



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

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

(11)



1-0036957

(51)^{2020.01} C12N 15/09; C07K 14/195; C12N 9/22; (13) B
C12N 15/31; C12N 15/55; C12N 15/63;
A01K 67/027; C12N 15/113

(21) 1-2020-01595 (22) 20/08/2018
(86) PCT/JP2018/030607 20/08/2018 (87) WO 2019/039417 28/02/2019
(30) 2017-158876 21/08/2017 JP; 2017-236518 08/12/2017 JP
(45) 25/09/2023 426 (43) 25/06/2020 387ASC
(73) TOKUSHIMA UNIVERSITY (JP)
2-24, Shinkura-cho, Tokushima-shi, Tokushima 7708501 Japan
(72) OSAKABE, Keishi (JP); OSAKABE, Yuriko (JP).
(74) Công ty TNHH một thành viên Sở hữu trí tuệ VCCI (VCCI-IP CO.,LTD)

(54) PHƯƠNG PHÁP HƯỚNG ĐÍCH, PHƯƠNG PHÁP BIẾN ĐỔI TRÌNH TỰ
NUCLEOTIT ĐÍCH, CHẾ PHẨM VÀ KIT HƯỚNG ĐÍCH ĐẾN TRÌNH TỰ
NUCLEOTIT ĐÍCH

(57) Sáng chế đề xuất phương pháp hướng đích trình tự nucleotit đích. Phương pháp này bao gồm bước đưa vào trong tế bào: (i) các protein liên kết CRIPR typ I-D Cas5d, Cas6d, và Cas7d, hoặc các axit nucleic mã hóa cho các protein này; và (ii) ARN dẫn hướng mà bao gồm trình tự bổ trợ cho trình tự nucleotit đích nêu trên và các trình tự lặp chung có nguồn gốc từ locus gen CRISPR trước và sau trình tự bổ trợ, hoặc ADN mã hóa cho ARN dẫn hướng. Sáng chế cũng đề cập đến phức hợp và vật truyền biểu hiện được sử dụng để hướng đích hoặc biến đổi trình tự nucleotit đích và phân tử ADN mã hóa phức hợp này.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp hướng đích trình tự nucleotit đích, phương pháp biến đổi đặc hiệu trình tự nucleotit đích, và phương pháp ức chế sự biểu hiện của gen đích, trong đó sự nhận biết nucleotit đích của hệ CRISPR typ I-D (Clustered Regularly Interspaced Short Palindromic Repeats – nhóm các đoạn ngắn đối xứng lặp lại thường xuyên) được sử dụng, và phức hợp bao gồm các protein Cas (liên kết CRISPR) và ARN dẫn hướng được sử dụng trong các phương pháp này.

Tình trạng kỹ thuật của sáng chế

Các vi khuẩn và archaea có các hệ CRISPR dưới dạng hệ miễn dịch thích ứng đối với các virus và các plasmit lạ dị tương đồng. Hệ CRISPR sử dụng ARN có trọng lượng phân tử thấp (được gọi là ARN dẫn hướng hoặc gARN) mà hỗ trợ cho trình tự ADN xâm lấn để thúc đẩy sự nhắm vào đích và sự suy thoái ADN lạ mong muốn. Đồng thời, protein Cas mà liên kết với gARN để tạo thành phức hợp được yêu cầu. Hệ CRISPR bao gồm các hệ typ I, typ II, typ III, và typ V. Trong hệ bất kỳ, phức hợp protein Cas-gARN tác động lên trình tự đích để can thiệp vào các virus và các plasmit lạ. Ở các hệ typ II và typ V, cơ chế can thiệp bao gồm các sự đứt gãy sợi kép ADN trên ADN đích nhờ protein được đính gắn có vùng protein mà giữ được sự liên kết gARN và vùng protein cắt ADN tương tự RuvC. Đối với hệ typ III, đã có sự giải thích in vitro và in vivo rằng sự can thiệp được gây ra bởi sự cắt chuỗi ARN đích bởi phức hợp gồm năm đến tám protein Cas và gARN, khác với hệ typ II.

Trong những năm gần đây, các kỹ thuật chỉnh sửa hệ gen sử dụng các hệ

CRISPR typ II và typ V đã được phát triển, trong đó Cas9 và Cpf1 được sử dụng dưới dạng protein Cas. Cas9 và Cpf1 cần có trình tự gồm có khoảng 2 đến 5 nucleotit ở vùng lân cận trình tự đích, mà được gọi là trình tự kiểu liên kề đoạn đệm (PAM - protospacer adjacent motif), để nhận biết ADN đích. Đã có sự giải thích in vitro và in vivo rằng phức hợp Cas9-gARN và phức hợp Cpf1-gARN là các endonucleaza dẫn hướng ARN đặc hiệu trình tự mà gây ra các sự đứt gãy sợi kép ADN ở các vị trí đích ở gần các trình tự PAM.

Mặt khác, về hệ CRISPR typ I, nhiều kiểu phụ đã được nhận biết trong các trình tự hệ gen từ các vi khuẩn khác nhau, và các kiểu phụ này đã được đặt tên là typ I-A, I-B, I-C, I-D, I-E, I-F, và I-U. Trong số các kiểu phụ này, hệ typ I-E có nguồn gốc từ *Escherichia coli* đã được nghiên cứu nhiều nhất, và đã được giải thích rằng phức hợp gồm có sáu protein Cas (Cas3, Cse1, Cse2, Cas7, Cas5, Cas6e) và gARN thúc đẩy sự suy thoái của trình tự ADN đích. Tuy nhiên, đối với các kiểu phụ khác trừ kiểu phụ (typ I-C), các thành phần protein Cas, các trình tự gARN, các trình tự PAM mà xác định các ADN đích, v.v. cần thiết cho hiệu quả can thiệp hầu như chưa được làm sáng tỏ. Ngoài ra, về các kỹ thuật sử dụng các protein Cas có nguồn gốc từ hệ CRISPR typ I, phương pháp ức chế sự biểu hiện gen đích mà bao gồm sử dụng các phân tử axit nucleic tái tổ hợp mã hóa cho các protein Cas có nguồn gốc từ hệ CRISPR typ I (tài liệu sáng chế 1), và phương pháp biến đổi axit nucleic đích bao gồm bước sử dụng phức hợp gồm các protein Cas có nguồn gốc từ hệ CRISPR typ I và các protein khác (tài liệu sáng chế 2 và tài liệu sáng chế 3) đã được báo cáo. Tuy nhiên, kỹ thuật cắt và biến đổi sợi kép của phân tử ADN đích bởi endonucleaza được dẫn hướng bởi ARN có nguồn gốc từ hệ CRISPR typ I chưa từng được báo cáo.

Danh sách tài liệu viện dẫn

Tài liệu sáng chế

Tài liệu sáng chế 1: WO2015/155686

Tài liệu sáng chế 2: JP-A 2015-503535

Tài liệu sáng chế 3: WO2017/043573

Bản chất kỹ thuật của sáng chế

Các vấn đề kỹ thuật

Ở các hệ CRISPR typ II và typ V thông thường, phân tử ARN được sử dụng để nhắm vào đích bị giới hạn ở phân tử ARN gồm khoảng 20 nucleotit trước hoặc sau trình tự PAM gồm khoảng 2 đến 5 nucleotit mà xác định tính đặc hiệu đích. Như vậy các hệ CRISPR typ II và typ V thông thường có các vấn đề là có nhiều ổ gen trong đó gen đích không thể tạo ra được và các trình tự tương tự có thể bị cắt. Sự phát triển hệ thống nhằm vào đích mới và endonucleaza được dẫn hướng bởi ARN mới mà không có các vấn đề nêu trên là được mong muốn.

Cách thức giải quyết vấn đề

Để giải quyết các vấn đề nêu trên, các tác giả sáng chế đã nghiên cứu sâu rộng. Nhờ đó, hệ thống nhằm vào đích mới và endonucleaza được dẫn hướng bởi ARN mới mà nhằm vào trình tự dài hơn trình tự đích của endonucleaza được dẫn hướng bởi các ARN của CRISPR typ II hoặc typ V thường được sử dụng trong công nghệ chỉnh sửa hệ gen được phát hiện từ CRISPR typ I-D, và sau đó nhận thấy rằng hệ nhằm vào đích mới và endonucleaza được dẫn hướng bởi ARN có thể được sử dụng trong các kỹ thuật chỉnh sửa hệ gen để cho phép biến đổi trên trình tự nucleotit đích. Như vậy, sáng chế được hoàn thành.

Tức là, sáng chế đề xuất:

[1] Phương pháp hướng đích trình tự nucleotit đích, phương pháp này bao gồm bước đưa vào trong tế bào:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và

các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ, hoặc ADN mã hóa ARN dẫn hướng này;

[2] Phương pháp biến đổi trình tự nucleotit đích, phương pháp này bao gồm bước đưa vào trong tế bào:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, hoặc các axit nucleic mã hóa cho các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ, hoặc ADN mã hóa ARN dẫn hướng này;

[3] Phương pháp ức chế sự biểu hiện gen đích, phương pháp này bao gồm bước đưa vào trong tế bào:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein, và

(ii) ARN dẫn hướng bao gồm trình tự hỗ trợ cho ít nhất là một phần của trình tự gen đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ, hoặc ADN mã hóa ARN dẫn hướng;

[4] Phương pháp theo mục bất kỳ trong số các mục từ [1] đến [3], trong đó ARN dẫn hướng bao gồm trình tự gồm có 20 đến 50 nucleotit mà hỗ trợ cho trình tự nucleotit đích;

[5] Phương pháp theo mục [2] hoặc [4], còn bao gồm bước đưa thể cho polynucleotit vào trong tế bào;

[6] Phương pháp theo mục bất kỳ trong số các mục [2], [4] và [5], trong đó sự biến đổi là sự khuyết đoạn nucleotit, xen đoạn nucleotit, hoặc thế đoạn nucleotit;

[7] Phương pháp theo mục bất kỳ trong số các mục từ [1] đến [6], trong đó Cas5d nhận biết 5'-GTH-3' (H = A, C, hoặc T) là trình tự kiểu liên kế đoạn đệm (PAM);

[8] Phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ;

[9] Phức hợp theo mục [8], còn bao gồm Cas3d và Cas10d;

[10] Phức hợp theo mục [8] hoặc [9], trong đó ARN dẫn hướng bao gồm trình tự gồm có 20 đến 50 nucleotit mà bổ trợ cho trình tự nucleotit đích;

[11] Vật truyền biểu hiện bao gồm:

(i) các axit nucleic mã hóa các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ADN mã hóa ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ;

[12] Vật truyền biểu hiện theo mục [11], còn bao gồm các axit nucleic mã hóa Cas3d và Cas10d;

[13] Phân tử ADN mã hóa cho phức hợp theo mục bất kỳ trong số các mục từ [8] đến [10];

[14] Sử dụng

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ, hoặc ADN mã hóa cho ARN dẫn hướng, để nhằm vào trình tự nucleotit đích;

[15] Sử dụng

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và

Cas10d, hoặc các axit nucleic mã hóa cho các protein, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ, hoặc ADN mã hóa cho ARN dẫn hướng, để biến đổi trình tự nucleotit đích;

[16] Sử dụng

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho ít nhất là một phần của trình tự gen đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ, hoặc ADN mã hóa cho ARN dẫn hướng, để ức chế sự biểu hiện của gen đích;

[17] Sử dụng theo mục bất kỳ trong số các mục từ [14] đến [16], trong đó ARN dẫn hướng bao gồm trình tự gồm có 20 đến 50 nucleotit bổ trợ cho trình tự nucleotit đích;

[18] Sử dụng theo mục [15] hoặc [17], trong đó sự biến đổi là sự khuyết đoạn nucleotit, xen đoạn nucleotit, hoặc thế đoạn nucleotit;

[19] Sử dụng theo mục bất kỳ trong số các mục từ [14] đến [18], trong đó Cas5d nhận biết 5'-GTH-3' (H = A, C, hoặc T) là trình tự kiểu liên kết đoạn đệm (PAM);

[20] Sử dụng phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ,

để nhắm vào trình tự nucleotit đích;

[21] Sử dụng phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ, để biến đổi trình tự nucleotit đích;

[22] Sử dụng phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho ít nhất là một phần của trình tự gen đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ, để ức chế sự biểu hiện của gen đích; và

[23] Sử dụng theo mục bất kỳ trong số các mục từ [20] đến [22], trong đó ARN dẫn hướng bao gồm trình tự gồm có 20 đến 50 nucleotit bổ trợ cho trình tự nucleotit đích.

Hiệu quả của sáng chế

Các trình tự PAM của hệ CRISPR typ I-D (sau đây, còn được gọi là “TiD”) khác với các trình tự PAM của hệ CRISPR typ II và hệ typ V. Do vậy, theo sáng chế, việc sử dụng các protein Cas của CRISPR typ I-D làm cho có thể nhắm vào các locus mà không thể nhắm vào được bằng các kỹ thuật chỉnh sửa hệ gen thông thường sử dụng endonucleaza được dẫn hướng bởi các ARN của CRISPR typ II hoặc typ V. Hơn nữa, các trình tự PAM của endonucleaza được dẫn hướng bởi ARN có nguồn gốc từ CRISPR typ I-D theo sáng chế được phát hiện thấy thường xuyên hơn trên các trình tự hệ gen của vài tổ chức khác các trình tự PAM của CRISPR typ II và typ V. Do vậy, theo sáng chế, có thể nhắm vào một số lượng các trình tự gen lớn hơn so với các kỹ thuật chỉnh sửa hệ gen thông thường sử dụng các hệ CRISPR typ II và typ V. Hơn nữa, các tác giả sáng

chế nhận thấy rằng gARN trong hệ CRISPR typ I-D có thể nhằm vào trình tự đích có độ dài là 30 nucleotit hoặc hơn nữa. Mặt khác, gARN trong hệ CRISPR typ II hoặc typ V có thể nhằm vào trình tự có độ dài khoảng 20 nucleotit. Như vậy, hệ CRISPR typ I-D theo sáng chế thể hiện các tính chất liên kết và tính đặc hiệu đích ổn định hơn so với các kỹ thuật thông thường.

Như vậy, theo sáng chế, có thể thực hiện sự tạo ra các alen đột biến, kiểm soát sự biểu hiện gen nhờ sự bất hoạt và kích hoạt quá trình phiên mã, và sự biến đổi hệ gen biểu sinh bằng cách nhằm vào vùng protein cải biến ADN/cải biến histon, trên các vùng gen mà không thể nhằm vào được bởi các kỹ thuật thông thường.

Mô tả vắn tắt các hình vẽ

FIG. 1 thể hiện thành phần của hệ CRISPR typ I-D theo sáng chế và các cơ chế nhằm vào đích và cắt trên trình tự đích.

FIG.2 thể hiện vật truyền biểu hiện TiD cho sự chỉnh sửa hệ gen E. coli: a) cấu trúc của plasmit pEcTiD2; b) cấu trúc của plasmit pEcTiD3, Pro: đoạn khởi đầu tổng hợp J23108, t1: trình tự đoạn kết thúc STOP767, RBS: trình tự liên kết ribosom, t2: trình tự đoạn kết thúc STOP768 (1), t3: trình tự đoạn kết thúc TOP768 (2), t7: trình tự đoạn kết thúc T7, 7d: Cas7d có nguồn gốc từ *Microcystis aeruginosa* (sau đây, được viết tắt là "Ma"), 6d: MaCas6d, 5d: MaCas5d, 3d: MaCas3d, 10d: MaCas10d, T7 pro: đoạn khởi đầu T7, crARN: trình tự lặp CRISPR có nguồn gốc từ TiD, Cm: trình tự kháng cloramphenicol, p15A ori: điểm khởi đầu sao chép có nguồn gốc từ plasmit p15A.

FIG. 3 thể hiện các cấu trúc của thư viện plasmit PAM-pMW_ccdB và pMW_ccdB: a) cấu trúc của pMW_ccdB, t2: trình tự đoạn kết thúc rrnB2, t1: trình tự đoạn kết thúc rrnB1, PAM: trình tự kiểu liên kế đoạn đệm, T7 pro: đoạn khởi đầu T7, ccdB: gen ccdB, Km: gen kháng kanamycin, pSC101 ori: điểm khởi đầu sao chép có nguồn gốc từ plasmit pSC101; b) trình tự đích của thư viện

plasmid pMW_ccdB-PAM, trong đó 4 nucleotit ngẫu nhiên được xen vào trong vị trí NNNN để thu thập plasmid tách trình tự PAM, vùng được đóng khung biểu thị đoạn khởi đầu T7, trình tự được gạch chân biểu thị trình tự đích TiD, và các chữ viết hoa biểu thị locus ccdB.

FIG. 4 thể hiện vật truyền biểu hiện TiD để chỉnh sửa hệ gen thực vật: a) cấu trúc của plasmid pEgPTiD1; b) cấu trúc của cassette biểu hiện crARN cho các thực vật; c) cấu trúc của plasmid pEgPTiD2, RB: trình tự mép bên phải, LB: trình tự mép bên trái, 2x35S: đoạn khởi đầu gen 35S của virus thể khảm của cải hoa lơ 2x và trình tự gen tăng cường sự dịch mã Ω , 3d: MaCas3d với trình tự mã hóa 2xNLS (tín hiệu định vị nhân), 10d: MaCas10d với 2xNLS, 7d: MaCas7d với 2xNLS, 6d: MaCas6d với 2xNLS, 5d: MaCas5d với 2xNLS, 2A (1)-(4): các trình tự peptit tự cắt 2A (1)-(4), Ter: đoạn kết thúc gen có kích cỡ 18,2 kDa của protein sốc nhiệt *Arabidopsis*, Km: cassette biểu hiện gen kháng kanamycin, U6-26: đoạn khởi đầu gen *Arabidopsis* U6 snRNA-26, crARN: các trình tự lặp CRISPR có nguồn gốc từ locus TiD.

FIG. 5-1 thể hiện sự phát sinh đột biến của gen PDS của cây thuốc lá sử dụng pEgPTiD2-pds: a) các trình tự đích trên gen PDS của cây thuốc lá, trong đó trình tự đích 1 được chọn từ exon thứ ba và trình tự đích 2 được chọn từ exon thứ sáu, các phần được đóng khung trong các trình tự đích được thể hiện trong ô ở dưới biểu thị các trình tự PAM và các phần được gạch chân biểu thị các trình tự đích; b) sự đưa vào pEgPTiD2-pds và vật truyền nhị thể biểu hiện GFP bằng phương pháp xâm nhiễm của *Agrobacterium*, trong đó vi khuẩn *Agrobacterium* mang pEgPTiD2-pds (1) hoặc pEgPTiD-pds (2) và *Agrobacteria* mang plasmid nhị thể biểu hiện GFP được xâm nhiễm bằng phương pháp xâm nhiễm của *Agrobacterium*, và các đĩa lá trong đó sự biểu hiện GFP được quan sát thấy được cắt và được sử dụng cho sự phân tích quá trình đưa đột biến PDS vào.

FIG. 5-2 thể hiện sự phát sinh đột biến định hướng điểm của gen PDS của

cây thuốc lá sử dụng pEgPTiD2-pds: c) sự phân tích quá trình đưa đột biến PDS vào nhờ thử nghiệm Cel-1, trong đó ADN hệ gen được bào chế từ các đĩa lá trong đó sự biểu hiện GFP được quan sát ở FIG. 5-b), và sự có mặt hoặc không có mặt các sự đột biến được phân tích bằng thử nghiệm Cel-1. Kí hiệu hình tam giác biểu thị các đoạn gen PDS đã đột biến mà được cắt bằng Cel-1 nucleaza.

FIG. 6 thể hiện sự phát sinh đột biến của gen IAA9 của cây cà chua sử dụng pEgPTiD2-iaa9: a) trình tự đích trên gen IAA9 của cây cà chua, trong đó trình tự đích 1 được chọn từ exon thứ hai, phần được đóng khung trong trình tự đích được thể hiện ở ô ở dưới biểu thị trình tự PAM, và phần được gạch chân biểu thị trình tự đích; b) pEgPTiD2-iaa9 được đưa vào trong đĩa lá cà chua bằng phương pháp *Agrobacterium* để thu các tế bào thể chai được biến nạp; c) sự phân tích đột biến bằng PCR-RFLP, trong đó một vùng chứa IAA9 trình tự đích được khuếch đại bằng PCR từ ADN hệ gen mà được bào chế từ các tế bào thể chai được biến nạp trong đó pEgPTiD2-iaa9 được đưa vào, và sự phân tích đột biến được thực hiện bằng PCR-RFLP sử dụng *AccI*. Các hình tam giác trống biểu thị các đoạn cắt *AccI* có nguồn gốc kiểu dại, và hình tam giác ở phía trên các hình tam giác trống biểu thị các đoạn đột biến mà không trải qua sự cắt bằng *AccI*.

FIG. 7 thể hiện sự phân tích đột biến bằng cách xác định trình tự trong các thể chai được đưa pEctiD2-iaa9 vào. Trình tự phía trên cùng thể hiện trình tự IAA9 kiểu dại, và phần được gạch chân biểu thị trình tự đích. Các trình tự được đóng khung biểu thị các trình tự PAM. Các vị trí, mà ở đó xuất hiện các sự đột biến, được thể hiện bằng cách xen các kí hiệu hoặc các dấu nối. Dấu nối biểu thị sự khuyết đoạn nucleotit.

FIG. 8 thể hiện sự phân tích đột biến ở thực vật tái sinh được đưa pEctiD2-iaa9- vào: a) sự phân tích đột biến bằng PCR-RFLP, trong đó các hình tam giác trống biểu thị các đoạn cắt *AccI* có nguồn gốc kiểu dại, và hình tam

giác ở phía trên các hình tam giác trông biểu thị các đoạn đột biến mà không trải qua sự cắt bằng AccI; và b) hình ảnh của cây cà chua được gây đột biến mà thể hiện hình thái bất thường của lá thực do sự đứt gãy gen IAA9.

FIG. 9 thể hiện sơ đồ thực nghiệm của sự chỉnh sửa hệ gen sử dụng dòng tế bào HEK293.

FIG. 10 thể hiện các kết quả của sự phân tích đột biến bằng sự phân tích tính linh hoạt của chuỗi kép dị hợp. Đoạn được cho là có nguồn gốc từ trình tự được gây đột biến được phát hiện (kí hiệu hình phím màu đen) từ hệ gen của tế bào trong đó crARN chứa trình tự của đích 1 trên gen EMX1 và các gen TiD được đưa vào.

FIG. 11 thể hiện các kết quả của sự phân tích đột biến bằng sự phân tích tính linh hoạt của chuỗi kép dị hợp. Đoạn được cho là có nguồn gốc từ trình tự được gây đột biến được phát hiện (kí hiệu hình phím màu đen) từ hệ gen của tế bào trong đó crARN chứa trình tự của đích 2 trên gen EMX1 và các gen TiD được đưa vào.

FIG. 12 thể hiện sự phân tích xác định trình tự của các trình tự được gây đột biến. Các chữ viết hoa màu trắng trên nền đen biểu thị PAM (protospacer adjacent sequence – trình tự liền kề đoạn đệm) được nhận diện bởi TiD. Các trình tự được đóng khung biểu thị các trình tự đích. Các dấu nối (-) biểu thị sự khuyết đoạn nucleotit. Chữ cái thường in đậm màu đen biểu thị sự chèn nucleotit. Ở phía bên phải của từng trình tự, hiệu quả đột biến soma (số lượng các dòng vô tính trong đó trình tự được gây đột biến được quan sát / tổng số các dòng vô tính được phân tích) được thể hiện.

FIG. 13 thể hiện sự phân tích xác định trình tự của các trình tự được gây đột biến. Các chữ viết hoa màu trắng trên nền đen biểu thị PAM (protospacer adjacent sequence – trình tự liền kề đoạn đệm) được nhận diện bởi TiD. Các trình tự được đóng khung biểu thị các trình tự đích. Các dấu nối (-) biểu thị sự

khuyết đoạn nucleotit. Các chữ cái thường in đậm màu đen biểu thị sự chèn nucleotit. Ở phía bên phải của từng trình tự, hiệu quả đột biến soma (số lượng các dòng vô tính trong đó trình tự được gây đột biến được quan sát / tổng số các dòng vô tính được phân tích) được thể hiện.

Mô tả chi tiết sáng chế

Sáng chế đề xuất kỹ thuật chỉnh sửa hệ gen sử dụng hệ CRISPR typ I-D. Cụ thể, trong số các Cas protein của CRISPR typ I-D, Cas3d, Cas5d, Cas6d, Cas7d và Cas10d được sử dụng trong sáng chế. Trong sáng chế, hệ CRISPR typ I-D được tìm thấy bao gồm môđun nhận biết đích gồm Cas5d, Cas6d và Cas7d và môđun cắt polynucleotit gồm Cas3d và Cas10d.

Cụ thể, nguyên lý hoạt động của sáng chế là như sau.

Phức hợp bao gồm:

1) gARN mà cần thiết cho sự nhắm vào trình tự nucleotit đích (sau đây còn được gọi là “nhắm vào đích”), gồm có trình tự bổ trợ cho trình tự nucleotit đích và trình tự lặp chung hiện diện trong locus CRISPR typ I-D,

2) Cas5d mà nhận biết trình tự PAM có mặt ở vùng lân cận trình tự nucleotit đích,

3) Cas7d mà liên kết với gARN của mục 1) và cần thiết cho sự nhắm vào trình tự nucleotit đích, và

4) Cas6d mà thực hiện sự xử lý gARN của mục 1), và

5) phức hợp bao gồm Cas10d mà tương tác với phức hợp bao gồm các mục 1) đến 4) và thực hiện sự tái mô hình hóa của trình tự nucleotit đích và Cas3d mà gây ra sự suy thoái của polynucleotit được tạo ra cho tế bào và trong tế bào,

6) nhắm vào trình tự nucleotit đích nhờ phức hợp bao gồm 1) đến 4), tức là,

7) nhắm vào trình tự nucleotit đích nhờ phức hợp gồm có gARN trưởng

thành mà thu được bằng cách xử lý gARN của mục 1) bằng Cas6d của mục 4), và 2) và 3)

được thực hiện, và

8) polynucleotit trên trình tự nucleotit đích được cắt bằng phức hợp của mục 5).

Do vậy, sáng chế đề xuất phương pháp hướng đích trình tự nucleotit đích (sau đây còn được gọi là “phương pháp hướng đích trình tự đích theo sáng chế”), phương pháp biến đổi trình tự nucleotit đích (sau đây được gọi là “phương pháp biến đổi trình tự đích theo sáng chế”), và phương pháp ức chế sự biểu hiện của gen đích (sau đây còn được gọi là “phương pháp ức chế sự biểu hiện gen đích theo sáng chế”), trong đó hệ CRISPR typ I-D được sử dụng trong các phương pháp này. Hơn nữa, sáng chế đề xuất phức hợp bao gồm CRISPR typ I-D- được kết hợp các protein Cas và gARN (sau đây còn được gọi là “phức hợp theo sáng chế”), và vật truyền bao gồm phân tử axit nucleic mã hóa cho phức hợp nêu trên, mà được sử dụng trong các phương pháp nêu trên của sáng chế.

(1) Tế bào

Trong sáng chế, tế bào có thể là tế bào không nhân hoặc tế bào có nhân điển hình, và không bị giới hạn một cách cụ thể. Các ví dụ về tế bào bao gồm vi khuẩn, archaea, nấm men, tế bào thực vật, tế bào côn trùng, và các tế bào động vật (ví dụ, tế bào người, tế bào không phải của người, tế bào động vật có xương sống không phải động vật có vú, tế bào động vật không xương sống, v.v.).

(2) Endonucleaza được dẫn hướng bởi ARN và protein Cas

Trong sáng chế, “endonucleaza được dẫn hướng bởi ARN” có nghĩa là endonucleaza bao gồm ít nhất là một vùng nucleaza và ít nhất là một vùng mà liên kết với gARN, mà được dẫn hướng đến trình tự nucleotit đích (hoặc vị trí nucleotit đích) bởi gARN. Endonucleaza được dẫn hướng bởi ARN được sử dụng trong sáng chế là endonucleaza được dẫn hướng bởi ARN có nguồn gốc từ

CRISPR typ I-D, và bao gồm các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d. Trong sáng chế, nhận thấy rằng Cas5d, Cas6d và Cas7d cấu thành “môđun nhận biết đích” mà góp phần vào sự nhận diện đích, và Cas3d và Cas10d cấu thành “môđun cắt polynucleotit” mà góp phần vào việc cắt polynucleotit. Cụ thể, endonucleaza được dẫn hướng bởi ARN được sử dụng trong sáng chế bao gồm môđun nhận biết đích gồm Cas5d, Cas6d và Cas7d và môđun cắt polynucleotit gồm Cas3d và Cas10d.

Cas3d, Cas5d, Cas6d, Cas7d và Cas10d được sử dụng trong sáng chế có thể có nguồn gốc từ vi khuẩn hoặc archaeum bất kỳ. Các ví dụ về ví dụ và archaeum bao gồm *Microcystis aeruginosa*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Anabaena cylindrica*, *Anabaena variabilis*, *Caldicellulosiruptor lactoaceticus*, *Caldilinea aerophila*, *Clostridium algicarnis*, *Crinalium epipsammum*, *Cyanothece Sp.*, *Cylindrospermum stagnale*, *Haloquadratum walsbyi*, *Halorubrum lacusprofundi*, *Methanocaldococcus vulcanius*, *Methanospirillum hungatei*, *Natrialba asiatica*, *Natronomonas pharaonis*, *Nostoc punctiforme*, *Phormidesmis priestleyi*, *Oscillatoria acuminata*, *Picrophilus torridus*, *Spirochaeta thermophila*, *Stanieria cyanosphaera*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Synechocystis Sp.*, *Thermacetogenium phaeum*, *Thermofilum pendens*, v.v.. Thông tin trình tự axit amin và trình tự nucleotit của các protein Cas thu được từ dữ liệu cơ sở công cộng, ví dụ, ngân hàng gen NCBI GenBank. Ngoài ra, các trình tự từ các loài vi khuẩn mới cũng có thể thu được từ dữ liệu hệ gen vi khuẩn thu được bằng cách phân tích đa hệ gen hoặc tương tự bằng cách sử dụng chương trình BLAST. Các axit nucleic mã hóa các protein Cas có thể được tạo thành, ví dụ, bằng phương pháp tổng hợp hóa học hoặc phương pháp tương tự sau khi chọn lọc các tối ưu để dịch mã trong tế bào chủ trong đó các axit nucleic được đưa vào trên cơ sở thông tin về trình tự axit amin. Việc sử dụng các codon mà

thường xuyên được sử dụng trong tế bào chủ làm cho có thể gia tăng mức biểu hiện của protein. Ví dụ, các protein Cas có thể được tổng hợp hóa học dựa trên thông tin trình tự axit amin, hoặc được tạo ra trong tế bào bằng cách đưa các axit nucleic mã hóa các protein Cas vào trong tế bào thông qua vật truyền thích hợp hoặc tương tự. Mỗi protein Cas trong số Cas3d, Cas5d, Cas6d, Cas7d và Cas10d có thể là protein Cas kiểu đột biến miễn là giữ được chức năng của từng protein Cas như được nêu trong phần nguyên lý hoạt động của sáng chế.

(3) ARN dẫn hướng

Trong sáng chế, ARN dẫn hướng (gARN) là phân tử mà tạo thành phức hợp với môđun nhận biết đích (Cas5d, Cas6d và Cas7d) để nhằm vào trình tự nucleotit đích cùng với các Cas protein này. Trong sáng chế, gARN liên kết với Cas7d của môđun nhận biết đích. Trong sáng chế, gARN liên kết với phức hợp bao gồm Cas5d, Cas6d và Cas7d để dẫn phức hợp đến trình tự nucleotit đích. Ví dụ, gARN liên kết với môđun nhận biết đích của endonucleaza được dẫn hướng bởi ARN để dẫn endonucleaza được dẫn hướng bởi ARN đến trình tự nucleotit đích. Khi môđun nhận biết đích có mặt như một phần của protein dung hợp khác với endonucleaza được dẫn hướng bởi ARN, gARN liên kết với môđun nhận biết đích để dẫn protein dung hợp đến trình tự nucleotit đích.

gARN bao gồm trình tự bổ trợ cho trình tự đích sao cho cặp bazơ có thể được tạo ra giữa gARN và trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR typ I-D trước và sau (ở phía đầu 5' và phía đầu 3' của) trình tự bổ trợ. Các phần trình tự lặp chung của gARN có thể có ít nhất là một cấu trúc cấu trúc hình cặp tóc. Ví dụ, phần trình tự lặp chung nằm ở phía đầu 5' của trình tự bổ trợ cho trình tự nucleotit đích có thể có cấu trúc hình cặp tóc, và phần trình tự lặp chung nằm ở phía đầu 3' của trình tự bổ trợ cho trình tự nucleotit đích có thể là sợi đơn. Trong sáng chế, gARN tốt hơn là có cấu trúc hình cặp tóc.

Trình tự lặp chung có nguồn gốc từ locus CRISPR typ I-D có thể được phát hiện thấy từ vùng trình tự gen gARN liền kề nhóm gen typ I-D bằng cách sử dụng chương trình tra cứu đoạn lặp liên tiếp. Độ dài nucleotit của trình tự lặp chung có trong gARN không bị giới hạn một cách cụ thể miễn là gARN tương tác với môđun nhận biết đích để nhằm vào trình tự nucleotit đích. Ví dụ, mỗi trình tự lặp chung trong số các trình tự lặp trước và sau trình tự hỗ trợ cho trình tự nucleotit đích có thể có độ dài khoảng 10 đến 70 nucleotit, ví dụ, độ dài từ 30 đến 50 nucleotit.

gARN có thể chứa trình tự gồm có khoảng 10 đến 70 nucleotit, mà hỗ trợ cho trình tự nucleotit đích. Trình tự hỗ trợ cho trình tự nucleotit đích có trong gARN tốt hơn là trình tự gồm có 20 đến 50 nucleotit, tốt hơn nữa là trình tự gồm có 25 đến 45 nucleotit, tốt hơn nữa là trình tự gồm có 30 đến 40 nucleotit, hoặc vẫn tốt hơn nữa là trình tự gồm có 32 đến 37 nucleotit, ví dụ, trình tự gồm có 32 nucleotit, 33 nucleotit, 34 nucleotit, 35 nucleotit, 36 nucleotit, hoặc 37 nucleotit. Tính đặc hiệu trình tự của sự nhận biết đích nhờ gARN được gia tăng nhiều hơn nữa vì trình tự đích mà có thể được nhằm vào là dài hơn. Ngoài ra, trị số Tm của cặp bazơ được tạo ra giữa gARN và trình tự đích là cao hơn và như vậy tính ổn định của quá trình nhận biết đích được gia tăng nhiều hơn nữa vì trình tự đích mà có thể được nhằm vào là dài hơn. Vì độ dài của trình tự mà có thể được nhằm vào bởi gARN cho endonucleaza được dẫn hướng bởi các ARN (ví dụ, Cas9 và Cpf1) được sử dụng trong các kỹ thuật chỉnh sửa hệ gen thông thường là độ dài khoảng 20 đến 24 nucleotit, sáng chế có tính đặc hiệu trình tự và tính ổn định tốt hơn so với các phương pháp thông thường.

(4) Trình tự nucleotit đích

Trong sáng chế, trình tự nucleotit đích (còn được gọi là “trình tự đích”, như được sử dụng ở đây) là trình tự axit nucleic bất kỳ, và không bị giới hạn một cách cụ thể miễn là nó là trình tự nằm ở vùng lân cận của mẫu gần đoạn đệm

(PAM). Trình tự nucleotit đích có thể là trình tự ADN sợi kép, trình tự ADN sợi đơn, hoặc trình tự ARN. Các ví dụ về ADN bao gồm ADN hệ gen có nhân điển hình, ADN ty thể, ADN lập thể, ADN hệ gen không có nhân điển hình, ADN thể thực khuẩn, và ADN plasmit. Trong sáng chế, trình tự nucleotit đích tốt hơn là ADN sợi kép trên hệ gen. Như được sử dụng ở đây, cụm từ “ở vùng lân cận” bao gồm cả nghĩa liền kề một vị trí và gần sát một vị trí. Như được sử dụng ở đây, thuật ngữ “lân cận” bao gồm cả nghĩa gần kề và bên cạnh.

Các trình tự PAM dùng để nhận biết đích của các hệ CRISPR thay đổi tùy thuộc vào các kiểu của các hệ CRISPR. Trong sáng chế, nhận thấy rằng trình tự PAM được sử dụng bởi hệ CRISPR typ I-D là 5'-GTH-3' (H = A, C hoặc T) (ví dụ 1). Tốt hơn là, trình tự nằm ở vùng lân cận phía xuôi về đầu 3' của trình tự PAM được chọn là trình tự nucleotit đích. Ví dụ, trình tự nucleotit đích có thể là trình tự nằm ở vùng lân cận trình tự PAM và hiện diện trong intron, vùng mã hóa, vùng không mã hóa, hoặc vùng kiểm soát của gen đích. Gen đích có thể là gen bất kỳ và được chọn tùy ý.

Các trình tự PAM dùng cho Cas9 và Cpf1 được sử dụng trong các kỹ thuật chỉnh sửa hệ gen thông thường là 5'-NGG-3' (N = A, C, G hoặc T) và 5'-TTTV-3' (V = A, C hoặc G). Tần suất xuất hiện của trình tự PAM cho TiD (tức là, số lượng đích tham gia của hệ CRISPR) được so sánh với các tần suất xuất hiện của các trình tự PAM cho Cas9 và Cpf1 trong các trình tự hệ gen của các thực vật bậc cao hơn. Do đó, nhận thấy rằng tần suất xuất hiện của trình tự PAM cho TiD là cao nhất, và TiD có số lượng đích lớn hơn so với các kỹ thuật chỉnh sửa hệ gen thông thường sử dụng Cas9 và Cpf1 (Bảng 1).

[Bảng 1]

So sánh số lượng trình tự PAM (= số lượng đích tham gia) trong các trình tự hệ gen của các thực vật bậc cao hơn			
Loài thực vật	TiD	Cas9	Cpf1
Cây cà chua	$5,99 \times 10^7$	$5,10 \times 10^7$	$3,97 \times 10^7$
Cây dâu tây	$1,70 \times 10^7$	$1,68 \times 10^7$	$0,99 \times 10^7$
Cây khoai tây	$5,45 \times 10^7$	$4,99 \times 10^7$	$3,71 \times 10^7$
Cây đậu tương	$7,20 \times 10^7$	$6,96 \times 10^7$	$5,38 \times 10^7$
Cây ớt tây	$21,58 \times 10^7$	$20,19 \times 10^7$	$13,60 \times 10^7$
Cây dương	$2,98 \times 10^7$	$2,77 \times 10^7$	$2,39 \times 10^7$
Cây cải bắp	$3,67 \times 10^7$	$3,19 \times 10^7$	$2,42 \times 10^7$
Cây củ cải	$4,11 \times 10^7$	$4,06 \times 10^7$	$2,70 \times 10^7$
Cỏ linh lăng	$3,08 \times 10^7$	$2,48 \times 10^7$	$2,31 \times 10^7$

(5) Phương pháp hướng đích trình tự đích theo sáng chế

Phương pháp hướng đích trình tự đích của sáng chế khác biệt ở chỗ đưa môđun nhận biết đích (Cas5d, Cas6d và Cas7d) và gARN vào trong tế bào. Cụ thể, phương pháp hướng đích trình tự đích theo sáng chế khác biệt ở chỗ đưa vào trong tế bào (i) Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein này, và (ii) gARN hoặc ADN mã hóa gARN. Phương pháp hướng đích trình tự đích theo sáng chế có thể được tiến hành in vitro hoặc in vivo.

Trong phương pháp hướng đích trình tự đích theo sáng chế, môđun nhận biết đích có thể được đưa vào trong tế bào dưới dạng phức hợp được phân tách bao gồm Cas5d, Cas6d và Cas7d, hoặc mỗi protein trong số Cas5d, Cas6d và Cas7d có thể được đưa vào trong tế bào dưới dạng protein đơn được phân tách. Trong phương pháp hướng đích trình tự đích theo sáng chế, môđun nhận biết đích cũng có thể được đưa vào trong tế bào dưới dạng các axit nucleic mã hóa cho các protein Cas Cas5d, Cas6d và Cas7d. Các ví dụ về axit nucleic bao gồm ARN như mARN và ADN.

Các ADN mã hóa cho các protein Cas có thể có trong, ví dụ, một vật truyền. Trình tự ADN tốt hơn là được liên kết linh hoạt với trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc. Khi tế bào trong đó môđun nhận biết đích được đưa vào là tế bào có nhân điển hình, tín hiệu định vị nhân trình tự tốt hơn

là được thêm vào ADN mã hóa protein Cas. Hai hoặc nhiều hoặc toàn bộ các ADN mã hóa cho các protein Cas Cas5d, Cas6d và Cas7d có thể có trong vật truyền đơn hoặc có thể có trong các vật truyền riêng biệt. Số lượng các vật truyền và loại và các tổ hợp của các protein Cas được mã hóa bởi các ADN được kết hợp vào trong mỗi vật truyền không bị giới hạn. Khi hai hoặc nhiều ADN mã hóa cho các protein Cas có trong vật truyền đơn, các trình tự ADN có thể được liên kết với nhau, ví dụ qua trình tự mã hóa cho peptit tự cắt, để được biểu hiện nhiều xistron. Hai hoặc nhiều ADN mã hóa cho các protein Cas có thể được liên kết theo trình tự bất kỳ.

GARN có thể được đưa vào trong tế bào dưới dạng ARN hoặc dưới dạng ADN mã hóa cho gARN. ADN mã hóa cho gARN có thể có, ví dụ, trong vật truyền. Trình tự ADN tốt hơn là được liên kết linh hoạt với trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc.

Các ADN mã hóa cho các protein Cas và ADN mã hóa cho gARN có thể có trong cùng một vật truyền hoặc có thể có trong các vật truyền riêng biệt. Ví dụ, một hoặc nhiều hoặc toàn bộ các ADN mã hóa Cas5d, Cas6d và Cas7d, và ADN mã hóa gARN có thể có trong vật truyền đơn.

Trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc và trình tự tín hiệu định vị nhân là đã biết trong lĩnh vực kỹ thuật này và có thể được chọn một cách phù hợp tùy thuộc vào các loài sinh vật trong đó tế bào trong đó môđun nhận biết đích và gARN được đưa vào có nguồn gốc từ các loài sinh vật đó. Vật truyền được dùng cho quá trình đưa vào có thể được chọn một cách phù hợp tùy thuộc vào các loài sinh vật trong đó tế bào trong đó vật truyền được đưa vào có nguồn gốc từ các loài sinh vật, và không bị giới hạn một cách cụ thể. Các ví dụ về vật truyền này bao gồm các vật truyền plasmit, các vật truyền virut, các phagemid, các cosmid, các nhiễm sắc thể nhân tạo/vi thể nhiễm sắc, và các gen nhảy.

Quá trình đưa môđun nhận biết đích và gARN vào trong tế bào có thể được thực hiện bằng nhiều phương thức khác nhau đã biết trong lĩnh vực kỹ thuật này. Các ví dụ về các phương thức như vậy bao gồm sự chuyển nhiễm, ví dụ, sự chuyển nhiễm được gây ra bởi canxi phosphat, sự mở lỗ bằng điện, sự chuyển nhiễm liposom, v.v., sự tải nạp virus, sự chuyển nhiễm nhờ liposom, súng bắn gen, vi tiêm, phương pháp *Agrobacterium*, phương pháp xâm nhiễm của *Agrobacterium*, và phương pháp PEG-canxi.

Môđun nhận biết đích và gARN có thể được đưa vào trong tế bào một cách đồng thời hoặc liên tiếp. Cas5d, Cas6d và Cas7d cấu thành môđun nhận biết đích, hoặc các axit nucleic mã hóa cho các protein Cas này có thể được đưa vào trong tế bào một cách đồng thời hoặc liên tiếp. Ví dụ, các protein Cas Cas5d, Cas6d và Cas7d được tổng hợp in vitro hoặc in vivo và gARN được tổng hợp in vitro hoặc in vivo có thể được ủ in vitro để tạo thành phức hợp, và phức hợp có thể được đưa vào trong tế bào.

Trong quá trình đưa vào môđun nhận biết đích và gARN, tế bào được nuôi cấy trong các điều kiện thích hợp cho phương pháp hướng đích trình tự nucleotit đích. Tế bào này sau đó được nuôi cấy trong các điều kiện thích hợp cho sự sinh trưởng và duy trì tế bào. Các điều kiện nuôi cấy có thể thích hợp cho các loài sinh vật trong đó tế bào trong đó môđun nhận biết đích và gARN được đưa vào có nguồn gốc từ, và có thể được xác định một cách phù hợp bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này, ví dụ, dựa trên các kỹ thuật nuôi cấy tế bào đã biết.

Theo phương pháp hướng đích trình tự đích theo sáng chế, gARN liên kết với Cas7d của môđun nhận biết đích để tạo ra phức hợp gồm môđun nhận biết đích và gARN. Đồng thời, gARN tạo thành cặp bazơ với trình tự nucleotit đích. Môđun nhận biết đích xác định trình tự nucleotit đích theo cách đặc hiệu trình tự bằng cách nhận diện trình tự PAM có mặt ở vùng lân cận trình tự nucleotit đích.

Trong phương pháp hướng đích trình tự đích theo sáng chế, Cas10d có thể tiếp tục được đưa vào trong tế bào.

(6) Phương pháp hướng đích trình tự đích theo sáng chế

Phương pháp biến đổi trình tự đích theo sáng chế khác biệt ở chỗ đưa endonucleaza được dẫn hướng bởi ARN và gARN vào trong tế bào. Cụ thể, phương pháp biến đổi trình tự đích theo sáng chế khác biệt ở chỗ đưa vào trong tế bào (i) Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, hoặc các axit nucleic mã hóa cho các protein, và (ii) gARN hoặc ADN mã hóa cho gARN. Phương pháp biến đổi trình tự đích theo sáng chế bao gồm bước cắt trình tự nucleotit đích bằng phương pháp hướng đích trình tự đích theo sáng chế với môđun cắt polynucleotit. Phương pháp biến đổi trình tự đích theo sáng chế có thể được thực hiện in vitro hoặc in vivo. Trong sáng chế, sự biến đổi bao gồm sự khuyết đoạn, xen đoạn, và thế đoạn của một hoặc nhiều nucleotit, và tổ hợp của chúng.

Trong phương pháp biến đổi trình tự đích theo sáng chế, ngoài endonucleaza được dẫn hướng bởi ARN và gARN, thể cho polynucleotit có thể được đưa vào trong tế bào. Thể cho polynucleotit bao gồm ít nhất là một trình tự thể cho mà chứa sự biến đổi mong muốn được đưa vào trong vị trí đích. Thể cho polynucleotit có thể bao gồm, ngoài trình tự thể cho, các trình tự có độ tương đồng cao với các trình tự phía trước và phía sau của trình tự đích (tốt hơn là, các trình tự gần giống với các trình tự phía trước và phía sau của trình tự đích) ở cả hai đầu của trình tự thể cho. Thể cho polynucleotit có thể là ADN sợi đơn hoặc ADN sợi kép. Thể cho polynucleotit có thể được thiết lập một cách phù hợp bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này dựa trên các kỹ thuật đã biết trong lĩnh vực kỹ thuật này.

Khi thể cho polynucleotit không có trong phương pháp biến đổi trình tự đích theo sáng chế, sự cắt trong trình tự nucleotit đích có thể được sửa chữa bằng cách ghép nối đầu không tương đồng (NHEJ - non-homologous end

joining). NHEJ được biết đến là quá trình dễ gây ra lỗi, và sự khuyết đoạn, xen đoạn, hoặc thế đoạn của một hoặc nhiều nucleotit, hoặc tổ hợp của chúng có thể xuất hiện trong khi sửa chữa chỗ cắt. Như vậy, trình tự có thể được biến đổi ở vị trí trình tự đích, và bằng cách đó sự dịch chuyển khung hoặc codon kết thúc chưa trưởng thành được gây ra để bất hoạt hoặc hạ gục sự biểu hiện của gen được mã hóa bởi vùng trình tự đích.

Khi thể cho polynucleotit hiện diện trong phương pháp biến đổi trình tự đích theo sáng chế, trình tự thể cho của thể cho polynucleotit được xen vào trong vị trí trình tự đích hoặc thay thế vị trí trình tự đích bằng sự sửa chữa tái tổ hợp tương đồng (HDR - homologous recombination repair) của trình tự nucleotit đích được cắt. Do đó, sự biến đổi mong muốn được đưa vào trong vị trí trình tự đích.

Endonucleaza được dẫn hướng bởi ARN có thể được đưa vào trong tế bào dưới dạng phức hợp đã phân tách bao gồm Cas5d, Cas6d, Cas7d, Cas3d và Cas10d, hoặc mỗi protein trong số Cas5d, Cas6d, Cas7d, Cas3d và Cas10d có thể được đưa vào trong tế bào dưới dạng protein đơn được phân tách. Endonucleaza được dẫn hướng bởi ARN cũng có thể được đưa vào trong tế bào dưới dạng các axit nucleic mã hóa các protein Cas Cas5d, Cas6d, Cas7d, Cas3d và Cas10d. Các ví dụ về axit nucleic bao gồm ARN như mARN và ADN.

ADN mã hóa protein Cas có thể có trong, ví dụ, vật truyền, và trình tự ADN tốt hơn là được liên kết linh hoạt với trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc. Khi tế bào trong đó endonucleaza được dẫn hướng bởi ARN được đưa vào là tế bào có nhân điển hình, trình tự tín hiệu định vị nhân tốt hơn là được thêm vào ADN mã hóa protein Cas. Hai hoặc nhiều hoặc toàn bộ các ADN mã hóa cho các protein Cas Cas3d, Cas5d, Cas6d, Cas7d và Cas10d có thể có trong vật truyền đơn hoặc có thể có trong các vật truyền riêng biệt. Số lượng các vật truyền và các loại và các tổ hợp của các protein Cas được mã hóa bởi các

ADN được kết hợp vào trong từng vật truyền không bị giới hạn. Khi hai hoặc nhiều ADN mã hóa cho các protein Cas có trong vật truyền đơn, các trình tự ADN có thể được liên kết với nhau, ví dụ thông qua trình tự mã hóa cho peptit tự cắt, để được biểu hiện nhiều xistron. Hai hoặc nhiều ADN mã hóa cho các protein Cas có thể được liên kết theo trình tự bất kỳ.

GARN có thể được đưa vào trong tế bào dưới dạng ARN hoặc dưới dạng ADN mã hóa cho gARN. ADN mã hóa cho gARN có thể có, ví dụ, trong vật truyền. ADN trình tự tốt hơn là được liên kết linh hoạt với trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc.

Các ADN mã hóa cho các protein Cas và ADN mã hóa cho gARN có thể có trong cùng một vật truyền hoặc có thể có trong các vật truyền riêng biệt. Ví dụ, một hoặc nhiều hoặc toàn bộ các ADN mã hóa cho Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và ADN mã hóa cho gARN có thể có trong vật truyền đơn.

Trình tự điều hòa như đoạn khởi đầu hoặc đoạn kết thúc và trình tự tín hiệu định vị nhân là đã biết trong lĩnh vực kỹ thuật này, và có thể được chọn một cách phù hợp tùy thuộc vào kiểu tế bào trong đó endonucleaza được dẫn hướng bởi RAN và gARN được đưa vào. Vật truyền được sử dụng trong quá trình đưa vào có thể được chọn một cách phù hợp tùy thuộc vào kiểu tế bào trong đó vật truyền được đưa vào, và không bị giới hạn một cách cụ thể. Các ví dụ về vật truyền bao gồm các vật truyền plasmit, các vật truyền virus, các phagemid, các cosmid, các nhiễm sắc thể nhân tạo/các vi thể nhiễm sắc, và các gen nhảy.

Quá trình đưa endonucleaza được dẫn hướng bởi ARN, gARN, và thể cho polynucleotit vào trong tế bào có thể được thực hiện bằng các phương thức khác nhau đã biết trong lĩnh vực kỹ thuật này. Các ví dụ về các phương thức này bao gồm sự chuyển nhiễm, ví dụ, sự chuyển nhiễm được gây ra bởi canxi phosphat, sự mở lỗ bằng điện, sự chuyển nhiễm liposom, v.v., sự tải nạp virus, sự chuyển nhiễm nhờ liposom, súng bắn gen, vi tiêm, phương pháp *Agrobacterium*,

phương pháp xâm nhiễm của *Agrobacterium*, và phương pháp PEG-canxi.

Endonucleaza được dẫn hướng bởi ARN, gARN, và thể cho polynucleotit có thể được đưa vào trong tế bào một cách đồng thời hoặc liên tiếp. Cas3d, Cas5d, Cas6d, Cas7d và Cas10d tạo thành endonucleaza được dẫn hướng bởi ARN, hoặc các axit nucleic mã hóa cho các protein Cas này có thể được đưa vào trong tế bào một cách đồng thời hoặc liên tiếp.

Trong khi đưa endonucleaza được dẫn hướng bởi ARN và gARN hoặc endonucleaza được dẫn hướng bởi ARN, gARN và thể cho polynucleotit vào, tế bào được nuôi cấy trong các điều kiện thích hợp cho quá trình cắt ở vị trí trình tự đích. Tế bào sau đó được nuôi cấy trong các điều kiện thích hợp cho sự sinh trưởng và duy trì tế bào. Các điều kiện nuôi cấy có thể thích hợp cho các loài sinh vật mà tế bào trong đó endonucleaza được dẫn hướng bởi ARN và gARN hoặc endonucleaza được dẫn hướng bởi ARN, gARN và thể cho polynucleotit được đưa vào có nguồn gốc từ, và có thể được xác định một cách thích hợp bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này, ví dụ, dựa trên các kỹ thuật nuôi cấy tế bào đã biết.

Theo phương pháp biến đổi trình tự đích của sáng chế, gARN tạo thành cặp bazơ với trình tự nucleotit đích, và đồng thời, gARN tương tác với môđun nhận biết đích của endonucleaza được dẫn hướng bởi ARN để dẫn endonucleaza được dẫn hướng bởi ARN đến vị trí trình tự đích. Sau đó, môđun cắt của endonucleaza được dẫn hướng bởi ARN cắt trình tự ở vị trí trình tự đích. Khi trình tự đã cắt được sửa chữa, trình tự đích bị thay đổi. Ví dụ, phương pháp biến đổi trình tự đích theo sáng chế có thể được sử dụng cho sự biến đổi trình tự nucleotit đích trên hệ gen. ADN sợi kép trên hệ gen được cắt và sau đó được biến đổi ở vị trí đích bằng phương pháp biến đổi trình tự đích theo sáng chế.

(7) Phương pháp ức chế gen đích theo sáng chế

Phương pháp ức chế sự biểu hiện gen đích theo sáng chế khác biệt ở chỗ

đưa môđun nhận biết đích (Cas5d, Cas6d và Cas7d) và gARN vào trong tế bào. Cụ thể, phương pháp ức chế gen đích theo sáng chế khác biệt ở chỗ đưa vào trong tế bào (i) Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa cho các protein, và (ii) gARN hoặc ADN mã hóa cho gARN. Trong phương pháp ức chế gen đích theo sáng chế, ít nhất là một phần trình tự gen đích được chọn là trình tự nucleotit đích, và gARN chứa trình tự bổ trợ cho trình tự đích được sử dụng. Phương pháp ức chế gen đích theo sáng chế bao gồm bước ức chế sự biểu hiện của gen chứa trình tự đích bằng cách liên kết phức hợp gồm môđun nhận biết đích và gARN với trình tự đích khi nhằm vào trình tự nucleotit đích bằng phương pháp ức chế gen đích theo sáng chế. Phương pháp ức chế gen đích theo sáng chế có thể được thực hiện in vitro hoặc in vivo. Theo phương pháp ức chế gen đích của sáng chế, mặc dù trình tự gen đích không bị cắt, chức năng của vùng gen chứa trình tự đích hoặc sự biểu hiện của gen bị ức chế bằng cách liên kết phức hợp gồm môđun nhận biết đích và gARN với trình tự nucleotit đích.

Môđun nhận biết đích và gARN, phương pháp đưa chúng vào trong tế bào, dịch nuôi cấy tế bào ở thời điểm đưa vào và sau khi đưa vào, và tương tự là như được mô tả ở phần “(5) Phương pháp hướng đích trình tự đích theo sáng chế”. Trong phương pháp ức chế gen đích theo sáng chế, Cas10d có thể tiếp tục được đưa vào trong tế bào.

(8) Phức hợp theo sáng chế

Phức hợp theo sáng chế bao gồm các protein Cas của CRISPR typ I-D và gARN. Sáng chế tạo ra phức hợp bao gồm môđun nhận biết đích và gARN, và phức hợp gồm endonucleaza được dẫn hướng bởi ARN và gARN. Cụ thể hơn, phức hợp bao gồm Cas5d, Cas6d, Cas7d và gARN, và phức hợp bao gồm Cas5d, Cas6d, Cas7d, Cas3d, Cas10d và gARN được đề xuất. Ngoài ra, phân tử ADN mã hóa cho phức hợp cũng được đề xuất. Phức hợp theo sáng chế có thể được sử dụng trong phương pháp biến đổi trình tự đích, phương pháp ức chế gen đích và

phương pháp hướng đích trình tự đích theo sáng chế. Trình tự đích trên hệ gen của tế bào có thể được biến đổi bằng cách đưa phức hợp bao gồm endonucleaza được dẫn hướng bởi ARN (phức hợp gồm Cas5d, Cas6d, Cas7d, Cas3d và Cas10d) và gARN vào trong tế bào để cho phép phức hợp thực hiện chức năng trong tế bào. Ngoài ra, trình tự đích trong tế bào có thể được nhắm vào và sự biểu hiện của gen được mã hóa bởi vùng trình tự đích có thể được ức chế bằng cách đưa phức hợp bao gồm môđun nhận biết đích (phức hợp gồm Cas5d, Cas6d và Cas7d) và gARN vào trong tế bào để cho phép phức hợp thực hiện chức năng trong tế bào. Phức hợp bao gồm môđun nhận biết đích và gARN có thể còn chứa Cas10d.

Phức hợp theo sáng chế có thể được tạo ra in vitro hoặc in vivo bằng phương pháp thông thường. Ví dụ, các axit nucleic mã hóa cho các protein Cas tạo thành endonucleaza được dẫn hướng bởi ARN hoặc môđun nhận biết đích, và gARN hoặc ADN mã hóa cho gARN có thể được đưa vào trong tế bào để cho tạo thành phép phức hợp trong tế bào.

Các ví dụ về phức hợp theo sáng chế bao gồm, nhưng không bị giới hạn ở, phức hợp bao gồm Cas5d (SEQ ID NO: 1), Cas6d (SEQ ID NO: 2) và Cas7d (SEQ ID NO: 3) từ *Microcystis aeruginosa*, và gARN gồm có trình tự được thể hiện bởi

GUUCCAAUUAUUCUUAAGCCCUAUUAGGGAUUGAAACNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNGUCCAAUUAUUCUUAAG

CCCUAUUAGGGAUUGAAAC (SEQ ID NO:6; N là nucleotit bất kỳ cấu

thành trình tự bổ trợ cho trình tự nucleotit đích), và phức hợp bao gồm Cas 5d

(SEQ ID NO: 1), Cas6d (SEQ ID NO: 2), Cas7d (SEQ ID NO: 3), Cas3d (SEQ

ID NO: 4), và Cas10d (SEQ ID NO: 5) từ *Microcystis aeruginosa*, và gARN

gồm có trình tự được thể hiện bởi

GUUCCAAUUAUUCUUAAGCCCUAUUAGGGAUUGAAACNNNNNNNNN

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGUCCAAUUAUCUUAAG
CCCUAUUAGGGAUUGAAAC (SEQ ID NO:6; N là nucleotit bất kỳ cấu
thành trình tự bổ trợ cho trình tự nucleotit đích). Trong trình tự gARN, số lượng
N có thể được thay đổi trong khoảng 10 đến 70, tốt hơn là 20 đến 50, tốt hơn
nữa là 25 đến 45, vẫn tốt hơn nữa là 30 đến 40, và vẫn tốt hơn nữa là 32 đến 37.

(9) Vật truyền biểu hiện theo sáng chế

Sáng chế còn đề xuất vật truyền biểu hiện chứa axit nucleic mã hóa cho
endonucleaza được dẫn hướng bởi ARN bao gồm Cas3d, Cas5d, Cas6d, Cas7d
và Cas10d, và ADN mã hóa cho gARN bao gồm trình tự bổ trợ cho trình tự đích
và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự
đích, và vật truyền biểu hiện chứa các axit nucleic mã hóa các protein liên kết
CRISPR typ I-D Cas5d, Cas6d và Cas7d, và ADN mã hóa cho gARN bao gồm
trình tự bổ trợ cho trình tự đích và các trình tự lặp chung có nguồn gốc từ locus
CRISPR, trước và sau trình tự đích.

Vật truyền theo sáng chế là vật truyền để đưa các protein Cas và gARN
vào trong tế bào, như được nêu ở phần “(5) Phương pháp hướng đích trình tự
đích theo sáng chế”, “(6) Phương pháp biến đổi trình tự đích theo sáng chế”, và
“(7) Phương pháp ức chế sự biểu hiện gen đích theo sáng chế”. Sau khi đưa vật
truyền vào trong tế bào, các protein Cas và gARN được biểu hiện trong tế bào.
Vật truyền theo sáng chế cũng có thể là vật truyền trong đó trình tự đích có trong
gARN được thay thế bằng trình tự chứa vị trí giới hạn. Vật truyền này được sử
dụng sau khi kết hợp trình tự nucleotit đích mong muốn vào trong vị trí giới hạn.
Trình tự bất kỳ có thể là, ví dụ, trình tự đoạn đệm có mặt trên locus CRISPR typ
I-D hoặc một phần của trình tự đoạn đệm.

(10) Protein dung hợp bao gồm môđun nhận biết đích theo sáng chế

Sáng chế còn điều chế protein dung hợp bao gồm môđun nhận biết đích
và polypeptit chức năng. Khi protein dung hợp và gARN được đưa vào vào

trong tế bào, protein dung hợp được dẫn đến trình tự nucleotit đích hoặc gen đích trong tế bào nhờ tác động của môđun nhận biết đích và gARN, và trình tự nucleotit đích hoặc gen đích bị biến đổi hoặc cải biến nhờ tác động của polypeptit chức năng. Như vậy, sáng chế còn đề xuất phương pháp biến đổi hoặc cải biến trình tự nucleotit đích hoặc gen đích, mà bao gồm bước đưa protein dung hợp và gARN vào trong tế bào. Hơn nữa, sáng chế đề xuất phức hợp bao gồm protein dung hợp và gARN.

Polypeptit chức năng là polypeptit mà thể hiện chức năng bất kỳ đối với trình tự đích và là polypeptit khác với Cas3d và Cas10d. Các ví dụ về polypeptit chức năng gồm, nhưng không bị giới hạn ở, các enzym giới hạn, các yếu tố phiên mã, các ADN metylaza, các histon axetylaza, các protein phát sáng huỳnh quang; môđun cắt polynucleotit, ví dụ, các môđun cắt nucleotit của các enzym giới hạn; các môđun điều chỉnh sự biểu hiện gen, ví dụ, các môđun kích hoạt sự phiên mã và các môđun ức chế sự phiên mã của các yếu tố phiên mã; và các môđun cải biến hệ gen biểu sinh, ví dụ, các môđun metyl hóa của các ADN metylaza, và các môđun axetyl hóa histon của các histon axetylaza. Ví dụ về protein phát sáng huỳnh quang là GFP. Ví dụ, trình tự đích có thể được biến đổi bằng cách đưa protein dung hợp bao gồm môđun nhận biết đích và môđun cắt polynucleotit cùng với gARN vào trong tế bào, theo cách thức giống như phương pháp biến đổi trình tự đích của sáng chế. Ví dụ, trình tự đích có thể được cải biến để điều chỉnh sự biểu hiện của gen đích bằng cách đưa protein dung hợp bao gồm môđun nhận biết đích và môđun điều chỉnh sự biểu hiện gen hoặc môđun cải biến hệ gen biểu sinh cùng với gARN vào trong tế bào. Ví dụ, vùng lân cận trình tự đích có thể được đánh dấu huỳnh quang bằng cách đưa protein dung hợp bao gồm môđun nhận biết đích và protein phát sáng huỳnh quang cùng với gARN vào trong tế bào.

Sau đây, các ví dụ về sáng chế được thể hiện. Tuy nhiên, sáng chế không

bị giới hạn ở các ví dụ này.

Ví dụ thực hiện sáng chế

Như một phương án, một nhóm gồm các gen (Cas3d, Cas5d, Cas6d, Cas7d, Cas10d) có nguồn gốc từ locus CRISPR typ I-D (sau đây còn được gọi là “TiD”) có nguồn gốc từ *Microcystis aeruginosa* được tách dòng và sau đó được sử dụng. Để xử lý và tạo ra các trình tự ADN trong các ví dụ, sự tổng hợp hóa học gen nhân tạo, PCR, quá trình xử lý enzym giới hạn, ghép nối, hoặc phương pháp lắp ráp Gibson được sử dụng. Ngoài ra, phương pháp Sanger hoặc phương pháp xác định trình tự thế hệ kế tiếp được sử dụng để xác định các trình tự nucleotit.

Ví dụ 1. Sự chỉnh sửa hệ gen trong E. coli

Trong ví dụ này, đã chứng tỏ rằng phương pháp theo sáng chế thực hiện chức năng một cách hiệu quả trong E. coli, là sinh vật mô hình vi khuẩn điển hình.

(1) Tạo ra plasmit biểu hiện gen TiD

Nhóm gen có nguồn gốc từ locus CRISPR typ I-D (sau đây còn được gọi là “locus TiD”) của *Microcystis aeruginosa* (sau đây còn được gọi là “*M. aeruginosa*”) được tách dòng. Các trình tự được tối ưu hóa bởi codon của E. coli (các SEQ ID NO: 7 đến 11) mã hóa cho từng protein Cas được tổng hợp hóa học nhân tạo dựa trên thông tin trình tự axit amin của Cas5d, Cas6d, Cas7d, Cas3d và Cas10d có nguồn gốc từ locus TiD từ *M. aeruginosa*. Đoạn ADN bao gồm đoạn khởi đầu tổng hợp J23108 (SEQ ID NO: 12) hoặc trình tự liên kết ribosom tổng hợp (SEQ ID NO: 13) phía trước của từng gen mã hóa protein Cas và trình tự đoạn kết thúc (SEQ ID NO: 14 đến 17) phía sau của từng gen mã hóa protein Cas được gắn vào trong plasmit vật truyền pACYC184 (do Nippon gene sản xuất) để tạo ra pEcTiD1. Ngoài ra, trình tự lặp CRISPR (crARN, SEQ ID NO: 18) hiện diện ở vùng lân cận locus CRISPR typ I-D từ *M. aeruginosa* được

trích rút, và casset biểu hiện crARN (SEQ ID NO: 20) chứa trình tự lặp CRISPR dưới sự kiểm soát của đoạn khởi đầu T7 (SEQ ID NO: 19) được tổng hợp. Casset biểu hiện crARN chứa trình tự vùng khởi đầu của gen *ccdB* của *E. coli* là trình tự đích trong ví dụ này. Trình tự casset biểu hiện crARN được kết hợp vào trong pEcTiD1 để tạo ra pEcTiD2 (FIG. 2a). Hơn nữa, pEcTiD3 chứa các casset biểu hiện gen *Cas5d*, *Cas6d* và *Cas7d* được tạo ra làm vật truyền plasmit biểu hiện TiD cho quá trình chỉnh sửa hệ gen mà không cần làm đứt gãy sợi kép ADN (FIG. 2b). Các đoạn khởi đầu, đoạn kết thúc, trình tự lặp CRISPR, và trình tự casset biểu hiện crARN được sử dụng trong ví dụ này được thể hiện trong Bảng 2.

[Bảng 2]

Đoạn khởi đầu tổng hợp J23108	5'-CTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC-3' (SEQ ID NO:12)
Trình tự liên kết ribosom (RBS)	5'-AATAATTTTGTTTAACTTTAAGAAGGAGATATACAT-3' (SEQ ID NO:13)
Trình tự đoạn kết thúc STOP767	5'- AGATCCTGTAAAACGACGGCCAGT-3' (SEQ ID NO:14)
Trình tự đoạn kết thúc STOP768(1)	5'-CGCCAGGGTTTTCCCAGTC-3' (SEQ ID NO:15)
Trình tự đoạn kết thúc TOP768(2)	5'- CGCCAGGGTTTTCCCAGTC-3' (SEQ ID NO:16)
Trình tự đoạn kết thúc T7	5'- TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG-3' (SEQ ID NO:17)
Trình tự lặp CRISPR	5'- GTTCCAATTAATCTTAAGCCCTATTAGGGATTGAAAC-3' (SEQ ID NO:18)
Trình tự đoạn khởi đầu T7	5'-TAATACGACTCACTATAGG-3' (SEQ ID NO:19)
Trình tự casset biểu hiện crARN	5'-GTTCCAATTAATCTTAAGCCCTATTAGGGATTGAAAC ggtaataatacgaactcactatagggagaaaggatcGTTCCAATTAATCTTAAGCCTATTAGGGATTGAAAC-3' (SEQ ID NO:20) (Các chữ in hoa biểu thị trình tự lặp CRISPR crARN của TiD. Các chữ in thường biểu thị trình tự gồm 35 nucleotit của vùng khởi đầu của gen <i>ccdB</i> , là trình tự đích.)

(2) Tạo ra thư viện mẫu liên kế đoạn đệm (PAM)

Trong ví dụ này, casset gen *ccdB* tổng hợp (SEQ ID NO: 21) (Bảng 3) trong đó trình tự đoạn khởi đầu T7 được liên kết với phía trước của gen *ccdB* của *E. coli* làm ADN đích được sử dụng. Trình tự đích của TiD là trình tự gồm

35 nucleotit bao gồm vùng khởi đầu T7 ở phía trước của gen *ccdB*. Casset gen *ccdB* tổng hợp được gắn vào vị trí đa tách dòng trong vật truyền plasmit pMW219 (do Nippon gene sản xuất) để tạo ra pMW_ccdB1 (FIG. 3a).

Hệ CRISPR nhận diện trình tự kiểu liên kề đoạn đệm (PAM) nằm ở vùng lân cận trình tự đích và liên kết với trình tự đích qua gARN. Vì trình tự PAM của *M. aeruginosa* TiD được sử dụng trong ví dụ này là chưa biết, plasmit thư viện trình tự PAM để xác định trình tự PAM của *M. aeruginosa* TiD được tạo ra. Bốn trình tự nucleotit ngẫu nhiên được đưa vào phần phía trước của đoạn khởi đầu T7 của pMW_ccdB1 bằng cách sử dụng quá trình tổng hợp ADN hóa học nhân tạo và PCR (FIG. 3b). Các plasmit thư viện pMW_ccdB-PAM đã tạo ra được đưa vào trong dòng tế bào *E. coli* kháng *ccdB* giữ được tính kháng CcdB (do Thermo Fisher Scientific sản xuất), và sau đó các plasmit được tạo ra.

[Bảng 3]

SEQ ID NO:21 Casset biểu hiện gen <i>ccdB</i> tổng hợp (trình tự được gạch chân biểu thị đoạn khởi đầu T7. Các chữ in hoa biểu thị vùng mã hóa cho gen <i>ccdB</i> .)
5'-aggettt <u>aatacga</u> ctcactataggagaaaggatccataaaggaggtaaataatgaagcagcgtattacagtgcaggtgacagc gacagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggttaagcacaacatgcagaatgaagcccgctcgtctgctg ccgaacgctggaaagcggaaaatcaggaagggtggtgaggtgccccggtttattgaaatgaacggctctttgctgacgagaaca gggactggtgaaATGCAGTTTAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTG TTTGTGGATGTACAGAGTGATATTATTGACACGCCCCGGGCGACGGATGGTGATCCC CCTGGCCAGTGACAGTCTGCTGTCAGATAAAGTCTCCCGTGAACCTTACCCGGTGG TGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCG GTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCA AAAACGCCATTAACCTGATGTTCTGGGGAATATAA-3'

(3) Xác định trình tự PAM trong hệ TiD của *M. aeruginosa*

Trình tự PAM cho TiD được xác định bằng cách sử dụng pEcTiD3-T7 mà đã chèn trình tự gồm 35 nucleotit hỗ trợ cho vùng khởi đầu T7 trên plasmit thư viện pMW_ccdB-PAM. pEcTiD3-T7 được đưa vào trong chủng *E. coli* BL21AI (do Thermo Fisher Scientific sản xuất) để thu chủng *E. coli* chủ cho quá trình chỉnh sửa hệ gen của gen *ccdB*. Chủng BL21AI [pEcTiD3-T7] biểu hiện các

protein Cas5d, Cas6d và Cas7d cần thiết cho sự nhận biết trình tự đích. Phức hợp Cas5d/Cas6d/Cas7d-crARN nhận biết trình tự đích liền kề trình tự PAM thích hợp và liên kết với trình tự đích để ức chế chức năng của đoạn khởi đầu T7, là trình tự đích, mặc dù có không cắt trình tự đích.

Sự biểu hiện ccdB nhờ pMW-ccdB-PAM được đưa vào trong chủng BL21AI được gây ra trong môi trường được bổ sung arabinoza, và các tế bào BL21AI không có tính kháng CcdB bị tiêu diệt. Khi plasmit thư viện pMW_ccdB-PAM được đưa vào trong các tế bào BL21AI trong đó plasmit biểu hiện TiD đã được đưa vào trước, Cas5d/Cas6d/Cas7-crARN được biểu hiện từ plasmit pEcTiD3 liên kết với đoạn khởi đầu T7 của plasmit thư viện pMW_ccdB-PAM có trình tự PAM thích hợp để được nhận biết bởi TiD, bằng cách đó sự sản sinh protein CcdB bị kìm hãm, và như vậy các tế bào E. coli có thể sinh trưởng. Từ các khuẩn lạc E. coli đã phát triển, plasmit thư viện pMW_ccdB-PAM được tạo ra, và trình tự PAM được phân tích bằng cách xác định trình tự để xác định trình tự PAM của TiD của M. aeruginosa.

Plasmit thư viện pMW_ccdB-PAMs, mà được tạo ra với lượng lớn, được đưa vào trong chủng BL21AI [pEcTiD3-T7] bằng phương pháp tế bào khả biến hóa học. Các tế bào BL21AI giữ lại được plasmit thư viện pMW_ccdB-PAM và pEcTiD3-T7 được chọn lọc trên môi trường thạch LB chứa 25mg/L cloramphenicol, 25mg/L kanamycin và 1% glucoza. Từ các khuẩn lạc E. coli thu được như vậy, khoảng 1×10^7 các khuẩn lạc được thu gom, được rửa vài lần với môi trường lỏng LB không chứa các thuốc kháng sinh và glucoza, và sau đó được tạo huyền phù trong môi trường lỏng LB chứa 1% arabinoza ở nồng độ 1×10^6 tế bào/mL. Huyền phù được nuôi cấy có rung lắc ở nhiệt độ 37°C trong 2 giờ để gây ra sự biểu hiện của crARN và ccdB dưới sự kiểm soát của đoạn khởi đầu T7 nhờ arabinoza. Sau đó, 200µL huyền phù được cấy mỏng trên môi trường thạch LB chứa 25mg/L cloramphenicol, 25mg/L kanamycin và 1% arabinoza.

Sau khi nuôi cấy qua đêm ở nhiệt độ 37°C, các khuẩn lạc vi khuẩn được thu gom. Từ khoảng 500 các khuẩn lạc thu gom như vậy, các plasmit được tạo ra, và vùng lân cận trình tự PAM được đem đi phân tích xác định trình tự. Các trình tự PAM của plasmit thư viện pMW_ccdB-PAMs được tách với sự có mặt của plasmit biểu hiện TiD có trình tự 5'-NGTH-3' (N = A, C, G hoặc T; H = A, C hoặc T). Tần suất sử dụng các trình tự PAM là 28% đối với NGTA, 33% đối với NGTC, và 38% đối với NGTT. Do vậy, nhận thấy rằng trình tự PAM được sử dụng bởi TiD là 5'-GTH-3' (H = A, C hoặc T).

(4) Sự chỉnh sửa hệ gen trong E. coli

Các plasmit pMW_ccdB-PAMgta, pMW_ccdB-PAMgtc và pMW_ccdB-PAMgtt chứa ba kiểu trình tự PAM đã xác định sử dụng các plasmit thư viện pEcTiD3-T7 và pMW_ccdB-PAM được tạo ra, và được đưa cùng với pEcTiD2-I7 vào trong chủng BL21AI. Các tế bào BL21AI giữ lại được pMW_ccdB-PAMgta/pEcTiD2-T7, pMW_ccdB-PAMgtc/pEcTiD2-T7 và pMW_ccdB-PAMgtt/pEcTiD2-T7 được lựa chọn trên môi trường thạch LB chứa 25mg/L cloramphenicol, 25mg/L kanamycin và 1% glucoza, và sau đó được nhận thấy chứa plasmit được đưa vào trong từng tế bào vi khuẩn bằng sự phân tích xác định trình tự. Sau đó, các tế bào BL21AI giữ được plasmit đúng được cấy mỏng trên môi trường thạch LB chứa 25mg/L cloramphenicol, 25mg/L kanamycin và 1% arabinoza và sau đó được nuôi cấy qua đêm ở nhiệt độ 37°C. Nhờ đó, tất cả các tế bào vi khuẩn đã không phát triển, mà được gây ra bởi sự đứt gãy ADN sợi kép trên ADN plasmit với sự có mặt của Cas3d và Cas10d.

Ví dụ 2. Quá trình chỉnh sửa hệ gen ở các thực vật bậc cao hơn

Trong ví dụ này, về phương án chỉnh sửa hệ gen ở các sinh vật có nhân điển hình bậc cao hơn, đã chứng tỏ rằng phương pháp theo sáng chế thực hiện chức năng một cách hữu hiệu trong *Nicotiana benthamiana* và *Solanum lycopersicum*.

(1) Tạo ra vật truyền nhị thể cho sự biểu hiện gen TiD ở tế bào thực vật bậc cao hơn

Theo tần suất trong *Arabidopsis* và cây thuốc lá, các trình tự được tối ưu hóa bởi codon hai lá mầm mã hóa cho mỗi protein Cas được tổng hợp hóa học nhân tạo dựa trên thông tin trình tự axit amin của Cas5d, Cas6d, Cas7d, Cas3d và Cas10d có nguồn gốc từ locus TiD của *M. aeruginosa*. Đoạn ADN bao gồm trình tự tín hiệu định vị nhân (SEQ ID NO: 22, SEQ ID NO: 23) chứa hai tín hiệu định vị nhân được sắp xếp ở đoạn phía trước 5' liên tiếp của từng gen mã hóa cho protein Cas, và trình tự peptit tự cắt 2A (các SEQ ID NO: 24-28) giữa các gen mã hóa cho protein Cas được tạo ra. Trình tự đoạn khởi đầu (đoạn khởi đầu $2 \times 35S$; SEQ ID NO: 29) bao gồm hai đoạn khởi đầu của gen của virus thể khảm của cải hoa lơ 35S được sắp xếp liên tiếp và trình tự gen tăng cường sự dịch mã Ω được liên kết với đầu phía trước 5' của năm đoạn gen TiD được dung hợp với nhau qua các trình tự peptit 2A, và trình tự đoạn kết thúc của gen có kích cỡ 18,2 kDa của protein sốc nhiệt của *Arabidopsis* (SEQ ID NO: 30) được liên kết với đầu phía sau 3' của năm đoạn gen TiD được dung hợp với nhau qua các trình tự peptit 2A, và nhờ đó cassette biểu hiện gen TiD được tạo ra. Cassette biểu hiện gen TiD được tách dòng vào trong vật truyền plasmit nhị thể pCAMBIA2300 để tạo ra pEgPTiD1 (FIG. 4a). Đối với cassette biểu hiện crARN cho các thực vật, ADN trong đó trình tự đoạn đệm chứa hai vị trí enzyme giới hạn BsaI được đặt giữa hai trình tự crARN sao cho trình tự bất kỳ gồm 35 nucleotit được tổng hợp hóa học nhân tạo có thể được gắn vào trong các vị trí BsaI (SEQ ID NO: 31). Trình tự đoạn khởi đầu gen U6 snRNA-26 của *Arabidopsis* (SEQ ID NO: 32) được gắn đầu phía trước 5' của trình tự cassette biểu hiện crARN, và trình tự poly T được gắn ở đầu phía sau 3' của trình tự cassette biểu hiện crARN (Fig. 4b). Cassette biểu hiện crARN cho các thực vật được gắn giữa trình tự RB và đoạn khởi đầu $2 \times 35S$ của pEgPTiD1 để tạo ra

pEgPTiD2, mà được sử dụng là vật truyền plasmit nhị thể biểu hiện gen TiD cho quá trình chỉnh sửa hệ gen thực vật (FIG. 4c). Các trình tự được tối ưu hóa bởi codon hai lá mầm mã hóa cho từng protein Cas mà có tín hiệu định vị nhân được gắn vào trong pEgPTiD1 và pEgPTiD2 được thể hiện trong các SEQ ID NO: 33 đến 37. Trình tự tín hiệu định vị nhân, trình tự peptit tự cắt 2A, đoạn khởi đầu, đoạn kết thúc, và trình tự cassette biểu hiện crARN được sử dụng trong ví dụ này được thể hiện trong Bảng 4.

[Bảng 4-1]

Trình tự axit amin tín hiệu định vị nhân	DPKKKRKVDPKKKRKVSG (SEQ ID NO:22)
Trình tự tín hiệu định vị nhân (mã hóa SEQ ID NO:22)	5'- GACCCAAAGAAGAAGCGGAAGGTAGACCCTAAG AAGAAGCGCAAGGTTTCTGGA-3' (SEQ ID NO:23)
Trình tự axit amin của peptit tự cắt 2A	GSEGRGSLTTCGDVEENPGP (SEQ ID NO:24)
Trình tự peptit tự cắt 2A(1) (mã hóa SEQ ID NO:24)	5'- GGCTCTGAGGGCAGAGGCAGCCTGCTGACCTGC GGCGACGTGGAGGAAAACCTGGCCCT-3' (SEQ ID NO:25)
Trình tự peptit tự cắt 2A(2) (mã hóa SEQ ID NO:24)	5'- GGGTCTGAGGGACGCGGCTCCCTGCTCACCTGTG GAGATGTGGAAGAGAACCCAGGCCCC-3' (SEQ ID NO:26)
Trình tự peptit tự cắt 2A(3) (mã hóa SEQ ID NO:24)	5'- GGTTCTGAAGGCAGAGGCTCTCTGCTGACATGTG GGGATGTGGAGGAAAATCCTGGCCCT-3' (SEQ ID NO:27)
Trình tự peptit tự cắt 2A(4) (mã hóa SEQ ID NO:24)	5'- GGATCCGAGGGCAGAGGAAGTCTGCTAACATGC GGTGACGTTGAGGAGAATCCCGGGCCA-3' (SEQ ID NO:28)
Trình tự 2 x đoạn khởi đầu +Ω của gen 35S của virus thể khảm của cải hoa lơ	5'- GCCAACATGGTGGAGCACGACACTCTCGTCTACT CCAAGAATATCAAAGATACAGTCTCAGAAGACC AAAGGGCTATTGAGACTTTTCAACAAAGGGTAA TATCGGGAAACCTCCTCGGATTCCATTGCCCAGC TATCTGTCACTTCATCAAAAGGACAGTAGAAAA GGAAGGTGGCACCTACAAATGCCATCATTGCGA TAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCC GACAGTGGTCCCAAAGATGGACCCCCACCCACG AGGAGCATCGTGGAAGAAAGACGTTCCAACC ACGTCTTCAAAGCAAGTGGATTGATGTGAACATG GTGGAGCACGACACTCTCGTCTACTCCAAGAATA TCAAAGATACAGTCTCAGAAGACCAAAGGGCTA TTGAGACTTTTCAACAAAGGGTAATATCGGGAA

	ACCTCCTCGGATTCCATTGCCAGCTATCTGTCA CTTCATCAAAAGGACAGTAGAAAAGGAAGGTGG CACCTACAAATGCCATCATTGCGATAAAGGAAA GGCTATCGTTCAAGATGCCTCTGCCGACAGTGGT CCCAAAGATGGACCCCCACCCACGAGGAGCATC GTGGAAAAAGAAGACGTTCCAACCACGTCTTCA AAGCAAGTGGATTGATGTGATATCTCCACTGACG TAAGGGATGACGCACAATCCCACTATCCTTCGCA AGACCCTTCCTCTATATAAGGAAGTTCATTTCAT TTGGAGAGGCCGGTCTAGAGTATTTTACAACAA TTACCAACAACAACAACAACAACAACATTAC AATTACTATTACAATT-3' (SEQ ID NO:29)
--	---

[Bảng 4-2]

Đoạn kết thúc của gen có kích cỡ 18,2 kDa của protein sốc nhiệt của Arabidopsis	5'- ATATGAAGATGAAGATGAAATATTTGGTGTGT CAAATAAAAAGCTTGTGTGCTTAAGTTTGTGT TTTTTCTTGGCTTGTGTGTTATGAATTTGTG GCTTTTTCTAATATTAAATGAATGTAAGATCTC ATTATAATGAATAAACAAATGTTTCTATAATC CATTGTGAATGTTTTGTTGGATCTCTTCTGCAG CATATACTACTGTATGTGCTATGGTATGGAC TATGGAATATGATTAAAGATAAGATGGGCTCA TAGAGTAAAACGAGGCGAGGGACCTATAAAC CTCCCTTCATCATGCTATTTTCATGATCTATTTT ATAAAATAAAGATGTAGAAAAAAGTAAGCGT AATAACCGCAAAACAAATGATTTAAACATG GCACATAATGAGGAGATTAAGTTCGGTTTACG TTTATTTTAGTACTAATTGTAACGTGAGAC-3' (SEQ ID NO:30)
Casset biểu hiện crARN cho các thực vật	5'- GTTCCAATTAATCTTAAGCCCTATTAGGGATT GAAACggagaccctcaattgtcggtctcGTTCCAATTAATC TTAAGCCCTATTAGGGATTGAAAC-3' (SEQ ID NO:31)
Trình tự đoạn khởi đầu của gen U6 snRNA-26 của Arabidopsis thaliana	5'- AAGCTTCGTTGAACAACGGAAACTCGACTTGC CTTCGCACAATACATCATTTCTTCTTAGCTTT TTTTCTTCTTCTTCGTTTCATACAGTTTTTTTTTG TTTATCAGCTTACATTTTCTTGAACCGTAGCTT TCGTTTTCTTCTTTTAACTTTCCATTTCGGAGTT TTTGTATCTTGTTCATAGTTTGTCCCAGGATT AGAATGATTAGGCATCGAACCTTCAAGAATTT GATTGAATAAAACATCTTCATTCTTAAGATAT GAAGATAATCTTCAAAAGGCCCTGGGAATCT GAAAGAAGAGAAGCAGGCCCATTTATATGGG AAAGAACAATAGTATTTCTTATATAGGCCCAT TTAAGTTGAAAACAATCTTCAAAAGTCCCA TCGCTTAGATAAGAAAACGAAGCTGAGTTTAT ATACAGCTAGAGTCGAAGTAGTGATT-3' (SEQ ID NO:32)

(2) Quá trình chỉnh sửa hệ gen trong *Nicotiana benthamiana*

Trong ví dụ về cây thuốc lá, gen phytoen desaturaza (PDS) được chọn làm trình tự đích để đưa các sự đột biến vào (FIG. 5-1 a). Trình tự đích 1 (đích 1, SEQ ID NO: 38) được chọn từ exon thứ ba trong gen PDS của cây thuốc lá, và ADN được tổng hợp hóa học nhân tạo của đích 1 được gắn vào trong casset biểu hiện crARN cho các thực vật để tạo ra pEgPTiD2-pds(1). Tương tự, trình tự đích 2 (đích 2, SEQ ID NO: 39) được chọn từ exon thứ sáu, và ADN được tổng hợp hóa học nhân tạo của đích 2 được gắn vào trong casset biểu hiện crARN cho các thực vật để tạo ra pEgPTiD2-pds(2). Các vật truyền nhị thể được tạo ra như vậy được đưa vào trong chủng *Agrobacterium tumefaciens* GV2260. Quá trình đưa vật truyền biểu hiện TiD nhằm vào PDS của cây thuốc lá vào trong các tế bào của cây thuốc lá được thực hiện bằng phương pháp xâm nhiễm của *Agrobacterium*. Các tế bào *Agrobacterium* còn giữ lại pEgPTiD2-pds(1) hoặc pEgPTiD2-pds(2) và các tế bào *Agrobacterium* còn giữ lại vật truyền nhị thể biểu thị GFP được nuôi cấy riêng biệt, và sau đó được cùng chuyển nhiễm vào lá thực của *Nicotiana benthamiana* (FIG. 5-1 b). Ba ngày sau khi đồng chuyển nhiễm, ADN hệ gen được bào chế từ vùng phát ánh sáng huỳnh quang GFP trong đĩa lá, và sau đó được sử dụng làm khuôn mẫu để khuếch đại PCR đoạn gen PDS có kích cỡ 300-500 bp chứa trình tự đích. Đoạn được khuếch đại PCR được sử dụng cho thử nghiệm Cel-1 để phân tích xem sự đột biến có được đưa vào trong gen PDS hay không. Để làm đối chứng, đĩa lá cây thuốc lá trong đó chỉ vật truyền nhị thể biểu hiện GFP được đưa vào được sử dụng. Khi chỉ vật truyền biểu hiện GFP được đưa vào, không quan sát thấy có sự đột biến gen PDS. Ngược lại, khi pEgPTiD2-pds và vật truyền biểu hiện GFP được đưa vào đồng thời, sự đưa (các) đột biến vào được quan sát thấy trên từng trình tự đích của gen PDS (FIG. 5-2 c). Các trình tự đích 1 và 2 được thể hiện trong Bảng 5.

[Bảng 5]

Trình tự đích 1 trên gen PDS của cây thuốc lá (đích 1)	5'- TGGGTGGTTTGTCTACAGCAAAATATCTGGCAGAT- 3' (SEQ ID NO:38)
Trình tự đích 2 trên gen PDS của cây thuốc lá (đích 2)	5'- AAATTTGCTATTGGACTCTTGCCAGCAATGCTTGG- 3' (SEQ ID NO:39)

(3) Quá trình chỉnh sửa hệ gen trong *Solanum lycopersicum*

Trong ví dụ của cây cà chua, gen IAA9 của yếu tố phiên mã Aux/IAA được chọn là trình tự đích để đưa các sự đột biến vào (FIG. 6a). Trình tự đích 1 (SEQ ID NO: 40) (Bảng 6) được chọn từ exon thứ hai của gen IAA9 của cây cà chua, và ADN được tổng hợp hóa học nhân tạo của đích 1 được gắn vào trong casset biểu hiện crARN cho các thực vật để tạo ra pEgPTiD2-iaa9. Vật truyền nhị thể đã tạo ra được đưa vào trong chủng *Agrobacterium tumefaciens* GV2260. Sự đưa vật truyền biểu hiện TiD nhằm vào gen IAA9 của cây cà chua vào trong các tế bào cây cà chua được thực hiện bằng phương pháp *Agrobacterium* sử dụng đĩa lá có nguồn gốc từ các lá mầm của cây cà chua. Các đĩa lá được đồng chuyển nhiễm với *Agrobacterium* được nuôi cấy trên môi trường hóa rắn MS chứa 100mg/L kanamycin và 1,5mg/L t-zeatin để thu thể chai trong đó sự đưa vào gen của vùng T-ADN trên pEgPTiD2-iaa9 xảy ra (FIG. 6b). Trình tự xác nhận cho enzym giới hạn AccI có trong trình tự đích của IAA9. Khi sự đột biến được đưa vào nhờ sự chỉnh sửa hệ gen bằng TiD, vị trí nhận diện AccI biến mất. Như vậy quá trình phân tích hiện tượng đa hình độ dài enzym giới hạn bằng PCR (RFLP - PCR-restriction enzyme length polymorphism) sử dụng AccI được thực hiện để phân tích các sự đột biến mà xảy ra trên trình tự đích của IAA9. ADN hệ gen được bào chế từ thể chai biến nạp thu được và được dùng làm khuôn mẫu để khuếch đại PCR vùng có kích cỡ khoảng 300 bazơ chứa trình tự đích IAA9. Đoạn PCR được cắt bằng AccI. Nhận thấy rằng đoạn PCR từ dịch nuôi cấy thể chai trong đó pEgPTiD2-iaa9 được đưa vào có trình tự mà không bị cắt bởi AccI nhờ việc đưa đột biến vào ở trình tự đích IAA9 (FIG. 6c). Trình tự

nucleotit của đoạn PCR từ thể chai trong đó pEgPTiD2-iaa9 được đưa vào được xác định. Kết quả là, nhận thấy rằng sự khuyết đoạn nucleotit của 1 đến 4 nucleotit được đưa vào ngay sau trình tự PAM trên trình tự đích IAA9 (FIG. 7).

Thể chai trong đó pEgPTiD2-iaa9 được đưa vào tiếp tục được nuôi cấy trên môi trường hóa rắn MS chứa 100mg/L kanamycin và 1,0mg/L t-zeatin để thu rễ được tái sinh và được biến nạp. ADN hệ gen được bào chế từ các rễ được tái sinh, và được dùng làm khuôn mẫu để tiến hành quá trình phân tích PCR-RFLP với AccI. Như được thể hiện ở FIG. 8a, các đoạn PCR mà không được cắt bằng AccI được quan sát. Nói cách khác, thu được các rễ được tái sinh và được biến nạp trong đó trình tự đích IAA9 bị đột biến gần như 100%. Trong số 14 rễ được tái sinh và được biến nạp, 13 rễ thể hiện các kết quả giống như được thể hiện trên FIG. 8a. Trong các thực vật tái sinh này, các lá thực là dạng lá đơn, là một trong số các kiểu hình được gây ra bởi sự thiếu hụt IAA9. Như vậy, đã nhận thấy rằng (các) sự đột biến có thể được đưa vào có hiệu quả cao nhờ quá trình chỉnh sửa hệ gen sử dụng TiD.

[Bảng 6]

Trình tự đích trên gen IAA9 của cây cà chua	5'- TACCTGGATCTCAGTCTCCCGAAAGAGGTGAGGAG- 3' (SEQ ID NO:40)
---	--

Ví dụ 3. Sự chỉnh sửa hệ gen trong các động vật bậc cao hơn

Trong ví dụ này, như là phương án về quá trình chỉnh sửa hệ gen ở động vật bậc cao hơn, người ta đã chứng tỏ rằng phương pháp theo sáng chế thực hiện chức năng một cách hiệu quả trong dòng tế bào có nguồn gốc từ tế bào thận phôi người HEK293.

(1) Tạo ra vật truyền để biểu hiện gen TiD ở các tế bào động vật bậc cao hơn

Các trình tự gen mã hóa cho mỗi protein Cas được tổng hợp hóa học nhân tạo dựa trên thông tin trình tự axit amin của Cas5d, Cas6d, Cas7d, Cas3d và Cas10d có nguồn gốc từ locus TiD từ *M. aeruginosa*. Đoạn ADN bao gồm trình tự tín hiệu định vị nhân (SEQ ID NO: 22, SEQ ID NO: 23) chứa hai tín hiệu

định vị nhân được bố trí liên tiếp ở đầu phía trước 5' của mỗi gen trong số các gen mã hóa cho protein Cas, và trình tự peptit tự cắt 2A (các SEQ ID NO: 24-28) giữa các gen mã hóa cho protein Cas được tạo ra. Trình tự lai đoạn khởi đầu của gen β -actin của gà + gen tăng cường của cytomegalovirus (đoạn khởi đầu CBh; SEQ ID NO: 41) được gắn ở đầu phía trước 5' của năm đoạn gen TiD được dung hợp với nhau nhờ các trình tự peptit 2A, và trình tự đoạn kết thúc gen hormon sinh trưởng của bò (đoạn kết thúc bGH; SEQ ID NO: 42) được gắn đầu phía sau 3' của năm đoạn gen TiD được dung hợp với nhau nhờ các trình tự peptit 2A, và nhờ đó cassette biểu hiện gen TiD được tạo ra. Cassette biểu hiện gen TiD được gắn vào trong vật truyền pCR8TOPO (do Thermo Fisher Scientific sản xuất) để tạo ra pCR_hTiD. Đối với cassette biểu hiện crARN, ADN trong đó trình tự đoạn đệm chứa hai vị trí enzym giới hạn BsaI được đặt giữa hai trình tự crARN sao cho trình tự bất kỳ gồm 35 nucleotit có thể được ghép nối được tổng hợp hóa học nhân tạo (SEQ ID NO: 31). Trình tự đoạn khởi đầu gen U6 snRNA của người (SEQ ID NO: 43) là trình tự kiểm soát sự biểu hiện được gắn ở đầu phía trước 5' của cassette biểu hiện crARN, và trình tự poly T được gắn ở đầu phía sau 3' của cassette biểu hiện crARN. Cassette biểu hiện crARN với đoạn khởi đầu gen U6 snRNA của người và trình tự poly T được gắn vào trong vật truyền pCR8TOPO (do Thermo Fisher Scientific sản xuất) để tạo ra pCR_crARN. Các trình tự mã hóa cho từng protein Cas cùng với các tín hiệu định vị nhân trong pCR_hTiD được thể hiện dưới dạng các SEQ ID NO: 33-37. Trình tự đoạn khởi đầu CBh, trình tự đoạn kết thúc bGH, và trình tự đoạn khởi đầu của gen U6 snRNA của người được thể hiện trong Bảng 7.

[Bảng 7]

Gen tăng cường của cytomegalovirus + đoạn khởi đầu lai gen β -actin của gà chung	5'- CGTTACATAACTTACGGTAAATGGCCCCGCCTGGC TGACCGCCCAACGACCCCCGCCATTGACGTCAA TAGTAACGCCAATAGGGACTTTCCATTGACGTCA ATGGGTGGAGTATTTACGGTAAACTGCCCACTTG GCAGTACATCAAGTGTATCATATGCCAAGTACGC CCCCTATTGACGTCAATGACGGTAAATGGCCCCG CTGGCATTGTGCCAGTACATGACCTTATGGGAC TTTCCTACTTGGCAGTACATCTACGTATTAGTCAT CGCTATTACCATGGTCGAGGTGAGCCCCACGTTC TGCTTCACTCTCCCCATCTCCCCCCCCCTCCCCACC CCCAATTTTGTATTTATTTATTTTAAATTATTTTG TGCAGCGATGGGGGCGGGGGGGGGGGGGGGGGCG CGCGCCAGGCGGGGCGGGGCGGGGCGAGGGGCG GGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGC CAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTT ATGGCGAGGCGGCGGGCGGGCGGGCCTATAAA AAGCGAAGCGCGCGGGCGGGCGGGAGTCGCTGCG ACGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCC GCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCG CGTTACTCCCACAGGTGAGCGGGCGGGACGGCC CTTCTCCTCCGGGCTGTAATTAGCTGAGCAAGAG GTAAGGGTTTAAGGGATGGTTGGTTGGTGGGGTA TTAATGTTTAATTACCTGGAGCACCTGCCTGAAA TCACTTTTTTTCAGGTTGGACCGGTGCCACC-3' (SEQ ID NO:41)
Trình tự đoạn kết thúc gen hormon sinh trưởng của bò	5'- GCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCA GCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCT TGACCCTGGAAGGTGCCACTCCCAGTGCCTTTC CTAATAAAATGAGGAAATTGCATCGCATTGTCTG AGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG GGCAGGACAGCAAGGGGGAGGATTGGGAAGAG AATAGCAGGCATGCTGGGGA-3' (SEQ ID NO:42)
Đoạn khởi đầu gen U6 snRNA của người	5'- GAGGGCCTATTTCCCATGATTCCTTCATATTTGCA TATACGATACAAGGCTGTTAGAGAGATAATTGGA ATTAATTTGACTGTAAACACAAAGATATTAGTAC AAAATACGTGACGTAGAAAGTAATAATTTCTTGG GTAGTTTGCAGTTTAAATTAATGTTTTAAATG GACTATCATATGCTTACCGTAACTTGAAAGTATT TCGATTTCTTGGCTTTATATATCTTGTGGAAAGGA CGAAACACC-3' (SEQ ID NO:43)

(2) Quá trình chỉnh sửa hệ gen trong các tế bào động vật được nuôi cấy

Trong ví dụ về các tế bào động vật được nuôi cấy, dòng tế bào có nguồn gốc từ tế bào thận phôi người (dòng tế bào HEK293) được sử dụng, và gen EMX1 được chọn là trình tự đích để đưa các đột biến vào. Đích 1 (SEQ ID NO:

44) và đích 2 (SEQ ID NO: 45) được chọn là các trình tự đích trong gen EMX1, và ADN được tổng hợp hóa học nhân tạo của đích 1 và đích 2 được gắn vào trong casset biểu hiện crARN cho các tế bào người được nuôi cấy được bào chế ở mục (1) nêu trên để tạo ra pUC_crRNA-T1 chứa đích 1 và pUC_crRNA-T2 chứa đích 2. Các plasmit được tạo ra được đem đi khuếch đại trong chủng *E. coli* HST08 (do Takara Bio Inc. sản xuất) và sau đó được tinh lọc sử dụng PureYield (nhãn hiệu thương mại đã đăng kí) Plasmid Miniprep System (do Promega Corp. sản xuất). Trong số các plasmit được tinh lọc, hỗn hợp gồm pCR_hTiD và pUC_crRNA-T1 hoặc hỗn hợp gồm pCR_hTiD và pUC_crRNA-T2 được đưa vào trong các tế bào HEK293 bằng cách chuyển nhiễm. Ba ngày sau khi đưa các vật truyền plasmit vào, các tế bào được thu gom, và ADN hệ gen được bào chế từ chúng sử dụng kit Blood & Cell Culture DNA Mini Kit (do Qiagen sản xuất). Bằng cách sử dụng ADN hệ gen được tạo ra như vậy làm khuôn mẫu, vùng trình tự hệ gen bao gồm đích 1 hoặc đích 2 được khuếch đại bằng PCR, và sự phân tích đột biến được thực hiện bằng cách phân tích tính linh hoạt của chuỗi kép dị hợp sử dụng thiết bị điện di tự động MultiNA (do Shimadzu Corporation sản xuất). Hơn nữa, đoạn PCR được khuếch đại được tách dòng vào trong vật truyền pNEB193 (do New England Biolab sản xuất), và trình tự được gây đột biến được nhận diện bằng cách phân tích xác định trình tự. Hiệu quả đột biến soma được tính dựa trên "số lượng các dòng vô tính trong đó trình tự được gây đột biến được quan sát / tổng số của các dòng vô tính được phân tích". Dòng tế bào trong đó plasmit không được đưa vào, hoặc pCR_hTiD, pUC_crRNA-T1 hoặc pUC_crRNA-T2 đã được đưa vào riêng biệt được sử dụng làm đối chứng để tiến hành phân tích đột biến theo cách thức tương tự. Sơ đồ thực nghiệm cho quá trình chỉnh sửa hệ gen sử dụng dòng tế bào HEK293 được thể hiện ở FIG. 9.

FIG. 10 và FIG. 11 thể hiện các kết quả thử nghiệm thu được khi dòng tế

bào HEK293 được chuyển nhiễm với hỗn hợp gồm pCR_hTiD và pUC_crRNA-T1 hoặc hỗn hợp gồm pCR_hTiD và pUC_crRNA-T2 hoặc khi dòng tế bào HEK293 không được chuyển nhiễm với plasmit (đối chứng). Như được thể hiện ở FIG. 10 và FIG. 11, các đỉnh biểu thị các sự đột biến được đưa vào trên trình tự đích được phát hiện trong dòng tế bào HEK293 được chuyển nhiễm với hỗn hợp gồm pCR_hTiD và pUC_crRNA-T1 hoặc hỗn hợp gồm pCR_hTiD và pUC_crRNA-T2. Mặt khác, không có đỉnh biểu thị sự đưa vào đột biến được phát hiện trong dòng tế bào trong đó plasmit không được đưa vào làm đối chứng. Tương tự với dòng tế bào trong đó plasmit không được đưa vào, không có đỉnh biểu thị sự đưa đột biến vào được phát hiện ở dòng tế bào trong đó pCR_hTiD, pUC_crRNA-T1 hoặc pUC_crRNA-T2 được đưa vào riêng biệt.

Sau đó, các mẫu trình tự trong đó đỉnh biểu thị sự đưa vào đột biến được phát hiện nhờ sự phân tích tính linh hoạt của chuỗi kép dị hợp được tách dòng vào trong vật truyền plasmit và được phân tích bằng cách xác định trình tự. Nhờ đó, như được thể hiện trong FIG. 12 và FIG 13, nhận thấy rằng các sự đột biến khuyết đoạn và/hoặc xen đoạn được đưa vào đích 1 và đích 2.

[Bảng 8]

Các trình tự đích trên gen EMX 1 của người	
Đích 1	5'-CCAGAACCGGAGGACAAAGTACAAACGGCAGAAGC-3' (SEQ ID NO:44)
Đích 2	5'-GATGTGATGGGAGCCCTTCTTCTGCTCGGACT-3' (SEQ ID NO:45)

Khả năng ứng dụng trong công nghiệp

Theo sáng chế, có thể nhằm vào trình tự gen mà không thể nhằm vào được bằng kỹ thuật chỉnh sửa hệ gen thông thường sử dụng endonucleaza được dẫn hướng bởi ARN có nguồn gốc từ CRISPR typ II hoặc typ V. Cụ thể, theo sáng chế, có thể tạo ra các alen đột biến, kiểm soát sự biểu hiện gen bằng cách kích hoạt và bất hoạt quá trình phiên mã, và thực hiện sự biến đổi hệ gen biểu sinh bằng cách nhằm vào vùng protein cải biến ADN/cải biến histon, trên các vùng gen mà không thể nhằm vào được bằng các kỹ thuật thông thường.

Danh mục trình tự

- SEQ ID NO: 1; trình tự axit amin Cas5d của *Microcystis aeruginosa*
- SEQ ID NO: 2; trình tự axit amin Cas6d của *Microcystis aeruginosa*
- SEQ ID NO: 3; trình tự axit amin Cas7d của *Microcystis aeruginosa*
- SEQ ID NO: 4; trình tự axit amin Cas3d của *Microcystis aeruginosa*
- SEQ ID NO: 5; trình tự axit amin S Cas10d của *Microcystis aeruginosa*
- EQ ID NO: 6; TiDcrRNA chứa đoạn lặp trực tiếp (37b) và đoạn đệm (35b của N). N là nucleotit bất kỳ cấu thành trình tự bổ trợ cho trình tự nucleotit đích.
- SEQ ID NO: 7; trình tự nucleotit Cas5d cho sự biểu hiện trong *Escherichia coli*
- SEQ ID NO: 8; trình tự nucleotit Cas6d cho sự biểu hiện trong *Escherichia coli*
- SEQ ID NO: 9; trình tự nucleotit Cas7d cho sự biểu hiện trong *Escherichia coli*
- SEQ ID NO: 10; trình tự nucleotit Cas3d cho sự biểu hiện trong *Escherichia coli*
- SEQ ID NO: 11; trình tự nucleotit Cas10d cho sự biểu hiện trong *Escherichia coli*
- SEQ ID NO: 12; đoạn khởi đầu tổng hợp J23108
- SEQ ID NO: 13; trình tự liên kết ribosom
- SEQ ID NO: 14; trình tự đoạn kết thúc STOP767
- SEQ ID NO: 15; trình tự đoạn kết thúc STOP768 (1)
- SEQ ID NO: 16; trình tự đoạn kết thúc TOP768 (2)
- SEQ ID NO: 17; trình tự đoạn kết thúc T7
- SEQ ID NO: 18; trình tự lặp CRISPR
- SEQ ID NO: 19; trình tự đoạn khởi đầu T7
- SEQ ID NO: 20; cassette biểu hiện crARN
- SEQ ID NO: 21; cassette biểu hiện gen ccdB tổng hợp
- SEQ ID NO: 22; trình tự axit amin của tín hiệu định vị nhân (NLS)
- SEQ ID NO: 23; trình tự nucleotit NLS
- SEQ ID NO: 24; trình tự axit amin của peptit tự cắt 2A

- SEQ ID NO: 25; trình tự mã hóa cho peptit tự cắt 2A(1)
- SEQ ID NO: 26; trình tự mã hóa cho peptit tự cắt 2A(2)
- SEQ ID NO: 27; trình tự mã hóa cho peptit tự cắt 2A(3)
- SEQ ID NO: 28; trình tự mã hóa cho peptit tự cắt 2A(4)
- SEQ ID NO: 29; 2 x đoạn khởi đầu gen 35S của virus thể khảm của cải hoa lơ + trình tự omega
- SEQ ID NO: 30; đoạn kết thúc gen 18.2kDa của protein sốc Arabidopsis
- SEQ ID NO: 31; cassette biểu hiện crARN
- SEQ ID NO: 32; trình tự đoạn khởi đầu gen U6 snRNA-26 của Arabidopsis
- SEQ ID NO: 33; 2xNLS + Cas5d
- SEQ ID NO: 34; 2xNLS + Cas6d
- SEQ ID NO: 35; 2xNLS + Cas7d
- SEQ ID NO: 36; 2xNLS + Cas3d
- SEQ ID NO: 37; 2xNLS + Cas10d
- SEQ ID NO: 38; trình tự đích 1 trên gen PDS của cây thuốc lá
- SEQ ID NO: 39; trình tự đích 2 trên gen PDS của cây thuốc lá
- SEQ ID NO: 40; trình tự đích trên gen IAA9 của cây cà chua
- SEQ ID NO: 41; đoạn khởi đầu lai gen beta-actin phổ biến của gà + gen tăng cường của Cytomegalovirus
- SEQ ID NO: 42; trình tự đoạn kết thúc gen hormon sinh trưởng có nguồn gốc từ bò
- SEQ ID NO: 43; đoạn khởi đầu gen U6 snRNA của người
- SEQ ID NO: 44; trình tự đích 1 trên gen EMX1 của người
- SEQ ID NO: 45; trình tự đích 2 trên gen EMX1 của người

YÊU CẦU BẢO HỘ

1. Phương pháp hướng đích đến trình tự nucleotit đích, phương pháp này bao gồm bước đưa vào trong tế bào:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng,

trong đó, trong trường hợp ở đó protein dung hợp của Cas5d, Cas6d hoặc Cas7d với polypeptit khác với Cas3d và Cas10d được hình thành trong tế bào, polypeptit khác với Cas3d và Cas10d được dẫn đến trình tự nucleotit đích nhờ tác dụng của Cas5d, Cas6d và Cas7d và ARN dẫn hướng, và

với điều kiện là phương pháp này không bao gồm phương pháp được tiến hành trên cơ thể người.

2. Phương pháp biến đổi trình tự nucleotit đích, phương pháp này bao gồm bước đưa vào trong tế bào:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng,

trong đó trình tự nucleotit đích là ADN sợi kép, và

với điều kiện là phương pháp này không bao gồm phương pháp được tiến hành trên cơ thể người.

3. Phương pháp theo điểm 1 hoặc 2, trong đó ARN dẫn hướng bao gồm trình tự gồm có từ 20 đến 50 nucleotit mà bổ trợ cho trình tự nucleotit đích.

4. Phương pháp theo điểm 2 hoặc 3, phương pháp này còn bao gồm bước đưa polynucleotit thể cho vào trong tế bào.

5. Phương pháp theo điểm bất kỳ trong số các điểm 2, 3 và 4, trong đó sự biến đổi là sự khuyết đoạn nucleotit, xen đoạn nucleotit, hoặc thế đoạn nucleotit.

6. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 5, trong đó Cas5d nhận diện 5'-GTH-3' (H = A, C, hoặc T) là trình tự kiểu liên kề đoạn đệm (PAM).

7. Phương pháp sản xuất tế bào chứa trình tự nucleotit đích biến đổi, phương pháp này bao gồm bước:

biến đổi trình tự nucleotit đích bằng phương pháp theo điểm 2.

8. Phương pháp theo điểm bất kỳ trong số các điểm 1, 3 và 6, trong đó phức hợp được đưa vào trong tế bào, phức hợp này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ này, và

bằng cách đó các protein Cas và ARN dẫn hướng được đưa vào trong tế bào.

9. Phương pháp theo điểm bất kỳ trong số các điểm 1, 3 và 6, trong đó vật truyền biểu hiện được đưa vào trong tế bào, vật truyền biểu hiện này bao gồm:

(i) các axit nucleic mã hóa các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ADN mã hóa ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ này, và

bằng cách đó các axit nucleic mã hóa các protein Cas và ADN mã hóa ARN dẫn hướng được đưa vào trong tế bào.

10. Phương pháp theo điểm bất kỳ trong số các điểm từ 2 đến 7, trong đó phức hợp được đưa vào trong tế bào, phức hợp này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, và

bằng cách đó các protein Cas và ARN dẫn hướng được đưa vào trong tế bào.

11. Phương pháp theo điểm bất kỳ trong số các điểm từ 2 đến 7, trong đó vật truyền biểu hiện được đưa vào trong tế bào, vật truyền biểu hiện này bao gồm:

(i) các axit nucleic mã hóa các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và

(ii) ADN mã hóa ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, và

bằng cách đó các axit nucleic mã hóa các protein Cas và ADN mã hóa ARN dẫn hướng được đưa vào trong tế bào.

12. Chế phẩm hướng đích đến trình tự nucleotit đích, chế phẩm này chứa phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ,

trong đó, trong trường hợp ở đó protein dung hợp của Cas5d, Cas6d hoặc Cas7d với polypeptit khác với Cas3d và Cas10d được tạo ra trong khi hướng đích, polypeptit khác với Cas3d và Cas10d được dẫn hướng đến trình tự nucleotit đích nhờ tác dụng của Cas5d, Cas6d và Cas7d và ARN dẫn hướng.

13. Chế phẩm hướng đích đến trình tự nucleotit đích, chế phẩm này chứa vật truyền biểu hiện bao gồm:

(i) các axit nucleic mã hóa các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

(ii) ADN mã hóa ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ, và

trong đó, trong trường hợp ở đó protein dung hợp của Cas5d, Cas6d hoặc Cas7d với polypeptit khác với Cas3d và Cas10d được tạo ra trong khi hướng đích, polypeptit khác với Cas3d và Cas10d được dẫn hướng đến trình tự nucleotit đích nhờ tác dụng của Cas5d, Cas6d và Cas7d và ARN dẫn hướng.

14. Chế phẩm biến đổi trình tự nucleotit đích, chế phẩm này chứa phức hợp bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và

(ii) ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ này,

trong đó trình tự nucleotit đích là ADN sợi kép.

15. Chế phẩm biến đổi trình tự nucleotit đích, chế phẩm này chứa vật truyền biểu hiện bao gồm:

(i) các axit nucleic mã hóa các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và

(ii) ADN mã hóa ARN dẫn hướng bao gồm trình tự hỗ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự hỗ trợ này,

trong đó trình tự nucleotit đích là ADN sợi kép.

16. Chế phẩm hướng đích đến trình tự nucleotit đích, chế phẩm này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng.

17. Kit hướng đích đến trình tự nucleotit đích, kit này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng.

18. Chế phẩm biến đổi hoặc phân cắt trình tự nucleotit đích, chế phẩm này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng, trong đó trình tự nucleotit đích là ADN sợi kép.

19. Kit biến đổi hoặc phân cắt trình tự nucleotit đích, kit này bao gồm:

(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, hoặc các axit nucleic mã hóa các protein này, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, hoặc ADN mã hóa ARN dẫn hướng, trong đó trình tự nucleotit đích là ADN sợi kép.

20. Chế phẩm hoặc kit theo điểm bất kỳ trong số các điểm từ 12 đến 19, trong đó ARN dẫn hướng bao gồm trình tự gồm có 20 đến 50 nucleotit mà bổ trợ cho trình tự nucleotit đích.

21. Phương pháp *in vitro* hướng đích đến trình tự nucleotit đích trong axit nucleic, phương pháp này bao gồm bước sử dụng

(i) các protein liên kết CRISPR typ I-D Cas5d, Cas6d và Cas7d, và

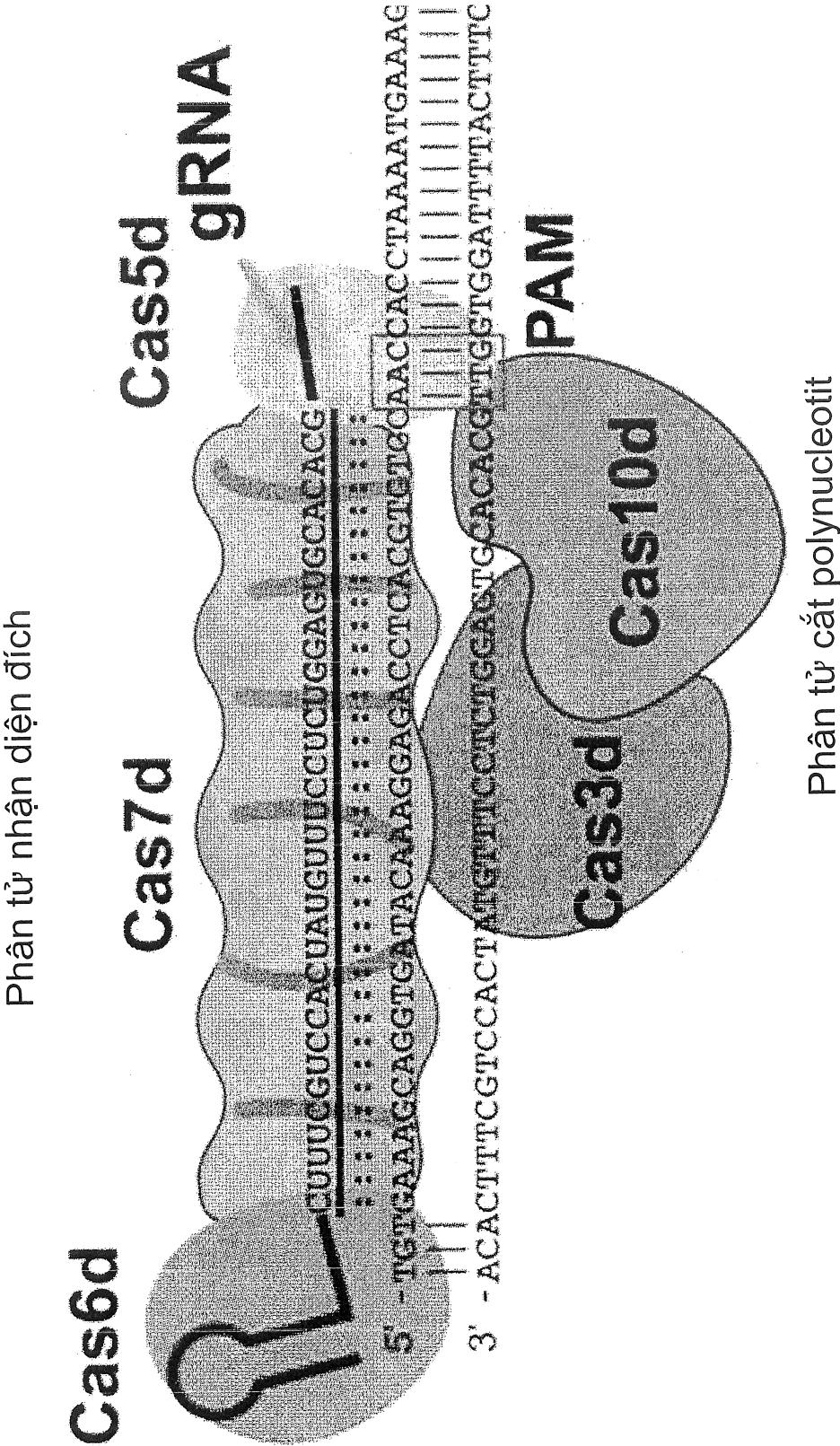
(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này.

22. Phương pháp *in vitro* phân cắt trình tự nucleotit đích trong axit nucleic, phương pháp này bao gồm bước sử dụng

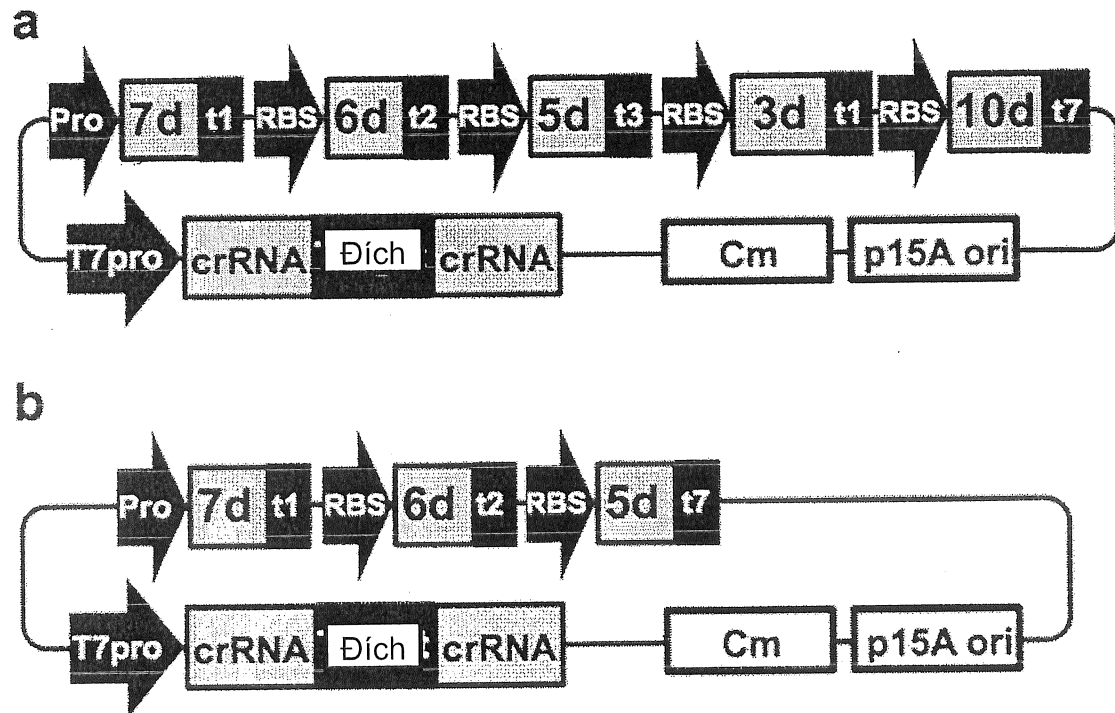
(i) các protein liên kết CRISPR typ I-D Cas3d, Cas5d, Cas6d, Cas7d và Cas10d, và

(ii) ARN dẫn hướng bao gồm trình tự bổ trợ cho trình tự nucleotit đích, và các trình tự lặp chung có nguồn gốc từ locus CRISPR, trước và sau trình tự bổ trợ này, trong đó trình tự nucleotit đích là ADN sợi kép.

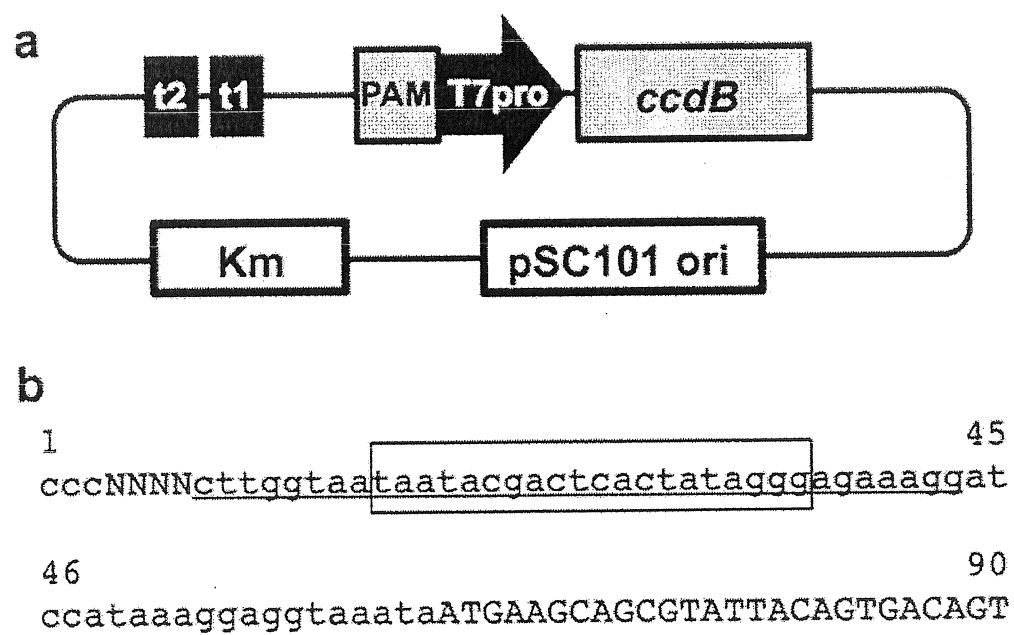
[Fig. 1]



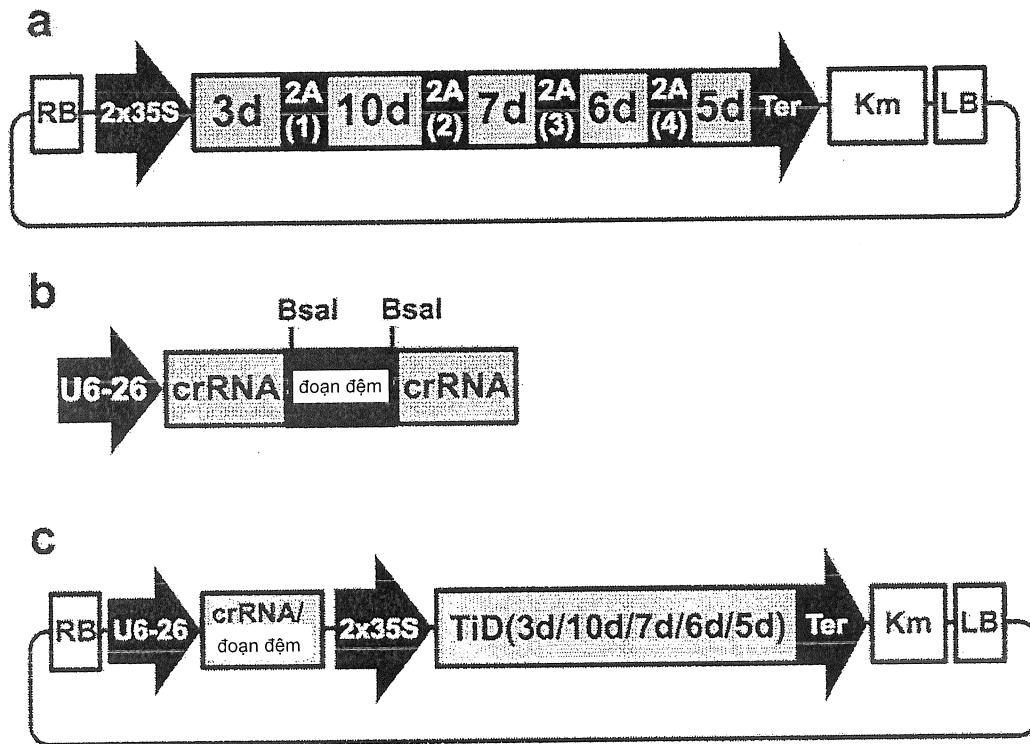
[FIG. 2]



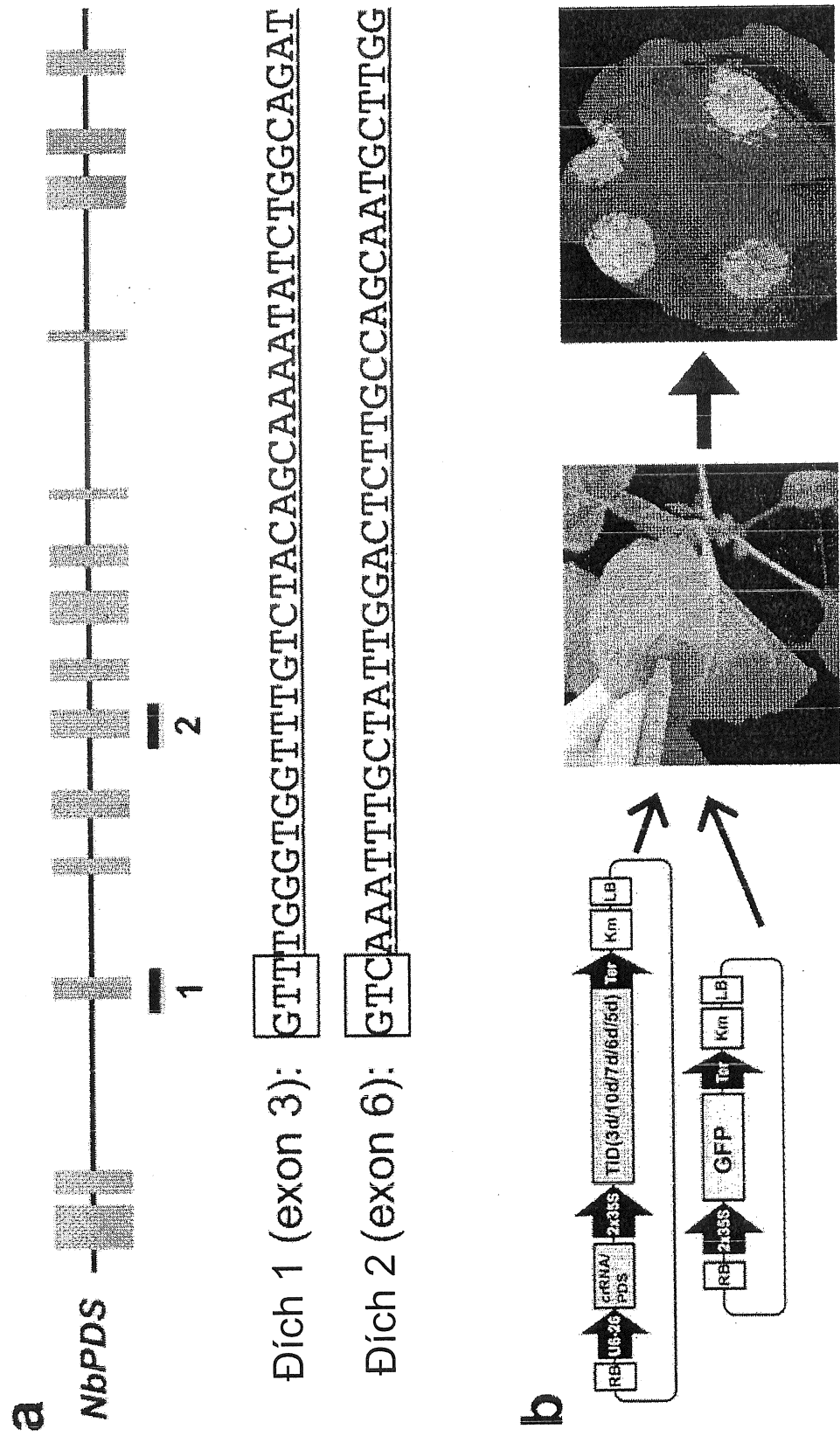
[FIG. 3]



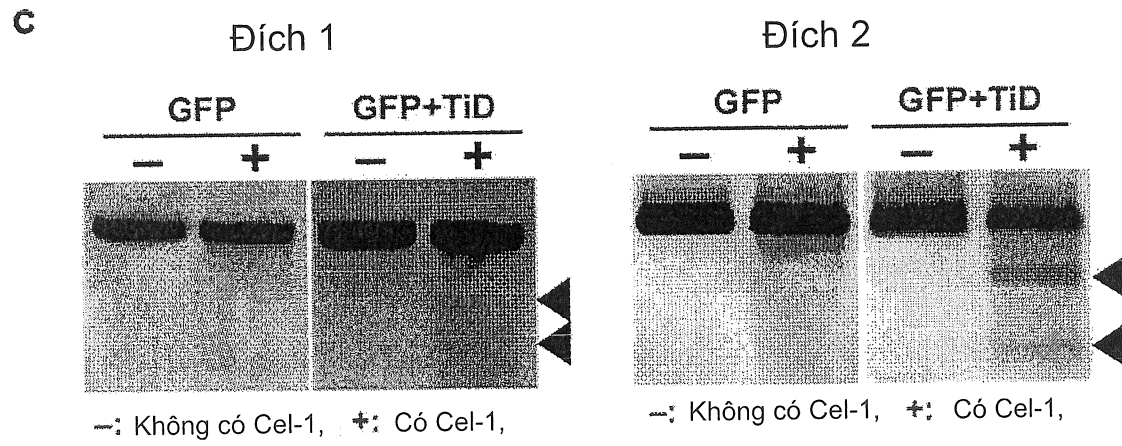
[FIG. 4]



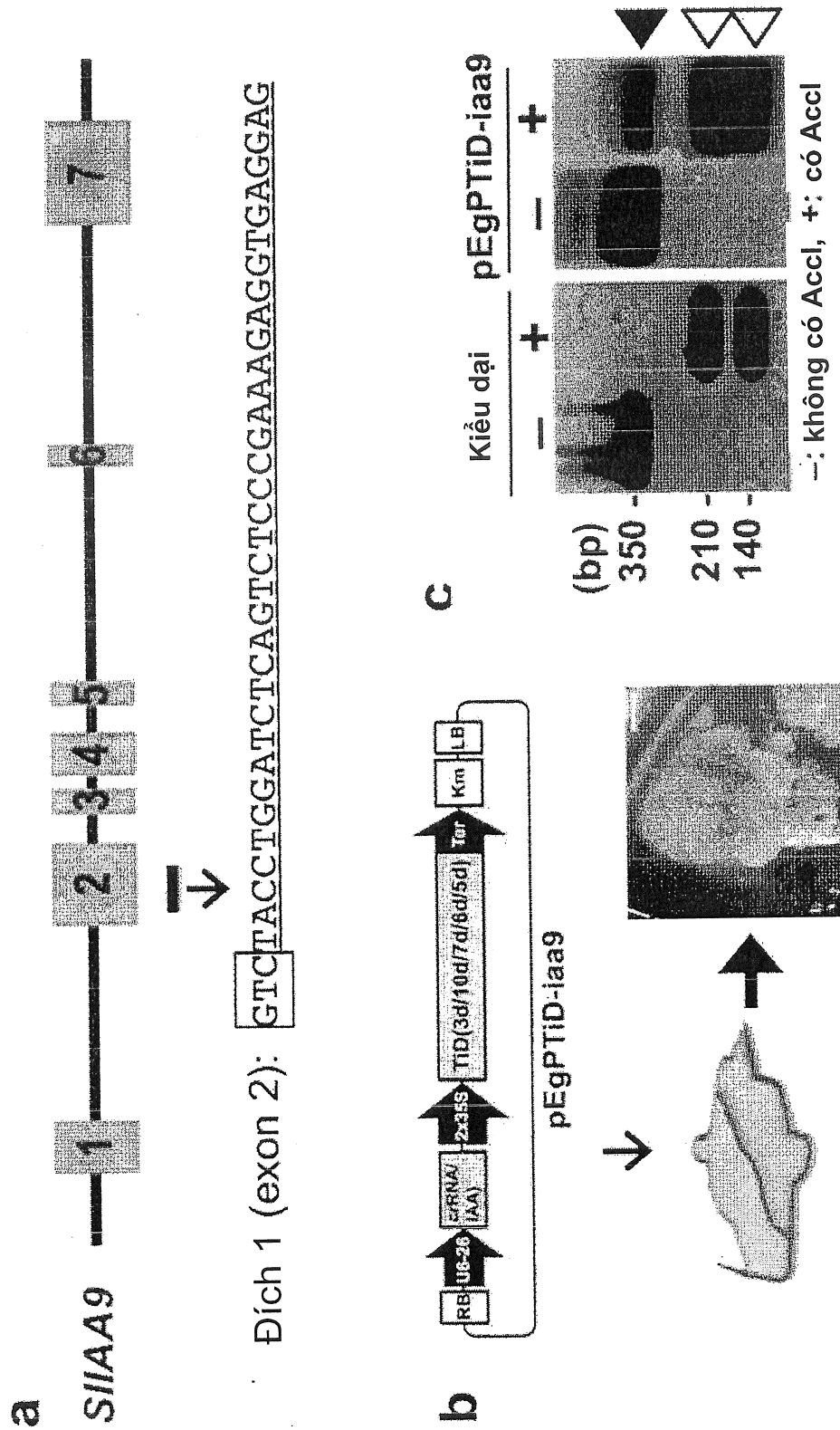
[FIG.5-1]



[FIG. 5-2]

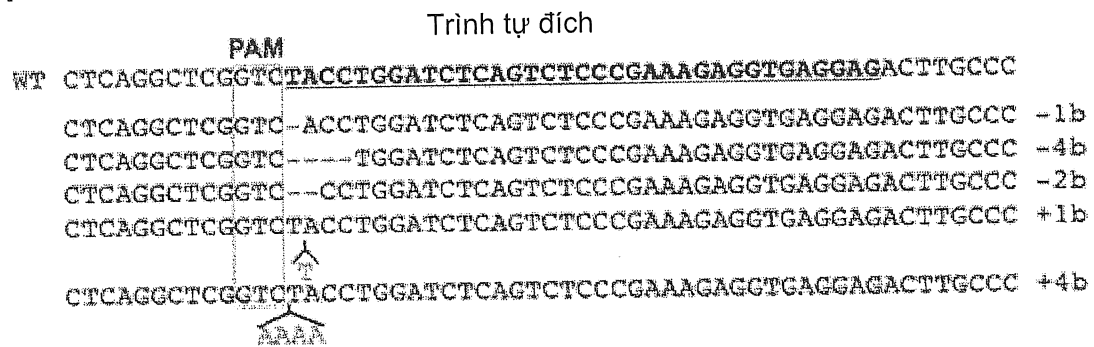


[FIG. 6]

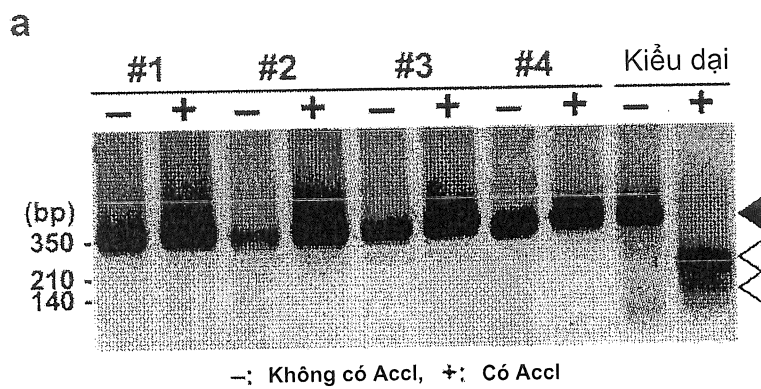
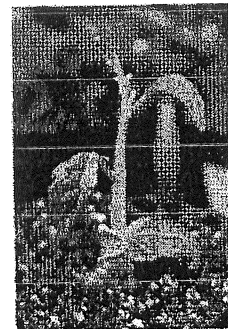


7/12

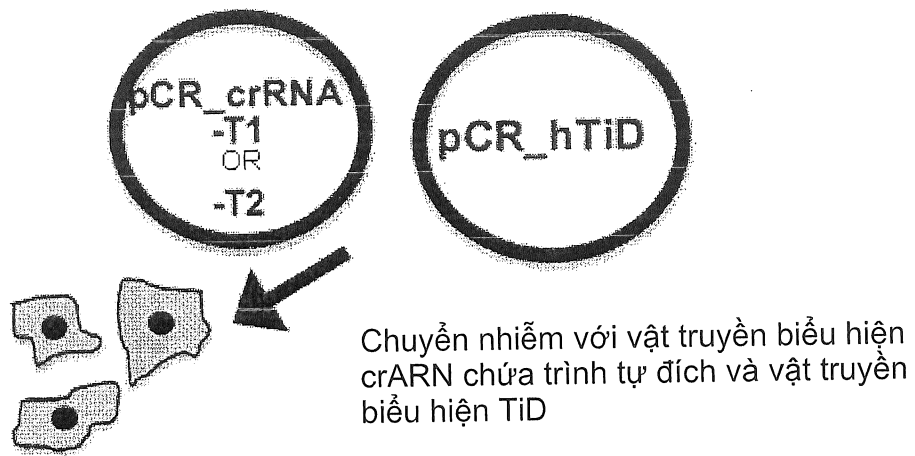
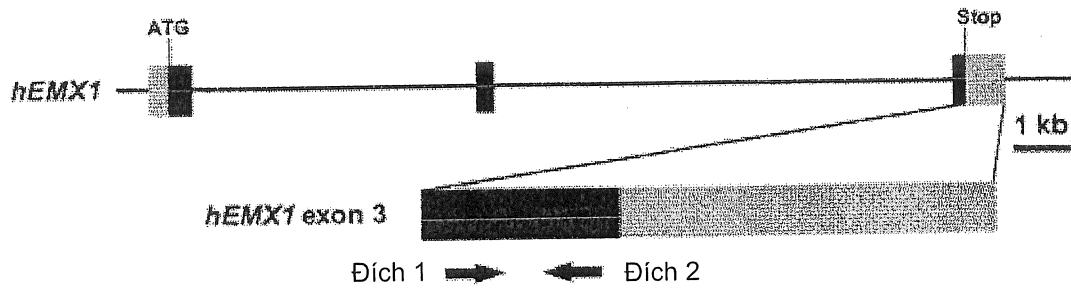
[FIG. 7]



[FIG. 8]

**b**

[FIG. 9]



Các tế bào được nuôi cấy **HEK293**

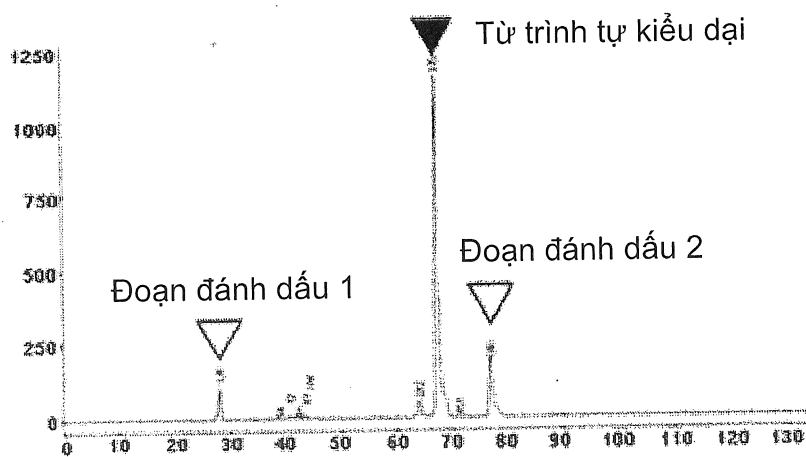
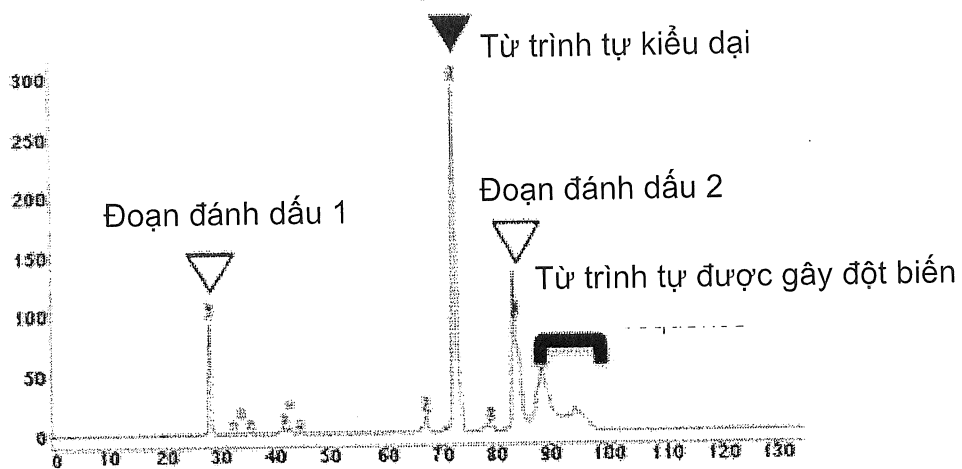
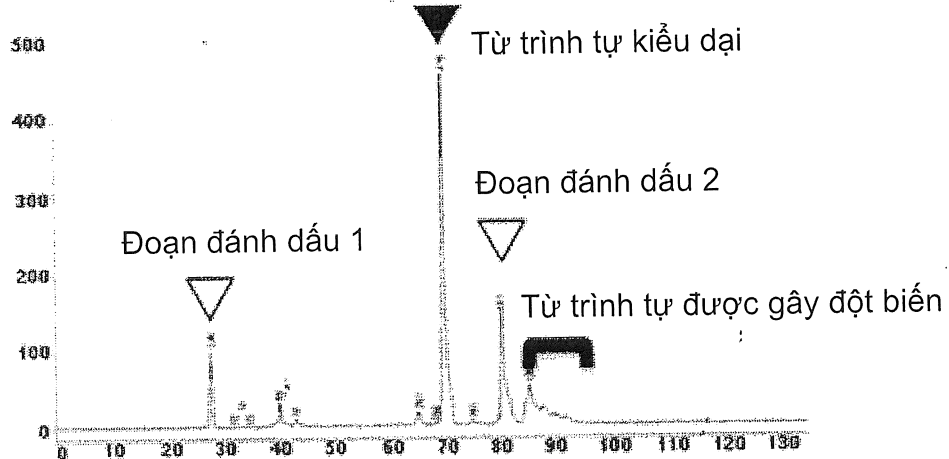


Bào chế ADN hệ gen từ các tế bào ở 3 ngày sau khi đưa gen vào

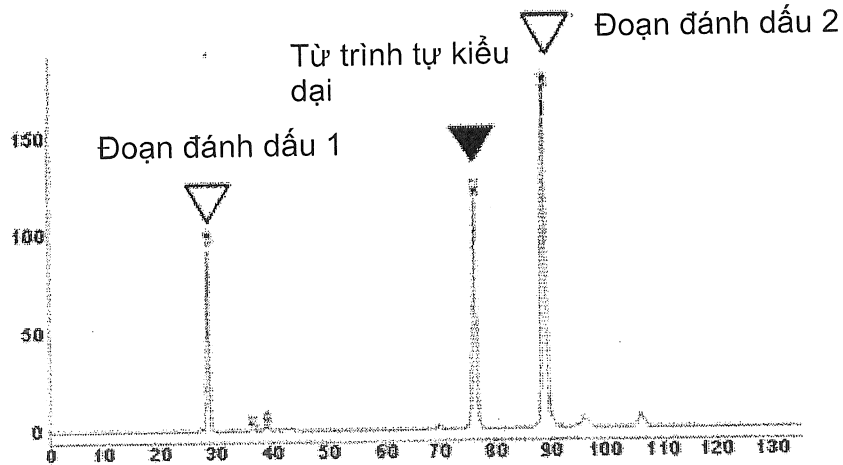
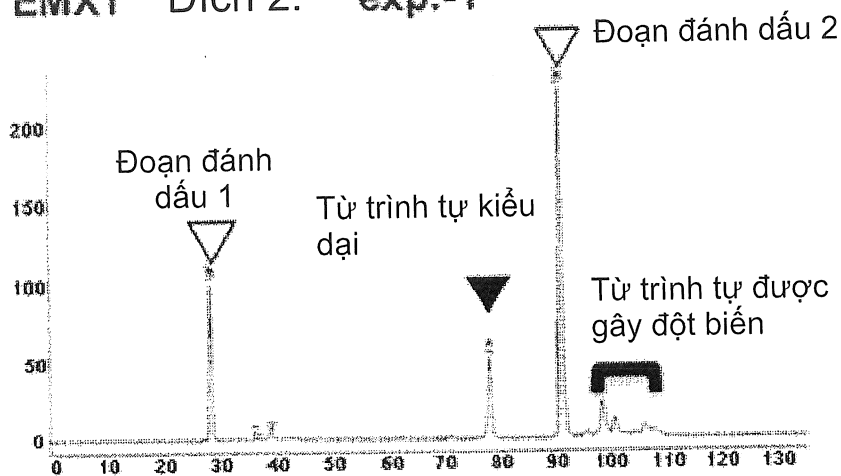
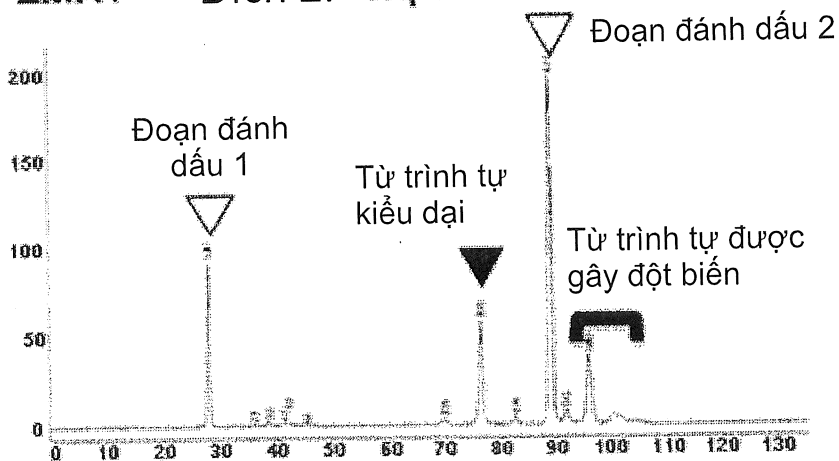


Tiến hành phân tích tính linh động của chuỗi kép dị hợp sử dụng thiết bị điện di tự động và sự phân tích xác định trình tự

[FIG.10]

EMX1 Đích 1: Đối chứng**EMX1 Đích 1: exp.-1****EMX1 Đích 1: exp.-2**

[FIG. 11]

EMX1 Đích 2: Đối chứng**EMX1 Đích 2: exp.-1****EMX1 Đích 2: exp.-2**

11/12

[FIG. 12]

EMX1 Dích 1: exp.-1

Dích 1

PAM

WT 5'-GGTGTG**GTTC**CCAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3'

Mut1 5'-GGTGTGGT--CAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (3/46)

Mut2 5'-GGTGTGGTT-CAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (2/46)

Mut3 5'-GGTGTGGT-----AACGGCAGAAGCTGGA-3' (1/46)

Mut4 5'-GGTGTGGA-----GGACAAAGTACAAACGGCAGAAGCTGGA-3' (1/46)

Mut5 5'-GGTGTG-----AGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (1/46)

Mut6 5'-GGTGTGGT-----AACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (1/46)

EMX1 Dích 1: exp.-2

Dích 1

PAM

WT 5'-GGTGTG**GTTC**CCAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3'

Mut1 5'-GGTGTGGTTC--GAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (2/30)

Mut2 5'-GGTGTGGTT-CAGAACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (1/30)

Mut3 5'-GGTGTGGT-----AACGGCAGAAGCTGGA-3' (1/30)

Mut4 5'-GGTGTGG-----AACCGGAGGACAAAGTACAAACGGCAGAAGCTGGA-3' (1/30)

12/12

[FIG. 13]

EMX1 Dích 2: exp.-1

Dích 2 PAM

WT 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGC-3'

Mut1 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT-AACCGGTGGC-3' (1/40)

Mut2 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACA--AACCGGTGGC-3' (1/40)

Mut3 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCC-----ACCGGTGGC-3' (1/40)

Mut4 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT---CCGGTGGC-3' (1/40)

Mut5 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACA---ttCGGTGGC-3' (1/40)

EMX1 Dích 2: exp.-2

Dích 2 PAM

WT 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGC-3'

Mut1 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT-AACCGGTGGC-3' (1/36)

Mut2 5'-CTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCAC---AACCGGTGGC-3' (1/36)

Mut3 5'-CTGAGTCCGAGCAGAAGA-----ACCGGTGGC-3' (1/36)

agcgggaactccgaaaggaggagacgagctcatccctcgggccttacttgcagctcgca

DANH MỤC TRÌNH TỰ

<110> Tokushima University

<120> PHUONG PHÁP HƯỚNG ĐÍCH, PHUONG PHÁP BIẾN ĐỔI TRÌNH TỰ NUCLEOTIT ĐÍCH, CHẾ PHẨM VÀ KIT HƯỚNG ĐÍCH ĐẾN TRÌNH TỰ NUCLEOTIT ĐÍCH

<130> 674277

<150> JP2017-158876

<151> 2017-08-21

<150> JP2017-236518

<151> 2017-12-08

<160> 45

<170> PatentIn version 3.5

<210> 1

<211> 224

<212> PRT

<213> Microcystis aeruginosa

<400> 1

Met	Val	His	Ile	Tyr	Ser	Cys	Gln	Leu	Glu	Leu	His	Asp	Ser	Leu	Tyr
1				5					10					15	

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

His	Asn	Tyr	Ala	Leu	Cys	Tyr	Ala	Leu	Gly	Leu	Val	Asn	Ser	Asp	Ser
		35					40					45			

Tyr	Arg	Tyr	Phe	Cys	Ser	Glu	Gln	Ile	Pro	Gln	Tyr	Gln	Glu	His	Leu
	50					55					60				

Asn	Pro	Leu	Asn	Glu	Glu	Lys	Ile	Tyr	Val	Thr	Pro	Ala	Arg	Ala	Ile
65				70						75					80

Ala	His	Thr	Ala	Val	Leu	Asn	Thr	Trp	Lys	Tyr	Ala	Asn	Asn	Asn	Tyr
				85					90					95	

His	Val	Glu	Met	Glu	Lys	Thr	Gln	Lys	Asn	Ile	Pro	Ser	Phe	Gly	Arg
			100					105						110	

Ala Lys Glu Ile Ala Pro Glu Ser Ile Phe Glu Cys Phe Ile Ile Ser
 115 120 125

His His Pro Leu Gln Leu Pro Lys Trp Ile Arg Leu Gly Lys Trp Met
 130 135 140

Ser Lys Ala Glu Val Lys Leu Thr Glu Leu Ser Leu Ser Lys Gln Lys
 145 150 155 160

Glu Asp Leu Phe Ile Tyr Pro Tyr Pro Leu Asn Pro Leu Asp Val Met
 165 170 175

Phe Thr His Gln Val Ile Gly Tyr Asp Val Ile Asn Met Pro Pro Val
 180 185 190

Ser Leu Ile Arg Asn Val Arg Met Arg Gly Glu Tyr Tyr Gln Ile Ser
 195 200 205

Asp Arg Pro Asp Leu Lys Ile Pro Ala Arg Leu Ser Tyr His Phe Gly
 210 215 220

<210> 2

<211> 277

<212> PRT

<213> Microcystis aeruginosa

<400> 2

Met Pro Tyr Ser Leu Val Leu Asn Leu Thr Pro Arg Ser Pro Ile Tyr
 1 5 10 15

Pro Asn Phe Leu Thr Gly Arg His Leu His Ala Leu Phe Leu Thr Leu
 20 25 30

Val Ser Ser Val Asp Gln Glu Leu Gly Asn Ile Leu His Thr Ala Glu
 35 40 45

Ala Asp Lys Ala Phe Thr Leu Ser Pro Leu Gln Met Gln Ser Gly Gly
 50 55 60

Lys Thr Ile Asn Ser Pro Gln Trp Arg His Glu Arg Glu Ile Ala Ser
65 70 75 80

Glu Thr Pro Cys Trp Trp Arg Ile Ser Leu Leu Asp Asp Arg Leu Phe
85 90 95

Gly Lys Leu Thr Ser Leu Trp Leu Asn Leu Asn Pro Lys Gln Pro Trp
100 105 110

His Leu Gly Ser Ala Asp Leu Val Ile Thr Ser Val Leu Ala Thr Pro
115 120 125

Gln Ser Val Gln Pro Trp Ala Asn Ser Cys Thr Tyr Gln Tyr Leu Tyr
130 135 140

Glu Asn Ala Ser Glu Thr Asn Arg Glu Phe Asp Phe Leu Phe Ala Thr
145 150 155 160

Pro Val Thr Phe Arg Gln Gly Lys Phe Asp Ser Ala Leu Pro Thr Arg
165 170 175

Glu Leu Val Phe Asn Ser Leu Leu Gly Arg Trp Asn Arg Tyr Ser Gly
180 185 190

Ile Pro Phe Asp Ser Ile Ala Leu Glu Ser Ile Phe Pro Ser Phe Phe
195 200 205

Asp Ile Gln Thr Lys Leu Ala Asp Glu Ala Tyr Lys Asn Gln Ser Ile
210 215 220

Gly Cys Val Gly Glu Ile His Tyr Arg Leu Leu Gly Glu Val Glu Pro
225 230 235 240

Ala Lys Ile Lys Ala Ile Asn Ala Leu Ala Asp Phe Ala Leu Tyr Ala
245 250 255

Gly Val Gly Arg Lys Thr Thr Met Gly Met Gly Met Thr Arg Arg Ile
260 265 270

Ser Lys Asp Lys Arg
275

<210> 3
<211> 332
<212> PRT
<213> Microcystis aeruginosa

<400> 3

Thr Phe Leu Thr Ser Val Asp Ala Lys Phe Phe His Ser Glu Ile Pro
1 5 10 15

Tyr Lys Pro Met Gly Lys Tyr Val His Phe Leu Thr Ile Arg Val Thr
20 25 30

Glu Ser Tyr Pro Leu Phe Gln Thr Asp Gly Glu Leu Asn Lys Ala Arg
35 40 45

Val Arg Ala Gly Ile Asp Ser Lys Lys Thr Ile Ser Arg Leu Ser Met
50 55 60

Phe Lys Arg Lys Gln Ser Thr Pro Glu Arg Leu Val Gly Arg Glu Leu
65 70 75 80

Leu Arg Asn Tyr Gly Leu Ile Thr Ala Glu Glu Cys Glu Tyr Asn Val
85 90 95

Lys Phe Ala Met Asn Asn Ala Asp Cys Ile Ile Tyr Gly Phe Ala Ile
100 105 110

Gly Asp Ser Gly Ser Glu Lys Ser Lys Val Val Val Asp Thr Ala Phe
115 120 125

Ser Ile Thr Pro Phe Asp Glu Ser His Glu Ser Phe Thr Leu Asn Ala
130 135 140

Pro Tyr Glu Asn Gly Thr Met Ala Ser Lys Gly Glu Asn Asn Thr Lys
145 150 155 160

Val Gly Glu Val Thr Ser Arg Ile Asn Gln Gln Asp His Ile Arg Pro

				165						170						175
Gln	Val	Phe	Phe	Pro	Ser	Ile	Val	Thr	Leu	Lys	Asp	Pro	Thr	Glu	Ala	
			180					185					190			
Ser	Phe	Leu	Tyr	Val	Phe	Asn	Asn	Ile	Leu	Arg	Thr	Arg	His	Tyr	Gly	
		195					200					205				
Ala	Gln	Thr	Thr	Arg	Thr	Gly	Arg	Val	Arg	Asn	Glu	Leu	Ile	Gly	Val	
	210					215					220					
Ile	Phe	Ala	Asp	Gly	Glu	Ile	Val	Ser	Asn	Leu	Arg	Trp	Thr	Gln	Ala	
225					230					235					240	
Ile	Tyr	Asp	Arg	Leu	Pro	Asp	Glu	Val	Leu	His	Ser	Ile	Asp	Pro	Leu	
				245					250					255		
Asp	Glu	Asp	Leu	Val	Met	Glu	Lys	Ala	Thr	Glu	Ala	Ile	Gln	Ala	Leu	
			260					265					270			
Met	Ala	Glu	Glu	Phe	Ile	Val	His	Thr	Asp	Phe	Ile	Gly	Glu	Asn	Phe	
		275					280					285				
Gln	Pro	Leu	Leu	Thr	Glu	Val	Lys	Thr	Leu	Thr	Gly	Thr	Glu	Ala	Gly	
	290					295					300					
Ile	Leu	Ser	Val	Leu	Asp	Gln	Ala	Asn	Lys	Glu	Ser	Lys	Lys	Tyr	Phe	
305					310					315					320	
Glu	Gln	Tyr	Ile	Glu	Lys	Lys	Lys	Ala	Glu	Lys	Lys					
				325					330							
<210>	4															
<211>	721															
<212>	PRT															
<213>	Microcystis aeruginosa															
<400>	4															
Met	Gly	Asn	Tyr	Gln	Val	Thr	Leu	Lys	Pro	Val	Tyr	Ser	Cys	Pro	Ala	
1				5					10					15		

Asp Glu Ile Pro Asp Gly Ile Lys Val Pro Gln Gly Trp Arg Leu Ser
 20 25 30

Trp His Gln Val Glu Thr Trp Lys Ala Leu Asn Asp Pro Asp Ile Asp
 35 40 45

Val Ile Phe Asn Thr Ala Met Thr Gly Asp Gly Lys Ser Leu Ala Ala
 50 55 60

Tyr Leu Arg Thr Leu Gln Gly Tyr Phe Pro Ile Met Gly Leu Tyr Pro
 65 70 75 80

Thr Asn Glu Leu Ala Arg Asp Gln Arg Gly Gln Ile Glu Ala Tyr Ile
 85 90 95

Gln Arg Phe Gln Pro Thr Asp Gln Pro Arg Val Asn Leu Leu Thr Gly
 100 105 110

Pro Glu Leu Glu Leu Tyr Ala Glu Arg Asp Gly Lys Thr Lys Ala Ile
 115 120 125

Ala Leu Glu Thr Arg Ser Lys Gln Ser Glu Ile Leu Leu Thr Asn Pro
 130 135 140

Asp Ile Phe His Tyr Leu His Arg Ala Ala Tyr Leu Thr Pro Tyr Asp
 145 150 155 160

Asn Pro Asp Gln Leu Trp Asn Arg Ile Asp Lys His Phe Asp Leu Phe
 165 170 175

Leu Phe Asp Glu Phe His Val Phe Gly Thr Pro Gln Val Ala Ser Ile
 180 185 190

Ile Asn Thr Met Leu Leu Ile Arg Arg Ala Asn Arg Gly Lys Arg Tyr
 195 200 205

Leu Phe Leu Ser Ala Thr Pro Asp Glu Gly Leu Leu Lys Arg Leu Asp
 210 215 220

Lys Ala Gly Phe Arg Tyr Arg Ser Ile Asp Pro Val Arg Glu Gly Lys
 225 230 235 240

Tyr Arg Phe Pro Asp Thr Pro Glu Glu Ala Asn Ser Leu Ala Gln Gln
 245 250 255

Gly Trp Arg Gln Val Thr Ser Glu Ile Glu Leu Ser Phe Ile Pro Leu
 260 265 270

Pro Ser Ser Phe Gln Thr Ser Glu Asn Trp Leu Lys Glu Asn Lys Glu
 275 280 285

Arg Ile Leu Asp Tyr Phe Lys Arg Tyr Pro Gly Ser Lys Gly Ala Ile
 290 295 300

Ile Leu Asn Ser Ile Ala Ser Val Lys Arg Leu Leu Pro Ile Phe Arg
 305 310 315 320

Glu Leu Leu Ala Thr Ile Gly Leu Thr Val Gly Glu Asn Thr Gly Leu
 325 330 335

Ser Gly Thr Arg Glu Lys Leu Ala Ser Leu Asn Arg Asp Leu Val Ile
 340 345 350

Gly Thr Ser Thr Ile Asp Val Gly Val Asp Phe Lys Ile Asn Phe Leu
 355 360 365

Ile Phe Glu Ser Ser Asp Ala Gly Asn Phe Ile Gln Arg Phe Gly Arg
 370 375 380

Leu Gly Arg His Ser Gly Tyr Asp Arg Lys Gly Thr Ala Val Lys Phe
 385 390 395 400

Thr Asn Phe Thr Ala Ile Ala Leu Val Pro Lys Phe Phe Leu Glu Arg
 405 410 415

Leu Phe Glu Lys Lys Asp Ala Pro Leu Gln Val Gly Glu Arg Tyr Asp
 420 425 430

Arg Ile Gln Leu Gln Glu Ala Ile Lys Ser Asn Tyr Arg His Ile Asn
 435 440 445

Asn Phe Glu Gly Tyr Tyr Gln Arg Trp Gly Ala Val Gln Ser Phe Gln
 450 455 460

Leu Trp Trp Asn Leu Gly Ser Pro Lys Ile Lys Ser Gln Tyr Gly Glu
 465 470 475 480

Ser Arg Gln Lys Phe Gln Gln Glu Cys Glu Glu Val Phe Asp Thr Ser
 485 490 495

Leu Lys Arg Val Ala Gly Arg Val Lys Gly Trp Ala Asp Glu Trp Lys
 500 505 510

Glu Leu Ser Gly Lys Asn Gly Asn Pro Ile Phe Glu Asp Ala Ser Ser
 515 520 525

Phe Arg Gly Ser Ser Pro Leu Leu Cys Gly Leu Tyr Asp Ser Thr Glu
 530 535 540

Pro Glu Glu Cys Asp Arg Phe Lys Thr Tyr Asp Leu Pro Ser Ile Leu
 545 550 555 560

Gly Asn Leu Glu Val Glu Val Trp Arg Lys Gly Glu Phe Lys Arg Gln
 565 570 575

Ile Glu Ala Thr Lys Thr Pro Ile Ala Arg Arg Arg Phe Asp Tyr Cys
 580 585 590

Leu Ala Phe Leu Asn Leu Lys Gly Tyr Arg Glu Glu Arg Leu Asn Trp
 595 600 605

Arg Phe Thr Tyr Asp Gly Asp Leu Gly Glu Ile Ala Ser Ala Trp Lys
 610 615 620

Val Gln Val Leu Thr Gly Ile Gly Val Trp Gln Pro Asp Asn Pro Trp
 625 630 635 640

Leu Asp Arg Ile Ser Arg Glu Leu Arg Asp Leu Ala Leu Val Ser Phe
 645 650 655

Val Phe Ala Tyr Pro Val Ala Ala Val Arg Gln Arg Leu Gln Leu Pro
 660 665 670

Met His Phe Gly Ile Tyr Pro Ile Ser Asp Glu Ser Ser Leu His Ser
 675 680 685

Pro Leu Ser Pro Tyr Ser Ile Ala Ile Gly Gln Ala Ala Leu Leu Leu
 690 695 700

Asp Thr Leu Ala His Arg Phe Lys Gly Lys Gly Gly Glu Val Trp Ile
 705 710 715 720

Cys

<210> 5
 <211> 1156
 <212> PRT
 <213> Microcystis aeruginosa

<400> 5

Met Pro Lys Lys Gln Lys Lys Leu Glu Glu Thr Gly Gln Leu Asn Leu
 1 5 10 15

Phe Asp Asn Thr Thr Glu Ile Asp Asp Glu Asp Leu Asp Phe Glu Phe
 20 25 30

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

Glu Ser Val Ser Asp Arg Arg Val Glu Thr Val Arg Gln Leu Leu Thr
 50 55 60

Leu Lys Leu Leu Arg Glu Ala Ile Arg Ala Glu Asn Pro Asp Asp Arg
 65 70 75 80

Val Met Ala Asp Phe Ala Glu Met Val Leu Pro Asn Leu Leu Arg Leu
85 90 95

Ala Ile Gly Val Thr Ala Lys Gly Gly Asn Phe Ile Glu Ala Val Asp
100 105 110

Arg Gly Arg Glu Leu Arg Asn Lys Pro Lys Ala Lys Arg Asp Asn Ala
115 120 125

Gly Asp Gln Ser Leu Asn Thr His Leu Leu Asn Gly Leu Phe Pro Ala
130 135 140

Asn Leu Ile Glu Lys Arg Leu Gln Lys Leu Asn Thr Thr Val Arg Arg
145 150 155 160

Ile Ile Lys Glu Phe Glu Arg Arg Leu Ala Ile Ala Gly Phe Leu Val
165 170 175

His Asp Phe Glu Lys Phe Ser Tyr Asp Arg Phe Pro Ser Met Ser Glu
180 185 190

Arg Tyr Ile Gln Ile Gln Arg Asp Phe Ile Gln Asp Pro Phe Lys Asn
195 200 205

Gln Asp Pro Arg Lys Leu Ser Arg Glu Glu His Arg Glu Ile Leu Gln
210 215 220

Val Leu Ile Pro Glu Leu Gly Leu Asp Arg Phe Leu Phe Pro Asp Asn
225 230 235 240

Pro Glu Arg Trp Leu Glu Tyr Leu Asp Asp Leu Leu Tyr Ile Ala Lys
245 250 255

Asn Thr Gln Arg Arg Asn Asp Thr Asp Leu Asn Thr Ser Glu Asp Gly
260 265 270

Leu Asn Val Arg Leu Asn Asp Arg Val Ile Glu Ser Leu Cys Asp Leu
275 280 285

Ala Cys Leu Ala Asp Arg Leu Ala Ser Ile Ile Lys His Pro His Asp
 290 295 300

Ala Glu Lys Ala Ser Leu Gln Asp Leu Leu Tyr Ser Leu Ser Asp Gly
 305 310 315 320

Glu Leu Lys Phe Thr Tyr His Ser Ile Ala Glu Asn Arg Gly Val Leu
 325 330 335

Thr Asn Val Leu Asn Asn Ala Val Met Glu Ala His Gln Glu Leu Asp
 340 345 350

Tyr Gln Pro Leu Leu Tyr Leu Pro Thr Gly Val Val Tyr Ile Ala Pro
 355 360 365

Lys Asn Ala Pro Glu Val Ser Leu Glu Thr Leu Pro Asn Arg Val Val
 370 375 380

Asp Thr Ile Lys Ser Leu Cys Ser Gly Glu Leu Gln Arg Lys Gln Thr
 385 390 395 400

Gly Phe Gly Arg Asp Gly Lys Gly Met Lys Tyr Ala Asp Tyr Tyr Ser
 405 410 415

Gln Phe Phe Asp Asp Ala Gly Leu Met Arg Ala Ala Leu Asn Ala Thr
 420 425 430

Leu Arg Ile Leu Gly Asp Asn Lys Ala Ser Val Ala Arg Ser Arg Gly
 435 440 445

Glu Asn Leu Ile Lys Phe Gln Gln Gln Gly Val Leu Pro Thr Asp Tyr
 450 455 460

Asp Phe His Cys Glu Asp Asp Ile Arg Ile Asp Arg Leu Ala Glu Phe
 465 470 475 480

Gly Asp Val Val Thr Arg Lys Ile Trp Gly Asp Arg Leu Glu Lys Ile
 485 490 495

Glu Gln Ala Arg Lys Leu Gln Lys Asn Leu Pro Ala Pro Pro Asp Leu
500 505 510

Asp Leu Ile Ser Glu Ile Ala His Tyr Trp Asn Leu Glu Asn Tyr Leu
515 520 525

Pro Gln Ile Arg Ala Ile Lys Arg Ile Asn Glu Ser Leu Lys Glu Leu
530 535 540

Lys Leu Lys Gly Asn Thr Gly Gly Val Pro Tyr Glu Trp Tyr Tyr Leu
545 550 555 560

Ala Ala Gln Tyr Leu Lys Gln His Pro Gly Ile Glu Asp Ile Arg Pro
565 570 575

Val Ala Glu Asp Leu Ile Ala Phe Leu Ala Ala Lys Ile Ala Ala Ile
580 585 590

Val Ala Gly Tyr Asn Leu Pro Asp Gly Trp Glu Asp Leu Arg Glu Trp
595 600 605

Val Asn Gln Val Val Gln Leu Pro Gly Arg Glu Leu Ala His Ser Ile
610 615 620

Glu Thr Phe Gln Lys Glu Leu Asn His Tyr Asn Ala Ala Lys Lys Gln
625 630 635 640

Gly Arg Gly Arg Gln Leu Leu Cys Ser Ile Ser His Ser Pro Tyr Ser
645 650 655

Val Ser Glu Gln Met Glu Ser Ala Val Leu Phe Thr Pro Gln Val Tyr
660 665 670

Thr Asn Lys Gln Met Leu Ala Gly Ser Asn Ala Lys Arg Asn Ile Ser
675 680 685

Ser Ile Ala Gly Thr Glu Met Met Leu Arg Gln Ile Leu Met Asn Gln
690 695 700

Thr Gln Ala Val Gly Lys Arg Phe Glu Asp Gly Lys Tyr Arg Tyr Leu
705 710 715 720

Tyr Phe Tyr Pro Thr Tyr Tyr Phe Thr Pro Glu Thr Asn Ser Phe Leu
725 730 735

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

Lys Leu His Phe Val Asp Lys Asn Leu Val Ala Asn Phe Asp Arg Thr
755 760 765

Arg Tyr Gln Ser Val Asp Ser Phe Leu Ile Asp Glu Lys Leu Arg Gln
770 775 780

Lys Lys Glu Thr Ile Asn Glu Glu Glu Asp Gly Lys Lys Asp Arg Thr
785 790 795 800

Phe Lys Leu Ser Tyr Pro Glu Asp Lys Pro Leu Thr Phe Tyr Phe Met
805 810 815

Ala Leu Pro Pro Gly Arg Asn Pro Thr Asp Thr Glu Ser Trp Val Met
820 825 830

Pro Ala Trp Leu Gly Leu Ala Phe Pro Met Ile Leu Asp Val Lys Thr
835 840 845

Val Val Ser Glu Ser Pro Ile Pro Pro Tyr Arg Asp Gly Ala Glu Phe
850 855 860

Glu Glu Thr Val Phe Leu Asp Ser Ala Pro Gln Ala Ile Arg Ser Leu
865 870 875 880

Thr Arg Cys Asp Arg Phe Arg Leu Asp Arg Val Leu Asn Pro Trp Gln
885 890 895

Asp Asn Asp Gly Lys Lys Tyr Ser Ala Pro Leu Asn Thr Leu Thr Ala
900 905 910

Ala Tyr Ser Ile His Leu Asp Val Asn Ser Lys Gln Gly Lys Thr Gly
 915 920 925

Tyr Asp Pro Asn Trp Gly Lys Leu Thr Glu Leu Ala Ile Asn Leu Glu
 930 935 940

Thr Ser Pro Leu Tyr Val Phe His Tyr Phe Lys Gln Trp Lys Arg Gly
 945 950 955 960

Lys Asp Ala Asp Ile Pro Ser Ala Asn Arg Ile Ala Leu Tyr Leu Tyr
 965 970 975

Asp Phe Tyr Pro Cys Phe Asp Pro Tyr Val Gln Ala Asn Arg Thr Asn
 980 985 990

Leu Thr Ile Asp Met Thr Ala Glu Ser Pro Leu Asn His Pro Lys Asn
 995 1000 1005

Leu Thr Glu Leu Tyr Arg Gln Phe Tyr Arg Ala Lys Ser Ser Lys
 1010 1015 1020

Gly Lys Pro Ile Lys Ala Asn Ala Ile Leu Lys Pro Ile Asp Glu
 1025 1030 1035

Ala Ala Asp Ile Ile Leu Lys Ala Asp Lys Ala Ile Ser Asp Asp
 1040 1045 1050

Leu Thr Ser Leu Val Ala Ala Arg Leu Phe Lys Leu Met Asp Arg
 1055 1060 1065

Val Arg Ser Gln Thr Ala Glu Gly Arg Tyr Val Ile Lys Glu Arg
 1070 1075 1080

Asp Gln Glu Arg Glu Lys Ile Leu Asp Phe Ala Lys Tyr Phe Val
 1085 1090 1095

Lys Asn Val Phe Glu Glu Ser Phe Glu Ser Asp Arg Ala Arg Leu
 1100 1105 1110

Ala Gly Arg Gln Leu Asn Ile Ile Arg Asp Thr Cys Glu Phe Leu
 1115 1120 1125

Tyr Arg Leu Glu Met Asp Lys Glu Arg Arg Gln Arg Gln Val Gln
 1130 1135 1140

Pro Leu Asp Thr Ser Asn Ser Ser Ser Glu Glu Glu Glu
 1145 1150 1155

<210> 6
 <211> 109
 <212> ARN
 <213> Nhân tạo

<220>
 <223> TiDcrRNA

<220>
 <221> misc_feature
 <222> (38)..(72)
 <223> n is a, c, g, or u

<400> 6
 guuccaauua aucuuaagcc cuauuaggga uugaaacnnn nnnnnnnnnn nnnnnnnnnn
 60

nnnnnnnnnn nnguuccaau uaaucuuaag ccuauuagg gauugaaac
 109

<210> 7
 <211> 675
 <212> ADN
 <213> Nhân tạo

<220>
 <223> Cas5d nucleotide sequence for expression in Escherichia coli

<400> 7
 atggtgcata tttatagctg ccagctggaa ctgcatgata gcctgtatta tgcgaccgcg
 60

gaaattggcc gcctgtatga aagcgaaccg gtgattcata actatgcgct gtgctatgcg
 120

ctgggcctgg tgaacagcga tagctatcgc tatttctgca gcgaacagat tccgcagtat
 180

caggaacatc tgaacccgct gaacgaagaa aaaatttatg tgaccccggc gcgcgcgatt
240

gcgcataccg cgggtgctgaa cacctggaaa tatgcgaaca acaactatca tgtggaaatg
300

gaaaaaaccc agaaaaacat tccgagcttc ggccgcgcga aagaaattgc gccggaaagc
360

attttcgaat gcttcattat tagccatcat ccgctgcagc tgccgaaatg gattcgctctg
420

ggcaaattgga tgagcaaagc ggaagtgaaa ctgaccgaac tgagcctgag caaacagaaa
480

gaagatctgt tcatttatcc gtatccgctg aaccgctgg atgtgatgtt caccatcag
540

gtgattggct atgatgtgat taacatgccg ccggtgagcc tgattcgcaa cgtgcgcatg
600

cgcggcgaat attatcagat tagcgatcgc ccggatctga aaattccggc acgtctgagc
660

tatcatttcg gctaa
675

<210> 8

<211> 834

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự nucleotit Cas6d để biểu hiện trong Escherichia coli

<400> 8

atgccgtata gcctgggtgct gaacctgacc ccgcgcagcc cgatttatcc gaacttcctg
60

accggccgcc atctgcatgc gctgttcctg accctgggtga gcagcgtgga tcaggaactg
120

ggcaacattc tgcataccgc ggaagcggat aaagcgttca ccctgagccc gctgcagatg
180

cagagcggcg gcaaaaccat taacagcccc cagtggcgcc atgaacgcga aattgcgagc
240

gaaaccccggt gctgggtggcg cattagcctg ctggatgacg gcctgttcgg caaactgacc
300

agcctgtggc tgaacctgaa cccgaaacag ccgtggcatc tgggcagcgc ggatctggtg
360

attaccagcg tgctggcgac cccgcagagc gtgcagccgt gggcgaacag ctgcacctat
420

cagtatctgt atgaaaacgc gagcgaaacc aaccgcgaat tcgatttcct gttcgcgacc
480

ccggtgacct tccgccaggg caaattcgat agcgcgctgc cgacccgcga actggtgttc
540

aacagcctgc tgggccgctg gaaccgctat agcggcattc cgttcgatag cattgcgctg
600

gaaagcattt tcccgagctt cttcgatatt cagaccaaac tggcggatga agcgtataaa
660

aaccagagca ttggctgcgt gggcgaaatt cattatcgcc tgctgggcga agtgggaaccg
720

gcgaaaatta aagcgattaa cgcgctggcg gatttcgcgc tgtatgcggg cgtgggccgc
780

aaaaccacca tgggcatggg catgaccgcg cgcattagca aagataaacg ctaa
834

<210> 9
<211> 1002
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự nucleotit Cas7d để biểu hiện trong Escherichia coli

<400> 9
atgaccttcc tgaccagcgt ggatgcgaaa ttcttccata gcgaaattcc gtataaaccg
60

atgggcaaatt atgtgcattt cctgaccatt cgcgtgaccg aaagctatcc gctgttccag
120

accgatggcg aactgaacaa agcgcgcgtg cgcgcgggca ttgacagcaa gaaaaccatt
180

agccgcctga gcatgttcaa acgcaaacag agcaccgccg aacgcctggt gggccgcgaa
240

ctgctgcgca actatggcct gattaccgcg gaagaatgcg aatataacgt gaaattcgcg
300

atgaacaacg cggattgcat tatttatggc ttgcgattg gcgatagcgg cagcgaaaaa
360

agcaaagtgg tgggtgatac cgcgttcagc attaccccg tcatgaaaag ccatgagagc
420

ttcaccctga acgcgccgta tgaaaacggc acgatggcga gcaaaggcga aaacaacacc
480

aaagtgggcg aagtgaccag ccgcattaac cagcaggatc atattcgccc gcaggtgttc
540

ttcccagaca ttgtgaccct gaaagatccg accgaagcga gcttcctgta tgtgttcaac
600

aacattctgc gcacccgcca ttatggcgcg cagaccaccc gcaccggccg cgtgcgcaac
660

gaactgattg gcgtgatttt cgcggatggc gaaattgtga gcaacctgcg ctggaccacg
720

gcgatttatg atcgctgcc ggatgaagtg ctgcatagca ttgatccgct ggatgaagat
780

ctggtgatgg aaaaagcgac cgaagcgatt caggcgctga tggcggaaga atttattgtg
840

cataccgatt tcattggcga aaacttccag ccgctgctga ccgaagtga aaccctgacc
900

ggcaccgaag cgggcattct gagcgtgctg gatcaggcga acaaagaaag caaaaaatat
960

ttcgaacagt atattgaaaa gaaaaaggcg gaaaagaaat aa
1002

<210> 10
<211> 2166
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự nucleotit Cas3d để biểu hiện trong Escherichia coli

<400> 10
atgggaaact atcaggtgac tcttaagcca gtgtattctt gccagctga cgagatccca
60

gacggaatca aggtgccaca gggatggaga ctttcttggc accaggtgga gacttggaag
120

gctcttaacg acccagacat cgacgtgatc ttcaacactg ctatgactgg agacggaaag
180

tctcttgctg cttatcttag aactcttcag ggatatttcc caatcatggg actttatcca
240

actaacgagc ttgctagaga tcagagagga cagatcgagg cttatatcca gagattccag
300

ccaactgacc agccaagagt gaaccttctt actggaccag agcttgagct ttatgctgag
360

agagacggaa agactaaggc tatcgctctt gagactagat ctaagcagtc tgagatcctt
420

cttactaacc cagacatctt ccactatctt cacagagctg cttatcttac tccatatgac
480

aaccagacc agctttggaa cagaatcgac aagcacttcg accttttcct ttctgacgag
540

ttccacgtgt tcggaactcc acaggtggct tctatcatca acactatgct tcttatcaga
600

agagctaaca gaggaagag atatcttttc ctttctgcta ctccagacga gggacttctt
660

aagagacttg acaaggctgg attcagatat agatctatcg acccagtgag agagggaaaag
720

tatagattcc cagacactcc agaggaggct aactctcttg ctccagcagg atggagacag
780

gtgacttctg agatcgagct ttctttcatc ccacttccat cttctttcca gacttctgag
840

aactggctta aggagaacaa ggagagaatc cttgactatt tcaagagata tccaggatct
900

aaggagacta tcatccttaa ctctatcgct tctgtgaaga gacttcttcc aatcttcaga
960

gagcttcttg ctactatcgg acttactgtg ggagagaaca ctggactttc tggaactaga
1020

gagaaacttg cttctcttaa cagagatctt gtgatcgga cttctactat cgacgtggga
1080

gtggacttca agatcaactt ccttatcttc gagtcttctg acgctggaaa cttcatccag
1140

agattcggaa gacttgggaag acactctgga tatgacagaa agggaactgc tgtgaagttc
1200

actaacttca ctgctatcgc tcttgtgcc aagttcttcc ttgagagact ttctcgagaag
1260

aaggacgctc cacttcaggt gggagagaga tatgacagaa tccagcttca ggaggctatc
1320

aagtctaact atagacacat caacaacttc gagggatatt atcagagatg gggagctgtg
1380

cagtctttcc agctttggtg gaaccttgga tctccaaaga tcaagtctca gtatggagag
1440

tcaagacaga agttccagca ggagtgcgag gaggtgttcg acacttctct taagagagtg
1500

gctggaagag tgaagggatg ggctgacgag tggaaggagc tttctggaaa gaacggaaac
1560

ccaatcttcg aggacgcttc ttctttcaga ggatcttctc cacttctttg cggactttat
1620

gactctactg agccagagga gtgcgacaga ttcaagactt atgaccttcc atctatcctt
1680

ggaaaccttg aggtggaggt gtggagaaag ggagagttca agagacagat cgaggctact
1740

aagactccaa tcgctagaag aagattcgac tattgccttg ctttccttaa ccttaaggga
1800

tatagagagg agagacttaa ctggagattc acttatgacg gagatcttgg agagatcgct
1860

tctgcttgga aggtgcaggt gcttactgga atcggagtgt ggcagccaga caacccttgg
1920

cttgacagaa tctcaagaga gcttagagat ctgctcttg tgtctttcgt gttcgcttat
1980

ccagtggctg ctgtgagaca gagacttcag ctccaatgc acttcggaat ctatccaatc
2040

tctgacgagt cttctcttca ctctccactt tctccatatt ctatcgctat cggacaggct
2100

gctcttcttc ttgacactct tgctcacaga ttcaaggga agggaggaga ggtgtggatc
2160

tgctag
2166

<210> 11
<211> 3471
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự nucleotit Cas10d để biểu hiện trong *Escherichia coli*

<400> 11
atgccaaaga agcagaagaa gcttgaggag actggacagc ttaacctttt cgacaacact
60

actgagatcg acgacgagga ccttgacttc gagttcgagg acatcgacct tgagtctctt
120

gtgtctgagg accttggaa cactgagtct gtgtctgaca gaagagtgga gactgtgaga
180

cagcttctta ctcttaagct tcttagagag gctatcagag ctgagaaccc agacgacaga
240

gtgatggctg acttcgctga gatggtgctt ccaaaccctt ttagacttgc tatcgagtg
300

actgctaagg gaggaactt catcgaggct gtggacagag gaagagagct tagaaacaag
360

ccaaaggcta agagagacaa cgctggagat cagtctctta aactcacct tcttaacgga
420

cttttcccag ctaaccttat cgagaagaga cttcagaagc ttaacactac tgtgagaaga
480

atcatcaagg agttcgagag aagacttgct atcgctggat tccttgtgca cgacttcgag
540

aagttctctt atgacagatt cccatctatg tctgagagat atatccagat ccagagagac
600

ttcatccagg acccattcaa gaaccaggac ccaagaaagc tttcaagaga ggagcacaga
660

gagatccttc aggtgcttat ccagagctt ggacttgaca gattcctttt ccagacaaac
720

ccagagagat ggcttgagta tcttgacgac cttctttata tcgctaagaa cactcagaga
780

agaaacgaca ctgaccttaa cacttctgag gacggactta acgtgagact taacgacaga
840

gtgatcgagt ctctttgcga ccttgcttgc cttgctgaca gacttgcttc tatcatcaag
900

caccacacag acgctgagaa ggcttctctt caggaccttc tttattctct ttctgacgga
960

gagcttaagt tcacttatca ctctatcgct gagaacagag gagtgccttac taacgtgctt
1020

aacaacgctg tgatggaggc tcaccaggag cttgactatc agccaacttct ttatcttcca
1080

actggagtgg tgtatatcgc tccaaagaac gctccagagg tgtctcttga gactcttcca
1140

aacagagtgg tggacactat caagtctctt tgctctggag agcttcagag aaagcagact
1200

ggattcggaa gagacggaaa gggatatgaag tatgctgact attattctca gttcttcgac
1260

gacgctggac ttatgagagc tgctcttaac gctactctta gaatccttgg agacaacaag
1320

gcttctgtgg ctagatctag aggagagaac cttatcaagt tccagcagca gggagtgctt
1380

ccaactgact atgacttcca ctgcgaggac gacatcagaa tcgacagact tgctgagttc
1440

ggagacgtgg tgactagaaa gatctgggga gacagacttg agaagatcga gcaggctaga
1500

aagcttcaga agaaccttcc agctccacca gaccttgacc ttatctctga gatcgctcac
1560

tattggaacc ttgagaacta tcttccacag atcagagcta tcaagagaat caacgagtct
1620

cttaaggagc ttaagcttaa gggaaacact ggaggagtgc catatgagtg gtattatctt
1680

gctgctcagt atcttaagca gcaccaggga atcgaggaca tcagaccagt ggctgaggac
1740

cttatcgctt tccttgctgc taagatcgct gctatcgctg ctggatataa ccttccagac
1800

ggatgggagg accttagaga gtgggtgaac caggtggtgc agcttccagg aagagagctt
1860

gtcactcta tcgagacttt ccagaaggag cttaccact ataacgctgc taagaagcag
1920

ggaagaggaa gacagcttct ttgctctatc tctcactctc catattctgt gtctgagcag
1980

atggagtctg ctgtgctttt cactccacag gtgtatacta acaagcagat gcttgctgga
2040

tctaacgcta agagaaacat ctcttctatc gctggaactg agatgatgct tagacagatc
2100

cttatgaacc agactcaggc tgtgggaaag agattcgagg acggaaagta tagatatctt
2160

tatttctatc caacttatta tttcactcca gagactaact ctttccttca gaaggcttat
2220

gctaacatcg ctcagactag attcgactct tctatcaagc ttcacttcgt ggacaagaac
2280

cttgtggcta acttcgacag aactagatat cagtctgtgg actctttcct tatcgacgag
2340

aagcttagac agaagaagga gactatcaac gaggaggagg acggaaagaa ggacagaact
2400

ttcaagcttt cttatccaga ggacaagcca cttactttct atttcatggc tcttccacca
2460

ggaagaaacc caactgacac tgagtcttgg gtaatgccag cttggcttgg acttgctttc
2520

ccaatgatcc ttgacgtgaa gactgtggtg tctgagtctc caatcccacc atatagagac
2580

ggagctgagt tcgaggagac tgtgttcctt gactctgctc cacaggctat cagatctctt
2640

actagatgcg acagattcag acttgacaga gtgcttaacc cttggcagga caacgacgga
2700

aagaagtatt ctgctccact taacactctt actgctgctt attctatcca ccttgacgtg
2760

aactctaagc agggaaagac tggatatgac ccaaactggg gaaagcttac tgagcttgct
2820

atcaaccttg agacttctcc actttatgtg ttccactatt tcaagcagtg gaagagagga
2880

aaggacgctg acatcccatc tgctaacaga atcgctcttt atctttatga cttctatcca
2940

tgcttcgacc catatgtgca ggctaacaga actaacctta ctatcgacat gactgctgag
3000

tctccactta accacccaaa gaaccttact gagctttata gacagttcta tagagctaag
3060

tcttctaagg gaaagccaat caaggctaac gctatcctta agccaatcga cgaggctgct
3120

gacatcatcc ttaaggctga caaggctatc tctgacgacc ttacttctct tgtggctgct
3180

agacttttca agcttatgga cagagtgaga tctcagactg ctgagggaag atatgtgatc
3240

aaggagagag atcaggagag agagaagatc cttgacttcg ctaagtatct cgtgaagaac
3300

gtgttcgagg agtctttcga gtctgacaga gctagacttg ctggaagaca gcttaacatc
3360

atcagagaca cttgcgagtt cctttataga cttgagatgg acaaggagag aagacagaga
3420

caggtgcagc cacttgacac ttctaactct tcttctgagg aggaggagta a
3471

<210> 12
<211> 35
<212> ADN
<213> Nhân tạo

<220>
<223> Đoạn khởi đầu tổng hợp J23108

<400> 12
ctgacagcta gctcagtcct aggtataatg ctagc
35

<210> 13
<211> 36
<212> ADN
<213> Nhân tạo

<220>

<223> Trình tự liên kết ribosom

<400> 13

aataattttg tttaacttta agaaggagat atacat
36

<210> 14

<211> 24

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn kết thúc STOP767

<400> 14

agatcctgta aaacgacggc cagt
24

<210> 15

<211> 19

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn kết thúc STOP768(1)

<400> 15

cgccagggtt ttcccagtc
19

<210> 16

<211> 19

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn kết thúc TOP768(2)

<400> 16

cgccagggtt ttcccagtc
19

<210> 17

<211> 47

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn kết thúc T7

<400> 17

tagcataacc ccttggggcc tctaaacggg tcttgagggg ttttttg
47

<210> 18

<211> 37

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn lắp CRISPR

<400> 18

gttccaatta atcttaagcc ctattaggga ttgaaac
37

<210> 19

<211> 19

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự đoạn khởi đầu T7

<400> 19

taatacgact cactatagg
19

<210> 20

<211> 109

<212> ADN

<213> Nhân tạo

<220>

<223> Casset biểu hiện crARN

<400> 20

gttccaatta atcttaagcc ctattaggga ttgaaacggg aataatacga ctcactatag
60

ggagaaagga tcgttccaat taatcttaag ccctattagg gattgaaac
109

<210> 21

<211> 581

<212> ADN

<213> Nhân tạo

<220>

<223> Casset biểu hiện gen ccdB tổng hợp

<400> 21

aggctttaat acgactcact atagggagaa aggatccata aaggaggtaa ataataagc
60

agcgtattac agtgacagtt gacagcgaca gctatcagtt gctcaaggca tatatgatgt
120

caatatctcc ggtctggtaa gcacaacat gcagaatgaa gcccgtcgtc tgcgtgccga
180

acgctggaaa gcggaaaatc aggaaggat ggctgaggtc gcccggttta ttgaaatgaa
240

cggctctttt gctgacgaga acagggactg gtgaaatgca gtttaagggtt tacacctata
300

aaagagagag ccgttatcgt ctgtttgtgg atgtacagag tgatattatt gacacgcccg
360

ggcgacggat ggtgatcccc ctggccagtg cacgtctgct gtcagataaa gtctcccgtg
420

aactttaccc ggtggtgcat atcggggatg aaagctggcg catgatgacc accgatatgg
480

ccagtgtgcc ggtctccgtt atcggggaag aagtggctga tctcagccac cgcgaaaatg
540

acatcaaaaa cgccattaac ctgatgttct ggggaatata a
581

<210> 22

<211> 18

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của tín hiệu định vị nhân (NLS)

<400> 22

Asp	Pro	Lys	Lys	Lys	Arg	Lys	Val	Asp	Pro	Lys	Lys	Lys	Arg	Lys	Val
1				5					10					15	

Ser Gly

<210> 23
 <211> 54
 <212> ADN
 <213> Nhân tạo

<220>
 <223> Trình tự nucleotit NLS

<400> 23
 gacccaaaga agaagcggaa ggtagaccct aagaagaagc gcaagggtttc tgga
 54

<210> 24
 <211> 20
 <212> PRT
 <213> Nhân tạo

<220>
 <223> Trình tự axit amin của peptit tự cắt 2A

<400> 24
 Gly Ser Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu
 1 5 10 15

Asn Pro Gly Pro
 20

<210> 25
 <211> 60
 <212> ADN
 <213> Nhân tạo

<220>
 <223> Trình tự mã hóa cho peptide tự cắt 2A(1)

<400> 25
 ggctctgagg gcagaggcag cctgctgacc tgcggcgacg tggaggaaaa ccctggccct
 60

<210> 26
 <211> 60
 <212> ADN
 <213> Nhân tạo

<220>

<223> Trình tự mã hóa cho peptide tự cắt 2A(2)

<400> 26

gggtctgagg gacgcggctc cctgctcacc tgtggagatg tggaagagaa cccaggcccc
60

<210> 27

<211> 60

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự mã hóa cho peptide tự cắt 2A(3)

<400> 27

ggttctgaag gcagaggctc tctgctgaca tgtggggatg tggaggaaaa tcctggccct
60

<210> 28

<211> 60

<212> ADN

<213> Nhân tạo

<220>

<223> Trình tự mã hóa cho peptide tự cắt 2A(4)

<400> 28

ggatccgagg gcagaggaag tctgctaaca tgcggtgacg ttgaggagaa tcccgggcca
60

<210> 29

<211> 820

<212> ADN

<213> Nhân tạo

<220>

<223> 2 x đoạn khởi đầu của gen 35S của virus thể khảm của cải hoa lơ
+ trình tự omega

<400> 29

gccaatggtg tggagcacga cactctcgtc tactccaaga atatcaaaga tacagtctca
60

gaagacaaaa gggctattga gacttttcaa caaagggtaa tatcgggaaa cctcctcgga
120

ttccattgcc cagctatctg tcaattcatc aaaaggacag tagaaaagga aggtggcacc
180

tacaaatgcc atcattgcga taaaggaaag gctatcggtc aagatgcctc tgccgacagt
240

ggtcccaaag atggaccccc acccacgagg agcatcgtgg aaaaagaaga cgttccaacc
300

acgtcttcaa agcaagtgga ttgatgtgaa catgggtggag cacgacactc tcgtctactc
360

caagaatatc aaagatacag tctcagaaga ccaaagggtc attgagactt ttcaacaaag
420

ggtaatatcg ggaaacctcc tcggattcca ttgccagct atctgtcact tcattcaaaag
480

gacagtagaa aaggaagggtg gcacctacaa atgccatcat tgcgataaag gaaaggctat
540

cgttcaagat gcctctgccg acagtgggtcc caaagatgga cccccacca cgaggagcat
600

cgtggaaaaa gaagacgttc caaccacgtc ttcaaagcaa gtggattgat gtgatatttc
660

cactgacgta agggatgaag cacaatccca ctatccttcg caagaccctt cctctatata
720

aggaagttca tttcatttgg agaggccggt ctagagtatt tttacaacaa ttaccaacaa
780

caacaaacaa caaacaacat tacaattact atttacaatt
820

<210> 30
<211> 447
<212> ADN
<213> Nhân tạo

<220>
<223> Đoạn kết thúc của gen có kích cỡ 18,2kDa của protein sốc
Arabidopsis

<400> 30
atatgaagat gaagatgaaa tatttgggtgt gtcaaataaa aagcttgtgt gcttaagttt
60

gtgttttttt cttggcttgt tgtgttatga atttgtggct ttttctaata ttaaatgaat
120

gtaagatctc attataatga ataaacaaat gtttctataa tccattgtga atgttttggt
180

ggatctcttc tgcagcatat aactactgta tgtgctatgg tatggactat ggaatatgat
240

taaagataag atgggctcat agagtaaaac gaggcgaggg acctataaac ctcccttcat
300

catgctattt catgatctat ttataaaaat aaagatgtag aaaaaagtaa gcgtaataac
360

cgcaaaacaa atgatttaaa acatggcaca taatgaggag attaagttcg gtttacgttt
420

attttagtac taattgtaac gtgagac
447

<210> 31
<211> 97
<212> ADN
<213> Nhân tạo

<220>
<223> Casset biểu hiện crARN

<400> 31
gttccaatta atcttaagcc ctattagga ttgaaacgga gacctcaat tgcggtctc
60

gttccaatta atcttaagcc ctattagga ttgaaac
97

<210> 32
<211> 448
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự đoạn khởi đầu của gen U6 snRNA-26 của Arabidopsis

<400> 32
aagcttcgtt gaacaacgga aactcgactt gccttccgca caatacatca tttcttctta
60

gctttttttc ttcttcttcg ttcatacagt ttttttttgt ttatcagctt acattttctt
120

gaaccgtagc tttcgttttc ttctttttta ctttccattc ggagtttttg tatcttggtt
180

catagtttgt cccaggatta gaatgattag gcatcgaacc ttcaagaatt tgattgaata
240

aaacatcttc attcttaaga tatgaagata atcttcaaaa ggcccctggg aatctgaaag
300

aagagaagca ggcccattta tatgggaaag aacaatagta tttcttatat aggcccattt
360

aagttgaaaa caatcttcaa aagtcccaca tcgcttagat aagaaaacga agctgagttt
420

atatacagct agagtcgaag tagtgatt
448

<210> 33
<211> 723
<212> ADN
<213> Nhân tạo

<220>
<223> 2xNLS + Cas5d

<400> 33
gacccaaaga agaagcggaa ggtagaccct aagaagaagc gcaaggtttc tggagtgcac
60

atctattctt gccagcttga gcttcacgac tctctttatt atgctactag agagatcgga
120

agactttatg agtctgagcc agtgatccac aactatgctc tttgctatgc tcttggaactt
180

gtgaactctg actcttatag atatttctgc tctgagcaga tcccacagta tcaggagcac
240

cttaaccac ttaacgagga gaagatctat gtgactccag ctagagctat cgctcacact
300

gctgtgctta acacttggaa gtatgctaac aacaactatc acgtggagat ggagaagact
360

cagaagaaca tcccatcttt cggaagagct aaggagatcg ctccagagtc tatcttcgag
420

tgcttcatca tctctcacca cccacttcag cttccaaagt ggatcagact tggaaagtgg
480

atgtctaagg ctgaggtgaa gcttactgag ctttctcttt ctaagcagaa ggaggacctt
540

ttcatctatc catatccact taaccacactt gacgtgatgt tcactcacca ggttatcgga
600

tatgacgtga tcaacatgcc accagtgtct cttatcagaa acgtgagaat gagaggagag
660

tattatcaga tctctgacag accagacctt aagatcccag ctagactttc ttatcacttc
720

gga
723

<210> 34
<211> 882
<212> ADN
<213> Nhân tạo

<220>
<223> 2xNLS + Cas6d

<400> 34
gacccaaaga agaagcggaa ggtagaccct aagaagaagc gcaagggttc tggaccatat
60

tctcttgtgc ttaaccttac tccaagatct ccaatctatc caaacttcct tactggaaga
120

caccttcacg ctcttttcct tactcttgtg tcttctgtgg accaggagct tggaaacatc
180

cttcacactg ctgaggctga caaggctttc actctttctc cacttcagat gcagtctgga
240

ggaaagacta tcaactctcc acagtggaga cacgagagag agatcgcttc tgagactcca
300

tgctggtgga gaatctctct tcttgacgac agacttttcg gaaagcttac ttctctttgg
360

cttaacctta acccaaagca gccttggcac cttggatctg ctgaccttgt gatcacttct
420

gtgcttgcta ctccacagtc tgtgcagcct tgggctaact cttgcaacta tcagtatctt
480

tatgagaacg cttctgagac taacagagag ttcgacttcc ttttcgctac tccagtgact
540

ttcagacagg gaaagttoga ctctgctctt ccaactagag agcttggtgt caactctctt
600

cttggaagat ggaacagata ttctggaatc ccattcgact ctatcgctct tgagtctatc
660

ttcccatctt tcttcgacat ccagactaag cttgctgacg aggcttataa gaaccagtct
720

atcggatgcg tgggagagat ccactataga cttcttggag aggtggagcc agctaagatc
780

aaggctatca acgctcttgc tgacttcgct ctttatgctg gagtgggaag aaagactact
840

atgggaatgg gaatgactag aagaatctct aaggacaaga ga
882

<210> 35
<211> 1053
<212> ADN
<213> Nhân tạo

<220>
<223> 2xNLS + Cas7d

<400> 35
gacccaaaga agaagcggaa ggtagaccct aagaagaagc gcaagggttc tggaactttc
60

cttacttctg tggacgctaa gttcttccac tctgagatcc catataagcc aatgggaaag
120

tatgtgcact tccttactat cagagtgcact gagtcttctc cactttttcca gactgacgga
180

gagcttaaca aggctagagt gagagctgga atcgactcta agaagactat ctcaagactt
240

tctatgttca agagaaagca gtctactcca gagagacttg tgggaagaga gcttcttaga
300

aactatggac ttatcactgc tgaggagtgc gagtataacg tgaagttcgc tatgaacaac
360

gctgactgca tcattctatgg attcgctatc ggagactctg gatctgagaa gtctaagggtg
420

gtggtggaca ctgctttctc tatcactcca ttcgacgagt ctcacgagtc tttcactctt
480

aacgctccat atgagaacgg aactatggct tctaaggag agaacaacac taagggtggga
540

gaggtgactt caagaatcaa ccagcaggac cacatcagac cacaggtgtt cttcccatct
600

atcgtgactc ttaaggaccc aactgaggct tctttccttt atgtgttcaa caacatcctt
660

agaactagac actatggtgc tcagactact agaactggaa gagtgagaaa cgagcttata
720

ggagtgatct tcgctgacgg agagatcgtg tctaacctta gatggactca ggctatctat
780

gacagacttc cagacgaggt gcttcaactc atcgacccac ttgacgagga ccttgtgatg
840

gagaaggcta ctgaggctat ccaggctctt atggctgagg agttcatcgt gcacactgac
900

ttcatcggag agaacttcca gccacttctt actgaggtga agactcttac tggaactgag
960

gctggaatcc tttctgtgct tgaccaggct aacaaggagt ctaagaagta tttcgagcag
1020

tatatcgaga agaagaaggc tgagaagaag taa
1053

<210> 36
<211> 2217
<212> ADN
<213> Nhân tạo

<220>
<223> 2xNLS + Cas3d

<400> 36
atggacccaa agaagaagcg gaaggtagac cctaagaaga agcgcaagggt ttctggagga
60

aactatcagg tgactcttaa gccagtgtat tcttgcccag ctgacgagat ccagacgga
120

atcaagggtgc cacagggatg gagactttct tggcaccagg tggagacttg gaaggctctt
180

aacgacccag acatcgacgt gatcttcaac actgctatga ctggagacgg aaagtctctt
240

gctgcttata ttagaactct tcagggatat ttcccaatca tgggacttta tccaactaac
300

gagcttgcta gagatcagag aggacagatc gaggcttata tccagagatt ccagccaact
360

gaccagccaa gagtgaacct tcttactgga ccagagcttg agctttatgc tgagagagac
420

ggaaagacta aggctatcgc tottgagact agatctaagc agtctgagat ccttcttact
480

aaccagaca tcttccacta tcttcacaga gctgcttata ttactccata tgacaaccca
540

gaccagcttt ggaacagaat cgacaagcac ttcgaccttt tccttttcga cgagttccac
600

gtgttcggaa ctccacaggt ggcttctatc atcaacacta tgcttcttat cagaagagct
660

aacagaggaa agagatatct tttcctttct gctactccag acgagggact tcttaagaga
720

cttgacaagg ctggattcag atatagatct atcgacctcag tgagagaggg aaagtataga
780

ttcccagaca ctccagagga ggctaactct cttgctcagc agggatggag acaggtgact
840

tctgagatcg agctttcttt catcccactt ccatcttctt tccagacttc tgagaactgg
900

cttaaggaga acaaggagag aatccttgac tatttcaaga gatatccagg atctaaggga
960

gctatcatcc ttaactctat cgcttctgtg aagagacttc ttccaatctt cagagagctt
1020

cttgctacta tcggacttac tgtgggagag aacactggac tttctggaac tagagagaaa
1080

cttgcttctc ttaacagaga tcttgtgatc ggaacttcta ctatcgacgt gggagtggac
1140

ttcaagatca acttccttat cttogagtct tctgacgctg gaaacttcat ccagagattc
1200

ggaagacttg gaagacactc tggatatgac agaaaggga ctgctgtgaa gttcactaac
1260

ttcactgcta tcgctcttgt gccaaagttc ttccttgaga gacttttcga gaagaaggac
1320

gctccacttc aggtgggaga gagatatgac agaatccagc ttcaggaggc tatcaagtct
1380

aactatagac acatcaacaa cttcgaggga tattatcaga gatggggagc tgtgcagtct
1440

ttccagcttt ggtggaacct tggatctcca aagatcaagt ctcagtatgg agagtcaaga
1500

cagaagttcc agcaggagtg cgaggaggtg ttcgacactt ctcttaagag agtggctgga
1560

agagtgaagg gatgggctga cgagtggaag gagctttctg gaaagaacgg aaaccaatc
1620

ttcgaggacg cttcttcttt cagaggatct tctccacttc tttgcggaact ttatgactct
1680

actgagccag aggagtgcga cagattcaag acttatgacc ttccatctat ccttggaac
1740

cttgaggtgg aggtgtggag aaaggagag ttcaagagac agatcgaggc tactaagact
1800

ccaatcgcta gaagaagatt cgactattgc cttgctttcc ttaaccttaa gggatataga
1860

gaggagagac ttaactggag attcacttat gacggagatc ttggagagat cgcttctgct
1920

tggaaggtgc aggtgcttac tggaatcgga gtgtggcagc cagacaacc ttggcttgac
1980

agaatctcaa gagagcttag agatcttgct cttgtgtctt tcgtgttcgc ttatccagtg
2040

gctgctgtga gacagagact tcagcttcca atgcacttcg gaatctatcc aatctctgac
2100

gagtcttctc ttcactctcc actttctcca tattctatcg ctatcggaca ggctgctctt
2160

cttcttgaca ctcttgctca cagattcaag ggaaaggag gagaggtgtg gatctgc
2217

<210> 37

<211> 3519

<212> ADN

<213> Nhân tạo

<220>

<223> 2xNLS + Cas10d

<400> 37

gacccaaaga agaagcggaa ggtagaccct aagaagaagc gcaaggtttc tggaccaaag
60

aagcagaaga agcttgagga gactggacag cttaaccttt tcgacaacac tactgagatc
120

gacgacgagg accttgactt cgagttcgag gacatcgacc ttgagtctct tgtgtctgag
180

gaccttgga tcaactgagtc tgtgtctgac agaagagtgg agactgtgag acagcttctt
240

actcttaagc ttcttagaga ggctatcaga gctgagaacc cagacgacag agtgatggct
300

gacttcgctg agatgggtgt tccaaacctt cttagacttg ctatcggagt gactgctaag
360

ggaggaaact tcatcgaggc tgtggacaga ggaagagagc ttagaaacaa gccaaaggct
420

aagagagaca acgctggaga tcagtctctt aacactcacc ttcttaacgg acttttccca
480

gctaacctta tcgagaagag acttcagaag cttaacacta ctgtgagaag aatcatcaag
540

gagttcgaga gaagacttgc tatcgctgga ttcttgtgc acgacttcga gaagttctct
600

tatgacagat tcccatctat gtctgagaga tatatccaga tccagagaga cttcatccag
660

gacccattca agaaccagga cccaagaaag ctttcaagag aggagcacag agagatcctt
720

caggtgetta tcccagagct tggacttgac agattccttt tcccagacaa cccagagaga
780

tggcttgagt atcttgacga ctttctttat atcgctaaga aactcagag aagaaacgac
840

actgacctta aacttctga ggacggactt aacgtgagac ttaacgacag agtgatcgag
900

tctctttgcg accttgcttg ccttgctgac agacttgctt ctatcatcaa gcaccacac
960

gacgctgaga aggccttctct tcaggacctt ctttattctc tttctgacgg agagcttaag
1020

ttcacttatc actctatcgc tgagaacaga ggagtgccta ctaacgtgct taacaacgct
1080

gtgatggagg ctcaccagga gcttgactat cagccacttc tttatcttcc aactggagtg
1140

gtgtatatcg ctccaaagaa cgctccagag gtgtctcttg agactcttcc aaacagagtg
1200

gtggacacta tcaagtctct ttgctctgga gagcttcaga gaaagcagac tggattcgga
1260

agagacggaa agggatatgaa gtatgctgac tattattctc agttcttcga cgacgctgga
1320

cttatgagag ctgctcttaa cgctactctt agaatccttg gagacaacaa ggcttctgtg
1380

gctagatcta gaggagagaa ccttatcaag ttccagcagc agggagtgct tccaactgac
1440

tatgacttcc actgcgagga cgacatcaga atcgacagac ttgctgagtt cggagacgtg
1500

gtgactagaa agatctgggg agacagactt gagaagatcg agcaggctag aaagcttcag
1560

aagaaccttc cagctccacc agaccttgac cttatctctg agatcgctca ctattggaac
1620

cttgagaact atcttccaca gatcagagct atcaagagaa tcaacgagtc tcttaaggag
1680

cttaagctta agggaaacac tggaggagtg ccatatgagt ggtattatct tgctgctcag
1740

tatcttaagc agcaccagc aatcgaggac atcagaccag tggctgagga ccttatcgct
1800

ttccttgctg ctaagatcgc tgctatcgtg gctggatata accttccaga cggatgggag
1860

gaccttagag agtgggtgaa ccagggtgtg cagcttccag gaagagagct tgctcactct
1920

atcgagactt tccagaagga gcttaaccac tataacgctg ctaagaagca gggaagagga
1980

agacagcttc tttgctctat ctctcactct ccatattctg tgtctgagca gatggagtct
2040

gctgtgcttt tcactccaca ggtgtatact aacaagcaga tgcttgctgg atctaacgct
2100

aagagaaaca tctcttctat cgctggaact gagatgatgc ttagacagat ccttatgaac
2160

cagactcagg ctgtgggaaa gagattcgag gacggaaagt atagatatct ttatttctat
2220

ccaacttatt atttcactcc agagactaac tctttccttc agaaggctta tgctaacatc
2280

gctcagacta gattcgactc ttctatcaag cttcacttcg tggacaagaa ccttgtggct
2340

aacttcgaca gaactagata tcagtctgtg gactctttcc ttatcgacga gaagcttaga
2400

cagaagaagg agactatcaa cgaggaggag gacggaaaga aggacagaac tttcaagctt
2460

tcttatccag aggacaagcc acttactttc tatttcatgg ctcttccacc aggaagaaac
2520

ccaactgaca ctgagtcttg ggtaatgcca gcttggcttg gacttgcttt cccaatgac
2580

cttgacgtga agactgtggt gtctgagtct ccaatccac catatagaga cggagctgag
2640

ttcgaggaga ctgtgttctt tgactctgct ccacaggcta tcagatctct tactagatgc
2700

gacagattca gacttgacag agtgcttaac ccttggcagg acaacgacgg aaagaagtat
2760

tctgctccac ttaacactct tactgctgct tattctatcc accttgacgt gaactctaag
2820

cagggaagaa ctggatatga cccaaactgg ggaaagctta ctgagcttgc tatcaacctt
2880

gagacttctc cactttatgt gttccactat ttcaagcagt ggaagagagg aaaggacgct
2940

gacatcccat ctgctaacag aatcgctctt tatctttatg accttatcc atgcttcgac
3000

ccatatgtgc aggctaacag aactaacctt actatcgaca tgactgctga gtotccactt
3060

aaccacccaa agaaccttac tgagctttat agacagttct atagagctaa gtcttctaag
3120

ggaaagccaa tcaaggctaa cgctatcctt aagccaatcg acgaggctgc tgacatcatc
3180

cttaaggctg acaaggctat ctctgacgac cttactttct ttgtggctgc tagacttttc
3240

aagcttatgg acagagtga atctcagact gctgagggaa gatatgtgat caaggagaga
3300

gatcaggaga gagagaagat ccttgacttc gctaagtatt tcgtgaagaa cgtgttcgag
3360

gagtctttcg agtctgacag agctagactt gctggaagac agcttaacat catcagagac
3420

acttgcgagt tcctttatag acttgagatg gacaaggaga gaagacagag acaggtgcag
3480

ccacttgaca cttctaactc ttcttctgag gaggaggag
3519

<210> 38
<211> 35
<212> ADN
<213> Nicotiana benthamiana

<400> 38
tgggtggttt gtctacagca aaatatctgg cagat
35

<210> 39
<211> 35
<212> ADN
<213> Nicotiana benthamiana

<400> 39
aaatttgcta ttggactctt gccagcaatg cttgg
35

<210> 40
<211> 35
<212> ADN
<213> Solanum lycopersicum

<400> 40

tacctggatc tcagtctccc gaaagaggtg aggag
35

<210> 41

<211> 811

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn khởi đầu lai gen beta-actin của gà phổ thông + gen tăng cường của Cytomegalovirus

<400> 41

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
60

gacgtcaata gtaacgcaa tagggacttt ccattgacgt caatgggtgg agtattttacg
120

gtaaactgcc cacttggcag tacatcaagt gtatcatatg ccaagtacgc cccctattga
180

cgtcaatgac ggtaaattggc ccgcctggca ttgtgcccag tacatgacct tatgggactt
240

tcctacttgg cagtacatct acgtattagt catcgctatt accatggtcg aggtgagccc
300

cacgtttctgc ttactctcc ccctctcccc cccctcccca cccccaattt tgtattttatt
360

tatttttttaa ttattttgtg cagcgatggg ggcggggggg gggggggggc gcgcgccagg
420

cggggcgggg cggggcgagg ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa
480

tcagagcggc gcgctccgaa agtttctttt tatggcgagg cggcggcggc ggcggcccta
540

taaaaagcga agcgcgcggc gggcgggagt cgctgcgacg ctgccttcgc cccgtgcccc
600

gctccgcgcg cgcctcgcgc cgcgcgcccc ggcctctgact gaccgcgtta ctcccacagg
660

tgagcggggc ggacggccct tctcctccgg gctgtaatta gctgagcaag aggtaagggt
720

ttaagggatg gttggttggt ggggtattaa tgtttaatta cctggagcac ctgcctgaaa
780

tcactttttt tcaggttgga ccggtgccac c
811

<210> 42
<211> 223
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự đoạn kết thúc gen hormon sinh trưởng có nguồn gốc từ
bò

<400> 42
gctgatcagc ctgcactgtg ctttctagtt gccagccatc tgttgtttgc ccctcccccg
60

tgcccttcctt gaccctggaa ggtgccactc ccactgtcct ttctaataa aatgaggaaa
120

ttgcatcgca ttgtctgagt aggtgtcatt ctattctggg ggggtggggtg gggcaggaca
180

gcaaggggga ggattgggaa gagaatagca ggcattgctgg gga
223

<210> 43
<211> 249
<212> ADN
<213> Nhân tạo

<220>
<223> Đoạn khởi đầu gen U6 snRNA của người

<400> 43
gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag
60

ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga
120

aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat
180

atgcttaccg taacttgaaa gtatttcgat ttcttggctt tatatatctt gtggaaagga
240

cgaaacacc
249

<210> 44
<211> 35
<212> ADN
<213> Homo sapiens

<400> 44
ccagaaccgg aggacaaagt acaaacggca gaagc
35

<210> 45
<211> 35
<212> ADN
<213> Homo sapiens

<400> 45
gatgtgatgg gagcccttct tcttctgctc ggact
35