



(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)



1-0035667

(51)⁷ C12N 15/87; C07H 21/04; C12N 5/04; (13) B
A01H 5/00; C12N 15/00

(21) 1-2019-04132

(22) 11/01/2018

(86) PCT/US2018/013298 11/01/2018

(87) WO2018/132556 19/07/2018

(30) 62/445,313 12/01/2017 US

(45) 25/05/2023 422

(43) 25/10/2019 379A

(73) MONSANTO TECHNOLOGY LLC (US)

800 North Lindbergh Boulevard, St. Louis, MO 63167, United States of America

(72) BAUM, James, A. (US); BOWEN, David, J. (US); CHAY, Catherine, A. (US); CHI, David, J. (US); CLINTON, William, P. (US); DART, Crystal, L. (US); ENGLISH, Leigh (US); FLASINSKI, Stanislaw (US); GUZOV, Victor, M. (US); JARRELL, Kevin, A. (US); KESANAPALLI, Uma, R. (US); MALVAR, Thomas, M. (US); MCCARROLL, Robert, M. (US); MILLIGAN, Jason, S. (US); MORGENSTERN, Jay, P. (US); RUCKER, Deborah, G. (US); SALVADOR, Sara, A. (US); SMITH, Temple, F. (US); SOTO, Carlos, E. (US); STULTZ, Collin, M. (US); TURCZYK, Brian, M. (US); VAUGHN, Ty, T. (US); VON RECHENBERG, Moritz, W., F., F. (DE).

(74) Công ty Luật TNHH T&G (TGVN)

(54) CẤU TRÚC POLYNUCLEOTIT, PHƯƠNG PHÁP SẢN XUẤT HẠT CHỨA CẤU TRÚC POLYNUCLEOTIT NÀY, PROTEIN GÂY ĐỘC CHO LOÀI LEPIDOPTERAN VÀ CHẾ PHẨM CHỨA PROTEIN NÀY ĐỂ PHÒNG TRỪ LOÀI GÂY HẠI LEPIDOPTERAN

(57) Sáng chế đề cập đến các cấu trúc polynucleotit chứa trình tự nucleotit mã hóa protein diệt côn trùng thể hiện hoạt tính ức chế *Lepidopteran*. Sáng chế cũng đề cập đến vector, tế bào chủ, cây và hạt chứa cấu trúc polynucleotit này. Sáng chế cũng đề cập đến phương pháp sản xuất hạt chứa cấu trúc polynucleotit này. Sáng chế cũng đề cập đến các protein gây độc cho loài *Lepidopteran* được gọi ở đây là protein gây độc BCW 001, BCW 002, BCW 003, và BCW và chế phẩm diệt côn trùng chứa protein này.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập chung đến lĩnh vực protein ức chế côn trùng, cụ thể là đến các protein thể hiện tác dụng ức chế côn trùng chống lại vật gây hại *Lepidopteran* liên quan đến nông nghiệp ở cây trồng và hạt giống, cụ thể là vật gây hại *Lepidopteran* như sâu xám (“black cutworm - BCW”, *Agrotis ipsilon*).

Tình trạng kỹ thuật của sáng chế

Protein ức chế côn trùng được sản xuất bởi loài vi khuẩn *Bacillus thuringiensis* (*Bt*) đã được biết trong lĩnh vực này. Một số protein *Bt* có thể được sử dụng để phòng trừ vật gây hại trong nông nghiệp ở cây trồng bằng cách phun các chế phẩm được chấp nhận trong nông nghiệp chứa một hoặc nhiều protein này lên thực vật, phủ hạt bằng chế phẩm được bào chế để chứa lượng có tác dụng diệt côn trùng của các protein này, hoặc bằng cách biểu hiện một hoặc nhiều protein hữu hiệu này ở cây/hạt.

Chỉ một số protein *Bt* đã được phát triển làm tính trạng chuyển gen để nông dân sử dụng trong ứng dụng thương mại nhằm phòng trừ côn trùng gây hại. Nông dân dựa trên các protein này để tạo ra phổ phòng trừ vật gây hại được chỉ định, và có thể tiếp tục dựa vào các hóa chất phổ rộng trong các ứng dụng cho lá và đất để phòng trừ vật gây hại. Một số côn trùng *Lepidopteran*, như loài *Agrotis* và loài *Striacosta*, đã được chứng minh là đặc biệt khó phòng trừ bằng cách sử dụng các tính trạng diệt côn trùng chuyển gen đang được sử dụng hiện nay bao gồm Cry1Ab, Cry1Ac, Cry1Fa, Cry2Ab, Cry2Ae, VIP3Aa, và nhiều độc tố *Bt* khác nhau khác mà được sử dụng ít thường xuyên hơn. Do đó, vẫn có nhu cầu về protein ức chế côn trùng mà có tác dụng chống lại phổ các loài côn trùng gây hại rộng hơn, và để sử dụng trong các độc tố để sử dụng trong việc khắc phục sự phát triển tính kháng của sâu bệnh với thuốc diệt côn trùng hiện nay, bao gồm các độc tố được sử dụng hiện nay trong các hệ thống quản lý sâu bệnh.

Sáng chế mô tả họ protein mới, các biến thể, và các cấu trúc protein độc tố khả

mà mỗi loại này thể hiện tác dụng diệt côn trùng hiệu quả bất ngờ kháng *Lepidoptera*, cụ thể là kháng vật gây hại thuộc loài *Agrotis*, như sâu xám.

Bản chất kỹ thuật của sáng chế

Nhóm mới của các polypeptit ức chế côn trùng (protein độc tố BCW 001, BCW 002 và BCW 003 và các mảnh có tác dụng diệt loài gây hại của chúng) được chỉ ra là thể hiện tác dụng ức chế chống lại một số vật gây hại *Lepidopteran* ở cây trồng, cụ thể là chống lại loài sâu xám (loài *Agrotis*). Mỗi protein này có thể được sử dụng một mình hoặc kết hợp với nhau và với các protein *Bt* khác và các chất ức chế côn trùng khác trong các chế phẩm và ở thực vật (*in planta*), do đó cung cấp giải pháp thay thế cho các protein *Bt* và các hóa chất diệt côn trùng hiện đang được sử dụng trong các hệ thống nông nghiệp.

Sáng chế đề xuất các cấu trúc polynucleotit chứa, trong mỗi liên kết hoạt động, đoạn trình tự khởi đầu khác loại được liên kết với trình tự nucleotit mã hóa protein diệt côn trùng có các đặc điểm Cry1A mà ngắn hơn chiều dài đầy đủ so với protein độc tố lớp Cry1A, và có trình tự axit amin từ khoảng vị trí 1 đến vị trí 607 như được chỉ ra trong SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO 6, SEQ ID NO:8, SEQ ID NO:10, và SEQ ID NO:12, hoặc mảnh có tác dụng diệt côn trùng của chúng. Polypeptit ngắn hơn độ dài đầy đủ thể hiện tác dụng diệt côn trùng như vậy phải có tỷ lệ đồng nhất ít nhất là khoảng 100%, 99%, 98%, 97%, 96%, 95%, 94%, 94%, 92%, 91%, hoặc 90% với trình tự axit amin BCW 001 như được chỉ ra trong SEQ ID NO:2 từ khoảng vị trí 1 đến khoảng vị trí 606, hoặc từ khoảng vị trí 5 đến khoảng vị trí 600. Nếu độ dài đầy đủ hoặc các mảnh độc tố lớn hơn đáng kể cần được sử dụng, tỷ lệ phần trăm đồng nhất nên ít nghiêm ngặt hơn và mở rộng đến tỷ lệ phần trăm đồng nhất từ khoảng 100, đến khoảng 95, đến khoảng 90, đến khoảng 85, hoặc thậm chí là 80% với các trình tự protein độc tố có độ dài đầy đủ như được chỉ ra trong SEQ ID NO:2, 4, và 6, vì các protein độc tố này có mức hoạt tính sinh học hữu ích về mặt thương mại khi được thử nghiệm kháng ấu trùng sâu xám trong các thử nghiệm sinh học về chế độ ăn, và khi được thử nghiệm trên thực vật ở ngô, bông và các trường hợp chuyển gen đậu tương biểu hiện các protein này.

Sáng chế cũng đề xuất các protein gây độc cho loài sâu xám *Lepidopteran*, bao gồm các protein có trình tự axit amin như được chỉ ra trong SEQ ID NO:2 từ vị trí 256 đến 606 (protein BCW 001), và các protein có trình tự axit amin như được chỉ ra trong một trong

số SEQ ID NO:4 và SEQ ID NO:6 từ vị trí axit amin 257 đến 607 (được gọi lần lượt ở đây là protein độc tố BCW 002 và protein độc tố BCW 003).

Các protein diệt côn trùng này cũng được quan sát thấy là thể hiện hoạt tính kháng loài *Lepidopteran* được chọn từ nhóm bao gồm *Spodoptera frugiperda*, *Spodoptera exigua*, *Spodoptera litura*, *Mamestra configurata*, *Striacosta albicosta*, *Trichoplusia ni*, *Pseudoplusia includens*, *Anticarsia gemmatilis*, *Hypena scabra*, *Heliothis virescens*, *Agrotis subterranea*, *Pseudaletia unipuncta*, *Agrotis ipsilon*, *Agrotis orthogonia*, *Ostrinia nubilalis*, *Amyelois transitella*, *Crambus caliginosellus*, *Herpetogramma licarsisalis*, *Homoeosoma electellum*, *Elasmopalpus lignosellus*, *Cydia pomonella*, *Endopiza viteana*, *Grapholita molesta*, *Suleima helianthana*, *Plutella xylostella*, *Pectinophora gossypiella*, *Lymantria dispar*, *Alabama argillacea*, *Archips argyrospila*, *Archips rosana*, *Chilo suppressalis*, *Cnaphalocrocis medinalis*, *Crambus caliginosellus*, *Crambus teterrellus*, *Diatraea grandiosella*, *Diatraea saccharalis*, *Earias insulana*, *Egrias vittella*, *Helicoverpa armigera*, *Helicoverpa zea*, *Heliothis virescens*, *Herpetogramma licarsisalis*, *Lobesia botrana*, *Pectinophora gossypiella*, *Phyllocnistis citrella*, *Pieris brassicae*, *Pieris rapae*, *Plutella xylostella*, và *Tuta absoluta*.

Các protein theo sáng chế cũng có thể thể hiện hoạt tính sinh học kháng loài *Lepidopteran* được chọn từ nhóm bao gồm *Spodoptera frugiperda*, *Spodoptera exigua*, *Spodoptera litura*, *Mamestra configurata*, *Striacosta albicosta*, *Trichoplusia ni*, *Pseudoplusia includens*, *Anticarsia gemmatilis*, *Hypena scabra*, *Heliothis virescens*, *Agrotis subterranea*, *Pseudaletia unipuncta*, *Agrotis ipsilon*, *Agrotis orthogonia*, *Ostrinia nubilalis*, *Amyelois transitella*, *Crambus caliginosellus*, *Herpetogramma licarsisalis*, *Homoeosoma electellum*, *Elasmopalpus lignosellus*, *Cydia pomonella*, *Endopiza viteana*, *Grapholita molesta*, *Suleima helianthana*, *Plutella xylostella*, *Pectinophora gossypiella*, *Lymantria dispar*, *Alabama argillacea*, *Archips argyrospila*, *Archips rosana*, *Chilo suppressalis*, *Cnaphalocrocis medinalis*, *Crambus caliginosellus*, *Crambus teterrellus*, *Diatraea grandiosella*, *Diatraea saccharalis*, *Earias insulana*, *Egrias vittella*, *Helicoverpa armigera*, *Helicoverpa zea*, *Heliothis virescens*, *Herpetogramma licarsisalis*, *Lobesia botrana*, *Pectinophora gossypiella*, *Phyllocnistis citrella*, *Pieris brassicae*, *Pieris rapae*, *Plutella xylostella*, và *Tuta absoluta*.

Các protein theo sáng chế, và các cấu trúc được đề xuất ở đây, có thể được bao gồm trong vectơ bất kỳ bao gồm plasmid, cosmid, bacmid, vectơ qua trung gian thể thực khuẩn, và tương tự.

Các vectơ này có thể được sử dụng để đưa các cấu trúc theo sáng chế vào số lượng tế bào chủ bất kỳ, bao gồm vào các tế bào vi khuẩn, các tế bào nấm men, và các tế bào thực vật.

Các tế bào chủ mà là tế bào nấm men có thể là *Saccharomyces cerevisiae* hoặc *Saccharomyces pombe* và các tế bào tương tự. Các tế bào chủ vi khuẩn có thể là số lượng bất kỳ các tế bào chủ đã biết này bao gồm nhưng không bị giới hạn ở *E. coli*, *B. thuringiensis*, và các vi khuẩn cùng họ khác. Các tế bào chủ thực vật có thể thu được từ số lượng bất kỳ các loài thực vật, bao gồm nhưng không bị giới hạn ở các tế bào thực vật từ cây cỏ linh lăng, chuối, lúa mạch, đậu, bông cải xanh, bắp cải, cải, cà rốt, sắn, thầu dầu, bông cải trắng, cần tây, đậu gà, cải thảo, cam, dứa, cà phê, ngô, cỏ ba lá, bông, bầu bí, dưa chuột, cây gỗ linh sam, cà tím, bạch đàn, cây lanh, tỏi, nho, cây hublông, tỏi tây, rau diếp, thông trầm hương, kê, quả dưa, quả hạch, yến mạch, oliu, hành tây, cây cảnh, cò, cỏ, đậu Hà Lan, củ lạc, cây ớt, đậu triều, thông, khoai tây, bạch dương, bí đỏ, thông Radiata, củ cải, hạt cải dầu, lúa, cây thân rễ, lúa mạch đen, hồng hoa, cây bụi, cây cao lương, thông phương Nam, đậu tương, rau chân vịt, quả bí, dâu tây, củ cải đường, mía, hoa hướng dương, ngô ngọt, gỗ gôm, khoai lang, cỏ kê Mỹ, chè, thuốc lá, cà chua, tiểu hắc mạch, cỏ, dưa hấu, và lúa mì.

Các dòng thực vật chuyển gen có thể được tạo ra, cụ thể là các giống cây trồng chuyển gen ngô, bông và đậu tương, bằng cách đưa các cấu trúc theo sáng chế chứa trình tự polynucleotit được biến đổi thích hợp như được chỉ ra trong SEQ ID NO:7, SEQ ID NO:9, và SEQ ID NO:11, ví dụ, vào bộ gen của tế bào thực vật này và thu hồi cây ngô, đậu tương hoặc bông chuyển gen có khả năng thụ phấn chứa trong bộ gen của nó cấu trúc di truyền để biểu hiện ít nhất là độc tố protein theo sáng chế, nghĩa là, độc tố protein BCW 001, BCW 002, hoặc BCW 003. Các cây trồng chuyển gen này sẽ đưa vào bộ gen thực vật của chúng, cấu trúc polynucleotit chứa ít nhất là đoạn trình tự khởi đầu khác loại được liên kết hoạt động với trình tự nucleotit mã hóa protein diệt côn trùng BCW 001, BCW 002, hoặc BCW 003 có trình tự axit amin được chọn từ nhóm bao gồm SEQ ID NO:8, SEQ ID

NO:10, và SEQ ID NO:12, hoặc mảnh protein diệt côn trùng của chúng.

Các hạt cũng được đề xuất như một dấu hiệu của sáng chế, trong đó hạt này được sản xuất từ các cây chuyển gen này và hạt này chứa lượng phát hiện được của cấu trúc polynucleotit được đưa vào bộ gen thực vật này. Phần hoa, hạt, các tế bào thực vật dòng con, mô thực vật và các sản phẩm hàng hóa được sản xuất từ mỗi cây chuyển gen này sẽ chứa lượng phát hiện được của cấu trúc polynucleotit.

Mẫu sinh học bất kỳ mà chứa ít nhất là lượng phát hiện được của cấu trúc polynucleotit mã hóa các protein BCW 001, BCW 002, hoặc BCW 003 sẽ nằm trong phạm vi của sáng chế.

Các chế phẩm cung cấp lượng có tác dụng diệt côn trùng của protein BCW 001, 002 hoặc 003 theo sáng chế được đề xuất, và được cung cấp để phòng trừ vật gây hại *Lepidopteran*. Các chế phẩm này cũng có thể chứa chất bổ sung khác với protein độc tố BCW. Các chất này cũng sẽ gây độc cho cùng loài *Lepidopteran* giống như protein độc tố BCW. Chất bổ sung này được chọn từ nhóm các chất bao gồm các protein hoặc polypeptit có trình tự axit amin khác với protein BCW, và cũng có thể là chất mà là phân tử ARN có các tác dụng gây độc trên côn trùng gây hại đích (như dsARN, miARN, hoặc siARN), và cũng có thể là hóa chất diệt côn trùng như pyrethrin, thuốc diệt côn trùng phosphat hữu cơ và tương tự. Theo cách khác, chất bổ sung này có thể là Cry thích hợp bất kỳ hoặc protein độc tố liên quan như Cry1A, Cry1Ab, Cry1Ac, Cry1A.105, Cry1Ae khác, (nhưng các chất này không được ưu tiên vì chúng có thể không tạo ra các tính chất phòng trừ tính kháng thích hợp), hoặc các protein diệt côn trùng Cry1B, Cry1C, Cry1D, Cry1E, Cry1F, Cry1G, Cry1H, Cry1I, Cry1J, Cry1K, Cry1L, Cry2A, Cry2Ab, Cry2Ae, Cry4B, Cry6, Cry7, Cry8, Cry9, Cry15, Cry43A, Cry43B, ET35, ET66, TIC400, TIC800, TIC807, TIC834, TIC853, TIC1415, VIP3A, VIP3Ab, Axmi, protein diệt côn trùng DIG, eHIP, và protein VIP và protein độc tố bất kỳ đã biết trong lĩnh vực này để gây ra các tính chất gây độc trên ấu trùng loài sâu xám hoặc các loài gây hại đích *Lepidopteran* khác có thể áp dụng.

Các chế phẩm này cũng có thể bao gồm các chất diệt côn trùng bổ sung mà không nhất định phải gây độc cho cùng một loài gây hại đích, như chất bổ sung được chọn từ nhóm bao gồm Cry1C, Cry3A, Cry3B, Cry34, Cry35, Cry51Aa1, ET29, ET33, ET34, ET70, TIC407, TIC417, TIC431, TIC901, TIC1201, TIC3131, 5307, DIG-10, Axmi184,

Axmi205 và AxmiR1.

Các phương pháp cũng được đề xuất để sản xuất hạt có được các đặc điểm diệt côn trùng của các protein theo sáng chế. Các phương pháp này bao gồm cấu trúc polynucleotit được thiết kế để biểu hiện protein BCW 001, BCW 002 hoặc BCW 003, hoặc protein có tỷ lệ đồng nhất ít nhất khoảng 90% với protein này, phương pháp này bao gồm bước gieo một hoặc nhiều hạt chứa polynucleotit biểu hiện một hoặc nhiều trong số các độc tố protein BCW theo sáng chế, trồng các cây từ hạt này và sau đó thu hoạch sản phẩm gồm các hạt này từ các cây trồng. Hạt thu hoạch được sẽ chứa cấu trúc polynucleotit và sẽ phát triển thành cây mà cũng sẽ kháng lại sự phá hoại của vật gây hại là sâu xám.

Các cây này có thể là ngô, bông, đậu tương, hoặc cây khác bất kỳ dễ bị nhiễm loài gây hại *Lepidopteran* mà đã được chứng minh là phòng trừ được bởi các protein theo sáng chế. Các cây này được đề xuất là các cây chuyển gen đã được tạo ra trước đó mà được hưởng lợi từ các tác dụng của các đặc điểm gây độc của các protein theo sáng chế. Cây ngô nằm trong nhóm này bao gồm nhưng không bị giới hạn ở các dòng chuyển gen được chọn từ nhóm bao gồm DKB89614-9, MON801, MON802, MON809, MON810, MON863, MON88017, MON89034, dòng 4114-3, dòng 5307, DAS59122-7, Bt10, Bt11, Bt176, CBH-351, DKB-83614-9, MIR162, MIR604, TC1507, TC6275, dòng 676, dòng 678, dòng 680, dòng 98140, DAS40278-9, DKB89790-5, MON21-9, HCEM485, MON832, MON87427, NK603, T14, T25 và VCO01981-5. Cây đậu tương nằm trong nhóm các cây chuyển gen này được chọn từ nhóm bao gồm MON87751, DAS81419-2, MON87701, A2704-12, A2704-21, A5547-127, A5547-35, CV127, DAS44406-6, DAS68416-4, DP356043, FG72, MON4032, ACS-GM003-1, MON87705, MON87708, MON89788, W62, W98 và GFM Cry1A. Cây bông chuyển gen nằm trong nhóm này được chọn từ nhóm bao gồm DAS24236-5, DAS21023-5, dòng 31707, dòng 31803, dòng 31807, dòng 31808, dòng 42317, BNLA-601, COT102, COT67B, dòng 1, GHB119, GK12, MON15985, MLS9124, MON1076, MON531, MON757, T303-3, T304-40, SGK321, dòng 19-51a, GHB614, LLCotton25, MON88701, MON88702, MON1445, MON1698 và MON88913. Cây mía chuyển gen nằm trong nhóm này bao gồm dòng chuyển gen của cây mía NXI-1T. Cây lúa đã được biết trong lĩnh vực này là hưởng lợi từ việc có cấu trúc mã hóa các độc tố protein BCW này bao gồm các dòng chuyển gen của cây lúa được chọn từ nhóm bao gồm

LLRICE06, LLRICE601, LLRICE62, GM-A17054 và GM-A17054.

Ngoài ra, sáng chế cũng đề xuất các sản phẩm thực vật đã được xử lý chứa lượng phát hiện được của polynucleotit tái tổ hợp đã được mô tả này. Các sản phẩm đã được xử lý này bao gồm, nhưng không giới hạn ở: sinh khối thực vật, dầu, bột xay thô, thức ăn cho động vật, bột mỳ, tấm, cám, xơ vải, vỏ trấu, và hạt đã được xử lý.

Các phương pháp tạo ra các cây chuyển gen cũng được đề xuất. Các phương pháp này bao gồm bước đưa polynucleotit tái tổ hợp vào tế bào thực vật và chọn lọc cây chuyển gen biểu hiện lượng có tác dụng ức chế côn trùng của polypeptit tái tổ hợp được mã hóa bởi polynucleotit tái tổ hợp này.

Các phương án, dấu hiệu và các ưu điểm khác của sáng chế sẽ rõ ràng từ phần mô tả chi tiết, ví dụ thực hiện sáng chế và yêu cầu bảo hộ sau đây.

Mô tả vắn tắt các hình vẽ

Fig. 1 thể hiện danh mục các loài côn trùng gây hại *Lepidopteran* mà được kiểm tra trong thử nghiệm sinh học với các protein độc tố BCW 001, BCW 002, và BCW 003. “+” chỉ ra tỷ lệ chết so với đối chứng đậm; “-” chỉ ra là không quan sát thấy tỷ lệ chết cao hơn đáng kể mức của đối chứng đậm; “ND” chỉ ra là không được kiểm tra bằng cách sử dụng protein độc tố thích hợp. BCW 001 thể hiện tác dụng gây chết cho *Agrotis ipsilon* (sâu xám), *Striacosta albicostsa* (sâu xám miền Tây), *Helicoverpa zea* (sâu đục bắp trên cây ngô), *Ostrinia nubilalis* (sâu đục thân bắp châu Âu), *Diatraea grandiosella* (sâu đục thân bắp Tây Nam), *Trichoplusia ni* (sâu đo bắp cải), *Pseudoplusia includens* (sâu đo đậu tương), và không gây ra tác dụng gây chết hoặc gây choáng khi được kiểm tra trên *Spodoptera frugiperda* (sâu xanh mùa thu). *Diatraea saccharalis* (sâu đục thân cây mía) không được kiểm tra với BCW 001. BCW 002 thể hiện tác dụng gây chết trên *Agrotis ipsilon* (sâu xám), *Striacosta albicostsa* (sâu ngài đêm trên cây đậu miền Tây), *Helicoverpa zea* (sâu đục bắp trên cây ngô), và *Ostrinia nubilalis* (sâu đục bắp trên cây ngô châu Âu), và không thể hiện tác dụng gây chết hoặc gây choáng khi được kiểm tra trên *Spodoptera frugiperda* (sâu xanh mùa thu). *Diatraea saccharalis* (sâu đục thân cây mía), *Diatraea grandiosella* (sâu đục thân bắp Tây Nam), *Trichoplusia ni* (sâu đo bắp cải), và *Pseudoplusia includens* (sâu đo đậu tương) không được kiểm tra với BCW 002. BCW 003

thể hiện tác dụng gây chết trên *Agrotis ipsilon* (sâu xám), *Striacosta albicostsa* (sâu ngài đêm trên cây đậu miền Tây), *Helicoverpa zea* (sâu đục bắp trên cây ngô), *Ostrinia nubilalis* (sâu đục bắp châu Âu), *Diatraea saccharalis* (sâu đục thân cây mía), *Diatraea grandiosella* (sâu đục thân bắp Tây Nam), *Trichoplusia ni* (sâu đo bắp cải), và *Pseudoplusia includens* (sâu đo đậu tương), và không thể hiện tác dụng gây chết hoặc gây choáng khi được thử nghiệm trên *Spodoptera frugiperda* (sâu xanh mùa thu).

Fig. 2 thể hiện sự sắp xếp thẳng hàng trình tự axit amin của BCW 001 (SEQ ID NO:2, hàng trên cùng) với BCW 002 (SEQ ID NO:4, hàng giữa), với BCW 003 (SEQ ID NO:6, hàng dưới cùng); các dấu hoa thị bên dưới mỗi nhóm ba hàng thể hiện sự khác nhau ở vị trí axit amin thích hợp trong ít nhất một trong số ba trình tự khác nhau.

Mô tả vắn tắt các trình tự

SEQ ID NO:1 là trình tự nucleotit EG4384 của giống *B. thuringiensis* tự nhiên mã hóa protein gây độc BCW 001 *Lepidopteran*.

SEQ ID NO:2 là trình tự axit amin được suy luận của BCW 001 từ khung đọc mở như được chỉ ra trong SEQ ID NO:1.

SEQ ID NO:3 là trình tự nhân tạo mã hóa protein gây độc BCW 002 *Lepidopteran* khảm.

SEQ ID NO:4 là trình tự axit amin được suy luận của BCW 002 từ khung đọc mở như được chỉ ra trong SEQ ID NO:3, trong đó protein BCW 002 này chứa miền I của Cry1Ac được liên kết hoạt động với các miền II và III của BCW 001 (vị trí axit amin 258 đến vị trí axit amin 606 như được chỉ ra trong SEQ ID NO:2) và được liên kết hoạt động với miền tiền độc tố Cry1Ac từ vị trí axit amin 608 đến 1177 như được chỉ ra trong SEQ ID NO:4 và SEQ ID NO:6. SEQ ID NO:4 (BCW 002) là khác với SEQ ID NO:6 (BCW 003) chỉ ở vị trí 259, BCW 002 chứa Isoleuxin ở vị trí này, BCW 003 chứa Threonin như trong BCW 001.

SEQ ID NO:5 là trình tự nhân tạo mã hóa protein gây độc BCW 003 *Lepidopteran* khảm.

SEQ ID NO:6 là trình tự axit amin được suy luận của BCW 003 từ khung đọc mở như được chỉ ra trong SEQ ID NO:3, trong đó protein BCW 003 này chứa miền I của

Cry1Ac được liên kết hoạt động với các miền II và III của BCW 001 (vị trí axit amin 258 đến vị trí axit amin 606 như được chỉ ra trong SEQ ID NO:2) và được liên kết hoạt động với miền tiền độc tố Cry1Ac từ vị trí axit amin 608 đến 1177 như được chỉ ra trong SEQ ID NO:4 và SEQ ID NO:6. SEQ ID NO:6 (BCW 003) là khác với SEQ ID NO:4 (BCW 002) chỉ ở vị trí 259, BCW 002 chứa Isoleuxin ở vị trí này, BCW 003 chứa Threonin như trong BCW 001.

SEQ ID NO:7 là trình tự nhân tạo mã hóa protein BCW 001 để biểu hiện ở thực vật.

SEQ ID NO:8 là trình tự axit amin được suy luận của BCW 001 thu được từ SEQ ID NO:7.

SEQ ID NO:9 là trình tự nhân tạo mã hóa protein BCW 002 để biểu hiện ở thực vật.

SEQ ID NO:10 là trình tự axit amin được suy luận của BCW 002 thu được từ SEQ ID NO:9.

SEQ ID NO:11 là trình tự nhân tạo mã hóa protein BCW 003 để biểu hiện ở thực vật.

SEQ ID NO:12 là trình tự axit amin được suy luận của BCW 003 thu được từ SEQ ID NO:11.

Mô tả chi tiết sáng chế

Một phương án thay thế cho việc phòng trừ vật gây hại trong nông nghiệp ở cây trồng bằng cách phun chế phẩm chứa protein diệt côn trùng lên cây/hạt là gắn chèn các polynucleotit mã hóa các protein này vào bộ gen thực vật để biểu hiện ở cây trồng hoặc các bộ phận của cây trồng. Các cây được biến nạp với các polynucleotit này mang tính kháng côn trùng được tạo ra bởi các protein được biểu hiện này dưới dạng các tính trạng chuyển gen.

Để tránh sự phát triển của, hoặc tránh việc côn trùng kháng lại các protein hiện đang được sử dụng, các protein mới có cơ chế tác dụng khác (MOA), cũng như có phổ rộng và hiệu lực là cần thiết để phòng trừ *Lepidoptera*. Một cách để giải quyết nhu cầu này là giải trình tự bộ gen *Bt* với hy vọng phát hiện các protein diệt côn trùng mới. Một phương pháp khác là trao đổi các đoạn từ nhiều protein *Bt* khác nhau để tạo ra các protein *Bt* khả năng mới

có các đặc điểm ức chế côn trùng. Khả năng tạo ra tùy ý protein khả năng có các đặc điểm được tăng cường từ việc tái sắp xếp các cấu trúc miền của số lượng lớn protein tinh thể diệt côn trùng tự nhiên đã biết trong lĩnh vực này là rất xa (*Xem*, ví dụ A Strategy for Shuffling Numerous *Bacillus thuringiensis* Crystal Protein Domains; J. Economic Entomology, 97 (6):1805-1813. 2004).

Được mô tả ở đây là các trình tự nucleotit mã hóa các protein diệt côn trùng, được xác định ở đây là các protein BCW, mà giải quyết nhu cầu về MOA thay thế, cung cấp hoạt tính kháng lại phổ côn trùng gây hại rộng hơn, và có tác dụng làm chậm hoặc tránh được sự phát triển của tính kháng, cụ thể là để sử dụng trong việc phòng trừ vật gây hại gây hại là sâu xám (BCW).

BCW 001 được phát hiện dưới dạng khung đọc mở dự đoán trình tự axit amin có các đặc điểm của protein typ Cry1A sau khi giải trình tự bộ gen của EG4384 chủng *Bacillus thuringiensis*. Khung đọc mở BCW 001 (ORF) mã hóa protein gồm 1180 axit amin và protein này được dự đoán là có nhiều đặc điểm của độc tố protein Cry1, bao gồm cấu trúc miền I, II, và III xác định được, và miền tiền độc tố typ Cry1A đặc trưng ở nửa đầu cùng carboxy của protein được dự đoán. Trình tự axit amin Miền I được dự đoán (gốc 1 đến khoảng 258 của SEQ ID NO:2) thể hiện tỷ lệ đồng nhất khoảng 67% với miền độc tố I protein Cry1Ac. Trình tự axit amin Miền II được dự đoán (các gốc từ khoảng 259 đến khoảng gốc 459 như được chỉ ra trong SEQ ID NO:2) thể hiện tỷ lệ đồng nhất hoàn hảo (100%) với Miền II của Cry1Ai2. Trình tự axit amin Miền III được dự đoán (các gốc từ khoảng 260 đến khoảng 606 như được chỉ ra trong SEQ ID NO:2) thể hiện tỷ lệ đồng nhất khoảng 63% với các gốc Miền III tương ứng trong Cry1Ah2. Cấu trúc miền tiền độc tố của protein được dự đoán BCW 001 (khoảng các gốc 607 đến 1180 như được chỉ ra trong SEQ ID NO:2) thể hiện tỷ lệ đồng nhất khoảng 96% với các gốc tương ứng trong Cry1Aa9. Nhìn chung protein có chiều dài đầy đủ được dự đoán này thể hiện tỷ lệ đồng nhất trình tự axit amin khoảng 83% với Cry1Ai, và vùng độc tố được dự đoán từ vị trí axit amin 1 đến khoảng gốc 607 như được chỉ ra trong SEQ ID NO:2 thể hiện tỷ lệ đồng nhất 76% với Cry1Ai1. Không chỉ định protein độc tố mới này vào lớp Cry1A cụ thể và do đó hội đồng danh pháp *Bacillus thuringiensis* sẽ được cung cấp trình tự này và sẽ xác định protein này có đủ tiêu chuẩn để tạo ra lớp mới và riêng của chính nó hay không.

Protein BCW 001 được sản sinh ra từ vector plasmit ở chủng không tạo tinh thể của *Bacillus thuringiensis* và các chế phẩm tinh thể bào tử được kiểm tra trên loài vật hại *Lepidopteran*. Xem số liệu trên Fig. 1, cột 2. Bằng chứng được chỉ ra là protein này là không điển hình cho bất kỳ trong số các protein mà từ đó thu được nguồn gốc của nó, nghĩa là, các protein độc tố Cry1Aa, Cry1Ah, hoặc Cry1Ai đã biết trong lĩnh vực này. Không có protein thuộc giải pháp đã biết nào thể hiện hoạt tính đáng kể bất kỳ khi được kiểm tra trên sâu xám, tuy nhiên, protein mới BCW 001 này gây độc trong các thử nghiệm sinh học khi được thử nghiệm trên sâu xám, và bất ngờ thể hiện các đặc điểm gây độc khi được thử nghiệm trên nhóm các vật gây hại *Lepidopteran* khác cũng như được chỉ ra trong Fig. 1.

Cụ thể là, trong các thử nghiệm sinh học so với đối chứng chế độ ăn của côn trùng không được xử lý, protein BCW 001 thể hiện hoạt tính kháng lại sâu ngài đêm trên cây đậu miền Tây (“WBC”; *Striacosta albicosta*), sâu đục bắp trên cây ngô (“CEW”, *Helicoverpa zea*), sâu đục bắp trên cây ngô châu Âu (“ECB”, *Ostrinia nubilalis*), sâu đục bắp trên cây ngô tây nam (“SWC”, *Diatraea grandiosella*), sâu đo đậu tương (“SL”, *Pseudoplusia includes*), sâu đo bắp cải (“CLW”, *Trichoplusia ni*), và sâu xám 1 tuổi và 3 tuổi (“BCW”, *Agrotis ipsilon*).

Như được mô tả thêm dưới đây, protein độc tố khảm được sản sinh bằng cách sử dụng Miền I của Cry1Ac và Miền II và III của BCW 001 (nghĩa là, BCW 002 và BCW 003 như được chỉ ra lần lượt trong SEQ ID NO:4 và SEQ ID NO:6), và các protein độc tố khảm này được đưa vào cây ngô và cây mía. Cả hai protein khảm đều thể hiện hoạt tính ở cây ngô kháng lại BCW, WBC, CEW, và SWC. BCW 003 thể hiện hoạt tính kháng SCB ở cây mía.

Cụm từ “protein BCW”, như được sử dụng ở đây, đề cập đến protein ức chế côn trùng mới bất kỳ mà chứa, mà bao gồm, mà gần như tương đồng với, hoặc mà thu được từ trình tự polypeptit ức chế côn trùng bất kỳ trong số: BCW 001 (SEQ ID NO:2), BCW 002 (SEQ ID NO:4), và BCW 003 (SEQ ID NO:6), và các đoạn ức chế côn trùng của chúng, hoặc tổ hợp của chúng, mà gây ra hoạt tính kháng *Lepidoptera*, cụ thể là, nhưng không bị giới hạn ở BCW, WBC và/hoặc SCB. Polynucleotit mã hóa BCW 001 thu được từ chủng EG4384. Trình tự axit amin gây độc lõi cho BCW 001 tương ứng với các axit amin từ khoảng vị trí 28 đến khoảng vị trí 606 và đến vị trí 618 như được chỉ ra trong SEQ ID

NO:2, và trình tự axit amin gây độc lõi cho BCW 002 và 003 tương ứng với các axit amin từ khoảng vị trí 29 đến khoảng vị trí 607 và đến khoảng vị trí 619 như được chỉ ra lần lượt trong SEQ ID NO:4 và SEQ ID NO:6.

Trong một phương án, các protein được mô tả ở đây có liên quan bởi cấu trúc nội độc tố delta bậc một, bởi chiều dài (khoảng 1176-1180 axit amin), bởi độ dài protein không có tiền độc tố (từ khoảng 600 đến khoảng 619 axit amin), bởi độ dài của lõi gây độc (khoảng 591 axit amin), hoặc bởi sự có mặt của ít nhất một đoạn đặc hiệu BCW.

Các protein được lấy làm ví dụ được sắp xếp thẳng hàng với nhau bằng cách sử dụng thuật toán Clustal W, tạo ra nhiều sự đồng nhất axit amin theo từng cặp và tỷ lệ phần trăm đồng nhất axit amin theo từng cặp cho mỗi cặp bằng cách sử dụng các tham số mặc định này: Ma trận khối lượng: blosum; Điểm phạt mở khoảng trống: 10,0; Điểm phạt kéo dài khoảng trống: 0,05; Khoảng trống ưa nước: On; Các gốc ưa nước: GPSNDQERK; Các điểm phạt khoảng trống đặc hiệu gốc: On. Thuật toán Clustal W được mô tả trong Thompson, J. D., Higgins, D.G. và Gibson, T.J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. "Nucleic Acids Research, 22:4673-4680.

Các thuật toán sắp xếp thẳng hàng khác cũng là sẵn có trong lĩnh vực này và cung cấp các kết quả tương tự với các kết quả thu được bằng cách sử dụng phương pháp sắp xếp thẳng hàng Clustal W.

Thuật ngữ “khoảng” được sử dụng ở đây để mô tả là các biên giới của đoạn có thể thay đổi 1, 2, 3, 4, 5, 10, 20, 25, 30, hoặc 35 gốc, tùy theo trình tự của protein mẹ và sự sắp xếp thẳng hàng với nhau của chúng. Để mô tả thêm về sự biến đổi và hình dạng của nhiều đoạn khác nhau này, Bảng 2 và 3 lập bảng các biên giới của đoạn của BCW 002 và 003 và các thể khả hoạt động của sâu xám khác.

Thuật ngữ “mảnh” được sử dụng ở đây để mô tả các trình tự axit amin hoặc axit nucleic liên tục mà ngắn hơn so với trình tự axit amin hoặc axit nucleic hoàn chỉnh thể hiện protein gây độc BCW.

Cụm từ “ức chế côn trùng” và “diệt côn trùng” được sử dụng ở đây có thể thay thế

lẫn nhau và đề cập đến protein, mảnh protein, đoạn protein hoặc polynucleotit mà gây ra tác dụng ức chế đo được bất kỳ đối với sự sống, sự sinh trưởng của côn trùng, sự phát triển của côn trùng, sự sinh sản của côn trùng, tập tính ăn của côn trùng, tập tính giao phối và/hoặc tác dụng làm giảm đo được bất kỳ đối với tác dụng bất lợi gây ra bởi việc cho côn trùng ăn trên protein, mảnh protein, đoạn protein hoặc polynucleotit.

Thuật ngữ “hoạt tính sinh học”, “có tác dụng”, “hoạt tính”, “hữu hiệu”, “hiệu quả” hoặc các biến thể của chúng được sử dụng thay thế lẫn nhau ở đây để mô tả các tác dụng của protein theo sáng chế trên côn trùng gây hại mục tiêu.

Cây là cây tự mọc hoặc được trồng mà có sản phẩm được thu hoạch ở một số thời điểm của giai đoạn sinh trưởng của nó. Ví dụ không giới hạn về các sản phẩm này là hạt, quả nang, lá, hoa, thân, rễ, hoặc bộ phận bất kỳ của chúng.

Mẫu sinh học thu được từ mô bất kỳ của thực vật, vi khuẩn, virus hoặc vector chứa polynucleotit hoặc biểu hiện protein như được minh họa ở đây, như, nhưng không bị giới hạn ở, hạt giống, quả nang, lá, hoa, thân, rễ, hoặc bộ phận bất kỳ của chúng và chứa lượng phát hiện được của polynucleotit, protein này, hoặc cả hai.

Cụm từ “lượng phát hiện được” được sử dụng ở đây để mô tả lượng tối thiểu của protein hoặc polynucleotit được mô tả ở đây mà có thể phát hiện được bằng các phương pháp phân tích chuẩn như, nhưng không bị giới hạn ở, kỹ thuật phản ứng chuỗi polymeraza (PCR) và thử nghiệm hấp thu miễn dịch liên kết enzym (ELISA), và các kỹ thuật tương tự.

Trong một phương án, các protein độc tố được mô tả ở đây có liên quan bởi chức năng chung và thể hiện hoạt tính diệt côn trùng đối với loài côn trùng *Lepidoptera*.

Các đoạn BCW 001 tạo ra hoạt tính kháng sâu xám cho thể khảm Cry1 chứa các đoạn này. Ví dụ về các đoạn BCW 001 mà tạo ra hoạt tính kháng sâu xám được chỉ ra trong SEQ ID NO:2 từ khoảng vị trí axit amin 250 đến khoảng 606 và cụ thể hơn là từ khoảng vị trí axit amin 255 đến khoảng 606. Thể khảm Cry1 chứa đoạn axit amin này tương ứng với Miền II và III của protein độc tố BCW 001 thường cũng sẽ tạo ra trên protein khảm này các đặc tính gây độc liên quan đến việc phòng diệt côn trùng xám, và điều này được kiểm tra trong các khung Cry1A, Cry1B, Cry1C, Cry1D, Cry1E, và Cry1F trong đó cấu trúc độc tố thích hợp đã có các thành phần Miền II và III của nó được thay thế bằng dãy

axit amin này từ BCW 001 và trong nhiều trường hợp, hoạt tính kháng sâu xám bất ngờ được duy trì ở cấu trúc khảm này (số liệu không được thể hiện).

Theo một khía cạnh của sáng chế, vật gây hại được phòng trừ bởi protein độc tố BCW thích hợp cụ thể là côn trùng gây hại thuộc bộ *Lepidoptera*, bao gồm con trưởng thành, nhộng, ấu trùng, và sơ sinh.

Các côn trùng thuộc bộ *Lepidoptera* bao gồm, nhưng không bị giới hạn ở, sâu khoang, sâu ngài đêm, sâu đo, và sâu xanh trong họ ngài đêm (ví dụ, sâu xanh mùa thu (*Spodoptera frugiperda*), sâu xanh da láng (*Spodoptera exigua*), sâu khoang bertha (*Mamestra configurata*), sâu xám (*Agrotis ipsilon*), sâu đo bắp cải (*Trichoplusia ni*), sâu đo đậu tương (*Pseudoplusia includens*), sâu bướm trên cây đậu mèo rừng (*Anticarsia gemmatilis*), sâu xanh (*Hypena scabra*), sâu chồi thuốc lá (*Heliothis virescens*), sâu ngài đêm dạng hạt (*Agrotis subterranea*), sâu ngài đêm phương tây (*Agrotis orthogonia*), sâu khoang (*Pseudaletia unipuncta*), sâu đục thân, sâu túi, sâu kéo màng, sâu hình nón, sâu bắp cải và sâu gặm thịt lá thuộc họ Pyralidae (ví dụ, sâu đục bắp trên cây ngô châu Âu (*Ostrinia nubilalis*), sâu cam navel (*Amyelois transitella*), sâu kéo màng rễ ngô (*Crambus caliginosellus*), sâu kéo màng cỏ (*Herpetogramma licarsisalis*), sâu bướm hoa hướng dương (*Homoeosoma electellum*), sâu đục thân ngô lesser (*Elasmopalpus lignosellus*), bướm cuốn lá, sâu đục chồi, sâu đục hạt, và sâu đục quả thuộc họ Tortricidae (ví dụ, sâu ngài codling (*Cydia pomonella*), sâu ngài hại nho (*Endopiza viteana*), sâu ngài đục trái phương Đông (*Grapholita molesta*), sâu chồi hoa hướng dương (*Suleima helianthana*); và nhiều *Lepidoptera* quan trọng về kinh tế khác (ví dụ, sâu tơ cải bắp (*Plutella xylostella*), sâu hồng hại bông (*Pectinophora gossypiella*), sâu bướm gypsy (*Lymantria dispar*). Các côn trùng gây hại khác thuộc bộ *Lepidoptera* bao gồm, ví dụ, *Alabama argillacea* (sâu lá bông), *Archips argyrospila* (bướm cuốn lá cây ăn quả), *A. rosana* (bướm cuốn lá châu Âu) và các loài *Archips* khác, *Chilo suppressalis* (sâu đục thân lúa Asiatic, hoặc sâu đục thân lúa), *Cnaphalocrocis medinalis* (sâu cuốn lá lúa), *Crambus caliginosellus* (sâu kéo màng rễ ngô), *C. teterrellus* (sâu kéo màng cỏ xanh), *Diatraea grandiosella* (sâu đục bắp trên cây ngô tây nam), *D. saccharalis* (sâu đục thân mía), *Earias insulana* (ấu trùng của sâu bướm có gai), *E. vittella* (sâu đốm đục quả), *Helicoverpa armigera* (sâu bướm Mỹ), *H. zea* (sâu đục bắp trên cây ngô hoặc sâu đục quả cây bông), *Heliothis virescens* (sâu chồi thuốc

lá), *Herpetogramma licarsisalis* (sâu kéo màng cỏ), *Lobesia botrana* (sâu bướm cây nho châu Âu), *Pectinophora gossypiella* (sâu hồng hại bông), *Phyllocnistis citrella* (sâu ăn lá cam), *Pieris brassicae* (bướm trắng lớn), *P. rapae* (sâu bắp cải nhập khẩu, hoặc bướm trắng nhỏ), *Plutella xylostella* (sâu tơ cải bắp), *Spodoptera exigua* (sâu xanh da láng), *S. litura* (sâu chồi thuốc lá, sâu bướm cụm), *S. frugiperda* (sâu xanh mùa thu), và *Tuta absoluta* (sâu ăn lá cà chua).

Các protein được mô tả ở đây cũng có thể được sử dụng để tạo ra các kháng thể liên kết đặc hiệu với các protein độc tố đặc hiệu BCW và có thể được sử dụng để sàng lọc và để phát hiện các thành viên khác của nhóm độc tố BCW.

Trong một phương án, các polynucleotit được lấy làm ví dụ mà mã hóa protein liên quan đến BCW 001 ức chế côn trùng được chỉ ra trong SEQ ID NO:1, 3, 5, 7, 9 và 11. Các trình tự nucleotit mã hóa các protein này có thể được sử dụng làm đoạn dò và đoạn mồi để sàng lọc để xác định các thành viên khác của nhóm này bằng cách sử dụng phương pháp lai và/hoặc khuếch đại nhiệt hoặc đẳng nhiệt và các phương pháp xác định khác đã biết đối với người có kiến thức trung bình trong lĩnh vực này.

Một khía cạnh của sáng chế đề xuất các phương pháp để phát hiện các protein liên quan, và các phương pháp này bao gồm bước giải trình tự bộ gen *Bt*, tập hợp dữ liệu trình tự, nhận diện và tách dòng các gen *Bt* mã hóa các protein diệt côn trùng bệnh này, và biểu hiện và kiểm tra các protein *Bt* mới để thử nghiệm hoạt tính diệt côn trùng. Một khía cạnh khác của sáng chế sử dụng các phương pháp phân tử để thao tác di truyền và tách dòng các protein hữu ích thương mại chứa các thể khảm của các protein thuộc nhóm protein diệt côn trùng, ví dụ, các thể khảm này có thể được lắp ráp từ các đoạn của các protein gây độc BCW để thu được các phương án bổ sung. Các protein được mô tả có thể được sắp xếp thẳng hàng với nhau và với các protein diệt côn trùng *Bt* khác, và các đoạn của mỗi protein này có thể được xác định là hữu ích để thay thế giữa các protein đã sắp xếp thẳng hàng, tạo ra cấu trúc của protein khảm. Các protein khảm này có thể được đưa vào phân tích thử nghiệm sinh học vật gây hại và được mô tả về sự có mặt của hoạt tính sinh học tăng cường và/hoặc phổ gây hại đích được mở rộng so với các protein mẹ mà từ đó mỗi đoạn trong thể khảm thu được. Hoạt tính diệt côn trùng của các polypeptit có thể được thao tác di truyền thêm cho hoạt tính đối với vật gây hại cụ thể hoặc đối với phổ gây hại rộng hơn bằng cách

hoán đổi các miền hoặc các đoạn với các protein khác.

Trong một phương án, các protein được mô tả ở đây bao gồm các mảnh tương đương về mặt chức năng (đột biến mất đầu cùng N hoặc C) của các protein được mô tả ở đây.

Các protein độc tố BCW được đề xuất ở đây. Trong một số phương án, protein độc tố liên quan đến BCW 001 có thể được phân lập, có thể được cung cấp trong chế phẩm, trong vi sinh vật chuyển gen, hoặc trong thực vật chuyển gen. Trong phương án này, protein BCW 002 và đặc biệt là protein BCW 003 gây ra hoạt tính ức chế *Lepidoptera*, cụ thể là hoạt tính ức chế kháng lại sâu xám và/hoặc sâu đục thân mía. Trong bản mô tả này khi đề cập đến “protein được phân lập”, hoặc thuật ngữ hoặc cụm từ tương đương, được dự định có nghĩa là protein này là protein tồn tại một mình hoặc kết hợp với các thành phần khác, nhưng không nằm trong môi trường tự nhiên của nó. Ví dụ, các protein độc tố theo sáng chế, và các protein tương tự, mà được tìm thấy tự nhiên trong sinh vật không được xem là “được phân lập” chỉ khi chúng nằm trong sinh vật mà trong đó chúng được tìm thấy một cách tự nhiên. Tuy nhiên, mỗi protein này sẽ được xem là “được phân lập” trong phạm vi sáng chế khi protein này không nằm trong sinh vật mà trong đó nó được tìm thấy một cách tự nhiên.

“Được liên kết hoạt động” như được sử dụng ở đây đề cập đến việc nối các trình tự axit nucleic hoặc các trình tự axit amin sao cho một trình tự có thể tạo ra chức năng cần thiết hoặc đặc điểm thích hợp hoặc hữu ích cho trình tự được liên kết.

Peptit, polypeptit, và protein tương đương về chức năng sinh học với BCW 001, BCW 002 và BCW 003 bao gồm, nhưng không bị giới hạn ở, các trình tự axit amin chứa các đột biến thể axit amin bảo tồn trong các trình tự protein độc tố BCW này. Trong các trình tự axit amin này, một hoặc nhiều axit amin trong trình tự này được thay thế bằng (các) axit amin khác, gây ra thay đổi âm hoặc thay đổi trình tự axit amin bảo tồn.

Mặc dù tốt hơn, nếu các polypeptit ức chế côn trùng đã mô tả chứa trình tự protein BCW 001, BCW 002 hoặc BCW 003, các mảnh và các biến thể của trình tự này có hoạt tính ức chế côn trùng giống hoặc tương tự với của protein ức chế côn trùng này cũng được mô tả ở đây. Ví dụ, các trình tự liên ứng chứa ít nhất là 30, 35, 38, 40, 45, 50, 55, 60, 65, 70, 75, 100, 150, 200, 500, 550 hoặc nhiều axit amin trong protein độc tố liên quan đến

BCW có hoạt tính ức chế côn trùng. Trong một phương án khác, các mảnh của protein độc tố liên quan đến BCW có hoạt tính ức chế côn trùng có thể chứa các đột biến thể, đột biến mất đoạn, đột biến gắn xen hoặc đột biến bổ sung axit amin trong trình tự protein độc tố BCW.

Trong một phương án, polypeptit ức chế côn trùng chứa đoạn ức chế côn trùng từ khoảng các gốc 28 đến khoảng gốc 618 của trình tự protein BCW 001 như được chỉ ra trong SEQ ID NO:2. Ví dụ không giới hạn bao gồm bất kỳ một trong số các SEQ ID NO:2, 4, hoặc 6, hoặc các mảnh ngắn hơn, hoặc các biến thể có hoạt tính ức chế côn trùng giống hoặc tương tự với của protein BCW 001 cụ thể chính nó hoặc trong liên kết hoạt động trong protein khảm. Trong một phương án khác, các đoạn có trình tự axit amin liên ứng chứa ít nhất khoảng 38 hoặc nhiều axit amin của bất kỳ một trong số SEQ ID NO:2, 4, hoặc 6 có hoạt tính ức chế côn trùng cũng tạo ra protein diệt côn trùng hữu ích. Các mảnh gây độc BCW 001 ức chế côn trùng cũng có thể bao gồm các đoạn với ít nhất là 30, 35, 38, 40, 45, 50, 100, 150, 200, 500, 550, 555, 560, 565, 570, 572, 574, 580 hoặc 585 gốc axit amin của vùng 591 axit amin tương ứng với khoảng các gốc 28 đến khoảng 618 của các trình tự của bất kỳ một trong số SEQ ID NO:2, 4, hoặc 6.

Trong một số phương án, các đoạn của protein BCW 001 thành thực (thành thực, có nghĩa là dạng tiền độc tố của protein là 1180 axit amin, bị cắt bởi quá trình phân giải protein ở ruột côn trùng gây hại để giải phóng độc tố lõi chỉ từ đầu cùng N đến các gốc 607 qua khoảng gốc 618, giải phóng đoạn độc tố hoạt động chứa, nhiều hơn hoặc ít hơn, các gốc 1 đến gốc 606 hoặc số lượng bất kỳ của các gốc từ khoảng 5 đến khoảng 618, như được chỉ ra trong SEQ ID NO:2, miễn là đoạn được giải phóng thể hiện đặc điểm gây độc cho ấu trùng sâu xám) có thể là dạng bị cắt trong đó một hoặc nhiều axit amin bị mất khỏi đầu cùng N, đầu cùng C, đoạn giữa của protein, hoặc kết hợp của chúng có hoạt tính ức chế côn trùng. Các đoạn này có thể là các biến thể tự nhiên hoặc tổng hợp của BCW 001, và vẫn duy trì hoạt tính ức chế côn trùng của protein độc tố BCW. Trong một số phương án, các mảnh của protein BCW 001, BCW 002, hoặc BCW 003 thành thực thể hiện hoạt tính diệt côn trùng của các phân tử protein khởi đầu mà từ đó chúng thu được. Mảnh hoặc biến thể được mô tả ở đây có thể chứa thêm miền được xác định ở đây mà chịu trách nhiệm về hoạt tính diệt côn trùng của protein này. Dẫn xuất bị cắt có hoạt tính ức chế côn trùng là

protein độc tố BCW tương ứng với các gốc từ khoảng 28 đến khoảng 606 hoặc đến khoảng 618 của trình tự protein độc tố BCW 001 như được chỉ ra trong SEQ ID NO:2 hoặc tương ứng với các gốc từ khoảng 29 đến khoảng 607 và đến gốc 619 của protein độc tố BCW như được chỉ ra trong SEQ ID NO:4 hoặc SEQ ID NO:6.

Vẫn trong một phương án khác, các đột biến mất đoạn đầu cùng N bị cắt bao gồm, nhưng không bị giới hạn ở, các protein gây độc BCW 001 mà thiếu các gốc axit amin từ đầu cùng N và/hoặc đầu cùng C của phần độc tố không có tiền độc tố, hoặc lõi gây độc của protein độc tố BCW. Ví dụ, từ 1 đến 6 gốc axit amin đầu cùng N của lõi gây độc của protein BCW 001 tương ứng với các gốc từ 28 đến 618 của các SEQ ID NO:2 hoặc tương ứng với các gốc 29 đến 619 của SEQ ID NO:4 hoặc 6 có thể bị xóa bỏ. Các đột biến mất đoạn đầu cùng C bị cắt của protein độc tố BCW tương ứng với các gốc 28 đến 618 của SEQ ID NO:2 hoặc các gốc 29 đến 619 của SEQ ID NO:4 hoặc 6 bao gồm, nhưng không bị giới hạn ở, các protein độc tố BCW thiếu từ 1 đến 6 gốc axit amin đầu cùng C. Trong các phương án khác, protein độc tố BCW với các gốc tương ứng 28 đến 618 của SEQ ID NO:2 hoặc các gốc tương ứng từ 29 đến 619 của SEQ ID NO:4 hoặc 6 có thể có cả sự cắt đầu cùng N từ 1 đến 6 gốc đầu cùng amino và sự cắt đầu cùng C từ 1 đến 6 gốc đầu cùng carboxy.

Trong một số phương án, các đoạn đơn lẻ từ 1 đến 6 của protein CPR24719, hoặc tổ hợp của các đoạn từ 1 đến 6, mà gây ra hoạt tính kháng sâu xám cho protein khác với CPR24719-1, cũng có thể thể hiện chức năng giống hoặc tương tự.

Các mảnh và các biến thể của protein độc tố BCW được mô tả ở đây có thể có tỷ lệ đồng nhất trình tự khoảng 62% hoặc lớn hơn, tỷ lệ đồng nhất trình tự khoảng 65%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98%, hoặc lớn hơn, và tỷ lệ đồng nhất trình tự axit amin khoảng 99%, 99,5%, 100%, với các đoạn tương ứng của protein độc tố BCW thành thực có các trình tự axit amin tương ứng được thể hiện trong các gốc 28 đến 618 của SEQ ID NO:2 hoặc với các gốc 29 đến 619 của SEQ ID NO:4 hoặc 6.

Một phương án của sáng chế bao gồm các chế phẩm polynucleotit tái tổ hợp mã hóa protein độc tố BCW. Ví dụ, protein độc tố BCW có thể được biểu hiện với các cấu trúc ADN tái tổ hợp trong đó phân tử polynucleotit đã phân lập với khung đọc mở mã hóa protein này được liên kết hoạt động với các yếu tố như trình tự khởi đầu và yếu tố điều hòa khác bất kỳ hoạt động để biểu hiện trong hệ thống mà cấu trúc này được dự định sử dụng

cho. Khi bản mô tả này đề cập đến “phân tử ADN được phân lập”, hoặc thuật ngữ hoặc cụm từ tương đương, được dự định có nghĩa là phân tử ADN này là phân tử ADN tồn tại một mình hoặc kết hợp với các thành phần khác, nhưng không nằm trong môi trường tự nhiên của nó. Ví dụ, các yếu tố axit nucleic như trình tự mã hóa, trình tự intron, trình tự dẫn đầu không được dịch mã, trình tự khởi đầu, trình tự kết thúc phiên mã, và tương tự, mà được tìm thấy tự nhiên trong ADN của bộ gen của sinh vật không được xem là được “phân lập” khi yếu tố này nằm trong bộ gen của sinh vật và ở vị trí trong bộ gen mà trong đó nó được tìm thấy một cách tự nhiên. Tuy nhiên, mỗi yếu tố này, và tiểu phần của các yếu tố này, sẽ là “được phân lập” trong phạm vi của phần mô tả này khi yếu tố này không nằm trong bộ gen của sinh vật và ở vị trí bên trong bộ gen mà trong đó nó được tìm thấy một cách tự nhiên. Tương tự, trình tự nucleotit mã hóa protein diệt côn trùng hoặc biến thể diệt côn trùng tự nhiên bất kỳ của protein đó sẽ là trình tự nucleotit được phân lập với điều kiện là trình tự nucleotit này không nằm trong ADN của vi khuẩn mà từ đó trình tự mã hóa protein này được tìm thấy tự nhiên. Trình tự nucleotit tổng hợp mã hóa trình tự axit amin của protein diệt côn trùng tự nhiên sẽ được coi là được phân lập với các mục đích của sáng chế. Với các mục đích của sáng chế, trình tự nucleotit chuyển gen bất kỳ, nghĩa là, trình tự nucleotit của ADN được gắn chèn vào bộ gen của các tế bào của thực vật hoặc vi khuẩn, hoặc có mặt trong vectơ ngoài nhiễm sắc thể, sẽ được coi là trình tự nucleotit được phân lập dù nó xuất hiện trong plasmid hay cấu trúc tương tự được sử dụng để biến nạp các tế bào này, trong bộ gen của thực vật hoặc vi khuẩn, hay có mặt với lượng phát hiện được trong mô, dòng con, các mẫu sinh học hoặc các thương phẩm thu được từ thực vật hoặc vi khuẩn này. Ví dụ không giới hạn bao gồm các trình tự khởi đầu hoạt động ở thực vật được liên kết hoạt động với các trình tự mã hóa protein độc tố BCW để biểu hiện protein này ở thực vật hoặc các trình tự khởi đầu hoạt động *Bt* được liên kết hoạt động với các trình tự mã hóa protein độc tố BCW để biểu hiện protein trong *Bt*. Các yếu tố khác mà có thể được liên kết hoạt động với các trình tự mã hóa protein độc tố BCW bao gồm, nhưng không bị giới hạn ở, yếu tố tăng cường, intron, trình tự dẫn đầu, các thể bất động protein được mã hóa (HIS-tag), các peptit chuyển vị dưới mức tế bào được mã hóa (ví dụ, các peptit chuyển lap thể, peptit tín hiệu), các vị trí polypeptit được mã hóa cho các enzym biến đổi sau dịch mã, vị trí liên kết ribosom, và các vị trí đích ARNi.

Như được sử dụng ở đây, “phân tử ADN tái tổ hợp” là phân tử ADN chứa tổ hợp của các phân tử ADN mà không xuất hiện tự nhiên cùng nhau nếu không có sự can thiệp của con người. Ví dụ, phân tử ADN tái tổ hợp có thể là phân tử ADN mà bao gồm ít nhất hai phân tử ADN khác loại với nhau, một phân tử ADN chứa trình tự ADN mà có nguồn gốc từ các trình tự ADN tồn tại trong tự nhiên, hoặc một phân tử ADN được đưa vào ADN của tế bào chủ bằng kỹ thuật biến nạp di truyền hoặc biên tập gen. Tương tự, “phân tử protein tái tổ hợp” là phân tử protein chứa tổ hợp của các axit amin mà không xuất hiện tự nhiên cùng nhau nếu không có sự can thiệp của con người. Ví dụ, phân tử protein tái tổ hợp có thể là phân tử protein mà bao gồm ít nhất hai phân tử axit amin khác loại đối với nhau, phân tử protein mà chứa trình tự axit amin có nguồn gốc từ các trình tự axit amin tồn tại trong tự nhiên, hoặc phân tử protein mà được biểu hiện trong tế bào chủ nhờ kỹ thuật biến nạp di truyền tế bào chủ hoặc bằng cách biên tập gen đối với bộ gen tế bào chủ.

Các phân tử polynucleotit tái tổ hợp được lấy làm ví dụ được đề xuất ở đây bao gồm, nhưng không bị giới hạn ở, SEQ ID NO:7, SEQ ID NO:9, và SEQ ID NO:11 cũng như mỗi đoạn nucleotit được chỉ ra trong SEQ ID NO:3 và SEQ ID NO:5 và mã hóa các polypeptit hoặc các protein tương ứng có trình tự axit amin như được chỉ ra trong SEQ ID NO:2 (BCW 001), SEQ ID NO:4 (BCW 002), SEQ ID NO:6 (BCW 003) và SEQ ID NO:8, SEQ ID NO:10 và SEQ ID NO:12. Các codon của phân tử polynucleotit tái tổ hợp mã hóa các protein được mô tả ở đây có thể được thay thế bằng các codon đồng nghĩa (còn được gọi là đột biến im lặng). Các polynucleotit tái tổ hợp mã hóa bất kỳ trong số các protein biến đổi độc tố BCW được mô tả ở đây cũng được đề xuất.

Cấu trúc ADN tái tổ hợp chứa trình tự mã hóa protein độc tố BCW cũng có thể bao gồm thêm vùng ADN mã hóa một hoặc nhiều chất ức chế côn trùng mà có thể được cấu tạo để biểu hiện đồng thời hoặc đồng biểu hiện với trình tự ADN mã hóa protein độc tố BCW, protein khác với protein độc tố BCW, phân tử dsARN ức chế côn trùng, hoặc hóa chất diệt côn trùng. Ví dụ không giới hạn về các hóa chất diệt côn trùng là các clorua hữu cơ, phosphat hữu cơ và carbamat, pyrethroid, neonicotinoit, và ryanoit.

Cấu trúc ADN tái tổ hợp có thể được lắp ráp nhờ đó toàn bộ các protein hoặc các phân tử dsARN được biểu hiện từ một trình tự khởi đầu hoặc mỗi protein hoặc phân tử dsARN là dưới sự kiểm soát của các trình tự khởi đầu riêng biệt hoặc một số tổ hợp của

chúng. Các protein được mô tả ở đây có thể được biểu hiện từ hệ biểu hiện đa gen trong đó một hoặc nhiều protein được mô tả ở đây được biểu hiện từ đoạn nucleotit chung trên đó cũng chứa các khung đọc mở và/hoặc các trình tự khởi đầu khác phụ thuộc vào kiểu của hệ biểu hiện được chọn. Ví dụ, hệ biểu hiện đa gen vi khuẩn có thể sử dụng một trình tự khởi đầu duy nhất để điều khiển sự biểu hiện của các khung đọc mở được liên kết bội/song song từ bên trong một operon duy nhất. Trong một ví dụ khác, hệ biểu hiện đa gen thực vật có thể sử dụng caxet biểu hiện được liên kết bội, mỗi caxet biểu hiện protein khác nhau hoặc chất khác như một hoặc nhiều phân tử dsARN. Trong một ví dụ khác, hệ biểu hiện đa gen thực vật có thể sử dụng caxet biểu hiện không được liên kết bội, mỗi caxet biểu hiện protein khác nhau hoặc chất khác như một hoặc nhiều phân tử dsARN. Trình tự khởi đầu để sử dụng trong axit nucleic tái tổ hợp được mô tả ở đây có thể bao gồm trình tự khởi đầu đầy đủ hoặc biến thể hoặc mảnh bất kỳ của nó có hoạt tính trình tự khởi đầu hoặc hoạt tính điều hòa gen.

Polynucleotit tái tổ hợp hoặc cấu trúc ADN tái tổ hợp chứa trình tự mã hóa protein độc tố BCW có thể được phân phối cho tế bào chủ bằng các vectơ, ví dụ, plasmid, baculovirus, nhiễm sắc thể nhân tạo, virion, cosmid, phagemid, thể thực khuẩn, hoặc vectơ virus. Các vectơ này có thể được sử dụng để đạt được sự biểu hiện ổn định hoặc tạm thời của trình tự mã hóa protein độc tố BCW trong tế bào chủ, hoặc sự biểu hiện tiếp thành polypeptit. Cấu trúc polynucleotit tái tổ hợp hoặc ADN tái tổ hợp ngoại sinh mà chứa trình tự mã hóa protein độc tố BCW và được đưa vào tế bào chủ được đề cập trong sáng chế là “gen chuyển”.

Cũng được đề xuất ở đây là vi khuẩn chuyển gen, tế bào thực vật chuyển gen, thực vật chuyển gen, nấm và nấm men, và các bộ phận thực vật chuyển gen mà chứa polynucleotit tái tổ hợp bất kỳ biểu hiện bất kỳ một hoặc nhiều trình tự mã hóa protein độc tố BCW được đề xuất ở đây. Thuật ngữ “tế bào vi khuẩn” hoặc “vi khuẩn” có thể bao gồm, nhưng không bị giới hạn ở, tế bào *Agrobacterium*, *Bacillus*, *Escherichia*, *Salmonella*, *Pseudomonas*, hoặc tế bào *Rhizobium*. Thuật ngữ “tế bào thực vật” hoặc “thực vật” có thể bao gồm nhưng không bị giới hạn ở, cây hoặc tế bào của cây cỏ linh lăng, chuối, lúa mạch, đậu, bông cải xanh, bắp cải, cải, cà rốt, sắn, thầu dầu, bông cải trắng, cần tây, đậu gà, cải thảo, cam, dưa, cà phê, ngô, cỏ ba lá, bông, bầu bí, dưa chuột, gỗ linh sam, cà tím, bạch

đàn, cây lanh, tỏi, nho, cây hublông, tỏi tây, rau diếp, thông trăm hương, kê, quả dưa, quả hạch, yến mạch, oliu, hành tây, cây cảnh, cò, cỏ, đậu Hà Lan, củ lạc, cây ớt, đậu triều, thông, khoai tây, bạch dương, bí đỏ, thông Radiata, củ cải, hạt cải dầu, lúa, cây thân rễ, lúa mạch đen, hồng hoa, cây bụi, cây cao lương, thông phương Nam, đậu tương, rau chân vịt, quả bí, dâu tây, củ cải đường, mía, hoa hướng dương, ngô ngọt, gỗ gôm, khoai lang, cỏ kê Mỹ, chè, thuốc lá, cà chua, tiểu hắc mạch, cỏ, dưa hấu, và lúa mỳ. Trong một số phương án, thực vật chuyển gen và các bộ phận của thực vật chuyển gen được tái sinh từ tế bào thực vật chuyển gen được đề xuất. Trong một số phương án, thực vật chuyển gen có thể thu được từ hạt chuyển gen, bằng cách nhân giống, cắt, bẻ gãy, nghiền hoặc theo cách khác tách bộ phận này ra khỏi thực vật. Trong một số phương án, bộ phận của thực vật có thể là hạt, quả nang, lá, hoa, thân, rễ hoặc phần bất kỳ của chúng, hoặc phần không thể tái sinh của bộ phận của thực vật chuyển gen. Như được sử dụng trong ngữ cảnh này, phần “không thể tái sinh” của bộ phận của thực vật chuyển gen là phần mà không thể được cảm ứng để tạo ra thực vật hoàn chỉnh hoặc phần không thể được cảm ứng để tạo ra thực vật hoàn chỉnh có khả năng sinh sản hữu tính và/hoặc vô tính. Trong một số phương án, phần không tái sinh được của bộ phận thực vật là một phần của hạt, quả nang, lá, hoa, thân hoặc rễ chuyển gen.

Cũng được đề xuất ở đây là các phương pháp tạo ra thực vật chuyển gen chứa lượng có tác dụng ức chế côn trùng hoặc *Lepidoptera* của protein độc tố BCW. Các thực vật này có thể được tạo ra bằng cách đưa polynucleotit tái tổ hợp mà mã hóa bất kỳ trong số các protein độc tố BCW được đề xuất ở đây vào tế bào thực vật, và chọn thực vật thu được từ tế bào thực vật này mà biểu hiện lượng có tác dụng ức chế côn trùng hoặc *Lepidoptera* của protein độc tố BCW. Các cây có thể thu được từ các tế bào thực vật này bằng cách tái sinh, gieo hạt, thụ phấn hoặc các kỹ thuật biến nạp mô phân sinh.

Các quy trình để biến nạp thực vật là đã biết trong lĩnh vực này. Ví dụ, quá trình biến nạp qua trung gian *Agrobacterium* được mô tả trong Công bố đơn yêu cầu cấp patent Mỹ 2009/0138985A1 (đậu tương), 2008/0280361A1 (đậu tương), 2009/0142837A1 (ngô), 2008/0282432 (bông), 2008/0256667 (bông), 2003/0110531 (lúa mỳ), 2001/0042257 A1 (củ cải đường), Patent Mỹ số 5,750,871 (cây cải dầu), 7,026,528 (lúa mỳ), và 6,365,807 (lúa), và trong Arencibia *et al.* (1998) Transgenic Res. 7:213-222 (cây mía).

Cũng được đề xuất ở đây là việc sử dụng thực vật chuyển gen biểu hiện lượng có tác dụng ức chế côn trùng hoặc *Lepidoptera* của protein độc tố BCW để phòng trừ sự phá hoại của côn trùng hoặc *Lepidoptera*. Bất kỳ trong số các thực vật chuyển gen đã đề cập trên đây có thể được sử dụng trong các phương pháp bảo vệ thực vật khỏi sự phá hoại của côn trùng hoặc *Lepidoptera* được đề xuất ở đây. Các phương pháp thu được thực vật chuyển gen biểu hiện protein có tác dụng chống *Lepidopteran* như protein Cry1A (patent Mỹ số 5,880,275), Cry1B (đơn yêu cầu cấp patent Mỹ số 10/525318), Cry1C (patent Mỹ số 6,033,874), Cry1F, thể khảm Cry1A/F (patent Mỹ số 7,070,982; 6,962,705, và 6,713,063), và protein Cry2Ab (patent Mỹ số 7,064,249) cũng được mô tả rõ ràng.

Cũng được đề xuất ở đây là việc sử dụng bất kỳ trong số các tế bào chủ chuyển gen đã đề cập trên đây để sản xuất protein độc tố BCW.

Các khía cạnh bổ sung của sáng chế bao gồm các kháng thể và các phương pháp phát hiện polynucleotit mã hóa protein độc tố BCW hoặc phân biệt các mảnh và các đoạn của chúng, các phương pháp nhận diện các thành viên ức chế côn trùng bổ sung của lớp protein này, các chế phẩm và phương pháp phòng trừ sự sinh trưởng và/hoặc sự phá hoại của côn trùng, và các phương pháp cung cấp tác dụng phòng trừ này cho thực vật và các vật chủ nhận khác.

Trong một số phương án, sản phẩm thực vật có thể bao gồm các thương phẩm hoặc các sản phẩm thương mại khác thu được từ thực vật chuyển gen hoặc bộ phận của thực vật chuyển gen, trong đó thương phẩm hoặc các sản phẩm khác có thể được theo dõi qua thương mại bằng cách phát hiện các đoạn nucleotit hoặc ARN được biểu hiện hoặc các protein mà mã hóa hoặc chứa các phần phân biệt của protein độc tố BCW. Hàng hóa hoặc các sản phẩm thương mại khác như vậy bao gồm, nhưng không bị giới hạn ở, bộ phận thực vật, sinh khối, dầu, bột xay thô, đường, thức ăn cho động vật, bột mỳ, tấm, cám, xơ vải, hạt đã xử lý, và hạt giống.

Cũng được đề xuất ở đây là các sản phẩm thực vật đã xử lý trong đó sản phẩm đã xử lý này chứa lượng phát hiện được của polynucleotit tái tổ hợp mã hóa protein độc tố BCW, một đoạn của nó, mảnh có tác dụng ức chế côn trùng của nó, hoặc phần phân biệt bất kỳ của nó. Trong một số phương án, sản phẩm đã xử lý này được chọn từ nhóm bao gồm: sinh khối thực vật, dầu, bột thô, thức ăn cho động vật, bột mỳ, tấm, cám, xơ vải, trấu,

và hạt đã xử lý. Trong một số phương án, sản phẩm đã xử lý này là không thể tái sinh được.

Cũng được đề xuất ở đây là các phương pháp phòng trừ côn trùng. Trong một số phương án, sự phá hoại của *Lepidoptera* ở cây trồng được phòng trừ. Các phương pháp này có thể bao gồm trồng cây chứa lượng có tác dụng ức chế côn trùng hoặc *Lepidoptera* của protein độc tố BCW. Trong một số phương án, các phương pháp này có thể bao gồm thêm bất kỳ một hoặc nhiều trong số: (i) sử dụng chế phẩm bất kỳ chứa hoặc mã hóa protein độc tố BCW cho cây hoặc hạt mà sẽ phát triển thành cây; và/hoặc (ii) biến nạp cây này hoặc tế bào thực vật mà sẽ phát triển thành cây này bằng polynucleotit mã hóa protein độc tố BCW. Trong một số phương án, cây này là cây chuyển gen được biến nạp ổn định hoặc tạm thời chứa gen chuyển biểu hiện lượng có tác dụng ức chế côn trùng hoặc *Lepidoptera* của protein độc tố BCW. Trong một số phương án, cây này là cây không chuyển gen mà chế phẩm chứa protein độc tố BCW đã được sử dụng cho. Trong một số phương án của các phương pháp này, cây này là cây ngô hoặc cây mía. Trong một số phương án, loài *Lepidoptera* là *Agrotis ipsilon*. Trong một số phương án, loài *Lepidoptera* là *Diatraea saccharalis*. Trong một số phương án, loài *Lepidoptera* là ở cánh đồng trồng cây.

Việc làm giàu các protein được mô tả ở đây trong các thực vật hoặc bằng quy trình có thể bao gồm nuôi cấy tế bào *Bt* tái tổ hợp trong các điều kiện để biểu hiện/sản sinh polypeptit/protein tái tổ hợp. Quy trình này có thể bao gồm bào chế bằng cách làm khô, làm đông khô, đông nhất hóa, chiết, lọc, ly tâm, làm sa lắng, hoặc cô đặc môi trường nuôi cấy các tế bào *Bt* tái tổ hợp biểu hiện/sản sinh polypeptit tái tổ hợp này. Quy trình này có thể tạo ra dịch chiết tế bào *Bt*, huyền phù tế bào, dịch đồng nhất tế bào, dịch dung giải tế bào, dịch nổi bề mặt chứa tế bào, dịch lọc tế bào, hoặc pelet tế bào *Bt*. Bằng cách thu được polypeptit/protein tái tổ hợp được sản xuất như vậy, chế phẩm chứa polypeptit/protein tái tổ hợp này có thể bao gồm các tế bào vi khuẩn, các bào tử vi khuẩn, và các thể vùi bao quanh bào tử và có thể được bào chế cho nhiều đường sử dụng khác nhau, bao gồm sản phẩm phun ức chế côn trùng nông nghiệp hoặc dưới dạng chế phẩm ức chế côn trùng trong thử nghiệm sinh học chế độ ăn.

Trong một phương án, chế phẩm/dạng bào chế ức chế côn trùng chứa polypeptit/protein tái tổ hợp đã mô tả có thể chứa thêm ít nhất một polypeptit bổ sung thể hiện hoạt tính ức chế côn trùng chống lại cùng một loài côn trùng *Lepidopteran*, nhưng

khác với polypeptit tái tổ hợp này để làm giảm tỷ lệ kháng của côn trùng *Lepidopteran* với protein độc tố BCW hoặc chế phẩm ức chế côn trùng *Lepidopteran* khác. Polypeptit này được chọn từ nhóm bao gồm: protein ức chế côn trùng, phân tử dsARN ức chế côn trùng, và hóa chất. Một ví dụ về việc sử dụng các trình tự ribonucleotit này để phòng trừ côn trùng gây hại được mô tả trong công bố đơn yêu cầu cấp patent Mỹ số 2006/0021087. Ví dụ về các chế phẩm khác này bao gồm, nhưng không bị giới hạn ở, Cry1A (patent Mỹ số 5,880,275), Cry1Ab, Cry1Ac, Cry1Ae, Cry1B (đơn patent Mỹ số 10/525,318), Cry1C (patent Mỹ số 6,033,874), Cry1E, Cry1F, và các thể khảm Cry1A/F (patent Mỹ số 7,070,982; 6,962,705; và 6,713,063), Cry1G, Cry1H, Cry1I, Cry1J, Cry1K, Cry1L, Cry2A, Cry2Ab (patent Mỹ số 7,064,249), Cry2Ae, Cry4B, Cry6, Cry7, Cry8, Cry9, Cry15, Cry43A, Cry43B, ET35, ET66, TIC400, TIC800, TIC807, TIC834, TIC853 và TIC1415. Các ví dụ không giới hạn khác là protein có tác dụng chống *Lepidoptera* VIP, Axmi và DIG như, nhưng không giới hạn ở, Vip3A, VIP3Ab, AXMI-184, AXMI-196, DIG-3, DIG-4, DIG-5, và DIG-11, mà có thể được kết hợp với các protein được mô tả ở đây.

Trong các phương án khác, chế phẩm này có thể bao gồm thêm ít nhất một polypeptit bổ sung thể hiện hoạt tính ức chế côn trùng đối với côn trùng không bị ức chế bởi protein độc tố BCW ức chế côn trùng khác để mở rộng phổ ức chế côn trùng thu được. Ví dụ, để phòng trừ vật gây hại *Coleopteran*, các hỗn hợp của các protein độc tố BCW ức chế côn trùng có thể được sử dụng với các protein có tác dụng chống *Coleopteran* như, nhưng không giới hạn ở, các biến thể Cry1C, các biến thể Cry3A, Cry3Bb (patent Mỹ số 6,501,009), Cry34/35, 5307, Axmi184, Axmi205, AxmiR1, TIC407, TIC417, TIC431, TIC901, TIC1201, TIC3131, DIG-10 và eHIP (Công bố đơn patent Mỹ số 2010/0017914).

Khả năng phát triển tính kháng của côn trùng với một số thuốc diệt côn trùng đã được ghi lại trong lĩnh vực này. Một chiến lược quản lý tính kháng của côn trùng là sử dụng cây trồng chuyển gen biểu hiện hai chất ức chế côn trùng khác nhau mà hoạt động nhờ các cơ chế hoạt động khác nhau. Do đó, côn trùng bất kỳ có khả năng kháng lại một trong các chất ức chế côn trùng có thể được phòng trừ bởi chất ức chế côn trùng còn lại. Một chiến lược quản lý tính kháng của côn trùng khác sử dụng ứng dụng của các thực vật không được bảo vệ cho các vật gây hại *Lepidopteran* để tạo ra nơi ẩn náu. Một ví dụ cụ thể được mô tả trong patent Mỹ số 6,551,962.

Các phương án khác được mô tả ở đây bao gồm các hóa chất diệt côn trùng sử dụng tại chỗ được thiết kế để phòng trừ vật gây hại mà cũng được phòng trừ bởi các protein được mô tả ở đây, để được sử dụng cùng với các protein trong các quy trình xử lý hạt, chế phẩm phun lên, nhỏ giọt lên hoặc lau có thể được sử dụng trực tiếp cho đất (thấm ướt đất), được sử dụng cho cây đang phát triển biểu hiện các protein được mô tả ở đây, hoặc được bào chế để sử dụng được cho hạt chứa một hoặc nhiều gen chuyển mã hóa một hoặc nhiều protein đã được mô tả. Các chế phẩm để sử dụng trong các quy trình xử lý hạt có thể được sử dụng cùng với nhiều nhân dính và chất kết dính đã biết trong lĩnh vực này. Các chế phẩm này có chứa thuốc diệt côn trùng tác dụng hiệp đồng về MOA với các protein đã được mô tả, nhờ đó thuốc diệt côn trùng trong chế phẩm tác dụng bằng MOA khác để phòng trừ cùng một loài sâu bệnh hoặc loài sâu bệnh tương tự mà có thể được phòng trừ bởi protein đã được mô tả, hoặc thuốc diệt côn trùng này tác dụng để phòng trừ vật gây hại trong phạm vi vật chủ rộng hơn, như loài *Lepidopteran* hoặc *Hemipteran* hoặc loài sâu bệnh trên thực vật khác như loài *Coleopteran* mà chưa được phòng trừ hiệu quả.

Chế phẩm/dạng bào chế đã đề cập trên đây có thể chứa thêm chất mang được chấp nhận trong nông nghiệp, như môi, bột, bụi, pelet, hạt, dạng xịt, nhũ tương, huyền phù keo, dung dịch nước, chế phẩm bào tử/tinh thể *Bacillus*, chất xử lý hạt giống, tế bào thực vật/mô thực vật/hạt giống/cây tái tổ hợp được biến nạp để biểu hiện một hoặc nhiều trong số các protein này, hoặc vi khuẩn được biến nạp để biểu hiện một hoặc nhiều trong số các protein này. Tùy theo mức độ ức chế côn trùng hoặc mức độ ức chế diệt côn trùng vốn có trong polypeptit tái tổ hợp và mức độ mà chế phẩm cần được sử dụng cho thực vật hoặc thử nghiệm chế độ ăn, chế phẩm/dạng bào chế này có thể bao gồm nhiều lượng khác nhau theo khối lượng của polypeptit tái tổ hợp này, ví dụ, từ 0,0001% đến 0,001% đến 0,01% đến 1% đến 99% theo khối lượng polypeptit tái tổ hợp này.

Các protein được mô tả ở đây có thể được kết hợp trong các chế phẩm để sử dụng tại chỗ cho bề mặt thực vật, cho đất, trong các chế phẩm để xử lý hạt giống, và trong các chế phẩm với các chất khác gây độc cho các sâu bệnh đích thuộc loài *Lepidopteran*. Các chất này bao gồm nhưng không bị giới hạn ở: các protein Cry1A, Cry1B, Cry1C, Cry1F, thể khảm Cry1A/F, và protein Cry2Ab.

Ví dụ thực hiện sáng chế

Các phương án được mô tả sau đây chỉ là đại diện cho sáng chế, có thể được bao gồm ở nhiều dạng khác nhau. Do đó các chi tiết về cấu trúc và chức năng cụ thể được mô tả ở đây không được hiểu là làm giới hạn. Nên được hiểu là toàn bộ phần mô tả của mỗi tài liệu tham khảo được trích dẫn ở đây được đưa vào đây bằng cách viện dẫn.

Ví dụ 1

Ví dụ này trình bày về việc phát hiện và phân tích về protein độc tố BCW 001 và cấu trúc của các độc tố khảm BCW 002 và BCW 003.

Bt chủng EG4384 được xác định là gây ra hoạt tính *Lepidopteran* trong thử nghiệm sinh học chế độ ăn bằng cách sử dụng các chế phẩm tinh thể bào tử. Trình tự của bộ gen của chủng này được tạo ra, các trình tự đọc thô được xử lý, các đoạn trình tự liên tục được lắp ráp từ các trình tự đọc đã xử lý, các khung đọc mở được xác định là thể hiện sự tương đồng với protein Cry1, và các trình tự axit amin đã suy luận được phân tích. Khung đọc mở cụ thể như được chỉ ra trong SEQ ID NO:1 được xác định là mã hóa trình tự axit amin đã suy luận của protein (BCW 001, SEQ ID NO:2) thể hiện trình tự axit amin mới so với hầu hết các protein Cry1 đã biết trong lĩnh vực này. Protein đã suy luận từ khung đọc mở này có tất cả các đặc điểm của protein typ Cry1 mới, vì nó dài 1180 axit amin, và việc sắp xếp thẳng hàng các protein Cry1 đã biết chỉ ra rằng protein này có cấu trúc ba miền đặc trưng nằm trong xấp xỉ 600-630 axit amin đầu cùng amino, và cấu trúc trình tự axit amin tiền độc tố đặc trưng typ Cry1A. Trình tự polynucleotit mã hóa trình tự axit amin được dự đoán này chứa khung đọc mở mà cũng đặc trưng cho Cry1, nghĩa là, vị trí giới hạn NheI trong đoạn ADN mã hóa vùng đầu cùng C của Miền I được dự đoán của độc tố này, và vị trí giới hạn KpnI trong đoạn ADN mã hóa phần đầu cùng N của miền tiền độc tố được dự đoán này.

Việc so sánh trình tự axit amin của độc tố BCW 001 với Cry1Ac cho thấy là đoạn axit amin tương ứng với Miền I (các axit amin từ khoảng vị trí 1 đến khoảng vị trí 258) chỉ thể hiện tỷ lệ đồng nhất khoảng 67% với cùng đoạn đó trong Cry1Ac, đoạn axit amin tương ứng với Miền II (các axit amin từ khoảng vị trí 58 đến khoảng vị trí 460) thể hiện tỷ lệ phần trăm đồng nhất rất thấp với cùng một đoạn đó trong Cry1Ac, và trình tự axit amin

tương ứng với Miền III (các axit amin từ khoảng vị trí 460 đến khoảng vị trí 607) thể hiện tỷ lệ đồng nhất khoảng 63% với đoạn Miền III của Cry1Ah2.

Đoạn ADN chủ yếu mã hóa miền II và III đã dự đoán, từ vị trí giới hạn NheI qua vị trí giới hạn KpnI, được cắt ra và được thay thế cho đoạn tương ứng của trình tự mã hóa Cry1Ac trong vector biểu hiện chứa đoạn ADN mã hóa Cry1Ac, tạo ra khung đọc mở chứa, và liên kết trong khung theo thứ tự liên tục từ đầu 5' đến đầu 3', đoạn thứ nhất mã hóa Miền I của Cry1Ac, đoạn thứ hai mã hóa miền II và III của BCW 001, và đoạn thứ ba mã hóa miền tiền độc tố của protein độc tố Cry1Ac. Cấu trúc khảm này (SEQ ID NO:3) mã hóa protein độc tố khảm được gọi ở đây là BCW 002 (SEQ ID NO:4). Độ dịch điểm gãy giữa Miền I và Miền II tạo ra ở mức độ không đáng kể khung đọc mở (SEQ ID NO:5) mã hóa protein độc tố khảm được gọi ở đây là BCW 003 (SEQ ID NO:6), có trình tự axit amin khác với BCW002 chỉ ở vị trí axit 259. BCW 003, tương tự BCW 001, chứa threonin (T) ở vị trí 259 trong khi BCW 002 chứa isoleuxin (I) ở vị trí đó. BCW 001 khác biệt với BCW 002 và BCW 003 chủ yếu trong Miền I của độc tố, nghĩa là các axit amin 1-202, và BCW 002 và BCW 003, như đã khẳng định ở trên, gần như là đồng nhất ngoại trừ sự khác biệt I/T ở vị trí 259.

Ví dụ 2

Ví dụ này đề cập đến hoạt tính sinh học phòng trừ vật gây hại *Lepidopteran* một cách hữu hiệu của protein BCW 001, 002 và 003.

Các cấu trúc biến nạp biểu hiện protein độc tố BCW 001, 002 và 003 trong *E. coli* hoặc trong *Bacillus thuringiensis* có thể sử dụng được hoặc *Bacilli* khác cho phép kiểm tra các protein đã biểu hiện trong thử nghiệm sinh học và so sánh với các protein đã biết trong lĩnh vực này là gây độc cho sâu xám, như Cry1Fa và Cry1Ac. Các chủng tái tổ hợp thu được được quan sát thấy có sự biểu hiện protein tái tổ hợp có hoạt tính kháng sâu bệnh *Lepidopteran*. Hoạt tính thử nghiệm sinh học là đặc biệt mạnh khi được thử nghiệm trên sâu xám và ấu trùng sâu đục bắp trên cây ngô. Như đã chỉ ra trên đây trong phần mô tả chi tiết, phần tình trạng kỹ thuật của sáng chế và bản chất kỹ thuật của sáng chế, có rất ít các protein độc tố được phát hiện ra là thể hiện mức hoạt tính sinh học đáng kể kháng lại sâu xám, và do đó có nhu cầu trong lĩnh vực này về việc nhận diện các protein này để sử dụng trong thực vật để bảo vệ các thực vật này khỏi sự nhiễm sâu xám, và để đảm bảo là có sự

cung cấp đầy đủ protein có hoạt tính kháng sâu xám bổ sung sẵn có để ngăn ngừa được sự phát triển tính kháng bất kỳ với protein có tác dụng chống sâu xám bất kỳ hiện đang được sử dụng, như protein độc tố Cry1Fa.

Ví dụ 3

Ví dụ này mô tả là Miền II và III của BCW 001 là đủ để gây ra hoạt tính sinh học trên sâu xám cho các protein độc tố Cry1 khác khi các miền này được thay thế cho các miền tương ứng trong các độc tố Cry1 khác này.

Nhiều thể khảm độc tố BCW có hoạt tính kháng *Lepidoptera* được nhận diện, cụ thể là hai thể khảm thể hiện hoạt tính mạnh chống lại BCW, WBC, và SCB.

Các cấu trúc có trình tự nucleotit mã hóa Cry1Ab, Cry1Ac, và Cry1Ca được sử dụng để tạo cấu trúc các thể khảm chứa các đoạn Miền II và Miền III BCW 001 được thay thế cho các miền có thể sử dụng của Cry1Ab, Cry1Ac, và Cry1Ca, và các protein tự nhiên và khảm thu được được kiểm tra trong các thử nghiệm sinh học tinh thể bào tử. Hoạt tính kháng BCW, FAW và CEW trong các thử nghiệm sinh học chế độ ăn được lập bảng. Trong các điều kiện thử nghiệm đã kiểm tra, Cry1Ac thể hiện hoạt tính chống lại FAW, BCW và CEW, Cry1Ab thể hiện hoạt tính chống FAW và CEW, nhưng không chống lại BCW, và Cry1Ca không thể hiện hoạt tính chống lại FAW, BCW, cũng như CEW. BCW 001 thể hiện hoạt tính chống lại BCW và CEW, nhưng không chống lại FAW. So với BCW 003, hoạt tính chống lại BCW đối với Cry1Ac thấp hơn khoảng mười lần đối với Cry1Ac. Các thể khảm Cry1Ab, và Cry1Ac chứa Miền II và III của BCW 001 thể hiện các đặc điểm gây độc khi được thử nghiệm trong thử nghiệm sinh học kháng FAW, BCW và CEW. Cry1Ac không gây độc kháng lại CEW và Cry1Ab không gây độc kháng lại BCW. Thể khảm Cry1Ca /BCW 001 được tạo cấu trúc, trong đó Miền III Cry1Ca được thay thế cho miền tương ứng của BCW 001, và độc tố khảm tạo thành thể hiện các đặc điểm gây độc cho FAW, BCW và CEW, trong khi đó độc tố Cry1Ca là không hữu hiệu khi được thử nghiệm kháng lại bất kỳ trong số các sâu bệnh này.

Ví dụ 4

Ví dụ này minh họa các đặc điểm gây độc của BCW 001, 002 và 003 khi được kiểm tra trong thử nghiệm sinh học chống lại nhiều loại sâu bệnh *Lepidopteran*.

Các quy trình cho thử nghiệm sinh học và tính điểm côn trùng về tỷ lệ chết và choáng đã được biết trong lĩnh vực này, ví dụ của chúng được mô tả trong Công bố đơn PCT số WO 2012/139004 và trong patent Mỹ số 7,927,598.

Fig. 1 tạo mối tương quan giữa nhiều protein độc tố BCW 001, 002 và 003 khác nhau với hoạt tính diệt côn trùng theo các loài côn trùng trong thử nghiệm sinh học chế độ ăn. Mỗi protein độc tố này chứng minh hoạt tính chống lại côn trùng *Lepidopteran*.

Ví dụ 5

Ví dụ này mô tả cấu trúc của các trình tự nhân tạo mã hóa các protein theo sáng chế để sử dụng ở thực vật, tạo ra vector thực vật và các cấu trúc để sử dụng trong thực vật, và sự sản sinh các thực vật biểu hiện các protein theo sáng chế.

Các trình tự nucleotit mã hóa protein BCW 001 (SEQ ID NO:1), protein BCW 002 (SEQ ID NO:3), và protein BCW 003 (SEQ ID NO:5) được thiết kế và được tổng hợp theo các phương pháp được mô tả trong Patent Mỹ số 5,500,365. Các vùng mã hóa được thao tác di truyền này được thiết kế để biểu hiện ở thực vật được đề xuất ở đây là SEQ ID NO:7 mã hóa BCW 001, SEQ ID NO:9 mã hóa BCW 002, và SEQ ID NO:11 mã hóa BCW 003.

Nhiều caxet biểu hiện thực vật khác nhau được tạo cấu trúc với các trình tự như được chỉ ra trong SEQ ID NO:7, 9, và 11. Các caxet biểu hiện này là hữu ích để biểu hiện tạm thời trong tế bào trần của thực vật hoặc để biến nạp tế bào thực vật. Các caxet biểu hiện điển hình được thiết kế đối với việc đưa thực tế protein này vào trong tế bào này. Một bộ caxet biểu hiện được thiết kế theo cách cho phép protein này được dịch mã và duy trì trong dịch bào tương. Một bộ caxet biểu hiện khác được thiết kế để có peptit vận chuyển liền kề với protein độc tố để cho phép gắn đích vào bào quan của tế bào như lục lạp hoặc lạp thể. Tất cả các caxet biểu hiện đều được thiết kế để bắt đầu ở đầu 5' bằng trình tự khởi đầu, mà có thể bao gồm nhiều yếu tố khởi đầu và yếu tố tăng cường được liên kết chặt chẽ để tăng cường sự biểu hiện của gen chuyển. Trình tự khởi đầu thường được theo sau liền kề bởi một hoặc nhiều trình tự dẫn đầu 3' đến trình tự khởi đầu này. Trình tự intron được cung cấp 3' cho trình tự dẫn đầu để cải thiện sự biểu hiện của gen chuyển này. Trình tự mã hóa cho độc tố này hoặc peptit vận chuyển và trình tự mã hóa cho độc tố được đặt ở đầu 3' của trình tự khởi đầu, cấu hình dẫn đầu và intron. Trình tự được cung cấp 3' của trình tự

mã hóa để tạo thuận lợi cho việc kết thúc phiên mã và cung cấp các trình tự quan trọng cho sự polyadenyl hóa của bản sao tạo thành. Tất cả các yếu tố được mô tả trên đây được sắp xếp liền kề, thường là với trình tự bổ sung được cung cấp để tạo cấu trúc của caxet biểu hiện như vị trí endonucleaza giới hạn hoặc vị trí tách dòng độc lập với sự nối.

Đối với cây ngô, bộ caxet biểu hiện được thiết kế để biểu hiện trong dịch bào tương BCW 001 chứa trình tự khởi đầu *Mexicana* ubiquitin 1, BCW 002 chứa trình tự khởi đầu *Oryza sativa* actin 15 hoặc trình tự khởi đầu 35S, và BCW 003 chứa trình tự khởi đầu 35S.

Một bộ caxet biểu hiện khác được thiết kế để biểu hiện gắn đích ở cây ngô của protein độc tố côn trùng BCW 002 và BCW 003 trong đó trình tự mã hóa peptit lục lạp (ví dụ, CTP2) được dung hợp trong khung ở đầu 5' của đoạn ADN mã hóa protein độc tố BCW, chứa trình tự khởi đầu *Oryza sativa* actin 15 hoặc trình tự khởi đầu 35S, và trình tự chứa trình tự khởi đầu 35S.

Các caxet biểu hiện cây mía chứa trình tự khởi đầu CaMV 35S hoặc trình tự khởi đầu PCISV. FLt được liên kết hoạt động với trình tự khởi đầu 35S được tạo cấu trúc trong các vector biến nạp thực vật. Trong một số trường hợp, caxet biểu hiện Cry2Ab gắn đích lục lạp được bao gồm.

Các thực vật biểu hiện các protein theo sáng chế được kiểm tra trên các ấu trùng BCW, WBC, CEW, SWC và SCB tuổi thứ ba. Caxet biểu hiện trong dịch bào tương cho BCW 001 và caxet biểu hiện gắn đích lục lạp thể và dịch bào tương cho BCW 002 và BCW 003 được tách dòng và được sử dụng để tạo ra các dòng ngô chuyển gen biểu hiện các protein này. Các tế bào đã biến nạp được cảm ứng để tạo ra thực vật bằng các quy trình đã biết trong lĩnh vực này. Các thử nghiệm sinh học sử dụng các đĩa lá cây được thực hiện tương tự như được mô tả trong Patent Mỹ số 8,344,207. Mức độ hủy hoại lá cây (leaf damage rating: LDR) được tính điểm đánh giá dựa vào tỷ lệ phần trăm đĩa lá cây bị ăn bởi côn trùng theo thang từ 0 (0% bị ăn) đến 11 (nhiều hơn 50%) bị ăn. Điểm đánh giá tăng theo bậc với số gia là 5%. Dòng ngô đồng hợp tử được sử dụng để thu được mô làm đối chứng âm và các kết quả được đánh giá. Sự biểu hiện gắn đích lục lạp thể và sự biểu hiện trong dịch bào tương của protein độc tố côn trùng BCW 002 và BCW 003 đều làm giảm mức độ ăn gây hại so với đối chứng không được biến nạp. Các kết quả của các thử nghiệm đĩa lá

cây trên các côn trùng này phù hợp với số liệu của thử nghiệm sinh học từ các ví dụ được trình bày trên đây. Một cấu trúc chứa casset BCW 003 để biểu hiện trong dịch bào tương tạo ra 34 dòng biến nạp, 25 trong số này biểu hiện sự phòng trừ hoàn toàn sâu mọt nở BCW. Các cây ngô biểu hiện BCW 001 và BCW 003 trong dịch bào tương cũng được thử nghiệm trên CEW, SWC và FAW. Các cây biểu hiện BCW 003 thể hiện tác dụng phòng trừ 100% CEW và SWC và các giá trị LDR nằm trong khoảng từ 1 đến 2. Ba dòng biến nạp biểu hiện BCW 001 tạo ra các cây có hiệu quả kháng CEW và SWC và các giá trị LDR nằm trong khoảng từ 1 đến 3. Điều này phù hợp với số liệu thử nghiệm sinh học chế độ ăn được trình bày trong các ví dụ trước.

Cây mía chuyển gen biểu hiện BCW 003 được tạo ra và được thử nghiệm trên SCB trong các thử nghiệm sinh học. Mỗi thử nghiệm sinh học bao gồm các đĩa lá cây từ mía kiểu dại làm đối chứng âm và đối chứng dương biểu hiện mức Cry2Ab cao. Tỷ lệ chết của côn trùng và mức gây hại cho lá được đo bốn (4) ngày sau khi tác nhân gây hại xâm nhập vào. Các đĩa lá cây từ một số dòng mía chuyển gen biểu hiện BCW 003 được phát hiện ra là phòng diệt côn trùng đục thân cây mía ở thực vật (*in planta*) và thể hiện điểm đánh giá mức hủy hoại lá là dưới 2, tương tự đối chứng dương và tỷ lệ chết trung bình của côn trùng là 90-100%.

Các dòng chuyển gen biểu hiện BCW 003 và Cry2Ab thể hiện tác dụng phòng trừ SCB tốt hơn so với các dòng chỉ biểu hiện BCW 003.

YÊU CẦU BẢO HỘ

1. Cấu trúc polynucleotit chứa trình tự nucleotit mã hóa:

(a) protein diệt côn trùng có trình tự axit amin chứa SEQ ID:NO 6 hoặc SEQ ID NO:12; hoặc

(b) polypeptit diệt côn trùng thể hiện tỷ lệ đồng nhất ít nhất là 98% với trình tự axit amin của (a) trong đó trình tự nucleotit này được liên kết hoạt động với trình tự khởi đầu khác loại.

2. Protein gây độc cho loài sâu xám *Lepidopteran* chứa trình tự axit amin như được nêu trong SEQ ID NO:6.

3. Cấu trúc polynucleotit theo điểm 1, trong đó protein diệt côn trùng này thể hiện hoạt tính kháng các loài *Lepidopteran* được chọn từ nhóm bao gồm *Spodoptera frugiperda*, *Spodoptera exigua*, *Spodoptera litura*, *Mamestra configurata*, *Striacosta albicosta*, *Trichoplusia ni*, *Pseudoplusia includens*, *Anticarsia gemmatalis*, *Hypena scabra*, *Heliothis virescens*, *Agrotis subterranea*, *Pseudaletia unipuncta*, *Agrotis ipsilon*, *Agrotis orthogonia*, *Ostrinia nubilalis*, *Amyelois transitella*, *Crambus caliginosellus*, *Herpetogramma licarsisalis*, *Homoeosoma electellum*, *Elasmopalpus lignosellus*, *Cydia pomonella*, *Endopiza viteana*, *Grapholita molesta*, *Suleima helianthana*, *Plutella xylostella*, *Pectinophora gossypiella*, *Lymantria dispar*, *Alabama argillacea*, *Archips argyrospila*, *Archips rosana*, *Chilo suppressalis*, *Cnaphalocrocis medinalis*, *Crambus teterrellus*, *Diatraea grandiosella*, *Diatraea saccharalis*, *Earias insulana*, *Egrias vittella*, *Helicoverpa armigera*, *Helicoverpa zea*, *Lobesia botrana*, *Pectinophora gossypiella*, *Phyllocnistis citrella*, *Pieris brassicae*, *Pieris rapae*, *Plutella xylostella*, và *Tuta absoluta*.

4. Protein theo điểm 2, trong đó protein này còn có hoạt tính sinh học kháng các loài *Lepidopteran* được chọn từ nhóm bao gồm *Spodoptera frugiperda*, *Spodoptera exigua*, *Spodoptera litura*, *Mamestra configurata*, *Striacosta albicosta*, *Trichoplusia ni*, *Pseudoplusia includens*, *Anticarsia gemmatalis*, *Hypena scabra*, *Heliothis virescens*, *Agrotis subterranea*, *Pseudaletia unipuncta*, *Agrotis ipsilon*, *Agrotis orthogonia*, *Ostrinia nubilalis*, *Amyelois transitella*, *Crambus caliginosellus*, *Herpetogramma licarsisalis*, *Homoeosoma electellum*, *Elasmopalpus lignosellus*, *Cydia pomonella*, *Endopiza viteana*,

Grapholita molesta, *Suleima helianthana*, *Plutella xylostella*, *Pectinophora gossypiella*, *Lymantria dispar*, *Alabama argillacea*, *Archips argyrospila*, *Archips rosana*, *Chilo suppressalis*, *Cnaphalocrocis medinalis*, *Crambus teterrellus*, *Diatraea grandiosella*, *Diatraea saccharalis*, *Earias insulana*, *Egrias vittella*, *Helicoverpa armigera*, *Helicoverpa zea*, *Lobesia botrana*, *Pectinophora gossypiella*, *Phyllocnistis citrella*, *Pieris brassicae*, *Pieris rapae*, *Plutella xylostella*, và *Tuta absoluta*.

5. Vectơ chứa cấu trúc polynucleotit theo điểm 1.

6. Tế bào chủ chứa cấu trúc polynucleotit theo điểm 1, trong đó tế bào chủ này được chọn từ nhóm bao gồm tế bào vi khuẩn, tế bào nấm men, và tế bào thực vật.

7. Tế bào chủ theo điểm 6, trong đó tế bào thực vật được chọn từ nhóm bao gồm tế bào cây cỏ linh lăng, chuối, lúa mạch, đậu, bông cải xanh, bắp cải, cải, cà rốt, sắn, thầu dầu, bông cải trắng, cần tây, đậu gà, cải thảo, cam, dừa, cà phê, ngô, cỏ ba lá, bông, bầu bí, dưa chuột, gỗ linh sam, cà tím, bạch đàn, cây lanh, tỏi, nho, cây hublông, tỏi tây, rau diếp, thông trầm hương, kê, quả dưa, quả hạch, yến mạch, oliu, hành tây, cây cảnh, cò, cỏ, đậu hà lan, lạc, hồ tiêu, đậu triều, thông, khoai tây, bạch dương, bí đỏ, thông Radiata, củ cải, cải dầu, lúa, cây thân rễ, lúa mạch đen, hồng hoa, cây bụi, cây cao lương, thông phương Nam, đậu tương, rau chân vịt, cây bí, dâu tây, củ cải đường, mía, hướng dương, ngô ngọt, cây gỗ gôm, khoai lang, cỏ kê Mỹ, chè, thuốc lá, cà chua, tiểu hắc mạch, cỏ, dưa hấu, và lúa mì.

8. Cây chứa cấu trúc polynucleotit theo điểm 1.

9. Hạt được tạo ra từ cây theo điểm 8, trong đó hạt này chứa lượng phát hiện được của cấu trúc polynucleotit này.

10. Cây theo điểm 8, trong đó hạt, phần hoa, cây thế hệ sau, tế bào thực vật, mô thực vật và sản phẩm hàng hóa được tạo ra từ cây này chứa lượng phát hiện được của cấu trúc polynucleotit này.

11. Mẫu sinh học chứa lượng phát hiện được của cấu trúc polynucleotit theo điểm 1.

12. Chế phẩm cung cấp lượng có tác dụng diệt côn trùng của protein theo điểm 2 để phòng trừ loài gây hại *Lepidopteran*, và một chất khác với protein này và cũng gây độc cho cùng loài *Lepidopteran*, trong đó chất này được chọn từ nhóm bao gồm: polypeptit có trình tự axit amin khác với protein này, phân tử ARN, và hóa chất.

13. Phương pháp sản xuất hạt chứa trình tự nucleotit theo điểm 1, trong đó phương pháp này bao gồm:

- a. gieo một hoặc nhiều hạt chứa trình tự nucleotit này;
- b. trồng cây từ hạt này; và
- c. thu hoạch vụ hạt từ cây này, trong đó vụ hạt này chứa trình tự nucleotit này.

14. Protein gây độc cho loài *Lepidopteran* chứa trong liên kết hoạt động:

- a. đoạn peptit thứ nhất chứa trình tự axit amin miền I của Cry1A;
- b. đoạn peptit thứ hai chứa các gốc từ 258 đến 606 của SEQ ID NO:6;

trong đó protein gây độc cho *Lepidopteran* này thể hiện hoạt tính sinh học gây độc khi được thử nghiệm trên vật gây hại *Lepidopteran* được chọn từ nhóm bao gồm sâu xám (*A. ipsilon*), sâu đục bắp trên cây ngô (*H. zea*), sâu ngài đêm trên đậu miền Tây (*S. albicosta*), sâu đục bắp châu Âu (*O. nubilalis*), sâu đục thân bắp Tây nam (*D. grandiosella*), sâu đo bắp cải (*T. ni*), sâu đo đậu tương (*P. includens*), sâu xanh mùa thu (*S. frugiperda*), và sâu đục thân cây mía (*D. saccharalis*).

Loài côn trùng	BCW 001	BCW 002	BCW 003
A. ipsilon	+	+	+
S. albicosta	+	+	+
H. zea	+	+	+
O. nubilalis	+	+	+
D. saccharalis	ND	ND	+
D. grandiosella	+	ND	+
T. ni	+	ND	+
P. includens	+	ND	+
S. frugiperda	-	-	-

FIG. 1

SEQ ID NO:2 MEEN-

NQNQCVPYNCLNNPAIEILEGDRISVGNTPIDISLSLVELLISEFVPGGGIITGL 59

SEQ ID NO:4

MDNNPNINECIPYNCLSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGFV
LGL 60

SEQ ID NO:6

MDNNPNINECIPYNCLSNPEVEVLGGERIETGYTPIDISLSLTQFLLSEFVPGAGFV
LGL 60

SEQ ID NO:2

LNIVWGFVGPSQWDAFLAQVEQLINQRRISEAVRNTAIQELEGMARVYRTYATAF
AEWERD 119

SEQ ID NO:4

VDIIWGIFGPSQWDAFLVQIEQLINQRRIEFARNQAISRLEGLSNLYQIYAESFREW
EAD 120

SEQ ID NO:6

VDIIWGIFGPSQWDAFLVQIEQLINQRRIEFARNQAISRLEGLSNLYQIYAESFREW
EAD 120

SEQ ID NO:2

PNNTDLREAVRTQFTATETYISGRISVLKIQNFEVQLLSVFAQAANLHLSLLRDV
FFGQ 179

SEQ ID NO:4

PTNPALREEMRIQFNDMNSALTTAIPLLAVQNYQVPLLSVYVQAANLHLSVLRD
VSVFGQ 180

SEQ ID NO:6

PTNPALREEMRIQFNDMNSALTTAIPLLAVQNYQVPLLSVYVQAANLHLSVLRD
VSVFGQ 180

SEQ ID NO:2

RWGFSTTTVNYYNDLTEEISTYTDYAVRWYNTGLERVWGPDSRDWVRYNQF
RRELTLTV 239

SEQ ID NO:4

RWGFDAATINSRYNDLTRLIGNYTDYAVRWYNTGLERVWGPDSRDWVRYNQF
RRELTLTV 240

SEQ ID NO:6

RWGFDAATINSRYNDLTRLIGNYTDYAVRWYNTGLERVWGPDSRDWVRYNQF
RRELTLTV 240

SEQ ID NO:2

LDIVALFPNYDSRRYPPIRTVSQLTREIYTNPVLENFDGSFRGMAQRIEQNIRQPHL
MDIL 299

SEQ ID NO:4

LDIVALFPNYDSRRYPPIRVSQLTREIYTNPVLENFDGSFRGMAQRIEQNIRQPHL
MDIL 300

SEQ ID NO:6

LDIVALFPNYDSRRYPPIRTVSQLTREIYTNPVLENFDGSFRGMAQRIEQNIRQPHL
MDIL 300

*

SEQ ID NO:2

NSITIYTDVHRGFNYWSGHQITASPVGFSGPEFAFPLFGNAGNAAPPVLVSLTGLG
IFRT 359

SEQ ID NO:4

NSITIYTDVHRGFNYWSGHQITASPVGFSGPEFAFPLFGNAGNAAPPVLVSLTGLG
IFRT 360

SEQ ID NO:6

NSITIYTDVHRGFNYWSGHQITASPVGFSGPEFAFPLFGNAGNAAPPVLVSLTGLG
IFRT 360

SEQ ID NO:2

LSSPLYRRIILGSGPNNQELFVLDGTEFSFASLTTNLPSTIYRQRGTVDSL DVIPPQD
NS 419

SEQ ID NO:4

LSSPLYRRIILGSGPNNQELFVLDGTEFSFASLTTNLPSTIYRQRGTVDSL DVIPPQD
NS 420

SEQ ID NO:6

LSSPLYRRIILGSGPNNQELFVLDGTEFSFASLTTNLPSTIYRQRGTVDSL DVIPPQD
NS 420

SEQ ID NO:2

VPPRAGFSHRLSHVTMLSQAAGAVYTLRAPTFWQHRSATTTNIIAADSITQIPAV
KGRS 479

SEQ ID NO:4

VPPRAGFSHRLSHVTMLSQAAGAVYTLRAPTFWQHRSATTTNIIAADSITQIPAV
KGRS 480

SEQ ID NO:6

VPPRAGFSHRLSHVTMLSQAAGAVYTLRAPTFWQHRSATTTNIIAADSITQIPAV
KGRS 480

SEQ ID NO:2

IINNGTVISGPGFTGGDLVRLYNADFNINNRAYLEVPIFFQSPSTNYRVRVRYAST
SSLP 539

SEQ ID NO:4

IINNGTVISGPGFTGGDLVRLYNADFNINNRAYLEVPIFFQSPSTNYRVRVRYAST
SSLP 540

SEQ ID NO:6

IINNGTVISGPGFTGGDLVRLYNADFNINNRAYLEVPIFFQSPSTNYRVRVRYAST
SSLP 540

SEQ ID NO:2

VDVVFVGNISHPTTFPATARSLDNLQSNDFGYIDIAGTFLPSLGPSIGIRPMLSTINLI
VD 599

SEQ ID NO:4

VDVVFVGNISHPTTFPATARSLDNLQSNDFGYIDIAGTFLPSLGPSIGIRPMLSTINLI
VD 600

SEQ ID NO:6

VDVVFVGNISHPTTFPATARSLDNLQSNDFGYIDIAGTFLPSLGPSIGIRPMLSTINLI
VD 600

SEQ ID NO:2

RFEFIPVTATFEAESDLERAQKAVNALFTSTNQLGIKTDVTDYHIDQVSNLVECLS
DEFY 659

SEQ ID NO:4

RFEFIPVTATFEAESDLERAQKAVNALFTSTNQLGIKTDVTDYHIDQVSNLVECLS
DEFY 660

SEQ ID NO:6

RFEFIPVTATFEAESDLERAQKAVNALFTSTNQLGIKTDVTDYHIDQVSNLVECLS
DEFY 660

FIG. 2

SEQ ID NO:2

LDEKRELSEKVKHAKRLSDERNLLQDPNFRGINRQPDRGWRGSTDITIQQGGDDV
FKENYV 719

SEQ ID NO:4

LDEKRELSEKVKHAKRLSDERNLLQDPNFRGINRQPDRGWRGSTDITIQQGGDDV
FKENYV 720

SEQ ID NO:6

LDEKRELSEKVKHAKRLSDERNLLQDPNFRGINRQPDRGWRGSTDITIQQGGDDV
FKENYV 720

SEQ ID NO:2

TLPGTFDECYPTYLYQKIDESKLKAYTRYQLRGYIEDSQDLEIYLIRYNAKHETV
NVPGT 779

SEQ ID NO:4

TLPGTFDECYPTYLYQKIDESKLKAFTRYQLRGYIEDSQDLEIYLIRYNAKHETVN
VPGT 780

SEQ ID NO:6

TLPGTFDECYPTYLYQKIDESKLKAFTRYQLRGYIEDSQDLEIYLIRYNAKHETVN
VPGT 780

*

SEQ ID NO:2

GSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLHCSCRDGEKCAHHSHHFSLDIDV
GCTDLN 839

SEQ ID NO:4

GSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLDCSCRDGEKCAHHSHHFSLDIDV
GCTDLN 840

SEQ ID NO:6

GSLWPLSAQSPIGKCGEPNRCAPHLEWNPDLDCSCRDGEKCAHHSHHFSLDIDV
GCTDLN 840

*

SEQ ID NO:2

EDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKL
QLETNIV 899

SEQ ID NO:4

EDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKL
EWETNIV 900

SEQ ID NO:6

EDLGVWVIFKIKTQDGHARLGNLEFLEEKPLVGEALARVKRAEKKWRDKREKL
EWETNIV 900

**

SEQ ID NO:2

YKEAKESVDALFVNSQYDQLQADTNAMIHAADKRVHRIREAYLPESVIPGVN
ADISEE 959

SEQ ID NO:4
 YKEAKESVDALFVNSQYDQLQADTNAMIHAADKRVHSIREAYLPELSVIPGVN
 AAIFEE 960
 SEQ ID NO:6
 YKEAKESVDALFVNSQYDQLQADTNAMIHAADKRVHSIREAYLPELSVIPGVN
 AAIFEE 960

* * *

SEQ ID NO:2
 LEGRIFTAFSLYDARNVIKNGDFNNGLLCWNVKGHVDVEEQNNHRSVLVVPEW
 EAEVSQE 1019
 SEQ ID NO:4
 LEGRIFTAFSLYDARNVIKNGDFNNGLLCWNVKGHVDVEEQNNQRSVLVVPEW
 EAEVSQE 1020
 SEQ ID NO:6
 LEGRIFTAFSLYDARNVIKNGDFNNGLLCWNVKGHVDVEEQNNQRSVLVVPEW
 EAEVSQE 1020

* *

SEQ ID NO:2
 VRVCPGRGYILRV TAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEVYPNNTVTC
 NDYTA 1079
 SEQ ID NO:4
 VRVCPGRGYILRV TAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEIYPNNTVTC
 NDYTV 1080
 SEQ ID NO:6
 VRVCPGRGYILRV TAYKEGYGEGCVTIHEIENNTDELKFSNCVEEEIYPNNTVTC
 NDYTV 1080

* *

SEQ ID NO:2
 NQEEYEGTYTSRNRGYDEAYESNSSVPAEYASVYEEKVYTDGRRGNPCESNRG
 YGDYTPL 1139
 SEQ ID NO:4 NQEEYGGAYTSRNRGYNEAP----
 SVPADYASVYEEKSYTDGRRENPCFNRGYRDYTPL 1136
 SEQ ID NO:6 NQEEYGGAYTSRNRGYNEAP----
 SVPADYASVYEEKSYTDGRRENPCFNRGYRDYTPL 1136

* * * * * * * * * * * *

SEQ ID NO:2 PAGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
 1180
 SEQ ID NO:4 PVGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
 1177
 SEQ ID NO:6 PVGYVTKELEYFPETDKVWIEIGETEGTFIVDSVELLLMEE
 1177

*

FIG. 2 (tiếp)

DANH MỤC TRÌNH TỰ

<110> Monsanto Technology LLC

<120> Cấu trúc polynucleotit, phương pháp sản xuất hạt chứa cấu trúc polynucleotit này, protein gây độc cho loài *Lepidopteran* và chế phẩm chứa protein này để phòng trừ loài gây hại *Lepidopteran*

<130> MONS:434WO

<150> US 62/445,313

<151> 12-01-2017

<160> 12

<170> PatentIn phiên bản 3.5

<210> 1

<211> 3543

<212> ADN

<213> *Bacillus thuringiensis*

<220>

<221> CDS

<222> (1)..(3540)

<400> 1

atg gag gaa aat aat cag aat caa tgc gtc cct tat aat tgt ttg aat
48

Met Glu Glu Asn Asn Gln Asn Gln Cys Val Pro Tyr Asn Cys Leu Asn
1 5 10 15

aat cct gca atc gaa ata tta gaa gga gac aga ata tca gtt ggt aac
96

Asn Pro Ala Ile Glu Ile Leu Glu Gly Asp Arg Ile Ser Val Gly Asn
20 25 30

act cca atc gat att tct cta tca ctt gtg gaa ctt ctt att agt gaa
144

Thr Pro Ile Asp Ile Ser Leu Ser Leu Val Glu Leu Leu Ile Ser Glu
35 40 45

ttt gtc cca ggc ggt gga ata ata aca gga ttg ttg aac ata gta tgg
192

Phe Val Pro Gly Gly Gly Ile Ile Thr Gly Leu Leu Asn Ile Val Trp
50 55 60

gga ttt gta ggg cct tcc caa tgg gac gca ttt ctt gct caa gtg gaa
240

Gly Phe Val Gly Pro Ser Gln Trp Asp Ala Phe Leu Ala Gln Val Glu

```

65              70              75              80
cag tta att aac caa agg ata tca gaa gct gta aga aat aca gca att
288
Gln Leu Ile Asn Gln Arg Ile Ser Glu Ala Val Arg Asn Thr Ala Ile
      85              90              95

cag gaa tta gag gga atg gcg cgg gtt tat aga acc tat gct act gct
336
Gln Glu Leu Glu Gly Met Ala Arg Val Tyr Arg Thr Tyr Ala Thr Ala
      100             105             110

ttt gct gag tgg gaa aga gat cct aat aac aca gat cta aga gaa gca
384
Phe Ala Glu Trp Glu Arg Asp Pro Asn Asn Thr Asp Leu Arg Glu Ala
      115             120             125

gta cgg aca cag ttt aca gca act gag act tat atc agt gga aga ata
432
Val Arg Thr Gln Phe Thr Ala Thr Glu Thr Tyr Ile Ser Gly Arg Ile
      130             135             140

tct gtt tta aaa att caa aat ttt gaa gtg cag ctg tta tcg gtg ttt
480
Ser Val Leu Lys Ile Gln Asn Phe Glu Val Gln Leu Leu Ser Val Phe
      145             150             155             160

gcc caa gct gcc aat tta cat tta tct tta tta aga gac gtt gtg ttt
528
Ala Gln Ala Ala Asn Leu His Leu Ser Leu Leu Arg Asp Val Val Phe
      165             170             175

ttt ggg caa aga tgg ggg ttt tca acg aca acc gta aat aat tac tac
576
Phe Gly Gln Arg Trp Gly Phe Ser Thr Thr Thr Val Asn Asn Tyr Tyr
      180             185             190

aat gat tta aca gaa gag att agt acc tat aca gat tat gca gta cgc
624
Asn Asp Leu Thr Glu Glu Ile Ser Thr Tyr Thr Asp Tyr Ala Val Arg
      195             200             205

tgg tac aat acg gga tta gag cgt gta tgg gga ccg gat tct aga gat
672
Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg Asp
      210             215             220

tgg gta agg tat aat caa ttt aga aga gag cta aca ctt act gta tta
720
Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val Leu
      225             230             235             240

```

gat atc gtt gct cta ttc cca aat tat gat agt cga agg tat cca att
768
Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro Ile
245 250 255

cga aca gtt tcc caa tta aca aga gaa att tat acg aac cca gta tta
816
Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val Leu
260 265 270

gaa aat ttt gat ggt agt ttt cgt gga atg gct cag aga ata gaa cag
864
Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu Gln
275 280 285

aat att agg caa cca cat ctt atg gat atc ctt aat agt ata acc att
912
Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr Ile
290 295 300

tat act gat gtg cat aga ggc ttt aat tat tgg tca ggg cat caa ata
960
Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln Ile
305 310 315 320

aca gct tct cct gta ggg ttt tca gga cca gaa ttc gca ttc cct tta
1008
Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro Leu
325 330 335

ttt ggg aat gcg gga aat gca gct cca ccc gta ctt gtc tca tta act
1056
Phe Gly Asn Ala Gly Asn Ala Ala Pro Pro Val Leu Val Ser Leu Thr
340 345 350

ggg ttg ggg att ttt aga aca tta tct tca cct tta tat aga aga att
1104
Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg Ile
355 360 365

ata ctt ggt tca ggc cca aat aat cag gaa ctg ttt gtc ctt gat gga
1152
Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp Gly
370 375 380

acg gag ttt tct ttt gcc tcc cta acg acc aac ttg cct tcc act ata
1200
Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr Ile
385 390 395 400

tat aga caa agg ggt aca gtc gat tca cta gat gta ata ccg cca cag
1248

Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro Gln
 405 410 415
 gat aat agt gta cca cct cgt gcg gga ttt agc cat cga ttg agt cat
 1296
 Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser His
 420 425 430
 gtt aca atg ctg agc caa gca gct gga gca gtt tac acc ttg aga gct
 1344
 Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg Ala
 435 440 445
 cca acg ttt tct tgg cag cat cgc agt gct acg aca act aat ata att
 1392
 Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile Ile
 450 455 460
 gca gcg gat agt att act caa att cct gct gtt aaa gga cgt tct att
 1440
 Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser Ile
 465 470 475 480
 att aat aat ggc acg gta att tca gga cca ggg ttt acc gga ggc gat
 1488
 Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly Asp
 485 490 495
 ttg gtt aga tta tac aat gct gat ttt aat att aat aat aga gca tac
 1536
 Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala Tyr
 500 505 510
 ctt gaa gtt ccg ata ttc ttc caa tca ccc tct aca aat tat cgt gtt
 1584
 Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg Val
 515 520 525
 cgt gtt cgt tat gct tct aca tct tca ctc cct gta gat gta gtt ttc
 1632
 Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val Phe
 530 535 540
 gga aat att agt cat cct act aca ttc cca gcc act gcc aga tca tta
 1680
 Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser Leu
 545 550 555 560
 gat aat cta caa tcc aat gat ttt gga tat att gat att gct gga act
 1728
 Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly Thr
 565 570 575

ttc tta cct tca cta ggg cct agt ata ggt atc aga cct atg tta tct
1776

Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu Ser
580 585 590

act att aat ttg ata gta gat aga ttt gaa ttt att cca gta act gca
1824

Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr Ala
595 600 605

acc ttt gaa gca gaa tcg gat tta gaa aga gca caa aag gcg gtg aat
1872

Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn
610 615 620

gcg ctg ttt act tct aca aac caa cta ggg ata aaa aca gat gtg acg
1920

Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val Thr
625 630 635 640

gat tat cat att gat caa gtg tcc aat tta gtg gag tgt tta tcg gat
1968

Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser Asp
645 650 655

gaa ttt tat ctg gat gaa aag cga gaa ttg tcc gag aaa gtc aaa cat
2016

Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys His
660 665 670

gcg aag cga ctc agt gat gag cga aat tta ctt caa gat cca aac ttc
2064

Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe
675 680 685

agg ggc atc aat aga caa cca gat cgt ggc tgg aga gga agt acg gat
2112

Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr Asp
690 695 700

att acc atc caa gga gga gat gac gta ttc aaa gag aat tac gtc aca
2160

Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr
705 710 715 720

cta cca ggt acc ttt gat gag tgc tat cca acg tat tta tat caa aaa
2208

Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Lys
725 730 735

ata gat gag tcg aaa tta aaa gcc tat acc cgt tat caa tta aga ggg
2256

Ile Asp Glu Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Gln Leu Arg Gly
740 745 750

tat atc gag gat agt caa gac tta gaa atc tat tta att cgc tac aat
2304

Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn
755 760 765

gca aaa cat gaa aca gta aat gtg cca ggt acg ggt tcc tta tgg ccg
2352

Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro
770 775 780

ctt tca gcc caa agt cca atc gga aag tgt gga gag ccg aat cga tgc
2400

Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys
785 790 795 800

gcg cca cac ctt gaa tgg aat cct gat tta cac tgt tcc tgc aga gac
2448

Ala Pro His Leu Glu Trp Asn Pro Asp Leu His Cys Ser Cys Arg Asp
805 810 815

ggg gaa aaa tgt gct cat cat tct cat cat ttc tcc ttg gac att gat
2496

Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile Asp
820 825 830

gtt gga tgt aca gac tta aat gag gat tta ggt gta tgg gtg ata ttc
2544

Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe
835 840 845

aag att aag acg caa gat ggc cat gca aga cta gga aat cta gag ttt
2592

Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe
850 855 860

ctc gaa gag aaa cca tta gta ggg gaa gca cta gct cgt gtg aaa aga
2640

Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys Arg
865 870 875 880

gcg gag aaa aaa tgg aga gac aaa cgc gaa aaa tta caa ttg gaa aca
2688

Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Gln Leu Glu Thr
885 890 895

aat atc gtt tat aaa gag gca aaa gaa tct gta gat gct tta ttt gta
2736

Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val
 900 905 910
 aac tct caa tat gat caa tta caa gcg gat acg aac atc gcg atg att
 2784
 Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met Ile
 915 920 925
 cat gcg gca gat aaa cgt gtt cat aga atc cga gaa gcg tac ctt cca
 2832
 His Ala Ala Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro
 930 935 940
 gag tta tct gtg att ccg ggt gta aat gca gac att tcc gaa gaa tta
 2880
 Glu Leu Ser Val Ile Pro Gly Val Asn Ala Asp Ile Ser Glu Glu Leu
 945 950 955 960
 gaa ggg cgt att ttc act gca ttc tct cta tat gat gcg aga aat gtc
 2928
 Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val
 965 970 975
 att aaa aat ggc gat ttc aat aat ggc tta tta tgc tgg aac gtg aaa
 2976
 Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys
 980 985 990
 ggg cat gta gat gta gaa gaa caa aat aac cac cgt tcg gtc ctt gtt
 3024
 Gly His Val Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val
 995 1000 1005
 gtt ccg gaa tgg gaa gca gaa gtg tca caa gag gtt cgt gtc tgt
 3069
 Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
 1010 1015 1020
 ccg ggg cgt ggc tat atc ctt cgt gtc aca gcg tac aag gag gga
 3114
 Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly
 1025 1030 1035
 tat gga gaa ggt tgc gta acc att cat gag atc gag aac aat aca
 3159
 Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr
 1040 1045 1050
 gac gaa ctg aag ttt agc aac tgt gta gaa gag gaa gtc tat cca
 3204
 Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Val Tyr Pro
 1055 1060 1065

aac aac acg gta acg tgt aat gat tat act gca aat caa gaa gaa
3249

Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Ala Asn Gln Glu Glu
1070 1075 1080

tat gag ggt acg tac act tct cgt aat cga gga tat gac gaa gcc
3294

Tyr Glu Gly Thr Tyr Thr Ser Arg Asn Arg Gly Tyr Asp Glu Ala
1085 1090 1095

tat gaa agc aat tct tct gta cca gct gag tat gcg tca gtc tat
3339

Tyr Glu Ser Asn Ser Ser Val Pro Ala Glu Tyr Ala Ser Val Tyr
1100 1105 1110

gaa gaa aaa gtg tat aca gat gga cga aga ggg aat cct tgt gaa
3384

Glu Glu Lys Val Tyr Thr Asp Gly Arg Arg Gly Asn Pro Cys Glu
1115 1120 1125

tct aac aga gga tat ggg gat tac aca cca cta cca gct ggc tat
3429

Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu Pro Ala Gly Tyr
1130 1135 1140

gtg aca aaa gaa tta gag tac ttc cca gaa acc gat aag gta tgg
3474

Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp
1145 1150 1155

att gag att gga gaa aca gaa gga aca ttc att gtg gat agt gtg
3519

Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val
1160 1165 1170

gaa tta ctc ctt atg gag gaa taa
3543

Glu Leu Leu Leu Met Glu Glu
1175 1180

<210> 2

<211> 1180

<212> PRT

<213> Bacillus thuringiensis

<400> 2

Met Glu Glu Asn Asn Gln Asn Gln Cys Val Pro Tyr Asn Cys Leu Asn
1 5 10 15

Asn Pro Ala Ile Glu Ile Leu Glu Gly Asp Arg Ile Ser Val Gly Asn
20 25 30

Thr Pro Ile Asp Ile Ser Leu Ser Leu Val Glu Leu Leu Ile Ser Glu
35 40 45

Phe Val Pro Gly Gly Gly Ile Ile Thr Gly Leu Leu Asn Ile Val Trp
50 55 60

Gly Phe Val Gly Pro Ser Gln Trp Asp Ala Phe Leu Ala Gln Val Glu
65 70 75 80

Gln Leu Ile Asn Gln Arg Ile Ser Glu Ala Val Arg Asn Thr Ala Ile
85 90 95

Gln Glu Leu Glu Gly Met Ala Arg Val Tyr Arg Thr Tyr Ala Thr Ala
100 105 110

Phe Ala Glu Trp Glu Arg Asp Pro Asn Asn Thr Asp Leu Arg Glu Ala
115 120 125

Val	Arg	Thr	Gln	Phe	Thr	Ala	Thr	Glu	Thr	Tyr	Ile	Ser	Gly	Arg	Ile
130						135					140				

Ser Val Leu Lys Ile Gln Asn Phe Glu Val Gln Leu Leu Ser Val Phe
145 150 155 160

Ala Gln Ala Ala Asn Leu His Leu Ser Leu Leu Arg Asp Val Val Phe
165 170 175

Phe Gly Gln Arg Trp Gly Phe Ser Thr Thr Thr Val Asn Asn Tyr Tyr
180 185 190

Asn Asp Leu Thr Glu Glu Ile Ser Thr Tyr Thr Asp Tyr Ala Val Arg
195 200 205

Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg Asp
210 215 220

Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val Leu
 225 230 235 240

Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro Ile
 245 250 255

Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val Leu
 260 265 270

Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu Gln
 275 280 285

Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr Ile
 290 295 300

Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln Ile
 305 310 315 320

Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro Leu
 325 330 335

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

Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg Ile
 355 360 365

Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp Gly
 370 375 380

Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr Ile
 385 390 395 400

Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro Gln
 405 410 415

Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser His
 420 425 430

Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg Ala
 435 440 445

Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile Ile
 450 455 460

Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser Ile
 465 470 475 480

Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly Asp
 485 490 495

Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala Tyr
 500 505 510

Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg Val
 515 520 525

Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val Phe
 530 535 540

Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser Leu
 545 550 555 560

Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly Thr
 565 570 575

Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu Ser
 580 585 590

Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr Ala
 595 600 605

Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn
 610 615 620

Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val Thr
 625 630 635 640

Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser Asp
 645 650 655

Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys His
 660 665 670

Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe
 675 680 685

Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr Asp
 690 695 700

Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr
 705 710 715 720

Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Lys
 725 730 735

Ile Asp Glu Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Gln Leu Arg Gly
 740 745 750

Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn
 755 760 765

Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro
 770 775 780

Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys
 785 790 795 800

Ala Pro His Leu Glu Trp Asn Pro Asp Leu His Cys Ser Cys Arg Asp
 805 810 815

Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile Asp
 820 825 830

Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe
 835 840 845

Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe
850 855 860

Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys Arg
865 870 875 880

Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Gln Leu Glu Thr
885 890 895

Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val
900 905 910

Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met Ile
915 920 925

His Ala Ala Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro
930 935 940

Glu Leu Ser Val Ile Pro Gly Val Asn Ala Asp Ile Ser Glu Glu Leu
945 950 955 960

Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val
965 970 975

Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys
980 985 990

Gly His Val Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val
995 1000 1005

Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly
1025 1030 1035

Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr
1040 1045 1050

Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Val Tyr Pro
 1055 1060 1065

Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Ala Asn Gln Glu Glu
 1070 1075 1080

Tyr Glu Gly Thr Tyr Thr Ser Arg Asn Arg Gly Tyr Asp Glu Ala
 1085 1090 1095

Tyr Glu Ser Asn Ser Ser Val Pro Ala Glu Tyr Ala Ser Val Tyr
 1100 1105 1110

Glu Glu Lys Val Tyr Thr Asp Gly Arg Arg Gly Asn Pro Cys Glu
 1115 1120 1125

Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu Pro Ala Gly Tyr
 1130 1135 1140

Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp
 1145 1150 1155

Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val
 1160 1165 1170

Glu Leu Leu Leu Met Glu Glu
 1175 1180

<210> 3
 <211> 3534
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> ADN tổng hợp hoàn toàn mã hóa BCW 002

<220>
 <221> CDS
 <222> (1)..(3531)

<400> 3

atg gat aac aat ccg aac atc aat gaa tgc att cct tat aat tgt tta
 48
 Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15
 agt aac cct gaa gta gaa gta tta ggt gga gaa aga ata gaa act ggt
 96
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30
 tac acc cca atc gat att tcc ttg tcg cta acg caa ttt ctt ttg agt
 144
 Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45
 gaa ttt gtt ccc ggt gct gga ttt gtg tta gga cta gtt gat ata ata
 192
 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60
 tgg gga att ttt ggt ccc tct caa tgg gac gca ttt ctt gta caa att
 240
 Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80
 gaa cag tta att aac caa aga ata gaa gaa ttc gct agg aac caa gcc
 288
 Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95
 att tct aga tta gaa gga cta agc aat ctt tat caa att tac gca gaa
 336
 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110
 tct ttt aga gag tgg gaa gca gat cct act aat cca gca tta aga gaa
 384
 Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125
 gag atg cgt att caa ttc aat gac atg aac agt gcc ctt aca acc gct
 432
 Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140
 att cct ctt ttg gca gtt caa aat tat caa gtt cct ctt tta tca gta
 480
 Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160
 tat gtt caa gct gca aat tta cat tta tca gtt ttg aga gat gtt tca
 528

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175
 gtg ttt gga caa agg tgg gga ttt gat gcc gcg act atc aat agt cgt
 576
 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190
 tat aat gat tta act agg ctt att ggc aac tat aca gat tat gct gta
 624
 Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205
 cgc tgg tac aat acg gga tta gaa cgt gta tgg gga ccg gat tct aga
 672
 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220
 gat tgg gta agg tat aat caa ttt aga aga gaa tta aca cta act gta
 720
 Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240
 tta gat atc gtt gct ctg ttc ccg aat tat gat agt aga aga tat cca
 768
 Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255
 att cga ata gtt tcc caa tta aca aga gaa att tat acg aac cca gta
 816
 Ile Arg Ile Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270
 tta gaa aat ttt gat ggt agt ttt cgt gga atg gct cag aga ata gaa
 864
 Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285
 cag aat att agg caa cca cat ctt atg gat atc ctt aat agt ata acc
 912
 Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300
 att tat act gat gtg cat aga ggc ttt aat tat tgg tca ggg cat caa
 960
 Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320
 ata aca gct tct cct gta ggg ttt tca gga cca gaa ttc gca ttc cct
 1008
 Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335

tta ttt ggg aat gcg gga aat gca gct cca ccc gta ctt gtc tca tta
1056

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

act ggt ttg ggg att ttt aga aca tta tct tca cct tta tat aga aga
1104

Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
355 360 365

att ata ctt ggt tca ggc cca aat aat cag gaa ctg ttt gtc ctt gat
1152

Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
370 375 380

gga acg gag ttt tct ttt gcc tcc cta acg acc aac ttg cct tcc act
1200

Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
385 390 395 400

ata tat aga caa agg ggt aca gtc gat tca cta gat gta ata ccg cca
1248

Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
405 410 415

cag gat aat agt gta cca cct cgt gcg gga ttt agc cat cga ttg agt
1296

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
420 425 430

cat gtt aca atg ctg agc caa gca gct gga gca gtt tac acc ttg aga
1344

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
435 440 445

gct cca acg ttt tct tgg cag cat cgc agt gct acg aca act aat ata
1392

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
450 455 460

att gca gcg gat agt att act caa att cct gct gtt aaa gga cgt tct
1440

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
465 470 475 480

att att aat aat ggc acg gta att tca gga cca ggg ttt acc gga ggc
1488

Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
485 490 495

gat ttg gtt aga tta tac aat gct gat ttt aat att aat aat aga gca
1536
Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
500 505 510

tac ctt gaa gtt ccg ata ttc ttc caa tca ccc tct aca aat tat cgt
1584
Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
515 520 525

gtt cgt gtt cgt tat gct tct aca tct tca ctc cct gta gat gta gtt
1632
Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
530 535 540

ttc gga aat att agt cat cct act aca ttc cca gcc act gcc aga tca
1680
Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
545 550 555 560

tta gat aat cta caa tcc aat gat ttt gga tat att gat att gct gga
1728
Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
565 570 575

act ttc tta cct tca cta ggg cct agt ata ggt atc aga cct atg tta
1776
Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
580 585 590

tct act att aat ttg ata gta gat aga ttt gaa ttt att cca gta act
1824
Ser Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr
595 600 605

gca acc ttt gaa gca gaa tcg gat tta gaa aga gca caa aag gcg gtg
1872
Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

aat gcg ctg ttt act tct aca aac caa cta ggg ata aaa aca gat gtg
1920
Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
625 630 635 640

acg gat tat cat att gat caa gtg tcc aat tta gtg gag tgt tta tcg
1968
Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
645 650 655

gat gaa ttt tat ctg gat gaa aag cga gaa ttg tcc gag aaa gtc aaa
2016

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
 660 665 670
 cat gcg aag cga ctc agt gat gag cga aat tta ctt caa gat cca aac
 2064
 His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
 675 680 685
 ttc agg ggc atc aat aga caa cca gat cgt ggc tgg aga gga agt acg
 2112
 Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
 690 695 700
 gat att acc atc caa gga gga gat gac gta ttc aaa gag aat tac gtc
 2160
 Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
 705 710 715 720
 aca cta cca ggt acc ttt gat gag tgc tat cca aca tat ttg tat caa
 2208
 Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
 725 730 735
 aaa atc gat gaa tca aaa tta aaa gcc ttt acc cgt tat caa tta aga
 2256
 Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
 740 745 750
 ggg tat atc gaa gat agt caa gac tta gaa atc tat tta att cgc tac
 2304
 Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
 755 760 765
 aat gca aaa cat gaa aca gta aat gtg cca ggt acg ggt tcc tta tgg
 2352
 Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
 770 775 780
 ccg ctt tca gcc caa agt cca atc gga aag tgt gga gag ccg aat cga
 2400
 Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
 785 790 795 800
 tgc gcg cca cac ctt gaa tgg aat cct gac tta gat tgt tcg tgt agg
 2448
 Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
 805 810 815
 gat gga gaa aag tgt gcc cat cat tcg cat cat ttc tcc tta gac att
 2496
 Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
 820 825 830

gat gta gga tgt aca gac tta aat gag gac cta ggt gta tgg gtg atc
2544

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

ttt aag att aag acg caa gat ggg cac gca aga cta ggg aat cta gag
2592

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

ttt ctc gaa gag aaa cca tta gta gga gaa gca cta gct cgt gtg aaa
2640

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

aga gcg gag aaa aaa tgg aga gac aaa cgt gaa aaa ttg gaa tgg gaa
2688

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

aca aat atc gtt tat aaa gag gca aaa gaa tct gta gat gct tta ttt
2736

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

gta aac tct caa tat gat caa tta caa gcg gat acg aat att gcc atg
2784

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

att cat gcg gca gat aaa cgt gtt cat agc att cga gaa gct tat ctg
2832

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

cct gag ctg tct gtg att ccg ggt gtc aat gcg gct att ttt gaa gaa
2880

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

tta gaa ggg cgt att ttc act gca ttc tcc cta tat gat gcg aga aat
2928

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

gtc att aaa aat ggt gat ttt aat aat ggc tta tcc tgc tgg aac gtg
2976

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

aaa ggg cat gta gat gta gaa gaa caa aac aac caa cgt tcg gtc ctt
 3024
 Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005

 gtt gtt ccg gaa tgg gaa gca gaa gtg tca caa gaa gtt cgt gtc
 3069
 Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
 1010 1015 1020

 tgt ccg ggt cgt ggc tat atc ctt cgt gtc aca gcg tac aag gag
 3114
 Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
 1025 1030 1035

 gga tat gga gaa ggt tgc gta acc att cat gag atc gag aac aat
 3159
 Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
 1040 1045 1050

 aca gac gaa ctg aag ttt agc aac tgc gta gaa gag gaa atc tat
 3204
 Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
 1055 1060 1065

 cca aat aac acg gta acg tgt aat gat tat act gta aat caa gaa
 3249
 Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
 1070 1075 1080

 gaa tac gga ggt gcg tac act tct cgt aat cga gga tat aac gaa
 3294
 Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
 1085 1090 1095

 gct cct tcc gta cca gct gat tat gcg tca gtc tat gaa gaa aaa
 3339
 Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
 1100 1105 1110

 tcg tat aca gat gga cga aga gag aat cct tgt gaa ttt aac aga
 3384
 Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
 1115 1120 1125

 ggg tat agg gat tac acg cca cta cca gtt ggt tat gtg aca aaa
 3429
 Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
 1130 1135 1140

 gaa tta gaa tac ttc cca gaa acc gat aag gta tgg att gag att
 3474

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
 1145 1150 1155
 gga gaa acg gaa gga aca ttt atc gtg gac agc gtg gaa tta ctc
 3519
 Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
 1160 1165 1170
 ctt atg gag gaa tag
 3534
 Leu Met Glu Glu
 1175
 <210> 4
 <211> 1177
 <212> PRT
 <213> Trình tự nhân tạo
 <220>
 <223> Cấu trúc tổng hợp
 <400> 4
 Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30
 Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45
 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60
 Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80
 Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95
 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140

Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255

Ile Arg Ile Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285

Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335

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

Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
 355 360 365

Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
 370 375 380

Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
 385 390 395 400

Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
 405 410 415

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
 420 425 430

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
 435 440 445

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
 450 455 460

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
 465 470 475 480

Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
 485 490 495

Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
 500 505 510

Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
 515 520 525

Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
 530 535 540

Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
 545 550 555 560

Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
 565 570 575

Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
 580 585 590

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

Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
 610 615 620

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
 625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
 645 650 655

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
 660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
 675 680 685

Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
 690 695 700

Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
 705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
 725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
1010 1015 1020

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
1025 1030 1035

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
1040 1045 1050

Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
1055 1060 1065

Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
1070 1075 1080

Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
1085 1090 1095

Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
1100 1105 1110

Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
1115 1120 1125

Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
1130 1135 1140

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
 1145 1150 1155

Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
 1160 1165 1170

Leu Met Glu Glu
 1175

<210> 5
 <211> 3534
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Trình tự tổng hợp hoàn toàn mã hóa BCW 003

<220>
 <221> CDS
 <222> (1)..(3531)

<400> 5
 atg gat aac aat ccg aac atc aat gaa tgc att cct tat aat tgt tta
 48

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15

agt aac cct gaa gta gaa gta tta ggt gga gaa aga ata gaa act ggt
 96

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30

tac acc cca atc gat att tcc ttg tcg cta acg caa ttt ctt ttg agt
 144

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45

gaa ttt gtt ccc ggt gct gga ttt gtg tta gga cta gtt gat ata ata
 192

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

tgg gga att ttt ggt ccc tct caa tgg gac gca ttt ctt gta caa att
 240

Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80

gaa cag tta att aac caa aga ata gaa gaa ttc gct agg aac caa gcc
 288
 Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95

att tct aga tta gaa gga cta agc aat ctt tat caa att tac gca gaa
 336
 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110

tct ttt aga gag tgg gaa gca gat cct act aat cca gca tta aga gaa
 384
 Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125

gag atg cgt att caa ttc aat gac atg aac agt gcc ctt aca acc gct
 432
 Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140

att cct ctt ttg gca gtt caa aat tat caa gtt cct ctt tta tca gta
 480
 Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160

tat gtt caa gct gca aat tta cat tta tca gtt ttg aga gat gtt tca
 528
 Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175

gtg ttt gga caa agg tgg gga ttt gat gcc gcg act atc aat agt cgt
 576
 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190

tat aat gat tta act agg ctt att ggc aac tat aca gat tat gct gta
 624
 Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205

cgc tgg tac aat acg gga tta gaa cgt gta tgg gga ccg gat tct aga
 672
 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220

gat tgg gta agg tat aat caa ttt aga aga gaa tta aca cta act gta
 720
 Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240

tta gat atc gtt gct ctg ttc ccg aat tat gat agt aga aga tat cca
 768

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255
 att cga aca gtt tcc caa tta aca aga gaa att tat acg aac cca gta
 816
 Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270
 tta gaa aat ttt gat ggt agt ttt cgt gga atg gct cag aga ata gaa
 864
 Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285
 cag aat att agg caa cca cat ctt atg gat atc ctt aat agt ata acc
 912
 Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300
 att tat act gat gtg cat aga ggc ttt aat tat tgg tca ggg cat caa
 960
 Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320
 ata aca gct tct cct gta ggg ttt tca gga cca gaa ttc gca ttc cct
 1008
 Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335
 tta ttt ggg aat gcg gga aat gca gct cca ccc gta ctt gtc tca tta
 1056
 Leu Phe Gly Asn Ala Gly Asn Ala Ala Pro Pro Val Leu Val Ser Leu
 340 345 350
 act ggt ttg ggg att ttt aga aca tta tct tca cct tta tat aga aga
 1104
 Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
 355 360 365
 att ata ctt ggt tca ggc cca aat aat cag gaa ctg ttt gtc ctt gat
 1152
 Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
 370 375 380
 gga acg gag ttt tct ttt gcc tcc cta acg acc aac ttg cct tcc act
 1200
 Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
 385 390 395 400
 ata tat aga caa agg ggt aca gtc gat tca cta gat gta ata ccg cca
 1248
 Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
 405 410 415

cag gat aat agt gta cca cct cgt gcg gga ttt agc cat cga ttg agt
1296

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
420 425 430

cat gtt aca atg ctg agc caa gca gct gga gca gtt tac acc ttg aga
1344

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
435 440 445

gct cca acg ttt tct tgg cag cat cgc agt gct acg aca act aat ata
1392

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
450 455 460

att gca gcg gat agt att act caa att cct gct gtt aaa gga cgt tct
1440

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
465 470 475 480

att att aat aat ggc acg gta att tca gga cca ggg ttt acc gga ggc
1488

Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
485 490 495

gat ttg gtt aga tta tac aat gct gat ttt aat att aat aat aga gca
1536

Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
500 505 510

tac ctt gaa gtt ccg ata ttc ttc caa tca ccc tct aca aat tat cgt
1584

Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
515 520 525

gtt cgt gtt cgt tat gct tct aca tct tca ctc cct gta gat gta gtt
1632

Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
530 535 540

ttc gga aat att agt cat cct act aca ttc cca gcc act gcc aga tca
1680

Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
545 550 555 560

tta gat aat cta caa tcc aat gat ttt gga tat att gat att gct gga
1728

Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
565 570 575

act ttc tta cct tca cta ggg cct agt ata ggt atc aga cct atg tta
 1776
 Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
 580 585 590

tct act att aat ttg ata gta gat aga ttt gaa ttt att cca gta act
 1824
 Ser Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr
 595 600 605

gca acc ttt gaa gca gaa tcg gat tta gaa aga gca caa aag gcg gtg
 1872
 Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
 610 615 620

aat gcg ctg ttt act tct aca aac caa cta ggg ata aaa aca gat gtg
 1920
 Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
 625 630 635 640

acg gat tat cat att gat caa gtg tcc aat tta gtg gag tgt tta tcg
 1968
 Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
 645 650 655

gat gaa ttt tat ctg gat gaa aag cga gaa ttg tcc gag aaa gtc aaa
 2016
 Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
 660 665 670

cat gcg aag cga ctc agt gat gag cga aat tta ctt caa gat cca aac
 2064
 His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
 675 680 685

ttc agg ggc atc aat aga caa cca gat cgt ggc tgg aga gga agt acg
 2112
 Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
 690 695 700

gat att acc atc caa gga gga gat gac gta ttc aaa gag aat tac gtc
 2160
 Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
 705 710 715 720

aca cta cca ggt acc ttt gat gag tgc tat cca aca tat ttg tat caa
 2208
 Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
 725 730 735

aaa atc gat gaa tca aaa tta aaa gcc ttt acc cgt tat caa tta aga
 2256

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
 740 745 750
 ggg tat atc gaa gat agt caa gac tta gaa atc tat tta att cgc tac
 2304
 Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
 755 760 765
 aat gca aaa cat gaa aca gta aat gtg cca ggt acg ggt tcc tta tgg
 2352
 Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
 770 775 780
 ccg ctt tca gcc caa agt cca atc gga aag tgt gga gag ccg aat cga
 2400
 Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
 785 790 795 800
 tgc gcg cca cac ctt gaa tgg aat cct gac tta gat tgt tcg tgt agg
 2448
 Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
 805 810 815
 gat gga gaa aag tgt gcc cat cat tcg cat cat ttc tcc tta gac att
 2496
 Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
 820 825 830
 gat gta gga tgt aca gac tta aat gag gac cta ggt gta tgg gtg atc
 2544
 Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
 835 840 845
 ttt aag att aag acg caa gat ggg cac gca aga cta ggg aat cta gag
 2592
 Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
 850 855 860
 ttt ctc gaa gag aaa cca tta gta gga gaa gca cta gct cgt gtg aaa
 2640
 Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
 865 870 875 880
 aga gcg gag aaa aaa tgg aga gac aaa cgt gaa aaa ttg gaa tgg gaa
 2688
 Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
 885 890 895
 aca aat atc gtt tat aaa gag gca aaa gaa tct gta gat gct tta ttt
 2736
 Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
 900 905 910

gta aac tct caa tat gat caa tta caa gcg gat acg aat att gcc atg
2784

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

att cat gcg gca gat aaa cgt gtt cat agc att cga gaa gct tat ctg
2832

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

cct gag ctg tct gtg att ccg ggt gtc aat gcg gct att ttt gaa gaa
2880

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

tta gaa ggg cgt att ttc act gca ttc tcc cta tat gat gcg aga aat
2928

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

gtc att aaa aat ggt gat ttt aat aat ggc tta tcc tgc tgg aac gtg
2976

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

aaa ggg cat gta gat gta gaa gaa caa aac aac caa cgt tcg gtc ctt
3024

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

gtt gtt ccg gaa tgg gaa gca gaa gtg tca caa gaa gtt cgt gtc
3069

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
1010 1015 1020

tgt ccg ggt cgt ggc tat atc ctt cgt gtc aca gcg tac aag gag
3114

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
1025 1030 1035

gga tat gga gaa ggt tgc gta acc att cat gag atc gag aac aat
3159

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
1040 1045 1050

aca gac gaa ctg aag ttt agc aac tgc gta gaa gag gaa atc tat
3204

Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
1055 1060 1065

cca aat aac acg gta acg tgt aat gat tat act gta aat caa gaa
3249

Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
1070 1075 1080

gaa tac gga ggt gcg tac act tct cgt aat cga gga tat aac gaa
3294

Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
1085 1090 1095

gct cct tcc gta cca gct gat tat gcg tca gtc tat gaa gaa aaa
3339

Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
1100 1105 1110

tcg tat aca gat gga cga aga gag aat cct tgt gaa ttt aac aga
3384

Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
1115 1120 1125

ggg tat agg gat tac acg cca cta cca gtt ggt tat gtg aca aaa
3429

Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
1130 1135 1140

gaa tta gaa tac ttc cca gaa acc gat aag gta tgg att gag att
3474

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
1145 1150 1155

gga gaa acg gaa gga aca ttt atc gtg gac agc gtg gaa tta ctc
3519

Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
1160 1165 1170

ctt atg gag gaa tag
3534

Leu Met Glu Glu
1175

<210> 6

<211> 1177

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Cấu trúc tổng hợp

<400> 6

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu

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

210		215		220
Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val				
225		230		235
Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro				
	245		250	255
Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val				
	260		265	270
Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu				
	275		280	285
Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr				
	290		295	300
Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln				
305		310		315
Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro				
	325		330	335
Leu Phe Gly Asn Ala Gly Asn Ala Ala Pro Pro Val Leu Val Ser Leu				
	340		345	350
Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg				
	355		360	365
Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp				
	370		375	380
Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr				
385		390		395
Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro				
	405		410	415
Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser				

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

625					630						635				640
Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Glu	Cys	Leu	Ser
				645					650					655	
Asp	Glu	Phe	Tyr	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys
			660					665					670		
His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Pro	Asn
		675					680					685			
Phe	Arg	Gly	Ile	Asn	Arg	Gln	Pro	Asp	Arg	Gly	Trp	Arg	Gly	Ser	Thr
	690					695					700				
Asp	Ile	Thr	Ile	Gln	Gly	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val
705					710					715					720
Thr	Leu	Pro	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln
				725					730					735	
Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	Arg
			740					745					750		
Gly	Tyr	Ile	Glu	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	Tyr
		755					760					765			
Asn	Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp
	770					775					780				
Pro	Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg
785					790					795					800
Cys	Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	Arg
				805					810					815	
Asp	Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	Ile
			820					825					830		
Asp	Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile

835	840	845
Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu		
850	855	860
Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys		
865	870	875
Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu		
	885	890
Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe		
	900	905
Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met		
	915	920
Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu		
	930	935
Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu		
	945	950
Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn		
	965	970
Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val		
	980	985
Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu		
	995	1000
Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val		
	1010	1015
Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu		
	1025	1030
Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn		

1040		1045		1050
Thr Asp	Glu Leu Lys Phe	Ser Asn Cys Val	Glu Glu	Glu Ile Tyr
1055		1060		1065
Pro Asn	Asn Thr Val Thr	Cys Asn Asp Tyr	Thr Val	Asn Gln Glu
1070		1075		1080
Glu Tyr	Gly Gly Ala Tyr	Thr Ser Arg Asn Arg	Gly Tyr	Asn Glu
1085		1090		1095
Ala Pro	Ser Val Pro Ala	Asp Tyr Ala Ser	Val Tyr	Glu Glu Lys
1100		1105		1110
Ser Tyr	Thr Asp Gly Arg	Arg Glu Asn Pro	Cys Glu	Phe Asn Arg
1115		1120		1125
Gly Tyr	Arg Asp Tyr Thr	Pro Leu Pro Val	Gly Tyr	Val Thr Lys
1130		1135		1140
Glu Leu	Glu Tyr Phe Pro	Glu Thr Asp Lys	Val Trp	Ile Glu Ile
1145		1150		1155
Gly Glu	Thr Glu Gly Thr	Phe Ile Val Asp	Ser Val	Glu Leu Leu
1160		1165		1170
Leu Met	Glu Glu			
1175				

<210> 7
 <211> 3543
 <212> ADN
 <213> Trình tự nhân tạo

<220>
 <223> Trình tự tổng hợp hoàn toàn để sử dụng trong thực vật mã hóa
 BCW 001

<220>
 <221> CDS
 <222> (1)..(3540)

<223> BCW 001

<400> 7

atg gag gag aac aac cag aac cag tgt gtc cca tac aac tgc ctc aac
48

Met Glu Glu Asn Asn Gln Asn Gln Cys Val Pro Tyr Asn Cys Leu Asn
1 5 10 15

aac cca gcc atc gaa atc ctt gag ggc gac cga att tca gtc ggc aac
96

Asn Pro Ala Ile Glu Ile Leu Glu Gly Asp Arg Ile Ser Val Gly Asn
20 25 30

acg ccc atc gac atc tcc ctg agt ctt gtg gaa ctc ctc atc tcg gag
144

Thr Pro Ile Asp Ile Ser Leu Ser Leu Val Glu Leu Leu Ile Ser Glu
35 40 45

ttc gtc cct ggc ggc ggc ata atc acc ggt ctg ctc aac atc gtg tgg
192

Phe Val Pro Gly Gly Gly Ile Ile Thr Gly Leu Leu Asn Ile Val Trp
50 55 60

gga ttc gtg ggc cca tcc cag tgg gat gcg ttc ctg gcc caa gtg gag
240

Gly Phe Val Gly Pro Ser Gln Trp Asp Ala Phe Leu Ala Gln Val Glu
65 70 75 80

cag ctc atc aac cag agg atc tcc gag gcc gtc cgc aat acc gcg atc
288

Gln Leu Ile Asn Gln Arg Ile Ser Glu Ala Val Arg Asn Thr Ala Ile
85 90 95

caa gag ctg gag ggc atg gcc cgc gtg tac cgc acc tac gcc acc gcc
336

Gln Glu Leu Glu Gly Met Ala Arg Val Tyr Arg Thr Tyr Ala Thr Ala
100 105 110

ttt gct gaa tgg gag cgc gac ccg aac aac act gac ctg cgc gag gcc
384

Phe Ala Glu Trp Glu Arg Asp Pro Asn Asn Thr Asp Leu Arg Glu Ala
115 120 125

gtc cga aca cag ttc acg gcg acc gag acc tac atc agc ggc cgg atc
432

Val Arg Thr Gln Phe Thr Ala Thr Glu Thr Tyr Ile Ser Gly Arg Ile
130 135 140

tca gtg ctc aag atc cag aac ttc gag gtg cag ctc cta tcg gtc ttc
480

Ser Val Leu Lys Ile Gln Asn Phe Glu Val Gln Leu Leu Ser Val Phe
145 150 155 160

```

gcc cag gcc gcc aac ttg cac ctg agc ctc ctg cgg gac gtt gtg ttc
528
Ala Gln Ala Ala Asn Leu His Leu Ser Leu Leu Arg Asp Val Val Phe
      165                      170                      175

ttc ggc cag cgg tgg ggc ttc tct act acg acc gtg aac aac tac tac
576
Phe Gly Gln Arg Trp Gly Phe Ser Thr Thr Thr Val Asn Asn Tyr Tyr
      180                      185                      190

aac gac ctg acg gag gaa atc agc acc tac aca gat tac gca gtt cgt
624
Asn Asp Leu Thr Glu Glu Ile Ser Thr Tyr Thr Asp Tyr Ala Val Arg
      195                      200                      205

tgg tac aac acc ggc ctt gag cgc gtg tgg ggc ccg gac tcc cgc gat
672
Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg Asp
      210                      215                      220

tgg gtc cgc tac aac cag ttc cgc cgc gag ctg acg ctt aca gtg ctg
720
Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val Leu
      225                      230                      235                      240

gac atc gtc gca ctc ttt cct aac tac gac tcc agg cgc tat ccc atc
768
Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro Ile
      245                      250                      255

agg aca gtg tca cag ctc acc cgc gag atc tac aca aac ccg gtg ctt
816
Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val Leu
      260                      265                      270

gag aac ttc gac ggc agc ttc cgt ggc atg gcg cag cgc att gaa cag
864
Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu Gln
      275                      280                      285

aac atc cgc cag ccg cac ctt atg gac atc ttg aac agt atc act atc
912
Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr Ile
      290                      295                      300

tac acc gac gtc cac aga ggc ttc aac tac tgg agc gga cac cag atc
960
Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln Ile
      305                      310                      315                      320

```

aca gcc agc ccg gta ggc ttc tcg ggt cca gag ttc gcc ttc ccg ctg
1008

Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro Leu
325 330 335

ttt ggg aac gct ggc aat gcc gcg ccg ccc gtg ctg gtc agc ctc act
1056

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

ggt ctc ggc atc ttc cgc aca ctt tcc tcg ccg ctg tac agg agg atc
1104

Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg Ile
355 360 365

atc ctc ggg tcc ggt ccg aac aac cag gag ctg ttc gtg ctc gac ggg
1152

Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp Gly
370 375 380

acc gag ttc agt ttc gcc agc ctc acg acg aac ctc ccg tcc acc atc
1200

Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr Ile
385 390 395 400

tat cgc cag cgc gga acg gtc gat tcc ctg gat gtt atc cca ccg caa
1248

Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro Gln
405 410 415

gac aat tct gtg ccg ccg agg gcc ggg ttc tcc cac cgg ctg tct cac
1296

Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser His
420 425 430

gtg act atg ctt tca cag gcc gcc gga gcc gtg tac acg ctc cgt gcg
1344

Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg Ala
435 440 445

cct act ttc tcc tgg cag cac cgc agc gcg acc acg acc aac atc atc
1392

Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile Ile
450 455 460

gca gca gac tcc atc acc cag atc ccg gcc gtt aag ggc cgc agc atc
1440

Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser Ile
465 470 475 480

atc aac aac gga act gtc atc agc ggt ccg ggc ttc acg ggc ggc gac
1488

Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly Asp
 485 490 495
 ctg gtc cgg ctc tac aac gca gac ttc aac atc aat aac cgc gct tat
 1536
 Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala Tyr
 500 505 510
 ctt gaa gta cct atc ttc ttc cag agc ccg agc act aac tac cgg gtt
 1584
 Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg Val
 515 520 525
 cgc gtc cgc tac gcc agc acc tcc agc ctc cct gtg gat gtc gtg ttc
 1632
 Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val Phe
 530 535 540
 gga aac ata agc cat ccg acc acg ttc cca gcc acg gct agg agc ctg
 1680
 Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser Leu
 545 550 555 560
 gac aac cta cag agc aac gac ttc ggc tac atc gac atc gcg ggc acc
 1728
 Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly Thr
 565 570 575
 ttt ctg cca agc ctg ggt ccg tct atc ggc atc cgc ccg atg ctg agc
 1776
 Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu Ser
 580 585 590
 act atc aac cta att gtg gac cgg ttc gag ttt atc ccg gtg acg gca
 1824
 Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr Ala
 595 600 605
 acg ttc gag gcg gag tct gac ctc gaa agg gca cag aag gcc gtg aac
 1872
 Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn
 610 615 620
 gcc ctg ttc acg agc acc aac cag ctt ggc att aag act gat gtc acc
 1920
 Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val Thr
 625 630 635 640
 gac tac cac att gac caa gtc agc aac ctg gtg gag tgc ctc tcg gac
 1968
 Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser Asp
 645 650 655

gag ttc tat ctt gat gag aaa cgg gaa cta agc gag aag gtg aag cac
2016

Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys His
660 665 670

gca aag cgc ttg agc gac gag cgg aac tta ctc cag gac cct aac ttc
2064

Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn Phe
675 680 685

cgt ggg att aac cgc cag ccg gat cgc ggg tgg cgc ggc tca acg gac
2112

Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr Asp
690 695 700

atc acc atc cag ggc ggc gat gac gtc ttc aag gag aac tac gtg acc
2160

Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val Thr
705 710 715 720

ctc cct ggc acg ttc gac gag tgc tac ccg acg tac ctt tat cag aag
2208

Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln Lys
725 730 735

att gac gaa agc aag ctg aaa gcc tac acc cgc tac cag ttg cgc ggc
2256

Ile Asp Glu Ser Lys Leu Lys Ala Tyr Thr Arg Tyr Gln Leu Arg Gly
740 745 750

tac atc gag gac tct caa gac ctg gag atc tac ttg att cga tac aac
2304

Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr Asn
755 760 765

gcg aaa cac gag acc gtc aac gtg ccg ggc act ggg agc ctg tgg ccg
2352

Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp Pro
770 775 780

ttg tct gca caa agt ccg atc ggc aag tgc ggc gag cca aac cgg tgc
2400

Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg Cys
785 790 795 800

gct ccg cac ctg gag tgg aac cca gac ctt cat tgc tcc tgt agg gat
2448

Ala Pro His Leu Glu Trp Asn Pro Asp Leu His Cys Ser Cys Arg Asp
805 810 815

ggc gag aag tgc gct cac cac agc cat cac ttc agc ctc gac att gac
2496

Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile Asp
820 825 830

gtc ggc tgt acc gac ctt aat gag gat ctg ggt gtg tgg gtg atc ttc
2544

Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile Phe
835 840 845

aag atc aag acc cag gac ggt cac gcc cgg ttg ggc aat ctg gag ttc
2592

Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu Phe
850 855 860

ctg gag gag aag ccg ctg gtt ggc gag gct ctc gcg cgg gtc aag cgg
2640

Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys Arg
865 870 875 880

gcg gag aag aag tgg cgg gac aaa cgc gag aag ctc cag tta gag acg
2688

Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Gln Leu Glu Thr
885 890 895

aac atc gtg tac aag gag gcg aag gaa tcc gtg gac gca cta ttc gtg
2736

Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe Val
900 905 910

aac agc cag tac gac caa ctc cag gcc gac acc aac atc gcc atg att
2784

Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met Ile
915 920 925

cac gca gcc gac aag agg gtg cac cgc atc cgc gaa gcc tac ctt ccc
2832

His Ala Ala Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro
930 935 940

gaa ctt tcg gtc atc cca ggc gtc aac gct gac atc tcg gag gaa ttg
2880

Glu Leu Ser Val Ile Pro Gly Val Asn Ala Asp Ile Ser Glu Glu Leu
945 950 955 960

gag ggc aga atc ttc acg gcc ttc tct ttg tac gat gcc agg aac gtc
2928

Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val
965 970 975

atc aag aac ggc gac ttc aac aac ggc ctg ctg tgc tgg aac gtg aag
2976

Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys
 980 985 990

ggc cac gtg gac gtc gag gag cag aac aac cac aga tca gtc ctg gtg
 3024

Gly His Val Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val
 995 1000 1005

gtg ccc gag tgg gaa gcc gaa gtc tca caa gaa gtc cgg gtg tgc
 3069

Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
 1010 1015 1020

cct gga cgc ggg tac att ctc cgc gtg acc gcc tac aag gag ggc
 3114

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly
 1025 1030 1035

tac ggt gag ggc tgc gtg acc atc cac gag atc gag aac aac acc
 3159

Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr
 1040 1045 1050

gac gag ctg aaa ttc agt aac tgt gtt gag gag gag gtg tac ccg
 3204

Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Val Tyr Pro
 1055 1060 1065

aac aac acc gtc acc tgc aac gac tac act gcg aac cag gag gaa
 3249

Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Ala Asn Gln Glu Glu
 1070 1075 1080

tac gag ggc acg tac acg agc cgc aat cgc ggg tac gac gag gcg
 3294

Tyr Glu Gly Thr Tyr Thr Ser Arg Asn Arg Gly Tyr Asp Glu Ala
 1085 1090 1095

tac gag agc aac tcc agc gtc ccg gcc gag tac gcc tcc gtg tac
 3339

Tyr Glu Ser Asn Ser Ser Val Pro Ala Glu Tyr Ala Ser Val Tyr
 1100 1105 1110

gag gag aag gtt tac acc gac ggg agg cgt ggc aac ccg tgc gag
 3384

Glu Glu Lys Val Tyr Thr Asp Gly Arg Arg Gly Asn Pro Cys Glu
 1115 1120 1125

agc aac aga ggc tac ggc gat tac act ccg ctt ccc gct ggc tac
 3429

Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu Pro Ala Gly Tyr
 1130 1135 1140

gtg acg aaa gag ctg gag tac ttc cca gag acc gac aag gtg tgg
3474

Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp
1145 1150 1155

atc gag atc gga gaa acg gag ggc acg ttc ata gtg gac tcc gtt
3519

Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val
1160 1165 1170

gag ctg ctg ctc atg gag gag tag
3543

Glu Leu Leu Leu Met Glu Glu
1175 1180

<210> 8

<211> 1180

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Cấu trúc tổng hợp

<400> 8

Met Glu Glu Asn Asn Gln Asn Gln Cys Val Pro Tyr Asn Cys Leu Asn
1 5 10 15

Asn Pro Ala Ile Glu Ile Leu Glu Gly Asp Arg Ile Ser Val Gly Asn
20 25 30

Thr Pro Ile Asp Ile Ser Leu Ser Leu Val Glu Leu Leu Ile Ser Glu
35 40 45

Phe Val Pro Gly Gly Gly Ile Ile Thr Gly Leu Leu Asn Ile Val Trp
50 55 60

Gly Phe Val Gly Pro Ser Gln Trp Asp Ala Phe Leu Ala Gln Val Glu
65 70 75 80

Gln Leu Ile Asn Gln Arg Ile Ser Glu Ala Val Arg Asn Thr Ala Ile
85 90 95

Gln Glu Leu Glu Gly Met Ala Arg Val Tyr Arg Thr Tyr Ala Thr Ala

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

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

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

				725						730						735
Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Tyr	Thr	Arg	Tyr	Gln	Leu	Arg	Gly	
			740					745					750			
Tyr	Ile	Glu	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	Tyr	Asn	
		755					760					765				
Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	Pro	
		770				775					780					
Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg	Cys	
785					790					795					800	
Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	His	Cys	Ser	Cys	Arg	Asp	
			805						810					815		
Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	Ile	Asp	
			820					825					830			
Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile	Phe	
		835					840					845				
Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	Phe	
	850					855					860					
Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	Arg	
865					870					875					880	
Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Gln	Leu	Glu	Thr	
				885					890					895		
Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	Val	
			900					905					910			
Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	Ile	
		915					920				925					
His	Ala	Ala	Asp	Lys	Arg	Val	His	Arg	Ile	Arg	Glu	Ala	Tyr	Leu	Pro	

930		935		940
Glu Leu Ser Val Ile Pro Gly Val Asn Ala Asp Ile Ser Glu Glu Leu				
945		950	955	960
Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn Val				
	965		970	975
Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys				
	980		985	990
Gly His Val Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val				
	995	1000		1005
Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys				
1010		1015		1020
Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly				
1025		1030		1035
Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr				
1040		1045		1050
Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Val Tyr Pro				
1055		1060		1065
Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Ala Asn Gln Glu Glu				
1070		1075		1080
Tyr Glu Gly Thr Tyr Thr Ser Arg Asn Arg Gly Tyr Asp Glu Ala				
1085		1090		1095
Tyr Glu Ser Asn Ser Ser Val Pro Ala Glu Tyr Ala Ser Val Tyr				
1100		1105		1110
Glu Glu Lys Val Tyr Thr Asp Gly Arg Arg Gly Asn Pro Cys Glu				
1115		1120		1125
Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu Pro Ala Gly Tyr				


```

1130          1135          1140
Val Thr  Lys Glu Leu Glu Tyr  Phe Pro Glu Thr Asp  Lys Val Trp
1145          1150          1155

Ile Glu  Ile Gly Glu Thr Glu  Gly Thr Phe Ile Val  Asp Ser Val
1160          1165          1170

Glu Leu  Leu Leu Met Glu Glu
1175          1180

<210>  9
<211>  3534
<212>  ADN
<213>  Trình tự nhân tạo

<220>
<223>  Trình tự tổng hợp hoàn toàn mã hóa CW 002 để sử dụng trong thực
vật

<220>
<221>  CDS
<222>  (1)..(3531)
<223>  BCW 002

<400>  9
atg gac aac aac ccg aac atc aac gag tgc atc ccc tac aac tgc ctc
48
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1          5          10          15

tcc aac ccg gag gtc gag gtg ctg ggc ggc gaa agg atc gag acc ggc
96
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20          25          30

tac act ccc atc gac atc agc ctc agc ctg acc cag ttc ctg ctc tct
144
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35          40          45

gag ttc gtg ccc ggc gcg ggg ttc gtt ctc ggc ctg gtc gac atc atc
192
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50          55          60

```

tgg ggc atc ttc ggt ccg agc cag tgg gac gcc ttt ctc gtt cag att
 240
 Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80
 gag cag ctg atc aac cag cgc atc gag gag ttc gcc cgc aac cag gcg
 288
 Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95
 atc tcc cgg ctg gag ggc ctc tcc aac ctg tac caa atc tac gcc gag
 336
 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110
 agc ttc cgg gag tgg gaa gcc gat ccg acc aac ccc gct ctc agg gag
 384
 Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125
 gag atg cgg att cag ttc aac gac atg aac tcc gct ctc acg act gcc
 432
 Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140
 atc cca ctc ctc gct gtg cag aac tac caa gtg ccg ctc ctg tcc gtg
 480
 Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160
 tac gtg cag gcc gcc aat ctg cac ctc tcc gtc ctc cgg gac gtt agc
 528
 Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175
 gtg ttc ggg cag cgc tgg ggc ttc gac gcc gct acc atc aac tcc cgt
 576
 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190
 tac aac gat ctc act cgc ctc atc ggc aac tac acc gac tat gcc gtg
 624
 Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205
 cgc tgg tac aac act ggt ctt gag aga gtc tgg ggc ccg gac agc cgc
 672
 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220
 gac tgg gtg cgc tac aac cag ttc cgg cgc gag ctg acc ctc acc gtg
 720

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240
 ctc gac atc gta gcc ctc ttt ccc aac tac gac tcc cgg cgc tac ccg
 768
 Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255
 att cgc atc gtc agc cag ctc acc agg gag atc tac acc aac cct gtg
 816
 Ile Arg Ile Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270
 ctg gag aac ttc gac ggc tcc ttt cgc ggg atg gcc caa cgc ata gag
 864
 Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285
 cag aac atc cgc caa cct cat ctg atg gac atc ctt aat tct atc acc
 912
 Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300
 atc tac act gac gtt cat cgc ggg ttt aac tac tgg tcg ggc cac caa
 960
 Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320
 atc act gcg tcg ccc gtt ggt ttc tcc ggc ccg gag ttc gcg ttc cct
 1008
 Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335
 ctg ttc gga aac gcg ggc aat gcc gct cca ccc gta ttg gtg agc ctg
 1056
 Leu Phe Gly Asn Ala Gly Asn Ala Ala Pro Pro Val Leu Val Ser Leu
 340 345 350
 acc ggc ctc ggc atc ttc cgt aca ctg tct agc cct ctg tac aga agg
 1104
 Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
 355 360 365
 atc att ctt ggc agc ggt ccc aat aac cag gaa ctc ttc gtg ttg gac
 1152
 Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
 370 375 380
 ggc acc gag ttc agc ttc gcc agt ctt acg acc aat ttg ccc tcc aca
 1200
 Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
 385 390 395 400

atc tat cgc cag cgc ggt act gtg gac tcc ctt gat gtg ata cca cct
 1248
 Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
 405 410 415
 cag gac aac tct gtc cca cct cgc gcc ggt ttc tcc cac cgc ctc agc
 1296
 Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
 420 425 430
 cac gtc act atg ctg agt cag gct gcg gga gcc gtg tac acc ctt cgg
 1344
 His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
 435 440 445
 gct ccg acg ttt agc tgg cag cac agg agc gcg act acc acg aac atc
 1392
 Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
 450 455 460
 att gcg gct gac tcc atc act caa atc cct gcc gtt aag ggt cgc tcc
 1440
 Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
 465 470 475 480
 atc atc aac aat ggg aca gtg atc tcg gga ccg ggc ttc acc ggc ggt
 1488
 Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
 485 490 495
 gac ctg gtg agg ctg tac aac gcg gac ttc aac atc aac aac agg gcg
 1536
 Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
 500 505 510
 tac ctc gaa gtc ccg atc ttc ttc cag tcg ccc agc acg aac tat cgt
 1584
 Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
 515 520 525
 gtc agg gtc cgg tac gcc tca acc tca tcc ctc ccg gtc gat gtg gtc
 1632
 Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
 530 535 540
 ttc ggc aac atc agc cac ccg acc acg ttt ccg gct acc gcc cga tcc
 1680
 Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
 545 550 555 560

ctg gac aat ctg caa agc aac gat ttc ggc tac att gac att gcc ggg
1728

Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
565 570 575

acg ttc ctc ccg agc ctc ggc cca tcc atc ggc atc cgg ccc atg ctc
1776

Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
580 585 590

tcc acc atc aac ctg atc gtg gat cgg ttt gag ttc atc cca gtg aca
1824

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

gcc act ttc gag gct gag tcc gac cta gag cgt gct cag aag gca gtc
1872

Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

aat gct ctg ttt acc tcc acc aat cag ctc ggc att aag acc gat gtg
1920

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
625 630 635 640

acc gat tac cac att gac caa gtc tca aac ctc gtt gag tgc ctc tcg
1968

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
645 650 655

gat gag ttc tac ctt gat gag aag agg gag ctt tca gag aaa gtt aag
2016

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

cac gct aag aga ctc tcg gac gaa cgc aat ctg ttg caa gat ccc aac
2064

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

ttc aga ggg atc aac cgt cag cca gac cgg gga tgg cgc ggg tcc acg
2112

Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

gac atc act atc cag ggc ggt gat gac gtc ttc aag gag aac tac gtg
2160

Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

acc ctg ccg ggc acc ttt gac gaa tgc tac ccc act tac ctc tac cag
2208

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
 725 730 735

aag att gac gag tcc aag ctc aag gcg ttc aca cgc tac cag ctc agg
 2256

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
 740 745 750

ggt tac atc gag gac tcc caa gac ctg gaa atc tac ctg atc cgc tac
 2304

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
 755 760 765

aac gct aag cac gag act gtc aac gtg ccc ggc acc ggc agc ctg tgg
 2352

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
 770 775 780

ccc ttg tcc gct cag agc cca atc ggc aag tgc ggc gag ccc aac cgc
 2400

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
 785 790 795 800

tgc gcg ccc cac ctg gaa tgg aac ccc gac ctc gac tgt agc tgc cgc
 2448

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
 805 810 815

gac gga gag aag tgc gcg cat cac tcc cac cac ttc agc ctc gac atc
 2496

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
 820 825 830

gac gtc ggt tgc acc gac ctt aac gag gat ctg ggc gtt tgg gtg atc
 2544

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
 835 840 845

ttc aag atc aag act cag gac ggc cac gcc cgc ctg gga aac ctg gag
 2592

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
 850 855 860

ttc ctg gag gag aag ccc ctc gtt ggc gag gcc ctg gcc cgc gtc aag
 2640

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
 865 870 875 880

agg gcc gag aag aaa tgg cgc gac aag cgc gag aag ctg gag tgg gag
 2688

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
 885 890 895

acc aac atc gtg tac aag gaa gcg aag gag tca gtt gac gcc ctg ttc
2736

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

gtc aac agc cag tac gac cag ctc cag gca gac aca aac atc gct atg
2784

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

atc cat gcg gcc gac aag cgc gtc cac tcc atc cgc gag gcg tac ctg
2832

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

ccc gag ctg tcc gtc atc ccc ggc gtc aac gcc gcg atc ttt gag gag
2880

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

ctg gag ggc cgc atc ttc acc gcc ttc tcc ctc tac gac gca cgc aac
2928

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
965 970 975

gtt atc aag aat ggc gac ttc aac aac ggg ctg tcc tgc tgg aat gtc
2976

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

aag ggc cac gtg gac gtc gag gag cag aac aac cag cgc tca gtc ctg
3024

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

gtc gtc ccg gag tgg gag gcc gaa gtc agc cag gaa gtc cgc gtc
3069

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
1010 1015 1020

tgc cct gga cgc ggg tac atc ctg cgc gtc act gcc tac aag gaa
3114

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
1025 1030 1035

ggc tac gga gag ggc tgc gtc acc atc cat gag atc gaa aac aac
3159

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
1040 1045 1050

acg gat gag ctt aag ttc agc aac tgt gtt gaa gag gaa atc tac
 3204
 Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
 1055 1060 1065

 ccg aac aac acg gtc acc tgc aat gat tac acc gtc aac cag gag
 3249
 Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
 1070 1075 1080

 gaa tac ggt gga gct tac acc tcc cgc aac agg ggc tac aac gag
 3294
 Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
 1085 1090 1095

 gca ccc tct gtc ccg gcc gac tac gct tca gtc tac gaa gag aag
 3339
 Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
 1100 1105 1110

 tcg tac acc gac gga cgc aga gag aac ccg tgt gag ttc aac cgc
 3384
 Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
 1115 1120 1125

 ggc tac cgc gat tac acc ccg ctg cct gtc ggg tac gtc acc aaa
 3429
 Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
 1130 1135 1140

 gag ctg gaa tac ttc cca gag acc gac aaa gtc tgg att gag atc
 3474
 Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
 1145 1150 1155

 ggc gag acc gag ggc acg ttc atc gtg gac tcc gtc gaa ctc ctt
 3519
 Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
 1160 1165 1170

 ctg atg gaa gag tga
 3534
 Leu Met Glu Glu
 1175

<210> 10
 <211> 1177
 <212> PRT
 <213> Trình tự nhân tạo

<220>

<223> Cấu trúc tổng hợp

<400> 10

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
 65 70 75 80

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
 85 90 95

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
 100 105 110

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140

Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255

Ile Arg Ile Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285

Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335

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

Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
 355 360 365

Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
 370 375 380

Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
 385 390 395 400

Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
 405 410 415

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
 420 425 430

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
 435 440 445

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
 450 455 460

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
 465 470 475 480

Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
 485 490 495

Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
 500 505 510

Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
 515 520 525

Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
 530 535 540

Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
 545 550 555 560

Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
 565 570 575

Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
 580 585 590

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

Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
610 615 620

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
645 650 655

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
 820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
 835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
 850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
 865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
 885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
 900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
 915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
 930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
 945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
 1010 1015 1020

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
1025 1030 1035

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
1040 1045 1050

Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
1055 1060 1065

Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
1070 1075 1080

Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
1085 1090 1095

Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
1100 1105 1110

Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
1115 1120 1125

Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
1130 1135 1140

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
1145 1150 1155

Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
1160 1165 1170

Leu Met Glu Glu
1175

<210> 11
<211> 3534
<212> ADN
<213> Trình tự nhân tạo

<220>
<223> Trình tự nucleotit tổng hợp hoàn toàn mã hóa BCW 003 để sử dụng trong thực vật

<220>

<221> CDS

<222> (1)..(3531)

<223> BCW 003

<400> 11

atg gac aac aac ccg aac atc aac gag tgc atc ccc tac aac tgc ctc
48

Met Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1 5 10 15

tcc aac ccg gag gtc gag gtg ctg ggc ggc gaa agg atc gag acc ggc
96

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20 25 30

tac act ccc atc gac atc agc ctc agc ctg acc cag ttc ctg ctc tct
144

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35 40 45

gag ttc gtg ccc ggc gcg ggg ttc gtt ctc ggc ctg gtc gac atc atc
192

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50 55 60

tgg ggc atc ttc ggt ccg agc cag tgg gac gcc ttt ctc gtt cag att
240

Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
65 70 75 80

gag cag ctg atc aac cag cgc atc gag gag ttc gcc cgc aac cag gcg
288

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85 90 95

atc tcc cgg ctg gag ggc ctc tcc aac ctg tac caa atc tac gcc gag
336

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
100 105 110

agc ttc cgg gag tgg gaa gcc gat ccg acc aac ccc gct ctc agg gag
384

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115 120 125

gag atg cgg att cag ttc aac gac atg aac tcc gct ctc acg act gcc
432

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

atc cca ctc ctc gct gtg cag aac tac caa gtg ccg ctc ctg tcc gtg
 480
 Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160
 tac gtg cag gcc gcc aat ctg cac ctc tcc gtc ctc cgg gac gtt agc
 528
 Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175
 gtg ttc ggg cag cgc tgg ggc ttc gac gcc gct acc atc aac tcc cgt
 576
 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190
 tac aac gat ctc act cgc ctc atc ggc aac tac acc gac tat gcc gtg
 624
 Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205
 cgc tgg tac aac act ggt ctt gag aga gtc tgg ggc ccg gac agc cgc
 672
 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220
 gac tgg gtg cgc tac aac cag ttc cgg cgc gag ctg acc ctc acc gtg
 720
 Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240
 ctc gac atc gta gcc ctc ttt ccc aac tac gac tcc cgg cgc tac ccg
 768
 Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255
 att cgc acc gtc agc cag ctc acc agg gag atc tac acc aac cct gtg
 816
 Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270
 ctg gag aac ttc gac ggc tcc ttt cgc ggg atg gcc caa cgc ata gag
 864
 Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
 275 280 285
 cag aac atc cgc caa cct cat ctg atg gac atc ctt aat tct atc acc
 912
 Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

atc tac act gac gtt cat cgc ggg ttt aac tac tgg tcg ggc cac caa
960

Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
305 310 315 320

atc act gcg tcg ccc gtt ggt ttc tcc ggc ccg gag ttc gcg ttc cct
1008

Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
325 330 335

ctg ttc gga aac gcg ggc aat gcc gct cca ccc gta ttg gtg agc ctg
1056

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

acc ggc ctc ggc atc ttc cgt aca ctg tct agc cct ctg tac aga agg
1104

Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
355 360 365

atc att ctt ggc agc ggt ccc aat aac cag gaa ctc ttc gtg ttg gac
1152

Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
370 375 380

ggc acc gag ttc agc ttc gcc agt ctt acg acc aat ttg ccc tcc aca
1200

Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
385 390 395 400

atc tat cgc cag cgc ggt act gtg gac tcc ctt gat gtg ata cca cct
1248

Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
405 410 415

cag gac aac tct gtc cca cct cgc gcc ggt ttc tcc cac cgc ctc agc
1296

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
420 425 430

cac gtc act atg ctg agt cag gct gcg gga gcc gtg tac acc ctt cgg
1344

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
435 440 445

gct ccg acg ttt agc tgg cag cac agg agc gcg act acc acg aac atc
1392

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
450 455 460

att gcg gct gac tcc atc act caa atc cct gcc gtt aag ggt cgc tcc
1440

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
 465 470 475 480
 atc atc aac aat ggg aca gtg atc tcg gga ccg ggc ttc acc ggc ggt
 1488
 Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
 485 490 495
 gac ctg gtg agg ctg tac aac gcg gac ttc aac atc aac aac agg gcg
 1536
 Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
 500 505 510
 tac ctc gaa gtc ccg atc ttc ttc cag tcg ccc agc acg aac tat cgt
 1584
 Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
 515 520 525
 gtc agg gtc cgg tac gcc tca acc tca tcc ctc ccg gtc gat gtg gtc
 1632
 Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
 530 535 540
 ttc ggc aac atc agc cac ccg acc acg ttt ccg gct acc gcc cga tcc
 1680
 Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
 545 550 555 560
 ctg gac aat ctg caa agc aac gat ttc ggc tac att gac att gcc ggg
 1728
 Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
 565 570 575
 acg ttc ctc ccg agc ctc ggc cca tcc atc ggc atc ccg ccc atg ctc
 1776
 Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
 580 585 590
 tcc acc atc aac ctg atc gtg gat ccg ttt gag ttc atc cca gtg aca
 1824
 Ser Thr Ile Asn Leu Ile Val Asp Arg Phe Glu Phe Ile Pro Val Thr
 595 600 605
 gcc act ttc gag gct gag tcc gac cta gag cgt gct cag aag gca gtc
 1872
 Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
 610 615 620
 aat gct ctg ttt acc tcc acc aat cag ctc ggc att aag acc gat gtg
 1920
 Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
 625 630 635 640

acc gat tac cac att gac caa gtc tca aac ctc gtt gag tgc ctc tcg
1968

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
645 650 655

gat gag ttc tac ctt gat gag aag agg gag ctt tca gag aaa gtt aag
2016

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
660 665 670

cac gct aag aga ctc tcg gac gaa cgc aat ctg ttg caa gat ccc aac
2064

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
675 680 685

ttc aga ggg atc aac cgt cag cca gac cgg gga tgg cgc ggg tcc acg
2112

Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
690 695 700

gac atc act atc cag ggc ggt gat gac gtc ttc aag gag aac tac gtg
2160

Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

acc ctg ccg ggc acc ttt gac gaa tgc tac ccc act tac ctc tac cag
2208

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

aag att gac gag tcc aag ctc aag gcg ttc aca cgc tac cag ctc agg
2256

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

ggt tac atc gag gac tcc caa gac ctg gaa atc tac ctg atc cgc tac
2304

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

aac gct aag cac gag act gtc aac gtg ccc ggc acc ggc agc ctg tgg
2352

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

ccc ttg tcc gct cag agc cca atc ggc aag tgc ggc gag ccc aac cgc
2400

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

tgc gcg ccc cac ctg gaa tgg aac ccc gac ctc gac tgt agc tgc cgc
2448

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

gac gga gag aag tgc gcg cat cac tcc cac cac ttc agc ctc gac atc
2496

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

gac gtc ggt tgc acc gac ctt aac gag gat ctg ggc gtt tgg gtg atc
2544

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

ttc aag atc aag act cag gac ggc cac gcc cgc ctg gga aac ctg gag
2592

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

ttc ctg gag gag aag ccc ctc gtt ggc gag gcc ctg gcc cgc gtc aag
2640

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

agg gcc gag aag aaa tgg cgc gac aag cgc gag aag ctg gag tgg gag
2688

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

acc aac atc gtg tac aag gaa gcg aag gag tca gtt gac gcc ctg ttc
2736

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

gtc aac agc cag tac gac cag ctc cag gca gac aca aac atc gct atg
2784

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
915 920 925

atc cat gcg gcc gac aag cgc gtc cac tcc atc cgc gag gcg tac ctg
2832

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
930 935 940

ccc gag ctg tcc gtc atc ccc ggc gtc aac gcc gcg atc ttt gag gag
2880

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
945 950 955 960

ctg gag ggc cgc atc ttc acc gcc ttc tcc ctc tac gac gca cgc aac
2928

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975

gtt atc aag aat ggc gac ttc aac aac ggg ctg tcc tgc tgg aat gtc
 2976

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990

aag ggc cac gtg gac gtc gag gag cag aac aac cag cgc tca gtc ctg
 3024

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005

gtc gtc ccg gag tgg gag gcc gaa gtc agc cag gaa gtc cgc gtc
 3069

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
 1010 1015 1020

tgc cct gga cgc ggg tac atc ctg cgc gtc act gcc tac aag gaa
 3114

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
 1025 1030 1035

ggc tac gga gag ggc tgc gtc acc atc cat gag atc gaa aac aac
 3159

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
 1040 1045 1050

acg gat gag ctt aag ttc agc aac tgt gtt gaa gag gaa atc tac
 3204

Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
 1055 1060 1065

ccg aac aac acg gtc acc tgc aat gat tac acc gtc aac cag gag
 3249

Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
 1070 1075 1080

gaa tac ggt gga gct tac acc tcc cgc aac agg ggc tac aac gag
 3294

Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
 1085 1090 1095

gca ccc tct gtc ccg gcc gac tac gct tca gtc tac gaa gag aag
 3339

Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
 1100 1105 1110

tcg tac acc gac gga cgc aga gag aac ccg tgt gag ttc aac cgc
 3384

Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
 1115 1120 1125

ggc tac cgc gat tac acc ccg ctg cct gtc ggg tac gtc acc aaa
3429

Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
1130 1135 1140

gag ctg gaa tac ttc cca gag acc gac aaa gtc tgg att gag atc
3474

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
1145 1150 1155

ggc gag acc gag ggc acg ttc atc gtg gac tcc gtc gaa ctc ctt
3519

Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
1160 1165 1170

ctg atg gaa gag tga
3534

Leu Met Glu Glu
1175

<210> 12

<211> 1177

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Cấu trúc tổng hợp

<400> 12

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1 5 10 15

Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20 25 30

Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35 40 45

Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50 55 60

Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
65 70 75 80

Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85 90 95

Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
100 105 110

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
115 120 125

Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130 135 140

Ile Pro Leu Leu Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165 170 175

Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
195 200 205

Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210 215 220

Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
245 250 255

Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Met Ala Gln Arg Ile Glu
275 280 285

Gln Asn Ile Arg Gln Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

Ile Tyr Thr Asp Val His Arg Gly Phe Asn Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Thr Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Ala Phe Pro
 325 330 335

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

Thr Gly Leu Gly Ile Phe Arg Thr Leu Ser Ser Pro Leu Tyr Arg Arg
 355 360 365

Ile Ile Leu Gly Ser Gly Pro Asn Asn Gln Glu Leu Phe Val Leu Asp
 370 375 380

Gly Thr Glu Phe Ser Phe Ala Ser Leu Thr Thr Asn Leu Pro Ser Thr
 385 390 395 400

Ile Tyr Arg Gln Arg Gly Thr Val Asp Ser Leu Asp Val Ile Pro Pro
 405 410 415

Gln Asp Asn Ser Val Pro Pro Arg Ala Gly Phe Ser His Arg Leu Ser
 420 425 430

His Val Thr Met Leu Ser Gln Ala Ala Gly Ala Val Tyr Thr Leu Arg
 435 440 445

Ala Pro Thr Phe Ser Trp Gln His Arg Ser Ala Thr Thr Thr Asn Ile
 450 455 460

Ile Ala Ala Asp Ser Ile Thr Gln Ile Pro Ala Val Lys Gly Arg Ser
 465 470 475 480

Ile Ile Asn Asn Gly Thr Val Ile Ser Gly Pro Gly Phe Thr Gly Gly
 485 490 495

Asp Leu Val Arg Leu Tyr Asn Ala Asp Phe Asn Ile Asn Asn Arg Ala
 500 505 510

Tyr Leu Glu Val Pro Ile Phe Phe Gln Ser Pro Ser Thr Asn Tyr Arg
 515 520 525

Val Arg Val Arg Tyr Ala Ser Thr Ser Ser Leu Pro Val Asp Val Val
 530 535 540

Phe Gly Asn Ile Ser His Pro Thr Thr Phe Pro Ala Thr Ala Arg Ser
 545 550 555 560

Leu Asp Asn Leu Gln Ser Asn Asp Phe Gly Tyr Ile Asp Ile Ala Gly
 565 570 575

Thr Phe Leu Pro Ser Leu Gly Pro Ser Ile Gly Ile Arg Pro Met Leu
 580 585 590

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

Ala Thr Phe Glu Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val
 610 615 620

Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Ile Lys Thr Asp Val
 625 630 635 640

Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Glu Cys Leu Ser
 645 650 655

Asp Glu Phe Tyr Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
 660 665 670

His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn
 675 680 685

Phe Arg Gly Ile Asn Arg Gln Pro Asp Arg Gly Trp Arg Gly Ser Thr
 690 695 700

Asp Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
705 710 715 720

Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln
725 730 735

Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg
740 745 750

Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr
755 760 765

Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp
770 775 780

Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg
785 790 795 800

Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg
805 810 815

Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile
820 825 830

Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile
835 840 845

Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu
850 855 860

Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys
865 870 875 880

Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu
885 890 895

Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe
900 905 910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
 915 920 925

Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
 930 935 940

Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
 945 950 955 960

Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990

Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val
 1010 1015 1020

Cys Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu
 1025 1030 1035

Gly Tyr Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn
 1040 1045 1050

Thr Asp Glu Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr
 1055 1060 1065

Pro Asn Asn Thr Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu
 1070 1075 1080

Glu Tyr Gly Gly Ala Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu
 1085 1090 1095

Ala Pro Ser Val Pro Ala Asp Tyr Ala Ser Val Tyr Glu Glu Lys
 1100 1105 1110

Ser Tyr Thr Asp Gly Arg Arg Glu Asn Pro Cys Glu Phe Asn Arg
1115 1120 1125

Gly Tyr Arg Asp Tyr Thr Pro Leu Pro Val Gly Tyr Val Thr Lys
1130 1135 1140

Glu Leu Glu Tyr Phe Pro Glu Thr Asp Lys Val Trp Ile Glu Ile
1145 1150 1155

Gly Glu Thr Glu Gly Thr Phe Ile Val Asp Ser Val Glu Leu Leu
1160 1165 1170

Leu Met Glu Glu
1175