



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

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

(11)



1-0045949

(51)^{2020.01} C12N 15/09; C12N 15/63; C12N 9/22; (13) B
C12N 15/55

(21) 1-2021-01594

(22) 23/08/2019

(86) PCT/JP2019/033045 23/08/2019

(87) WO 2020/045281 05/03/2020

(30) 2018-158710 27/08/2018 JP

(45) 26/05/2025 446

(43) 27/09/2021 402A

(73) HIROSHIMA UNIVERSITY (JP)

1-3-2, Kagamiyama, Higashi-Hiroshima-shi, Hiroshima 739-8511 Japan

(72) YAMAMOTO Takashi (JP); SAKUMA Tetsushi (JP); SAITO Masakazu (JP).

(74) Công ty TNHH một thành viên Sở hữu trí tuệ VCCI (VCCI-IP CO.,LTD)

(54) ENZYM PHÂN CẮT AXIT NUCLEIC NHÂN TẠO, AXIT NUCLEIC ĐƯỢC
PHÂN LẬP, PHƯƠNG PHÁP CẢI BIẾN AXIT NUCLEIC ĐÍCH VÀ KIT CẢI
BIẾN AXIT NUCLEIC ĐÍCH

(21) 1-2021-01594

(57) Sáng chế đề cập đến enzym phân cắt axit nucleic nhân tạo bao gồm: vùng nucleaza là polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó; và vùng liên kết axit nucleic.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến vùng nucleaza mới và enzym phân cắt axit nucleic nhân tạo chứa vùng nucleaza này.

Tình trạng kỹ thuật của sáng chế

Trong những năm gần đây, để làm công cụ chỉnh sửa hệ gen, các nucleaza mà nhận biết đặc hiệu trình tự đích và có thể phân cắt trong vùng lân cận của chúng đã trở nên sẵn có, như nucleaza ngón tay kẽm (zinc-finger nucleaza, ZFN), nucleaza thực hiện giống chất hoạt hóa phiên mã (transcription activator-like effector nucleaza, TALEN), và hệ thống cụm lặp gập đôi ngắn cách đều/liên quan đến CRISPR (clustered regularly interspaced short palindromic repeat/CRISPR-associated, CRISPR/Cas). Các nucleaza này bao gồm vùng liên kết axit nucleic và vùng phân cắt axit nucleic, và có thể dẫn đến sự đứt sợi kép (double-strand break, DSB) của ADN ở trình tự bazơ đặc hiệu của ADN hệ gen. Do việc sửa chữa DSB gây ra bởi các nucleaza này, sự nối đầu không tương đồng (non-homologous end joining, NHEJ) và sự tái tổ hợp tương đồng (homologous recombination, HR), mà dễ bị các lỗi sửa chữa, là đã biết. Khi sửa chữa DSB, các tế bào chủ yếu sử dụng con đường NHEJ để sửa chữa, nên các lỗi sửa chữa có khả năng xảy ra, sự dịch khung xảy ra, và kết quả là, chức năng gen bị mất. Theo cách này, công nghệ sửa chữa bộ gen mà có thể thực hiện việc làm đứt gãy gen và tương tự bằng cách sử dụng cơ

chế của các tế bào đang được sử dụng rộng rãi trong khoa học đời sống như nghiên cứu bệnh và ứng dụng vào các sản phẩm nông nghiệp.

Trong ZFN và TALEN, vùng nucleaza FokI (sau đây còn được gọi là "FokI-ND") thường được sử dụng làm vùng phân cắt axit nucleic. FokI-ND đã được chứng minh là có khả năng liên kết với các axit nucleic, nhưng không nhận biết trình tự bazơ và phân cắt axit nucleic đích bằng cách tạo ra dimer. Do đó, ZFN và TALEN cần hai trình tự đích của vùng liên kết axit nucleic, và FokI-ND tác dụng trong trình tự dimer nằm giữa hai trình tự đích này để gây ra DSB. Trong ZFN, hai trình tự đích này được gọi là vị trí nửa đích L và vị trí nửa đích R.

ZFN bao gồm FokI-ND, vùng liên kết axit nucleic được tạo bởi protein ngón tay kẽm (zinc-finger protein, ZFP), và trình tự nối giữa các phân tử để nối chúng. Trong ZFN, có những hạn chế rõ rệt do độ dài của trình tự dimer và các đặc tính của trình tự nối, và cần nghĩ ra thiết kế. Do đó, để làm kỹ thuật liên quan đến các ZFN, việc thay đổi độ dài của trình tự dimer và trình tự nối đã được thông báo (tài liệu phi sáng chế 1).

Ngoài ra, protein dung hợp của vùng liên kết axit nucleic như TALE và vùng nucleaza có nguồn gốc từ chủng *Clostridium spec. 7_2_43* FAA đã được báo cáo (tài liệu sáng chế 1).

Danh mục tài liệu viện dẫn:

Tài liệu sáng chế:

[Tài liệu sáng chế 1] Patent Mỹ số 9410134

Tài liệu phi sáng chế:

[Tài liệu phi sáng chế 1] Handel, E. M., Alwin, S. and Cathomen, T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. Mol Ther, 17, 104-111

Bản chất kỹ thuật của sáng chế

Vấn đề kỹ thuật:

Mục đích của sáng chế là đề xuất vùng nucleaza mới có hoạt tính phân cắt và hiệu quả phân cắt cao cũng như có ít hạn chế hơn do trình tự đích, trình tự đệm, trình tự nối, và tương tự, và enzym phân cắt axit nucleic nhân tạo chứa vùng nucleaza này.

Giải pháp cho vấn đề:

Các tác giả sáng chế đã tìm kiếm vùng nucleaza khác với FokI-ND, mà hiện nay được sử dụng làm vùng phân cắt axit nucleic chuẩn cho các enzym chỉnh sửa hệ gen, và kết quả là đã phát hiện được vùng nucleaza mới vượt trội hơn FokI-ND hiện có về chức năng như hoạt tính phân cắt, tính đặc hiệu, tính độc tế bào, và tính chọn lọc trình tự đích. Do đó, sáng chế đã được hoàn thành.

Tức là, sáng chế đề xuất các khía cạnh sau.

(1) Enzym phân cắt axit nucleic nhân tạo bao gồm:

vùng nucleaza là polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó; và

vùng liên kết axit nucleic.

(2) Enzym phân cắt axit nucleic nhân tạo theo mục (1) nêu trên, trong đó enzym

này còn bao gồm: trình tự nối nằm giữa vùng nucleaza và vùng liên kết axit nucleic.

(3) Enzym phân cắt axit nucleic nhân tạo theo mục (1) hoặc (2) nêu trên, trong đó vùng liên kết axit nucleic chứa ngón tay kẽm, TALE, CRISPR/Cas, hoặc PPR.

(4) Axit nucleic được phân lập bao gồm: trình tự axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo mục bất kỳ trong số các mục từ (1) đến (3) nêu trên.

(5) Vector bao gồm: axit nucleic theo mục (4) nêu trên, hoặc sản phẩm phiên mã hoặc sản phẩm dịch mã của nó.

(6) Phương pháp cải biến axit nucleic đích bao gồm: đưa vào tế bào enzym phân cắt axit nucleic nhân tạo theo mục bất kỳ trong số các mục từ (1) đến (3) nêu trên, axit nucleic theo mục (4) nêu trên, hoặc vector theo mục (5) nêu trên.

(7) Phương pháp theo mục (6) nêu trên, trong đó enzym phân cắt axit nucleic nhân tạo, axit nucleic, hoặc vector là hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, axit nucleic mã hóa hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, hoặc vector chứa axit nucleic, sản phẩm phiên mã của nó, hoặc sản phẩm dịch mã của nó.

(8) Kit cải biến axit nucleic đích bao gồm: enzym phân cắt axit nucleic nhân tạo theo mục bất kỳ trong số các mục từ (1) đến (3) nêu trên, axit nucleic theo mục (4) nêu trên, hoặc vector theo mục (5) nêu trên.

(9) Kit theo mục (8) nêu trên, trong đó enzym phân cắt axit nucleic nhân tạo, axit nucleic, hoặc vector là hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, axit nucleic mã hóa hai hoặc nhiều loại trong số các

enzym phân cắt axit nucleic nhân tạo, hoặc vectơ chứa axit nucleic, sản phẩm phiên mã của nó, hoặc sản phẩm dịch mã của nó.

(10) Polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó.

(11) Axit nucleic được phân lập bao gồm: trình tự axit nucleic mã hóa polypeptit hoặc polypeptit đột biến của nó theo mục (10) nêu trên.

(12) Vectơ bao gồm: axit nucleic theo mục (11) nêu trên, hoặc sản phẩm phiên mã của nó, hoặc sản phẩm dịch mã của nó.

Các hiệu quả có lợi của sáng chế:

Vùng nucleaza theo sáng chế tạo ra tác dụng phân cắt axit nucleic tốt hơn FokI-ND thông thường về chức năng như hoạt tính phân cắt, tính đặc hiệu, và tính chọn lọc trình tự đích. Bằng cách liên kết vùng nucleaza theo sáng chế với vùng liên kết axit nucleic, enzym phân cắt axit nucleic có hoạt tính và tính linh hoạt cao trong trình tự đích được tạo ra mà phân cắt đặc hiệu trình tự đích. Enzym phân cắt axit nucleic này là rất hữu ích để làm công cụ chỉnh sửa hệ gen.

Mô tả vắn tắt các hình vẽ

Fig.1A là hình vẽ sơ lược của ZFA36-ND. Phần bên trên thể hiện rằng hai vị trí đích ZFN được ngăn cách bởi trình tự đệm. Phần bên dưới thể hiện rằng mỗi ZFN chứa ba phân tử ngón tay kềm, trình tự nối ngắn gồm 7 axit amin TGAAARA, và vùng nucleaza FokI. Khi một cặp ZFN liên kết với vị trí nửa đích được ngăn cách bởi trình tự đệm có độ dài cho trước, vùng nucleaza được dime hóa và axit nucleic có thể được phân cắt.

Fig.1B thể hiện việc đóng thẳng hàng trình tự axit amin của trình tự tương đồng vùng nucleaza FokI. Các trình tự axit amin được đóng thẳng hàng bằng cách sử dụng MAFFT. Các axit amin mà tạo thành vùng nucleaza heterodime và các đột biến điểm tương ứng được chỉ ra bằng các mũi tên, dấu hoa thị, và dấu hai chấm. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.1C thể hiện các kết quả hoạt tính tế bào bằng thử nghiệm SSA trên cơ sở tế bào HEK293T được thực hiện trong ví dụ 2. Các dữ liệu được thể hiện dưới dạng giá trị trung bình $\pm$ SEM ($n = 3$), trong đó + và - để chỉ sự có mặt hoặc không có mặt của plasmit gen thông báo trình tự đích ngón tay kẽm.

Fig.2A thể hiện hoạt tính phân cắt của vùng nucleaza đối với hệ gen đích. Các tế bào CHO-K1 được chuyển nhiễm ZFA36-ND được ủ ở nhiệt độ 37°C và thu hoạch sau 72 giờ để thu được khuôn ADN hệ gen. Sản phẩm PCR chứa vị trí đích ZFA36 được phân cắt bởi kit phát hiện phân cắt hệ gen GeneArt. Các đoạn được phân cắt được phân tích bằng cách điện di (hệ thống MultiNA). Hiệu quả phân cắt được tính từ các nồng độ mol của băng không được cắt và các đoạn được phân cắt lớn hơn. Trên hình vẽ, "FokI" có nghĩa là FokI-ND.

Fig.2B thể hiện các kết quả của thử nghiệm hoạt tính phân cắt đối với hệ gen đích của vùng nucleaza được thực hiện trong ví dụ 3 về hiệu quả phân cắt.

Fig.3 thể hiện các kết quả của phân tích thấm Western cho thấy mức biểu hiện của ZFA36-ND. Đuôi AcV5-ZFA36-ND được biểu hiện bởi các tế bào HEK293T. Các tế bào được chuyển nhiễm được thu hoạch sau 72 giờ, và dịch phân giải tế bào được thăm dò bằng các kháng thể kháng cả đuôi AcV5 và α -tubulin. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.4 thể hiện các ảnh LSM của ZFA36-ND-EGFP trong các tế bào HEK293T, và ảnh chụp hiển vi của các tế bào biểu hiện quá mức ZFA36-ND-EGFP. ADN được quan sát bằng cách nhuộm DAPI. Thanh tỷ lệ: 40 μ m. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.5 thể hiện profin hoạt tính gen thông báo của các thể đột biến trình tự nối ZFA36-ND ở các độ dài trình tự đệm khác nhau, và thể hiện các kết quả của thử nghiệm gen thông báo SSA được thực hiện trong ví dụ 5. Các thể đột biến trình tự nối ZFN và các độ dài của trình tự đệm được thể hiện là 6 bp (A), 5 bp (B), và 7 bp (C). Dữ liệu được chuẩn hóa theo hoạt tính luxiferaza của trình tự nối TGAAARA của FokI có độ dài trình tự đệm là 6bp. Các dữ liệu được thể hiện dưới dạng giá trị trung bình $\pm$ SEM (n = 3). Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.6 thể hiện profin hoạt tính gen thông báo của thể đột biến ZFA36-ND-Sharkey. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND. (A) Hình vẽ dạng lược đồ của ZFN. Các trình tự của các thể đột biến ND1, ND2, và Sharkey tương ứng với trình tự axit amin ở các vị trí 414 đến 445 của FokI-ND được thể hiện. Trên hình vẽ này, số của axit amin là dựa trên axit amin có độ dài đầy đủ của mỗi vùng nucleaza. (B) Các kết quả của thử nghiệm gen thông báo SSA được thực hiện trong ví dụ 6 được thể hiện, trong đó - để chỉ rằng plasmit ZFN là không có mặt. Các dữ liệu được thể hiện dưới dạng giá trị trung bình $\pm$ SEM (n = 3).

Fig.7 thể hiện hoạt tính nucleaza trong một lần sử dụng của các ZFN có các đột biến kiểu heterodime được xác định bằng thử nghiệm SSA trên cơ sở tế bào. Các tế bào HEK293T được đồng chuyển nhiễm bằng mỗi plasmit

biểu hiện ZFN, plasmit gen thông báo, và plasmit tham chiếu để thực hiện thử nghiệm luxiferaza kép. Sau khi chuyển nhiễm, các tế bào được ủ trong 24 giờ, phân giải, và phân tích hoạt tính luxiferaza. (A) Gen thông báo hPGRN_ZFL1-6bp-hPGRN_ZFL1. (B) Gen thông báo ZFA36-6bp-ZFA36. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.8A thể hiện việc phân tích các phản ứng phân cắt homodime và heterodime bằng thử nghiệm gen thông báo SSA. Các tế bào HEK293T được đồng chuyển nhiễm bằng mỗi cặp plasmit biểu hiện ZFN, các plasmit gen thông báo, và các plasmit tham chiếu để thực hiện thử nghiệm luxiferaza kép. Sau khi chuyển nhiễm, các tế bào được ủ trong 24 giờ, phân giải, và phân tích hoạt tính luxiferaza. Các dữ liệu được thể hiện dưới dạng giá trị trung bình $\pm$ SEM ($n = 3$). Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.8B thể hiện việc phân tích các phản ứng phân cắt homodime và heterodime bằng thử nghiệm Cel-I. Các tế bào CHO-K1 được chuyển nhiễm bằng mỗi plasmit ZFN được ủ ở nhiệt độ 37°C trong 72 giờ. Sau đó, ADN hệ gen được tạo ra từ các tế bào được chuyển nhiễm để khuếch đại các đoạn chứa vị trí đích ZFA36-ZFL1. Sản phẩm PCR được phân cắt bằng kit phát hiện phân cắt hệ gen GeneArt. Các đoạn được phân cắt được phân tích bằng cách điện di trên gel agarosa. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.9 thể hiện hoạt tính nucleaza của các ZFN lai như được xác định bằng thử nghiệm SSA trên cơ sở tế bào. Thử nghiệm SSA trên cơ sở tế bào HEK293T. Các tế bào HEK293T được đồng chuyển nhiễm bằng mỗi cặp plasmit biểu hiện ZFN, plasmit gen thông báo, và plasmit tham chiếu để thực hiện thử nghiệm luxiferaza kép. Sau khi chuyển nhiễm, các tế bào được ủ trong

24 giờ, phân giải, và phân tích hoạt tính luxiferaza. Các cột trắng và xám lần lượt để chỉ sự không có mặt của ZFA36-FokI và ZFN. (A) ZFL1-FokI-DDD và vùng nucleaza thích hợp. (B) ZFL1-ND1-DDD và vùng nucleaza thích hợp. (C) ZFL1-ND2-DDD và vùng nucleaza thích hợp. Các dữ liệu được thể hiện dưới dạng giá trị trung bình $\pm$ SEM ($n = 3$). Trên hình vẽ, "FokI" có nghĩa là FokI-ND.

Fig.10 thể hiện các kết quả của phân tích trình tự của vị trí đích được thực hiện trong ví dụ 9. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.11A thể hiện các kết quả của thử nghiệm phân cắt của enzym phân cắt axit nucleic nhân tạo chứa vùng nucleaza mới theo sáng chế và TALE, được thực hiện trong ví dụ 10. Trên hình vẽ này, "FokI" có nghĩa là FokI-ND.

Fig.11B thể hiện các kết quả của thử nghiệm phân cắt của enzym phân cắt axit nucleic nhân tạo chứa vùng nucleaza mới theo sáng chế và TALE, được thực hiện trong ví dụ 10. Trên hình vẽ, "FokI" có nghĩa là FokI-ND.

Fig.12A thể hiện các kết quả của việc phát hiện, bằng thử nghiệm gen thông báo SSA, hoạt tính nucleaza của ZF-ND1 và ZF-FokI được đưa vào các tế bào thực vật bằng phương pháp tạo lỗ bằng điện.

Fig.12B là đồ thị của các kết quả của Fig.12A.

Fig.13 thể hiện các kết quả của việc phát hiện, bằng thử nghiệm gen thông báo SSA, hoạt tính nucleaza của ZF-ND1 và ZF-FokI được đưa vào các tế bào thực vật bằng phương pháp tải nạp protein.

Mô tả chi tiết sáng chế

1. Các vùng nucleaza mới

Để phát triển vùng nucleaza mới có tính năng cao hơn so với FokI-ND, các tác giả sáng chế đã tìm kiếm các vật liệu tự nhiên tương đồng bằng công cụ BLAST protein-protein dựa trên trình tự axit amin của FokI-ND. Các tác giả sáng chế đã chọn một số loại từ các trình tự tương đồng có mức độ tương đồng với FokI-ND nằm trong khoảng từ 35% đến 70% để thực hiện tiếp phân tích này, và đã phát hiện được vùng nucleaza 1 (sau đây còn được gọi là "ND1") và vùng nucleaza 2 (sau đây còn được gọi là "ND2") có hoạt tính nucleaza tốt hơn so với hoạt tính của FokI-ND.

Trình tự axit amin của protein có độ dài đầy đủ chứa ND1 (có nguồn gốc từ *Bacillus* SGD-V-76) được nêu trong SEQ ID NO: 1, và trình tự bazơ của nó được nêu trong SEQ ID NO: 2. Trình tự axit amin của protein có độ dài đầy đủ chứa ND2 (có nguồn gốc từ *Clostridium botulinum*) được nêu trong SEQ ID NO: 3, và trình tự bazơ của nó được nêu trong SEQ ID NO: 4. ND1 thường là một phần peptit tương ứng với các vị trí 391 đến 585 của SEQ ID NO: 1, và ND2 thường là một phần peptit tương ứng với các vị trí 389 đến 579 của SEQ ID NO: 3.

ND1 có mức độ giống trình tự axit amin 70% với FokI-ND, và ND2 có mức độ giống trình tự axit amin 57% với FokI-ND, cả hai vùng này là các polypeptit mới được phân lập lần đầu tiên theo sáng chế.

Do đó, sáng chế đề xuất vùng nucleaza là polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó (sau đây còn được gọi là "vùng nucleaza theo sáng chế"). Vùng nucleaza theo sáng chế bao gồm vùng nucleaza là polypeptit bao gồm trình tự axit amin được nêu ở các vị trí

391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó. Nói cách khác, sáng chế đề xuất việc sử dụng polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó, ví dụ polypeptit bao gồm trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó, làm vùng nucleaza.

Theo sáng chế, polypeptit đột biến của polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 là polypeptit có trình tự axit amin giống với trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 và có hoạt tính nucleaza. Các ví dụ về polypeptit đột biến bao gồm polypeptit chứa trình tự axit amin trong đó một đến một số gốc axit amin được thế, xóa, chèn, hoặc bổ sung vào trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 và có hoạt tính nucleaza, và polypeptit bao gồm trình tự axit amin trong đó một đến một số axit amin được thế, xóa, chèn, hoặc bổ sung vào trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 và có hoạt tính nucleaza. Ở đây, "một đến một số" là, ví dụ, 1 đến 10, tốt hơn là 1 đến 6, và là, ví dụ, 1, 2, 3, 4, hoặc 5.

Ngoài ra, theo một ví dụ khác, polypeptit đột biến có thể là polypeptit chứa trình tự axit amin có mức độ giống trình tự ít nhất 70%, ít nhất 75%, ít nhất 80%, ít nhất 85%, ít nhất 90%, ít nhất 95%, và ít nhất 98% với trình tự

axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 và có hoạt tính nucleaza, ví dụ có thể là polypeptit bao gồm trình tự axit amin có mức độ giống trình tự ít nhất 70%, ít nhất 75%, ít nhất 80%, ít nhất 85%, ít nhất 90%, ít nhất 95%, và ít nhất 98% với trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 và có hoạt tính nucleaza. Ngoài ra, các ví dụ khác về polypeptit đột biến bao gồm polypeptit chứa trình tự axit amin được mã hóa bởi trình tự bazơ có mức độ giống trình tự ít nhất 50%, ít nhất 55%, ít nhất 60%, ít nhất 65%, ít nhất 70%, ít nhất 75%, ít nhất 80%, ít nhất 85%, ít nhất 90%, ít nhất 95%, và ít nhất 98% với trình tự bazơ được nêu ở các vị trí 1171 đến 1755 của SEQ ID NO: 2 hoặc các vị trí 1165 đến 1737 của SEQ ID NO: 4 và có hoạt tính nucleaza, và polypeptit bao gồm trình tự axit amin được mã hóa bởi trình tự bazơ có mức độ giống trình tự ít nhất 50%, ít nhất 55%, ít nhất 60%, ít nhất 65%, ít nhất 70%, ít nhất 75%, ít nhất 80%, ít nhất 85%, ít nhất 90%, ít nhất 95%, và ít nhất 98% với trình tự bazơ được nêu ở các vị trí 1171 đến 1755 của SEQ ID NO: 2 hoặc các vị trí 1165 đến 1737 của SEQ ID NO: 4 và có hoạt tính nucleaza.

Theo sáng chế, "mức độ giống trình tự" của trình tự axit amin được xác định bằng cách so sánh hai trình tự được dóng thẳng hàng để tối đa hóa mức độ so khớp trình tự. Các phương pháp xác định giá trị bằng số (%) của mức độ giống trình tự là đã biết đối với người có hiểu biết trong lĩnh vực này. Để làm thuật toán để thu được sự dóng thẳng hàng tối ưu và mức độ giống trình tự, thuật toán bất kỳ mà người có hiểu biết trong lĩnh vực này đã biết (như thuật toán BLAST và thuật toán FASTA) có thể được sử dụng. Mức độ giống trình

tự của trình tự axit amin được xác định bằng cách sử dụng, ví dụ, chương trình phần mềm phân tích trình tự như BLASTP hoặc FASTA.

Ngoài ra, polypeptit đột biến nêu trên có thể tốt hơn là chứa trình tự axit amin có axit aspartic ở vị trí tương ứng với vị trí 483 trong trình tự axit amin của FokI (sau đây còn được gọi là "D483") và/hoặc arginin ở vị trí tương ứng với vị trí 487 trong trình tự axit amin của FokI (sau đây còn được gọi là "R487"). Ngoài ra, polypeptit đột biến nêu trên có thể tốt hơn là chứa trình tự axit amin có axit aspartic, axit aspartic, và lysin ở các vị trí tương ứng với các vị trí 450, 467 và 469 trong trình tự axit amin của FokI (sau đây còn được gọi lần lượt là "D450," "D467," và "K469"). Tốt hơn nữa nếu, polypeptit đột biến chứa trình tự axit amin có D483 và/hoặc R487 và có D450, D467, và K469. D483 và R487 là các gốc axit amin được cho là cần thiết đối với FokI-ND để tạo thành dime có hoạt tính phân cắt, và được giữ lại trong các trình tự axit amin của ND1 và ND2. Ngoài ra, liên quan đến vị trí xúc tác của FokI-ND, việc phân tích bằng cách sử dụng các sự thay thế axit amin đã cho thấy rằng D450 và D467 có liên quan đến sự phân cắt của các axit nucleic. Ngoài ra, trình tự bao gồm K469 ở vùng lân cận của các axit amin này là motif được tìm thấy trong EcoRI và EcoRV, và đã được làm rõ là các axit amin này có liên quan đến hoạt tính phân cắt liên kết phosphat dieste. D450, D467, và K469 được giữ lại trong các trình tự axit amin của ND1 và ND2.

Ngoài ra, polypeptit đột biến nêu trên có thể chứa axit aspartic ở các vị trí tương ứng với các vị trí 483, 487, và 496 trong trình tự axit amin của FokI (sau đây còn được gọi là "D483, D487, và D496") (sau đây còn được gọi là "thể đột biến kiểu DDD"), hoặc có thể chứa arginin ở các vị trí tương ứng với

các vị trí 483, 487, và 537 trong trình tự axit amin của FokI (sau đây còn được gọi là "R483, R487, và R537") (sau đây, còn được gọi là "thể đột biến kiểu RRR"), để cho vùng nucleaza theo sáng chế tạo thành heterodime khi phân cắt axit nucleic.

Độ dài mạch của polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3 hoặc polypeptit đột biến theo sáng chế thường là 500 axit amin hoặc ít hơn, ví dụ, 400 axit amin hoặc ít hơn, 300 axit amin hoặc ít hơn, hoặc 200 axit amin hoặc ít hơn.

Hơn nữa, theo sáng chế, các polypeptit đột biến nêu trên còn bao gồm các polypeptit chứa axit amin được cải biến và/hoặc axit amin không tự nhiên. Axit amin được cải biến bao gồm, nhưng không chỉ giới hạn ở, metyl hóa, este hóa, amit hóa, axetyl hóa, alkyl hóa, halogen hóa, và tương tự. Axit amin được cải biến và axit amin không tự nhiên có thể được đưa vào bằng phương pháp đã biết.

Vùng nucleaza theo sáng chế có hoạt tính nucleaza bằng hoặc cao hơn so với hoạt tính của FokI-ND. Ví dụ, vùng nucleaza theo sáng chế có hoạt tính nucleaza ít nhất 0,8 lần, ít nhất 0,9 lần, ít nhất 1 lần, ít nhất 1,3 lần, ít nhất 1,5 lần, ít nhất 1,8 lần, ít nhất 2 lần, ít nhất 2,3 lần, ít nhất 2,5 lần, ít nhất 2,8 lần, ít nhất 3 lần, ít nhất 3,3 lần, ít nhất 3,5 lần, ít nhất 3,8 lần, hoặc ít nhất 4 lần so với FokI-ND. Hoạt tính nucleaza có thể được khẳng định bằng các phương pháp đã biết trong lĩnh vực này. Ví dụ, vùng nucleaza có thể được đưa vào hệ thống ZFN, và được khẳng định bằng cách xác định hoạt tính phân cắt axit nucleic, như hoạt tính phân cắt ADN, bằng phương pháp gắn sợi đơn (single-

strand annealing, SSA) và phương pháp Cel-I.

Như được mô tả sau đây, vùng nucleaza theo sáng chế có các đặc tính khác với các đặc tính của FokI-ND về tính linh hoạt trong việc chọn trình tự nối trong enzym phân cắt axit nucleic nhân tạo và tính linh hoạt trong việc chọn trình tự đệm làm vị trí phân cắt đích (ví dụ 5). Ngoài ra, ngay cả khi đột biến Sharkey được biết là gây ra sự tăng hoạt tính trong FokI-ND được đưa vào vùng nucleaza theo sáng chế, không có sự tăng hoạt tính nào được quan sát (ví dụ 6). Ngoài ra, khác với FokI-ND, ND1 theo sáng chế không làm giảm hoạt tính nucleaza ngay cả khi heterodime được tạo ra (ví dụ 7). Do đó, vùng nucleaza theo sáng chế là vùng nucleaza có các đặc tính tốt hơn và/hoặc các đặc tính khác so với FokI-ND thông thường.

Vùng nucleaza theo sáng chế có thể là vùng nucleaza được phân lập. Sáng chế còn đề xuất axit nucleic được phân lập mã hóa vùng nucleaza theo sáng chế. Như được sử dụng ở đây, thuật ngữ "được phân lập" để chỉ trạng thái được tách ra từ thể giới tự nhiên hoặc cơ thể sống.

2. Enzym phân cắt axit nucleic nhân tạo

Enzym phân cắt axit nucleic nhân tạo theo sáng chế bao gồm vùng nucleaza và vùng liên kết axit nucleic theo sáng chế. Enzym phân cắt axit nucleic nhân tạo liên kết với trình tự đích trên axit nucleic thông qua vùng liên kết axit nucleic, và phân cắt axit nucleic ở vị trí phân cắt đích bằng vùng nucleaza. Do đó, enzym phân cắt axit nucleic nhân tạo là enzym phân cắt axit nucleic đặc hiệu trình tự có khả năng phân cắt đặc hiệu vị trí đích trong axit nucleic.

Theo sáng chế, "axit nucleic" bao gồm cả ADN và ARN. Các ví dụ về

axit nucleic được phân cắt bởi enzym phân cắt axit nucleic nhân tạo theo sáng chế bao gồm ADN sợi kép, ADN sợi đơn, và ARN sợi đơn. Các ví dụ về ADN bao gồm, nhưng không chỉ giới hạn ở, ADN hệ gen của nhân của tế bào chưa có nhân điển hình, ADN ty thể, ADN lập thể, ADN hệ gen của tế bào chưa có nhân điển hình, ADN thể thực khuẩn, ADN plasmit, và tương tự. Tốt hơn nếu enzym phân cắt axit nucleic nhân tạo theo sáng chế phân cắt ADN sợi kép trên hệ gen. Các ví dụ về ARN bao gồm, nhưng không chỉ giới hạn ở, ARN sợi đơn, ARN sợi kép, hoặc ADN-ARN sợi kép lai bao gồm ADN sợi đơn và ARN sợi đơn.

Vùng liên kết axit nucleic có thể là vùng protein mà liên kết đặc hiệu với trình tự axit nucleic tùy ý (trình tự đích), và các ví dụ về chúng bao gồm, nhưng không chỉ giới hạn ở, các ngón tay kẽm, TALE, CRISPR/Cas (phức chất của protein Cas và ARN dẫn đường), đoạn lặp pentatricopeptit (pentatricopeptide repeat, PPR), và tương tự, và các vùng liên kết ADN chứa các tổ hợp của chúng. Các ví dụ về CRISPR/Cas bao gồm CRISPR-Cas9, CRISPR-Cpf1 (Cas12a), CRISPR-Cas12b, CRISPR-CasX (Cas12e), CRISPR-Cas14, CRISPR-Cas3, và tương tự. Vùng liên kết axit nucleic chứa các ngón tay kẽm có thể tốt hơn là chứa hai hoặc nhiều ngón tay kẽm, và các ví dụ về chúng có thể bao gồm, nhưng không chỉ giới hạn ở, mảng ngón tay kẽm bao gồm 3 đến 9 ngón tay kẽm (sau đây còn được gọi là "ZFA"). Đã biết rằng một ngón tay kẽm nhận biết trình tự có ba bazơ, ví dụ, ngón tay kẽm nhận biết GNN, ANN, CNN, hoặc tương tự. Tốt hơn nếu CRISPR/Cas bao gồm Cas có hoạt tính nucleaza bị bất hoạt, như dCas (Cas có hoạt tính nucleaza bị bất hoạt, như dCas9). Các vùng liên kết axit nucleic được thiết kế thích hợp bởi người

có hiểu biết thông thường trong lĩnh vực này để liên kết với trình tự đích mong muốn. Lưu ý rằng theo sáng chế, "CRISPR/Cas" có nghĩa là phức chất của ARN dẫn đường và protein Cas, và ARN dẫn đường này nhận biết và liên kết với trình tự đích.

Trong enzym phân cắt axit nucleic nhân tạo theo sáng chế, vùng nucleaza và vùng liên kết axit nucleic có thể được liên kết trực tiếp hoặc có thể được liên kết qua trình tự nối. Trình tự nối này bao gồm, ví dụ, hai hoặc nhiều gốc axit amin, và độ dài của nó là không bị giới hạn cụ thể, nhưng có thể là, ví dụ, độ dài từ 2 đến 20 axit amin, ví dụ 2, 4, 6, 8, 9, 9, 12, 16, hoặc 20 axit amin. Loại trình tự nối cũng là không bị giới hạn cụ thể, và các ví dụ về chúng bao gồm TGAAARA, GS, RPGEKP, TGPGAAARA, và tương tự. Sự có mặt hoặc không có mặt của trình tự nối và độ dài và loại trình tự nối được chọn thích hợp bởi người có hiểu biết thông thường trong lĩnh vực này khi cân nhắc loại và yếu tố tương tự của vùng liên kết axit nucleic. Đã biết rằng hoạt tính nucleaza của các ZFN thông thường bao gồm FokI-ND bị ảnh hưởng nhiều bởi trình tự nối để nối ZF và FokI-ND, nhưng trong trường hợp enzym phân cắt axit nucleic nhân tạo theo sáng chế, sự phụ thuộc của hoạt tính nucleaza của nó vào trình tự nối là thấp hơn so với của các ZFN thông thường. Cụ thể, trong ND1, hầu như không có sự giảm hoạt tính nucleaza được quan sát bằng cách thay đổi trình tự nối (ví dụ 5). Do đó, enzym phân cắt axit nucleic nhân tạo theo sáng chế có tính linh hoạt cao trong trình tự nối.

Trình tự đích của enzym phân cắt axit nucleic nhân tạo theo sáng chế là trình tự tùy ý trên axit nucleic. Trình tự đích tốt hơn là hai trình tự được kẹp giữa trình tự đệm. Độ dài của trình tự đệm, độ dài của các trình tự đích, và các

trình tự bazơ là không bị giới hạn cụ thể. Người có hiểu biết thông thường trong lĩnh vực này có thể thiết lập thích hợp các trình tự đích có độ dài mong muốn và bao gồm các trình tự bazơ. Hai trình tự đích có thể là các trình tự gấp đôi hoặc các trình tự không gấp đôi với nhau. Khi hai trình tự đích là các trình tự không gấp đôi, hai loại enzym phân cắt axit nucleic nhân tạo hướng đích trình tự tương ứng được tạo ra. Ngoài ra, trong trường hợp sử dụng vùng liên kết axit nucleic chứa CRISPR/Cas, các trình tự nằm ở vùng lân cận của trình tự motif tiền trình tự nối (protospacer adjacent motif, PAM) được chọn làm các trình tự đích. Tức là, các trình tự đích được thiết lập sao cho PAM có mặt bên ngoài của cả hai trình tự đích hoặc trong trình tự đệm. Các trình tự đích có thể là hoặc trình tự sợi kép hoặc trình tự sợi đơn. Ngoài ra, trong bản mô tả này, "vùng lân cận" bao gồm cả các vị trí liền kề hoặc các vị trí gần.

Trong enzym phân cắt axit nucleic nhân tạo theo sáng chế, khi vùng liên kết axit nucleic liên kết với trình tự đích trên axit nucleic, vùng nucleaza phân cắt axit nucleic ở vị trí phân cắt đích. Vị trí phân cắt đích nằm trong hoặc ở gần trình tự đệm liền kề với các trình tự đích. Độ dài của mảng trình tự đệm là không bị giới hạn cụ thể, và các ví dụ về chúng bao gồm 1 đến 20 bp, tốt hơn là 3 đến 8 bp, và tốt hơn nữa là 5 đến 7 bp. Trong trường hợp các ZFN thông thường bao gồm FokI-ND, đã biết rằng độ dài của trình tự đệm tối ưu là 6 bp, và hoạt tính nucleaza của nó bị ảnh hưởng nhiều bởi độ dài của trình tự đệm, nhưng trong trường hợp enzym phân cắt axit nucleic nhân tạo theo sáng chế, độ dài của trình tự đệm là linh hoạt hơn độ dài của các ZFN thông thường. Enzym phân cắt axit nucleic nhân tạo theo sáng chế có hoạt tính phân cắt cao ngay cả đối với các trình tự đệm ngắn và các trình tự đệm dài bằng cách thay

đổi độ dài của trình tự nối (ví dụ 5). Mặt khác, trong các ZFN thông thường bao gồm FokI-ND, ngay cả khi độ dài của trình tự nối được thay đổi, hoạt tính phân cắt là thấp đáng kể đối với các trình tự đệm ngắn hơn hoặc dài hơn độ dài tối ưu của trình tự đệm. Do độ dài của trình tự đệm xác định khoảng cách giữa các trình tự nhận biết, enzym phân cắt axit nucleic nhân tạo theo sáng chế có nhiều sự lựa chọn của các vị trí phân cắt đích và các trình tự nhận biết do tính linh hoạt cao của độ dài trình tự đệm.

Enzym phân cắt axit nucleic nhân tạo theo sáng chế có thể được tạo ra in vitro hoặc in vivo bằng phương pháp đã biết trong lĩnh vực này. Các ví dụ về chúng bao gồm phương pháp tổng hợp nhân tạo dựa trên thông tin trình tự axit amin, và phương pháp trong đó axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế hoặc mỗi axit nucleic mã hóa vùng nucleaza và vùng liên kết axit nucleic theo sáng chế được tổng hợp nhân tạo, được chèn vào vectơ biểu hiện thích hợp, và sau đó được đưa vào tế bào chủ thích hợp để biểu hiện enzym phân cắt axit nucleic nhân tạo trong tế bào. Theo cách khác, phương pháp có thể được kể đến trong đó axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế được tổng hợp bằng cách dịch mã in-vitro hoặc in-vivo để tổng hợp protein mà được đưa vào tế bào thích hợp để cho phép enzym phân cắt axit nucleic nhân tạo tác dụng trong tế bào, hoặc ARN mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế được tổng hợp bằng cách phiên mã in-vitro, được đưa vào tế bào chủ thích hợp để cho phép enzym phân cắt axit nucleic nhân tạo tác dụng trong tế bào. Vectơ biểu hiện là không bị giới hạn cụ thể, và có thể được chọn từ các vectơ được sử dụng trong lĩnh vực này. Các ví dụ về chúng bao gồm các vectơ plasmit, các vectơ virus,

các vectơ thể thực khuẩn, các vectơ phagemit, các vectơ BAC, các vectơ YAC, các vectơ MAC, các vectơ HAC, và tương tự.

Tế bào chủ là không bị giới hạn cụ thể, và các ví dụ về chúng bao gồm tế bào chưa có nhân điển hình như *Escherichia coli*, xạ khuẩn, và cổ khuẩn, tế bào có nhân điển hình như nấm men, nhím biển, con tằm, cá ngựa vằn, chuột nhắt, chuột, ếch, cây thuốc lá, cây cải xoong (*Arabidopsis thaliana*), và cây lúa, và các tế bào nuôi cấy.

Một khía cạnh nữa của sáng chế còn đề xuất axit nucleic được phân lập chứa trình tự axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế, cũng như sản phẩm phiên mã và sản phẩm dịch mã của axit nucleic này. Axit nucleic nêu trên có thể là axit nucleic được phân lập bao gồm trình tự axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế. Ngoài ra, việc tối ưu hóa codon có thể được thực hiện. Theo sáng chế, "axit nucleic" bao gồm cả ADN và ARN.

3. Phương pháp cải biến axit nucleic đích

Khi enzym phân cắt axit nucleic nhân tạo theo sáng chế được đưa vào tế bào, enzym này phân cắt vị trí đích trên axit nucleic trong tế bào, và sau đó, sự phân cắt này được sửa chữa bằng sự nối đầu không tương đồng (NHEJ), sự tái tổ hợp tương đồng (HR), hoặc tương tự. Lúc này, trong quá trình sửa chữa bằng NHEJ là con đường sửa chữa chính, một hoặc nhiều đột biến được chèn vào vị trí phân cắt để cải biến axit nucleic. Do đó, một khía cạnh nữa của sáng chế đề xuất phương pháp cải biến axit nucleic đích, bao gồm việc đưa vào tế bào enzym phân cắt axit nucleic nhân tạo theo sáng chế hoặc axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo này. Enzym hoặc axit nucleic được

mô tả ở trên có thể là sản phẩm phiên mã của axit nucleic hoặc sản phẩm dịch mã của nó. Axit nucleic cũng có thể được tối ưu hóa codon để biểu hiện ở mức cao trong tế bào. Ngoài ra, theo sáng chế, "cải biến" bao gồm việc xóa, chèn, hoặc thay thế ít nhất một nucleotit, hoặc tổ hợp của chúng. Ngoài ra, theo sáng chế, thuật ngữ "đột biến" được sử dụng đối với các axit nucleic có nghĩa là xóa, chèn, hoặc thay thế nucleotit.

Việc đưa enzym hoặc axit nucleic nêu trên vào tế bào có thể là việc đưa vào bằng phương pháp vật lý, đưa vào thông qua sự nhiễm virus hoặc sinh vật, hoặc tương tự, và các phương pháp khác nhau đã biết trong lĩnh vực này có thể được sử dụng. Các ví dụ về phương pháp đưa vào vật lý bao gồm, nhưng không chỉ giới hạn ở, phương pháp tạo lỗ bằng điện, phương pháp dùng súng bắn hạt, phương pháp vi tiêm, phương pháp chuyển gen qua liposom, phương pháp tải nạp protein, và tương tự. Các ví dụ về phương pháp đưa vào thông qua sự nhiễm virus hoặc sinh vật bao gồm, nhưng không chỉ giới hạn ở, tải nạp virus, phương pháp chuyển gen thông qua *Agrobacterium*, gây nhiễm thể thực khuẩn, tiếp hợp, và tương tự. Các phương pháp đưa vào nêu trên liên quan đến việc sử dụng thích hợp các vector. Các vector có thể được chọn thích hợp theo loại tế bào để đưa vào, và các ví dụ về chúng bao gồm, nhưng không chỉ giới hạn ở, các vector plasmit, các vector virus, các vector thể thực khuẩn, các vector phagemit, các vector BAC, các vector YAC, các vector MAC, các vector HAC, và tương tự. Các ví dụ về chúng còn bao gồm các vector như liposom và lentivirus có khả năng mang các sản phẩm dịch mã (protein) và sản phẩm phiên mã (mARN), và các vector peptit như các peptit có khả năng thấm tế bào có khả năng đưa các phân tử được dung hợp hoặc liên kết vào tế bào.

Do đó, một khía cạnh nữa của sáng chế đề xuất vector chứa axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo sáng chế, hoặc vector chứa sản phẩm phiên mã của axit nucleic hoặc sản phẩm dịch mã của nó. Vector chứa sản phẩm phiên mã của axit nucleic hoặc sản phẩm dịch mã của nó bao gồm các sản phẩm được phiên mã bởi axit nucleic trong vector chứa axit nucleic, và các sản phẩm trong đó sản phẩm dịch mã được tạo ra.

Các tế bào nêu trên có thể là hoặc tế bào chưa có nhân điển hình hoặc tế bào có nhân điển hình, và là không bị giới hạn cụ thể. Các ví dụ về chúng bao gồm vi khuẩn, cổ khuẩn, nấm men, tế bào thực vật, tế bào côn trùng, và các tế bào động vật (như tế bào của người, tế bào không phải của người, tế bào của động vật có xương sống không phải động vật có vú, và tế bào của động vật không có xương sống). Các tế bào cũng có thể là tế bào in-vivo, tế bào được phân lập, tế bào sơ cấp, hoặc các tế bào nuôi cấy. Các tế bào cũng có thể là các tế bào sinh dưỡng, các tế bào mầm, hoặc các tế bào gốc.

Cụ thể, các ví dụ về sinh vật chưa có nhân điển hình mà từ đó thu được các tế bào nêu trên bao gồm *Escherichia coli*, xạ khuẩn, cổ khuẩn, và tương tự. Ngoài ra, các ví dụ về sinh vật có nhân điển hình mà từ đó thu được các tế bào nêu trên bao gồm nấm như nấm men, nấm, và nấm mốc, động vật ngành da gai như nhím biển, sao biển, và hải sâm, côn trùng như con tằm và con ruồi, cá như cá ngừ, cá tráp, cá xem sao, và cá ngựa vằn, động vật gặm nhấm như chuột nhắt, chuột, chuột lang, chuột túi má, và sóc, động vật móng guốc chẵn như bò, lợn rừng, lợn, cừu, và dê, động vật móng guốc lẻ như ngựa, động vật bò sát như con thằn lằn, động vật lưỡng cư như ếch, bộ thỏ như thỏ, bộ ăn thịt như chó, mèo, và chồn sương, chim như gà, đà điểu, và chim cun cút, và các

thực vật như cây thuốc lá, cây *Arabidopsis thaliana*, cây lúa, cây ngô, cây chuối, cây lạc, cây hướng dương, cây cà chua, cây cải dầu, cây lúa mì, cây lúa mạch, cây khoai tây, cây đậu nành, cây bông, và cây cẩm chướng. Các ví dụ về "tế bào động vật" bao gồm tế bào phôi của phôi ở các giai đoạn khác nhau (như phôi ở giai đoạn 1 tế bào, phôi ở giai đoạn 2 tế bào, phôi ở giai đoạn 4 tế bào, phôi ở giai đoạn 8 tế bào, phôi ở giai đoạn 16 tế bào, và phôi ở giai đoạn phôi dâu); các tế bào gốc như tế bào gốc vạn năng cảm ứng (induced pluripotent stem, iPS), tế bào gốc phôi (embryonic stem, ES), các tế bào gốc tạo máu; tế bào sinh dưỡng như nguyên bào sợi, các tế bào tạo máu, tế bào thần kinh, tế bào cơ, tế bào xương, tế bào gan, tế bào tụy, tế bào não, và tế bào thận; trứng thụ tinh, và tương tự.

Axit nucleic đích có thể là gen bất kỳ trong ADN hệ gen trong tế bào hoặc ADN trong vùng ngoài gen. Để phân cắt trình tự đích trong axit nucleic đích, enzym phân cắt axit nucleic nhân tạo theo sáng chế được thiết kế sao cho vùng liên kết axit nucleic chứa trong enzym phân cắt axit nucleic nhân tạo theo sáng chế liên kết với trình tự ở vùng lân cận của trình tự đích (được chọn làm trình tự đích của vùng liên kết axit nucleic). Kết quả là, trình tự mong muốn trong axit nucleic đích được phân cắt, ví dụ, sự biểu hiện của gen được giảm đi hoặc được loại bỏ, hoặc chức năng của gen được giảm đi hoặc không chức năng nào được thể hiện.

Theo phương pháp cải biến axit nucleic đích theo sáng chế, polynucleotit cho có thể được đưa vào tế bào ngoài enzym phân cắt axit nucleic nhân tạo nêu trên. Polynucleotit cho này chứa ít nhất một trình tự cho chứa cải biến cần được đưa vào ở vị trí phân cắt đích. Polynucleotit cho có thể được

thiết kế thích hợp bởi người có hiểu biết thông thường trong lĩnh vực này dựa trên các kỹ thuật đã biết trong lĩnh vực này. Khi polynucleotit cho có mặt trong phương pháp cải biến axit nucleic đích theo sáng chế, sự tái tổ hợp tương đồng xảy ra ở vị trí phân cắt đích, và polynucleotit cho được chèn vào vị trí này hoặc vị trí này được thay bằng trình tự cho. Kết quả là, cải biến mong muốn được đưa vào ở vị trí phân cắt đích.

Khi không có mặt polynucleotit cho trong phương pháp cải biến axit nucleic đích theo sáng chế, vị trí phân cắt đích được sửa chữa chủ yếu bằng cách nối đầu không tương đồng (NHEJ). Do NHEJ dễ mắc lỗi, việc xóa, chèn, hoặc thay thế ít nhất một nucleotit, hoặc tổ hợp của chúng, có thể xảy ra trong quá trình sửa chữa vị trí phân cắt này. Theo cách này, axit nucleic được cải biến ở vị trí phân cắt đích.

Theo phương pháp cải biến axit nucleic đích theo sáng chế, hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo nêu trên có thể được đưa vào tế bào. Các ví dụ về hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo bao gồm hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các vùng liên kết axit nucleic khác nhau, hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các vùng nucleaza khác nhau, hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các vùng nucleaza và các vùng liên kết các axit nucleic khác nhau, hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các trình tự nối khác nhau, và tương tự. Hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo có các vùng liên kết axit nucleic khác nhau bao gồm hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các trình tự đích khác nhau được liên kết với các vùng liên kết axit

nucleic, và hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các loại khác nhau của vùng liên kết axit nucleic, và để làm ví dụ, có thể nêu ra hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo có các vùng nucleaza khác nhau và các trình tự đích khác nhau được liên kết với các vùng liên kết axit nucleic. Ví dụ, khi không có trình tự gấp đôi nào có mặt trong vùng cần cải biến trên axit nucleic, hai loại enzym phân cắt axit nucleic nhân tạo hướng đích các trình tự khác nhau có thể được sử dụng. Trong trường hợp này, đột biến mà thúc đẩy sự tạo thành heterodime của các vùng nucleaza có thể được đưa thêm vào các vùng nucleaza của hai loại enzym phân cắt axit nucleic nhân tạo. Điều này làm cho có thể làm tăng các trình tự nhận biết của các enzym phân cắt axit nucleic nhân tạo và làm giảm khả năng chệch đích. Các đột biến này là, ví dụ, các thay thế vào D483, D487, và D496 (sau đây còn được gọi là “các đột biến kiểu DDD”) và thay thế vào R483, R487, và R537 (sau đây còn được gọi là “các đột biến kiểu RRR”), trong đó đột biến kiểu DDD được đưa vào vùng nucleaza của một enzym phân cắt axit nucleic nhân tạo, và đột biến kiểu RRR được đưa vào vùng nucleaza của enzym phân cắt axit nucleic nhân tạo còn lại. Lưu ý rằng nếu trình tự vùng nucleaza trước khi đưa đột biến vào đã có gốc axit amin của hoặc đột biến kiểu DDD hoặc đột biến kiểu RRR được thể hiện ở trên, thì chỉ các sự thay thế axit amin còn lại cần được thực hiện. Hai loại vùng nucleaza được đưa vào với hai đột biến (như đột biến kiểu DDD và đột biến kiểu RRR) mà thúc đẩy sự tạo ra heterodime này có thể là các vùng nucleaza giống nhau ngoại trừ đột biến nêu trên, hoặc có thể là có các vùng nucleaza khác nhau ban đầu trước khi các đột biến nêu trên được đưa vào. Ví dụ, enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu DDD của ND1

và enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu RRR của ND1 có thể được sử dụng trong phương pháp theo sáng chế. Ngoài ra, ví dụ, enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu DDD của ND2 và enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu RRR của ND1 có thể được sử dụng trong phương pháp theo sáng chế. Theo sáng chế, tổ hợp của thể đột biến kiểu DDD của ND1 và thể đột biến kiểu RRR của ND1, hoặc tổ hợp của thể đột biến kiểu DDD của ND2 và thể đột biến kiểu RRR của ND1 có tính đặc hiệu và hoạt tính nucleaza đặc biệt cao (ví dụ 8).

4. Kit

Một khía cạnh nữa của sáng chế đề xuất kit cải biến axit nucleic đích, bao gồm enzym phân cắt axit nucleic nhân tạo theo sáng chế, axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo này, hoặc vector chứa axit nucleic này, sản phẩm phiên mã của axit nucleic, hoặc sản phẩm dịch mã của nó. Ngoài ra, kit này có thể bao gồm hai hoặc nhiều loại enzym phân cắt axit nucleic nhân tạo như được mô tả ở trên. Ngoài ra, kit này có thể bao gồm axit nucleic mã hóa hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, hoặc vector chứa axit nucleic này, sản phẩm phiên mã của axit nucleic này, hoặc sản phẩm dịch mã của nó. Tốt hơn nếu kit theo sáng chế bao gồm một hoặc hai loại enzym phân cắt axit nucleic nhân tạo, axit nucleic mã hóa một hoặc hai loại enzym phân cắt axit nucleic nhân tạo, hoặc vector chứa axit nucleic này, sản phẩm phiên mã của axit nucleic này, hoặc sản phẩm dịch mã của nó. Hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo có thể là, ví dụ, enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu DDD của ND1 và enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu

RRR của ND1. Hơn nữa, hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo có thể là, ví dụ, enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu DDD của ND2 và enzym phân cắt axit nucleic nhân tạo chứa thể đột biến kiểu RRR của ND1. Kit này có thể còn bao gồm một hoặc nhiều chất phản ứng bổ sung, và các ví dụ về chất phản ứng bổ sung bao gồm, nhưng không chỉ giới hạn ở, dung dịch đệm pha loãng, dung dịch tái cấu trúc, dung dịch đệm rửa, chất phản ứng chuyển axit nucleic, chất phản ứng chuyển protein, và chất phản ứng kiểm soát. Kit này thường đi kèm với sách hướng dẫn.

Nếu không được nói rõ theo cách khác, nghĩa của các thuật ngữ theo sáng chế được hiểu là các nghĩa thường được công nhận trong các lĩnh vực sinh học, sinh học phân tử, sinh hóa học, di truyền, và tương tự.

Ví dụ thực hiện sáng chế

Sau đây, sáng chế được mô tả cụ thể dựa vào các ví dụ, nhưng sáng chế không bị giới hạn ở các ví dụ này.

Lưu ý rằng các tế bào được sử dụng trong các ví dụ được nuôi cấy như sau.

Nuôi cấy tế bào

Các tế bào HEK293T được duy trì trong môi trường D-MEM (Wako Pure Chemical Industries, Ltd.) chứa 10% FBS (Difco), 1% streptomycin-penicillin (Wako Pure Chemical Industries, Ltd.), và 1% NEAA (Difco). Các tế bào CHO-K1 được duy trì bằng Ham-F12 (Difco) chứa 10% FBS, 1% streptomycin-penicillin, và 5% kanamycin (Wako Pure Chemical Industries, Ltd.). Sau đây, nếu không được chỉ rõ theo cách khác, các tế bào nuôi cấy được duy trì ở 37°C và nồng độ cacbon dioxit 5%, và nếu nhiệt độ được chỉ ra, nhiệt

độ này được sử dụng. Ngoài ra, đối với dung dịch nuôi cấy, mỗi môi trường nuôi cấy nêu trên được sử dụng nếu không được chỉ rõ theo cách khác.

Ví dụ 1: tìm kiếm các vùng nucleaza mới

Dựa trên trình tự axit amin ở phía đầu C của FokI, việc tìm kiếm các vật liệu tự nhiên tương đồng được thực hiện bằng công cụ protein-protein BLAST. Trong các kết quả tìm kiếm thu được, bốn loại ứng viên được chọn tùy ý từ các trình tự có mức độ tương đồng với FokI-ND nằm trong khoảng từ 35% đến 70%. Theo thứ tự giảm dần của trình tự có mức độ tương đồng cao, chúng được gọi là ND1 (mức độ giống trình tự axit amin 70%; trình tự axit amin có độ dài đầy đủ: SEQ ID NO: 1; trình tự bazơ tương ứng: SEQ ID NO: 2), ND2 (mức độ giống trình tự axit amin 57%; trình tự axit amin có độ dài đầy đủ: SEQ ID NO: 3; trình tự bazơ tương ứng: SEQ ID NO: 4), ND3 (mức độ giống trình tự axit amin 49%; trình tự axit amin có độ dài đầy đủ: SEQ ID NO: 5; trình tự bazơ tương ứng: SEQ ID NO: 6), và ND4 (mức độ giống trình tự axit amin 45%; trình tự axit amin có độ dài đầy đủ: SEQ ID NO: 7; trình tự bazơ tương ứng: SEQ ID NO: 8).

Khi các trình tự axit amin của ND1 đến ND4 được so sánh với trình tự axit amin của FokI-ND, đã nhận thấy rằng các độ dài của FokI-ND và vùng nucleaza gần như giống nhau, mặc dù có các sự xóa và thay thế axit amin. Ngoài ra, ND1 đến ND4 mang các axit amin (D483, R487) được cho là cần thiết đối với FokI-ND để tạo thành đime có hoạt tính phân cắt, và các axit amin (D450, D467, và K469) được cho là có liên quan đến sự phân cắt gây bởi sự thủy phân của các liên kết phosphodiester trong ADN (Fig.1B). Do đó, ND1 đến ND4 được cho là có hoạt tính nucleaza.

Ví dụ 2: Hoạt tính phân cắt của các vùng nucleaza mới (phương pháp SSA)

Để đánh giá hoạt tính nucleaza của ND1 đến ND4, plasmit ZFN (pSTL-ZFA36-ND) chứa mỗi vùng trong số ND1 đến ND4 được tạo ra, và hoạt tính phân cắt của mỗi vùng nucleaza mới được xác định bằng thử nghiệm gen thông báo SSA.

(1) Thiết kế plasmit ZFN

Các trình tự bazơ của vùng ND1 đến ND4 được tối ưu hóa codon và được tổng hợp nhân tạo bởi IDT. Plasmit ZFN được sử dụng (pSTL-ZFA36) là cấu trúc được tạo ra bởi Ochiai và các đồng tác giả (Ochiai et al. (2010) Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes to Cells* 15: 875-885). Để thay thế FokI-ND của pSTL-ZFA36 bằng ND1 đến ND4 mới xác định được, PCR được thực hiện bằng PrimeSTAR Max (Takara) với mỗi cặp mồi thay thế vùng nucleaza. PCR được thực hiện ở 98°C trong 2 phút, sau đó là 40 chu trình 98°C trong 10 giây, 60°C trong 5 giây, và 68°C trong 40 giây. Sản phẩm PCR thu được và ND1 đến ND4 được tổng hợp nhân tạo được cho phản ứng bằng cách sử dụng In-Fusion (Clontech) và được biến nạp vào *Escherichia coli*. Plasmit được tách chiết từ thể biến nạp thu được và trình tự bazơ được khẳng định để thu được mỗi plasmit pSTL-ZFA36-ND mong muốn. Fig.1A thể hiện hình vẽ dạng lược đồ của ZFA (mảng ngón tay kềm) 36-ND thu được. ZFA36-ND này có trình tự nối TGAAARA.

(2) Thử nghiệm gen thông báo SSA

7,2µl dung dịch plasmit được tạo ra chứa 120ng plasmit gen thông báo SSA, 240ng mỗi plasmit pSTL-ZFA36-ND nêu trên, và 24ng plasmit tham chiếu pRL-CMV (Promega) (plasmit để chuẩn hóa hiệu quả tải nạp). Plasmit

gen thông báo SSA được tạo ra theo phương pháp đã biết để trình tự đích của mảng ZF (ZFA36) chứa trình tự đệm 6bp được chèn vào sao cho nằm giữa các vùng chồng nhau của trình tự Luc (Ochiai et al. (2010) Targeted mutagenesis in the sea urchin embryo using zinc-finger nucleases. *Genes to Cells* 15: 875-885). Đĩa nhiều lỗ có 96 lỗ (Iwaki) đã xử lý bằng cách phủ polylysin được bổ sung 25µl dung dịch nuôi cấy D-MEM không chứa huyết thanh, và 6µl dung dịch plasmit được trộn vào. 25,7µl dung dịch chứa 0,7µl Lipofectamine LTX (Life Technologies) cho 25µl dung dịch nuôi cấy D-MEM không chứa huyết thanh được bổ sung vào đĩa này, và hỗn hợp này được để yên ở nhiệt độ trong phòng trong 30 phút. Trong thời gian này, các tế bào HEK293T được tách ra khỏi đĩa nuôi cấy mô 100mm (Iwaki) bằng cách xử lý bằng trypsin, ly tâm ở tốc độ 1000 vòng/phút trong 3 phút, dịch nổi bề mặt được loại bỏ, và các tế bào được tạo hỗn dịch lại trong 5ml dung dịch nuôi cấy. 10µl hỗn dịch tế bào được lấy, và số lượng tế bào được đếm. Từ số lượng tế bào thu được, dung dịch nuôi cấy được bổ sung vào và được điều chỉnh để số lượng tế bào trong hỗn dịch tế bào là 6×10^5 tế bào/ml. Đĩa 96 lỗ nêu trên được để yên ở nhiệt độ trong phòng trong 30 phút, 100µl hỗn dịch tế bào được bổ sung vào. Sau đó, đĩa có 96 lỗ được chuyển sang 37°C và nồng độ cacbon dioxit 5%, và được nuôi cấy để thực hiện quá trình tải nạp.

Đối với dung dịch nuôi cấy của các tế bào HEK293T được nuôi cấy trong 24 giờ sau khi tải nạp, 75µl của nó được loại bỏ, 75µl cơ chất Dual Glo luciferaza được gắn với hệ thống phân tích Dual Glo Luciferaza (Promega) được bổ sung vào mỗi lỗ, và các tế bào được xử lý trong chỗ tối ở nhiệt độ trong phòng trong 10 phút. Sau đó, hoạt tính luxiferaza được xác định bằng

TriStar LB941 của máy đọc đĩa nhiều lỗ. Sau khi đo, 75 μ l dung dịch đệm kép Dual Glo Stop & Glo chứa cơ chất kép Glo Stop & Glo được pha loãng 100 lần được bổ sung vào mỗi mẫu, và hỗn hợp được xử lý trong chỗ tối ở nhiệt độ trong phòng trong 10 phút, và hoạt tính luxiferaza của plasmid tham chiếu được đo bằng máy đọc đĩa theo cách giống như trên. Để làm đối chứng, ZFA36-FokI chứa FokI-ND được sử dụng cho phép đo này. Đối với các kết quả đo, sự thay đổi của giá trị thu được bằng cách chia hoạt tính luxiferaza của gen thông báo SSA cho hoạt tính luxiferaza của plasmid tham chiếu được tính, và giá trị tương đối trong mỗi mẫu khi giá trị của ZFA36-FokI được điều chỉnh bằng 1 đã thu được.

(3) Các kết quả

Fig.1C thể hiện các kết quả này. Như đã rõ từ Fig.1C, ND1 và ND2 có hoạt tính phân cắt cao hơn khoảng 20% hoặc trên 20% so với FokI-ND được sử dụng thông thường. Mặt khác, đã nhận thấy rằng ND3 và ND4 có hoạt tính phân cắt thấp và chỉ có khoảng 20% hoạt tính của FokI-ND.

Ngoài ra, đã khẳng định được, khi vùng nucleaza không được biểu hiện trong tế bào, nucleaza có mặt ban đầu trong tế bào có hướng đích gen thông báo luciferaza hay không. Kết quả là, đã phát hiện được rằng gen thông báo luxiferaza không thể là đích của các nucleaza khác nhau có mặt trong các tế bào HEK293T do hoạt tính tương đối đối với FokI-ND là vô cùng thấp. Phân tích này đã cho thấy rằng ND1 và ND2 là các vùng nucleaza mới có nguồn gốc từ các vật liệu tự nhiên vượt trội so với hoạt tính phân cắt của FokI-ND được sử dụng thông thường trong các ZFN.

Ví dụ 3: Hoạt tính phân cắt của các vùng nucleaza mới đối với hệ gen đích

(phương pháp Cel-I)

Để khẳng định mức độ mà các vùng nucleaza mới ND1 và ND2 được khẳng định là có hoạt tính nucleaza trong ví dụ 2 gây ra sự đột biến trong ADN hệ gen in vivo, thử nghiệm Cel-I được sử dụng để phân tích tỷ lệ đột biến vào ADN hệ gen trong các tế bào nuôi cấy của các vùng nucleaza mới. Để làm đối chứng, FokI-ND, ND3, và ND4 được phân tích theo cách giống hệt.

(1) Thử nghiệm Cel-I

Trong các tế bào được tải nạp bằng công cụ chỉnh sửa hệ gen như TALEN và ZFN, ADN đích được phân cắt bởi vùng nucleaza của chúng và vị trí phân cắt được sửa chữa bằng cơ chế sửa chữa có trong tế bào, nhưng lỗi sửa chữa xảy ra trong quá trình này. Khi vùng chứa trình tự đích của ADN hệ gen này được khuếch đại bằng PCR, sản phẩm hỗn hợp của đoạn PCR của alen kiểu dại và đoạn PCR chứa đột biến đã thu được. Sự tái liên kết của các sản phẩm này sau khi phân ly tạo ra ADN sợi kép bắt cặp sai bao gồm đoạn PCR của alen kiểu dại và alen chứa đột biến, chúng được phân cắt bằng endonucleaza đặc hiệu bắt cặp sai (nucleaza Cel-I). Sự gây đột biến có thể được đánh giá bằng cách phát hiện kiểu phân cắt.

Các tế bào CHO-K1 được phân ly bằng cách xử lý bằng trypsin, ly tâm ở tốc độ 1000 vòng/phút trong 3 phút, dịch nổi bề mặt được loại bỏ, và các tế bào được tạo hỗn dịch trong 10ml dung dịch nuôi cấy. Số lượng tế bào được xác định bằng cách sử dụng 10µl hỗn dịch tế bào, dung dịch nuôi cấy được bổ sung vào để số lượng tế bào là 1×10^5 tế bào/ml, và các tế bào được nuôi cấy qua đêm ở 37°C và nồng độ cacbon dioxit 5%.

Mỗi plasmit ZFN được tạo ra trong ví dụ 2 được bổ sung vào 500µl

Opti-MEM (Difco) để có khối lượng là 800ng, và hỗn hợp được để yên ở nhiệt độ trong phòng trong 5 phút. 500 μ l Opti-MEM được bổ sung 32 μ l Lipofectamine LTX, và toàn bộ lượng này được trộn với môi trường Opti-MEM chứa plasmit nêu trên, và được để yên ở nhiệt độ trong phòng trong 30 phút. Trong thời gian này, dung dịch nuôi cấy của các tế bào CHO-K1 được phát triển qua đêm được thay bằng 4ml Opti-MEM. Dung dịch nuôi cấy tế bào được bổ sung toàn bộ lượng này của dung dịch plasmit đã được để yên ở nhiệt độ trong phòng trong 30 phút, và các tế bào được nuôi cấy qua đêm ở 37°C và nồng độ cacbon dioxid 5%. Buổi sáng ngày tiếp theo, dung dịch nuôi cấy được thay từ Opti-MEM thành 10ml dung dịch nuôi cấy, và quá trình nuôi cấy được tiếp tục thêm ở 37°C và nồng độ cacbon dioxid 5%. Ngoài ra, puromycin (Wako Pure Chemical Industries, Ltd.) được bổ sung vào để cho nồng độ cuối là 500 μ g/ml sau 24 giờ. Sau đó, quá trình nuôi cấy được thực hiện hàng ngày trong 3 ngày trong khi thay dung dịch nuôi cấy và bổ sung puromycin.

Các tế bào CHO-K1 được tải nạp bằng plasmit ZFN như được mô tả ở trên được thu hoạch từ đĩa nuôi cấy bằng cách xử lý bằng trypsin. Các tế bào thu hoạch được rửa một lần bằng PBS (-). Các tế bào thu được được xử lý bằng dung dịch đệm phân giải tế bào và chất phân hủy protein chứa trong hoặc kit tinh chế ADN (Qiagen) hoặc kit phát hiện phân cắt hệ gen GeneArt (Life Technologies) để thu được ADN hệ gen. Với ADN hệ gen thu được làm khuôn, PCR được thực hiện bằng cách sử dụng KOD Fx Neo (ToYoBo) hoặc AmpliTaq Gold (Life Technologies). Trong trường hợp KOD Fx Neo, PCR được thực hiện ở 95°C trong 2 phút, sau đó là 40 chu trình 95°C trong 30 giây, 60°C trong 30 giây, và 68°C trong 30 giây để khuếch đại gen đích. Trong

trường hợp AmpliTaq Gold, việc xử lý được thực hiện ở 95°C trong 10 phút, sau đó là 40 chu trình 95°C trong 10 giây, 60°C trong 30 giây, và 72°C trong 40 giây, và cuối cùng xử lý ở 72°C trong 7 phút được thực hiện để khuếch đại gen đích. Lượng tương đương của mỗi sản phẩm PCR thu được được phản ứng phân cắt bằng enzym phát hiện bằng cách sử dụng kit phát hiện phân cắt hệ gen GeneArt theo sách hướng dẫn. Trong dung dịch sau khi phản ứng, các đoạn được phân cắt được phát hiện bằng cách điện di trên gel agarosa hoặc MultiNA (Shimadzu Corporation). Hiệu quả phân cắt được tính từ các nồng độ mol của các băng không được phân cắt và các đoạn được phân cắt lớn hơn.

(2) Các kết quả

Fig.2 thể hiện các kết quả của phương pháp điện di bằng MultiNA. Bảng 1 thể hiện các trình tự ứng viên của đích hệ gen ZFN (ZFA36-ZFA36) trong các tế bào CHO-K1 được sử dụng trong phân tích này. Bảng 1 còn thể hiện các trình tự ứng viên của đích hệ gen ZFN (ZFL1-ZFA36) được sử dụng trong ví dụ 7 được mô tả sau đây.

Bảng 1

Trình tự của các vị trí đích	
Vị trí đích	Trình tự của vị trí đích
ZFA36-ZFA36	ACCATCTTCnnnnnnGAAGATGGT
vị trí thứ nhất	ACCATCTTCcactctGAAGATGGA
vị trí thứ hai	ACCATCTTCaagagaGAAGATGGC
vị trí thứ ba	ACCATCTTCcttgatGAAGATGCC
ZFL1-ZFA36	GTCACCTTCnnnnnnGAAGATGGT
vị trí thứ nhất	GTCACCTTCaagtctGAAGATGGT
vị trí thứ hai	TTCACCTTCttaagtGAAGATGGT

Trong tất cả các trình tự đích được phân tích, vị trí nửa đích (L) hoàn

toàn phù hợp với trình tự đích của mảng ZF. Mặt khác, vị trí nửa đích (R) không phù hợp hoàn toàn với trình tự đích của mảng ZF, và có sự bất cặp sai ở phía đầu 3'. Sự phân cắt của các trình tự ứng viên đích này bằng Cel-I cho thấy rằng hiệu quả phân cắt của các vùng nucleaza mới ND1 và ND2 cao hơn khoảng 1,5 lần so với hiệu quả của FokI-ND, giống với các kết quả của phân tích gen thông báo luciferaza (Fig.2A và Fig.2B). Cụ thể, hoạt tính gây đột biến của ND1 có hoạt tính phân cắt thích hợp hơn và cao hơn so với của ND2 trong tất cả các trình tự ứng viên đích được phân tích. Các kết quả này đã cho thấy rằng ND1 và ND2 là các vùng nucleaza mới có nguồn gốc từ các vật liệu tự nhiên hoạt tính hơn so với FokI-ND thông thường cũng đối với các trình tự hệ gen đích in vivo, giống với các kết quả của việc phân tích gen thông báo bằng cách sử dụng sự phát quang của luciferaza trong ví dụ 2.

Mặt khác, trong ND3, mặc dù hoạt tính gây đột biến thay đổi phụ thuộc vào trình tự ứng viên đích, hoạt tính này bằng khoảng 25 đến 45% hoạt tính của FokI-ND. Ngoài ra, trong ND4, không có đột biến nào trong các trình tự ứng viên đích có thể phát hiện được quan sát trong tất cả các trình tự ứng viên đích được phân tích.

Ví dụ 4: Phân tích lượng tích lũy trong tế bào của các vùng nucleaza mới

Để làm rõ việc có hay không nguyên nhân mà tỷ lệ gây đột biến của vùng nucleaza mới là cao hơn so với của FokI-ND là tỷ lệ gây đột biến biểu kiến là cao hơn do lượng tích lũy trong tế bào là lớn hơn so với của FokI-ND, lượng của mỗi vùng nucleaza được tích lũy được phân tích bằng cách bổ sung đuôi vào ZFN và sử dụng kháng thể nhận biết nó. Để thực hiện phân tích này, cấu trúc trong đó đuôi AcV5 được bổ sung mới vào đầu N của ZFN được tạo

ra và được sử dụng cho phân tích này.

(1) Phân tích lượng tích lũy trong tế bào (phân tích thâm Western)

Để phân tích lượng ZFN được tích lũy trong tế bào, PCR được thực hiện bằng cách sử dụng mỗi pSTL-ZFA36-ND được tạo ra trong ví dụ 2 làm khuôn, và sau phản ứng In-Fusion, sự biến nạp vào *Escherichia coli* được thực hiện để thu được mỗi plasmid ZFN có đuôi AcV5 được bổ sung vào phía đầu N của mảng ZF. Dung dịch plasmid với lượng 7,2 μ l chứa 240ng plasmid ZFN được tạo ra. Dung dịch plasmid này được tải nạp vào các tế bào HEK293T theo cách giống như trong thử nghiệm SSA được mô tả trong ví dụ 2, và được nuôi cấy ở 37°C và nồng độ cacbon dioxit 5% trong 3 ngày. Dung dịch nuôi cấy được thay đổi mỗi ngày.

Các tế bào HEK293T được nuôi cấy được xử lý bằng trypsin và ly tâm, và dịch nổi bề mặt được loại bỏ. Chất kết tủa thu được được tạo hỗn dịch trong PBS, và bổ sung dung dịch xử lý mẫu bao gồm 2% SDS, 10% glyxerol, 10mM DTT, 0,005% BPB, và 62,5mM Tris-HCl (pH=6,8), và hỗn hợp được đun sôi ở 95°C trong 5 phút. Sau đó, các axit nucleic và tương tự được cắt bằng cách xử lý bằng siêu âm. Các mẫu protein thu được được định lượng bằng cách sử dụng kit thử nghiệm protein (Bio-Rad) bằng cách sử dụng BSA làm chất chuẩn ngoại. Mỗi mẫu được điều chế 10 μ g được tách bằng SDS-PAGE và chuyển sang màng PVDF bằng cách chuyển bán khô. Các màng PVDF sau phiên mã được khóa bằng PBS chứa 5% sữa gầy và ủ với kháng thể α -tubulin được pha loãng 2000 lần (Abcam) và kháng thể đuôi AcV5 (Sigma). Sau khi rửa bằng PBST, chúng được ủ với HRP liên hợp với kháng thể IgG của dê kháng chuột nhất được pha loãng 2000 lần (Bio-Rad). Sau khi rửa bằng PBST, chúng được

làm cho phát sáng bằng cách sử dụng cơ chất có thời gian tín hiệu dài Supersignal West Dura (Thermo Fisher) và tiếp xúc với phim X quang.

(2) Các kết quả

Fig.3 thể hiện các kết quả. Khi lượng protein trong dung dịch phá vỡ tế bào được so sánh bằng cách sử dụng kháng thể đơn dòng kháng α -tubulin làm chất chuẩn nội, lượng tubulin được tích lũy là lớn trong các tế bào đối chứng không được tải nạp bằng ZFN. Khi ZFN chứa mỗi vùng nucleaza được biểu hiện, FokI-ND và ND1 dường như ít hơn một chút, nhưng có lượng tích tụ tubulin gần bằng nhau, và lượng tách protein gần bằng nhau bằng SDS-PAGE và phiên mã sang màng PVDF.

Mặt khác, như có thể thấy từ lượng được tích lũy của tubulin, trong mẫu đối chứng, mặc dù việc phân tích được thực hiện với lượng lớn hơn của protein so với dung dịch phá vỡ tế bào biểu hiện mỗi vùng nucleaza, không có tín hiệu không đặc hiệu nào có nguồn gốc từ kháng thể đuôi AcV5 được quan sát, và đã phát hiện được rằng kháng thể AcV5 nhận biết đặc hiệu đuôi này. Khi lượng protein được tích lũy được phân tích bằng cách sử dụng kháng thể AcV5 này, trong dung dịch phá vỡ tế bào được tải nạp bằng mỗi vùng nucleaza, tín hiệu có nguồn gốc từ đuôi AcV5 đã thu được trong FokI-ND và ND1 đến ND3 ở khoảng 35 kDa là kích thước được mong đợi. Ngoài ra, các tín hiệu của FokI-ND, ND1, và ND2 gần như giống nhau. Khi được xem xét kết hợp với lượng tubulin được sử dụng làm chất chuẩn nội, lượng tích lũy trong tế bào của mỗi vùng nucleaza FokI-ND, ND1, và ND2 gợi ý rằng FokI-ND được tích lũy nhiều nhất trong tế bào, sau đó là ND1 và ND2 theo thứ tự này. Mặt khác, đối với ND3, do tín hiệu có nguồn gốc từ đuôi AcV5 đã thu được mạnh hơn

đáng kể so với của FokI-ND và tương tự, đã nhận thấy rằng lượng ND3 được tích lũy trong tế bào là lớn hơn so với của FokI-ND và tương tự.

(3) Phân tích lượng tích lũy trong tế bào và định vị (phân tích sự biểu hiện ZFN-EGFP)

Ngoài ra, mặc dù hoạt tính phân cắt bị mất, cấu trúc trong đó EGFP được dung hợp với đầu C được tạo ra để làm rõ lượng tích lũy trong tế bào và định vị mỗi ZFN. Cấu trúc thu được này được sử dụng để thực hiện việc tải nạp vào các tế bào HEK293T, và sự phát huỳnh quang có nguồn gốc từ GFP được quan sát bằng kính hiển vi quét laze đồng tiêu để phân tích mức biểu hiện của ZFN và đồng thời, nhuộm nhân bằng DAPI để làm rõ định vị trong tế bào.

PCR được thực hiện bằng cách sử dụng mỗi pSTL-ZFA36-ND thu được trong ví dụ 2 làm khuôn, và sau khi phản ứng in-Fusion, sự biến nạp vào *Escherichia coli* được thực hiện để thu được mỗi plasmit ZFN có EGFP được bổ sung vào phía đầu C của mảng ZF. Dung dịch plasmit với lượng 7,2μl chứa 240ng plasmit ZFN được tạo ra. Dung dịch plasmit này được tải nạp vào các tế bào HEK293T theo cách giống như trong thử nghiệm SSA được mô tả trong ví dụ 2, và được nuôi cấy ở 37°C và nồng độ cacbon dioxit 5% trong 3 ngày. Các tế bào HEK293T được gắn vào đĩa có đáy thủy tinh có 24 lỗ được xử lý bằng collagen.

Các tế bào HEK293T được nuôi cấy này được cố định bằng paraformaldehyt. Sau khi rửa bằng PBS (-), axit nucleic được nhuộm bằng DAPI. Sau khi rửa lại bằng PBS (-), việc quan sát được thực hiện bằng kính hiển vi quét laze đồng tiêu (FD-1000D, Olympus). Các ảnh huỳnh quang đã thu được ở Ex: 473 nm/Em: 510 nm đối với EGFP và ở Ex: 405 nm/Em: 473

nm đối với DAPI.

(4) Các kết quả

Fig.4 thể hiện các kết quả. Trong đối chứng âm trong đó ZFN không được biểu hiện, sự phát huỳnh quang có nguồn gốc từ GFP không được quan sát, và chỉ sự phát huỳnh quang có nguồn gốc từ axit nucleic bởi DAPI được quan sát. Trong các tế bào được đưa vào với ZFN có EGFP được dung hợp trong mỗi vùng nucleaza, sự phát huỳnh quang có nguồn gốc GFP được quan sát trong chất tế bào và nhân, mặc dù có sự khác nhau về cường độ phát quang có nguồn gốc GFP.

Khi FokI-ND-EGFP được biểu hiện, sự phát huỳnh quang của GFP được quan sát trong chất tế bào đối với một số tế bào và cả trong nhân đối với một số tế bào khác. Khi ND1-EGFP được biểu hiện, sự phát huỳnh quang có nguồn gốc GFP được quan sát trong nhiều tế bào hơn so với trong FokI-ND, nhưng cường độ của mỗi sự phát huỳnh quang là yếu hơn so với trong FokI-ND. Khi ND2-EGFP được biểu hiện, đã phát hiện được rằng trong các tế bào được quan sát, số lượng tế bào có sự phát huỳnh quang có nguồn gốc GFP là nhỏ hơn so với trong FokI-ND, nhưng nó nằm trong chất tế bào và nhân. Khi ND3-EGFP được biểu hiện, sự phát huỳnh quang có nguồn gốc GFP được quan sát trong nhân và chất tế bào, giống như trong FokI-ND và tương tự. Ngoài ra, do sự phát huỳnh quang có nguồn gốc GFP trong ND3 được quan sát mạnh hơn so với trong các vùng nucleaza khác, đã cho rằng mức biểu hiện là cao hơn trong các tế bào. Phân tích ND3 bằng phân tích thấm Western cũng cho thấy rằng lượng tích lũy trong tế bào là cao hơn so với của FokI-ND, nhưng mặc dù nó có thể nằm trong nhân, tỷ lệ gây đột biến biểu kiến là thấp, và đã cho rằng các

đặc tính sinh hóa của enzym cũng là khác nhau, như nhiệt độ tối ưu khác với của FokI-ND.

(5) Tóm lại

Từ các kết quả của phân tích kháng thể và phân tích sự phát huỳnh quang GFP, đã nhận thấy rằng FokI-ND, ND1, và ND2 được tích lũy trong các tế bào ở mức độ giống nhau. Điều này đã cho thấy rằng hoạt tính phân cắt cao của các vùng nucleaza mới ND1 và ND2 không phải do lượng tích lũy trong tế bào lớn, mà là do hoạt tính phân cắt cao hơn so với của FokI-ND, có nghĩa là ND1 và ND2 là các vùng nucleaza có nguồn gốc từ các vật liệu tự nhiên có hoạt tính phân cắt vượt trội hơn FokI-ND mà được sử dụng thông thường trong các ZFN và tương tự.

Ví dụ 5: Phân tích tính chọn lọc trình tự đích của các vùng nucleaza mới

Để xác nhận thêm tính hữu ích của ND1 và ND2, đã thử cải biến trình tự nối giữa mảng ZF và vùng nucleaza. Trong các ZFN thông thường, đã được thông báo rằng khi trình tự nối được cải biến, độ dài của trình tự đệm có hoạt tính phân cắt và tính độc được tạo ra cho tế bào là khác nhau. Do đó, ngoài trình tự nối TGAAARA, là trình tự nối thông thường được sử dụng trong phân tích này cho đến nay, các vector ZFN được cải biến vào trình tự nối GS, trình tự nối RPGEKP, và trình tự nối TGPGAAARA được thiết kế và được sử dụng cho phân tích này.

(1) Phương pháp thử nghiệm

Để làm vector ZFN có trình tự nối TGAAARA, mỗi pSTL-ZFA36-ND thu được trong ví dụ 2 được sử dụng. Việc chọn trình tự nối GS, trình tự nối RPGEKP, và trình tự nối TGPGAAARA là dựa trên các thông báo trong tài

liệu: Handel et al. (2009) (NPL 1), Wilson et al. (2013) (Wilson et al. (2013) Expanding the Repertoire of Target Sites for Zinc Nuclease-mediated Genome modification. *Mol. Ther-Nucleic Acids Ther-Các axit nucleic* 2: e88), và Nomura et al. (2012) (Nomura et al. (2012) Effects of DNA Binding of the Zinc Finger and Linkers for Domain Fusion on the Catalytic Activity of Sequence-Specific Chimeric Recombinase Determined by a Facile Fluorescent System. *Biochemistry*. 51: 1510-1517). Vector ZFN với mỗi trình tự nối được cải biến được PCR bằng PrimeSTAR Max (Takara) với mỗi cặp đoạn mỗi cải biến trình tự nối bằng cách sử dụng pSTL-ZFA36-ND làm khuôn. Sản phẩm PCR thu được được cho phản ứng bằng cách sử dụng In-Fusion (Clontech) và được biến nạp vào *Escherichia coli*. Plasmid được tách chiết từ thể biến nạp thu được, và trình tự bazơ được khẳng định để thu được plasmit mong muốn. Ngoài ra, các plasmit gen thông báo SSA có các độ dài của trình tự đệm được cải biến thành 5 bp hoặc 7 bp được sử dụng để thực hiện thử nghiệm gen thông báo SSA như được mô tả trong ví dụ 2 để xác định hoạt tính nucleaza của mỗi ZFN.

(2) Các kết quả

Fig.5 thể hiện các kết quả. Lưu ý rằng hoạt tính phân cắt bao gồm trình tự nối thông thường và được thể hiện bằng hoạt tính tương đối khi hoạt tính phân cắt của ZFA-FokI-ND của trình tự đệm 6bp là 100%.

Khi độ dài của trình tự đệm là 6 bp, trong FokI-ND, hoạt tính này được giảm tới 20% hoặc ít hơn khi được cải biến thành trình tự nối GS hoặc trình tự nối RPGEKP, so với hoạt tính phân cắt khi mảng ZF và vùng nucleaza được nối bằng trình tự nối thông thường (TGAAARA). Mặt khác, khi được cải biến

thành trình tự nối TGPGAAARA, hoạt tính phân cắt của FokI-ND là khoảng 90% của hoạt tính trước khi cải biến trình tự nối (Fig.5A).

Trong ND1, khi trình tự đệm là 6 bp, hoạt tính phân cắt không bị ảnh hưởng ngay cả khi trình tự nối được cải biến, và hoạt tính phân cắt là khoảng 90% hoạt tính của FokI-ND trước khi cải biến trình tự nối. Trong ND2, khi trình tự đệm là 6 bp, hoạt tính phân cắt được giảm tới khoảng 60% khi được cải biến thành trình tự nối GS và tới khoảng 35% khi được cải biến thành trình tự nối RPGEKP, so với hoạt tính phân cắt của FokI-ND trước khi cải biến trình tự nối. Từ các kết quả này, đã thấy rõ rằng, như được thông báo trước đây, ngay cả khi các trình tự đệm có độ dài giống nhau, trình tự nối được sử dụng trong phân tích này ảnh hưởng đến hoạt tính phân cắt của vùng nucleaza.

Khi độ dài của trình tự đệm được điều chỉnh bằng 5 bp, hoạt tính phân cắt của FokI-ND được giảm tới khoảng 40% trong trình tự nối thông thường (trình tự nối TGAAARA) so với trường hợp 6 bp, nhưng khi được cải biến thành trình tự nối GS, khoảng 50% hoạt tính phân cắt vẫn giữ được. Tuy nhiên, khi được cải biến thành trình tự nối RPGEKP hoặc trình tự nối TGPGAAARA, hoạt tính phân cắt của FokI-ND là khoảng 20% hoạt tính của trình tự nối thông thường, và hầu như không có sự phân cắt nào có thể thực hiện. Đã nhận thấy rằng trong ND1, khi trình tự đệm là 5 bp, hoạt tính phân cắt vẫn còn cao khi được cải biến thành trình tự nối GS, nhưng trình tự nối thông thường giảm hoạt tính phân cắt tới khoảng 30%, và khi được biến đổi thành các trình tự nối khác, hầu như không có sự phân cắt nào có thể thực hiện.

Trong ND2, khi độ dài của trình tự đệm được điều chỉnh bằng 5 bp, hầu như không có hoạt tính phân cắt nào được thể hiện ngay cả khi trình tự nối

được cải biến. Khi độ dài của trình tự đệm là ngắn, sự ảnh hưởng của trình tự nối là lớn, và đã cho rằng trình tự nối càng ngắn, càng dễ duy trì hoạt tính phân cắt.

Khi độ dài của trình tự đệm được đặt là 7 bp, trong FokI-ND, hoạt tính phân cắt được quan sát chỉ khi được cải biến thành trình tự nối TGPGAAARA, và hoạt tính này là khoảng 50% hoạt tính phân cắt của trình tự nối thông thường có độ dài trình tự đệm 6 bp. Hầu như không có hoạt tính phân cắt nào được thể hiện khi được biến đổi thành các trình tự nối khác.

Trong ND1, đã phát hiện được rằng khi độ dài của trình tự đệm được đặt là 7 bp, hoạt tính phân cắt không được thể hiện khi được cải biến thành trình tự nối GS, nhưng khi được cải biến thành các trình tự nối khác, hoạt tính phân cắt bằng hoặc cao hơn khoảng 50% được thể hiện. Cụ thể, trình tự nối TGPGAAATA được cho là có hoạt tính phân cắt cao hơn hoạt tính phân cắt của FokI-ND đích. Mặt khác, đối với ND2, trình tự nối thông thường có khoảng 40% hoạt tính phân cắt, nhưng khi được cải biến thành các trình tự nối khác, hầu như không có hoạt tính phân cắt nào được thể hiện. Khi độ dài của trình tự đệm là dài, trình tự nối càng dài thì hoạt tính phân cắt có xu hướng càng cao.

Từ các kết quả nêu trên, đã được làm rõ là ND1 có ít khả năng bị cải biến trình tự nối hơn FokI-ND, có ít hạn chế hơn đối với trình tự đệm, và có hoạt tính phân cắt cao hơn so với FokI-ND. Điều này cho thấy rằng ND1 là vùng nucleaza linh hoạt hơn so với FokI-ND cả về tính chọn lọc trình tự đích và tương tự.

Cho đến nay, đã có các thông báo về việc cải biến FokI-ND đối với

ZFN. Cũng đã có thông báo rằng chính vùng nucleaza được cải biến từ FokI-ND thành endonucleaza như PvuII và I-TevI. Mặc dù ZFN được cải biến bằng cách sử dụng I-TevI này có hoạt tính phân cắt cao hơn so với FokI-ND, trình tự đệm có độ dài khoảng 30 bp mà dài hơn so với của FokI do trình tự nối nội phân tử và motif liên kết ADN trong I-TevI, ngoài mảng ZF. Tương tự, khi được cải biến thành PvuII, trình tự đệm là dài. Kết quả là, đã cho rằng trình tự ADN tác dụng lên ZFN cũng dài và dễ dàng được cắt bằng trình tự bazơ. Mặt khác, khi được sử dụng trong ZFN, vùng nucleaza theo sáng chế có cấu trúc protein giống như FokI-ND thông thường nhưng có tỷ lệ đột biến cao, và trình tự ADN đích tác dụng lên ZFN được rút ngắn tới 24 bp toàn bộ. Từ điều này, vùng nucleaza theo sáng chế ít bị cắt bởi trình tự bazơ hơn I-TevI và tương tự, và dễ dàng xử lý như một công cụ chỉnh sửa hệ gen đơn giản hơn.

Ví dụ 6: Hoạt tính phân cắt của vùng nucleaza mới được đưa vào bằng đột biến Sharkey

Sự thay thế axit amin (đột biến Sharkey) được biết là làm tăng hoạt tính phân cắt khi được đưa vào FokI-ND (Guo, J., Gaj, T. and Barbas, C. F., 3rd. (2010) Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol*, 400, 96-107). Do đó, đã thử đưa các sự thay thế axit amin giống như đột biến Sharkey vào các vùng nucleaza ND1 và ND2 mới nhằm mục đích làm tăng hơn nữa hoạt tính phân cắt.

(1) Phương pháp thử nghiệm

Các vectơ ZFN chứa vùng nucleaza được đưa vào bằng các đột biến Sharkey (các sự thay thế axit amin S418P và K441E trong FokI-ND, và các sự thay thế axit amin tương ứng trong các vùng nucleaza khác) được thực hiện

PCR bằng PrimeSTAR Max (Takara) với cặp đoạn mỗi đột biến Sharkey bằng cách sử dụng mỗi pSTL-ZFA36-ND làm khuôn dựa vào thông báo của Guo và các đồng tác giả. Sản phẩm PCR thu được được cho phản ứng bằng cách sử dụng In-Fusion (Clontech) và được biến nạp vào *Escherichia coli*. Plasmid được tách chiết từ thể biến nạp thu được, và trình tự bazơ được khẳng định để thu được plasmid mong muốn. Thử nghiệm gen thông báo SSA được thực hiện như được mô tả trong ví dụ 2 để xác định hoạt tính nucleaza của mỗi ZFN. Để làm đối chứng không có đột biến Sharkey, mỗi pSTL-ZFA36-ND thu được trong ví dụ 2 được sử dụng.

(2) Các kết quả

Bằng cách so sánh các trình tự axit amin của FokI-ND, đột biến Sharkey, và ND1 và ND2, vị trí thay thế axit amin được tạo ra bởi đột biến Sharkey được xác định (Fig.6). Fig.6 thể hiện các kết quả. Lưu ý rằng mỗi hoạt tính phân cắt được thể hiện dưới dạng hoạt tính tương đối khi hoạt tính phân cắt của ZFA-FokI-ND không có đột biến Sharkey bằng 1.

Trong ND1, các axit amin tương ứng với các vị trí S418 và K441 của FokI-ND, mà là các vị trí axit amin cần thiết để đưa đột biến Sharkey vào, là giống như của FokI-ND và lần lượt là S424 và K447. Mặt khác, trong ND2, axit amin tương ứng với vị trí S418 của FokI-ND là giống như của đột biến Sharkey (P422), nhưng axit amin tương ứng với vị trí K441 của FokI-ND là S445. Do đó, ND1 và ND2, trong đó các sự thay thế axit amin giống như đột biến Sharkey được đưa vào các vị trí axit amin nêu trên, lần lượt được thiết kế là ND1-Sharkey và ND2-Sharkey. Khi hoạt tính phân cắt được xác định bằng cách sử dụng gen thông báo luciferaza bằng cách sử dụng các vùng này, hoạt

tính phân cắt được tăng khoảng 1,5 lần trong đột biến Sharkey của FokI-ND (Fig.6B). Như đã thông báo, đột biến Sharkey đã được chứng minh là làm tăng hoạt tính phân cắt của FokI-ND. Tuy nhiên, trong cả ND1-Sharkey và ND2-Sharkey, hoạt tính phân cắt là thấp hơn so với trước khi đưa các axit amin thay thế vào, và hoạt tính phân cắt cao kỳ vọng đã không thu được. Kết quả này có nghĩa là cơ chế phân cắt của các nucleaza ND1 và ND2 là khác với của FokI-ND. Do đó, các vùng nucleaza ND1 và ND2 mới là các vùng nucleaza khác với FokI-ND.

Ví dụ 7: Hoạt tính phân cắt bằng cách sử dụng các vùng nucleaza kiểu heterodime

Khi trình tự đích được phân cắt bởi ZFN, FokI-ND tạo thành dime. Ngoài ra, đã biết rằng, bằng cách thay thế axit amin cụ thể, FokI-ND có thể được biến đổi chức năng thành thể đột biến như kiểu DDD/kiểu RRR và tương tự mà chỉ có hoạt tính phân cắt ở thời điểm tạo heterodime. Do đó, việc phân tích thêm được thực hiện để làm rõ xem các cải biến chức năng để duy trì hoạt tính phân cắt sẽ có thể có trong các vùng nucleaza mới ND1 và ND2 hay không.

Bằng cách dóng thẳng hàng với FokI-ND, đã kiểm tra xem các axit amin liên quan đến sự tạo heterodime cũng được bảo toàn trong ND1 và ND2 hay không (Fig.1B). Kết quả là, axit amin khác với gốc histidin có mặt phần lớn ở phía đầu C trong vị trí thay thế axit amin của nucleaza dạng heterodime được bộc lộ bởi FokI-ND cũng được bảo toàn trong ND1 và ND2. Từ điều này, có thể suy ra rằng ND1 và ND2 có thể được biến đổi chức năng thành dạng heterodime theo cách giống như FokI-ND. Do đó, trong ND1 và ND2, hoạt tính phân cắt trong quá trình tạo heterodime được xác định.

(1) Phương pháp thử nghiệm

Các thay thế được đưa vào ở các vị trí axit amin thích hợp như được thể hiện trong Bảng 2 để thực hiện việc biến đổi chức năng thành dạng heterodime cho ND1 và ND2. Lưu ý rằng trong bảng này, "FokI" có nghĩa là FokI-ND. Trong bảng này, các số của axit amin là dựa trên axit amin có độ dài đầy đủ của mỗi vùng nucleaza.

Bảng 2

Các trình tự axit amin của các vùng nucleaza FokI và các trình tự tương đồng

Nucleaza		Trình tự axit amin của vùng nucleaza
FokI	-	GQADEMQRYVE.. TRNKH ..RLNHITN
FokI-DDD	R487D/N496D	GQADEMQDYVE.. TRDKH ..RLNHITN
FokI-RRR	D483R/H537R	GQAREMQRYVE.. TRNKH ..RLNRITN
ND1	-	SQADEMQRYVD.. NRNAI ..RVSNLTK
ND1-DDD	R493D/N502D	SQADEMQDYVD.. NRDAI ..RVSNLTK
ND1-RRR	D489R/N543R	SQAREMQRYVD.. NRNAI ..RVSRLTK
ND2	-	SQADEMERYRE.. DENEH ..RISIDTG
ND2-DDD	R487D/N496D	SQADEMEDYRE.. DEDEH ..RISIDTG
ND2-RRR	D483R/I537R	SQAREMERYRE.. DENEH ..RISRDTG

Để phân tích sự tạo heterodime, vị trí nửa đích L và vị trí nửa đích R, là các trình tự đích của mảng ZF, phải là các trình tự khác nhau. Cấu trúc được thiết kế bằng cách sử dụng mảng ZF mà trình tự đích của nó là trình tự khác với mảng ZF được sử dụng trong có ví dụ trên đây. Để làm mảng ZF, hPGRN_ZFL1 được tổng hợp nhân tạo bởi IDT (sau đây còn được gọi là "ZFL1"). Để thay thế mảng ZF của mỗi pSTL-ZFA36-ND thu được trong ví dụ 2 bằng hPGRN_ZFL1, mỗi pSTL-ZFA36-ND được sử dụng làm khuôn với cặp

mỗi thay thế mảng ZF, và PCR được thực hiện bằng PrimeSTAR Max. Sản phẩm PCR thu được và hPGRN_ZFL1 được cho phản ứng bằng cách sử dụng In-Fusion và được biến nạp vào *Escherichia coli*. Plasmid được tách chiết từ thể biến nạp thu được, và trình tự bazơ được khẳng định để thu được mỗi plasmid pSTL-hPGRN_ZFL1-ND mong muốn.

Ngoài ra, để biến đổi chức năng các axit amin của các vùng nucleaza mới tương ứng với các axit amin (D483, R487, và N496) liên quan đến sự heterodimer hóa FokI-ND thành kiểu DDD, mỗi pSTL-hPGRN_ZFL1-ND được sử dụng làm khuôn và cặp đoạn mồi được sử dụng để thực hiện PCR bằng PrimeSTAR Max. Hơn nữa, để biến đổi chức năng thành kiểu RRR, mỗi pSTL-ZFA36-ND được sử dụng làm khuôn để thực hiện PCR theo cách giống hệt. Mỗi sản phẩm PCR thu được được biến nạp vào *Escherichia coli* sau khi phản ứng In-Fusion. Các plasmid được tách chiết từ thể biến nạp thu được, và sau khi khẳng định trình tự, các cấu trúc đích (sau đây còn được gọi là "thể đột biến kiểu DDD ZFL1-ND" và "thể đột biến kiểu RRR ZFA36-ND") được tạo ra. ZFL1-ND kiểu DDD và ZFA36-ND kiểu RRR thu được được sử dụng để thực hiện thử nghiệm gen thông báo SSA như được mô tả trong ví dụ 2 để xác định hoạt tính nucleaza của mỗi ZFN. Lưu ý rằng lượng mỗi plasmid ZFN trong 7,2µl dung dịch plasmid được sử dụng trong thử nghiệm gen thông báo SSA được thay đổi thành 120ng đối với vị trí nửa đích L và 120ng đối với vị trí nửa đích R. Ngoài ra, hoạt tính nucleaza được xác định bằng thử nghiệm Cel-I theo cách giống như được mô tả trong ví dụ 3.

(2) Các kết quả

Fig.7 và Fig.8A thể hiện các kết quả của thử nghiệm gen thông báo SSA.

Hoạt tính phân cắt được tính bằng cách sử dụng trình tự đệm 6bp giống như trong phân tích gen thông báo cho đến nay, và ở dạng hoạt tính tương đối so với hoạt tính phân cắt của ZFA36-FokI mà nhận biết hai trình tự đích bên trái và bên phải. Fig.8B thể hiện các kết quả của thử nghiệm Cel-I.

Hoạt tính phân cắt của trình tự đích (ZFL1-ZFL1) đối với gen thông báo được xác định với cấu trúc trong đó mỗi vùng nucleaza được nối với ZFL1 để trở thành homodime, và hoạt tính của ND1 và ND2 là cao hơn so với hoạt tính của FokI-ND, giống như các kết quả được thể hiện trên Fig.1C (Fig.7A). Từ kết quả này, đã được làm rõ là hoạt tính phân cắt cao của các vùng nucleaza mới ND1 và ND2 là phổ biến ngay cả khi các mảng ZF khác nhau được sử dụng.

Khi homodime được tạo ra bằng cách sử dụng cấu trúc để nối ZFL1 và các vùng nucleaza để xác định hoạt tính phân cắt của trình tự đích (ZFL1-ZFA36) đối với gen thông báo, ND1 có khoảng 40% hoạt tính phân cắt, nhưng trong FokI-ND và ND2, hầu như không có hoạt tính phân cắt nào được quan sát (Fig.8A). Mặt khác, khi homodime được tạo ra bởi cấu trúc nối mảng ZF (ZFA36) được sử dụng trong các ví dụ 2 đến 6 và các vùng nucleaza để thực hiện việc phân tích theo cách giống hệt, hoạt tính phân cắt cũng được quan sát trong FokI-ND và ND2 mà lần lượt là khoảng 35% và khoảng 50%. Hơn nữa, trong ND1, hoạt tính phân cắt là cao hơn, bằng khoảng 90% (Fig.8A).

Khi heterodime được tạo ra bằng cách sử dụng cấu trúc ZFN, trong đó các mảng ZF của các vùng nucleaza kiểu homodime là ZFA36 và ZFL1, để phân tích hoạt tính phân cắt, hoạt tính của ND1 và ND2 là cao hơn so với của FokI-ND như trong trường hợp phân tích với một mình ZFA36 (Fig.8A). Từ

điều này, cũng đã phát hiện được rằng ND1 và ND2 có hoạt tính phân cắt cao và là phổ biến bất kể mảng ZF.

Tiếp theo, để kiểm tra xem dime có được tạo ra và có hoạt tính phân cắt khi chỉ mỗi vùng nucleaza kiểu heterodime được biểu hiện không, hoạt tính phân cắt của trình tự đích (ZFL1-ZFL1 hoặc ZFA36-ZFA36) đối với gen thông báo được xác định. Đột biến kiểu DDD được đưa vào mỗi vùng nucleaza nối với ZFL1, và đột biến kiểu RRR được đưa vào mỗi vùng nucleaza nối với ZFA36, và theo kết quả của phân tích gen thông báo, hầu như không có hoạt tính phân cắt nào được quan sát ở một trong mỗi thể đột biến kiểu DDD ZFL1-ND và thể đột biến kiểu RRR ZFA36-ND (Fig.7A và Fig.7B). Từ điều này, đã được làm rõ là ngay cả trong các vùng nucleaza ND1 và ND2 mới, các homodime không thể được tạo ra bằng cách đưa các sự thay thế axit amin vào kiểu DDD/kiểu RRR, như trong trường hợp FokI-ND.

Mỗi vùng nucleaza được tải nạp để tạo thành kiểu heterodime để xác định hoạt tính phân cắt của trình tự đích (ZFL1-ZFA36) đối với gen thông báo. Fig.8A thể hiện các kết quả này. Hoạt tính phân cắt trong FokI-ND được giảm tới khoảng 60% so với của đích. Tuy nhiên, trong ND1, ngay cả trong kiểu heterodime, không có sự giảm hoạt tính phân cắt nào được quan sát, và hoạt tính này là cao bằng khoảng 1,6 lần hoạt tính của đích. Mặt khác, trong ND2, hoạt tính phân cắt hầu như biến mất khi dạng heterodime được sử dụng. Từ các kết quả này, đã được làm rõ là ND1 có khả năng biến đổi chức năng thành kiểu heterodime có hoạt tính phân cắt cao hơn so với FokI-ND.

Ngoài ra, để làm rõ hiệu quả gây đột biến của mỗi nucleaza kiểu heterodime vào trình tự đích của hệ gen, việc phân tích bằng Cel-I được thực

hiện (Fig.8B). Như được thể hiện trong Bảng 1, các trình tự bazơ đích trên hệ gen là hoàn toàn giống nhau ở vị trí thứ nhất. Các hệ gen thu được từ các tế bào CHO-K1 biểu hiện các nucleaza kiểu homodime và kiểu heterodime, và các mảnh PCR chứa các trình tự đích này được xử lý bằng Cel-I. Sau đó, trong kiểu homodime, mỗi vùng nucleaza phân cắt ADN hệ gen đích, giống với các kết quả thu được ở trên. Mặt khác, trong kiểu heterodime, sự gây đột biến được quan sát trong FokI-ND cũng như các kết quả của phân tích gen thông báo SSA, nhưng tỷ lệ gây đột biến là cao hơn trong ND1. Trong ND2, không đột biến có thể phát hiện được nào quan sát được ngay cả trong trình tự hệ gen đích, và đã cho rằng hoạt tính phân cắt bị mất.

Các kết quả này gợi ý rằng, khác với FokI-ND, ND1 có khả năng biến đổi chức năng thành kiểu heterodime mà duy trì hoạt tính cao, và ngoài ra sự biến đổi thành dạng heterodime có thể dẫn đến các thay đổi cấu trúc đáng kể trong ND2, điều này có nghĩa là mỗi đặc tính cấu trúc là khác với FokI-ND.

Ví dụ 8: sự ảnh hưởng đối với hoạt tính phân cắt khi sử dụng tổ hợp các vùng nucleaza kiểu heterodime

Việc phân tích thêm được thực hiện để xác định xem có thể có hoạt tính phân cắt cao hơn so với FokI-ND thông thường bằng cách thay đổi tổ hợp của các vùng nucleaza trong quá trình tạo heterodime hay không.

Thể đột biến kiểu DDD ZFL1-ND và thể đột biến kiểu RRR ZFA36-ND thu được trong ví dụ 7 được sử dụng trong các tổ hợp khác nhau, và hoạt tính nucleaza được xác định bằng thử nghiệm gen thông báo SSA theo cách giống như được mô tả trong ví dụ 2. Fig.9 thể hiện các kết quả này.

Khi việc phân tích được thực hiện với đột biến kiểu DDD FokI-ND và

mỗi vùng nucleaza được đưa vào với đột biến kiểu RRR, không có sự tăng nào của hoạt tính phân cắt của gen thông báo được quan sát ngay cả khi được kết hợp với ND1 có hoạt tính phân cắt cao cho đến nay, và hoạt tính này là khoảng 60% hoạt tính phân cắt của FokI-ND kiểu homodime. Ngoài ra, hầu như không có hoạt tính phân cắt nào được quan sát khi kết hợp với ND2 (Fig.9A).

Thể đột biến kiểu DDD ND1 có hoạt tính phân cắt khoảng 80% khi kết hợp với thể đột biến kiểu RRR FokI-ND, và trong khi đó, khi kết hợp với thể đột biến kiểu RRR ND2, hoạt tính phân cắt hầu như bị mất giống như trong trường hợp thể đột biến kiểu DDD FokI-ND (Fig.9B).

Khác với các kết quả của thể đột biến kiểu DDD FokI-ND và thể đột biến kiểu DDD ND1, thể đột biến kiểu DDD ND2 có hoạt tính phân cắt cao chỉ khi kết hợp với thể đột biến kiểu RRR ND1, và hầu như mất hoạt tính phân cắt của nó khi kết hợp với thể đột biến kiểu RRR FokI-ND hoặc thể đột biến kiểu RRR ND2 (Fig.9C). Kết quả là, đã được làm rõ là trong trường hợp dạng heterodime của thể đột biến kiểu RRR ND2, hoạt tính phân cắt bị mất bất kể tổ hợp của các vùng nucleaza.

Đã phát hiện được rằng tổ hợp của các vùng nucleaza kiểu heterodime có hoạt tính phân cắt là có thể có, và trong thay thế axit amin được phân tích lúc này, có thể làm tăng hoạt tính phân cắt chỉ trong tổ hợp của thể đột biến kiểu DDD ND1 và thể đột biến kiểu RRR ND1. Ngoài ra, đã cho thấy rằng tổ hợp của thể đột biến kiểu DDD ND2 và thể đột biến kiểu RRR ND1 cũng có hoạt tính phân cắt giống như homodime của FokI-ND. Từ các kết quả nêu trên, đã phát hiện được rằng khi ND1 và ND2 tạo thành đime, các đặc tính cấu trúc khác với các đặc tính của FokI-ND xảy ra.

Ví dụ 9: phân tích ở cấp độ trình tự bazơ đối với việc gây đột biến bằng ZFN

Việc nghiên cứu được thực hiện về sự thay đổi nào sẽ xảy ra trong trình tự bazơ do sự gây đột biến vào trình tự đích bởi ZFN.

(1) Phân tích trình tự

Để làm rõ sự thay đổi nào sẽ được gây ra cho trình tự bazơ bằng cách gây đột biến bởi ZFN vào các trình tự đích (vị trí thứ nhất và vị trí thứ hai) của ZFA36 có trên trình tự hệ gen của các tế bào CHO-K1, việc phân tích trình tự của vùng hệ gen chứa các trình tự đích được thực hiện như sau. Theo cách giống như trong thử nghiệm Cel-I được mô tả trong ví dụ 3, PCR được thực hiện bằng cách sử dụng KOD Fx Neo với ADN hệ gen của CHO-K1 được tải nạp bằng mỗi ZFN làm khuôn. Sản phẩm PCR thu được được tạo dòng TA sau khi thêm A vào sản phẩm PCR bằng TArget clone plus (Toyobo). Sản phẩm PCR được tạo dòng trong *Escherichia coli*, ít nhất 16 thể biến nạp *Escherichia coli* độc lập được chọn ngẫu nhiên, và mỗi plasmit của nó được tách chiết. Các đoạn mồi M13-f được bổ sung để plasmit thu được có nồng độ cuối 6,4 pmol ở 400ng, và Fasmac là cần thiết để thực hiện PCR và phân tích trình tự. Từ trình tự thu được, trình tự thể đột biến của trình tự đích trong mỗi ZFN và tỷ lệ của nó được tính.

(2) Các kết quả

Fig.10 thể hiện các kết quả này. Ở vị trí thứ nhất của ZFA36-ZFA36, FokI-ND thu được 1 dòng có đoạn chèn 1 bazơ và 3 dòng có đoạn chèn 4 bazơ trong số 17 dòng được phân tích. Trong số 14 dòng có thể được phân tích trong ND1, có 1 dòng đối với mỗi đoạn chèn 1 bazơ, đoạn xóa 2 bazơ, và đoạn xóa 8 bazơ. Ngoài ra, 2 dòng có đột biến đã thu được với đoạn chèn 2 bazơ và đoạn

xóa 1 bazơ, và 6 dòng có đột biến đã thu được với đoạn chèn 4 bazơ mà là đột biến có số lượng lớn nhất trong phân tích này. Trong số 13 dòng có thể được phân tích trong ND2, có 1 dòng thu được đối với mỗi đoạn chèn 1 bazơ, đoạn chèn 2 bazơ, đoạn xóa 2 bazơ, và đoạn xóa 4 bazơ. Trong 15 dòng của ND3 và 16 dòng của ND4 có thể được phân tích, không có sự chèn hoặc xóa bazơ nào được quan sát trong phân tích vị trí thứ nhất ZFA36-ZFA36 lúc này. Ngoài ra, sự gây đột biến của mỗi vùng nucleaza ở vị trí thứ hai của trình tự đích ZFA36 được phân tích. Sau đó, trong FokI, trong số 20 dòng được phân tích, có 1 dòng thu được đối với mỗi đoạn chèn 2 bazơ, đoạn chèn 5 bazơ, và đoạn xóa 2 bazơ, và 1 dòng đã thu được với đoạn chèn vượt quá 230 bazơ. Ngoài ra, giống như với vị trí thứ nhất ZFA36, phần lớn đột biến là đoạn chèn 4 bazơ, và 3 dòng đã thu được. Trong ND1, trong số 17 dòng được phân tích, có 2 dòng thu được đối với mỗi trong số đoạn xóa 2 bazơ và đoạn xóa 4 bazơ. Ngoài ra, 3 dòng đã thu được với đoạn chèn 4 bazơ, và 6 dòng đã thu được với đoạn chèn 2 bazơ mà là đột biến thường xuyên nhất trong phân tích này. Trong số 16 dòng được phân tích trong ND2, 1 dòng đã thu được với đoạn chèn 1 bazơ, 2 dòng đã thu được với đoạn chèn 2 bazơ, và 5 dòng đã thu được với đoạn chèn 4 bazơ mà là đột biến thường xuyên nhất trong phân tích này. Ngoài ra, 1 dòng đã thu được với đoạn chèn lớn tới 71 bazơ. Trong ND3, khác với vị trí thứ nhất, đột biến có mặt trong trình tự đích ở vị trí thứ hai, và trong 16 dòng được phân tích, có 1 dòng thu được đối với mỗi đoạn chèn 2 bazơ, đoạn xóa 2 bazơ, và đoạn xóa 8 bazơ. Trong ND4, trong tất cả 14 dòng có thể được phân tích, không đột biến nào được quan sát trong trình tự đích ở vị trí thứ hai cũng như ở vị trí thứ nhất.

Theo kết quả của phân tích trình tự, đối với việc gây đột biến vào trình tự đích bằng ND1 và ND2, nhiều dòng được quan sát có đoạn chèn 4 bazơ giống với FokI-ND, nhưng các kiểu khác nhau không được thấy trong FokI-ND đã được thấy là đột biến của hệ gen. Phần nhô ra của bazơ tạo bởi các vùng nucleaza mới có thể khác với vùng nucleaza FokI. Bảng 3 thể hiện tỷ lệ đột biến của ZFA36 thu được bằng cách phân tích trình tự ở mỗi vị trí đích. Trong các điều kiện phân tích này, đã thấy rõ rằng, cũng trong phân tích trình tự, tỷ lệ đột biến của mỗi vùng nucleaza trong trình tự hệ gen đích là giống với tỷ lệ trong thử nghiệm Cel-I.

Bảng 3

Tỷ lệ đột biến chỉnh sửa hệ gen trong phân tích trình tự		
ZFA36-ND	ZFA36-ZFA36 vị trí thứ nhất	ZFA36-ZFA36 vị trí thứ hai
FokI-ND	4/17 (23,5%)	7/20 (35,0%)
ND1	11/14 (78,6%)	13/17 (76,5%)
ND2	4/13 (30,8%)	9/16 (56,3%)
ND3	0/15 (0,0%)	3/16 (18,8%)
ND4	0/16 (0,0%)	0/14 (0,0%)

Ví dụ 10: Hoạt tính phân cắt của các vùng nucleaza mới được kết hợp với vùng liên kết axit nucleic khác với ZFA

Để làm rõ xem các vùng nucleaza mới theo sáng chế có hoạt tính phân cắt cao hơn so với FokI-ND ngay cả khi sử dụng vùng liên kết axit nucleic khác với ZFA hay không, enzym phân cắt axit nucleic nhân tạo chứa ND1 và TALE được tạo ra, và hoạt tính phân cắt của nó được xác định. Cụ thể, TALE-FokI-ND và TALE-ND1 thu được bằng cách sử dụng công nghệ platin TALEN

(doi: 10,1038/srep03379) được phát triển trong phòng thí nghiệm của các tác giả sáng chế, và hoạt tính phân cắt của chúng được đánh giá bằng phương pháp thử nghiệm gen thông báo SSA.

(1) Thiết kế vector đích của TALE-ND1

Vector đích của platin TALEN ptCMV-153/47-VR-NG (<https://www.addgene.org/50704/>) không được chèn môđun liên kết axit nucleic được sử dụng làm khuôn, và cặp mỗi thay thế vùng ptTALE-nucleaza (để khuếch đại vector) được sử dụng để khuếch đại bằng PrimeStar Max, để bằng cách đó thu được các mảnh vector. Ngoài ra, cặp mỗi thay thế vùng ptTALE-nucleaza (để khuếch đại đoạn chèn) được sử dụng để thu được các đoạn chèn của ND1. Các mảnh vector và các đoạn chèn được trộn lẫn để thực hiện phản ứng In-Fusion. Sau khi biến nạp *Escherichia coli*, plasmit được tách chiết từ thể biến nạp thu được để thu được vector đích mong muốn TALE-ND1 (vector trong đó FokI-ND trong ptCMV-153/47-VR-NG được thay bằng ND1).

(2) Đưa môđun liên kết axit nucleic vào và thiết kế plasmit gen thông báo SSA

Môđun liên kết axit nucleic để nhận biết các trình tự của locut ROSA26 và locut HPRT1 của các tế bào CHO-K1 được đưa vào vector mong muốn thu được trong mục (1) nêu trên bằng phương pháp cổng Golden (dựa trên phương pháp được mô tả trong doi: 10,1038/srep03379). Phương pháp này được sử dụng để biến nạp *Escherichia coli*, và các plasmit mong muốn (TALE-ROSA26-ND1 và TALE-HPRT1-ND1) đã thu được từ thể biến nạp thu được. Ngoài ra, để TALEN (TALE-FokI) giữ được FokI-ND thông thường, các cấu trúc (TALE-ROSA26-FokI và TALE-HPRT1-FokI) để đưa môđun liên kết axit nucleic vào đã thu được theo cách giống hệt. Đối với các plasmit gen thông

báo SSA, các plasmit chứa các trình tự đích tương ứng (gen thông báo SSA-CHO-ROSA26 và gen thông báo SSA-CHO-HPRT1) được tạo ra bằng phương pháp thông thường. Bảng 4 dưới đây thể hiện các trình tự đích của locut ROSA26 và locut HPRT1.

Bảng 4

Các trình tự đích ở locut ROSA26 và locut HPRT1 của các tế bào CHO-K1
ROSA26: 5'-<u>TGCCCAGAAGACTCCCG</u>gcccatctcccagaaaGCCTCGACTTGCAGATCA-3' 3'-ACGGGTCTTCTGAGGGCgggtagaggggtcttt<u>CGGAGCTGAACGTCTAGT</u>-5'
HPRT1: 5'-<u>TGAACCAGGCTATGACCT</u>agattattttgtatTCCTAATCACTATGTCGA-3' 3'-ACTTGGTCCGATACTGGAtctaaataaaacata<u>AGGATTAGTGATACAGCT</u>-5'
Các phần được gạch chân có nghĩa là các trình tự nhận biết ADN của TALEN (bên trái và bên phải) ở mỗi locut, và các chữ cái thường có nghĩa là trình tự đệm giữa các trình tự nhận biết ADN.

(3) Đưa vào tế bào và thử nghiệm gen thông báo

Bằng cách sử dụng các plasmit TALEN (TALE-FokI-ND và TALE-ND1) để đưa môđun liên kết axit nucleic vào để nhận biết trình tự đích, và các plasmit gen thông báo SSA để đưa trình tự đích thu được trong mục (2) nêu trên vào, việc tải nạp được thực hiện đối với các tế bào HEK293T bằng cách sử dụng Lipofectamine LTX, và bằng cách sử dụng các tế bào này 24 giờ sau khi tải nạp, thử nghiệm gen thông báo SSA được thực hiện theo cách giống như trong ví dụ 2. Để làm đối chứng, plasmit ZFN chứa FokI-ND (pSTL-ZFA36) và plasmit pSTL chỉ biểu hiện FokI-ND được sử dụng để thực hiện

việc xác định theo cách tương tự. Hoạt tính phân cắt của mỗi TALE-ND được tính là hoạt tính tương đối khi giá trị của hoạt tính phân cắt của ZFA36-FokI đối với gen thông báo chứa trình tự đích của ZFA36 được đặt bằng 1.

(4) Các kết quả

Fig.11A và Fig.11B thể hiện các kết quả này. Fig.11A thể hiện các kết quả khi trình tự có nguồn gốc từ locut ROSA26 là đích, và Fig.11B thể hiện các kết quả khi trình tự có nguồn gốc từ locut HPRT1 là đích.

Như đã rõ từ Fig.11A, trong TALE-ROSA26-FokI trong đó vùng nucleaza của platin TALEN là FokI-ND thông thường, đã cho thấy rằng hoạt tính phân cắt của gen thông báo hướng đích trình tự có nguồn gốc từ locut ROSA26 có hoạt tính gần với hoạt tính phân cắt của ZFA36-FokI hướng đích gen thông báo ZFA36. Mặt khác, trong TALE-ROSA26-ND1 trong đó vùng nucleaza của platin TALEN được thay bằng ND1, đã cho thấy rằng hoạt tính phân cắt của gen thông báo hướng đích trình tự có nguồn gốc từ locut ROSA26 bằng khoảng hai lần hoạt tính của TALE-ROSA26-FokI. Dựa vào điều trên, khi hướng đích trình tự có nguồn gốc locut ROSA26 của hệ gen nội sinh của các tế bào CHO-K1, đã thấy rằng TALE-ND1 có hoạt tính phân cắt trong vùng nucleaza cao hơn so với FokI-ND thông thường là TALE-FokI.

Ngoài ra, trình tự có nguồn gốc từ locut HPRT1, là locut khác với locut ROSA26, được hướng đích và được phân tích theo cách giống hệt. Kết quả là, như đã rõ từ Fig.11B, trong TALE-HPRT1-FokI trong đó vùng nucleaza của platin TALEN là FokI-ND thông thường, đã cho thấy rằng hoạt tính phân cắt của gen thông báo hướng đích trình tự có nguồn gốc locut HPRT1 bằng khoảng một nửa hoạt tính phân cắt của ZFA36-FokI hướng đích gen thông báo ZFA36.

Mặt khác, trong TALE-HPRT1-ND1 trong đó vùng nucleaza của platin TALEN được thay bằng ND1, hoạt tính phân cắt của gen thông báo hướng đích trình tự có nguồn gốc locut HPRT là bằng khoảng hai lần hoạt tính của TALE-FokI, và kết quả giống như của trình tự có nguồn gốc locut ROSA26 đã thu được.

Từ các kết quả này, đã được làm rõ là TALE-ND1 có ND1 làm vùng nucleaza của platin TALEN có hoạt tính phân cắt cao hơn so với TALE-FokI có FokI-ND thông thường. Từ điều trên, đã gợi ý rằng các vùng nucleaza mới theo sáng chế là các vùng nucleaza có hoạt tính phân cắt tốt hơn so với FokI-ND ngay cả trong trường hợp sử dụng không chỉ ZFN mà còn các vùng liên kết axit nucleic khác.

Từ điều trên, đã chứng tỏ rằng các vùng nucleaza mới theo sáng chế có các đặc trưng chính sau đây.

(i) ND1 và ND2 theo sáng chế, đặc biệt là ND1, có hoạt tính phân cắt cao hơn đáng kể so với FokI-ND. Ngoài ra, do ND1 có hoạt tính phân cắt cao hơn FokI-ND ngay cả trong mảng ZF nhận biết các trình tự khác và trong TALE, tác dụng làm tăng hoạt tính phân cắt bởi ND1 là phổ biến.

(ii) Bằng cách cải biến trình tự nối giữa mảng ZF và ND, ND1 có hoạt tính phân cắt gần giống như trình tự đệm 6bp ngay cả với trình tự đệm 5bp hoặc 7bp, và do đó ND1 có tính linh hoạt cao hơn FokI-ND trong quá trình dimer hóa và có thể làm giảm hạn chế của trình tự đích.

(iii) Bằng cách đưa đột biến kiểu DDD và đột biến kiểu RRR vào ND1 và ND2, sự homodimer hóa có thể được ngăn chặn giống như trong FokI-ND. Ngoài ra, bằng cách kết hợp ZFL1-DDD và ZFA36-RRR, việc phân cắt bằng heterodimer là có thể thực hiện trong ND1. Hơn nữa, do hoạt tính phân cắt ở

thời điểm đó là gần giống như của ND1 kiểu dại, có thể thu được vùng nucleaza có cả tính đặc hiệu cao và hoạt tính phân cắt cao bằng cách sử dụng ND1-DDD/RRR.

(iv) Khi hoạt tính này được kiểm tra bằng cách kết hợp FokI-ND kiểu DDD/RRR, ND1, và ND2 với nhau, đã nhận thấy rằng cặp ND1-DDD và ND1-RRR có hoạt tính cao nhất trong tất cả các tổ hợp, chứng tỏ tính vượt trội của hệ thống heterodime ND1-DDD/ND1-RRR.

Ví dụ 11: điều chế protein trong Escherichia Coli

(1) Thiết kế các vector biểu hiện của ZF-ND1 và ZF-FokI

Vị trí XhoI và vị trí SalI của plasmit pET-MCS chứa đuôi His ở đầu N được phân cắt, và đoạn ZF-ND1 được chèn vào. Đoạn ZF-ND1 chứa tín hiệu định vị nhân NLS, ngón tay kẽm (Zinc-Finger, sau đây được viết tắt là ZF), và vùng nucleaza ND1. Để làm ZF, vùng nhận biết hai loại trình tự được sử dụng. Vùng thứ nhất là ZFA36 mà nhận biết trình tự ADN có 12 cặp bazơ GAAGATGGT. Vùng còn lại là ZFL1 mà nhận biết trình tự ADN có 12 cặp bazơ GAAGGTGAC. Do đó, để làm các plasmit biểu hiện ZF (ZFA36A)-ND1 và ZF (ZFL1)-ND1, pET-ZF (ZFA36A) ND1 và pET-ZF (ZFL1) ND1 lần lượt được tạo ra.

Dựa trên các plasmit nêu trên, các plasmit trong đó vùng nucleaza của enzym giới hạn FokI (sau đây được viết tắt là FokI) được chèn vào thay cho ND1, pET-ZF (ZFA36A)-FokI, và pET-ZF (ZFL1)-FokI được tạo ra. Các trình tự axit amin của ZF (ZFA36A)-ND1, ZF (ZFL1)-ND1, ZF (ZFA36A)-FokI, và ZF (ZFL1)-FokI được nêu trong SEQ ID NO: 119 đến 122.

(2) Biểu hiện và tinh chế protein ZF-ND1 trong Escherichia Coli

Plasmid pET biểu hiện ZF-ND1 hoặc ZF-FokI, và pRARE2 (Stratagene) được biến nạp vào chủng *Escherichia coli* BL21 (DE3) và được nuôi cấy trong môi trường LB chứa kanamycin và chloramphenicol. *Escherichia coli* được biến nạp được nuôi cấy ở 37°C cho đến khi OD (năng suất hấp thụ ở 600 nm) trở thành 0,6, và sau đó được nuôi cấy ở 18°C trong 1 giờ. Isopropyl β -D-1-thiogalactopyranosit (IPTG) được bổ sung vào tới nồng độ cuối 0,1mM để cảm ứng sự biểu hiện protein. Sau khi lắc thêm và nuôi cấy ở 18°C trong 17 giờ, *Escherichia coli* được phân giải bằng cách sử dụng dung dịch đệm phân giải (20mM Tris-HCl, 500mM NaCl, 10% glyxerol, 10mM imidazol, 1mM benzyl sulfonyl florua, 1mM dithiothreitol, pH=8,0). Sau đó, protein được hấp thụ trên cột niken NTA, và dung dịch đệm rửa (20mM Tris-HCl, 500mM NaCl, 10% glyxerol, 20mM imidazol, pH=8,0) được sử dụng để loại bỏ các tạp chất. Quá trình rửa giải protein được thực hiện bằng cách sử dụng dung dịch đệm rửa giải (20mM Tris-HCl, 500mM NaCl, 10% glyxerol, 500mM imidazol, pH=8,0). Một phần của protein được rửa giải này được hòa tan trong dung dịch đệm của mẫu SDS để phân tích hiệu suất protein. Protein được rửa giải được tinh chế bằng cách sử dụng sắc ký lọc gel HiPrep 16/60 Sephacryl S-200 HR (GE healthcare, United States, Illinois) và dung dịch đệm A (20mM HEPES, 150mM NaCl, 10% glyxerol, pH =7,4), làm đông lạnh ngay bằng nitơ lỏng, và sau đó được bảo quản ở -80°C. Để phân tích hiệu suất protein, phân đoạn imidazol-protein được rửa giải và protein tinh chế lần cuối được hòa tan trong dung dịch đệm của mẫu SDS và điện di bằng SDS-PAGE. Trong quá trình điện di, nồng độ đã biết của protein BSA được điện di trong một làn khác, gel sau khi PAGE được nhuộm bằng Coomassie, băng của trọng lượng phân tử của

protein đích được định lượng bằng ChemiDoc XRS+, và hiệu suất protein được phân tích từ trọng lượng phân tử protein.

(3) Các kết quả

Khi biểu hiện và tinh chế các protein ZFN trong *Escherichia coli*, có vấn đề là hiệu suất protein là thấp. Đã cho rằng các đặc tính cơ bản của các protein ZFN là chúng có độ tan thấp dẫn đến sự kết tập. Đã kiểm tra protein của ND1 mới tan như thế nào so với FokI theo lĩnh vực thông thường.

Bảng 5

	Hiệu suất protein cho 1l dịch nuôi cấy <i>Escherichia Coli</i> (nmol/1l dịch nuôi cấy)	Mức tăng (lần) của hiệu suất protein của ND1 khi được so sánh với FokI
Protein ZF(ZFA36A)-ND1	104,8	11
Protein ZF(ZFA36A)-FokI	9,3	(1)
Protein ZF(ZFL1)-ND1	67,6	13
Protein ZF(ZFL1)-FokI	5,0	(1)

Như được thể hiện trong Bảng 5, khi ZF-ND1 được biểu hiện trong *Escherichia coli*, nó thường dễ tan hơn ZF-FokI, và hiệu suất cho 1l dịch nuôi cấy *Escherichia coli* được cải thiện. Độ tan càng cao, nồng độ protein có thể được đưa vào càng cao, do đó là đặc biệt thuận lợi cho việc chỉnh sửa hệ gen bằng cách sử dụng protein.

Ví dụ 12: Hoạt tính của ZF-ND1 và ZF-FokI trong tế bào thực vật

(1) Thiết kế các tế bào gen thông báo

Để đánh giá hoạt tính của ZF-ND1 hoặc ZF-FokI, các tế bào gen thông

báo được thiết kế để biểu hiện gen GUS (β -glucuronidaza) có độ dài đầy đủ 2,1 kb khi gắn sợi đơn (single-strand annealing, SSA) sau khi sự phân cắt ADN được gây ra. Cụ thể, 1,8 kb từ phía ngược dòng của gen GUS, và 0,8 kb từ phía xuôi dòng của gen GUS, trình tự nhận biết ZFA36 (GAAGCTGGT) được đưa vào pCAMBIA 1305.2 (Marker Gene). Đối với 1,8 kb từ phía ngược dòng của gen GUS và 0,8 kb từ phía xuôi dòng của gen GUS, chúng có 500 bp là trình tự chung (vùng chồng nhau). Vector này được biến nạp thông qua *Agrobacterium* vào dòng tế bào được nuôi cấy *Arabidopsis thaliana* T87 (RIKEN), và dòng tế bào được thiết lập này được ký hiệu là T87-GUUS (ZFA36). Do sự biểu hiện của GUS có thể được quan sát và phát hiện dễ dàng bằng cách sử dụng chất phản ứng nhuộm xanh da trời X-Gluc, việc phát hiện protein GUS cho phép đánh giá việc đưa ZF-ND1 hoặc ZF-FokI vào nhân và hoạt tính cảm ứng SSA.

(2) Đưa vào bằng phương pháp tạo lỗ bằng điện

Protein thu được trong ví dụ 11 được đưa vào bằng phương pháp tạo lỗ bằng điện. Tế bào T87-GUUS (ZFA36) đã được thiết lập được sử dụng để thử đưa ZF-ND1 và ZF-FokI bằng cách tạo lỗ bằng điện vào các tế bào T87 có thành tế bào. NEPA21 Typ II là sản phẩm của Neppa Gene được sử dụng để tạo lỗ bằng điện. Việc tạo lỗ bằng điện được thực hiện bằng cách sử dụng dung dịch đệm Opti-MEMI, và hoạt tính GUS được đánh giá hai ngày sau. Kết quả là, như được thể hiện trên Fig.12, khi quá trình tạo lỗ bằng điện được thực hiện bằng cách sử dụng Opti-MEMI, một lượng rất lớn tế bào biểu hiện GUS. Trong thử nghiệm bằng cách sử dụng protein ZF-ND1, đã phát hiện được rằng hoạt tính cảm ứng SSA là cao bằng khoảng hai lần hoạt tính trong trường hợp

sử dụng protein ZF-FokI. Trong thử nghiệm này, đã phát hiện được rằng, khi được đưa vào không chỉ dưới dạng vector biểu hiện gen mà cả dưới dạng protein, ZF-ND1 có hoạt tính cao hơn so với của ZF-FokI, và rằng ZF-ND1 có hoạt tính cao hơn so với ZF-FokI không chỉ trong các tế bào động vật mà cả trong tế bào thực vật.

(3) Đưa vào bằng phương pháp tải nạp protein

Protein thu được trong ví dụ 11 được sử dụng để thử đưa vào bằng phương pháp tải nạp protein. Tức là, T87-GUUS (ZFA36) có thành tế bào được bổ sung protein ZF (ZFA36A)-ND1 hoặc protein ZF (ZFA36A)-FokI có nồng độ cuối 0,1 đến 2 μ M, và sau đó, sau 1 giờ, hỗn hợp được rửa bằng môi trường nuôi cấy và được nuôi cấy trong 3 ngày. Khi protein này được bổ sung vào, không quá trình xử lý sinh hóa nào như xử lý bằng proteaza hoặc xử lý lý hóa như tạo lỗ bằng điện được thực hiện. Do đó, lượng hấp thụ protein tự phát được mong đợi. Kết quả là, như được thể hiện trên Fig.13, mặc dù đây là tỷ lệ nhỏ so với phương pháp tạo lỗ bằng điện, đã cho thấy rằng protein ZF (ZFA36A)-ND1 và protein ZF (ZFA36A)-FokI được đưa vào bằng phương pháp tải nạp protein, cho phép chỉnh sửa hệ gen. Khi phương pháp tải nạp protein được sử dụng, hoạt tính của ZF (ZFA36A)-ND1 là bằng hoặc cao hơn so với hoạt tính của ZF (ZFA36A)-FokI.

Khả năng áp dụng công nghiệp

Các vùng nucleaza mới theo sáng chế có các đặc tính tuyệt vời khác với vùng nucleaza của FokI, và enzym phân cắt axit nucleic nhân tạo chứa các vùng nucleaza theo sáng chế là rất hữu ích để làm công cụ chỉnh sửa hệ gen.

Văn bản không chứa danh mục trình tự

SEQ ID NO: 9; đoạn mỗi xuôi thay thế ND1

SEQ ID NO: 10; đoạn mỗi ngược thay thế ND1

SEQ ID NO: 11; đoạn mỗi xuôi thay thế ND2

SEQ ID NO: 12; đoạn mỗi ngược thay thế ND2

SEQ ID NO: 13; đoạn mỗi xuôi thay thế ND3

SEQ ID NO: 14; đoạn mỗi ngược thay thế ND3

SEQ ID NO: 15; đoạn mỗi xuôi thay thế ND4

SEQ ID NO: 16; đoạn mỗi ngược thay thế ND4

SEQ ID NO: 17; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-FokI-GS-F

SEQ ID NO: 18; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

SEQ ID NO: 19; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND1-GS-F

SEQ ID NO: 20; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

SEQ ID NO: 21; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND2-GS-F

SEQ ID NO: 22; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

SEQ ID NO: 23; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-FokI-RPGEKP-F

SEQ ID NO: 24; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

SEQ ID NO: 25; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND1-RPGEKP-F

SEQ ID NO: 26; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

SEQ ID NO: 27; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND2-RPGEKP-F

SEQ ID NO: 28; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

SEQ ID NO: 29; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-F

SEQ ID NO: 30; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

SEQ ID NO: 31; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND1-TGPGAAARA-F

SEQ ID NO: 32; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

SEQ ID NO: 33; đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND2-TGPGAAARA-F

SEQ ID NO: 34; đoạn mỗi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

SEQ ID NO: 35; đoạn mỗi xuôi để đưa đột biến Sharkey vào FokI-ND

SEQ ID NO: 36; đoạn mỗi ngược để đưa đột biến Sharkey vào FokI-ND

SEQ ID NO: 37; đoạn mỗi xuôi để đưa đột biến Sharkey vào ND1

SEQ ID NO: 38; đoạn mỗi ngược để đưa đột biến Sharkey vào ND1

SEQ ID NO: 39; đoạn mỗi xuôi để đưa đột biến Sharkey vào ND2

SEQ ID NO: 40; đoạn mỗi ngược để đưa đột biến Sharkey vào ND2

SEQ ID NO: 41; đoạn mỗi xuôi thay thế ZFA

SEQ ID NO: 42; đoạn mỗi ngược thay thế ZFA

SEQ ID NO: 43; đoạn mỗi xuôi để đưa đột biến FokI-DDD vào

SEQ ID NO: 44; đoạn mỗi ngược để đưa đột biến FokI-DDD vào

- SEQ ID NO: 45; đoạn mỗi xuôi để đưa đột biến ND1-DDD vào
- SEQ ID NO: 46; đoạn mỗi ngược để đưa đột biến ND1-DDD vào
- SEQ ID NO: 47; đoạn mỗi xuôi để đưa đột biến ND2-DDD vào
- SEQ ID NO: 48; đoạn mỗi ngược để đưa đột biến ND2-DDD vào
- SEQ ID NO: 49; đoạn mỗi xuôi để đưa vector đột biến FokI-RRR vào
- SEQ ID NO: 50; đoạn mỗi ngược để đưa đột biến vector FokI-RRR vào
- SEQ ID NO: 51; đoạn mỗi xuôi để đưa đột biến chèn đoạn FokI-RRR vào
- SEQ ID NO: 52; đoạn mỗi ngược để đưa đột biến chèn đoạn FokI-RRR vào
- SEQ ID NO: 53; đoạn mỗi xuôi để đưa đột biến vector ND1-RRR vào
- SEQ ID NO: 54; đoạn mỗi ngược để đưa đột biến vector ND1-RRR vào
- SEQ ID NO: 55; đoạn mỗi xuôi để đưa đột biến chèn đoạn ND1-RRR vào
- SEQ ID NO: 56; đoạn mỗi ngược để đưa đột biến chèn đoạn ND1-RRR vào
- SEQ ID NO: 57; đoạn mỗi xuôi để đưa đột biến vector ND2-RRR vào
- SEQ ID NO: 58; đoạn mỗi ngược để đưa đột biến vector ND2-RRR vào
- SEQ ID NO: 59; đoạn mỗi xuôi để đưa đột biến chèn đoạn ND2-RRR vào
- SEQ ID NO: 60; đoạn mỗi ngược để đưa đột biến chèn đoạn ND2-RRR vào
- SEQ ID NO: 61; đoạn mỗi M13-F
- SEQ ID NO: 62; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36
- SEQ ID NO: 63; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36
- SEQ ID NO: 64; vị trí thứ ba của đích CHO-K1 ZFA36-ZFA36
- SEQ ID NO: 65; vị trí thứ nhất của đích CHO-K1 ZFL1-ZFA36
- SEQ ID NO: 66; vị trí thứ hai của đích CHO-K1 ZFL1-ZFA36

SEQ ID NO: 67; trình tự axit amin của FokI-DDD

SEQ ID NO: 68; trình tự axit amin của FokI-RRR

SEQ ID NO: 69; trình tự axit amin của ND1-DDD

SEQ ID NO: 70; trình tự axit amin của ND1-RRR

SEQ ID NO: 71; trình tự axit amin của ND2-DDD

SEQ ID NO: 72; trình tự axit amin của ND2-RRR

SEQ ID NO: 73; trình tự axit amin của FokI-S418P/K441E

SEQ ID NO: 74; trình tự axit amin của ND1-S418P/K441E

SEQ ID NO: 75; trình tự axit amin của ND2-S441E

SEQ ID NO: 76; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi FokI

SEQ ID NO: 77; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi FokI

SEQ ID NO: 78; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi FokI

SEQ ID NO: 79; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 80; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 81; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 82; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 83; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 84; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND1

SEQ ID NO: 85; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 86; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 87; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 88; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 89; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 90; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND3

SEQ ID NO: 91; vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND4

SEQ ID NO: 92; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi FokI

SEQ ID NO: 93; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi FokI

SEQ ID NO: 94; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi FokI

SEQ ID NO: 95; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi FokI

SEQ ID NO: 96; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi FokI

SEQ ID NO: 97; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi FokI

SEQ ID NO: 98; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi ND1

SEQ ID NO: 99; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt

bởi ND1

SEQ ID NO: 100; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND1

SEQ ID NO: 101; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND1

SEQ ID NO: 102; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND1

SEQ ID NO: 103; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND2

SEQ ID NO: 104; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND2

SEQ ID NO: 105; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân

cắt bởi ND2

SEQ ID NO: 106; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 107; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND2

SEQ ID NO: 108; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND3

SEQ ID NO: 109; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND3

SEQ ID NO: 110; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND3

SEQ ID NO: 111; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND3

SEQ ID NO: 112; vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bởi ND4

SEQ ID NO: 113; đoạn môi xuôi để khuếch đại vector để thay thế ptTALE-ND, ptTALEN-ngược-F

SEQ ID NO: 114; đoạn môi ngược để khuếch đại vector để thay thế ptTALE-ND, NC-ngược-R_2

SEQ ID NO: 115; đoạn môi xuôi để khuếch đại đoạn chèn để thay thế ptTALE-ND, ND1-NC-F_v2

SEQ ID NO: 116; đoạn môi ngược để khuếch đại đoạn chèn để thay thế ptTALE-ND, ND1-ptTALEN-R

SEQ ID NO: 117; trình tự đích trên locus ROSA26 trong tế bào CHO-K1

SEQ ID NO: 118; trình tự đích trên locut HPRT1 trong tế bào CHO-K1

SEQ ID NO: 119; trình tự axit amin của ZF(ZFA36A)-ND1

SEQ ID NO: 120; trình tự axit amin của ZF(ZFL1)-ND1

SEQ ID NO: 121; trình tự axit amin của ZF(ZFA36A)-FokI

SEQ ID NO: 122; trình tự axit amin của ZF(ZFL1)-FokI

YÊU CẦU BẢO HỘ

1. Enzym phân cắt axit nucleic nhân tạo bao gồm:

vùng nucleaza là polypeptit chứa trình tự axit amin được nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO: 3, hoặc polypeptit đột biến của nó; và

vùng liên kết axit nucleic,

trong đó polypeptit đột biến là polypeptit chứa trình tự axit amin có ít nhất 90% trình tự tương đồng với trình tự axit amin nêu ở các vị trí 391 đến 585 của SEQ ID NO: 1 hoặc các vị trí 389 đến 579 của SEQ ID NO:3 và có hoạt tính nucleaza, và

trong đó vùng liên kết axit nucleic chứa ngón tay kẽm, TALE, CRISPR/Cas, hoặc PPR.

2. Enzym phân cắt axit nucleic nhân tạo theo điểm 1, trong đó enzym này còn bao gồm: trình tự nối nằm giữa vùng nucleaza và vùng liên kết axit nucleic.

3. Axit nucleic được phân lập bao gồm: trình tự axit nucleic mã hóa enzym phân cắt axit nucleic nhân tạo theo điểm 1 hoặc 2.

4. Vector bao gồm: axit nucleic theo điểm 3, hoặc sản phẩm phiên mã hoặc sản phẩm dịch mã của nó.

5. Phương pháp cải biến axit nucleic đích, bao gồm: đưa vào tế bào (ngoại trừ tế bào có trong cơ thể người, tế bào mầm người và tế bào phôi người), enzym phân cắt axit nucleic nhân tạo theo điểm 1 hoặc 2, axit nucleic theo điểm 3, hoặc vector theo điểm 4.

6. Phương pháp theo điểm 5, trong đó enzym phân cắt axit nucleic nhân tạo,

axit nucleic, hoặc vectơ là hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, axit nucleic mã hóa hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, hoặc vectơ chứa axit nucleic này, sản phẩm phiên mã của nó, hoặc sản phẩm dịch mã của nó.

7. Kit cải biến axit nucleic đích bao gồm: enzym phân cắt axit nucleic nhân tạo theo điểm 1 hoặc 2, axit nucleic theo điểm 3, hoặc vectơ theo điểm 4.

8. Kit theo điểm 7, trong đó enzym phân cắt axit nucleic nhân tạo, axit nucleic, hoặc vectơ là hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, axit nucleic mã hóa hai hoặc nhiều loại trong số các enzym phân cắt axit nucleic nhân tạo, hoặc vectơ chứa axit nucleic này, sản phẩm phiên mã của nó, hoặc sản phẩm dịch mã của nó.

1/14

Fig.1A

(A) ACCATCTTCnnnnnnnGAAGATGGT
 TGGTAGAAGnnnnnnnCTTCTACCA

└──────────┘ └──────────┘ └──────────┘
 Vị trí nửa đích (L) Trình tự đệm Vị trí nửa đích (R)

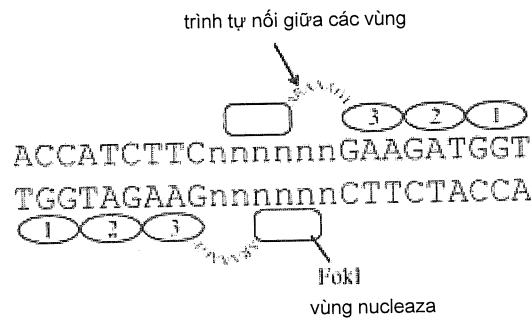


Fig.1B

(B) FokI LVKSELEENKSELPHKLYYPHEYTELIEIARNSTODRILEMKVMEFFMKYYGYRGNHLG
 ND1 LVKGEEXKNSDLRHLKHYVPEYIELIEIADDSKONRLFEPKYVEFLKEVYDYNGRHLG
 ND2 LVKSSIEMSKANMRDNLHLPHDYIELIEISODPYONRIFENKYNOLFINEYGFSSGHLG
 ND3 GIKSNISLLKDELGGIISHISHETLSLIDLAFOSKONRLFENKYLELLYNEYGFRGNHLG
 ND4 ISKTNILELKDKYRDLKYVDHRLALIDLAYDGTANRDFEIoTIDLLINELKFKGNHLG

FokI GSRKPDGAI-YTVGSPIDVGVIVDTKAYSQGYNLPISQADEMORYVEENOTRKHINPNE
 ND1 GSRKPDGAL-YINGLKTDYGIILDTRAYKDCVSLPISQADEMORYVDENNRNAIINPNE
 ND2 GSRKPDGAM-YAHG---FGVIVDTKAYKQGYNLPISQADEMERYVRENIDRNEHYNSNR
 ND3 GSRKPDGIY-YSTTLEDNFGIIVDTKAYSEGYSLPISQADEMERYVRENSNRDEEVNPNK
 ND4 GSRKPDGIISYNIN-----GVVIDKAYSTGYNLPINQADEMIRYIEENOTRDEKINSNK

FokI WVKYYPSSYTEFKFLFVSGFFKGNFYKAOLTRLNHITNCNGAVLSYEELLIGGEMIKAGTL
 ND1 WVKYYPNSILDFFKFLFVSGFFKGDYKOLARYSNLTKRKGAVLSVEOLLGGELKDGSL
 ND2 WVNIFPEDTNEYKFLFVSGFFKGNFEKOLERISIDTCYGGALSVEHLLLGAEYIKRGIL
 ND3 WVENFSEEVKAYYFVFISGSFGKGFEEOLRLSMTTCYNGSAYNVYNLLGAEKIRSGEM
 ND4 WVESFDEKVKDFNYLFVSSFFKGNFKNMLKHIANRTGYNGGAINYENLLYFAELKAGRI

FokI TLEEYRRK-FNNGEINF
 ND1 TLEDVGRK-FNNGEIIIF
 ND2 TLYDFKNS-FLNKEIOF
 ND3 TIEELERANFNNEFTI-
 ND4 SYLDGFKM-YNNGEITL

2/14

Fig.1C

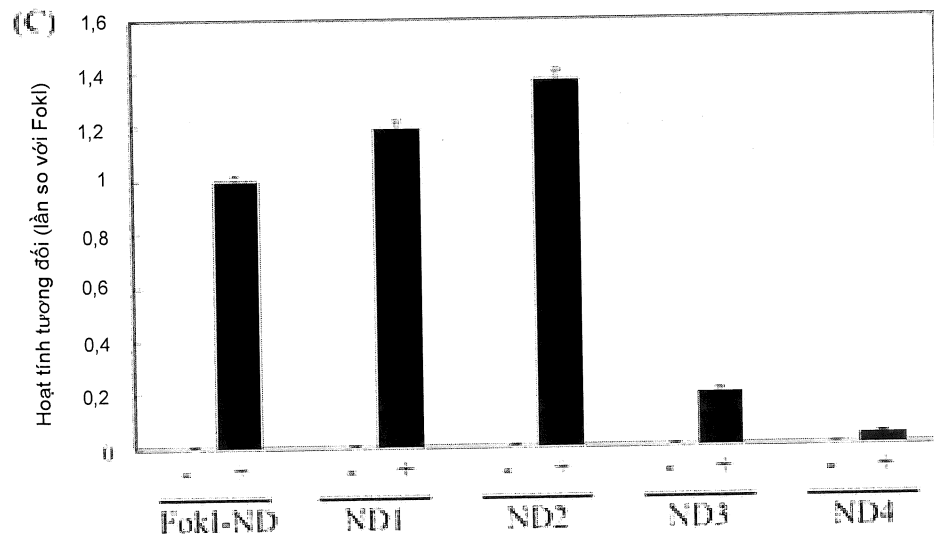
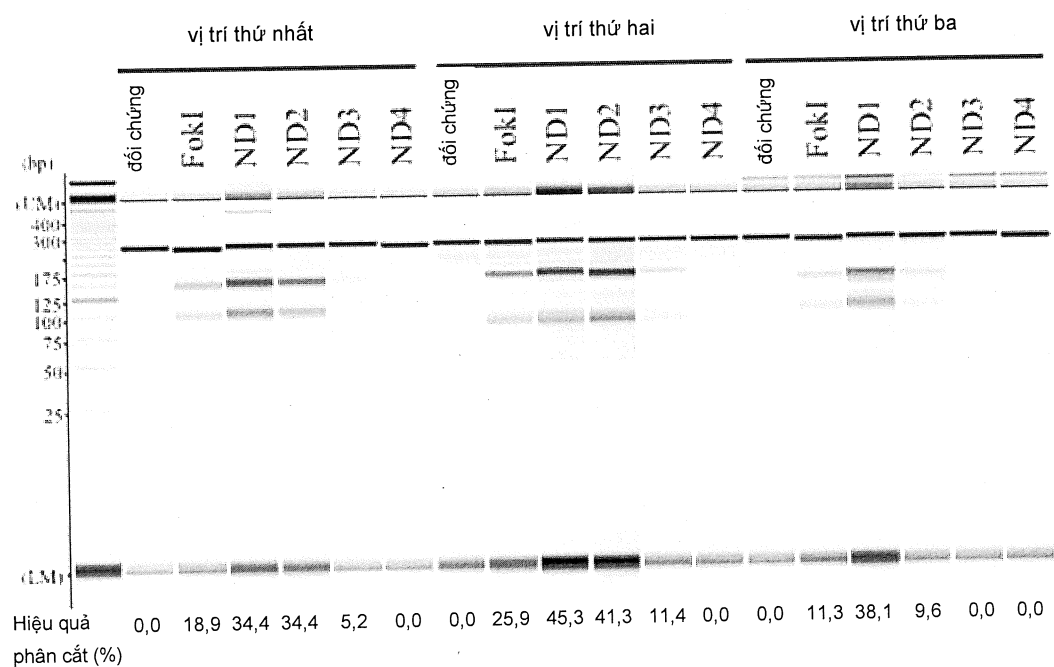


Fig.2A



3/14

Fig.2B

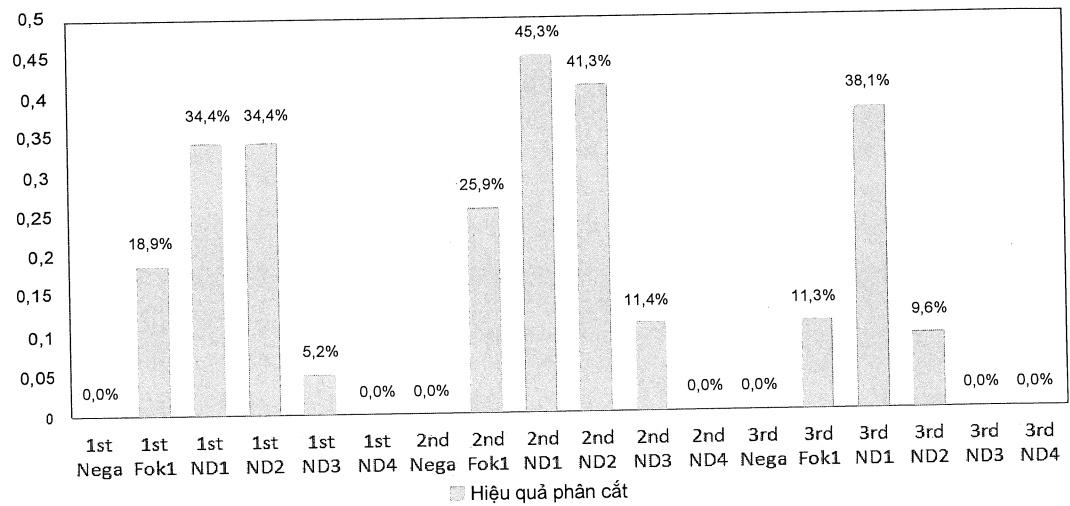
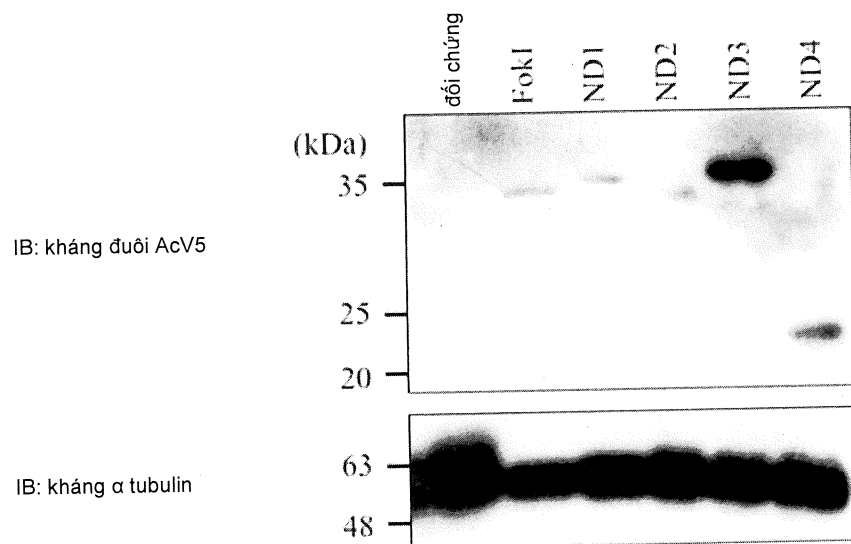
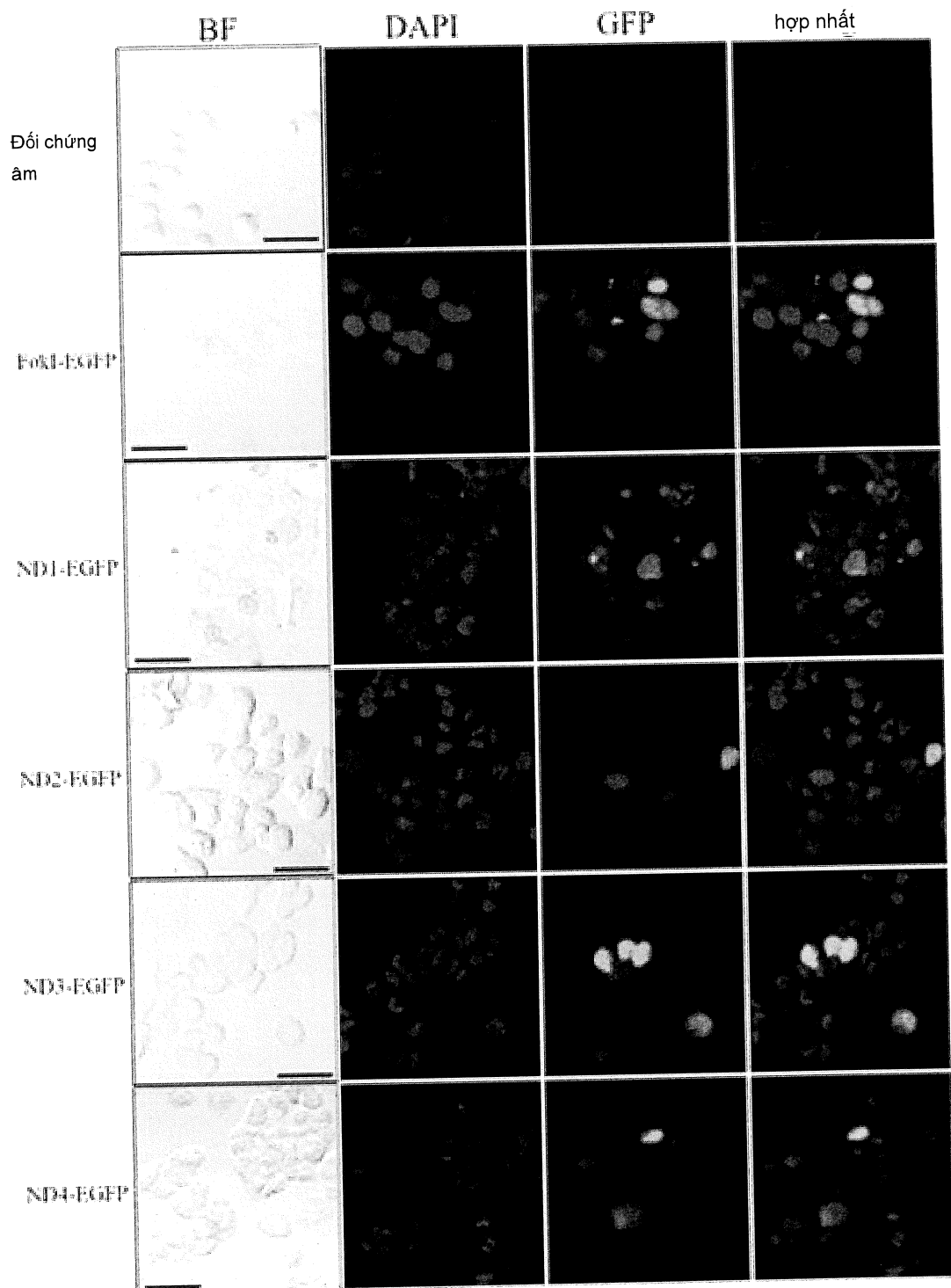


Fig.3



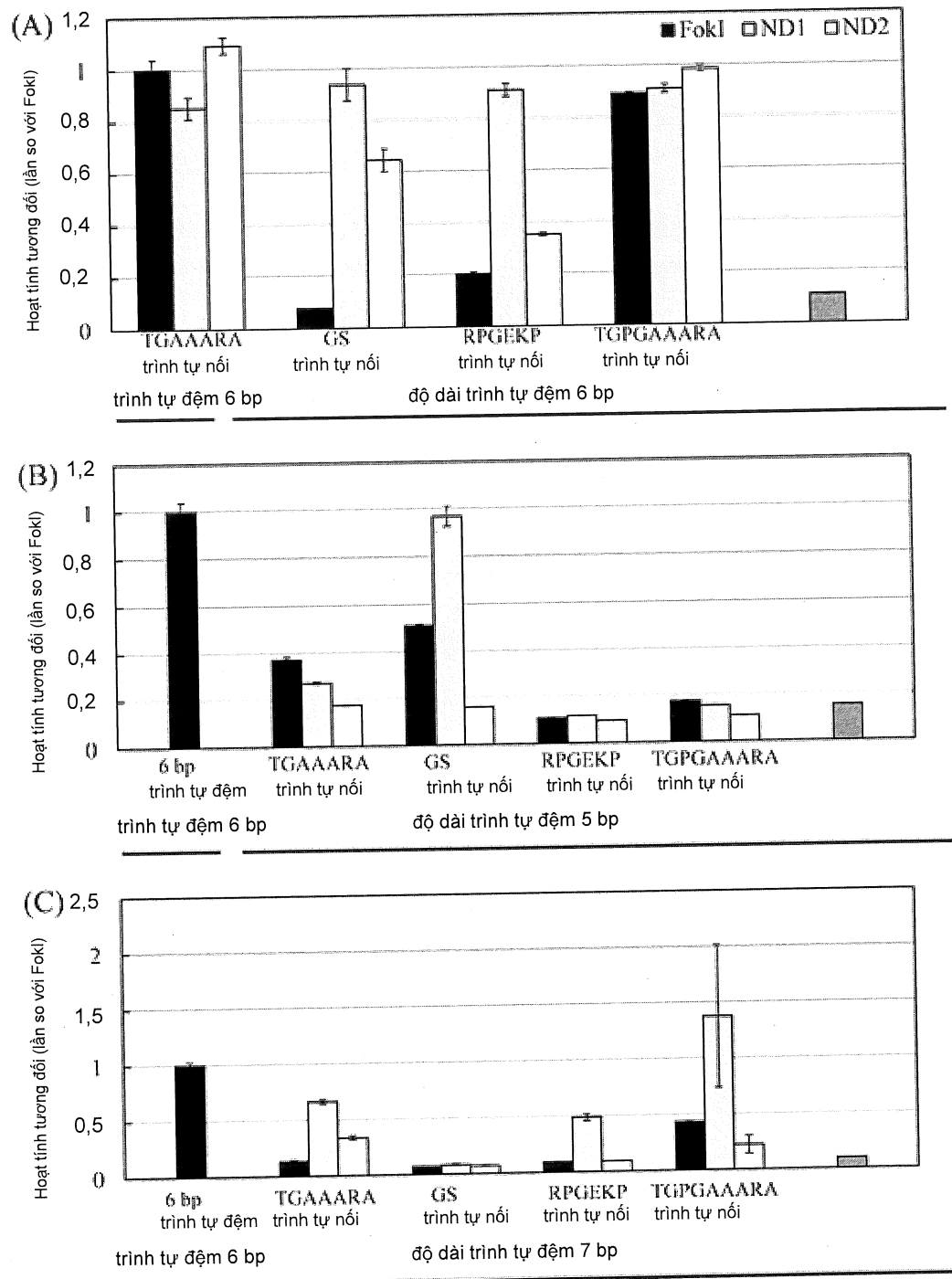
4/14

Fig.4



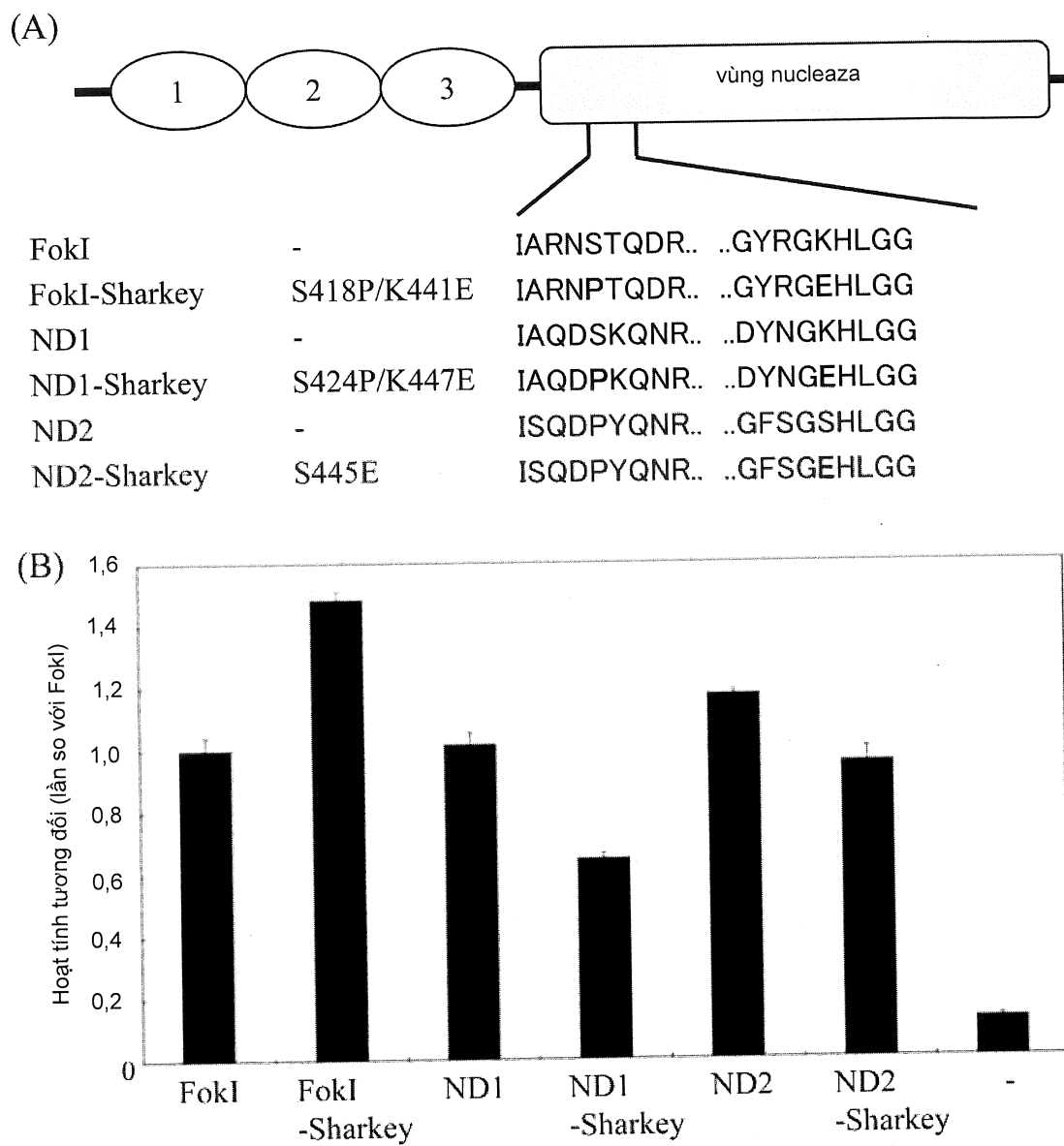
5/14

Fig.5



6/14

Fig.6



7/14

Fig.7

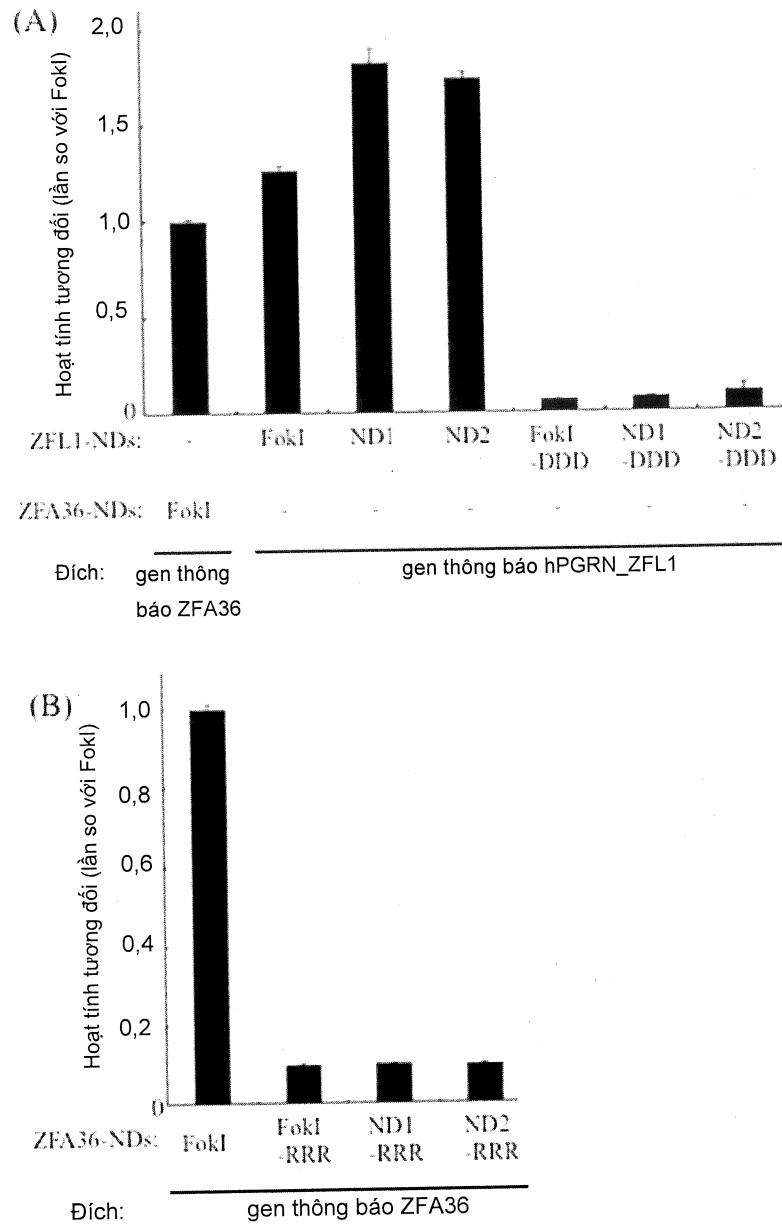


Fig.8A

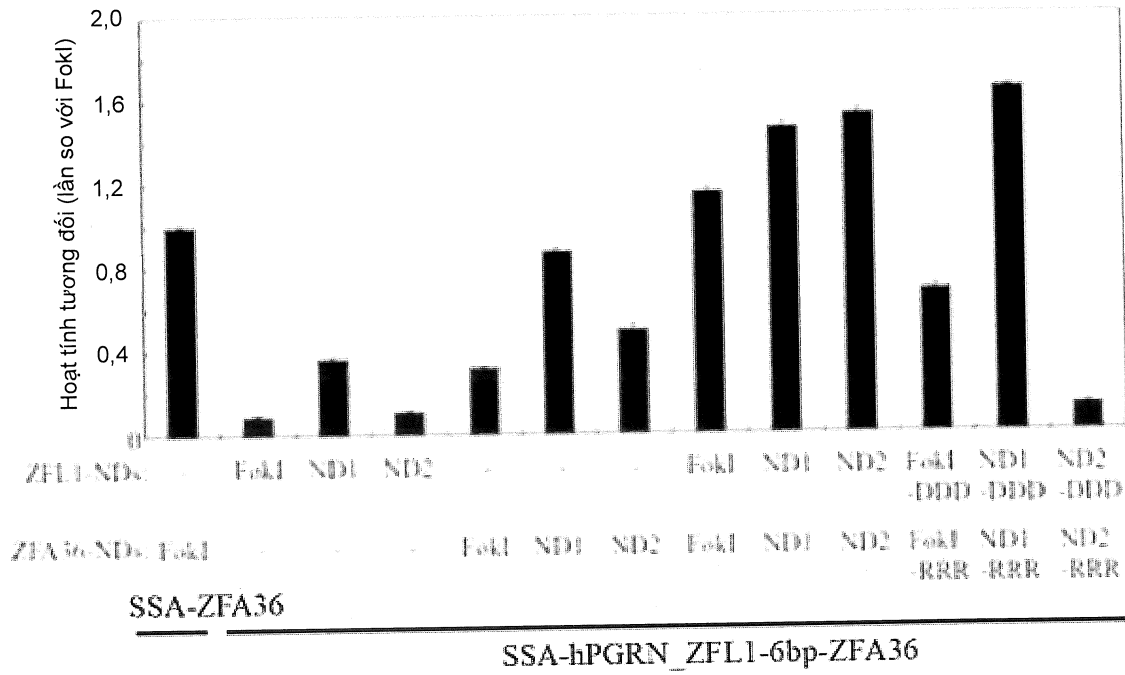
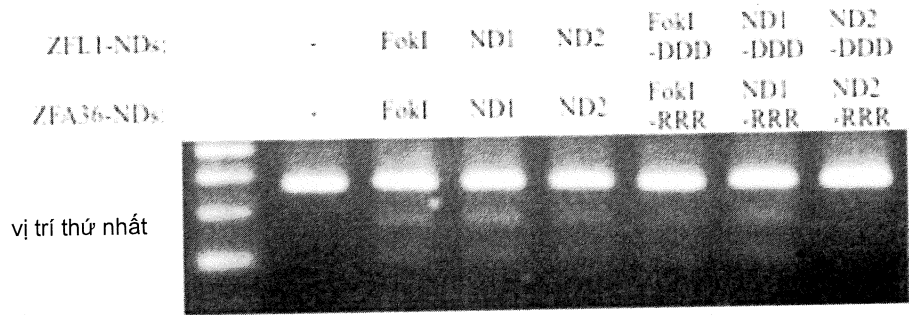
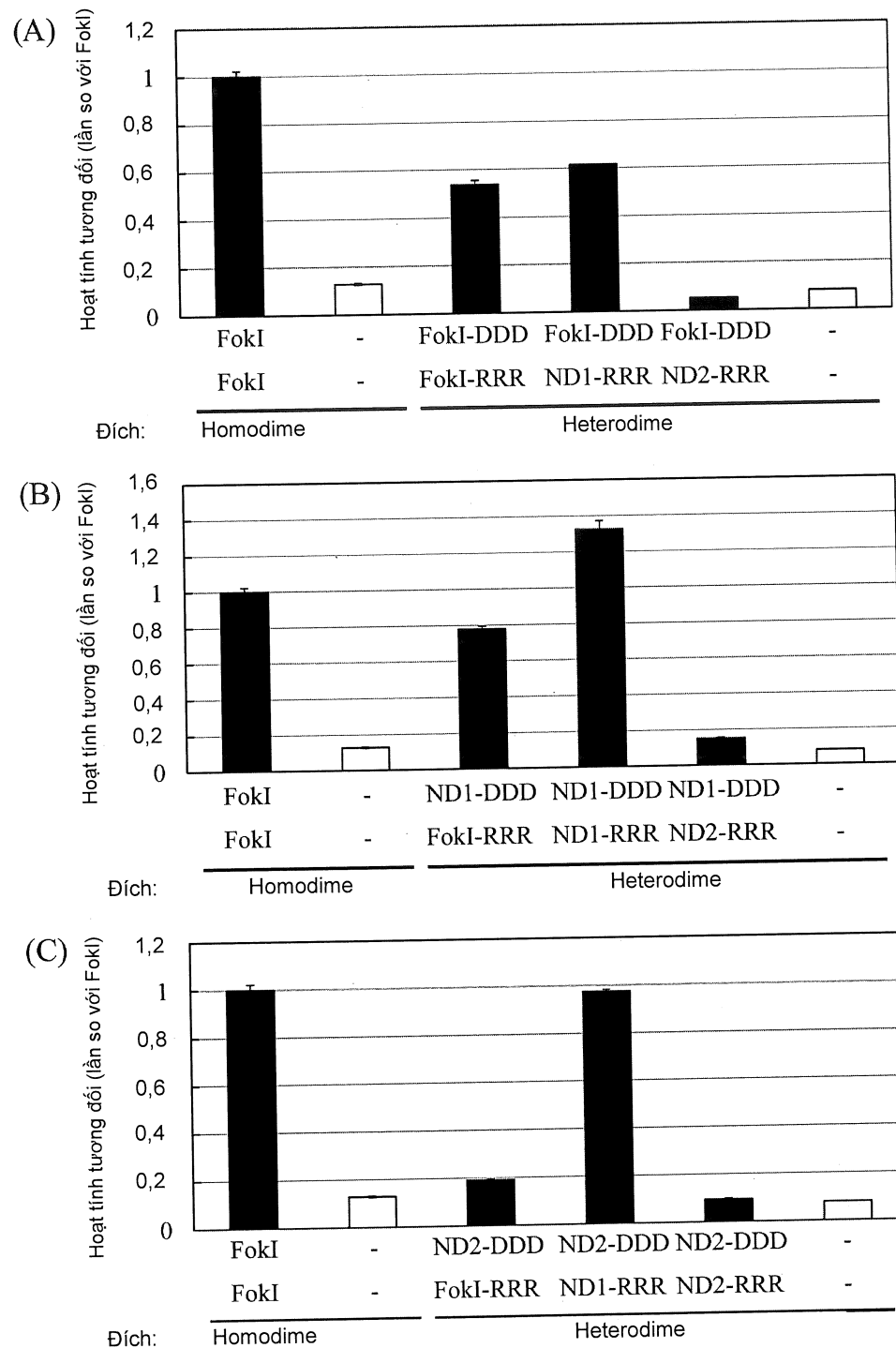


Fig.8B



9/14

Fig.9



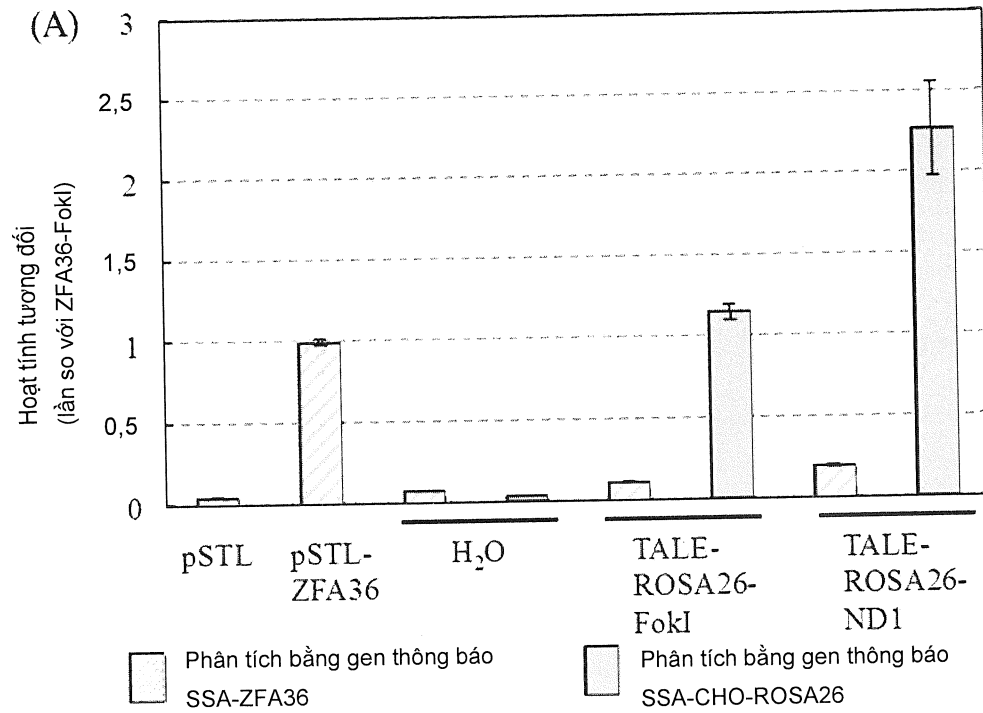
10/14

Fig.10

ZFA36-ZFA36 vị trí thứ nhất		ZFA36-ZFA36 vị trí thứ hai	
FokI (N=17) (tham chiếu) wt 13/17 CAGCCACCATCTTCC-----ACTCT-GAAGATGGGA +1 1/17 CAGCCACCATCTTCC-----ACTCT-GAAGATGGACAGGT +4 3/17 CAGCCACCATCTTCC-----ACTCTGAAAGATGGACAGGT +4 3/17 CAGCCACCATCTTCCACTTCCACTTCT-GAAGATGGACAGGT		FokI (N=20) (tham chiếu) wt 13/20 ACCTTACCATCTTCAA-----GAGA-GAAGATGGC +2 1/20 ACCTTACCATCTTCAA-----GAGA-GAAGATGGCTTAAA +4 3/20 ACCTTACCATCTTCAAGA-GAGAGA-GAAGATGGCTTAAA +5 1/20 ACCTTACCATCTTCAAGAGGAGAGA-GAAGATGGCTTAAA +232 1/20 ACCTTACCATCTTCAA(231bp)GAGTGAAGATGGCTTAAA Δ2 1/20 ACCTTACCATCTTCAA-----GA-GAAGATGGCTTAAA	
ND1 (N=14) (tham chiếu) wt 3/14 CAGCCACCATCTTCC-----ACTCTGAAGATGGGA +1 1/14 CAGCCACCATCTTCCA-----ACTCTGAAGATGGACAGGT -2/Δ1 2/14 CAGCCACCATCTTCCACT--CTCTGAAGATGGACAGGT +4 6/14 CAGCCACCATCTTCCACTCACTCTCTGAAGATGGACAGGT Δ2 1/14 CAGCCACCATCTTCC-----TCTGAAGATGGACAGGT Δ8 1/14 CAGCC-----C-----ACTCTGAAGATGGACAGGT		ND1 (N=17) (tham chiếu) wt 4/17 ACCTTACCATCTTCAA-----GAGAGAGATGGC +2 6/17 ACCTTACCATCTTCAA-----GAGAGAGATGGCTTAAA +4 3/17 ACCTTACCATCTTCAAGAGAGAGAGAGATGGCTTAAA Δ2 2/17 ACCTTACCATCTTCAA-----TAGAGATGGCTTAAA Δ4 2/17 ACCTTACCATCTTCAA-----ACAAGATGGCTTAAA	
ND2 (N=13) (tham chiếu) wt 9/13 CAGCCACCATCTTCCACTCT--GAAGATGGGA +1 1/13 CAGCCACCATCTTCCACTCT--TGAGATGGACAGGT +2 1/13 CAGCCACCATCTTCCACTCTCTGAAGATGGACAGGT Δ2 1/13 CAGCCACCATCTTCC--TCT--GAAGATGGACAGGT Δ4 1/13 CAGCCACCATCTT-----TCT--GAAGATGGACAGGT		ND2 (N=16) (tham chiếu) wt 7/16 ACCTTACCATCTTCAA-----GAGAGAGATGGC +1 1/16 ACCTTACCATCTTCAA-----GAGAGAGATGGCTTAAA +2 2/16 ACCTTACCATCTTCAA-----GAGAGAGATGGCTTAAA +4 5/16 ACCTTACCATCTTCAA-----GAGAGAGAGAGATGGCTTAAA +21 1/16 ACCTTACCATCTTCAA(20bp)AGAGAGAGATGGCTTAAA	
ND3 (N=15) (tham chiếu) wt 15/15 CAGCCACCATCTTCCACTCTGAAAGATGGGA		ND3 (N=16) (tham chiếu) wt 13/16 ACCTTACCATCTTCAA-----GAGAGAGATGGC +2 1/16 ACCTTACCATCTTCAAGAGAGAGAGATGGCTTAAA Δ2 1/16 ACCTTACCATCTTCAA-----GAGAGATGGCTTAAA Δ8 1/16 ACCTTACCATCTT-----AGTGGCTTAAA	
ND4 (N=16) (tham chiếu) wt 16/16 CAGCCACCATCTTCCACTCTGAAAGATGGACAGGT		ND4 (N=14) (tham chiếu) wt 14/14 CAGCCACCATCTTCAAGAGAGAGATGGCTTAAA	

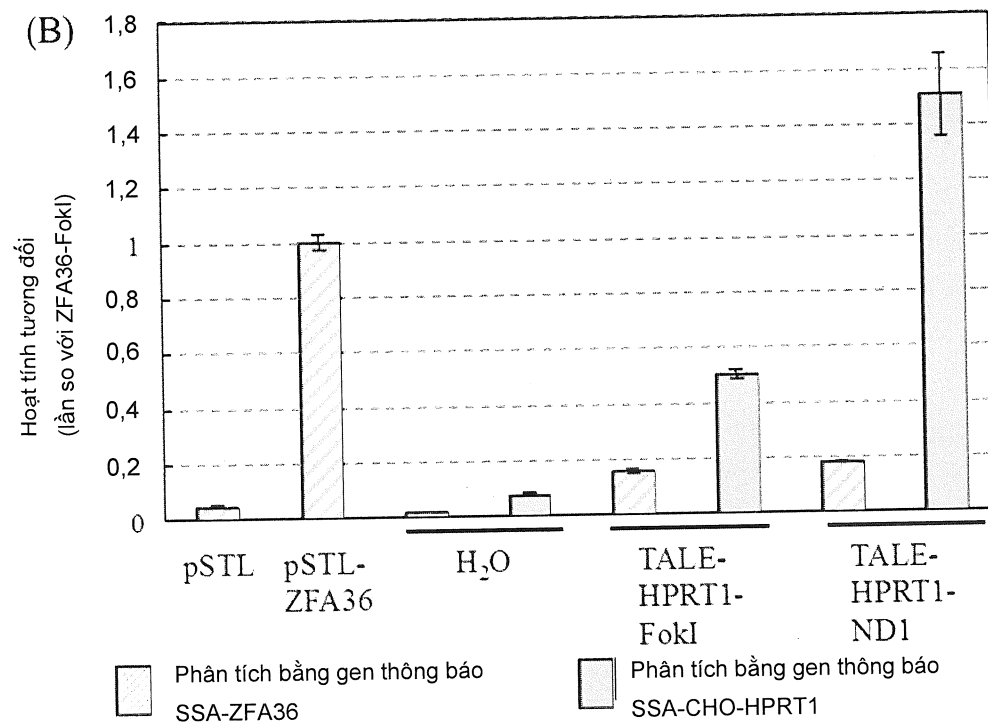
11/14

Fig.11A



12/14

Fig.11B



13/14

Fig.12A

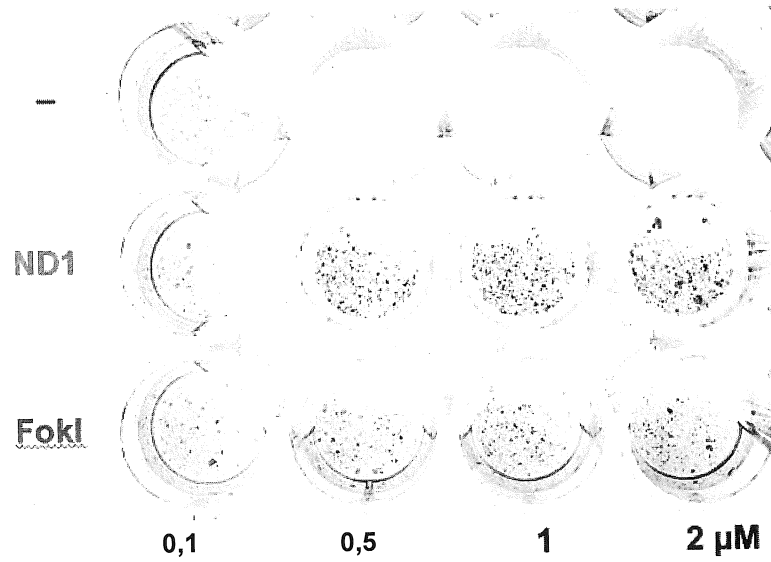
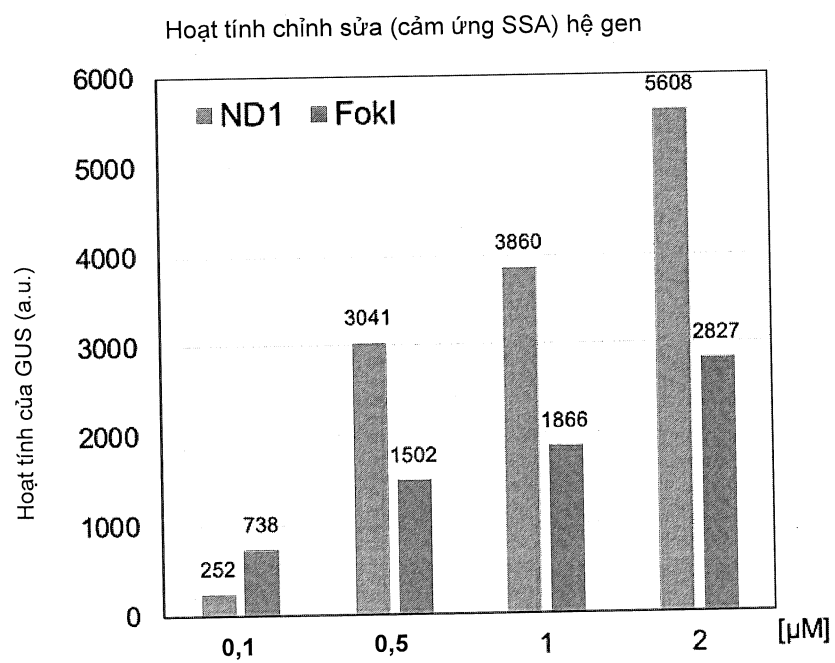
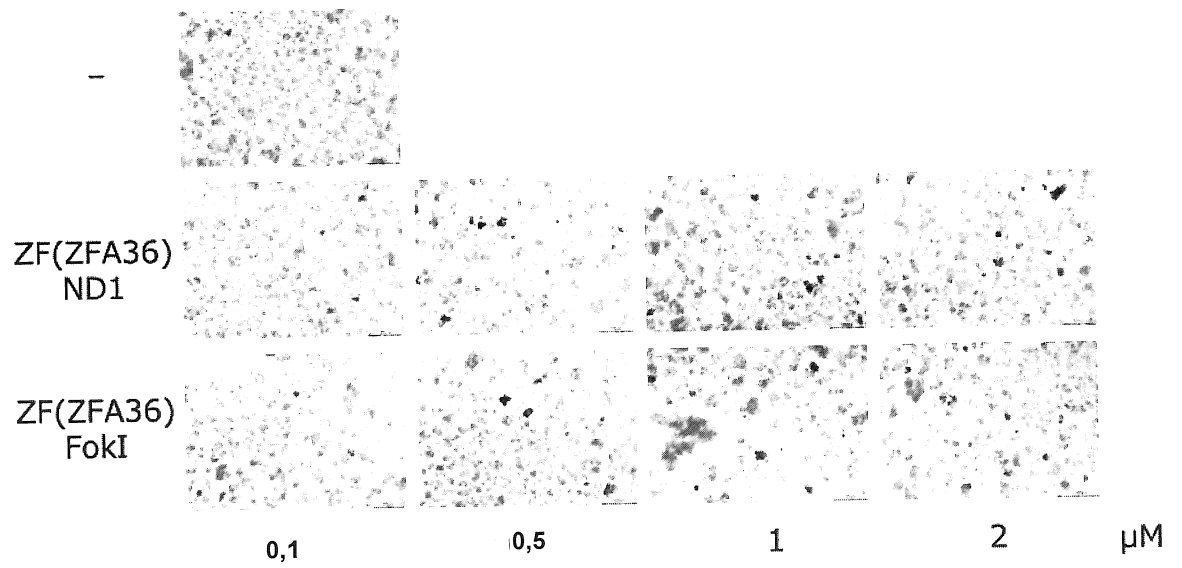


Fig.12B



14/14

Fig.13



DANH MỤC TRÌNH TỰ

<110> HIROSHIMA UNIVERSITY

<120> Vùng nucleaza mới và sử dụng chúng

<130> NEDO002WO

<150> JP2018-158710

<151> 2018-08-27

<160> 122

<170> PatentIn phiên bản 3.5

<210> 1

<211> 585

<212> PRT

<213> Bacillus sp. SGD-V-76

<400> 1

Met Ala Thr Lys Arg Val Val Arg Thr Phe Gly Trp Ile Gln Asn Pro
 1 5 10 15

Gly Lys Phe Glu Asn Leu Lys Arg Thr Val Gln Val Phe Asp Glu Asn
 20 25 30

Ser Lys Val Tyr Lys Glu Val Arg Asp Ile Lys Ile Pro Asn Leu Val
 35 40 45

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

Glu Asn Tyr Tyr Ser Tyr Lys Glu Leu Val Gly Thr Gly Thr Ser Val
 65 70 75 80

Arg Ala Gln Ala Pro Cys Asp Ala Ile Ile Gln Ala Ala Ile Gln Asp
 85 90 95

Gln Gly Asn Lys Lys Gly Tyr Ile Asp Asn Trp Ser Ser Asp Gly Phe
 100 105 110

Val Arg Trp Ala His Ala Leu Gly Phe Leu Glu Tyr Asn Ser Asp Lys
 115 120 125

Asp Ser Phe Tyr Leu Thr Asp Ile Gly Leu Glu Tyr Ser Arg Ser Thr
 130 135 140

Asp Asp Ser Leu Asn Glu Lys Ile Ile Leu Ile Glu Ala Leu Ser Ser
 145 150 155 160

Tyr Pro Pro Ala Ile Arg Ile Leu Thr Leu Leu Glu Asn Gly Glu His
 165 170 175

Leu Thr Lys Phe Asp Leu Gly Lys Gln Leu Gly Phe Ser Gly Glu Ser
 180 185 190

Gly Phe Thr Ser Leu Pro Gln Asn Leu Phe Leu Asp Ala Leu Ala Asn
 195 200 205

Ala Pro Ser Ser Pro Pro Thr Ile Lys Asn Ser Ile Arg Ser Asp Trp
 210 215 220

Glu Gly Ser Ala Asp Lys Tyr Ala Arg Met Ile Ser Gly Trp Leu Val
 225 230 235 240

Lys Met Gly Leu Val Ala Gln Gly Arg Lys Asn Phe Val Val Pro Thr
 245 250 255

Val Gly Gly Ser Thr Arg Asn Glu Tyr Ile Ser His Ala Phe Lys Ile
 260 265 270

Thr Lys Asp Gly Leu Asp Val Leu Arg Arg Ala Lys Gly Ala Ser Lys
 275 280 285

His Val Arg Val Pro Lys Arg Val Tyr Trp Glu Met Phe Ala Thr Asn
 290 295 300

Ile Asn Asp Arg Val Tyr Val Arg Thr Arg Arg Ala Tyr Thr Leu Tyr
 305 310 315 320

Leu Leu Met Arg Ser Asn Thr Ala Leu Asn Thr Val Gln Leu Gln Val
 325 330 335

Lys Leu Lys Ala Leu Gly Phe Glu Glu Leu Glu Glu Thr Ile Glu Asn
 340 345 350

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

Lys Gly Tyr His Phe Lys Asp Thr Ile Lys Pro Phe Glu Ile Pro Glu
 370 375 380

Phe Gly Val Thr Glu Pro Leu Val Lys Gly Glu Met Glu Lys Lys Lys
 385 390 395 400

Ser Asp Leu Arg His Lys Leu Lys His Val Pro His Glu Tyr Ile Glu
 405 410 415

Leu Ile Glu Ile Ala Gln Asp Ser Lys Gln Asn Arg Leu Phe Glu Phe
 420 425 430

Lys Val Val Glu Phe Leu Lys Glu Val Tyr Asp Tyr Asn Gly Lys His
 435 440 445

Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Leu Tyr Thr Asn Gly Leu
 450 455 460

Lys Thr Asp Tyr Gly Ile Ile Leu Asp Thr Lys Ala Tyr Lys Asp Gly
 465 470 475 480

Tyr Ser Leu Pro Ile Ser Gln Ala Asp Glu Met Gln Arg Tyr Val Asp
 485 490 495

Glu Asn Asn Asn Arg Asn Ala Ile Ile Asn Pro Asn Glu Trp Trp Lys
 500 505 510

Val Tyr Pro Asn Ser Ile Leu Asp Phe Lys Phe Leu Phe Val Ser Gly
 515 520 525

Phe Phe Lys Gly Asp Tyr Lys Lys Gln Leu Ala Arg Val Ser Asn Leu
 530 535 540

Thr Lys Arg Lys Gly Ala Val Leu Ser Val Glu Gln Leu Leu Leu Gly
 545 550 555 560

Gly Glu Lys Ile Lys Asp Gly Ser Leu Thr Leu Glu Asp Val Gly Asp
 565 570 575

Lys Phe Asn Asn Asp Glu Ile Ile Phe
 580 585

<210> 2

<211> 1758

<212> ADN

<213> Bacillus sp. SGD-V-76

<400> 2

atggcaacaa aacgagtggg tcgtacattt ggttggtac agaattccagg taaattcgag 60
 aacctaaaac gaacagtcca agtattcgat gaaaactcaa aagtatataa agaggtagg 120
 gatataaaga ttccaaatct agtgaaagat aaagatgtaa gaactacttt aatagctgct 180
 atgaatcgat tagaaaatta ttactcatat aaagagttgg ttggtacagg tacttcagtt 240
 agagctcaag cgccttgtga tgccattatt caagcagcga ttcaagacca aggtaataaa 300
 aaaggataca ttgataattg gtcttctgat ggattcgttc gttgggcaca tgcattgggg 360

tttcttgagt ataactcaga taaggactca ttttacctaa ctgatattgg attggaatat	420
tctcgttcta cagatgattc actaaatgaa aaaatcatac ttattgaagc attgtcatca	480
tatccaccag caattcgaat attgacatta cttgaaaatg gtgagcattt aacaaaattc	540
gacctaggta aacagttagg ctttagtgga gagagtgggt ttacttcatt accacaaaac	600
ttatTTTTtag atgcgttagc aaatgcacct tcatcaccac caacaattaa gaattctatt	660
cgttctgatt gggaagggtc tgcggataaa tatgcacgga tgatttcagg ctggcttgta	720
aagatggggc ttgttgcaaca aggtcgtaaa aattttgttg taccaacggg aggtggttct	780
actcgaatg agtatatatc tcatgctttt aaaattacaa aagatgggtc tgatgtgta	840
aggagagcaa aaggtgccag taaacatgta cgagtaccaa agcgtgtata ctgggagatg	900
ttcgctacca acattaacga tagagtgtat gtaagaacaa gacgggctta tacactttac	960
ttgttaatga gaagtaacac tgctctaaat actgttcaac ttcaagtga attaaaagct	1020
ttgggctttg aagaattgga agaaacaatt gagaatgatg taaaagggtt aattaactct	1080
ggaattttta tcaatgtgac gaataaagggt tatcatttta aagatacaat taaaccattc	1140
gaaattcctg aatttggcgt aacagaacca ttagtaaagg gggaaatgga gaagaagaaa	1200
tccgatttgc gtcataagtt aaaacacgta ccgcatgaat atattgaact aatagaaatt	1260
gcacaggatt caaagcaaaa tcgtttgttc gaatttaaag tagttgagtt tttgaaagaa	1320
gtttatgatt ataatggaaa gcacctaggg gggagtcgaa aaccagatgg cgactatat	1380
accaatgggt taaaaacaga ttatggaata attttagata ccaaagctta caaagatggt	1440
tatagtttac caatctccca ggcagatgaa atgcagcgtt atgtcgaatg aaataataat	1500

agaaatgcta ttattaatcc taacgaatgg tggaaagtgt atcctaattc tatcttggac 1560
 tttaaattct tattcgtaag tggttttttt aaaggtgact ataaaaaaca attagctcgt 1620
 gtaagtaatc ttacaaagcg taaaggtgct gtcctaagcg tagagcaact attattggga 1680
 ggggaaaaaa ttaaagatgg tagtttaaca ttagaagatg ttggtgacaa attcaataat 1740
 gatgaaatta ttttctaa 1758

<210> 3

<211> 579

<212> PRT

<213> Clostridium botulinum

<400> 3

Val Glu Ser Leu Arg Leu Ser Asp Ile Thr Ile Arg Thr Phe Gly Trp
 1 5 10 15

Ile Gln Asn Pro Ser Asp Phe Ser Lys Leu Lys Lys Ile Val Gln Leu
 20 25 30

Phe Asp His Thr Ser Ser Ile His Val Asp Leu Lys Glu Asn Arg Ile
 35 40 45

Pro Arg Leu Val Glu Lys Asp Asp Gly Arg Asp Arg Phe Ile Glu Ile
 50 55 60

Leu Asn Arg Ile Pro Leu Lys Ile Lys Tyr Ile Asp Leu Val Gly Thr
 65 70 75 80

Ser Phe Lys Pro Arg Ser Ser Gly Arg Cys Asn Gly Ile Ala Gln Ala
 85 90 95

Met Ile Glu Gly Gln Gly Lys Ser Phe Thr Asp Asn Trp Thr Ala Asp
 100 105 110

Gly Phe Ile Arg Trp Ala His Ala Leu Gly Phe Ile Gln Tyr Asp Tyr
 115 120 125

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

Ser Ile Asn Asp Ser Glu Gly Glu Asn Glu Ile Leu Ile Glu Ala Met
 145 150 155 160

Leu Ser Tyr Pro Pro Ala Val Arg Val Leu Asp Leu Leu Ser Leu Gly
 165 170 175

Glu His Leu Thr Lys Phe Asp Ile Gly Lys Asn Leu Gly Phe Asn Gly
 180 185 190

Glu Ser Gly Phe Thr Ser Leu Pro Leu Asn Ile Met Leu Gln Thr Leu
 195 200 205

Ala Asp Asn Asp Leu Val Ser Glu Lys Thr Lys Ile Arg Gln Asp Trp
 210 215 220

Glu Gly Ser Ser Asp Lys Tyr Ser Arg Met Ile Ala Gly Trp Leu Ser
 225 230 235 240

Lys Leu Gly Leu Val Lys Lys Gly Lys Lys Tyr Phe Asp Val Asn Ile
 245 250 255

Gly Gly Ile Ile His Arg Glu Tyr Ile Ser His Ala Tyr Lys Ile Thr
 260 265 270

Gly Glu Gly Leu Lys Gln Leu Arg Arg Ala Lys Gly Thr Thr Ser Ala
 275 280 285

Arg Arg Ile Glu Lys Arg Val Cys Trp Glu Met Phe Ala Thr Lys Asn
 290 295 300

Leu Asp Arg Thr Tyr Ile Arg Thr Arg Arg Ala Tyr Ile Leu Lys Phe
 305 310 315 320

Leu Glu Val Ser Ser Ser Leu Ile Thr Ile Glu Gln Ile Lys Thr Lys
325 330 335

Leu Leu Glu Lys Gly Tyr Asp Glu Thr Val Glu Thr Ile Lys Ser Asp
340 345 350

Ile Lys Gly Leu Ile Asn Ile Gly Leu Asn Ile Gln Glu Thr Thr Arg
355 360 365

Gly Tyr Ser Leu Lys Asp Ser Ile Asn Asp Phe Met Ile Pro Val Thr
370 375 380

Asp Leu Asn Gln Ile Val Lys Ser Ser Ile Glu Met Ser Lys Ala Asn
385 390 395 400

Met Arg Asp Asn Leu Gln Met Leu Pro His Asp Tyr Ile Glu Leu Ile
405 410 415

Glu Ile Ser Gln Asp Pro Tyr Gln Asn Arg Ile Phe Glu Met Lys Val
420 425 430

Met Asp Leu Phe Ile Asn Glu Tyr Gly Phe Ser Gly Ser His Leu Gly
435 440 445

Gly Ser Arg Lys Pro Asp Gly Ala Met Tyr Ala His Gly Phe Gly Val
450 455 460

Ile Val Asp Thr Lys Ala Tyr Lys Asp Gly Tyr Asn Leu Pro Ile Ser
465 470 475 480

Gln Ala Asp Glu Met Glu Arg Tyr Val Arg Glu Asn Ile Asp Arg Asn
485 490 495

Glu His Val Asn Ser Asn Arg Trp Trp Asn Ile Phe Pro Glu Asp Thr
500 505 510

Asn Glu Tyr Lys Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asn Phe
515 520 525

Glu Lys Gln Leu Glu Arg Ile Ser Ile Asp Thr Gly Val Gln Gly Gly
 530 535 540

Ala Leu Ser Val Glu His Leu Leu Leu Gly Ala Glu Tyr Ile Lys Arg
 545 550 555 560

Gly Ile Leu Thr Leu Tyr Asp Phe Lys Asn Ser Phe Leu Asn Lys Glu
 565 570 575

Ile Gln Phe

<210> 4

<211> 1740

<212> ADN

<213> Clostridium botulinum

<400> 4

gtggaatcat taagattatc agacataaca attagaacat ttgggtggat tcagaatcct	60
agtgatttta gtaaattaaa aaaaatagtt cagttatttg atcatacatc atctattcat	120
gtggatttga aagaaaaccg cataccacgg ttagtggaga aagatgacgg acgtgatagg	180
tttattgaga tattaaatcg gattccgtta aagataaagt atattgattt agtaggaaca	240
tcattttaac ctagaagtag tggtagatgt aatggaattg ctcaagcaat gattgaaggt	300
caaggtaaga gttttactga taattggaca gctgatgggt ttattaggtg ggctcatgca	360
cttggcttta tacaatatga ttatttagaa gatgagtttt caataacaca agcaggcttt	420
gatttttcaa ggtcaatcaa tgatagtga ggggaaaatg aaatattaat agaggcaatg	480
ttgtcttatc cacctgcagt tagagtttta gatttacttt cttaggaga gcatcttaca	540
aaatttgata ttggaaaaaa tctgggcttc aatggcgaat caggatttac atctttacct	600
ttaaatatca tgcttcaaac attagcagac aatgatttag taagtgaaaa gactaaaatt	660

cgtcaggatt gggaagggttc ttctgataag tattcacgaa tgatagccgg ctggttatcg	720
aaacttggtt tggtaaaaaa gggaaaaaag tattttgatg taaatattgg aggaataatt	780
catagagaat atatctctca tgcatataaa attactggag aaggtttaaa acaactccgt	840
agggctaaag gaactacaag tgcaaggcgt atagaaaaac gggtttgctg ggaaatgttc	900
gccacaaaga atctagatag gacatatatt cgtacaagaa gagcttatat tttaaagttt	960
ttagaagttt caagcagctt aattactatt gagcaaatta agactaagtt attggaaaaa	1020
ggttatgacg aaacagtaga aactattaaa tctgatatta aggggttaat taacattggt	1080
ttaaatattc aagaaaccac tcgaggttat tcattaaagg attcaattaa tgactttatg	1140
attccagtaa ccgacttaaa tcaaattggt aaaagtagta tagagatgtc taaagctaac	1200
atgagagaca atcttcagat gttgcctcat gattatattg agctaattga aatatcacia	1260
gatccatatc aaaatagaat atttgaaatg aaagttatgg atttatttat taatgagtat	1320
ggtttcagtg gttcgactt aggtggaagt aggaaacctg atggtgctat gtatgcacat	1380
ggttttggtg ttattgttga tacaaaagca tataaagatg ggtacaatct tcctataagt	1440
caagctgatg aaatggaacg atatgtaagg gaaaatattg atagaaatga acatgtgaac	1500
tcaaatagat ggtggaatat atttcccgaa gatacaaatg aatataagtt ttatttcgta	1560
agtggtttct ttaaaggaaa ttttgagaaa caattagaaa ggataagtat agatactggt	1620
gtacaaggag gggctttaag tgttgaacat ttattattag gtgctgagta cataaagagg	1680
ggtattctta ctttatatga ttttaagaat agctttttaa ataaggaaat tcaattttaa	1740

<210> 5

<211> 587

<212> PRT

<213> Clostridium sp. 7_2_43FAA

<400> 5

Met Ile Asn Ile Ile Asp Val Asn Asn Lys Thr Ile Arg Thr Phe Gly
 1 5 10 15

Trp Val Gln Asn Pro Ser Asn Phe Glu Ser Leu Lys Lys Val Val Ala
 20 25 30

Ile Phe Asp Asn Thr Ser Lys Thr Tyr Asn Glu Leu Lys Asp Lys Lys
 35 40 45

Ile Lys Lys Leu Val Asp Glu Arg Asp Gly Gln Lys Glu Leu Leu Asn
 50 55 60

Ala Leu Asn Ala Asn Pro Leu Lys Ile Lys Tyr Cys Asn Leu Val Gly
 65 70 75 80

Thr Ser Phe Thr Pro Arg Ser Ser Ala Arg Cys Asn Gly Ile Val Gln
 85 90 95

Ala Thr Val Lys Gly Gln Arg Lys Glu Phe Ile Asp Asp Trp Ser Ser
 100 105 110

Asp Asn Phe Val Arg Trp Ala His Ala Leu Gly Phe Ile Lys Tyr Asn
 115 120 125

Tyr Asp Thr Asp Thr Phe Glu Ile Thr Asp Val Gly Arg Lys Tyr Val
 130 135 140

Gln Ser Glu Asp Asp Ser Asn Glu Glu Ser Thr Ile Leu Glu Glu Ala
 145 150 155 160

Met Leu Ser Tyr Pro Pro Val Ala Arg Val Leu Thr Leu Leu Ser Asn
165 170 175

Gly Glu His Leu Thr Lys Tyr Glu Ile Gly Lys Lys Leu Gly Phe Val
180 185 190

Gly Glu Ala Gly Phe Thr Ser Leu Pro Leu Asn Val Leu Ile Met Thr
195 200 205

Leu Ala Thr Thr Asp Glu Pro Lys Glu Lys Asn Lys Ile Lys Thr Asp
210 215 220

Trp Asp Gly Ser Ser Asp Lys Tyr Ala Arg Met Ile Ser Gly Trp Leu
225 230 235 240

Val Lys Leu Gly Leu Leu Val Gln Arg Pro Lys Leu Val Thr Val Asp
245 250 255

Phe Gly Gly Glu Leu Tyr Ser Glu Thr Ile Gly His Ala Tyr Met Ile
260 265 270

Thr Asp Arg Gly Leu Lys Ala Val Arg Arg Leu Leu Gly Ile Asn Lys
275 280 285

Val Ala Arg Val Ser Lys Asn Val Phe Trp Glu Met Leu Ala Thr Lys
290 295 300

Gly Ile Asp Lys Asn Tyr Ile Arg Thr Arg Arg Ala Tyr Ile Leu Lys
305 310 315 320

Ile Leu Ile Glu Ser Asn Lys Val Leu Thr Leu Glu Asp Ile Lys Gly
325 330 335

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

Asp Ile Asn Gly Leu Ile Asn Thr Gly Ile Asn Ile Lys Ser Glu Thr
 355 360 365

Thr Gly Tyr Lys Ile Tyr Asp Ser Ile Asn Asp Phe Ile Ile Pro Lys
 370 375 380

Thr Gly Asp Thr Glu Gly Ile Lys Ser Asn Ile Ser Leu Leu Lys Asp
 385 390 395 400

Glu Leu Arg Gly Gln Ile Ser His Ile Ser His Glu Tyr Leu Ser Leu
 405 410 415

Ile Asp Leu Ala Phe Asp Ser Lys Gln Asn Arg Leu Phe Glu Met Lys
 420 425 430

Val Leu Glu Leu Leu Val Asn Glu Tyr Gly Phe Lys Gly Arg His Leu
 435 440 445

Gly Gly Ser Arg Lys Pro Asp Gly Ile Val Tyr Ser Thr Thr Leu Glu
 450 455 460

Asp Asn Phe Gly Ile Ile Val Asp Thr Lys Ala Tyr Ser Glu Gly Tyr
 465 470 475 480

Ser Leu Pro Ile Ser Gln Ala Asp Glu Met Glu Arg Tyr Val Arg Glu
 485 490 495

Asn Ser Asn Arg Asp Glu Glu Val Asn Pro Asn Lys Trp Trp Glu Asn
 500 505 510

Phe Ser Glu Glu Val Lys Lys Tyr Tyr Phe Val Phe Ile Ser Gly Ser
 515 520 525

Phe Lys Gly Lys Phe Glu Glu Gln Leu Arg Arg Leu Ser Met Thr Thr
 530 535 540

Gly Val Asn Gly Ser Ala Val Asn Val Val Asn Leu Leu Leu Gly Ala
 545 550 555 560

Glu Lys Ile Arg Ser Gly Glu Met Thr Ile Glu Glu Leu Glu Arg Ala
 565 570 575

Met Phe Asn Asn Ser Glu Phe Ile Leu Lys Tyr
 580 585

<210> 6

<211> 1764

<212> ADN

<213> Clostridium sp. 7_2_43FAA

<400> 6

atgattaaca taattgacgt aaataataaa actattagaa cttttggatg gggttcagaat	60
ccaagtaatt ttgaaagctt aaaaaaagtt gtagcaatat ttgataatac atcaaagaca	120
tataatgaat taaaagataa gaaaataaaa aagttagtag acgaaagaga tgggcaaaaa	180
gaactattaa atgcttttaa tgctaatacct ttaaagataa agtattgtaa tttagtaggg	240
acatctttta ctccacgaag ttctgcaaga tgtaatggaa tagtacaagc cacagtaaaa	300
ggacaaagaa aagagttcat agatgattgg tcatctgata actttgtaag atgggcacat	360
gcttttaggct ttattaaata taattatgat acagatacat ttgagataac cgatgtagga	420
agaaagtatg tacagagtga agatgatagc aatgaggaaa gtacaatttt agaagaagca	480
atgttatcat atccaccagt agcaagggtt ttgacattgc tttccaatgg agaacattta	540
acaaaatatg aaataggtaa gaaattagga tttgtagggg aagctggttt tactagtttg	600
ccattgaatg tattaattat gacattagct actacagatg agccgaagga aaaaaataaa	660
attaaaacag actgggatgg gtcttctgat aaatatgcaa gaatgatttc tggttggtta	720
gtaaaattag gattattggt acaaagacct aaattgggta ctgtagattt tgggtggtgaa	780

ttatatagcg aaacaatagg acatgcttat atgattacgg ataggggatt aaaagctggt	840
agaagattat taggaataaa taaagttgcg agggtttcaa aaaatgtttt ttgggaaatg	900
cttgcaacaa aaggaataga taaaaattat ataagaactc gaagagcata tattttaaaa	960
atattaattg aatctaataa ggtattaact ttagaggata taaaaggaaa gttgaaatta	1020
gcgagtatta atgaatctat taatactatt aaagacgata ttaatggatt aattaatata	1080
ggtataaata ttaagagtga aacaactggg tataaaatat atgattcaat taatgatttt	1140
attataccta aaacaggaga tacagaggga ataaagagta atattagctt attaaaagat	1200
gaattaagag ggcaaataag tcatatttca catgagtatt tatcattaat agatttagca	1260
tttgatagta agcaaaatag gttatttgaa atgaaagttt tagagttggt agttaatgag	1320
tatggattta aaggaaggca tctaggtgga agtagaaaac ctgatggtat agtttattca	1380
acaactttag aagataattt tggaataata gtagatacta aagcttattc agaagggtat	1440
agcttaccta taagtcaagc tgatgaaatg gagagatatg ttagagaaaa ctctaataga	1500
gatgaggaag taaaccctaa taagtgggtg gaaaactttt cagaagaagt caaaaagtat	1560
tattttgtat tcatatcagg ttcctttaag ggcaaatttg aagaacaatt aagaagattg	1620
agcatgacta ctgggggttaa cggttcagct gttaatgtag taaatttatt actaggagct	1680
gaaaaaataa ggtctggaga aatgactatt gaagaattag agagagctat gttaataaac	1740
tcggaattta tattgaagta ttaa	1764

<210> 7

<211> 589

<212> PRT

<213> Clostridiales bacterium KA00134

<400> 7

Met Ala Glu Arg Thr Leu Gly Trp Ile Gln Asn Pro Ser Ser Phe Glu
 1 5 10 15

Asn Leu Lys Asn Val Val Ser Val Phe Asp Lys Asn Thr Asp Ile Tyr
 20 25 30

Lys Glu Ile Leu Asn Thr Lys Leu Pro Lys Leu Val Lys Asp Leu Asp
 35 40 45

Leu Gln Asn Lys Leu Ile Ser Glu Met Lys Lys Asp Pro Leu Glu Met
 50 55 60

Asp Tyr Val Leu Leu Lys Gly His Gly Ile Lys Ser Gly Gln Lys Arg
 65 70 75 80

Ser Asp Ala Ala Cys Ser Gly Ile Val Gln Ala Ala Ile Thr Thr Gln
 85 90 95

Gly Gly Arg Ala Tyr Thr Asp Asp Trp Thr Ala Asp Gly Phe Leu Arg
 100 105 110

Trp Gly Ile Ser Ile Gly Leu Leu Asp Tyr Asp Thr Glu Asn Asp Thr
 115 120 125

Val Ser Ile Thr Lys Leu Gly Glu Lys Phe Val Lys Ser Asn Ser Glu
 130 135 140

Asp Ser Asp Lys Glu Ile Leu Ile Ser Ala Phe Leu Ser Tyr Pro Pro
 145 150 155 160

Ala Val Arg Ile Leu Thr Leu Leu Glu Ser Gly Asp His Leu Thr Lys
 165 170 175

Phe Glu Leu Gly Lys Gln Leu Gly Gly Leu Gly Glu Ala Gly Phe Thr
 180 185 190

Ser Ile Pro Gln Asp Leu Tyr Ile Gln Ala Ile Glu Leu Ala Ser Glu
195 200 205

Thr Asp Lys Ala Ser Ile Arg Ser Asn Thr Glu Gly Ser Ala Asp Lys
210 215 220

Tyr Ala Arg Met Ile Ser Gly Trp Leu Ser Lys Val Gly Leu Val Gln
225 230 235 240

Arg Val Glu Lys Glu Val Ser Thr Lys Ile Gly Asn Val Glu Tyr Met
245 250 255

Val Asn Ile Gly His Ser Phe Arg Ile Thr Leu Asn Gly Ile Lys Glu
260 265 270

Leu Lys Arg Ala Arg Gly Leu Ser Ser Tyr Pro Lys Thr Asp Lys Ile
275 280 285

Val Tyr Trp Gln Met Leu Ala Thr Lys Gly Lys Asp Arg Asp Tyr Ile
290 295 300

Arg Asn Arg Arg Gly Tyr Ile Ile Lys Ala Ile Asn Asn Arg Glu Arg
305 310 315 320

Asn Leu Glu Gly Ile Lys Ala Tyr Leu Met Glu Asn Asn Ile Asp Glu
325 330 335

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

Leu Ser Leu Lys His Ser Arg Asn Gly Tyr Val Ile Asp Asp Asn Ile
355 360 365

Ile Lys Leu Glu Ile Pro Arg Thr Lys Ile Ser Lys Thr Asn Ile Leu
370 375 380

Glu Leu Lys Asp Lys Val Arg Asp Lys Leu Lys Tyr Val Asp His Arg
385 390 395 400

Tyr Leu Ala Leu Ile Asp Leu Ala Tyr Asp Gly Thr Ala Asn Arg Asp
405 410 415

Phe Glu Ile Gln Thr Ile Asp Leu Leu Ile Asn Glu Leu Lys Phe Lys
420 425 430

Gly Val Arg Leu Gly Glu Ser Arg Lys Pro Asp Gly Ile Ile Ser Tyr
435 440 445

Asn Ile Asn Gly Val Ile Ile Asp Asn Lys Ala Tyr Ser Thr Gly Tyr
450 455 460

Asn Leu Pro Ile Asn Gln Ala Asp Glu Met Ile Arg Tyr Ile Glu Glu
465 470 475 480

Asn Gln Thr Arg Asp Glu Lys Ile Asn Ser Asn Lys Trp Trp Glu Ser
485 490 495

Phe Asp Glu Lys Val Lys Asp Phe Asn Tyr Leu Phe Val Ser Ser Phe
500 505 510

Phe Lys Gly Asn Phe Lys Asn Asn Leu Lys His Ile Ala Asn Arg Thr
515 520 525

Gly Val Asn Gly Gly Ala Ile Asn Val Glu Asn Leu Leu Tyr Phe Ala
530 535 540

Glu Glu Leu Lys Ala Gly Arg Ile Ser Tyr Leu Asp Ser Phe Lys Met
545 550 555 560

Tyr Asn Asn Asp Glu Ile Tyr Leu Gly Asp Ile Ser Asp Tyr Ser Tyr
565 570 575

Val Lys Phe Ala Ala Glu Glu Glu Gly Glu Tyr Leu Thr
580 585

<210> 8

<211> 1770

<212> ADN

<213> Clostridiales bacterium KA00134

<400> 8

atggcagaaa gaacattggg ttggatacaa aatccatcaa gttttgaaaa tttaaaaaat	60
gtagtttctg tttttgataa aaatactgat atatataagg aaatattaaa tactaaactt	120
cctaaattag taaaagattt agattttacaa aataaactca tttcagaaat gaaaaaagat	180
ccacttgaaa tggactatgt ttattataaaa ggtcatggta taaaatcagg tcagaaaaga	240
tccgatgcag cgtgtagtgg tattgttcaa gctgccatta caacacaagg cggtagggct	300
tatacagacg attggactgc tgatggattt ttaagatggg gaatatccat aggactttta	360
gattatgata cagaaaatga tacagtttct ataacaaaac taggagaaaa gtttgtaaaa	420
tctaattctg aagattcaga taaagaaata ctaatttcag cttttttgtc ataccaccca	480
gccgtaagaa ttttaacttt attggaaagt ggagaccatt taacaaaatt tgagcttggt	540
aagcaattag gtggacttgg agaagctggg ttacatcaa ttcctcaaga tttatatatt	600
caagcaatag aacttgcttc tgaaacagat aaagcaagta ttcgatcaaa cacagaggga	660
tcagctgata aatatgccag aatgatttca ggatggttat ctaaggttgg tttagttcaa	720
agagttgaaa aagaagtttc aacaaaaatt ggaaatgttg aatacatggg aaatataggt	780
cattccttta gaattacttt gaatgggata aaagaactta aaagagcaag gggattatct	840
tcatatccaa agactgataa aattgtttat tggcaaatgc tagcaaccaa gggtaaagat	900
agagattata ttagaaacag acgtggatat ataatcaaag ctatcaaaa cagagaaaaga	960

aatttagaag gcattaaggc ttatttaatg gaaaataata ttgatgaaag tattacaact 1020
atagaagacg aactaaaagt tcttaaagca atgggtcttt cacttaaaca tagtaggaat 1080
ggctatgtaa ttgatgataa cattataaaa ttagaaatac ctagaactaa gatttctaag 1140
acaaatatac tagaattaaa ggataaagta agagataaat taaaatatgt tgaccacaga 1200
tatcttgccg taattgattt agcttatgac ggaactgcc aatagagattt tgagattcaa 1260
acaatagacc ttttaattaa tgaacttaag tttaaagggg ttcgtctagg agaaagcaga 1320
aaaccagatg gaattatatt ttataatatt aatggcgtaa ttatagataa caaggcttat 1380
tctacgggtt ataatctacc aattaatcaa gcagatgaaa tgattagata tatcgaagaa 1440
aatcaaacta gagatgaaaa aataaattca aataaatggt gggaaagctt tgatgaaaaa 1500
gttaaggatt tcaactatct atttgtatca tctttcttta agggaaactt taagaataat 1560
ttgaaacata tagccaatag aactggtgta aatggtggag ctattaatgt agaaaattta 1620
ttatatcttg cagaagaatt aaaagctgga agaatatctt atcttgactc atttaaaatg 1680
tataacaatg atgaaatata tttaggagat attagtactc atagttatgt taaatttgct 1740
gcagaagaag aaggcgaata tttacataa 1770

<210> 9

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi xuôi thay thế ND1

<400> 9

gatgagatca tcttctaagg gcccttcgaa ggtaagcc

38

<210> 10

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi ngược thay thế ND1

<400> 10

ctctcccttt actaaagctc ttgccgcagc accc

34

<210> 11

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi xuôi thay thế ND2

<400> 11

aaggagattc agttctaagg gcccttcgaa ggtaagcc

38

<210> 12

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi ngược thay thế ND2

<400> 12

actacttttta acaatagctc ttgccgcagc accc

34

<210> 13

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi xuôi thay thế ND3

<400> 13

aacagcgagt tcatataagg gcccttcgaa ggtaagcc

38

<210> 14

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi ngược thay thế ND3

<400> 14

gttactcttg attccagctc ttgccgcagc accc

34

<210> 15

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi xuôi thay thế ND4

<400> 15

gatgaaatct acctgtaagg gcccttcgaa ggtaagcc

38

<210> 16

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi ngược thay thế ND4

<400> 16

gtttgtcttg gagatagctc ttgccgcagc accc

34

<210> 17

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-FokI-GS-F

<400> 17

agaacacacg gatccctagt caaaagtga ctggaggaga agaaatctg

49

<210> 18

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

<400> 18

ggatccgtgt gttctctgggt gtctcacgag attg

34

<210> 19

<211> 48

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND1-GS-F

<400> 19

agaacacacg gatccttagt aaaggagag atggagaaga aaaagagc

48

<210> 20

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

<400> 20

ggatccgtgt gttctctggt gtctcacgag attg

34

<210> 21

<211> 47

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để biến đổi trình tự nối, ZFN-ND2-GS-F

<400> 21

agaacacacg gatccattgt taaaagtagt attgagatgt ccaaggc

47

<210> 22

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-GS-ngược-R

<400> 22

ggatccgtgt gttctctggt gtctcacgag attg

34

<210> 23

<211> 51

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để biến đổi trình tự nối, ZFN-FokI-RPGEKP-F

<400> 23

gacccgggga gaagccccta gtcaaaagtg aactggagga gaagaaatct g

51

<210> 24

<211> 44

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

<400> 24

gcttctcccc gggctctgtgt gttctctggg gtctcacgag attg

44

<210> 25

<211> 50

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để biến đổi trình tự nối, ZFN-ND1-RPGEKP-F

<400> 25

gacccgggga gaagccctta gtaaaggag agatggagaa gaaaaagagc

50

<210> 26

<211> 44

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

<400> 26

gcttctcccc gggctctgtgt gttctctggt gtctcacgag attg

44

<210> 27

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-ND2-RPGEKP-F

<400> 27

gacccgggga gaagcccatt gttaaaagta gtattgagat gtccaaggc

49

<210> 28

<211> 44

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để biến đổi trình tự nối, ZFN-RPGEKP-R

<400> 28

gcttctcccc gggctctgtgt gttctctggt gtctcacgag attg

44

<210> 29

<211> 55

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-F

<400> 29

cccggggctg cggcaagagc tctagtcaaa agtgaactgg aggagaagaa atctg

55

<210> 30

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

<400> 30

tgccgcagcc ccgggacccg tgtgtgttct ctggtgtctc acgagattg

49

<210> 31

<211> 54

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để biến đổi trình tự nối, ZFN-ND1-TGPGAAARA-F

<400> 31

cccggggctg cggcaagagc tttagtaaag ggagagatgg agaagaaaaa gagc

54

<210> 32

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

<400> 32

tgccgcagcc ccgggacccg tgtgtgttct ctggtgtctc acgagattg

49

<210> 33

<211> 53

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để biến đổi trình tự nối, ZFN-ND2-TGPGAAARA-F

<400> 33

cccggggctg cggcaagagc tattgttaaa agtagtattg agatgtccaa ggc

53

<210> 34

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để biến đổi trình tự nối, ZFN-FokI-TGPGAAARA-R

<400> 34

tgccgcagcc ccgggacccg tgtgtgttct ctggtgtctc acgagattg

49

<210> 35

<211> 59

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến Sharkey vào FokI-ND

<400> 35

aaggtaatgg aattttttat gaaagtttat ggatatagag gtgaacattt gggtggatc

59

<210> 36

<211> 57

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến Sharkey vào FokI-ND

<400> 36

aaattccatt accttcattt caagaattct atcctgagta ggatttctgg caatttc

57

<210> 37

<211> 59

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến Sharkey vào ND1

<400> 37

ttaaggtagt ggaatttctg aaagaagtgt acgattacaa tggtagcat ctgggcggc 59

<210> 38

<211> 56

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến Sharkey vào ND1

<400> 38

attccactac cttaaactcg aagagtctgt tctgcttggg gtcctgagca atttcg 56

<210> 39

<211> 30

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến Sharkey vào ND2

<400> 39

tccggggagc atctgggggg tagccggaag 30

<210> 40

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến Sharkey vào ND2

<400> 40

cagatgctcc ccggagaagc catactcggt gatg

34

<210> 41

<211> 19

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi xuôi thay thế ZFA

<400> 41

acgggtgctg cggcaagag

19

<210> 42

<211> 43

<212> ADN

<213> Nhân tạo

<220>

<223> Đoạn mỗi ngược thay thế ZFA

<400> 42

aggtctctcg tgggtctag atttatcgtc gtcctccttg tag

43

<210> 43

<211> 54

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến FokI-DDD vào

<400> 43

gaagaaaatc aaacacgaga caaacatatc aaccctaagc aatggtggaa agtc

54

<210> 44

<211> 50

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến FokI-DDD vào

<400> 44

tgtttgattt tcttcgacat agtcttgcat ttcattctgct tggccaattg

50

<210> 45

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến ND1-DDD vào

<400> 45

cgacgaaaat aacaataggg acgcaatcat caaccctaatt gagtggtgg

49

<210> 46

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến ND1-DDD vào

<400> 46

ttgttatattt cgtcgacata gtcttgcatc tcattcgccct gagaaattg

49

<210> 47

<211> 49

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến ND2-DDD vào

<400> 47

tcagggagaa cattgacaga gacgagcacg tcaattctaa ccgctggtg

49

<210> 48

<211> 44

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến ND2-DDD vào

<400> 48

caatgttctc cctgacgtag tcttccattt catcggcctg gctg

44

<210> 49

<211> 43

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến vectơ FokI-RRR vào

<400> 49

ttaaatacgga tcactaattg taatggagct gttcttagtg tag

43

<210> 50

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến vectơ FokI-RRR vào

<400> 50

atttcccggg cttggccaat tggcagatta taac

34

<210> 51

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến chèn đoạn FokI-RRR vào

<400> 51

ccaagcccgg gaaatgcaac gatatgtcga agaaaatc

38

<210> 52

<211> 42

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến chèn đoạn FokI-RRR vào

<400> 52

agtgatccga tttaatcgtg taagctgagc tttgtagttt cc

42

<210> 53

<211> 28

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến vector ND1-RRR vào

<400> 53

gtgtccagac tgaccaagcg caagggcg

28

<210> 54

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến vector ND1-RRR vào

<400> 54

catctcccgg gcctgagaaa ttgggaggct gtag

34

<210> 55

<211> 45

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến chèn đoạn ND1-RRR vào

<400> 55

caggcccggg agatgcaaag atatgttgac gaaaataaca atagg

45

<210> 56

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để đưa đột biến chèn đoạn ND1-RRR vào

<400> 56

ggtcagtctg gacacccttg ccagctgttt cttg

34

<210> 57

<211> 28

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để đưa đột biến vector ND2-RRR vào

<400> 57

atctcaaggg ataccggggt gcagggcg

28

<210> 58

<211> 31

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến vectơ ND2-RRR vào

<400> 58

catttcccgg gcctggctga tggggagatt g

31

<210> 59

<211> 40

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để đưa đột biến chèn đoạn ND2-RRR vào

<400> 59

caggcccggg aaatggaaag gtacgtcagg gagaacattg

40

<210> 60

<211> 42

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để đưa đột biến chèn đoạn ND2-RRR vào

<400> 60

ggtatccctt gagattcttt ccaactgttt ctcaaagttt cc

42

<210> 61

<211> 17

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi M13-F

<400> 61

gtaaaacgac ggccagt

17

<210> 62

<211> 24

<212> ADN

<213> Cricetulus griseus

<400> 62

accatcttcc actctgaaga tgga

24

<210> 63

<211> 24

<212> ADN

<213> Cricetulus griseus

<400> 63

accatcttca agagagaaga tggc

24

<210> 64

<211> 24

<212> ADN

<213> Cricetulus griseus

<400> 64

accatcttcc ttgatgaaga tgcc

24

<210> 65

<211> 24

<212> ADN

<213> Cricetulus griseus

<400> 65

gtcaccttca agtctgaaga tggt

24

<210> 66

<211> 24

<212> ADN

<213> *Cricetulus griseus*

<400> 66

ttcaccttct taagtgaaga tggc

24

<210> 67

<211> 195

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của FokI-DDD

<400> 67

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys
1 5 10 15

Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg
20 25 30

Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe
35 40 45

Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys
50 55 60

Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val
65 70 75 80

Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly
85 90 95

Gln Ala Asp Glu Met Gln Asp Tyr Val Glu Glu Asn Gln Thr Arg Asp
100 105 110

Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val
115 120 125

Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr
 130 135 140

Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala
 145 150 155 160

Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala
 165 170 175

Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu
 180 185 190

Ile Asn Phe
 195

<210> 68
 <211> 195
 <212> PRT
 <213> Nhân tạo

<220>
 <223> Trình tự axit amin của FokI-RRR

<400> 68

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys
 1 5 10 15

Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg
 20 25 30

Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe
 35 40 45

Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val
65 70 75 80

Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly
85 90 95

Gln Ala Arg Glu Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn
100 105 110

Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val
115 120 125

Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr
130 135 140

Lys Ala Gln Leu Thr Arg Leu Asn Arg Ile Thr Asn Cys Asn Gly Ala
145 150 155 160

Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala
165 170 175

Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu
180 185 190

Ile Asn Phe
195

<210> 69

<211> 195

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND1-DDD

<400> 69

Leu Val Lys Gly Glu Met Glu Lys Lys Lys Ser Asp Leu Arg His Lys
1 5 10 15

Leu Lys His Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Gln
20 25 30

Asp Ser Lys Gln Asn Arg Leu Phe Glu Phe Lys Val Val Glu Phe Leu
35 40 45

Lys Glu Val Tyr Asp Tyr Asn Gly Lys His Leu Gly Gly Ser Arg Lys
50 55 60

Pro Asp Gly Ala Leu Tyr Thr Asn Gly Leu Lys Thr Asp Tyr Gly Ile
65 70 75 80

Ile Leu Asp Thr Lys Ala Tyr Lys Asp Gly Tyr Ser Leu Pro Ile Ser
85 90 95

Gln Ala Asp Glu Met Gln Asp Tyr Val Asp Glu Asn Asn Asn Arg Asp
100 105 110

Ala Ile Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Asn Ser Ile
115 120 125

Leu Asp Phe Lys Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asp Tyr
130 135 140

Lys Lys Gln Leu Ala Arg Val Ser Asn Leu Thr Lys Arg Lys Gly Ala
145 150 155 160

Val Leu Ser Val Glu Gln Leu Leu Leu Gly Gly Glu Lys Ile Lys Asp
165 170 175

Gly Ser Leu Thr Leu Glu Asp Val Gly Asp Lys Phe Asn Asn Asp Glu
180 185 190

Ile Ile Phe
195

<210> 70

<211> 195

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND1-RRR

<400> 70

Leu Val Lys Gly Glu Met Glu Lys Lys Lys Ser Asp Leu Arg His Lys
 1 5 10 15

Leu Lys His Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Gln
 20 25 30

Asp Ser Lys Gln Asn Arg Leu Phe Glu Phe Lys Val Val Glu Phe Leu
 35 40 45

Lys Glu Val Tyr Asp Tyr Asn Gly Lys His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Leu Tyr Thr Asn Gly Leu Lys Thr Asp Tyr Gly Ile
 65 70 75 80

Ile Leu Asp Thr Lys Ala Tyr Lys Asp Gly Tyr Ser Leu Pro Ile Ser
 85 90 95

Gln Ala Arg Glu Met Gln Arg Tyr Val Asp Glu Asn Asn Asn Arg Asn
 100 105 110

Ala Ile Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Asn Ser Ile
 115 120 125

Leu Asp Phe Lys Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asp Tyr
 130 135 140

Lys Lys Gln Leu Ala Arg Val Ser Arg Leu Thr Lys Arg Lys Gly Ala
 145 150 155 160

Val Leu Ser Val Glu Gln Leu Leu Leu Gly Gly Glu Lys Ile Lys Asp
 165 170 175

Gly Ser Leu Thr Leu Glu Asp Val Gly Asp Lys Phe Asn Asn Asp Glu
 180 185 190

Ile Ile Phe
 195

<210> 71

<211> 191

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND2-DDD

<400> 71

Ile Val Lys Ser Ser Ile Glu Met Ser Lys Ala Asn Met Arg Asp Asn
 1 5 10 15

Leu Gln Met Leu Pro His Asp Tyr Ile Glu Leu Ile Glu Ile Ser Gln
 20 25 30

Asp Pro Tyr Gln Asn Arg Ile Phe Glu Met Lys Val Met Asp Leu Phe
 35 40 45

Ile Asn Glu Tyr Gly Phe Ser Gly Ser His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Met Tyr Ala His Gly Phe Gly Val Ile Val Asp Thr
 65 70 75 80

Lys Ala Tyr Lys Asp Gly Tyr Asn Leu Pro Ile Ser Gln Ala Asp Glu
 85 90 95

Met Glu Asp Tyr Val Arg Glu Asn Ile Asp Arg Asp Glu His Val Asn
 100 105 110

Ser Asn Arg Trp Trp Asn Ile Phe Pro Glu Asp Thr Asn Glu Tyr Lys
 115 120 125

Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asn Phe Glu Lys Gln Leu
 130 135 140

Glu Arg Ile Ser Ile Asp Thr Gly Val Gln Gly Gly Ala Leu Ser Val
 145 150 155 160

Glu His Leu Leu Leu Gly Ala Glu Tyr Ile Lys Arg Gly Ile Leu Thr
 165 170 175

Leu Tyr Asp Phe Lys Asn Ser Phe Leu Asn Lys Glu Ile Gln Phe
 180 185 190

<210> 72

<211> 191

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND2-RRR

<400> 72

Ile Val Lys Ser Ser Ile Glu Met Ser Lys Ala Asn Met Arg Asp Asn
 1 5 10 15

Leu Gln Met Leu Pro His Asp Tyr Ile Glu Leu Ile Glu Ile Ser Gln
 20 25 30

Asp Pro Tyr Gln Asn Arg Ile Phe Glu Met Lys Val Met Asp Leu Phe
 35 40 45

Ile Asn Glu Tyr Gly Phe Ser Gly Ser His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Met Tyr Ala His Gly Phe Gly Val Ile Val Asp Thr
 65 70 75 80

Lys Ala Tyr Lys Asp Gly Tyr Asn Leu Pro Ile Ser Gln Ala Arg Glu
 85 90 95

Met Glu Arg Tyr Val Arg Glu Asn Ile Asp Arg Asn Glu His Val Asn
 100 105 110

Ser Asn Arg Trp Trp Asn Ile Phe Pro Glu Asp Thr Asn Glu Tyr Lys
 115 120 125

Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asn Phe Glu Lys Gln Leu
 130 135 140

Glu Arg Ile Ser Arg Asp Thr Gly Val Gln Gly Gly Ala Leu Ser Val
 145 150 155 160

Glu His Leu Leu Leu Gly Ala Glu Tyr Ile Lys Arg Gly Ile Leu Thr
 165 170 175

Leu Tyr Asp Phe Lys Asn Ser Phe Leu Asn Lys Glu Ile Gln Phe
 180 185 190

<210> 73

<211> 195

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của FokI-S418P/K441E

<400> 73

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys
 1 5 10 15

Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg
20 25 30

Asn Pro Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe
35 40 45

Met Lys Val Tyr Gly Tyr Arg Gly Glu His Leu Gly Gly Ser Arg Lys
50 55 60

Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val
65 70 75 80

Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly
85 90 95

Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn
100 105 110

Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val
115 120 125

Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr
130 135 140

Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala
145 150 155 160

Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala
165 170 175

Gly Thr Leu Thr Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu
180 185 190

Ile Asn Phe
195

<210> 74

<211> 195

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND1-S418P/K441E

<400> 74

Leu Val Lys Gly Glu Met Glu Lys Lys Lys Ser Asp Leu Arg His Lys
 1 5 10 15

Leu Lys His Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Gln
 20 25 30

Asp Pro Lys Gln Asn Arg Leu Phe Glu Phe Lys Val Val Glu Phe Leu
 35 40 45

Lys Glu Val Tyr Asp Tyr Asn Gly Glu His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Leu Tyr Thr Asn Gly Leu Lys Thr Asp Tyr Gly Ile
 65 70 75 80

Ile Leu Asp Thr Lys Ala Tyr Lys Asp Gly Tyr Ser Leu Pro Ile Ser
 85 90 95

Gln Ala Asp Glu Met Gln Arg Tyr Val Asp Glu Asn Asn Asn Arg Asn
 100 105 110

Ala Ile Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Asn Ser Ile
 115 120 125

Leu Asp Phe Lys Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asp Tyr
 130 135 140

Lys Lys Gln Leu Ala Arg Val Ser Asn Leu Thr Lys Arg Lys Gly Ala
 145 150 155 160

Val Leu Ser Val Glu Gln Leu Leu Leu Gly Gly Glu Lys Ile Lys Asp
 165 170 175

Gly Ser Leu Thr Leu Glu Asp Val Gly Asp Lys Phe Asn Asn Asp Glu
 180 185 190

Ile Ile Phe
 195

<210> 75

<211> 191

<212> PRT

<213> Nhân tạo

<220>

<223> Trình tự axit amin của ND2-S441E

<400> 75

Ile Val Lys Ser Ser Ile Glu Met Ser Lys Ala Asn Met Arg Asp Asn
 1 5 10 15

Leu Gln Met Leu Pro His Asp Tyr Ile Glu Leu Ile Glu Ile Ser Gln
 20 25 30

Asp Pro Tyr Gln Asn Arg Ile Phe Glu Met Lys Val Met Asp Leu Phe
 35 40 45

Ile Asn Glu Tyr Gly Phe Ser Gly Glu His Leu Gly Gly Ser Arg Lys
 50 55 60

Pro Asp Gly Ala Met Tyr Ala His Gly Phe Gly Val Ile Val Asp Thr
 65 70 75 80

Lys Ala Tyr Lys Asp Gly Tyr Asn Leu Pro Ile Ser Gln Ala Asp Glu
 85 90 95

Met Glu Arg Tyr Val Arg Glu Asn Ile Asp Arg Asn Glu His Val Asn
 100 105 110

Ser Asn Arg Trp Trp Asn Ile Phe Pro Glu Asp Thr Asn Glu Tyr Lys
 115 120 125

Phe Leu Phe Val Ser Gly Phe Phe Lys Gly Asn Phe Glu Lys Gln Leu
 130 135 140

Glu Arg Ile Ser Ile Asp Thr Gly Val Gln Gly Gly Ala Leu Ser Val
 145 150 155 160

Glu His Leu Leu Leu Gly Ala Glu Tyr Ile Lys Arg Gly Ile Leu Thr
 165 170 175

Leu Tyr Asp Phe Lys Asn Ser Phe Leu Asn Lys Glu Ile Gln Phe
 180 185 190

<210> 76

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 76

cagccacat cttccactct gaagatggac aggt

34

<210> 77

<211> 35

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 77

cagccaccat cttccactct tgaagatgga caggt

35

<210> 78

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 78

cagccaccat cttccactca ctctgaagat ggacaggt

38

<210> 79

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 79

cagccaccat cttccactct gaagatggac aggt

34

<210> 80

<211> 35

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 80

cagccaccat cttccaactc tgaagatgga caggt

35

<210> 81

<211> 36

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 81

cagccaccat cttccactct ctgaagatgg acaggt

36

<210> 82

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 82

cagccaccat cttccactca ctctgaagat ggacaggt

38

<210> 83

<211> 32

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 83

cagccaccat cttcctctga agatggacag gt

32

<210> 84

<211> 25

<212> ADN

<213> Nhân tạo

<220>

<223> vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 84

cagcccactc tgaagatgga caggt

25

<210> 85

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 85

cagccaccat cttccactct gaagatggac aggt

34

<210> 86

<211> 35

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 86

cagccaccat cttccactct tgaagatgga caggt

35

<210> 87

<211> 36

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 87

cagccaccat cttccactct ctgaagatgg acaggt

36

<210> 88

<211> 32

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 88

cagccaccat cttcctctga agatggacag gt

32

<210> 89

<211> 30

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 89

cagccaccat ctttctgaag atggacaggt

30

<210> 90

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND3

<400> 90

cagccaccat cttccactct gaagatggac aggt

34

<210> 91

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ nhất của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND4

<400> 91

cagccaccat cttccactct gaagatggac aggt

34

<210> 92

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 92

accttaccat cttcaagaga gaagatggct taaa

34

<210> 93

<211> 36

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 93

accttaccat cttcaagaga gagaagatgg cttaaa

36

<210> 94

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 94

accttaccat cttcaagaga gagagaagat ggcttaaa

38

<210> 95

<211> 39

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 95

accttaccat cttcaagagg agagagaaga tggcttaaa

39

<210> 96

<211> 32

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 96

accttaccat cttcaagaga agatggctta aa

32

<210> 97

<211> 268

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng FokI

<400> 97

accttaccat cttcaagagc atgaacacat caccattcat taccaaaaat aaaaaatctt

60

actatgaaga gctcttgga atgggagtgt tagcctaattg agagaacaga cacttagcat

120

attctaggcc ttaggatcag tcaccaccac caaaaaacaa acaaacaaaa aacttaacac

180

atgttcatcc ccagttagtt taagtcaggg accaagtaag attcccattt tgcagtagaa

240

tggatatcact taggtaagat ggcttaaa

268

<210> 98

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 98

accttaccat cttcaagaga gaagatggct taaa

34

<210> 99

<211> 36

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 99

accttaccat cttcaagaga gagaagatgg cttaaa

36

<210> 100

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 100

accttaccat cttcaagaga gagagaagat ggcttaaa

38

<210> 101

<211> 32

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 101

accttaccat cttcaataga agatggctta aa

32

<210> 102

<211> 30

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND1

<400> 102

accttaccat cttcaagagg atggcttaaa

30

<210> 103

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 103

accttaccat cttcaagaga gaagatggct taaa

34

<210> 104

<211> 35

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 104

accttaccat cttcaaagag agaagatggc ttaaa

35

<210> 105

<211> 36

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 105

accttaccat cttcaagaga gagaagatgg cttaaa

36

<210> 106

<211> 38

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 106

accttaccat cttcaagaga gagagaagat ggcttaaa

38

<210> 107

<211> 105

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND2

<400> 107

accttaccat cttcaggtct tcctgggtgta gaatccctag tctttgggtgg tgtcatataa

60

gtcttttgtt gttgaatgca ttcttaagag agaagatggc ttaaa

105

<210> 108

<211> 34
<212> ADN
<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND3

<400> 108

accttaccat cttcaagaga gaagatggct taaa

34

<210> 109
<211> 36
<212> ADN
<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND3

<400> 109

accttaccat cttcaagaga gagaagatgg cttaaa

36

<210> 110
<211> 32
<212> ADN
<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND3

<400> 110

accttaccat cttcaagaga agatggctta aa

32

<210> 111
<211> 26
<212> ADN
<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND3

<400> 111

accttaccat cttcagttgg cttaaa

26

<210> 112

<211> 34

<212> ADN

<213> Nhân tạo

<220>

<223> Vị trí thứ hai của đích CHO-K1 ZFA36-ZFA36 được phân cắt bằng ND4

<400> 112

cagccaccat cttcaagaga gaagatggct taaa

34

<210> 113

<211> 30

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi xuôi để khuếch đại vector để thay thế ptTALE-ND,
ptTALEN-ngược-F

<400> 113

taaaaaatca gcctcgactg tgccttctag

30

<210> 114

<211> 22

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mồi ngược để khuếch đại vector để thay thế ptTALE-ND,
NC-ngược-R_2

<400> 114

ctgggatctg atcaattccg gc

22

<210> 115

<211> 48

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi xuôi để khuếch đại đoạn chèn để thay thế ptTALE-ND,
ND1-NC-F_v2

<400> 115

ttgatcagat cccagttagt aaaggagag atggagaaga aaaagagc

48

<210> 116

<211> 46

<212> ADN

<213> Nhân tạo

<220>

<223> đoạn mỗi ngược để khuếch đại đoạn chèn để thay thế ptTALE-ND,
ND1-ptTALEN-R

<400> 116

gaggctgatt ttttagaaga tgatctcatc gttgttgaac ttatcc

46

<210> 117

<211> 50

<212> ADN

<213> Cricetulus griseus

<400> 117

tgcccagaag actcccgccc atctcccaga aagcctcgac ttgcagatca

50

<210> 118

<211> 51

<212> ADN

<213> Cricetulus griseus

<400> 118

tgaaccaggc tatgacctag atttatatttg tatttcctaat cactatgtcg a

51

<210> 119

<211> 318

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Trình tự axit amin của ZF(ZFA36A)-ND1

<400> 119

Met	Gly	His	His	His	His	His	His	Met	Arg	Leu	Glu	Pro	Lys	Lys	Lys
1				5					10					15	

Arg	Lys	Val	Ala	Ala	Ala	Asp	Tyr	Lys	Asp	Asp	Asp	Asp	Lys	Ser	Arg
			20					25					30		

Pro	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser
			35				40					45			

Thr	Ser	Gly	His	Leu	Val	Arg	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys
			50				55					60			

Pro	Tyr	Lys	Cys	Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Thr	Ser	Gly	Asn
65					70				75					80	

Leu	Val	Arg	His	Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys
				85					90					95	

Pro	Glu	Cys	Gly	Lys	Ser	Phe	Ser	Gln	Ser	Ser	Asn	Leu	Val	Arg	His
			100					105					110		

Gln	Arg	Thr	His	Thr	Gly	Ala	Ala	Ala	Arg	Ala	Leu	Val	Lys	Gly	Glu
			115				120						125		

Met Glu Lys Lys Lys Ser Asp Leu Arg His Lys Leu Lys His Val Pro
 130 135 140

His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Gln Asp Ser Lys Gln Asn
 145 150 155 160

Arg Leu Phe Glu Phe Lys Val Val Glu Phe Leu Lys Glu Val Tyr Asp
 165 170 175

Tyr Asn Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Leu
 180 185 190

Tyr Thr Asn Gly Leu Lys Thr Asp Tyr Gly Ile Ile Leu Asp Thr Lys
 195 200 205

Ala Tyr Lys Asp Gly Tyr Ser Leu Pro Ile Ser Gln Ala Asp Glu Met
 210 215 220

Gln Arg Tyr Val Asp Glu Asn Asn Asn Arg Asn Ala Ile Ile Asn Pro
 225 230 235 240

Asn Glu Trp Trp Lys Val Tyr Pro Asn Ser Ile Leu Asp Phe Lys Phe
 245 250 255

Leu Phe Val Ser Gly Phe Phe Lys Gly Asp Tyr Lys Lys Gln Leu Ala
 260 265 270

Arg Val Ser Asn Leu Thr Lys Arg Lys Gly Ala Val Leu Ser Val Glu
 275 280 285

Gln Leu Leu Leu Gly Gly Glu Lys Ile Lys Asp Gly Ser Leu Thr Leu
 290 295 300

Glu Asp Val Gly Asp Lys Phe Asn Asn Asp Glu Ile Ile Phe
 305 310 315

<210> 120

<211> 318

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Trình tự axit amin của ZF(ZFL1)-ND1

<400> 120

Met Gly His His His His His His Met Arg Leu Glu Pro Lys Lys Lys

1

5

10

15

Arg Lys Val Ala Ala Ala Asp Tyr Lys Asp Asp Asp Asp Lys Ser Arg

20

25

30

Pro His Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser

35

40

45

Asp Arg Ser Asn Leu Thr Arg His Thr Arg Thr His Thr Gly Glu Lys

50

55

60

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg His His

65

70

75

80

Leu Thr Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys

85

90

95

Arg Ile Cys Met Arg Asn Phe Ser Gln Lys Ala Asn Leu Thr Arg His

100

105

110

Leu Arg Thr His Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Gly Glu

115

120

125

Met Glu Lys Lys Lys Ser Asp Leu Arg His Lys Leu Lys His Val Pro

130

135

140

His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Gln Asp Ser Lys Gln Asn

145

150

155

160

Arg Leu Phe Glu Phe Lys Val Val Glu Phe Leu Lys Glu Val Tyr Asp
 165 170 175

Tyr Asn Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Leu
 180 185 190

Tyr Thr Asn Gly Leu Lys Thr Asp Tyr Gly Ile Ile Leu Asp Thr Lys
 195 200 205

Ala Tyr Lys Asp Gly Tyr Ser Leu Pro Ile Ser Gln Ala Asp Glu Met
 210 215 220

Gln Arg Tyr Val Asp Glu Asn Asn Asn Arg Asn Ala Ile Ile Asn Pro
 225 230 235 240

Asn Glu Trp Trp Lys Val Tyr Pro Asn Ser Ile Leu Asp Phe Lys Phe
 245 250 255

Leu Phe Val Ser Gly Phe Phe Lys Gly Asp Tyr Lys Lys Gln Leu Ala
 260 265 270

Arg Val Ser Asn Leu Thr Lys Arg Lys Gly Ala Val Leu Ser Val Glu
 275 280 285

Gln Leu Leu Leu Gly Gly Glu Lys Ile Lys Asp Gly Ser Leu Thr Leu
 290 295 300

Glu Asp Val Gly Asp Lys Phe Asn Asn Asp Glu Ile Ile Phe
 305 310 315

<210> 121

<211> 318

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Trình tự axit amin của ZF(ZFA36A)-FokI

<400> 121

Met Gly His His His His His His Met Arg Leu Glu Pro Lys Lys Lys
 1 5 10 15

Arg Lys Val Ala Ala Ala Asp Tyr Lys Asp Asp Asp Asp Lys Ser Arg
 20 25 30

Pro Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser
 35 40 45

Thr Ser Gly His Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys
 50 55 60

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Ser Gly Asn
 65 70 75 80

Leu Val Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
 85 90 95

Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Val Arg His
 100 105 110

Gln Arg Thr His Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Ser Glu
 115 120 125

Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro
 130 135 140

His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp
 145 150 155 160

Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly
 165 170 175

Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile
 180 185 190

Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys
 195 200 205

Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met
 210 215 220

Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro
 225 230 235 240

Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe
 245 250 255

Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr
 260 265 270

Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu
 275 280 285

Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu
 290 295 300

Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
 305 310 315

<210> 122

<211> 318

<212> PRT

<213> Trình tự nhân tạo

<220>

<223> Trình tự axit amin của ZF(ZFL1)-FokI

<400> 122

Met Gly His His His His His His Met Arg Leu Glu Pro Lys Lys Lys
 1 5 10 15
 Arg Lys Val Ala Ala Ala Asp Tyr Lys Asp Asp Asp Asp Lys Ser Arg
 20 25 30

Pro His Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
 35 40 45

Asp Arg Ser Asn Leu Thr Arg His Thr Arg Thr His Thr Gly Glu Lys
 50 55 60

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg His His
 65 70 75 80

Leu Thr Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
 85 90 95

Arg Ile Cys Met Arg Asn Phe Ser Gln Lys Ala Asn Leu Thr Arg His
 100 105 110

Leu Arg Thr His Thr Gly Ala Ala Ala Arg Ala Leu Val Lys Ser Glu
 115 120 125

Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro
 130 135 140

His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp
 145 150 155 160

Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly
 165 170 175

Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile
 180 185 190

Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys
 195 200 205

Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met
 210 215 220

Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro
225 230 235 240

Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe
245 250 255

Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr
260 265 270

Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu
275 280 285

Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu
290 295 300

Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
305 310 315