



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) CỤC SỞ HỮU TRÍ TUỆ



1-0037899

(51)^{2020.01} C12N 9/22; C12N 15/79; C07K 14/435; (13) B
C12N 15/52

-
- (21) 1-2021-01866 (22) 13/09/2019
(86) PCT/RU2019/050154 13/09/2019 (87) WO 2020/055293 19/03/2020
(30) 2018132816 14/09/2018 RU
(45) 25/12/2023 429 (43) 26/07/2021 400
(73) JOINT STOCK COMPANY "BIOCAD" (RU)
Liter A, bld. 34, Svyazi st., Strelna, Petrodvortsoviy district, St.Petersburg, 198515,
Russian Federation
(72) MADERA, Dmitriy Aleksandrovich (RU); KARABELSKII, Aleksandr
Vladimirovich (RU); IVANOV, Roman Alekseevich (RU); MOROZOV, Dmitry
Valentinovich (RU); SEVERINOV, Konstantin Viktorovich (RU); SHMAKOV,
Sergey Anatolevich (RU); SUTORMIN, Dmitrii Aleksandrovich (RU);
POBEGALOV, Georgii Evgenevich (RU); VASILEVA, Aleksandra Andreevna
(RU); SELKOVA, Polina Anatolevna (RU); ARSENIYEV, Anatolii Nikolaevich
(RU); ZYUBKO, Tatyana Igorevna (RU); FEDOROVA, Iana Vitalevna (RU).
(74) Công ty Luật TNHH WINCO (WINCO LAW FIRM)
-

(54) ENZYM PACAS9 NUCLEAZA, PHƯƠNG PHÁP TẠO RA VÀ PHÂN TỬ AXIT
NUCLEIC ĐƯỢC PHÂN LẬP MÃ HÓA ENZYM NÀY, VECTƠ BIỂU HIỆN VÀ
TẾ BÀO CHỦ CHỨA AXIT NÀY

(57) Sáng chế này đề cập đến lĩnh vực công nghệ sinh học, sinh học phân tử và y học, cụ thể là đề cập đến enzym nucleaza và mô tả việc sử dụng enzym này. Cụ thể hơn, sáng chế đề cập đến enzym PaCas9 nucleaza. Sáng chế cũng đề cập đến axit nucleic mã hóa nucleaza nêu trên, cấu trúc di truyền, vector biểu hiện, vector vận chuyển, chứa axit nucleic nêu trên, liposom chứa nucleaza nêu trên hoặc axit nucleic mã hóa nucleaza này, phương pháp tạo ra nucleaza, phương pháp vận chuyển, và tế bào chủ chứa axit nucleic mã hóa nucleaza nêu trên.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực công nghệ sinh học, sinh học phân tử và y học, cụ thể là đề cập đến enzym nucleaza và mô tả việc sử dụng enzym này. Cụ thể hơn, sáng chế đề cập đến enzym PaCas9 nucleaza. Sáng chế cũng đề cập đến axit nucleic mã hóa nucleaza nêu trên, cấu trúc di truyền, vector biểu hiện, vector vận chuyển, chứa axit nucleic nêu trên, liposom chứa nucleaza nêu trên hoặc axit nucleic mã hóa nucleaza nêu trên, phương pháp tạo ra nucleaza, phương pháp vận chuyển, và tế bào chủ chứa axit nucleic mã hóa nucleaza nêu trên.

Tình trạng kỹ thuật của sáng chế

CRISPR-Cas là hệ miễn dịch thích ứng trong nhiều vi khuẩn và phần lớn vi khuẩn cổ được phát hiện lần đầu tiên vào năm 2007 (Barrangou et al., 2007, Science 315: 1709-1712, Brouns et al., 2008, Science 321: 960-964). Cho đến nay, dựa trên các tiêu chuẩn về cấu trúc và chức năng, có ba loại hệ CRISPR-Cas đã được xác định tính chất, phần lớn trong số chúng sử dụng các phân tử ARN nhỏ làm phân tử dẫn đường để hướng đích các trình tự ADN bổ trợ (Makarova et al., 2011, Nat Rev Microbiol 9: 467-477, Van der Oost et al., 2014, Nat Rev Microbiol 12: 479-492).

Trong nghiên cứu gần đây bởi phòng thí nghiệm Doudna/Charpentier, việc xác định đặc tính của enzym tác động của hệ CRISPR-Cas typ II (Cas9) được tiến hành kỹ lưỡng, bao gồm cả việc chứng minh rằng sự đưa vào các phân tử ARN dẫn đường CRISPR đã thiết kế (với các trình tự của đoạn đệm cụ thể) hướng đích các trình tự bổ trợ (các đoạn tiền đệm) đối với plasmit, gây ra sự đứt gãy sợi kép của plasmit này (Jinek et al., 2012, Science 337: 816- 821). Tài liệu: Later, Jinek et al., 2012 sử dụng Cas9 làm công cụ để chỉnh sửa hệ gen.

Cas9 đã được sử dụng để thiết kế hệ gen của nhiều tế bào có nhân điển hình (ví dụ, cá, thực vật, người) (Charpentier and Doudna, 2013, Nature 495: 50-51).

Ngoài ra, Cas9 đã được sử dụng để cải thiện năng suất tái tổ hợp tương đồng ở vi khuẩn bằng cách chọn lọc các sự kiện tái tổ hợp riêng biệt (Jiang et al., 2013,

Nature Biotechnol 31: 233-239). Để đạt được mục đích này, mảnh có độc tố (cấu trúc hướng đích) được đồng chuyển nhiễm với mảnh phục hồi mang sự biến đổi mong muốn (cấu trúc chỉnh sửa, mang sự đột biến điểm hoặc sự mất đoạn). Cấu trúc hướng đích gồm Cas 9 kết hợp với CRISPR thiết kế và gen đánh dấu sự kháng thuốc kháng sinh, để xác định vị trí của sự tái tổ hợp mong muốn trên nhiễm sắc thể của vật chủ; khi có mặt chất kháng sinh tương ứng, sự dính gán cấu trúc hướng đích trong nhiễm sắc thể của vật chủ được chọn lọc. Chỉ khi xuất hiện sự tái tổ hợp bổ sung của cấu trúc chỉnh sửa với vị trí đích CRISPR ở vị trí khác trên nhiễm sắc thể của vật chủ, vật chủ này có thể tránh được vấn đề về tính tự miễn dịch. Do đó, khi có mặt chất kháng sinh, chỉ các thể đột biến mong muốn (không có gen đánh dấu) có thể sống sót và phát triển. Chiến lược liên quan để chọn lọc sự loại bỏ cấu trúc hướng đích đã dính gán sau đó ra khỏi nhiễm sắc thể cũng được thể hiện, tạo ra thể đột biến không có gen đánh dấu thực.

Trong những năm gần đây, đã xác định được rằng việc chỉnh sửa hệ gen qua trung gian CRISPR-Cas tạo thành công cụ hữu ích cho kỹ thuật di truyền. Đã xác định được rằng các hệ CRISPR của sinh vật có nhân nguyên thủy sử dụng vật chủ của chúng làm hệ miễn dịch thích ứng (Jinek et al., 2012, Science 337: 816-821) và có thể được sử dụng cho kỹ thuật di truyền hữu hiệu và nhanh chóng (ví dụ, Mali et al., 2013, Nat Methods 10: 957-963), chỉ cần cải biến trình tự dẫn đường để hướng đích các trình tự quan tâm.

Tuy nhiên, vẫn cần phát triển các chất có sự phát hiện, phân giải và thao tác axit nucleic đặc hiệu trình tự được cải tiến trong các điều kiện thử nghiệm khác nhau để sử dụng trong lĩnh vực nghiên cứu di truyền và chỉnh sửa hệ gen.

Bản chất kỹ thuật của sáng chế

Sáng chế đề cập đến enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2.

Theo một khía cạnh, sáng chế đề cập đến phân tử axit nucleic được phân lập mã hóa enzym PaCas9 nucleaza có trình tự nucleotit nêu trong SEQ ID NO: 1.

Theo một khía cạnh, sáng chế đề cập đến vector biểu hiện chứa axit nucleic có trình tự nucleotit nêu trong SEQ ID NO: 1.

Theo một số phương án, vector biểu hiện có cấu trúc di truyền như được thể

hiện trên Fig.1, PpCas9-T2A-GFP-sgARN1-MCS-sgARN2-MCS.

Theo một khía cạnh, sáng chế đề cập đến vectơ để vận chuyển chất điều trị chứa axit nucleic có trình tự nucleotit nêu trong SEQ ID NO: 1.

Theo một phương án của sáng chế, vectơ này vận chuyển chất điều trị đến các mô đích hoặc tế bào đích.

Theo một khía cạnh, sáng chế đề cập đến liposom để vận chuyển chất điều trị chứa enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2 hoặc axit nucleic có trình tự nucleotit nêu trong SEQ ID NO: 1.

Theo một phương án của sáng chế, liposom vận chuyển chất điều trị đến các mô đích hoặc tế bào đích.

Theo một khía cạnh, sáng chế đề cập đến phương pháp vận chuyển chất điều trị đến các mô đích hoặc tế bào đích bằng cách sử dụng vectơ hoặc liposom nêu trên.

Theo một phương án của phương pháp này, chất điều trị được vận chuyển đến các tế bào đích hoặc mô đích bằng cách sử dụng vectơ hoặc liposom nêu trên vào cơ thể của động vật có vú.

Theo một khía cạnh, sáng chế đề cập đến phương pháp tạo ra tế bào chủ để tiết ra enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2, phương pháp này bao gồm bước biến nạp tế bào bằng cách sử dụng vectơ bất kỳ trong số các vectơ nêu trên.

Theo một khía cạnh, sáng chế đề cập đến phương pháp tạo ra enzym PaCas9 nucleaza, phương pháp này bao gồm bước nuôi cấy tế bào chủ nêu trên trong môi trường sinh trưởng trong điều kiện đủ để tiết ra PaCas9 enzym nucleaza nêu trên, nếu cần, sau đó phân lập và tinh chế enzym PaCas9 nucleaza thu được.

Mô tả vắn tắt các hình vẽ

Fig.1 thể hiện sơ đồ dạng vòng của plasmid PpCas9-T2A-GFP-sgARN1-MCS-sgARN2-MCS dùng để tạo ra enzym PpCas9 nucleaza ở tế bào của động vật có vú.

AmpR là gen beta-lactamaza tạo ra sự kháng ampicillin,

CMV promoter là gen khởi đầu sớm của virus cự bào,

Trình tự Kozak được dùng để làm gia tăng hiệu quả dịch mã của protein,

START codon là codon khởi đầu,

NLS để chỉ tín hiệu khu trú trong nhân (nuclear localization signal: NLS),

PaCas9 là trình tự nucleotit nêu trong SEQ ID NO: 1 mã hóa enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2,

FLAG là trình tự epitop FLAG được sử dụng để phát hiện protein,

GFP là protein phát huỳnh quang màu xanh lục được cải biến,

TK pA để chỉ trình tự tín hiệu của poly-A thymidin kinaza được sử dụng để làm gia tăng tính ổn định của mRNA

F1 ori là nguồn gốc sao chép để cho phép đóng gói phagemit thành các tiểu phần của thể thực khuẩn khi được đồng chuyển nhiễm với các thể thực khuẩn hỗ trợ,

polIII term + U6 promotor để chỉ các băng biểu hiện của các phân tử RNA nhỏ, mỗi băng chứa gen khởi đầu U6 và gen kết thúc phiên mã RNA polymeraza III.

pUC origin là nguồn gốc sao chép pUC ở vi khuẩn.

Fig.2 thể hiện trình tự axit amin của enzym PaCas9 nucleaza có sự phân bố theo miền.

Mô tả chi tiết sáng chế

Các định nghĩa và phương pháp chung

Nếu không được định nghĩa theo cách khác, tất cả các thuật ngữ khoa học và kỹ thuật được sử dụng ở đây có nghĩa giống như thường được hiểu bởi người có hiểu biết trung bình trong lĩnh vực này.

Ngoài ra, trừ khi có yêu cầu theo cách khác do ngữ cảnh, các thuật ngữ số ít sẽ bao gồm cả nghĩa số nhiều và các thuật ngữ số nhiều sẽ bao gồm cả nghĩa số ít. Thông thường, sự sắp xếp và phương pháp nuôi cấy tế bào, sinh học phân tử, miễn dịch học, vi sinh học, di truyền học, hóa học phân tích, hóa hữu cơ tổng hợp, y học và hóa dược, cũng như kỹ thuật lai và hóa học protein và axit nucleic được mô tả ở đây là đã được biết rõ và sử dụng rộng rãi bởi người có hiểu biết trung bình trong lĩnh vực này. Các phản ứng enzym và phương pháp tinh chế được thực hiện theo các hướng dẫn của nhà sản xuất, như thường được tiến hành trong lĩnh vực này, hoặc như được mô tả ở đây.

Thuật ngữ "động vật có vú" để chỉ động vật bất kỳ được phân loại làm động vật có vú, bao gồm động vật linh trưởng, người, động vật gặm nhấm, chó, mèo, gia súc lớn, gia súc nhỏ, ngựa, lợn, v.v..

Nucleaza

Các enzym nucleaza là nhóm các enzym lớn thủy phân liên kết phosphodiester giữa các tiểu đơn vị axit nucleic.

Tùy thuộc vào độ đặc hiệu và hoạt tính, các enzym nucleaza có thể thuộc về một trong số các loại sau: exonucleaza và endonucleaza, ribonucleaza và deoxyribonucleaza, restrictaza và một số enzym khác. Các enzym restrictaza là thành phần quan trọng được sử dụng trong lĩnh vực sinh học phân tử.

PaCas9 nucleaza để chỉ loại enzym deoxyribonucleaza.

PaCas9 nucleaza có khả năng phân giải ADN chứa trình tự axit nucleic đích mà khi gắn kết với ít nhất một phân tử ARN sẽ nhận biết trình tự đích.

PaCas9 nucleaza chứa hai miền endonucleaza, mỗi miền gây ra sự đứt gãy sợi đơn, và khi tác dụng cùng nhau sẽ gây ra sự đứt gãy sợi kép.

PaCas9 nucleaza là enzym tác động của hệ CRISPR-Cas typ II (nucleaza typ 2).

PaCas9 nucleaza có khả năng gây ra sự đứt gãy của ADN sợi kép với vị trí nhận biết có độ đặc hiệu cao (từ 16 đến 20 chữ cái).

ADN của PaCas9 nucleaza được thể hiện trong SEQ ID NO:1.

Trình tự axit amin của PaCas9 nucleaza được thể hiện trong SEQ ID NO:2.

Fig.2 thể hiện trình tự axit amin của PaCas9 nucleaza có sự phân bố theo miền.

PaCas9 nucleaza có liên quan đến các đoạn lặp lại đối xứng ngắn cách đều thành cụm (clustered regularly interspaced short palindromic repeat: CRISPR), cũng như các thành phần liên kế khác của hệ CRISPR-Cas: các trình tự crARN và tracrARN.

Trình tự nucleotit mã hóa tracrARN được thể hiện trong SEQ ID NO:3.

Trình tự nucleotit mã hóa đoạn lặp xuôi (DR) được thể hiện trong SEQ ID NO:4.

crARN gồm phần biến đổi phụ thuộc đích và đoạn lặp xuôi DR được thể hiện trong trình tự SEQ ID NO:4.

Thuật ngữ "chất điều trị" trong bản mô tả này để chỉ enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2 hoặc phân tử axit nucleic được phân lập mã hóa enzym PaCas9 nucleaza và có trình tự nucleotit nêu trong SEQ ID NO: 1.

tracrARN (crARN chuyển hoạt hóa) là ARN nhỏ được chuyển mã hóa.

CRISPR (đoạn lặp lại đối xứng ngắn cách đều thành cụm) là locus của vi khuẩn và vi khuẩn cổ đặc biệt gồm các đoạn lặp xuôi được ngăn cách bằng các trình tự đặc biệt (đoạn đệm).

Phân tử axit nucleic

Các thuật ngữ "axit nucleic", "trình tự nucleic", "trình tự axit nucleic", "polynucleotit", "oligonucleotit", "trình tự polynucleotit" và "trình tự nucleotit", được sử dụng thay thế lẫn nhau trong bản mô tả này, để chỉ trình tự chính xác của các nucleotit, được cải biến hoặc không được cải biến, xác định mảnh hoặc vùng của axit nucleic, chứa hoặc không chứa các nucleotit không tự nhiên, và là ADN hoặc ARN sợi kép, ADN hoặc ARN sợi đơn, hoặc các sản phẩm phiên mã của các ADN nêu trên.

Cũng cần nêu ở đây rằng sáng chế không đề cập đến các trình tự nucleotit trong môi trường nhiễm sắc thể tự nhiên của chúng, nghĩa là ở trạng thái tự nhiên. Các trình tự của sáng chế đã được phân lập và/hoặc tinh chế, nghĩa là chúng được lấy mẫu theo cách trực tiếp hoặc gián tiếp, ví dụ, bằng các bản sao, môi trường của chúng đã được cải biến ít nhất một phần. Do đó, các axit nucleic phân lập thu được bằng phương pháp di truyền tái tổ hợp, ví dụ, của các tế bào chủ, hoặc thu được bằng cách tổng hợp hóa học cũng cần được đề cập ở đây.

Phân tử axit nucleic "được phân lập" là phân tử axit nucleic được xác định và tách ra khỏi ít nhất một phân tử axit nucleic- tạp chất mà phân tử này được gắn kết với nguồn tự nhiên của axit nucleaza nucleic. Phân tử axit nucleic được phân lập khác biệt về hình dạng hoặc tập hợp mà trong đó nó được tìm thấy trong các điều kiện tự nhiên. Do đó, phân tử axit nucleic được phân lập khác với phân tử axit nucleic tồn tại trong các tế bào trong điều kiện tự nhiên. Tuy nhiên, phân tử axit nucleic được phân lập bao gồm phân tử axit nucleic nằm trong các tế bào mà trong đó kháng thể này được biểu hiện bình thường, ví dụ, nếu phân tử axit nucleic có sự định vị nhiễm sắc thể khác với sự định vị của nó trong các tế bào trong điều kiện bình thường.

Thuật ngữ "trình tự nucleotit" bao gồm cả thành phần hỗ trợ của nó, trừ khi có quy định khác. Do đó, axit nucleic có trình tự cụ thể cần được hiểu là trình tự bao gồm cả sợi hỗ trợ của nó cùng với trình tự bổ sung của nó.

Thuật ngữ "các trình tự kiểm soát" để chỉ các trình tự ADN cần thiết cho sự biểu hiện của trình tự mã hóa được liên kết khi hoạt động trong sinh vật chủ cụ thể. Các trình tự kiểm soát là thích hợp đối với các sinh vật có nhân nguyên thủy, ví dụ,

bao gồm gen khởi đầu, tùy ý là trình tự gen điều khiển, và vị trí gắn kết ribosom. Các tế bào có nhân điển hình đã được biết là sử dụng các gen khởi đầu, tín hiệu polyadenyl hóa, và gen tăng cường.

Axit nucleic được "liên kết khi hoạt động" khi nó được đưa vào mối quan hệ về mặt chức năng với trình tự axit nucleic khác. Ví dụ, ADN của tiền trình tự hoặc trình tự dẫn đầu tiết được liên kết khi hoạt động với ADN của polypeptit nếu nó được biểu hiện dưới dạng tiền protein tham gia vào sự tiết polypeptit; gen khởi đầu hoặc gen tăng cường được liên kết khi hoạt động với trình tự mã hóa nếu nó ảnh hưởng đến sự phiên mã trình tự; vị trí gắn kết ribosom được liên kết khi hoạt động với trình tự mã hóa nếu nó nằm ở vị trí sao cho tạo điều kiện thuận lợi cho sự dịch mã. Nói chung, thuật ngữ "liên kết khi hoạt động" có nghĩa là các trình tự ADN được liên kết ở gần nhau, và trong trường hợp trình tự dẫn đầu tiết, ở gần nhau và trong pha đọc. Tuy nhiên, gen tăng cường không nhất thiết phải ở gần nhau.

Vector

Khi được sử dụng trong bản mô tả này, thuật ngữ "vector" có nghĩa là phân tử axit nucleic có khả năng vận chuyển axit nucleic khác mà nó đã được liên kết với. Theo một số phương án, vector là plasmid, nghĩa là mảnh ADN sợi kép mạch vòng mà các đoạn ADN bổ sung có thể được gắn với. Theo một số phương án, vector là vector của virus, trong đó các đoạn ADN bổ sung có thể được gắn với hệ gen của virus. Theo một số phương án, các vector có khả năng tự sao chép trong tế bào chủ mà chúng được đưa vào (ví dụ, vector của vi khuẩn có vị trí sao chép có nguồn gốc từ vi khuẩn và các vector của động vật có vú thể bổ sung). Theo các phương án khác, vector (ví dụ, các vector không phải của động vật có vú thể bổ sung) có thể được đính gắn vào hệ gen của tế bào chủ khi đưa vào tế bào chủ, và nhờ đó được sao chép cùng với gen của vật chủ. Ngoài ra, một số vector có khả năng định hướng sự biểu hiện của các gen mà chúng được liên kết với khi hoạt động. Các vector này được gọi ở đây là "vector biểu hiện tái tổ hợp" (hoặc đơn giản là "vector biểu hiện").

Theo một khía cạnh, sáng chế đề cập đến vector thích hợp cho sự biểu hiện của trình tự bất kỳ trong số các trình tự nucleotit được mô tả trong bản mô tả này.

Sáng chế đề cập đến các vector chứa phân tử axit nucleic mã hóa enzym PaCas9 nucleaza.

Theo một số phương án, enzym PaCas9 nucleaza của sáng chế được biểu hiện

bằng cách chèn ADN vào các vector biểu hiện sao cho các gen này được liên kết về mặt chức năng với các trình tự kiểm soát sự biểu hiện cần thiết, như các trình tự kiểm soát sự dịch mã và sự phiên mã. Các vector biểu hiện bao gồm các plasmid, retrovirus, adenovirus, virus kết hợp adeno (adeno-associated virus: AAV), các virus ở thực vật như virus bệnh đốm ở cây súp lơ, virus bệnh đốm ở cây thuốc lá, cosmit, YAC, episom có nguồn gốc từ virus EBV, và dạng tương tự. Các phân tử ADN có thể được gắn vào vector sao cho các trình tự kiểm soát sự phiên mã và sự dịch mã trong vector này thực hiện chức năng dự định của chúng là điều hòa sự phiên mã và sự dịch mã của ADN. Vector biểu hiện và các trình tự kiểm soát sự biểu hiện có thể được chọn để tương hợp được với tế bào chủ biểu hiện được sử dụng. Các phân tử ADN có thể được đưa vào vector biểu hiện bằng các phương pháp tiêu chuẩn (ví dụ, phương pháp gắn các vị trí giới hạn bổ trợ trên mảnh gen PaCas9 nucleaza và vector, hoặc gắn đầu bằng nếu không có mặt các vị trí giới hạn).

Ngoài gen PaCas9 nucleaza, sự biểu hiện vector tái tổ hợp của sáng chế có thể mang trình tự điều hòa để kiểm soát sự biểu hiện của gen PaCas9 nucleaza trong tế bào chủ. Người có hiểu biết trung bình trong lĩnh vực này sẽ hiểu rằng việc thiết kế vector biểu hiện, bao gồm cả sự chọn lọc các trình tự điều hòa, có thể phụ thuộc vào các yếu tố như sự lựa chọn tế bào chủ cần được biến nạp, mức độ biểu hiện của protein mong muốn, v.v.. Các trình tự kiểm soát được ưu tiên để biểu hiện tế bào chủ ở các động vật có vú bao gồm các thành phần của virus để đảm bảo sự biểu hiện ở mức cao của protein trong các tế bào của động vật có vú, như các gen khởi đầu và/hoặc gen tăng cường có nguồn gốc từ LTR của retrovirus, virus cự bào (cytomegalovirus: CMV) (như gen khởi đầu/gen tăng cường của CMV), virus SV40 ở khỉ (simian virus 40: SV40) (như gen khởi đầu/gen tăng cường của SV40), adenovirus, (ví dụ, adenovirus của gen khởi đầu muộn chính (AdMLP)), polyomavirus và các gen khởi đầu của động vật có vú mạnh như gen khởi đầu của globulin tự nhiên hoặc gen khởi đầu actin. Phần mô tả thêm các thành phần kiểm soát của virus và các trình tự của chúng, ví dụ, xem tài liệu: các Bằng độc quyền sáng chế Mỹ số US 5.168.062, 4.510.245 và 4.968.615. Các phương pháp biểu hiện polypeptit trong tế bào vi khuẩn hoặc tế bào nấm, ví dụ, các tế bào nấm men, cũng đã được biết rõ trong lĩnh vực này.

Ngoài gen PaCas9 nucleaza và các trình tự điều hòa, các vector biểu hiện tái tổ hợp của sáng chế có thể mang các trình tự bổ sung, như các trình tự điều hòa sự sao

chép của vector trong các tế bào chủ (ví dụ, nguồn gốc sao chép) và các gen đánh dấu có thể chọn lọc. Gen đánh dấu có thể chọn lọc tạo điều kiện thuận lợi cho sự chọn lọc tế bào chủ mà vector đã được đưa vào đó (ví dụ, xem các Bằng độc quyền sáng chế Mỹ số US 4.399.216, 4.634.665 và 5.179.017). Ví dụ, thường là gen đánh dấu có thể chọn lọc tạo ra sự kháng lại các chất dùng làm thuốc, như G418, hygromycin hoặc methotrexat, trên tế bào chủ mà vector đã được đưa vào. Ví dụ, các gen đánh dấu có thể chọn lọc bao gồm gen dihydrofolat reductaza (dihydrofolate reductase: DHFR) (để sử dụng trong các tế bào chủ trong khi chọn lọc/khuếch đại bằng methotrexat), gen neo (để chọn lọc G418), và gen glutamat synthetaza.

Khi được sử dụng trong bản mô tả này, thuật ngữ "trình tự kiểm soát sự biểu hiện" được dùng để chỉ các trình tự polynucleotit cần thiết để tác động đến sự biểu hiện và xử lý trình tự mã hóa mà chúng được gắn vào. Các trình tự kiểm soát sự biểu hiện bao gồm sự khởi mào, sự kết thúc quá trình phiên mã thích hợp, các trình tự của gen khởi đầu và gen tăng cường; các tín hiệu xử lý ARN hiệu quả như tín hiệu ghép và tín hiệu polyadenyl hóa; các trình tự làm ổn định mRNA trong bào chất; các trình tự làm gia tăng hiệu quả dịch mã (nghĩa là trình tự liên ứng Kozak); các trình tự làm gia tăng tính ổn định của protein; và nếu muốn, các trình tự làm gia tăng sự tiết protein. Tính chất của các trình tự kiểm soát này khác nhau tùy thuộc vào sinh vật chủ; ở các sinh vật có nhân nguyên thủy, các trình tự kiểm soát này thường bao gồm gen khởi đầu của vị trí gắn kết ribosom, và các trình tự kết thúc phiên mã; ở các sinh vật có nhân chuẩn, các trình tự kiểm soát này thường bao gồm gen khởi đầu và trình tự kết thúc phiên mã. Thuật ngữ "trình tự kiểm soát" được dự định bao gồm ít nhất là tất cả các thành phần mà sự có mặt của chúng là thiết yếu cho sự biểu hiện và xử lý, và cũng có thể bao gồm các thành phần bổ sung mà sự có mặt của chúng là có lợi, ví dụ, các trình tự đoạn dẫn đầu và trình tự của đối tác dung hợp.

Tế bào chủ

Khi được sử dụng trong bản mô tả này, thuật ngữ "tế bào chủ tái tổ hợp" (hoặc đơn giản là "tế bào chủ") được dùng để chỉ tế bào mà vector biểu hiện tái tổ hợp đã được đưa vào đó. Sáng chế đề cập đến tế bào chủ, có thể bao gồm, ví dụ, vector theo sáng chế được mô tả trên đây. Cần hiểu rằng "tế bào chủ tái tổ hợp" và "tế bào chủ" được dự định để chỉ không chỉ tế bào của đối tượng cụ thể mà còn để chỉ thế hệ con của tế bào này. Do các cải biến có thể xuất hiện trong các thế hệ tiếp theo do sự đột

biến hoặc sự ảnh hưởng của môi trường, trên thực tế, thể hệ con này có thể không giống hệt với tế bào bố-mẹ, tuy nhiên, các tế bào này vẫn được bao gồm trong phạm vi của thuật ngữ "tế bào chủ" khi được sử dụng trong bản mô tả này.

Các phân tử axit nucleic mã hóa PaCas9 nucleaza của sáng chế và các vector chứa các phân tử axit nucleic này có thể được sử dụng để chuyển nhiễm động vật có vú thích hợp hoặc tế bào của nó, thực vật hoặc tế bào của nó, tế bào chủ của vi khuẩn hoặc nấm men. Quá trình biến nạp có thể được thực hiện bằng kỹ thuật đã biết bất kỳ để đưa các polynucleotit vào tế bào chủ. Các phương pháp để đưa polynucleotit khác loài vào tế bào của động vật có vú là đã được biết rõ trong lĩnh vực này và bao gồm phương pháp chuyển nhiễm qua trung gian dextran, phương pháp chuyển nhiễm bằng phức chất polyme cation-axit nucleic, phương pháp kết tủa bằng canxi phosphat, phương pháp chuyển nhiễm qua trung gian polybren, phương pháp dung hợp tế bào trần, phương pháp bao nang (các) polynucleotit trong liposom, và phương pháp vi tiêm trực tiếp ADN vào nhân. Ngoài ra, các phân tử axit nucleic có thể được đưa vào tế bào của động vật có vú bằng vector của virus. Các phương pháp chuyển nhiễm tế bào là đã được biết rõ trong lĩnh vực này. Ví dụ, xem các Bằng độc quyền sáng chế Mỹ số US 4.399.216, 4.912.040, 4.740.461 và 4.959.455. Các phương pháp biến nạp tế bào thực vật là đã được biết rõ trong lĩnh vực này bao gồm, ví dụ, phương pháp biến nạp qua trung gian vi khuẩn *Agrobacterium*, phương pháp biến nạp bằng cách bắn gen, phương pháp tiêm trực tiếp, phương pháp gây xung điện và phương pháp biến nạp bằng virus. Các phương pháp biến nạp tế bào vi khuẩn và nấm men cũng đã được biết rõ trong lĩnh vực này.

Các dòng tế bào của động vật có vú được sử dụng làm vật chủ để biến nạp là đã được biết rõ trong lĩnh vực này và bao gồm nhiều dòng tế bào bất tử có sẵn. Các dòng tế bào này bao gồm, ví dụ, tế bào buồng trứng chuột đồng Trung Quốc (Chinese hamster ovary: CHO), tế bào NS0, tế bào SP2, tế bào HEK-293T, tế bào FreeStyle 293 (Invitrogen), tế bào NIH-3T3, tế bào HeLa, tế bào thận chuột đồng sơ sinh (baby hamster kidney: BHK), tế bào thận khỉ xanh châu Phi (COS), tế bào caxinom tế bào gan của người (ví dụ, Hep G2), tế bào A549, và nhiều dòng tế bào khác. Các dòng tế bào được chọn lọc bằng cách xác định xem các dòng tế bào này có mức độ biểu hiện cao và tạo ra các đặc tính cần thiết của protein được tạo ra hay không. Các dòng tế bào khác có thể được sử dụng là dòng tế bào côn trùng, như các tế bào Sf9 hoặc Sf21. Khi

các vector biểu hiện tái tổ hợp mã hóa PaCas9 nucleaza được đưa vào tế bào chủ của động vật có vú, PaCas9 nucleaza được tạo ra bằng cách nuôi cấy các tế bào chủ trong một khoảng thời gian đủ để cho phép sự biểu hiện của PaCas9 nucleaza trong các tế bào chủ hoặc tốt hơn nữa là sự tiết ra PaCas9 nucleaza vào môi trường nuôi cấy mà trong đó các tế bào chủ được sinh trưởng. Enzym PaCas9 nucleaza có thể được phân lập ra khỏi môi trường nuôi cấy này bằng cách sử dụng các kỹ thuật tinh chế protein tiêu chuẩn. Các tế bào chủ của thực vật bao gồm, ví dụ, giống *Nicotiana*, *Arabidopsis*, bèo tấm, ngô, lúa mì, khoai tây, v.v.. Các tế bào chủ của vi khuẩn bao gồm các loài *Escherichia* và *Streptomyces*. Các tế bào chủ của nấm men bao gồm *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae* và *Pichia pastoris*.

Ngoài ra, mức độ tiết ra PaCas9 nucleaza của sáng chế từ các dòng tế bào có thể được tăng lên bằng cách sử dụng nhiều kỹ thuật đã biết. Ví dụ, hệ thống biểu hiện gen glutamin synthetaza (hệ thống GS) là phương pháp chung để làm gia tăng mức độ biểu hiện trong một số điều kiện. Hệ thống GS được bàn luận trong toàn bộ hoặc một phần nội dung của các Bằng độc quyền sáng chế châu Âu số EP 0216846, 0256055, 0323997 và 0338841.

Enzym PaCas9 nucleaza có thể thu được từ các dòng tế bào khác nhau hoặc từ các động vật chuyển gen sẽ có profin glycosyl hóa khác nhau khi so sánh với nhau. Tuy nhiên, PaCas9 nucleaza được mã hóa bởi các phân tử axit nucleic được mô tả trong bản mô tả này là một phần của sáng chế, bất kể trạng thái glycosyl hóa, và nói chung là bất kể sự có mặt hoặc không có mặt của các cải biến sau dịch mã.

Liposom

Theo một khía cạnh, sáng chế đề cập đến liposom bao nang PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2 hoặc phân tử axit nucleic được phân lập mã hóa PaCas9 nucleaza và có trình tự nucleotit nêu trong SEQ ID NO: 1.

Liposom là các nang kín về mặt tế vi có pha phân tán, được bao quanh bằng một hoặc nhiều lớp lipid, và có khả năng giữ lại chất tan trong nước trong pha phân tán này, và chất tan trong dầu trong hai lớp phospholipit. Khi gắn hợp chất có hoạt tính trong liposom, và vận chuyển nó đến mô đích, cách gắn hợp chất có hoạt tính trong liposom với hiệu quả cao, và cách đảm bảo sự duy trì ổn định hợp chất có hoạt tính bằng liposom là các vấn đề quan trọng.

Nói chung, liposom được cho là tiểu phần có kích thước chủ yếu từ vài chục

nanomet đến tối đa là một phần mười micron, vỏ của nó chứa phân tử của (các) chất khác. Vỏ liposom có "tính bán thấm" với các phân tử nước và ion.

Liposom được đặc trưng bởi khả năng chứa và giữ lại các chất có tính chất khác nhau. Các loại chất được đưa vào liposom là rất đa dạng, từ các ion vô cơ và các hợp chất hữu cơ có trọng lượng phân tử thấp đến các protein và axit nucleic có kích thước lớn.

Liposom tạo ra sự giải phóng kéo dài của chất được kết hợp trong chất mang.

Liposom có thể được tạo ra từ phospholipit, cụ thể là từ phosphatidylcholin, phosphatidyletanolamin, phosphatidylserin, phosphatidylinositol, phosphatidylglyxerol, axit phosphatidic, sphingophospholipit, phospholipit từ trứng/đậu tương hoặc các hỗn hợp của chúng.

Ví dụ thực hiện sáng chế

Các ví dụ sau đây được đưa ra để hiểu rõ hơn sáng chế. Các ví dụ này được đưa ra chỉ nhằm mục đích minh họa và không được hiểu là làm giới hạn phạm vi của sáng chế theo cách bất kỳ.

Tất cả các tài liệu công bố, bằng độc quyền sáng chế, và đơn đăng ký sáng chế đã viện dẫn trong bản mô tả này được kết hợp vào đây để tham khảo. Mặc dù sáng chế nêu trên đã được mô tả trong một số chi tiết để minh họa và ví dụ nhằm mục đích hiểu rõ ràng, người có hiểu biết trung bình trong lĩnh vực này sẽ dễ dàng hiểu rõ sáng chế dựa vào phần bộc lộ của nó, và có thể tiến hành một số thay đổi và cải biến mà vẫn nằm trong phạm vi của các phương án kèm theo.

Vật liệu và phương pháp chung

Kỹ thuật ADN tái tổ hợp

Các phương pháp tiêu chuẩn được sử dụng để thao tác ADN như được mô tả trong tài liệu: Sambrook, J. et al, Molecular cloning: A laboratory manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. Các chất phản ứng sinh học phân tử được sử dụng theo các hướng dẫn của nhà sản xuất.

Tổng hợp gen

Tiến hành tạo ra các đoạn gen mong muốn từ oligonucleotit bằng phương pháp tổng hợp hóa học. Các đoạn gen có chiều dài 300-4000 kb, nằm ở vùng rìa của các vị

trí giới hạn đơn lẻ, được lắp ráp bằng cách ghép và gắn các oligonucleotit bao gồm cả khuếch đại bằng phản ứng PCR và sau đó được tách dòng bằng các vị trí giới hạn được chỉ định. Các trình tự ADN của các mảnh gen đã tách dòng phụ được khẳng định bằng cách giải trình tự ADN.

Xác định trình tự ADN

Các trình tự ADN được xác định bằng phương pháp giải trình tự Sanger.

Phân tích trình tự ADN và protein và quản lý số liệu trình tự

Phần mềm Vector NTI Advance suite, phiên bản 8.0 của Infomax được sử dụng để tạo ra, vẽ bản đồ, phân tích, giải thích và minh họa trình tự.

Vector biểu hiện

Để biểu hiện PaCas9 nucleaza, biến thể của các plasmid biểu hiện dùng để biểu hiện trong các tế bào của sinh vật có nhân nguyên thủy (*E.coli*), sự biểu hiện tạm thời trong các tế bào có nhân điển hình (ví dụ, trong các tế bào CHO) được sử dụng. Ngoài cat xet biểu hiện PaCas9 nucleaza, các vector chứa: nguồn gốc sao chép để cho phép sao chép plasmid nêu trên trong *E. coli*, các gen tạo ra tính kháng trong *E. coli* với các chất kháng sinh khác nhau (ví dụ, với ampicillin và/hoặc kanamycin).

Ví dụ 1

Phương pháp tạo ra enzym PaCas9 nucleaza

Để tạo ra các trình tự hệ gen meta, các mẫu bọt biển *Homoeodictya palmata* được thu thập từ vùng Biển Trắng, nguyên liệu này được cắt phân đoạn bằng cách ly tâm, sau đó ADN toàn phần được phân lập và tiếp đó được giải trình tự.

Khung đọc mở của protein PaCas9, cũng như các thành phần liên kế của hệ CRISPR-Cas (bằng CRISPR, cũng như các trình tự crARN và tracrARN) được phát hiện trong các trình tự của hệ gen meta bằng cách sử dụng các phương pháp tin sinh học.

ADN của PaCas9 nucleaza được thể hiện trong SEQ ID NO:1.

Trình tự axit amin của PaCas9 nucleaza được thể hiện trong SEQ ID NO: 2.

Trình tự nucleotit mã hóa tracrARN được thể hiện trong SEQ ID NO: 3.

Trình tự nucleotit mã hóa đoạn lặp xuôi DR được thể hiện trong SEQ ID NO:4.

Ví dụ 2

Phát hiện sự tách dòng

Trình tự của gen PaCas9 nucleaza thu được bằng phương pháp tin sinh học. Trình tự này được tối ưu hóa codon để đảm bảo sự biểu hiện tối ưu trong các tế bào của động vật có vú, và sau đó được lắp ráp *de novo* từ các oligonucleotit được tổng hợp bằng phương pháp hóa học bằng cách sử dụng phương pháp Gibson. Gen PaCas9 tổng hợp được được tách dòng trong cấu trúc di truyền từ đầu 3' của gen khởi đầu CMV. Trình tự Kozak và tín hiệu khu trú trong nhân (NLS) được bổ sung từ đầu 5' của gen này, và trình tự epitop FLAG để phát hiện protein được bổ sung từ đầu 3'. Sau khi trình tự PaCas9 và các thành phần liên quan của nó như đã liệt kê trên đây, các thành phần T2A và khung đọc mở của protein phát huỳnh quang màu xanh lục (EGFP) làm gen đánh dấu sự biểu hiện được đưa vào cấu trúc trong cùng khung đọc.

Sau các khung đọc, trình tự tín hiệu của poly-A thymidin kinaza được đưa vào từ đầu 3' để làm gia tăng tính ổn định của mARN. Có hai băng nối tiếp trong khu vực vỏ vi khuẩn của cấu trúc di truyền dùng cho sự biểu hiện của các phân tử ARN nhỏ. Mỗi băng chứa gen khởi đầu U6 và gen kết thúc phiên mã ARN polymeraza III. Các băng này là cần thiết cho sự biểu hiện của các phân tử ARN để tạo ra sự tương tác đặc hiệu của protein PaCas9 với phân tử ADN đích (hệ gen của tế bào). Bản đồ cấu trúc được thể hiện trên Fig.1. Cấu trúc này cho phép biểu hiện cả protein PaCas9 (được vận chuyển đến nhân bằng tín hiệu NLS) và các phân tử ARN dẫn đường protein (ARN dẫn đường), cũng như phát hiện protein bằng epitop FLAG và xác định hiệu quả của sự vận chuyển cấu trúc di truyền bằng cách phát hiện EGFP.

Ví dụ 3

Hoạt tính enzym của protein PaCas9

Các axit amin tham gia vào quá trình thủy phân bằng enzym của ADN/ARN được xác định bằng cách so sánh tính tương đồng của các miền HNH và RuvC của các protein họ Cas9 khác nhau với các miền của PaCas9 (sự phân bố theo miền được thể hiện trên Fig.2). Các axit amin bảo toàn, có liên quan đến hoạt tính enzym của các protein Cas9 đã được chứng minh trước đó, được phân lập trong PaCas9. Do đó, bằng các phương pháp phân tích, đã phát hiện được rằng các gốc axit amin của protein này là cần thiết cho hoạt tính enzym của protein PaCas9 (axit amin – vị trí): D 9; E 527; H 750; D 753; H 613; N 636.

Ví dụ 4

Xác định hoạt tính enzym của protein PaCas9

Để xác định trình tự của motif liền kề đoạn tiền đệm (Protospacer Adjacent Motif: PAM), các tác giả sáng chế tiến hành các phản ứng *in vitro* để cắt ngắn thư viện ADN bằng cách sử dụng protein nucleaza tái tổ hợp (SEQ ID NO: 2), crARN (gồm phần biến đổi phụ thuộc đích và đoạn lặp xuôi được thể hiện trong trình tự SEQ ID NO:4) và tracrARN (SEQ ID NO:3). Thư viện ADN là mảnh PCR chứa trình tự được chọn ngẫu nhiên có bảy chữ cái, và trình tự có thể nhận biết, đoạn tiền đệm.

Sau khi ủ phức hợp PaCas9-ARN-protein với thư viện ADN, các sản phẩm của phản ứng được tải vào thiết bị điện di trên gel. Các mảnh không bị cắt ngắn được chiết ra khỏi gel và giải trình tự trên thiết bị Illumina. So sánh các trình tự PAM có mặt trong các sản phẩm của phản ứng PaCas9 không bị cắt ngắn và phản ứng đối chứng sẽ cho phép xác định PAM của protein quan tâm.

Sau khi nhận biết trình tự PAM, hoạt tính của nucleaza *in vitro* được đánh giá. Nhằm mục đích này, protein trong phức chất với ARN dẫn đường được ủ với mảnh ADN mang trình tự của đoạn tiền đệm và PAM được nhận biết. Tỷ lệ tối ưu của phức chất ARN-protein với ADN đã cắt ngắn được xác định. Hoạt tính của nucleaza được đánh giá dựa trên lượng protein PaCas9 cần thiết để cắt ngắn 50% 200ng ADN đích có chiều dài khoảng 400 cặp bazơ, chứa PAM tối ưu.

Do đó, đã khẳng định được rằng PaCas9 nucleaza có hoạt tính enzym và gây ra sự đứt gãy sợi kép trong ADN.

Ngoài ra, đã khẳng định được rằng PaCas9 nucleaza có thể gây ra sự đứt gãy sợi kép trong ADN với vị trí nhận biết đặc hiệu ở mức cao (từ 16 đến 20 chữ cái).

YÊU CẦU BẢO HỘ

1. Enzym PaCas9 nucleaza có trình tự axit amin nêu trong SEQ ID NO: 2.
2. Phân tử axit nucleic được phân lập mã hóa enzym PaCas9 nucleaza theo điểm 1 và có trình tự nucleotit nêu trong SEQ ID NO: 1.
3. Vector biểu hiện chứa axit nucleic theo điểm 2.
4. Vector biểu hiện theo điểm 3, trong đó vector này có cấu trúc di truyền được thể hiện trên Fig.1.
5. Vector để vận chuyển chất điều trị chứa axit nucleic theo điểm 2.
6. Vector theo điểm 5, trong đó chất điều trị được vận chuyển đến các tế bào đích hoặc mô đích.
7. Phương pháp vận chuyển chất điều trị đến tế bào đích hoặc mô đích bằng cách sử dụng vector theo điểm bất kỳ trong số các điểm từ 3 đến 6.
8. Phương pháp tạo ra tế bào chủ để tiết ra enzym PaCas9 nucleaza theo điểm 1, bao gồm bước biến nạp tế bào này với vector theo điểm bất kỳ trong số các điểm từ 3 đến 6.
9. Tế bào chủ tạo ra enzym PaCas9 nucleaza theo điểm 1, trong đó tế bào này chứa axit nucleic theo điểm 2.
10. Phương pháp tạo ra enzym PaCas9 nucleaza theo điểm 1, bao gồm bước nuôi cấy tế bào chủ theo điểm 9 trong môi trường sinh trưởng trong điều kiện đủ để tiết ra enzym PaCas9 nucleaza nêu trên, nếu cần, sau đó phân lập và tinh chế enzym PaCas9 nucleaza thu được.

DANH MỤC TRÌNH TỰ

<110> JOINT STOCK COMPANY "BIOCAD"

<120> ENZYM PACAS9 NUCLEAZA, PHƯƠNG PHÁP TẠO RA VÀ PHÂN TỬ AXIT NUCLEIC
ĐƯỢC PHÂN LẬP MÃ HÓA ENZYM NÀY, VECTƠ BIỂU HIỆN VÀ TẾ BÀO CHỦ
CHỨA AXIT NÀY

<140> RU2018132816

<141> 2018-09-14

<160> 4

<170> BiSSAP 1.3.6

<210> 1

<211> 3135

<212> ADN

<213> Homoeodictya palmata

<400> 1

```

atggcaaaaa cacggcttgg tattgatatg gggactaatt ctatcggttg gatgttgtat      60
gaattggaca aaaacggtga aatttcggaa gttttaaaagg caggggtgcg cattttcctt      120
gatgggcgcg aggataaaac ccaagcatcc aaaaacgcgg cacggcgggt agcacgaatg      180
aatcgccgtc agcgagaccg ctatcttcag cgccggacag cgattttacg ctatttggtg      240
aaatttgggc tgatgccaga agataaaaca gaacaacgaa aattgcagga tattgacccc      300
tatagtattc gtgccaaggc aatggaagaa gaaattccac cccatcatct tgggcgggca      360
atttttcata tcagccaaag acgtggctat aagagtagcc gtcgtaatga agaaaacgat      420
aaagatggtc cagtaaaaag ttccattgaa gaattccggc ggagctggg cgacaaaagt      480
gtgggacaat ttttatctga attacaccaa gaaaataaac caatccgcgc ccgccgtgat      540
ggcgtaacta ataatgatct ttttcattat ttcccagacc gcgaactaat tgaaaaagag      600
tttaacgata tttggcaaaa acaacaacaa atccgctcgc aacaaccaga taaaaataaa      660
caaatTTTTat ctgatataatt aacaaatgaa aataaacaaa ctttatttga agttattttt      720
catcagcgct ctttaaagcc gccaatatc ggcaattgtc aattttttcc gaccaaaaaa      780
agaatcgcta aggcgttgcc ttcctttcag cgattccgta ttttgcagga tattaataac      840
cttgaatatt ttgatgaaaa cgagtggcac cttttgccgt cgtctattcg tgattgggca      900
ttgcacattc tttttgcggg aggtaatcta acttttaaga agttaaatag ccaaatgaag      960
aaagatgggtg aaattaacga atctgcattt ttttaacttag aagatgaaaa acgcaatgat      1020
attaaaggcg actttaccac caaaaccatg aaagaaatta tccccgcatt gtgggataac      1080
tgggatTTTgc acaacaaga cagcatgatt ttattactaa taggagataa aagtatagat      1140

```

gaagataaaa tgtttagatga agataaaatg ttagatgaac tttcgtctca ttataattta	1200
tctgaagaag aggcaacaaga atgtgcgaat gccaatattg acaacaggca aggcgttagt	1260
ggaagagcga gtttatcgct agaagcgata agcatactaa tacctatatt ggaaaaagga	1320
aaccactatg ataaagcact tgctgagtcg ggaattgaaa aagataagtc ctctaaacat	1380
gatgaccatt tctatatgca gccctacccc gaaatattag ggcaatgggtg tttgccgcgc	1440
aaaagcgagg atgaaagcaa caaagaattg tggcgtatcc cgaatccaac agtgcattgt	1500
gcattaaatc aattgcgggc agttgttaac gactgcataa gaatcaatgg cggagaaaaa	1560
ccttcacaag ttgttggtga gttggcacgt gacttgccaa taggtattgc aacacggcaa	1620
gaaattagaa aaaaacagtc agaaaaatcaa agggcacgaa cacaacgccg caataaaatt	1680
gaagaactcg gagagcgtgc gactgctaaa tctatgttgc gcatgcaatt gtgggaagaa	1740
tcgtctccca aaaaatgcaa taatcactgt tgcctatatt gtggaaaaca aataggctgc	1800
gctgccgtg tctcttctcc cgattttgag attgaccata ttctaccttt ttctaaaact	1860
ttggatgata gcgccgcaa caaaacgctt tgctgcgtcc aatgcaaccg cgaaaaagga	1920
aataaaactc cttatgaagc atggggcagt aatgaaaaac ggtatgatga aattaaaacg	1980
agagcaagtg ccctatcgcc aaaaaaacga cggcgTTTTT tgccagacgc aatgaaacat	2040
tttgacggtg ataacgattt tctcgtcgc caactaacgc atacgcaata tatcgccaaa	2100
gtgacgaagc gttattttga atctattata aaacccaatg atgtttatgt tattcctggt	2160
cgtttaactg aaatattaag acgcaagtgg ggattgaata atattttgaa tgatgacgga	2220
cacaaaaacc gtgatgacca tagacatcat gccgttgatg cagcagtcac tggcgcgacc	2280
actcgttcta cgctccaaaa atttgctact gaggcaggca aagacaactc cgattttacca	2340
aatgtctcta tcaccgcacc aatgaaagta tttcgtgaaa aggtggaaaa gaccgtaaaa	2400
aatattgttg tctcacacaa accagatcgt ttgagagtg gcgcgttaca taatgatact	2460
gcttatgggt tgcctgctga ttataaagaa ggtgctgggt cacaattcgt aagacatcgg	2520
attatgttgt ctgatataac tacaaaattt tctcaaaaag tggttaattc tttcttgcgc	2580
gaagaattga aaacattatg cgactatgca acggataaaa aatctttaga tcaagaatta	2640
aaacgctatg gagaaaaaaa taatattcgc cgtgttttaa tagaagaaaa attgtctgtt	2700
attgctgttt gcgataagga cggcaaacc tataaagggt ataaaagcga cagtaattgg	2760
gcctatgaga tttttgagaa atttcccaat ggaaattggg atggcgaaat tatttccaca	2820
tttaatgcta atcaagaaaa atttacgccc ttatggaaag aaaaaatcc caaggctaaa	2880
cttatcatgc gcttgacaaa aaatgacatg gtagcattag atgacaatgg actacgaaaa	2940

atatgttatg taaaaacatt aagtgctagc aaaattgccca tggttgaaca ttttgatgcc 3000
 actcctggta gaaatccacc tactattata accaaaagtc caaacgaatt tcgcaaaatc 3060
 aatggacgaa aaatccatat tagccctggc gggttagtaa gagacactaa aaaagatggc 3120
 cgatcaaatc attga 3135

<210> 2
 <211> 1044
 <212> PRT
 <213> Homoeodictya palmata

<400> 2
 Met Ala Lys Thr Arg Leu Gly Ile Asp Met Gly Thr Asn Ser Ile Gly
 1 5 10 15
 Trp Met Leu Tyr Glu Leu Asp Lys Asn Gly Glu Ile Ser Glu Val Leu
 20 25 30
 Lys Ala Gly Val Arg Ile Phe Pro Asp Gly Arg Glu Asp Lys Thr Gln
 35 40 45
 Ala Ser Lys Asn Ala Ala Arg Arg Val Ala Arg Met Asn Arg Arg Gln
 50 55 60
 Arg Asp Arg Tyr Leu Gln Arg Arg Thr Ala Ile Leu Arg Tyr Leu Val
 65 70 75 80
 Lys Phe Gly Leu Met Pro Glu Asp Lys Thr Glu Gln Arg Lys Leu Gln
 85 90 95
 Asp Ile Asp Pro Tyr Ser Ile Arg Ala Lys Ala Met Glu Glu Ile
 100 105 110
 Pro Pro His Leu Gly Arg Ala Ile Phe His Ile Ser Gln Arg Arg
 115 120 125
 Gly Tyr Lys Ser Ser Arg Arg Asn Glu Glu Asn Asp Lys Asp Gly Pro
 130 135 140
 Val Lys Ser Ser Ile Glu Glu Phe Arg Arg Gln Leu Gly Asp Lys Ser
 145 150 155 160
 Val Gly Gln Phe Leu Ser Glu Leu His Gln Glu Asn Lys Pro Ile Arg
 165 170 175
 Ala Arg Arg Asp Gly Val Thr Asn Asn Asp Leu Phe His Tyr Phe Pro
 180 185 190
 Asp Arg Glu Leu Ile Glu Lys Glu Phe Asn Asp Ile Trp Gln Lys Gln
 195 200 205
 Gln Gln Ile Arg Ser Gln Gln Pro Asp Lys Asn Lys Gln Ile Leu Ser
 210 215 220
 Asp Ile Leu Thr Asn Glu Asn Lys Gln Thr Leu Phe Glu Val Ile Phe
 225 230 235 240
 His Gln Arg Ser Leu Lys Pro Pro Ile Ile Gly Asn Cys Gln Phe Phe
 245 250 255
 Pro Thr Lys Lys Arg Ile Ala Lys Ala Leu Pro Ser Phe Gln Arg Phe
 260 265 270
 Arg Ile Leu Gln Asp Ile Asn Asn Leu Glu Tyr Phe Asp Glu Asn Glu
 275 280 285
 Trp His Pro Leu Pro Ser Ser Ile Arg Asp Trp Ala Leu His Ile Leu
 290 295 300
 Phe Ala Gly Gly Asn Leu Thr Phe Lys Lys Leu Asn Ser Gln Met Lys
 305 310 315 320
 Lys Asp Gly Glu Ile Asn Glu Ser Ala Phe Phe Asn Leu Glu Asp Glu

37899

				325					330					335		
Lys	Arg	Asn	Asp	Ile	Lys	Gly	Asp	Phe	Thr	Thr	Lys	Thr	Met	Lys	Glu	
			340					345					350			
Ile	Ile	Pro	Ala	Leu	Trp	Asp	Asn	Trp	Asp	Leu	His	Lys	Gln	Asp	Ser	
		355					360					365				
Met	Ile	Leu	Leu	Leu	Ile	Gly	Asp	Lys	Ser	Ile	Asp	Glu	Asp	Lys	Met	
	370					375					380					
Leu	Asp	Glu	Asp	Lys	Met	Leu	Asp	Glu	Leu	Ser	Ser	His	Tyr	Asn	Leu	
385				390						395					400	
Ser	Glu	Glu	Glu	Ala	Gln	Glu	Cys	Ala	Asn	Ala	Asn	Ile	Asp	Asn	Arg	
				405					410						415	
Gln	Gly	Val	Ser	Gly	Arg	Ala	Ser	Leu	Ser	Leu	Glu	Ala	Ile	Ser	Ile	
			420					425					430			
Leu	Ile	Pro	Tyr	Leu	Glu	Lys	Gly	Asn	His	Tyr	Asp	Lys	Ala	Leu	Ala	
		435					440					445				
Glu	Ser	Gly	Ile	Glu	Lys	Asp	Lys	Ser	Ser	Lys	His	Asp	Asp	His	Phe	
	450					455					460					
Tyr	Met	Gln	Pro	Tyr	Pro	Glu	Ile	Leu	Gly	Gln	Trp	Cys	Leu	Pro	Arg	
465					470					475					480	
Lys	Ser	Glu	Asp	Glu	Ser	Asn	Lys	Glu	Leu	Trp	Arg	Ile	Pro	Asn	Pro	
				485					490					495		
Thr	Val	His	Val	Ala	Leu	Asn	Gln	Leu	Arg	Ala	Val	Val	Asn	Asp	Cys	
			500					505					510			
Ile	Arg	Ile	Asn	Gly	Gly	Glu	Lys	Pro	Ser	Gln	Val	Val	Val	Glu	Leu	
		515					520					525				
Ala	Arg	Asp	Leu	Pro	Ile	Gly	Ile	Ala	Thr	Arg	Gln	Glu	Ile	Arg	Lys	
	530					535				540						
Lys	Gln	Ser	Glu	Asn	Gln	Arg	Ala	Arg	Thr	Gln	Arg	Arg	Asn	Lys	Ile	
545					550					555					560	
Glu	Glu	Leu	Gly	Glu	Arg	Ala	Ser	Ala	Lys	Ser	Met	Leu	Arg	Met	Gln	
				565					570					575		
Leu	Trp	Glu	Glu	Ser	Ser	Pro	Lys	Asn	Ala	Asn	Asn	His	Cys	Cys	Leu	
			580					585				590				
Tyr	Cys	Gly	Lys	Gln	Ile	Gly	Cys	Ala	Ala	Ala	Val	Ser	Ser	Pro	Asp	
		595				600					605					
Phe	Glu	Ile	Asp	His	Ile	Leu	Pro	Phe	Ser	Lys	Thr	Leu	Asp	Asp	Ser	
	610					615				620						
Ala	Ala	Asn	Lys	Thr	Leu	Cys	Cys	Val	Gln	Cys	Asn	Arg	Glu	Lys	Gly	
625					630					635					640	
Asn	Lys	Thr	Pro	Tyr	Glu	Ala	Trp	Gly	Ser	Asn	Glu	Lys	Arg	Tyr	Asp	
				645					650					655		
Glu	Ile	Lys	Thr	Arg	Ala	Ser	Ala	Leu	Ser	Pro	Lys	Lys	Arg	Arg		
			660		</											

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

<210> 3
 <211> 120
 <212> RNA
 <213> Homoeodictya palmata

<400> 3
 tataacaaac aaagaaatgc gggtttattta acattaacat gtgataaatt cgcaacaacc 60
 tttagggcaa cttcggttgc cctattgctt ttaatggcga tataaatatt ttattaatcc 120

<210> 4
 <211> 36
 <212> ADN
 <213> Homoeodictya palmata

<400> 4
 gttgtgttat gaccgcattt ctttgtttgg tataat 36

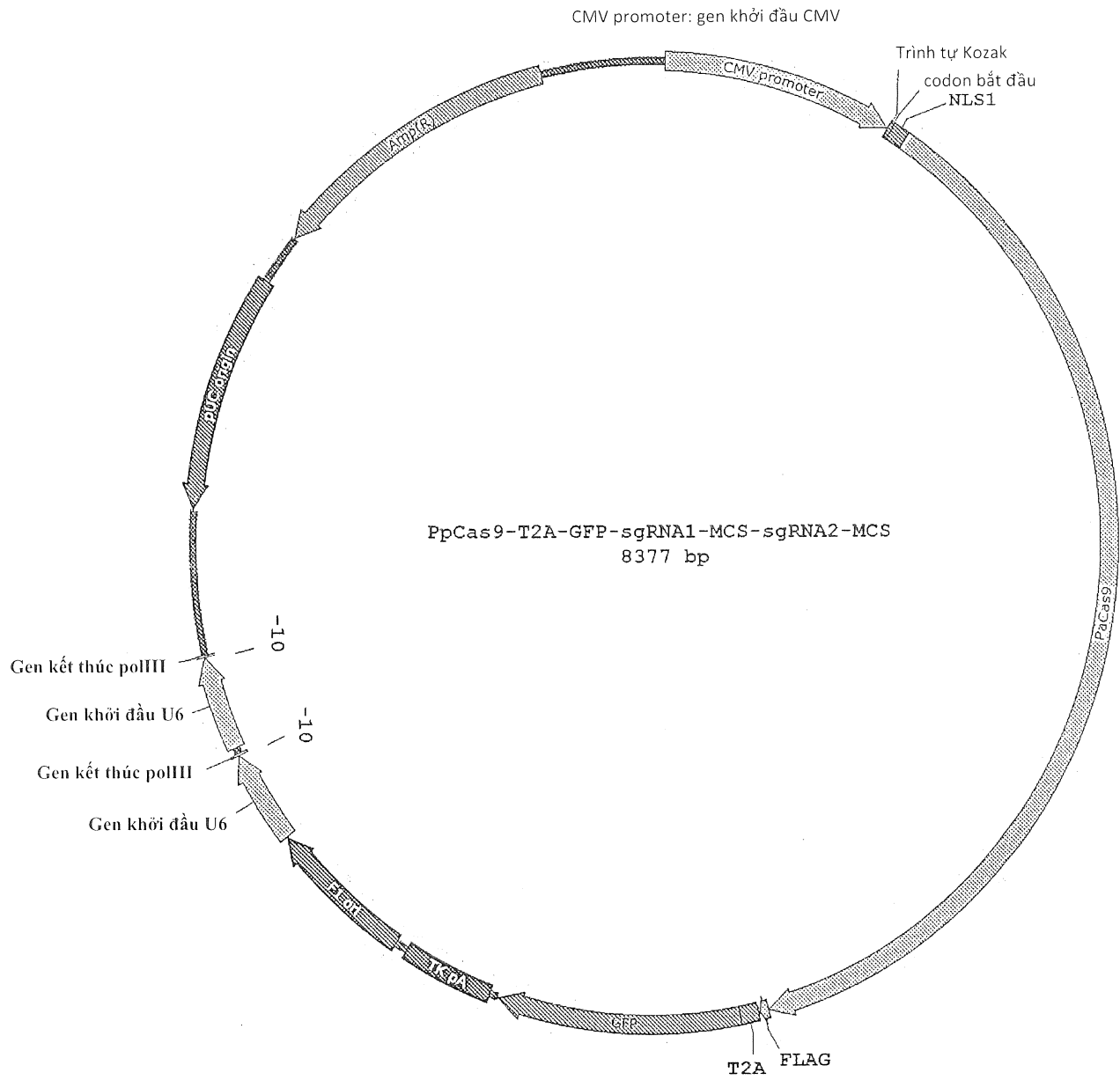


Fig.1

RuvC I
 PaCas9 1 MAKTRLGIDMGTSIGWMLYELDKNGEISEVLKAGVRIFPDGREDKTQASKNAARRVARM
 PaCas9 61 NRRQRDRYLQRRTAILRYLVKFGIMPEDKTEQRKLQDIDPYSIRAKAMEEEIPPHLGRA
 PaCas9 121 IFHISQRRGYKSSRRNEENDKDGVPKSSIEEFRRQLGDKSVGQFLSELHQENKPIRARRD
 PaCas9 181 GVTNNDLFHYFPDRELIEKEFNDIWQKQQQIRSQQPDKNKQILSDILTNNENKQTLFEVIF
 Rec lobe
 PaCas9 241 HQRSCLKPPIIGNCQFFPTKKRIAKALPSFQRFRIQDINNLEYFDENEWHPLPSSIRDWA
 PaCas9 301 LHILFAGGNLTFKKLNSQMKKDGESAFNLEDEKRNDIKGDFTTKTMKEIIPALWDN
 PaCas9 361 WDLHKQDSMILLIGDKSIDEDKMLDEDKMLDELSSHYNLSEEEAQECANANIDNRQGVVS
 PaCas9 421 GRASLSLEAISILIPYLEKGNHYDKALAESGIEKDKSSKHDDHFYMQPYPEILGQWCLPR
 RuvC II
 PaCas9 481 KSEDESNKELWRI PNPTVHVALNQLRAVVND CIRINGGEKPSQVVVELARDLP IGIATRQ
 PaCas9 541 EIRKKQSENQRARTQRRNKIEELGERASAKSMLRMQLWEESSPKNANNHCCLYCGKQIGC
 HNH
 PaCas9 601 AAAVSSPDFEIDHILPFSKTLDDSAANKTLCCVQCNREKGNKTPYEAWSNEKRYDEIKT
 PaCas9 661 RASALSPKKRRRFLPDAMKHFDGDNDFLARQLTDTQYIAKVTKRYFESIIPNDVYVIPG
 RuvC III
 PaCas9 721 RLTEILRRKWGLNNILNDDGHKNRDDHRHHAVDAAVIGATTRSTLQKFATEAGKDNSDLP
 PaCas9 781 NVSITAPMKVFREKVEKTVKNIVVSHKPDRFESGALHNDTAYGLPADYKEGAGAQFVRHR
 PaCas9 841 IMLSDITTNFSQKVNSFLREELKTLCDYATDKKSLDQELKRYGEKNIRRVLIEEKLVS
 PaCas9 901 IAVCDKDGKPYKGYKSDSNWAYEIFEKFPNGNWDGEIISTFNANQEKFTPLWKEKNPKAK
 PaCas9 961 LIMRLHKNDMVALDDNGLRKICYVKTL SASKIAMVEHFDATPGRNPPTIITKSPNEFRKI
 PaCas9 1021 NGRKIHISPGGLVRD TKKDGSRNH

Fig.2