgtaca gcaacaagtg ccagaccccc ctgggcatgg ccagcggcca catcagggac 3480

ttccagatca ccgcctctgg ccagtacggc cagtgggccc ccaagctggc caggctgcac 3540

tacagcggca gcatcaacgc ctggagcacc aaggagccct tcagctggat caaggtggac 3600

ctgctggccc ccatgatcat ccacggcatc aagacccagg gcgccaggca gaagttcagc 3660

agcctgtaca tcagccagtt catcatcatg tacagcctgg acggcaagaa gtggcagacc 3720

tacaggggca acagcaccgg caccctgatg gtgttcttcg gcaacgtgga cagcagcggc 3780

atcaagcaca acatcttcaa ccccccccatc atcgccaggt acatcaggct gcaccccacc 3840

cactacagca tcaggagcac cctgcggatg gaactgatgg gctgcgacct gaacagctgc 3900

agcatgcccc tgggcatgga gagcaaggcc atctctgacg cccagatcac cgccagcagc 3960

tacttcacca acatgttcgc cacctggagc cccagcaagg ccaggctgca cctgcagggc 4020

aggagcaacg cctggaggcc ccaggtgaac aaccccaagg agtggctgca ggtggacttc 4080

cagaagacca tgaaggtgac cggcgtgacc acccagggcg tgaagagcct gctgaccagc 4140

atgtacgtga aggagttcct gatcagcagc agccaggacg gccaccagtg gaccctgttc 4200

ttccagaacg gcaaagtgaa ggtgttccag ggcaaccagg acagcttcac ccccgtggtg 4260

aacagcctgg acccccccct gctgaccagg tatctgagga tccaccccca gagctgggtg 4320

caccagatcg ccctgagaat ggaagtgctg ggatgcgagg cccaggacct gtactga 4377

<210> 104
<211> 1458
<212> PRT
<213> Trình tự nhân tạo

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

<400> 104
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65 70 75 80

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

Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala Ser
100 105 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ser Ser
115 120 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130 135 140

Asp Lys Val Phe Pro Gly Lys Ser His Thr Tyr Val Trp Gln Val Leu
145 150 155 160

Lys Glu Asn Gly Pro Thr Ala Ser Asp Pro Pro Cys Leu Thr Tyr Ser
165 170 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
195 200 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
210 215 220

```
Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
                405                 410                 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
            420                 425                 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
        435                 440                 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
    450                 455                 460
```

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465 470 475 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
485 490 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly

705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Thr Thr Tyr Val Asn Arg Ser Leu Ser Gln Asn Pro Pro Val
755 760 765

Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp
770 775 780

Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys
785 790 795 800

Glu Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser
805 810 815

Phe Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu
820 825 830

Trp Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala
835 840 845

Gln Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe
850 855 860

Thr Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu
865 870 875 880

His Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn
885 890 895

Ile Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr
900 905 910

Ser Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro
915 920 925

Arg Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
930 935 940

Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala
945 950 955 960

Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly
965 970 975

Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala
980 985 990

His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile
995 1000 1005

Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg
1010 1015 1020

Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe
1025 1030 1035

Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp
1040 1045 1050

Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp
1055 1060 1065

Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His
1070 1075 1080

Phe Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys
1085 1090 1095

Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu
1100 1105 1110

Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile
1115 1120 1125

Gly Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr
1130 1135 1140

Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile
1145 1150 1155

Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala
1160 1165 1170

Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp
1175 1180 1185

Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala
1190 1195 1200

Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys
1205 1210 1215

Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu
1220 1225 1230

Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr
1235 1240 1245

Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His
1250 1255 1260

Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His
1265 1270 1275

Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met
1280 1285 1290

Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser
1295 1300 1305

Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr
1310 1315 1320

Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu
1325 1330 1335

Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys
1340 1345 1350

Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly
1355 1360 1365

Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val
1370 1375 1380

Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr
1385 1390 1395

Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
1400 1405 1410

Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
1415 1420 1425

Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
1430 1435 1440

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 105
<211> 1458
<212> PRT
<213> Trình tự nhân tạo

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

<400> 105
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65 70 75 80

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

Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser
100 105 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser
115 120 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
130 135 140

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu
145 150 155 160

```
Lys Glu Asn Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His
                325                 330                 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
            340                 345                 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
        355                 360                 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
    370                 375                 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385                 390                 395                 400
```

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
405 410 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
420 425 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
435 440 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
450 455 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465 470 475 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
485 490 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp

645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Thr Thr Tyr Val Asn Arg Ser Leu Ser Gln Asn Pro Pro Val
755 760 765

Leu Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp
770 775 780

Gln Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys
785 790 795 800

Glu Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser
805 810 815

Phe Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu
820 825 830

Trp Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala
835 840 845

Gln Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe
850 855 860

Thr Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu
865 870 875 880

His Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn
885 890 895

Ile Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr
900 905 910

Ser Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro
915 920 925

Arg Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys
930 935 940

Val Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala
945 950 955 960

Trp Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly
965 970 975

Leu Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala
980 985 990

His Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile
995 1000 1005

Phe Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg
1010 1015 1020

Asn Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe
1025 1030 1035

Lys Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp
1040 1045 1050

Thr Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp
1055 1060 1065

Tyr Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His
1070 1075 1080

Phe Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys
1085 1090 1095

Met Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu
1100 1105 1110

Met Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile
1115 1120 1125

Gly Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr
1130 1135 1140

Ser Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile
1145 1150 1155

Arg Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala
1160 1165 1170

Pro Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp
1175 1180 1185

Ser Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala
1190 1195 1200

Pro Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys
1205 1210 1215

Phe Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu
1220 1225 1230

Asp Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr
1235 1240 1245

Leu Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His
1250 1255 1260

Asn Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His
1265 1270 1275

Pro Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met
1280 1285 1290

Gly Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser
1295 1300 1305

Lys Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr
1310 1315 1320

Asn Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu
1325 1330 1335

Gln Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys
1340 1345 1350

Glu Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly
1355 1360 1365

Val Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val
1370 1375 1380

Lys Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr
1385 1390 1395

Leu Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln
1400 1405 1410

Asp Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu
1415 1420 1425

Thr Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile
1430 1435 1440

Ala Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 106
<211> 1457
<212> PRT
<213> Trình tự nhân tạo

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

<400> 106
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
35 40 45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
50 55 60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65 70 75 80

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

```
Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ser Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Lys Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Thr Ala Ser Asp Pro Pro Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
        275                 280                 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
    290                 295                 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305                 310                 315                 320

Asp Leu Gly Gln Phe Leu Leu Ser Cys His Ile Ser Ser His Gln His
                325                 330                 335
```

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
370 375 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385 390 395 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
405 410 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
420 425 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
435 440 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
450 455 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465 470 475 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
485 490 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe
515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val

580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
755 760 765

Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
770 775 780

Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
785 790 795 800

Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
805 810 815

Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
820 825 830

Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
835 840 845

Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
850 855 860

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
865 870 875 880

Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
885 890 895

Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
900 905 910

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
915 920 925

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
930 935 940

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
980 985 990

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
995 1000 1005

Asp Glu Thr Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn
1010 1015 1020

Cys Arg Ala Pro Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys
1025 1030 1035

Glu Asn Tyr Arg Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr
1040 1045 1050

Leu Pro Gly Leu Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr
1055 1060 1065

Leu Leu Ser Met Gly Ser Asn Glu Asn Ile His Ser Ile His Phe
1070 1075 1080

Ser Gly His Val Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met
1085 1090 1095

Ala Leu Tyr Asn Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met
1100 1105 1110

Leu Pro Ser Lys Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly
1115 1120 1125

Glu His Leu His Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser
1130 1135 1140

Asn Lys Cys Gln Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg
1145 1150 1155

Asp Phe Gln Ile Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro
1160 1165 1170

Lys Leu Ala Arg Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser
1175 1180 1185

Thr Lys Glu Pro Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro
1190 1195 1200

Met Ile Ile His Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe
1205 1210 1215

Ser Ser Leu Tyr Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp
1220 1225 1230

Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu
1235 1240 1245

Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn
1250 1255 1260

Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro
1265 1270 1275

Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly
1280 1285 1290

Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys
1295 1300 1305

Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn
1310 1315 1320

Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln
1325 1330 1335

Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu
1340 1345 1350

Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val
1355 1360 1365

Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys
1370 1375 1380

Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu
1385 1390 1395

Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp
1400 1405 1410

Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr
1415 1420 1425

Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala
1430 1435 1440

Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 107
<211> 1457
<212> PRT
<213> Trình tự nhân tạo

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

<400> 107
Met Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe
1 5 10 15

Cys Phe Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser
20 25 30

```
Trp Asp Tyr Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg
        35                  40                  45

Phe Pro Pro Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val
    50                  55                  60

Tyr Lys Lys Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile
65                  70                  75                  80

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

Ala Glu Val Tyr Asp Thr Val Val Val Thr Leu Lys Asn Met Ala Ser
            100                 105                 110

His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ser Ser
        115                 120                 125

Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp
    130                 135                 140

Asp Lys Val Phe Pro Gly Lys Ser His Thr Tyr Val Trp Gln Val Leu
145                 150                 155                 160

Lys Glu Asn Gly Pro Thr Ala Ser Asp Pro Pro Cys Leu Thr Tyr Ser
                165                 170                 175

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

Gly Ala Leu Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr
        195                 200                 205

Gln Thr Leu His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly
    210                 215                 220

Lys Ser Trp His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp
225                 230                 235                 240

Ala Ala Ser Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr
                245                 250                 255

Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val
            260                 265                 270
```

Tyr Trp His Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile
275 280 285

Phe Leu Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser
290 295 300

Leu Glu Ile Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met
305 310 315 320

Asp Leu Gly Gln Phe Leu Leu Ser Cys His Ile Ser Ser His Gln His
325 330 335

Asp Gly Met Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro
340 345 350

Gln Leu Arg Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp
355 360 365

Leu Thr Asp Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser
370 375 380

Pro Ser Phe Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr
385 390 395 400

Trp Val His Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro
405 410 415

Leu Val Leu Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn
420 425 430

Asn Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
435 440 445

Ala Tyr Thr Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu
450 455 460

Ser Gly Ile Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu
465 470 475 480

Leu Ile Ile Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro
485 490 495

His Gly Ile Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys
500 505 510

Gly Val Lys His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe

515 520 525

Lys Tyr Lys Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp
530 535 540

Pro Arg Cys Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg
545 550 555 560

Asp Leu Ala Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu
565 570 575

Ser Val Asp Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val
580 585 590

Ile Leu Phe Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu
595 600 605

Asn Ile Gln Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp
610 615 620

Pro Glu Phe Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val
625 630 635 640

Phe Asp Ser Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp
645 650 655

Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe
660 665 670

Ser Gly Tyr Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr
675 680 685

Leu Phe Pro Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro
690 695 700

Gly Leu Trp Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly
705 710 715 720

Met Thr Ala Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp
725 730 735

Tyr Tyr Glu Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys
740 745 750

Asn Asn Ala Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu
755 760 765

Lys Arg His Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln
770 775 780

Glu Glu Ile Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu
785 790 795 800

Asp Phe Asp Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe
805 810 815

Gln Lys Lys Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp
820 825 830

Asp Tyr Gly Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln
835 840 845

Ser Gly Ser Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr
850 855 860

Asp Gly Ser Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His
865 870 875 880

Leu Gly Leu Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile
885 890 895

Met Val Thr Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser
900 905 910

Ser Leu Ile Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg
915 920 925

Lys Asn Phe Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val
930 935 940

Gln His His Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp
945 950 955 960

Ala Tyr Phe Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu
965 970 975

Ile Gly Pro Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His
980 985 990

Gly Arg Gln Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe
995 1000 1005

```
Asp Glu  Thr Lys Ser Trp Tyr  Phe Thr Glu Asn Met  Glu Arg Asn
    1010                1015                1020

Cys Arg  Ala Pro Cys Asn Ile  Gln Met Glu Asp Pro  Thr Phe Lys
    1025                1030                1035

Glu Asn  Tyr Arg Phe His Ala  Ile Asn Gly Tyr Ile  Met Asp Thr
    1040                1045                1050

Leu Pro  Gly Leu Val Met Ala  Gln Asp Gln Arg Ile  Arg Trp Tyr
    1055                1060                1065

Leu Leu  Ser Met Gly Ser Asn  Glu Asn Ile His Ser  Ile His Phe
    1070                1075                1080

Ser Gly  His Val Phe Thr Val  Arg Lys Lys Glu Glu  Tyr Lys Met
    1085                1090                1095

Ala Leu  Tyr Asn Leu Tyr Pro  Gly Val Phe Glu Thr  Val Glu Met
    1100                1105                1110

Leu Pro  Ser Lys Ala Gly Ile  Trp Arg Val Glu Cys  Leu Ile Gly
    1115                1120                1125

Glu His  Leu His Ala Gly Met  Ser Thr Leu Phe Leu  Val Tyr Ser
    1130                1135                1140

Asn Lys  Cys Gln Thr Pro Leu  Gly Met Ala Ser Gly  His Ile Arg
    1145                1150                1155

Asp Phe  Gln Ile Thr Ala Ser  Gly Gln Tyr Gly Gln  Trp Ala Pro
    1160                1165                1170

Lys Leu  Ala Arg Leu His Tyr  Ser Gly Ser Ile Asn  Ala Trp Ser
    1175                1180                1185

Thr Lys  Glu Pro Phe Ser Trp  Ile Lys Val Asp Leu  Leu Ala Pro
    1190                1195                1200

Met Ile  Ile His Gly Ile Lys  Thr Gln Gly Ala Arg  Gln Lys Phe
    1205                1210                1215

Ser Ser  Leu Tyr Ile Ser Gln  Phe Ile Ile Met Tyr  Ser Leu Asp
    1220                1225                1230
```

Gly Lys Lys Trp Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu
1235 1240 1245

Met Val Phe Phe Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn
1250 1255 1260

Ile Phe Asn Pro Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro
1265 1270 1275

Thr His Tyr Ser Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly
1280 1285 1290

Cys Asp Leu Asn Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys
1295 1300 1305

Ala Ile Ser Asp Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn
1310 1315 1320

Met Phe Ala Thr Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln
1325 1330 1335

Gly Arg Ser Asn Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu
1340 1345 1350

Trp Leu Gln Val Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val
1355 1360 1365

Thr Thr Gln Gly Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys
1370 1375 1380

Glu Phe Leu Ile Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu
1385 1390 1395

Phe Phe Gln Asn Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp
1400 1405 1410

Ser Phe Thr Pro Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr
1415 1420 1425

Arg Tyr Leu Arg Ile His Pro Gln Ser Trp Val His Gln Ile Ala
1430 1435 1440

Leu Arg Met Glu Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1445 1450 1455

<210> 108

<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 108
atgcagattg agctgtccac ctgcttcttt ctgtgcctgc tgagattctg cttctctgcc 60

accaggagat actacctggg ggctgtggaa ctttcttggg actacatgca gtctgacctg 120

ggagagctgc ctgtggatgc caggttccca cccagagtgc ccaagtcctt cccattcaac 180

acctctgtgg tctacaagaa gacactcttt gtggaattca ctgaccacct gttcaacatt 240

gcaaaaccca gaccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcatcccacc ctgtgtctct gcatgctgtg 360

ggagtctcat actggaaatc ctctgaaggg gctgagtatg atgaccagac atcccagaga 420

gagaaagagg atgacaaggt gttccctggg aagtctcaca cctatgtgtg gcaagtcctc 480

aaggagaatg gacccactgc atctgaccca ccctgcctga catactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg cactgctggt gtgcagggaa 600

ggatccctgg ccaaggagaa aacccagaca ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagtcttg gcactctgaa acaaagaact ccctgatgca agacagggat 720

gctgcctctg ccagggcatg gcccaagatg cacactgtga atggctatgt gaacagatca 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aagtgcactc cattttcctg gagggacaca ccttcctggt caggaaccac 900

agacaagcct ctctggagat ctctcccatc accttcctca ctgcacagac actgctgatg 960

gaccttggac agttcctgct gtcctgccac atctcttccc accagcatga tggcatggaa 1020

gcctatgtca aggtggactc atgccctgag gaaccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt cagatttgat 1140

gatgacaact ctccatcctt cattcagatc aggtctgtgg caaagaaaca ccccaagaca 1200

tgggtgcact acattgctgc tgaggaagag gactgggact atgcaccact ggtcctggcc 1260

cctgatgaca ggagctacaa gtctcagtac ctcaacaatg gcccacaaag aattggaaga 1320

aagtacaaga aagtcagatt catggcctac actgatgaaa ccttcaagac aagagaagcc 1380

attcagcatg agtctggcat tctgggacca ctcctgtatg gggaagtggg agacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag caggagactg ccaaaagggg tgaaacacct caaggacttc 1560

cccattctgc ctggagagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

acaaagtctg accccaggtg cctcaccaga tactactcct cttttgtgaa catggagaga 1680

gacctggcat ctggactgat tggaccactg ctcatctgct acaaggagtc tgtggaccag 1740

agaggcaacc agatcatgtc tgacaagaga aatgtgattc tgttctctgt ctttgatgag 1800

aacagatcat ggtacctgac tgagaacatt cagagattcc tgcccaaccc tgctggggtg 1860

caactggaag accctgagtt ccaggcaagc aacatcatgc actccatcaa tggctatgtg 1920

tttgactctc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggcac aaactgactt cctttctgtc ttcttctctg gatacacctt caagcacaag 2040

atggtgtatg aggacaccct gacactcttc ccattctctg gggaaactgt gttcatgagc 2100

atggagaacc ctggactgtg gattctggga tgccacaact ctgacttcag aaacagggga 2160

atgactgcac tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

tcttatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccagaagc 2280

ttctctcaga atccacctgt cctgaagaga caccagagag agatcaccag gacaaccctc 2340

cagtctgacc aggaagagat tgactatgat gacaccattt ctgtggagat gaagaaggag 2400

gactttgaca tctatgatga ggacgagaac cagtctccaa gatcattcca gaagaagaca 2460

agacactact tcattgctgc tgtggaaaga ctgtgggact atggcatgtc ttcctctccc 2520

catgtcctca ggaacagggc acagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caggagttca ctgatggctc attcacccag cccctgtaca gaggggaact gaatgagcac 2640

ctgggactcc tgggaccata catcagggct gaggtggaag acaacatcat ggtgacattc 2700

agaaaccagg cctccaggcc ctacagcttc tactcttccc tcatcagcta tgaggaagac 2760

cagagacaag gggctgagcc aagaaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggcaccc accaaggatg agtttgactg caaggcctgg 2880

gcatacttct ctgatgtgga cctggagaaa gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcacatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagtcat ggtacttcac tgagaacatg 3060

gagagaaact gcagagcacc atgcaacatt cagatggaag accccacctt caaggagaac 3120

tacaggttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggca 3180

caggaccaga gaatcagatg gtacctgctt tctatgggat ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg agaaagaagg aggaatacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg aatgcctcat tggggagcac ctgcatgctg gcatgtcaac cctgttcctg 3420

gtctacagca acaagtgcca gacacccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggcaccca aactggccag gctccactac 3540

tctggctcca tcaatgcatg gtcaaccaag gagccattct cttggatcaa ggtggacctg 3600

ctggcaccca tgatcattca tggcatcaag acacaggggg caagacagaa attctcctct 3660

ctgtacatct cacagttcat catcatgtac tctctggatg gcaagaagtg gcagacatac 3720

agaggcaact ccactggcac cctcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc tcccatcatt gccagataca tcaggctgca ccccacccac 3840

tactcaatca gatcaaccct caggatggaa ctgatgggat gtgacctgaa ctcctgctca 3900

atgcccctgg gaatggagag caaggccatt tctgatgccc agatcactgc atcctcttac 3960

ttcaccaaca tgtttgccac ctggtcacca tcaaaagcca ggctgcacct ccagggaaga 4020

agcaatgcct ggagacccca ggtcaacaac ccaaaggaat ggctgcaagt ggacttccag 4080

aagacaatga aagtcactgg ggtgacaacc caggggtca agtctctgct cacctcaatg 4140

tatgtgaagg agttcctgat ctcttcctca caggatggcc accagtggac actcttcttc 4200

cagaatggca aagtcaaggt gttccagggc aaccaggact ctttcacacc tgtggtgaac 4260

tcactggacc ccccctcct gacaagatac ctgagaattc acccccagtc ttgggtccac 4320

cagattgccc tgagaatgga agtcctggga tgtgaggcac aagacctgta ctga 4374

<210> 109
<211> 4374
<212> ADN
<213> Trình tự nhân tạo

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

<400> 109
atgcagattg agctgagcac ctgcttcttc ctgtgcctgc tgaggttctg cttctctgcc 60

accaggagat actacctggg ggctgtggag ctttcttggg actacatgca gtctgacctg 120

ggggagctgc ctgtggatgc caggttccca cccagagtgc ccaaatcctt cccattcaac 180

acctctgtgg tctacaagaa gaccctcttt gtggagttca ctgaccacct gttcaacatt 240

gccaaaccca ggccaccctg gatgggactc ctgggaccca ccattcaggc tgaggtgtat 300

gacactgtgg tcgtcaccct caagaacatg gcctcccacc ctgtgagcct gcatgctgtg 360

ggggtcagct actggaagtc ctctgagggg gctgagtatg atgaccagac ctcccagagg 420

gagaaggagg atgacaaagt gttccctggg aagagccaca cctatgtgtg gcaggtcctc 480

aaggagaatg gccccactgc ctctgaccca ccctgcctga cctactccta cctttctcat 540

gtggacctgg tcaaggacct caactctgga ctgattgggg ccctgctggt gtgcagggag 600

ggctccctgg ccaaagagaa gacccagacc ctgcacaagt tcattctcct gtttgctgtc 660

tttgatgagg gcaagagctg gcactctgaa accaagaact ccctgatgca ggacagggat 720

gctgcctctg ccagggcctg gcccaagatg cacactgtga atggctatgt gaacaggagc 780

ctgcctggac tcattggctg ccacaggaaa tctgtctact ggcatgtgat tggcatgggg 840

acaacccctg aggtgcactc cattttcctg gagggccaca ccttcctggt caggaaccac 900

agacaggcca gcctggagat cagccccatc accttcctca ctgcccagac cctgctgatg 960

gacctcggac agttcctgct gtcctgccac atcagctccc accagcatga tggcatggag 1020

gcctatgtca aggtggacag ctgccctgag gagccacagc tcaggatgaa gaacaatgag 1080

gaggctgagg actatgatga tgacctgact gactctgaga tggatgtggt ccgctttgat 1140

gatgacaaca gcccatcctt cattcagatc aggtctgtgg ccaagaaaca ccccaagacc 1200

tgggtgcact acattgctgc tgaggaggag gactgggact atgccccact ggtcctggcc 1260

cctgatgaca ggagctacaa gagccagtac ctcaacaatg gcccacagag gattggacgc 1320

aagtacaaga aagtcaggtt catggcctac actgatgaaa ccttcaagac cagggaggcc 1380

attcagcatg agtctggcat cctgggccca ctcctgtatg gggaggtggg ggacaccctg 1440

ctcatcatct tcaagaacca ggcctccagg ccctacaaca tctacccaca tggcatcact 1500

gatgtcaggc ccctgtacag ccgcaggctg ccaaaggggg tgaaacacct caaggacttc 1560

cccattctgc ctggggagat cttcaagtac aagtggactg tcactgtgga ggatggacca 1620

accaaatctg accccaggtg cctcaccaga tactactcca gctttgtgaa catggagagg 1680

gacctggcct ctggcctgat tggcccactg ctcatctgct acaaggagtc tgtggaccag 1740

aggggaaacc agatcatgtc tgacaagagg aatgtgattc tgttctctgt ctttgatgag 1800

aacaggagct ggtacctgac tgagaacatt cagcgcttcc tgcccaaccc tgctggggtg 1860

cagctggagg accctgagtt ccaggccagc aacatcatgc actccatcaa tggctatgtg 1920

tttgacagcc tccagctttc tgtctgcctg catgaggtgg cctactggta cattctttct 1980

attggggccc agactgactt cctttctgtc ttcttctctg gctacacctt caaacacaag 2040

atggtgtatg aggacaccct gaccctcttc ccattctctg gggagactgt gttcatgagc 2100

atggagaacc ctggcctgtg gattctggga tgccacaact ctgacttccg caacaggggc 2160

atgactgccc tgctcaaagt ctcctcctgt gacaagaaca ctggggacta ctatgaggac 2220

agctatgagg acatctctgc ctacctgctc agcaagaaca atgccattga gcccaggagc 2280

ttcagccaga atccacctgt cctgaaacgc caccagaggg agatcaccag gaccaccctc 2340

cagtctgacc aggaggagat tgactatgat gacaccattt ctgtggagat gaagaaagag 2400

gactttgaca tctatgacga ggacgagaac cagagcccaa ggagcttcca gaagaagacc 2460

aggcactact tcattgctgc tgtggagcgc ctgtgggact atggcatgag ctccagcccc 2520

catgtcctca ggaacagggc ccagtctggc tctgtgccac agttcaagaa agtggtcttc 2580

caagagttca ctgatggcag cttcacccag cccctgtaca gaggggagct gaatgagcac 2640

ctgggactcc tgggcccata catcagggct gaggtggagg acaacatcat ggtgaccttc 2700

cgcaaccagg cctccaggcc ctacagcttc tacagctccc tcatcagcta tgaggaggac 2760

cagaggcagg gggctgagcc acgcaagaac tttgtgaaac ccaatgaaac caagacctac 2820

ttctggaaag tccagcacca catggccccc accaaggatg agtttgactg caaggcctgg 2880

gcctacttct ctgatgtgga cctggagaag gatgtgcact ctggcctgat tggcccactc 2940

ctggtctgcc acaccaacac cctgaaccct gcccatggaa ggcaagtgac tgtgcaggag 3000

tttgccctct tcttcaccat ctttgatgaa accaagagct ggtacttcac tgagaacatg 3060

gagcgcaact gcagggcccc atgcaacatt cagatggagg accccacctt caaagagaac 3120

taccgcttcc atgccatcaa tggctacatc atggacaccc tgcctgggct tgtcatggcc 3180

caggaccaga ggatcaggtg gtacctgctt tctatgggct ccaatgagaa cattcactcc 3240

atccacttct ctgggcatgt cttcactgtg cgcaagaagg aggagtacaa gatggccctg 3300

tacaacctct accctggggt ctttgagact gtggagatgc tgccctccaa agctggcatc 3360

tggagggtgg agtgcctcat tggggagcac ctgcatgctg gcatgagcac cctgttcctg 3420

gtctacagca acaagtgcca gacccccctg ggaatggcct ctggccacat cagggacttc 3480

cagatcactg cctctggcca gtatggccag tgggccccca agctggccag gctccactac 3540

tctggatcca tcaatgcctg gagcaccaag gagccattca gctggatcaa agtggacctg 3600

ctggccccca tgatcatcca tggcatcaag acccaggggg ccaggcagaa gttctccagc 3660

ctgtacatca gccagttcat catcatgtac agcctggatg gcaagaaatg gcagacctac 3720

agaggcaact ccactggaac actcatggtc ttctttggca atgtggacag ctctggcatc 3780

aagcacaaca tcttcaaccc cccaatcatc gccagataca tcaggctgca ccccacccac 3840

tacagcatcc gcagcaccct caggatggag ctgatgggct gtgacctgaa ctcctgcagc 3900

atgcccctgg gcatggagag caaggccatt tctgatgccc agatcactgc ctccagctac 3960

ttcaccaaca tgtttgccac ctggagccca agcaaggcca ggctgcacct ccagggaagg 4020

agcaatgcct ggaggcccca ggtcaacaac ccaaaggagt ggctgcaggt ggacttccag 4080

aagaccatga aggtcactgg ggtgaccacc caggggtca agagcctgct caccagcatg 4140

tatgtgaagg agttcctgat cagctccagc caggatggcc accagtggac cctcttcttc 4200

cagaatggca aggtcaaggt gttccagggc aaccaggaca gcttcacccc tgtggtgaac 4260

agcctggacc ccccccctcct gaccagatac ctgaggattc acccccagag ctgggtccac 4320

cagattgccc tgaggatgga ggtcctggga tgtgaggccc aggacctgta ctga 4374