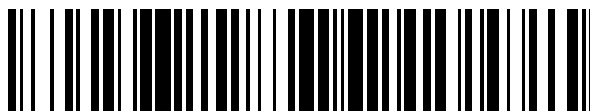


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 735 773**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/11 (2006.01)

C12N 15/113 (2010.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.12.2016** **PCT/US2016/064860**

87 Fecha y número de publicación internacional: **08.06.2017** **WO17096328**

96 Fecha de presentación y número de la solicitud europea: **02.12.2016** **E 16822804 (7)**

97 Fecha y número de publicación de la concesión europea: **01.05.2019** **EP 3371310**

54 Título: **Ácidos nucleicos manipulados dirigidos a ácidos nucleicos**

30 Prioridad:

04.12.2015 US 201562263232 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

20.12.2019

73 Titular/es:

CARIBOU BIOSCIENCES, INC. (100.0%)
2929 7th Street, Suite 105
Berkeley, CA 94710, US

72 Inventor/es:

DONOHUE, PAUL DANIEL y
MAY, ANDREW PAUL

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 735 773 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Ácidos nucleicos manipulados dirigidos a ácidos nucleicos

Campo técnico

5 La presente divulgación se refiere en general a secuencias de polinucleótidos manipuladas que forman andamios y complejos de nucleoproteínas que comprenden tales andamios y proteínas de unión a ácidos nucleicos. Se describen las secuencias de ácidos nucleicos que codifican los componentes de polinucleótido de andamio, así como los casetes de expresión, vectores y células que comprenden los componentes de polinucleótido. La descripción también se refiere a procedimientos para hacer y usar las secuencias de ácidos nucleicos manipuladas que forman andamios y los complejos de nucleoproteínas de la presente invención.

10 Antecedentes

Las repeticiones palindrómicas cortas agrupadas en intervalos regulares (CRISPR) y las proteínas asociadas a CRISPR (Cas) constituyen el sistema CRISPR-Cas. El sistema CRISPR-Cas proporciona inmunidad adaptativa contra el ADN extraño en bacterias (véase, por ejemplo, Barrangou, R., et al., Science 315: 1709-1712 (2007); Makarova, K.S., et al., Nature Reviews Microbiology 9: 467-477 (2011); Garneau, J.E., et al., Nature 468: 67-71 (2010); Sapranasauskas, R., et al., Nucleic Acids Research 39: 9275-9282 (2011)).

15 Los sistemas CRISPR-Cas se han reclasificado recientemente en dos clases, que comprenden cinco tipos y dieciséis subtipos (véase Makarova, K., et al., Nature Reviews Microbiology 13: 1-15 (2015)). Esta clasificación se basa en la identificación de todos los genes Cas en un locus CRISPR-Cas y en la determinación de los genes distintivos en cada locus CRISPR-Cas, colocando finalmente los sistemas CRISPR-Cas en la Clase 1 o en la Clase 2 en función de los genes que codifican el módulo efector, es decir, las proteínas involucradas en la etapa de interferencia. Recientemente, se ha identificado un sexto sistema CRISPR-Cas (Tipo VI) (véase Abudayyeh O., et al., Science 353(6299): aaf5573 (2016)). Ciertas bacterias poseen más de un tipo de sistema CRISPR-Cas.

20 Los sistemas de Clase 1 tienen un complejo multisubunidad ARN-cr-efector, mientras que los sistemas de Clase 2 tienen una sola proteína, como Cas9, Cpf1, C2c1, C2c2, C2c3, o un complejo ARN-cr-efector. Los sistemas de Clase 1 comprenden los sistemas Tipo I, Tipo III y Tipo IV. Los sistemas de Clase 2 comprenden los sistemas Tipo II, Tipo V y Tipo VI.

25 Los sistemas de tipo II tienen genes cas1, cas2 y cas9. El gen cas9 codifica una proteína de varios dominios que combina las funciones del complejo ARN-cr-efector con la escisión de la secuencia diana del ADN. Los sistemas de Tipo II se dividen en tres subtipos, los subtipos II-A, II-B y II-C. El subtipo II-A contiene un gen adicional, csn2. Ejemplos de organismos con sistemas de subtipo II-A incluyen, entre otros, *Streptococcus pyogenes*, *Streptococcus thermophilus* y *Staphylococcus aureus*. El subtipo II-B carece de la proteína csn2, pero tiene la proteína cas4. Un ejemplo de un organismo con un sistema de subtipo II-B es la *Legionella pneumophila*. El subtipo II-C es el sistema de Tipo II más común que se encuentra en las bacterias y tiene solo tres proteínas, Cas1, Cas2 y Cas9. Un ejemplo de un organismo con un sistema de subtipo II-C es *Neisseria lactamica*.

30 Los sistemas Tipo V tienen un gen cpf1 y genes cas1 y cas2 (véase Zetsche, B., et al., Cell 163: 1-13 (2015)). El gen cpf1 codifica una proteína, Cpf1, que tiene un dominio de nucleasa similar a RuvC que es homólogo al dominio respectivo de Cas9, pero carece del dominio de nucleasa HNH que está presente en las proteínas Cas9. Se han identificado sistemas de Tipo V en varias bacterias, entre las que se incluyen, entre otras, *Parcubacteria bacterium*, *Lachnospiraceae bacterium*, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium*, *Acidaminococcus spp.*, *Porphyromonas macacae*, *Porphyromonas crevioricanis*, *Prevotella disiens*, *Moraxella bovoculi*, *Smithella spp.*, *Leptospira inadai*, *Franciscella tularensis*, *Franciscella novicida*, *Candidatus methanoplasma termitum*, y *Eubacterium eligens*. Recientemente se ha demostrado que Cpf1 también tiene actividad de ARNasa y es responsable del procesamiento de pre-ARNc (véase Fonfara, I., et al., Nature 532(7600): 517-521 (2016)).

35 En los sistemas de Clase 2, el ARN-cr está asociado con una proteína única y logra la interferencia al combinar la actividad de la nucleasa con los dominios de unión a RNA y la formación de pares de bases entre el ARN-cr y una secuencia diana de ácidos nucleicos.

40 En los sistemas de Tipo II, la unión de la secuencia diana de ácidos nucleicos implica Cas9 y el ARN-cr, al igual que la escisión de la secuencia diana de ácidos nucleicos. En los sistemas de Tipo II, el dominio de nucleasa de tipo RuvC (pliegue de ARNasa H) y el dominio de nucleasa HNH (de tipo McrA) de Cas9 rompen una de las cadenas de la secuencia diana de ácidos nucleicos de doble cadena. La actividad de escisión de Cas9 de los sistemas de Tipo II también requiere la hibridación de ARN-cr a un traARN-cr para formar un dúplex que facilita la unión de la secuencia diana de ácidos nucleicos y ARN-cr por la proteína Cas9.

45 En los sistemas de Tipo V, la unión de la secuencia diana de ácidos nucleicos implica Cpf1 y el ARN-cr, al igual que la escisión de la secuencia diana de ácidos nucleicos. En los sistemas de Tipo V, el dominio de nucleasa de tipo RuvC de Cpf1 escinde una cadena de la secuencia diana de ácidos nucleicos bicatenario, y un dominio de nucleasa putativo escinde la otra cadena de la secuencia diana de ácidos nucleicos bicatenarios en una configuración

escalonada, que produce 5' sobresale, que está en contraste con los extremos romos generados por la división de Cas9.

La actividad de escisión de CpfI de los sistemas de Tipo V no requiere la hibridación del ARNcr con el traARN-cr para formar un dúplex, sino que el ARNcr de los sistemas de Tipo V utiliza un solo ARNcr que tiene una estructura de tallo-bucle que forma un dúplex interno. CpfI se une al ARNcr en una secuencia y estructura específica que reconoce el bucle del tallo y las secuencias adyacentes al bucle del tallo, especialmente los nucleótidos 5' de las secuencias espaciadoras que hibridan con la secuencia diana del ácido nucleico. Esta estructura de tallo-bucle está típicamente en el rango de 15 a 19 nucleótidos de longitud. Las sustituciones que interrumpen este dúplex tallo-bucle anulan la actividad de escisión, mientras que otras sustituciones que no interrumpen el dúplex tallo-bucle no anulan la actividad de escisión. Los nucleótidos 5' del bucle del tallo adoptan una estructura de pseudonudo que estabiliza aún más la estructura del bucle del tallo con el emparejamiento de bases de Watson-Crick no canónico, la interacción triple y el emparejamiento de bases de Hoogsteen inverso (véase Yamano, T., et al. 165(4): 949-962 (2016)). En los sistemas de Tipo V, el ARNcr forma una estructura tallo-bucle en el extremo 5', y la secuencia en el extremo 3' es complementaria a una secuencia en una secuencia diana de ácidos nucleicos.

Otras proteínas asociadas con el ARNcr de Tipo V y la unión y escisión de la secuencia diana del ácido nucleico incluyen al candidato 1 de Clase 2 (C2c1) y al candidato 3 de Clase 2 (C2c3). Las proteínas C2c1 y C2c3 tienen una longitud similar a las proteínas Cas9 y CpfI, que van desde aproximadamente 1.100 aminoácidos hasta aproximadamente 1.500 aminoácidos. Las proteínas C2c1 y C2c3 también contienen dominios de nucleasa tipo RuvC y tienen una arquitectura similar a CpfI. Las proteínas C2c1 son similares a las proteínas Cas9 que requieren un ARNcr y un traARN-cr para la unión y escisión de la secuencia diana de los ácidos nucleicos, pero tienen una temperatura de escisión óptima de 50°C. Las proteínas C2c1 se dirigen a un motivo protoespaciador rico en AT adyacente (PAM), similar al PAM de CpfI, que es 5' de la secuencia diana de ácidos nucleicos (véase, por ejemplo, Shmakov, S., et al., *Molecular Cell* 60(3): 385-397 (2015)).

El candidato de clase 2 2 (C2c2) no comparte la similitud de secuencia con otras proteínas efectoras CRISPR y fue identificado recientemente como un sistema de Tipo VI (véase Abudayyeh, O., et al., *Science* 353(6299): aaf5573 (2016)). Las proteínas C2c2 tienen dos dominios HEPN y demuestran una actividad de escisión del ARN monocatenario. Las proteínas C2c2 son similares a las proteínas CpfI que requieren un ARNcr para la unión y escisión de la secuencia diana de los ácidos nucleicos, aunque no requieren el traARN-41cr. Además, al igual que CpfI, el ARNcr para las proteínas C2c2 forma una horquilla estable o estructura de bucle-tallo, que ayuda en asociación con la proteína C2c2. Los sistemas de tipo VI tienen una única endonucleasa de ARN polipeptídica que utiliza un único ARNcr para dirigir la escisión específica del sitio. Además, después de la hibridación con el ARN objetivo complementario del espaciador, C2c2 se convierte en una endonucleasa de ARN promiscua que muestra una actividad endonucleasa no específica hacia cualquier ARN monocatenario en una secuencia independiente (véase East-Seletsky, A., et al., *Nature* 538 (7624): 270-273 (2016)).

Con respecto a los sistemas de Clase 2 Tipo II CRISPR-Cas, se conoce un gran número de ortólogos Cas9 en la técnica, así como sus componentes polinucleótidos asociados (traARN-cr y ARNcr) (véase, por ejemplo, Fonfara, I., et al., *Nucleic Acids Research* 42(4): 2577-2590 (2014), incluidos todos los datos suplementarios; Chylinski K., et al., *Nucleic Acids Research* 42(10): 6091-6105 (2014), incluidos todos los datos suplementarios). Además, las proteínas sintéticas similares a Cas9 son conocidas en la técnica (véase la Solicitud de Patente publicada de los Estados Unidos No. 2014-0315985, publicada el 23 de octubre de 2014).

Cas9 es un ejemplo de proteína Cas CRISPR de tipo II. Cas9 es una endonucleasa que puede ser programada por el traARN-cr/ARNcr para escindir, de una manera específica al sitio, una secuencia diana de ADN que utiliza dos dominios de endonucleasa distintos (HNH y RuvC/RNase H-like como dominios) (véase la Patente de los Estados Unidos 2014-0068797, publicado el 6 de marzo de 2014; véase también Jinek, M., et al., *Science* 337: 816-821 (2012)).

Típicamente, cada sistema CRISPR-Cas9 de tipo salvaje incluye un ARNcr y un traARN-cr. El ARNcr tiene una región de complementariedad con una posible secuencia diana de ADN y una segunda región que forma enlaces de hidrógeno de pares de bases con el traARN-cr para formar una estructura secundaria, típicamente para formar al menos una estructura de tallo. La región de complementariedad con la secuencia diana de ADN es el espaciador. El traARN-cr y el ARNcr interactúan a través de varios enlaces de hidrógeno de pares de bases para formar estructuras de ARN secundarias. La formación de complejos entre el traARN-cr/ARNcr y la proteína Cas9 da como resultado un cambio conformacional de la proteína Cas9 que facilita la unión al ADN, las actividades de endonucleasa de la proteína Cas9 y la escisión del ADN guiada por el ARNcr guiado por la endonucleasa Cas9. Para un complejo de proteína Cas9/traARN-cr/ARNcr para escindir una secuencia diana de ADN de doble cadena, la secuencia diana de ADN es adyacente a una PAM cognada. Al diseñar un ARNcr para que tenga una secuencia espaciadora apropiada, el complejo puede dirigirse a escindirse en un lugar de interés, por ejemplo, un lugar en el que se desea la modificación de la secuencia. Los documentos WO 2014/093712 A1 y WO 2014/065596 A1 describen composiciones que comprenden un ARN de guía dual diseñado que consiste en un ARNcr y un ARNtrac que es capaz de formar complejos con una proteína Cas9 para la escisión específica del objetivo.

Se conocen en la técnica una variedad de secuencias de ARNcr y traARN-cr del sistema CRISPR-Cas de Tipo II, así como las estructuras secundarias predichas (véase, por ejemplo, Ran, F.A., et al., *Nature* 520(7546): 186-191 (2015), incluidos todos los Datos Suplementarios, en particular los Datos Ampliados de la Figura 1; Fonfara, I., et al., *Nucleic Acids Research* 42(4): 2577-2590 (2014), incluidos todos los Datos Suplementarios, en particular la Figura Suplementaria S11). Las estructuras secundarias de traARN-cr predichas se basaron en el modelo de plegamiento de ARN de generación de constricción (Zuker, M., *Nucleic Acids Research* 31: 3406-3415 (2003). Las estructuras secundarias de dúplex de ARN se predijeron utilizando RNAcofold del paquete Vienna RNA (Bernhart, S.H., et al., *Algorithms for Molecular Biology* 1(1): 3 (2006); Hofacker, I.L., et al., *Journal of Molecular Biology* 319: 1059-1066 (2002)) y RNAhybrid (bibiserv.techfak.uni-bielefeld.de/rnahybrid/). Las predicciones de la estructura se visualizaron utilizando VARNA (Darty, K., et al., *Bioinformatics* 25: 1974-1975 (2009)). Fonfara, I. et al., muestran que el complejo ARNcr/traARN-cr para *Campylobacter jejuni* no tiene la región de protuberancia, sin embargo, el complejo retiene una estructura del tallo ubicada en 3' del espaciador que se sigue en la dirección 3' con otra estructura del tallo.

El espaciador de los sistemas CRISPR-Cas de Clase 2 puede hibridarse con una secuencia diana de ácidos nucleicos que está ubicada en 5' o 3' de una PAM, dependiendo de la proteína Cas que se utilice. Un PAM puede variar dependiendo del polipéptido Cas a usar. Por ejemplo, si se usa Cas9 de *S. pyogenes*, la PAM puede ser una secuencia en la secuencia diana de ácidos nucleicos que comprende la secuencia 5'-NRR-3', en el que R puede ser A o G, N es cualquier nucleótido, y N es inmediatamente 3' de la secuencia diana de ácidos nucleicos dirigida por la secuencia de unión a diana de ácidos nucleicos. Una proteína Cas puede modificarse de tal manera que una PAM puede ser diferente en comparación con una PAM para una proteína Cas no manipulada. Si, por ejemplo, se usa Cas9 de *S. pyogenes*, la proteína Cas9 puede modificarse de tal manera que la PAM ya no comprende la secuencia 5'-NRR-3', sino que comprende la secuencia 5'-NNR-3', en el que R puede ser A o G, N es cualquier nucleótido, y N es inmediatamente 3' de la secuencia diana de ácidos nucleicos dirigida por la secuencia diana de ácidos nucleicos.

Otras proteínas Cas reconocen otras PAM, y un experto en la técnica es capaz de determinar la PAM para cualquier proteína Cas en particular. Por ejemplo, Cpf1 tiene un sitio PAM rico en timina que se dirige, por ejemplo, a una secuencia TTTN (véase Fagerlund, R., et al., *Genome Biology* 16: 251 (2015)).

La endonucleasa Cas9 guiada por ARN se ha utilizado ampliamente para la edición del genoma programable en una variedad de organismos y sistemas modelo (véase, por ejemplo, Jinek M., et al., *Science* 337: 816-821 (2012); Jinek M., et al., *eLife* 2: e00471. doi: 10.7554/eLife.00471 (2013); Solicitud de Patente Publicada de los Estados Unidos No. 2014-0068797, publicada el 6 de marzo de 2014).

La modificación del genoma incluye alterar el genoma eliminando, insertando, mutando o sustituyendo secuencias específicas de ácidos nucleicos. La alteración puede ser específica de gen o ubicación. La ingeniería genómica puede usar nucleasas dirigidas al sitio, como las proteínas Cas y sus polinucleótidos afines, para cortar el ADN, generando así un sitio para la alteración. En ciertos casos, la división puede introducir una ruptura de doble cadena (DSB) en la secuencia diana de ADN. Los DSB pueden repararse, por ejemplo, mediante unión final no homóloga (NHEJ), unión final mediada por microhomología (MMEJ) o reparación dirigida por homología (HDR). La HDR se basa en la presencia de una plantilla para su reparación. En algunos ejemplos manipulado genómica, un polinucleótido donante o una porción del mismo puede insertarse en la ruptura.

Sumario de la invención

La presente invención se refiere en general a una composición de polinucleótido de ácidos nucleicos que comprende un complejo de polinucleótido que forma un andamio que es capaz de unirse a una proteína de unión a ácidos nucleicos. Típicamente, una composición de polinucleótido NASC es un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio que comprende: un elemento 1 de repetición, un elemento de repetición 2, un elemento 1 de unión a proteína de unión a ácidos nucleicos, un elemento de unión de proteína de unión a ácidos nucleicos 2, un elemento espaciador 1 (por ejemplo, que comprende una secuencia de unión a diana de ácidos nucleicos, y un elemento espaciador 2 (por ejemplo, que comprende una secuencia 2 de unión a diana de ácidos nucleicos). La composición del polinucleótido NASC es capaz de asociarse con una proteína de unión a ácidos nucleicos.

La invención está definida por las reivindicaciones adjuntas.

En un aspecto, la presente invención se refiere a una composición de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio ("NASC") que comprende un primer ácido nucleico manipulado genéticamente "NASC-PC1" y un segundo componente de ácidos nucleicos manipulado ("NASC-PC2"). El NASC-P1 comprende, en una dirección 5' a 3', un elemento espaciador 1 que comprende una secuencia 1 de unión a diana de ácidos nucleicos, un elemento de repetición 1 que comprende una secuencia 1 de ácidos nucleicos repetida y una proteína de unión a ácidos nucleicos elemento de unión 1, en el que el elemento espaciador 1 está conectado covalentemente con el elemento de repetición 1, y el elemento de repetición 1 está conectado covalentemente con el elemento de unión a la proteína de unión a ácidos nucleicos 1 que comprende una secuencia de unión a la proteína de unión a ácidos nucleicos bicatenarios 1. El segundo componente de ácidos nucleicos manipulado genéticamente ("NASC-PC2") comprende, en una dirección 5' a 3', un elemento espaciador 2 que comprende una secuencia 2 de unión a diana de ácidos nucleicos, un elemento de repetición 2 que comprende una secuencia 1C de ácidos

nucleicos repetida y una unión a proteína de unión a ácidos nucleicos el elemento 2 que comprende una secuencia 2 de unión a la proteína de unión a ácidos nucleicos bicatenario, en el que el elemento espaciador 2 está conectado covalentemente con el elemento de repetición 2, y el elemento de repetición 2 está conectado covalentemente con el elemento 2 de unión a la proteína de unión a ácidos nucleicos. La secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida están conectadas a través de pares de bases con enlaces de hidrógeno y la conexión forma la composición NASC. La composición NASC es capaz de unirse a una primera proteína CRISPR-Cas9 Clase 2 Tipo II y a una segunda clase de CRISPR-Cas9 Clase 2 Tipo II.

En algunas realizaciones, el elemento espaciador 1 y el elemento espaciador 2 comprenden secuencias de ácidos nucleicos adicionales. Por ejemplo, el elemento espaciador 1 puede comprender además una secuencia de ácidos nucleicos del elemento enlazador 3' de la secuencia de unión a la diana del ácido nucleico 1 y 5' del elemento de repetición 1. El elemento espaciador 2 puede comprender además una secuencia 3' de ácidos nucleicos del elemento enlazador de la secuencia 2 de unión a ácidos nucleicos diana y 5' del elemento de repetición 1.

En realizaciones adicionales, la secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida comprenden secuencias adicionales como sigue. El elemento de repetición 1 comprende, además, en una dirección de 5' a 3', una secuencia 1b de ácidos nucleicos repetida, una secuencia 1-2 de ácidos nucleicos de elemento enlazador, y una secuencia 1a de ácidos nucleicos repetida. El elemento de repetición 1C comprende, además, en una dirección 5' a 3', una secuencia 1aC de ácidos nucleicos repetida, una secuencia de ácidos nucleicos de elemento enlazador 2-2, y una secuencia 1bC de ácidos nucleicos repetida. La secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida están conectadas a través de pares de bases con enlaces de hidrógeno, y la secuencia 1a de ácidos nucleicos repetida y la secuencia de ácidos nucleicos repetida 1aC están conectadas a través de pares de bases con enlaces de hidrógeno.

En realizaciones adicionales, la secuencia repetida de ácidos nucleicos 1b y la secuencia repetida 1a comprenden componentes adicionales como sigue. La secuencia 1b de ácidos nucleicos repetida puede comprender, además, en una dirección 5' a 3', una secuencia 1b2 de ácidos nucleicos repetida, una secuencia 1b1 de ácidos nucleicos de protuberancia y una secuencia 1b1 de ácidos nucleicos repetida. La secuencia 1a de ácidos nucleicos repetida puede comprender, además, en una dirección 5' a 3', una secuencia 1a2 de ácidos nucleicos repetida, una secuencia 1a1 de ácidos nucleicos de protuberancia y una secuencia 1a1 de ácidos nucleicos repetida. La secuencia 1aC de ácidos nucleicos repetida puede comprender, además, en una dirección 5' a 3', una secuencia 1a1C de ácidos nucleicos repetida, una secuencia 2a2 de ácidos nucleicos de protuberancia y una secuencia 1a2C de ácidos nucleicos repetida. La secuencia 1bC de ácidos nucleicos repetida puede comprender, además, en una dirección 5' a 3', una secuencia 1b1C de ácidos nucleicos repetida, una secuencia 2b2 de ácidos nucleicos de protuberancia y una secuencia 1b2C de ácidos nucleicos repetida. La secuencia 1a1 de ácidos nucleicos repetida y la secuencia 1a1C de ácidos nucleicos repetida están conectadas a través de pares de bases enlazadas con hidrógeno, la secuencia 1a2 de ácidos nucleicos repetida y la secuencia 1a2C de ácidos nucleicos repetida están conectadas a través de pares de bases enlazadas a hidrógeno, la secuencia 1b1 de ácidos nucleicos repetida y la secuencia 1b1C de ácidos nucleicos repetida está conectada a través de pares de bases unidas a hidrógeno, y la secuencia 1b2 de ácidos nucleicos repetida y la secuencia 1b2C de ácidos nucleicos repetida están conectadas a través de pares de bases unidas a hidrógeno.

En realizaciones adicionales, la secuencia 1-2 de ácidos nucleicos del elemento enlazador y la secuencia de nucleótido del elemento enlazador 2-2 comprenden secuencias de ácidos nucleicos añadidas. El elemento enlazador 1-1 puede comprender, además, en una dirección de 5' a 3', una secuencia de ácidos nucleicos de elemento enlazador 1-2-2, una secuencia 1-2a de ácidos nucleicos repetida, y una secuencia de ácidos nucleicos de elemento enlazador 1-2-1. La secuencia 2-2 de ácidos nucleicos del elemento enlazador puede comprender, además, en una dirección 5' a 3', una secuencia 2-2 de ácidos nucleicos del elemento enlazador-1, una secuencia 1-2aC de ácidos nucleicos repetida y una secuencia 2-2-2 de ácidos nucleicos del elemento enlazador. La secuencia 1-2a de ácidos nucleicos repetida y la secuencia 1-2aC de ácidos nucleicos repetida están conectadas a través de pares de bases con enlaces de hidrógeno y forman una región de ácidos nucleicos bicatenarios 1-2. En algunas realizaciones, la región de ácidos nucleicos bicatenarios 1-2 comprende además un sitio de unión a la proteína efectora 1. La secuencia 1-2a de ácidos nucleicos repetida comprende además una secuencia 1-2a de ácidos nucleicos del sitio de unión a la proteína efectora. La repetida secuencia 2 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora 1-2aC. El sitio de unión del efector se forma mediante un enlace de par de bases a través de hidrógeno entre la secuencia 1-2a de ácidos nucleicos del sitio de unión de la proteína efectora y la secuencia 1-2aC de ácidos nucleicos del sitio de unión de la proteína efectora. El sitio de unión a la proteína Csy4 es un ejemplo de un sitio de unión a la proteína efectora. Una proteína Csy4 o una proteína Csy4 inactiva enzimáticamente son capaces de unirse al sitio de unión del efector.

En realizaciones adicionales, la secuencia 1C de ácidos nucleicos repetida comprende además una etiqueta de afinidad 1 y la secuencia de ácidos nucleicos repetida comprende además una etiqueta de afinidad 2, y la etiqueta de afinidad 1 está conectada con la etiqueta de afinidad 2.

Las composiciones de NASC pueden comprender, por ejemplo, ARN, ADN o ARN y ADN. En algunas realizaciones, NASC-PC1, NASC-PC2 o NASC-PC1 y NASC-PC2 comprenden ARN, ADN o ARN y ADN.

En otro aspecto, la presente invención incluye una composición de ácidos nucleicos/proteína que comprende una composición NASC y una primera proteína CRISPR-Cas9 Clase 2 Tipo II y una segunda proteína CRISPR-Cas9 Clase 2 Tipo II por ejemplo, la primera Clase 2 Tipo II CRISPR la proteína Cas9 es la misma que la segunda proteína CRISPR-Cas9 Clase 2 Tipo II, y la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II se seleccionan del grupo que consiste en una proteína Cas9 de *S. pyogenes*, una proteína *S. thermophilus* Cas9, una proteína *S. aureus* Cas9 y una proteína *C. jejuni* Cas9. En otras realizaciones, la primera proteína CRISPR-Cas9 Clase 2 Tipo II es diferente de la segunda proteína CRISPR-Cas9 Clase 2 Tipo II, y la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II se seleccionan del grupo que consiste en una proteína Cas9 de *S. pyogenes*, una proteína Cas9 de *S. thermophilus*, una proteína Cas9 de *S. aureus* y una proteína Cas9 de *C. jejuni*. Además, la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda CRISPR-Cas9 Clase 2 Tipo II pueden seleccionarse del grupo que consiste en proteína Cas9/proteína Cas9, proteína Cas9/proteína dCas9, proteína dCas9/proteína Cas9 y Proteína dCas9/proteína dCas9, respectivamente.

En un aspecto adicional, la presente invención se refiere a kits que comprenden uno o más componentes de una composición NASC. En algunas realizaciones, la composición NASC comprende una NASC-PC1 y una NASC-PC2, o una o más secuencias de ácidos nucleicos que codifican el NASC-PC1 y el NASC-PC2, y un tampón. Los kits pueden comprender además una o más proteínas de Clase 2 CRISPR-Cas9 Tipo II o una o más secuencias de ácidos nucleicos que codifican la una o más proteínas CRISPR-Cas9 Clase 2 Tipo II. En realizaciones adicionales, un kit puede comprender complejos de nucleoproteínas que comprenden una composición NASC y una o más proteínas CRISPR-Cas9 Clase 2 Tipo II.

En un aspecto adicional, se describe un vector de expresión que comprende una o más secuencias de ácidos nucleicos que codifican uno o más componentes de una composición NASC.

En otro aspecto más, se describe una célula recombinante que comprende una o más secuencias de ácidos nucleicos que codifican uno o más componentes de una composición NASC.

Además, se describen procedimientos para usar la composición NASC, como se describe en el presente documento. Un procedimiento es un procedimiento de unión de ADN. El procedimiento comprende poner en contacto una primera secuencia diana de ADN en un polinucleótido de ADN y una segunda secuencia diana de ADN en el polinucleótido de ADN con una composición de ácidos nucleicos/proteína que comprende la composición NASC y una proteína de unión a ácidos nucleicos (por ejemplo, una proteína Cas9 y/o Proteína Cpf1), facilitando así la unión de la composición de ácidos nucleicos/proteína a la primera secuencia diana de ADN en el ADN y la segunda secuencia diana de ADN en el ADN. Un elemento espaciador NASC-PC1 de la composición NASC puede ser complementario a la primera secuencia diana de ADN, y el espaciador NASC-PC2 de la composición NASC puede ser complementario a la segunda secuencia diana de ADN.

Otro procedimiento es un procedimiento de corte de ADN. El procedimiento comprende poner en contacto una primera secuencia diana de ADN en el polinucleótido de ADN y una segunda secuencia diana de ADN en el polinucleótido de ADN con una composición de ácidos nucleicos/proteína que comprende una composición NASC y una proteína de unión a ácidos nucleicos (por ejemplo, una proteína Cas9, y/o una proteína Cpf1), facilitando así la unión de la composición de ácidos nucleicos/proteína a la primera secuencia diana de ADN y la segunda secuencia diana de ADN. La unión da como resultado el corte de la primera secuencia diana de ADN y la segunda secuencia diana de ADN. Un elemento espaciador NASC-PC1 de la composición NASC puede ser complementario a la primera secuencia diana de ADN, y el espaciador NASC-PC2 de la composición NASC puede ser complementario a la segunda secuencia diana de ADN.

Breve descripción de las Figuras

Las Figuras que no pertenecen a la invención son solo para fines ilustrativos.

Las Figuras no están representadas proporcionalmente, ni las Figuras están a escala. Las ubicaciones de los indicadores son aproximadas.

La **Figura 1A, Figura 1B, Figura 1C**, y Figura 1D presenta ejemplos de ARN de guía asociados a CRISPR de clase 2 de Tipo II de guía dual.

La **Figura 2A, Figura 2B**, y **Figura 2C** presentan ejemplos de ARN de guía asociados a CRISPR de clase 2 de Tipo II de guía única.

La **Figura 3A y Figura 3B** presentan ejemplos de ARN de guía de ARNc de clase 2 de Tipo V.

La **Figura 4A, Figura 4B, Figura 4C, Figura 4D, Figura 4E, Figura 4F, Figura 4G, Figura 4H, Figura 4I, a Figura 4J, Figura 4K, Figura 4L, Figura 4M, Figura 4N, y Figura 4O** (esta última Figura es la Figura 4 "O" y no la Figura 4 "cero") ilustran ejemplos de disposiciones genéricas de composiciones de polinucleótidos de andamio de ácidos nucleicos manipuladas. Las secuencias ilustradas no se representan en una orientación de 5' a 3' o de 3' a 5' y no tienen polaridad.

La **Figura 5A, Figura 5B, Figura 5C, Figura 5D, Figura 5E, Figura 5F, Figura 5G, Figura 5H, y Figura 5I** ilustran ejemplos y elementos de composiciones de polinucleótidos de andamio de ácidos nucleicos.

La **Figura 6A, Figura 6B, Figura 6C, Figura 6D, Figura 6E, Figura 6F, Figura 6G, Figura 6H, Figura 6I, Figura 6J, Figura 6K, Figura 6L, y Figura 6M** ilustran ejemplos y elementos de composiciones de polinucleótidos de andamio de ácidos nucleicos de la presente invención.

La **Figura 7A, Figura 7B, Figura 7C, Figura 7D, Figura 7E, Figura 7F, Figura 7G, Figura 7H, y Figura 7I** ilustran ejemplos y elementos de composiciones de polinucleótidos de andamio de ácidos nucleicos concatenado.

La **Figura 8A, Figura 8B, Figura 8C, Figura 8D, Figura 8E, Figura 8F, Figura 8G, Figura 8H, Figura 8I, Figura 8J, Figura 8K, Figura 8L, Figura 8M, y Figura 8N** ilustra ejemplos y elementos de composiciones de polinucleótidos de andamio de ácidos nucleicos de nexo dividido concatenado manipulado.

La **Figura 9A y Figura 9B** ilustra ejemplos y elementos de composiciones de polinucleótidos de andamio de ácidos nucleicos.

La **Figura 10** ilustra un ejemplo y los elementos de una composición polinucleotídica de andamio de ácidos nucleicos manipulada.

La **Figura 11** ilustra un complejo de nucleoproteínas que comprende una composición de polinucleótido de andamio de ácidos nucleicos manipulada mediante ingeniería genética, la formación de un complejo de nucleoproteína y el complejo de nucleoproteínas de unión a dos secuencias diana de ácidos nucleicos.

La **Figura 12** ilustra un complejo de nucleoproteínas que comprende una composición polinucleotídica de andamio de ácidos nucleicos manipulada por ingeniería genética de unión a una primera secuencia diana de ácidos nucleicos y se une a una segunda secuencia diana de ácidos nucleicos que es cortada por una nucleasa del complejo.

La **Figura 13** ilustra un complejo de nucleoproteínas que comprende una composición de polinucleótido de andamio de ácidos nucleicos manipulada genéticamente de unión a tres secuencias diana de ácidos nucleicos en un polinucleótido.

La **Figura 14** ilustra un complejo de nucleoproteínas que comprende una composición polinucleotídica de andamio de ácidos nucleicos manipulada por ingeniería genética a una primera secuencia diana de ácidos nucleicos en un primer polinucleótido y se une a una segunda secuencia diana de ácidos nucleicos y una tercera secuencia diana de ácidos nucleicos en un segundo polinucleótido, en el que el segundo y las secuencias diana del tercer ácido nucleico son cortadas mediante una nucleasa del complejo.

La **Figura 15** ilustra un complejo de nucleoproteínas que comprende una composición polinucleotídica de andamio de ácidos nucleicos manipulada para unirse a una primera secuencia diana de ácidos nucleicos en un primer polinucleótido, de unión a una segunda secuencia diana de ácidos nucleicos en un segundo polinucleótido, y de unión a una tercera secuencia diana de ácidos nucleicos en un tercer polinucleótido, en el que las secuencias diana de ácidos nucleicos segundo y tercero son cortadas por una nucleasa del complejo.

La **Figura 16 A, Figura 16B, y Figura 16C** ilustran una composición polinucleotídica de andamio de ácidos nucleicos manipulada formando un complejo de nucleoproteínas con dos proteínas diferentes y uniéndose a una primera secuencia diana de ácidos nucleicos en un primer polinucleótido y una segunda secuencia de ácidos nucleicos en un segundo polinucleótido. Se ilustran tres combinaciones de resultados de unión y escisión.

Descripción detallada de la invención

La invención esta definida por las reivindicaciones adjuntas.

Debe entenderse que la terminología utilizada en este documento tiene el propósito de describir solo realizaciones particulares, y no pretende ser limitante. Tal como se utiliza en esta especificación y en las reivindicaciones adjuntas, las formas singulares "un", "una" y "el/la" incluyen referentes plurales, a menos que el contexto indique claramente lo contrario. Así, por ejemplo, la referencia a "un polinucleótido" incluye uno o más polinucleótidos, y la referencia a "un vector" incluye uno o más vectores.

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que entiende comúnmente un experto en la técnica a la que es pertinente la invención. Aunque otros procedimientos y materiales similares, o equivalentes, a los descritos en el presente documento pueden ser útiles en la presente invención, los materiales y procedimientos preferidos se describen en el presente documento.

En vista de las enseñanzas de la presente especificación, un experto en la técnica puede emplear técnicas convencionales de inmunología, bioquímica, química, biología molecular, microbiología, biología celular, genómica y

polinucleótidos recombinantes, como se enseña, por ejemplo, por los siguientes textos estándar: Antibodies: A Laboratory Manual, Second edition, E. A. Greenfield, Cold Spring Harbor Laboratory Press, ISBN 978-1-936113-81-1 (2014); Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th Edition, R. I. Freshney, Wiley-Blackwell, ISBN 978-0-470-52812-9 (2010); Transgenic Animal Technology, Third Edition: A Laboratory Handbook, C. A. Pinkert, Elsevier, ISBN 978-0124104907 (2014); The Laboratory Mouse, Second Edition, H. Hedrich, Academic Press, ISBN 978-0123820082 (2012); Manipulating the Mouse Embryo: A Laboratory Manual, R. Behringer, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1936113019 (2013); PCR 2: A Practical Approach, M. J. McPherson, et al., IRL Press, ISBN 978-0199634248 (1995); Methods in Molecular Biology (Series), J.M. Walker, ISSN 1064-3745, Humana Press; RNA: A Laboratory Manual, D. C. Rio, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879698911 (2010); Methods in Enzymology (Series), Academic Press; Molecular Cloning: A Laboratory Manual (Fourth Edition), M. R. Green, et al., Cold Spring Harbor Laboratory Press, ISBN 978-1605500560 (2012); Bioconjugate Techniques, Third Edition, G. T. Hermanson, Academic Press, ISBN 978-0123822390 (2013); Methods in Plant Biochemistry and Molecular Biology, W. V. Dashek, CRC Press, ISBN 978-0849394805 (1997); Plant Cell Culture Protocols (Methods in Molecular Biology), V. M. Loyola-Vargas, et al., Humana Press, ISBN 978-1617798177 (2012); Plant Transformation Technologies, C. N. Stewart, et al., Wiley-Blackwell, ISBN 978-0813821955 (2011); Recombinant Proteins from Plants (Methods in Biotechnology), C. Cunningham, et al., Humana Press, ISBN 978-1617370212 (2010); Plant Genomics: Methods and Protocols (Methods in Molecular Biology), D. J. Somers, et al., Humana Press, ISBN 978-1588299970 (2009); Plant Biotechnology: Methods in Tissue Culture and Gene Transfer, R. Keshavachandran, et al., Orient Blackswan, ISBN 978-8173716164 (2008).

Las repeticiones palindrómicas cortas agrupadas en intervalos regulares (CRISPR) y las proteínas asociadas a CRISPR relacionadas (proteínas Cas) constituyen sistemas CRISPR-Cas (véase, por ejemplo, Barrangou, R., et al., Science 315: 1709-1712 (2007)).

Como se usa en este documento, "proteína Cas" y "proteína CRISPR-Cas" se refieren a proteínas Cas que incluyen, entre otras, proteínas Cas de Clase 1 Tipo I, proteínas Cas de Clase 1 Tipo III, proteínas Cas de Clase 1 Tipo IV, Proteínas Cas de Clase 2 Tipo II, proteínas Cas de Clase 2 Tipo V y proteínas Cas de Clase 2 Tipo VI. Las proteínas Cas de clase 2 incluyen proteínas Cas9, proteínas similares a Cas9 codificadas por ortólogos Cas9, proteínas sintéticas similares a Cas9, proteínas Cpf1, proteínas codificadas por ortólogos Cpf1, proteínas sintéticas similares a Cpf1, proteínas C2c1, proteínas C2c2, proteínas C2c3 y variantes y modificaciones de los mismos. En algunos aspectos, las proteínas Cas son proteínas Cas de clase 2, por ejemplo, una o más proteínas Cas de clase 2 tipo II, como Cas9, una o más proteínas Cas de clase 2 tipo V, como Cpf1, o una o más de clase 2 Tipo VI. Proteínas cas, tales como C2c2. En los aspectos preferidos, las proteínas Cas son una o más proteínas Cas de Clase 2 Tipo II, como Cas9, y una o más proteínas Cas de Clase 2 Tipo V, como Cpf1. Típicamente, una proteína Cas es capaz de interactuar con uno o más polinucleótidos afines (más típicamente, ARN) para formar un complejo de nucleoproteínas (más típicamente, un complejo de ribonucleoproteínas).

"Proteína Cas9", como se usa en el presente documento, se refiere a una proteína de tipo salvaje Cas9 derivada de los sistemas CRISPR-Cas9 Clase 2 Tipo II, modificaciones de las proteínas Cas9, variantes de proteínas Cas9, ortólogos Cas9 y combinaciones de los mismos. Las proteínas Cas9 incluyen, pero no se limitan a, Cas9 de *Streptococcus pyogenes* (UniProtKB - Q99ZW2 (CAS9_STRP1)), *Streptococcus thermophilus* (UniProtKB - G3ECR1 (CAS9_STRTR)), y *Staphylococcus aureus* (UniProtKB - J7RUA5 (CAS9_STAAU)). Los homólogos de Cas9 pueden identificarse utilizando procedimientos de búsqueda de similitud de secuencias conocidos por los expertos en la técnica. "dCas9", como se usa en el presente documento, se refiere a variantes de la proteína Cas9 que son proteínas Cas9 desactivadas por nucleasa, también denominadas "proteína Cas9 catalíticamente inactiva", "Cas9 enzimáticamente inactiva", "Cas9 catalíticamente muerta" o "Cas9 muerto". Dichas moléculas carecen de toda o una parte de la actividad de la endonucleasa y, por lo tanto, pueden usarse para regular los genes de una manera guiada por el ARN (véase Jinek M., et al., Science 337: 816-821 (2012)). Esto se logra mediante la introducción de mutaciones en los residuos catalíticos, como D10A en el dominio RuvC-1 y H840A en el dominio HNH (numerado con relación a la proteína Cas9 de *S. pyogenes*), que inactivan la función de nucleasa Cas9. Se entiende que la mutación de otros residuos catalíticos para reducir la actividad de cualquiera o ambos de los dominios de nucleasa también puede llevarse a cabo por un experto en la técnica. El dCas9 resultante es incapaz de escindir el ADN de doble cadena, pero conserva la capacidad de formar complejos con un ácido nucleico guía y unirse a una secuencia diana de ADN. El doble mutante de Cas9 con cambios en las posiciones de aminoácidos D10A y H840A inactiva tanto las actividades de nucleasa como de níqueasa. La especificidad de la orientación está determinada por la unión de la proteína Cas9 a la secuencia PAM y por el apareamiento de bases complementarias del ARN guía (típicamente, un ARN de guía única) al locus genómico. Cas9 es la característica de la proteína para los sistemas CRISPR de Clase 2 Tipo II.

"Proteína Cpf1", como se usa en el presente documento, se refiere a una proteína de tipo salvaje Cpf1 derivada de los sistemas de Clase 2 Tipo V CRISPR-Cpf1, modificaciones de las proteínas Cpf1, variantes de proteínas Cpf1, ortólogos Cpf1 y combinaciones de los mismos. "dCpf1", como se usa en el presente documento, se refiere a variantes de la proteína Cpf1 que son proteínas Cpf1 desactivadas por nucleasa, también denominadas "proteína Cpf1 catalíticamente inactiva" o "Cpf1 enzimáticamente inactiva". Las proteínas Cpf1 incluyen, pero no se limitan a, *Francisella novicida* (UniProtKB - A0Q7Q2 (CPF1_FRATN)), *Lachnospiraceae* bacterium (UniProtKB - A0A182DWE3 (A0A182DWE3_9FIRM)), y *Acidaminopoccus* sp. (UniProtKB - U2UMQ6 (CPF1_ACISB)). Cpf1 es la

característica distintiva de la proteína para los sistemas de Clase 2 Tipo V CRISPR. Los homólogos de CpfI pueden identificarse usando procedimientos de búsqueda de similitud de secuencia conocidos por un experto en la técnica.

"Ácido nucleico que se dirige a ácido nucleico" (NATNA), como se usa en el presente documento, se refiere a uno o más polinucleótidos que guían una proteína, como una proteína Cas (por ejemplo, una proteína Cas9 o una proteína CpfI), para unirse preferentemente a una secuencia diana de ácidos nucleicos en un polinucleótido (en relación con un polinucleótido que no comprende la secuencia diana de ácidos nucleicos). Las NATNA pueden comprender bases de ribonucleótidos (por ejemplo, ARN), bases de desoxirribonucleótidos (por ejemplo, ADN), combinaciones de bases de ribonucleótidos y bases de desoxirribonucleótidos (por ejemplo, ARN/ADN), nucleótidos, análogos de nucleótidos, nucleótidos manipulados y similares, así como residuos o enlaces manipulados sintéticos, naturales y no naturales, por ejemplo, como se describe en el presente documento. Ejemplos de ácidos nucleicos que se dirigen a ácidos nucleicos incluyen, pero no se limitan a, moléculas de ARNcr/TraARNcr de Cas9 (véase, por ejemplo, la Figura 1A, la Figura 1B, la Figura 1C, y la Figura 1D), el ARNsg de Cas9 (véase, por ejemplo, la Figura 2A y la Figura 2B), y CpfI-ARNcr (véase, por ejemplo, la Figura 3A y la Figura 3B).

Tal como se usa en el presente documento, "ARN de guía dual" y "ARN de guía dual de Cas9" generalmente se refieren a un sistema de ARN de dos componentes para un componente de polinucleótido capaz de asociarse con una proteína Cas9 relacionada. La Figura 1A y Figura 1B presenta ejemplos ilustrativos de los ARN de guía dual asociados con CRISPR-Cas9 de Clase 2 Tipo II. La Figura 1A ilustra un ARN de dos componentes del sistema CRISPR-Cas9 de Tipo II que comprende un Cas9-ARNcr (Figura 1A, 101) y un Cas9-traARNcr (Figura 1A, 102). La Figura 1B ilustra la formación de enlaces de hidrógeno de pares de bases entre Cas9-ARNcr y Cas9-traARNcr para formar una estructura secundaria (véase, por ejemplo, la Solicitud de Patente Publicada de los Estados Unidos No. 2014-0068797, publicada el 6 de marzo de 2014; ver también Jinek M., et al., Science 337: 816-21 (2012)). La Figura 1B presenta una visión general y nomenclatura de los elementos estructurales secundarios del Cas9-ARNcr y Cas9-traARNcr del Cas9 de *S. pyogenes* que incluye lo siguiente: un elemento espaciador (Figura 1B, 103) que comprende una secuencia espaciadora (también denominada aquí como una secuencia de unión a un ácido nucleico); un primer elemento de tallo (Figura 1B, 104, 105, 106) que comprende un elemento de tallo inferior (Figura 1B, 104), un elemento de protuberancia que comprende nucleótidos no emparejados (Figura 1B, 105) y un elemento de tallo superior (Figura 1B, 106); un elemento nexa (Figura 1B, 107) que comprende un segundo elemento de tallo; un primer elemento de horquilla 3' (Figura 1B, 108) que comprende un tercer elemento de tallo; y un segundo elemento de horquilla 3' (Figura 1B, 109) que comprende un cuarto elemento de tallo. En algunos sistemas CRISPR-Cas9 Clase 2 Tipo II, el primer elemento de tallo no tiene un elemento de protuberancia (por ejemplo, *C. jejuni*). La Figura 1C ilustra un sistema de componente de dos ARN CRISPR-Cas9 de Tipo II que comprende un Cas9-ARNcr (Figura 1C, 101) y un Cas9-traARNcr (Figura 1C, 102). La Figura 1D ilustra la formación de enlaces de hidrógeno de pares de bases entre Cas9-ARNcr y Cas9-traARNcr para formar una estructura secundaria. La Figura 1D presenta una visión general y una nomenclatura de lo siguiente: un elemento espaciador (Figura 1D, 103); un primer elemento de tallo (Figura 1D, 110); un elemento nexa (Figura 1D, 107) que comprende un segundo elemento de tallo; un primer elemento de horquilla 3' (Figura 1D, 108) que comprende un tercer elemento de tallo; y un segundo elemento de horquilla 3' (Figura 1D, 109) que comprende un cuarto elemento de tallo. Un ARN de guía dual Cas9 es capaz de formar un complejo de nucleoproteínas con una proteína Cas9 relacionada, en el que el complejo es capaz de dirigirse a una secuencia diana de ácidos nucleicos complementaria a la secuencia espaciadora. Las modificaciones de las guías duales Cas9 son conocidas en la técnica, incluida la eliminación de uno o más elementos de horquilla 3' (Figura 1B, 108, 109; Figura 1D, 108, 109), modificaciones del primer elemento de tallo (Figura 1B, 104, 105, 106; Figura 1D 110), y modificaciones del tallo superior, la protuberancia y el tallo inferior (Figura 1B, 106, 105, 104, respectivamente) (véase, por ejemplo, la Publicación de Patente de los Estados Unidos No. 2014-0315985, publicado el 23 de octubre de 2014; Publicación de Patente de los Estados Unidos No. 2015-0376586, publicado el 31 de diciembre de 2015). Tal como se usa en el presente documento, un "polinucleótido Cas9 de guía dual" se refiere a un sistema de dos componentes que tiene un polinucleótido con los mismos elementos estructurales que un ARNcr (Figura 1A, 101) y un polinucleótido con los mismos elementos estructurales que un TraARNcr (Figura 1A 102). Un sistema polinucleotídico Cas9 de guía dual es capaz de asociarse con una proteína Cas9 relacionada.

Como se usa en el presente documento, "ARN de guía única" (ARNsg) y "Cas9-ARNsg" normalmente se refieren a un sistema de ARN de un componente como se describe más detalladamente en el presente documento, en el que el sistema es capaz de asociarse con una proteína Cas9 relacionada.

La Figura 2A, Figura 2B y Figura 2C muestran ejemplos de ARN asociado a CRISPR-Cas9 de Clase 2 Tipo II. Estas Figuras ilustran los ARN de guía única de Cas9 (ARNsg de Cas9), en los que el ARNcr de Cas9 se une covalentemente al ARN-tracr de Cas9, a menudo a través de un tetrabucle, y forma una estructura secundaria de polinucleótido de ARN a través del enlace de hidrógeno de par de bases (véase, por ejemplo, Solicitud de Patente Publicada de los Estados Unidos No. 2014-0068797, publicada el 6 de marzo de 2014). La Figura 2A presenta una descripción general y una nomenclatura de los elementos estructurales secundarios de un ARNsg de Cas9 para *S. pyogenes* que incluye lo siguiente: un elemento espaciador (Figura 2A, 201) que comprende una secuencia espaciadora (también denominada aquí como un ácido nucleico que se dirige al ácido nucleico secuencia); un primer elemento de tallo-bucle (Figura 2A, 202, 205, 203, 204) que comprende un elemento de tallo inferior (Figura 2A, 202), un elemento de protuberancia que comprende nucleótidos no apareados (Figura 2A, 205) y un tallo superior elemento (Figura 2A, 203), y un elemento de bucle (Figura 2A, 204) que comprende nucleótidos desapareados; un elemento nexa (Figura 2A, 206) que comprende un segundo elemento de tallo-bucle; un primer elemento de

horquilla 3' (Figura 2A, 207) que comprende un tercer elemento de tallo-bucle; y un segundo elemento de horquilla 3' que comprende un tercer elemento de tallo (Figura 2A, 208) que comprende un cuarto elemento de tallo-bucle. (véase, por ejemplo, las Figuras 1 y 3 de Briner, A. E., et al., Molecular Cell 56(2): 333-339 (2014)).

La Figura 2B presenta una descripción general y una nomenclatura de los elementos estructurales secundarios de un ARNsg de Cas9 para *C. jejuni*, que incluye lo siguiente: un elemento espaciador (Figura 2B, 201); un primer elemento de tallo (Figura 2B, 209) y un elemento de bucle (Figura 2B, 204) que comprenden nucleótidos no apareados (es decir, el primer elemento de tallo-bucle comprende el primer elemento de tallo y el elemento de bucle); un elemento nexa (Figura 2B, 206) que comprende un segundo elemento de tallo-bucle; un primer elemento de horquilla 3' (Figura 2B, 207) que comprende un tercer elemento de tallo-bucle; y un segundo elemento de horquilla 3' que comprende un tercer elemento de tallo (Figura 2B, 208) que comprende un cuarto elemento de tallo-bucle. Un Cas9-ARNsg es capaz de formar un complejo de nucleoproteínas con una proteína Cas9 afín, en el que el complejo es capaz de dirigirse a una secuencia de ácidos nucleicos complementaria a la secuencia espaciadora.

Se conocen en la técnica modificaciones de Cas9 de guía única que incluyen, entre otras, la eliminación de uno o más elementos de horquilla 3' (Figura 2, 207, 208), modificaciones del primer elemento de tallo (Figura 1B, 104, 105, 106; Figura 1D 110), y modificaciones del tallo superior, la protuberancia y el tallo inferior (Figura 1B, 106, 105, 104, respectivamente) (véase, por ejemplo, la Publicación de Patente de los Estados Unidos No. 2014-0315985, publicada el 23 de octubre de 2014; Publicación de Patente de los Estados Unidos No. 2015-0376586, publicada el 31 de diciembre de 2015).

Como se usa en el presente documento, un "polinucleótido de guía única de Cas9" se refiere a un sistema de un componente que tiene los mismos elementos estructurales que un ARNsg (Figura 2). Un sistema polinucleotídico Cas9 de guía única es capaz de asociarse con una proteína Cas9 relacionada.

La Figura 2C presenta una ilustración más detallada de la Figura 2A. La Tabla 1 presenta una serie de indicadores numéricos utilizados para ilustrar regiones de secuencias de ácidos nucleicos asociadas con un ARNsg de CRISPR-Cas9 Clase 2 Tipo II. En la Tabla 1, "": "es el equivalente del término "que comprende".

Tabla 1

Indicadores numéricos utilizados para ilustrar regiones de secuencias de ácidos nucleicos en un ARNsg	
Indicador	Región correspondiente
	Secuencias de unión a proteínas de unión a ácidos nucleicos
210	un término 3'
210-211	una secuencia de ácidos nucleicos de elemento enlazador: una secuencia de ácidos nucleicos de elemento enlazador terminal 3'
	elementos de horquilla 3' (compárense Figura 2A, 207, 208)
211-214	una secuencia 1-2 de ácidos nucleicos de horquilla: un elemento 1-2 de horquilla 3'
211-212	a 3' una secuencia 2 de ácidos nucleicos del elemento de tallo 1-2 de horquilla 3'
212-213	a 3' una secuencia de ácidos nucleicos del elemento de bucle 1-2 de horquilla 3'
213-214	a 3' una secuencia 1 de ácidos nucleicos del elemento de tallo 1-2 de horquilla 3'
214-215	una secuencia de ácidos nucleicos de elemento enlazador: una secuencia 1-2 de ácidos nucleicos de elemento enlazador de horquilla 3'
215-218	una secuencia 1-1 de ácidos nucleicos de horquilla: un elemento 1-1 de horquilla 3'

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de secuencias de ácidos nucleicos en un ARNsg	
Indicador	Región correspondiente
	Secuencias de unión a proteínas de unión a ácidos nucleicos
215-216	a 3' una secuencia 2 de ácidos nucleicos del elemento de tallo 1-1 de horquilla 3'
216-217	a 3' una secuencia de ácidos nucleicos del elemento de bucle 1-1 en horquilla 1
217-218	a 3' una secuencia 1 de ácidos nucleicos del elemento de tallo 1-1 de horquilla 3'
218-219	una secuencia de ácidos nucleicos de elemento enlazador: una secuencia 1-1 de ácidos nucleicos de elemento enlazador de horquilla de 3'
	Elemento nexa
219-220	una secuencia de ácidos nucleicos del elemento nexa
219-230	una secuencia 1-1 de ácidos nucleicos del elemento de tallo nexa dividido
230-220	una secuencia de ácidos nucleicos del elemento de tallo nexa 1-2 dividido
	Elemento enlazador de tallo 1
220-221	una secuencia 1 de ácidos nucleicosonectivo: una secuencia 1 de ácidos nucleicos de elemento enlazador
221-228	primer elemento de tallo-bucle (véase también Figura 2A, 202, 203, 204, 205; Figura 2B, 204, 209)
221-224/225-228	primer elemento de tallo (véase también Figura 1A, 104, 105, 106; Figura 1D, 110; 2A, 202, 205, 203)
221-222/227-228	un elemento 1 de tallo inferior
222-223/226-227	un elemento 1 de protuberancia
223-224/225-226	un elemento 1 del tallo superior
221-224	una secuencia 1-2 de ácidos nucleicos del primer elemento de tallo
221-222	una secuencia 1-2 de ácidos nucleicos del elemento de tallo inferior
222-223	una secuencia 1-2 de ácidos nucleicos del elemento de protuberancia
223-224	una secuencia 1-2 de ácidos nucleicos del elemento de tallo superior

224-225	una secuencia 1 de ácidos nucleicos del elemento de bucle
224-225	una secuencia 1-1 de ácidos nucleicos del elemento de tallo superior
224-225	una secuencia 1-1 de ácidos nucleicos del elemento de protuberancia
224-225	una secuencia 1-1 de ácidos nucleicos del elemento de tallo inferior
225-228	una primera secuencia 1-1 de ácidos nucleicos del elemento de tallo
228-229	un elemento espaciador 1 que comprende una secuencia 1 de ácidos nucleicos: una secuencia espaciadora de ácidos nucleicos: una secuencia de unión a diana de ácidos nucleicos (por ejemplo, una secuencia de unión a diana de ADN)
229	un terminal 5'

"Clase 2 Tipo V guía ARNcr" y "CpfI-ARNcr", como se usa en este documento, generalmente se refiere a un sistema

de ARN de un componente para un componente polinucleótido capaz de asociarse con una proteína CpfI afín (véase, por ejemplo, Zetsche, B., et al., Cell 163: 1-13 (2015)). La Figura 3A presenta un ejemplo de un ARN asociado a CRISPR-CpfI de Tipo V (CpfI-ARNcr), así como una descripción general y nomenclatura de los elementos estructurales secundarios de un ARNcr de CpfI como sigue: un elemento de tallo-bucle (Figura 3A, 301) y un elemento espaciador (Figura 3A, 302) que comprende una secuencia de unión a diana de ácidos nucleicos. El elemento de tallo-bucle comprende, en una dirección 5' a 3', una secuencia 1C de ARN del tallo CpfI (Figura 3A, 303), un elemento de bucle (Figura 3A, 304) y una secuencia complementaria de ARN del tallo CpfI 1C (Figura 3A, 305), en el que la secuencia 1 de ARN del tallo CpfI y la secuencia 1C complementaria del ARN del tallo CpfI forman un dúplex. La Figura 3B presenta una modificación del CpfI-ARNcr en el que el elemento de bucle se elimina del elemento de tallo-bucle de la Figura 3A. La Figura 3B ilustra un elemento de tallo (Figura 3B, 301) que comprende, en una dirección 5' a 3', una secuencia de ácidos nucleicos de tallo CpfI 1 (Figura 3B, 303); una secuencia de ácidos nucleicos de tallo CpfI complementaria 1C (Figura 3B, 305), en la que la secuencia de ácidos nucleicos de tallo CpfI 1 y la secuencia de ácidos nucleicos de tallo CpfI complementaria 1C forman un dúplex; y un elemento espaciador (Figura 3A, 302) que comprende una secuencia de unión a diana de ácidos nucleicos. Una guía ARNcr es capaz de formar un complejo de nucleoproteínas con una proteína CpfI relacionada, en el que el complejo es capaz de dirigirse a una secuencia diana de ácidos nucleicos complementaria a la secuencia de unión a diana de ácidos nucleicos.

Como se usa en el presente documento, "una secuencia de unión a ácidos nucleicos diana" y "secuencia espaciadora de ácidos nucleicos" se refieren a secuencias de ácidos nucleicos capaces de hibridar con una secuencia diana de ácidos nucleicos en un polinucleótido. Un "elemento espaciador" comprende una secuencia de unión a diana de ácidos nucleicos.

Como se usa en el presente documento, "un andamio de ácidos nucleicos", "NASC", "una composición de polinucleótidos de NASC", "una composición NASC" y "una composición de polinucleótidos de NASC" se refieren a un complejo de polinucleótidos que forma un andamio. En realizaciones preferidas, el andamio es capaz de unirse a una proteína CRISPR-Cas9 Clase 2 Tipo II. Típicamente, una composición de polinucleótido NASC es un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio que comprende: (i) un elemento de repetición 1 (por ejemplo, que comprende una secuencia 1 de ácidos nucleicos repetida) y un elemento de repetición 2 (por ejemplo, que comprende una repetir la secuencia 2 de ácidos nucleicos); (ii) un elemento de unión a proteína de unión a ácidos nucleicos 1 (por ejemplo, que comprende una secuencia de unión a proteína de unión a ácidos nucleicos 1) y un elemento de unión a proteína de unión a ácidos nucleicos 2 (por ejemplo, que comprende una secuencia de unión a proteína de unión a ácidos nucleicos 2); y (iii) un elemento espaciador 1 (por ejemplo, que comprende una secuencia 1 de unión a diana de ácidos nucleicos) y un elemento espaciador 2 (por ejemplo, que comprende una secuencia 2 de unión a diana de ácidos nucleicos). En una composición de polinucleótido NASC, el elemento de repetición 1 está conectado con el elemento de repetición 2.

La composición de polinucleótido de NASC es capaz de asociarse con una proteína de unión a ácidos nucleicos. En algunas realizaciones, la composición de polinucleótido NASC es capaz de asociarse con dos o más proteínas CRISPR-Cas9 Clase 2 Tipo II (por ejemplo, proteínas CRISPR-Cas9 Clase 2 Tipo II que tienen motivos estructurales y motivos funcionales similares) para formar un complejo de nucleoproteína. Ejemplos de proteínas CRISPR-Cas9 Clase 2 Tipo II se explican a continuación.

En algunas realizaciones de composiciones de polinucleótidos de NASC, cada uno de un primer componente de polinucleótido de NASC (por ejemplo, una NASC-PC1 que comprende un elemento de repetición 1, un elemento de unión a proteína de unión a ácidos nucleicos 1 y un elemento espaciador 1) y un segundo componente de polinucleótido de NASC (por ejemplo, NASC-PC2 que comprende un elemento de repetición 2, un elemento de unión a la proteína de unión a ácidos nucleicos y un elemento espaciador 2) es capaz de asociarse con la misma clase de proteína CRISPR-Cas9 de Clase 2 Tipo II (por ejemplo, de Clase 2 Tipo II proteínas CRISPR-Cas9 que tienen motivos estructurales y motivos funcionales similares para formar un complejo de nucleoproteínas.

En otras realizaciones de composiciones de polipéptidos NASC, un complejo de nucleoproteínas puede formarse por una unión de proteína CRISPR-Cas9 de Tipo II de Clase 2 a una macromolécula que comprende una secuencia 1 de unión a diana de ácidos nucleicos, una secuencia 1 de ácidos nucleicos repetida, a repetir la secuencia 1 de ácidos nucleicos, y una secuencia 1 de unión a diana de ácidos nucleicos.

Una composición de polinucleótido NASC/proteína de unión a ácidos nucleicos 1/complejo de proteína de unión a ácidos nucleicos 2 es capaz de unirse preferentemente a una secuencia diana de ácidos nucleicos en un polinucleótido (en relación con un polinucleótido que no comprende la secuencia diana de ácidos nucleicos).

Un polinucleótido NASC (NASC-PC) que comprende múltiples elementos espaciadores se denomina genéricamente aquí como "NASC-PC-MTS", y se refiere específicamente con referencia al número de elementos espaciadores "a NASC-PC- (número de elementos espaciadores)TS" (por ejemplo, para dos elementos espaciadores, la designación utilizada es NASC-PC-2TS).

Los componentes de una NASC-PC que comprenden múltiples polinucleótidos se mencionan aquí con referencia al número de polinucleótidos "a NASC-PC-(número de polinucleótidos)" (por ejemplo, para dos polinucleótidos, la designación utilizada es una NASC-PC1-1 y una NASC-PC 1-2).

Un componente de polinucleótido NASC-PC que comprende elementos concatenados se denomina en este documento "NASC-PC-CE". En realizaciones particulares que comprenden polinucleótidos de nexo dividido, un componente de polinucleótido NASC que comprende elementos de nexo dividido concatenado se denomina en el presente documento "NASC-PC-SCE".

Como se usa en el presente documento, "una secuencia de refuerzo de ácidos nucleicos" es una secuencia de ácidos nucleicos que comprende al menos dos secuencias diana de ácidos nucleicos distintas: una secuencia diana de ácidos nucleicos complementaria de una secuencia de unión diana de ácidos nucleicos de una primera composición polinucleotídica NASC y una secuencia diana de ácidos nucleicos 2 complementaria a una secuencia 2 de unión a diana de ácidos nucleicos de una segunda composición de polinucleótido NASC. Un ejemplo de una secuencia de refuerzo de ácidos nucleicos es una secuencia de refuerzo de ADN.

Como se usa en el presente documento, "una Composición NASC-Cage (NASC-CC)" comprende al menos una primera composición polinucleotídica NASC conectada por secuencias de refuerzo de ácidos nucleicos a una segunda composición polinucleotídica NASC para formar una estructura similar a una jaula que típicamente tiene un espacio interno para el empaquetado de moléculas.

Como se usa en el presente documento, el término "cognado" se refiere típicamente a una proteína Cas (por ejemplo, proteína Cas9 o una proteína Cpf1) y uno o más polinucleótidos de Cas (por ejemplo, NATNA asociada a CRISPR-Cas9 Clase 2 Tipo II o Clase 2 Tipo V CRISPR-NATNA asociadas a Cpf1, respectivamente) que son capaces de formar un complejo de nucleoproteínas capaz de unirse al sitio a una secuencia diana de ácidos nucleicos complementaria a la secuencia de unión diana de ácidos nucleicos presente en uno de los uno o más polinucleótidos de Cas.

Los términos "de tipo salvaje", "de origen natural" y "sin modificar" se usan en el presente documento para referirse a la forma, apariencia, fenotipo o cepa típica (o más común) que existe en la naturaleza; por ejemplo, la forma típica de células, organismos, características, polinucleótidos, proteínas, complejos macromoleculares, genes, ARN, ADN o genomas a medida que aparecen y se pueden aislar de una fuente en la naturaleza. La forma, la apariencia, el fenotipo o la cepa de tipo salvaje sirven como el padre original antes de una modificación intencional. Por lo tanto, las formas mutantes, variantes, manipuladas, recombinantes y modificados no son formas de tipo salvaje.

Como se usa en el presente documento, los términos "manipulados", "manipulados genéticamente", "recombinantes", "modificados", "no naturales", y "no nativos" son intercambiables e indican intencionalmente la manipulación humana.

Como se usa en el presente documento, "interrumpido", "roto" y "discontinuo" se usan indistintamente para significar una ruptura en la continuidad, por ejemplo, en enlaces covalentes de un esqueleto de polinucleótido. Por ejemplo, un primer polinucleótido y un segundo polinucleótido que son discontinuos tienen cada uno un extremo 5' y un extremo 3' (término 5' terminal primero polinucleótido-3' y terminal 5' terminal-segundo polinucleótido-3'). Por ejemplo, el término 5' de una molécula de ADN o ARN es típicamente el quinto carbono en el anillo de azúcar y el término 3' es típicamente el grupo hidroxilo en el tercer carbono en el anillo de azúcar. Dos polinucleótidos, cada uno con un extremo 5' y un extremo 3', se forman cuando el esqueleto de un solo polinucleótido se rompe en un sitio. Un

extremo 5' y/o 3' puede modificarse covalentemente, por ejemplo, mediante la adición de una unidad estructural (por ejemplo, una unidad estructural que proporciona resistencia a los efectos degradantes de las exonucleasas).

"Enlace covalente", "unido covalentemente", "enlazado covalentemente", "ligado covalentemente", "conectado covalentemente" y "enlace molecular" se usa de manera intercambiable en el presente documento, y se refiere a un enlace químico que implica el intercambio de pares de electrones entre los átomos. Ejemplos de enlaces covalentes incluyen, pero no se limitan a, enlaces fosfodiéster y enlaces fosforotioato.

"Enlace no covalente", "unido no covalentemente", "enlazado no covalentemente", "ligado no covalentemente", "interacción no covalente", y "conectado no covalentemente" se usan de manera intercambiable en este documento, y hacen referencia a cualquier enlace químico relativamente débil que no implique compartir un par de electrones. Los enlaces no covalentes múltiples a menudo estabilizan la conformación de las macromoléculas y median las interacciones específicas entre las moléculas. Ejemplos de enlaces no covalentes incluyen, entre otros, enlaces de hidrógeno, interacciones iónicas (por ejemplo, Na^+Cl^-), interacciones de van der Waals y enlaces hidrófobos.

Como se usa en este documento, "enlace de hidrógeno", "apareamiento de bases con hidrógeno", "apareamiento de bases con enlace de hidrógeno", "enlace de hidrógeno" y "pares de bases enlazados con hidrógeno" se usan de manera intercambiable y se refieren a enlaces de hidrógeno canónicos y enlaces de hidrógeno no canónicos que incluyen, pero no se limitan a, "pares de bases enlazados con hidrógeno de Watson-Crick" (pares de bases enlazados con hidrógeno-W-C o enlaces de hidrógeno con W-C); "Pares de bases unidas con hidrógeno de Hoogsteen" (enlace de hidrógeno de Hoogsteen); y "pares de bases con uniones de hidrógeno ondulantes" (enlace de hidrógeno ondulante). El enlace de hidrógeno W-C, incluido el enlace de hidrógeno W-C inverso, se refiere al emparejamiento de bases purina-pirimidina, que es adenina:timina, guanina:citosina y uracilo:adenina. El enlace de hidrógeno de Hoogsteen, incluido el enlace de hidrógeno de Hoogsteen inverso, se refiere a una variación del emparejamiento de bases en los ácidos nucleicos en el que dos nucleobases, uno en cada cadena, se mantienen unidos por enlaces de hidrógeno en el surco principal. Este enlace de hidrógeno no W-C puede permitir que una tercera hebra se enrolle alrededor de un dúplex y forme hélices de triple hebra. El enlace de hidrógeno oscilante, incluido el enlace de hidrógeno oscilante inverso, se refiere a un emparejamiento entre dos nucleótidos en moléculas de ARN que no siguen las reglas del par de bases de Watson-Crick. Hay cuatro pares de bases oscilantes principales: guanina:uracilo, inosina (hipoxantina):uracilo, inosina-adenina e inosina-citosina. Los expertos en la técnica conocen las reglas para el enlace de hidrógeno canónico y el enlace de hidrógeno no canónico (véase, por ejemplo, *The RNA World*, Third Edition (Cold Spring Harbor Monograph Series), R.F. Gesteland, Cold Spring Harbor Laboratory Press, ISBN 978-0879697396 (2005); *The RNA World*, Second Edition (Cold Spring Harbor Monograph Series), R.F. Gesteland, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879695613 (1999); *The RNA World* (Cold Spring Harbor Monograph Series), R.F. Gesteland, et al., Cold Spring Harbor Laboratory Press, ISBN 978-0879694562 (1993) (véase, por ejemplo, Appendix 1: Structures of Base Pairs Involving at Least Two Hydrogen Bonds, I. Tinoco); *Principles of Nucleic Acid Structure*, W. Saenger, Springer International Publishing AG, ISBN 978-0-387-90761-1 (1988); *Principles of Nucleic Acid Structure*, First Edition, S. Neidle, Academic Press, ISBN 978-01236950791 (2007)).

"Conectar", "conectado" y "que conecta" se usan de manera intercambiable en este documento, y se refieren a un enlace covalente o un enlace no covalente entre dos macromoléculas (por ejemplo, polinucleótidos, proteínas y similares).

Como se usa en el presente documento, "complementariedad" se refiere a la capacidad de una secuencia de ácidos nucleicos para formar uno o más enlaces de hidrógeno con otra secuencia de ácidos nucleicos (por ejemplo, a través del emparejamiento de bases canónico de Watson-Crick). Un porcentaje de complementariedad indica el porcentaje de residuos en una molécula de ácidos nucleicos que puede formar enlaces de hidrógeno con una segunda secuencia de ácidos nucleicos. Si dos secuencias de polinucleótidos tienen un 100% de complementariedad, las dos secuencias son perfectamente complementarias, es decir, todos los residuos contiguos de un primer enlace de hidrógeno de polinucleótido con el mismo número de residuos contiguos en un segundo polinucleótido.

Como se usa en este documento, "unión" se refiere a una interacción no covalente entre macromoléculas (por ejemplo, entre una proteína y un polinucleótido, entre un polinucleótido y un polinucleótido, o entre una proteína y una proteína, y similares). Dicha interacción no covalente también se conoce como "asociación" o "interacción" (por ejemplo, si una primera macromolécula interactúa con una segunda macromolécula, la primera macromolécula se une a la segunda macromolécula de una manera no covalente). Algunas porciones de una interacción de unión pueden ser específicas de secuencia. La "unión específica de secuencia", como se usa en el presente documento, típicamente se refiere a una o más composiciones de polipéptidos NASC capaces de formar un complejo con una o más proteínas (por ejemplo, una proteína Cas9 y/o una proteína CpfI) para hacer que las proteínas se unan a una secuencia de ácidos nucleicos (por ejemplo, una secuencia de ADN) que comprende una secuencia diana de ácidos nucleicos (por ejemplo, una secuencia diana de ADN) preferentemente en relación con una segunda secuencia de ácidos nucleicos (por ejemplo, una segunda secuencia de ADN) sin la secuencia de unión diana de ácidos nucleicos (por ejemplo, la secuencia de unión a la diana del ADN). No es necesario que todos los componentes de una interacción de unión sean específicos de secuencia, como la unión de proteínas con residuos de fosfato en un esqueleto de ADN. Las interacciones de unión pueden caracterizarse por una constante de disociación (Kd).

"Afinidad de unión" se refiere a la fuerza de la interacción de unión. Una afinidad de unión aumentada se correlaciona con una Kd inferior.

Como se usa en este documento, se dice que una proteína Cas (por ejemplo, una proteína Cas9) "apunta" a un polinucleótido si un complejo de nucleoproteínas dirigido al sitio que comprende una proteína Cas se une o escinde un polinucleótido en la secuencia diana del ácido nucleico dentro del polinucleótido.

Tal como se usa en el presente documento, "ruptura de doble cadena" (DSB) se refiere a ambas cadenas de un segmento de doble cadena de ADN que se está cortando. En algunos casos, si se produce una ruptura de este tipo, se puede decir que una cadena tiene un "extremo pegajoso" en el que los nucleótidos están expuestos y no están unidos por hidrógeno a los nucleótidos en la otra cadena. En otros casos, puede ocurrir un "extremo romo" en el que ambas hebras permanecen completamente emparejadas en la base entre sí.

"Polinucleótido donante", "oligonucleótido donante" y "plantilla donante" se usan indistintamente en este documento y puede ser un polinucleótido bicatenario (por ejemplo, un ADN bicatenario), un polinucleótido monocatenario (por ejemplo, ADN trenzado), o una combinación de los mismos. Los polinucleótidos donantes comprenden brazos de homología que flanquean la secuencia de inserción (por ejemplo, DSB en el ADN). Los brazos de homología en cada lado pueden variar en longitud. Los parámetros para el diseño y la construcción de polinucleótidos donantes son bien conocidos en la técnica (véase, por ejemplo, Ran, F., et al., *Nature Protocols* 8(11): 2281-2308 (2013); Smithies, O. et al. al., *Nature* 317: 230-234 (1985); Thomas, K., et al., *Cell* 44: 419-428 (1986); Wu, S., et al., *Nature Protocols* 3: 1056-1076 (2008); Singer, B., et al., *Cell* 31: 25-33 (1982); Shen, P., et al., *Genetics* 112: 441-457 (1986); Watt, V., et al., *Proceedings of the National Academy of Sciences of the United States of America* 82: 4768-4772 (1985); Sugawara, N., et al., *Journal of Molecular Cell Biology* 12(2): 563-575 (1992); Rubnitz, J., et al., *Journal of Molecular Cell Biology* 4(11): 2253-2258 (1984); Ayares, D., et al., *Proceedings of the National Academy of Sciences of the United States of America* 83(14): 5199-5203 (1986); Liskay, R., et al., *Genetics* 115(1): 161-167 (1987)).

Como se usa en este documento, "reparación dirigida por homología" (HDR) se refiere a la reparación del ADN que tiene lugar en las células, por ejemplo, durante la reparación de un DSB en el ADN. HDR requiere homología de secuencia de nucleótidos y utiliza un polinucleótido donante para reparar la secuencia en la que ocurrió el DSB (por ejemplo, dentro de una secuencia diana de ADN). El polinucleótido donante generalmente tiene la homología de secuencia requerida con la secuencia que flanquea la DSB, de modo que el polinucleótido donante puede servir como una plantilla adecuada para la reparación. Los resultados de HDR en la transferencia de información genética desde, por ejemplo, el polinucleótido donante a la secuencia diana de ADN. La HDR puede resultar en la alteración de la secuencia diana del ADN (por ejemplo, inserción, eliminación o mutación) si la secuencia del polinucleótido donante difiere de la secuencia diana del ADN y parte o todo el polinucleótido donante se incorpora a la secuencia diana del ADN. En algunas realizaciones, un polinucleótido donante completo, una porción del polinucleótido donante o una copia del polinucleótido donante se integran en el sitio de la secuencia diana de ADN. Por ejemplo, se puede usar un polinucleótido donante para reparar la ruptura en la secuencia diana del ADN, en el que la reparación da como resultado la transferencia de información genética (es decir, secuencias de polinucleótidos) desde el polinucleótido donante en el sitio o muy cerca de la ruptura en el ADN. Por consiguiente, la nueva información genética (es decir, las secuencias de polinucleótidos) se puede insertar o copiar en una secuencia diana de ADN.

Una "región genómica" es un segmento de un cromosoma en el genoma de una célula huésped que está presente en cualquiera de los lados del sitio de secuencia diana de ácidos nucleicos o, alternativamente, también incluye una parte del sitio de secuencia diana de ácidos nucleicos. Los brazos de homología del polinucleótido donante tienen suficiente homología para experimentar una recombinación homóloga con las regiones genómicas correspondientes. En algunas realizaciones, los brazos de homología del polinucleótido donante comparten una homología de secuencia significativa con la región genómica que flanquea inmediatamente el sitio de secuencia diana de ácidos nucleicos; se reconoce que los brazos de homología pueden diseñarse para tener suficiente homología con las regiones genómicas más alejadas del sitio de secuencia diana de ácidos nucleicos.

Como se usa en este documento, "unión final no homóloga" (NHEJ) se refiere a la reparación de un DSB en el ADN mediante la ligadura directa de un extremo de la ruptura al otro extremo de la ruptura sin el requisito de un polinucleótido donante. NHEJ es una vía de reparación de ADN disponible para las células para reparar el ADN sin el uso de una plantilla de reparación. El NHEJ en ausencia de un polinucleótido donante a menudo da como resultado la inserción o eliminación aleatoria de nucleótidos en el sitio de la DSB.

La "unión de extremo mediada por microhomología" (MMEJ) es una vía para reparar un DSB en el ADN. MMEJ implica eliminaciones que flanquean un DSB y alineación de secuencias microhomólogas internas a los extremos rotos antes de unirse. MMEJ se define genéticamente y requiere la actividad de, por ejemplo, CtIP, Poli(ADP-Ribosa) Polimerasa 1 (PARP1), ADN polimerasa theta (Pol θ), ADN Ligasa 1 (Lig 1) o ADN Ligasa 3 (Lig 3). Se conocen componentes genéticos adicionales en la técnica (véase, por ejemplo, Sfeir, A., et al., *Trends in Biochemical Sciences* 40: 701-714 (2015)).

Como se usa en el presente documento, la "reparación de ADN" abarca cualquier proceso por el cual la maquinaria celular repara el daño a una molécula de ADN contenida en la célula. El daño reparado puede incluir roturas de una

sola hebra o roturas de doble hebra. Existen al menos tres mecanismos para reparar DSB: HDR, NHEJ y MMEJ. La "reparación de ADN" también se usa en el presente documento para referirse a la reparación de ADN resultante de la manipulación humana, en el que un locus objetivo se modifica, por ejemplo, insertando, eliminando o sustituyendo nucleótidos, todos los cuales representan formas de edición del genoma.

- 5 Como se usa en el presente documento, "recombinación" se refiere a un proceso de intercambio de información genética entre dos polinucleótidos.

Como se usa en este documento, los términos "secuencias tampónas", "elementos tampones" y "elementos de control" son intercambiables y se refieren a secuencias polinucleotídicas que están en sentido ascendente (secuencias no codificantes en 5'), dentro o corriente abajo (secuencias no traducidas en 3') de un polinucleótido diana por expresar. Las secuencias tampónas influyen, por ejemplo, en el momento de la transcripción, la cantidad o el nivel de transcripción, el procesamiento o la estabilidad del ARN y/o la traducción de la secuencia de nucleótidos estructurales relacionada. Las secuencias tampónas pueden incluir secuencias de unión a activadores, potenciadores, intrones, secuencias de reconocimiento de poliadenilación, promotores, sitios de inicio de la transcripción, secuencias de unión al represor, estructuras de tallo-bucle, secuencias de iniciación de la traducción, sitios de entrada de ribosomas internos (IRES), secuencias líderes de la traducción, secuencias de terminación de la transcripción (por ejemplo, señales de poliadenilación y secuencias de poli-U), secuencias de terminación de la traducción, sitios de unión a cebadores, y similares.

Los elementos tampones incluyen aquellos que dirigen la expresión constitutiva, inducible y reprimible de una secuencia de nucleótidos en muchos tipos de células huésped y aquellos que dirigen la expresión de la secuencia de nucleótidos solo en ciertas células huésped (por ejemplo, secuencias tampónas específicas de tejido). En algunas realizaciones, un vector comprende uno o más promotores pol III, uno o más promotores pol II, uno o más promotores pol I, o combinaciones de los mismos. Ejemplos de promotores pol III incluyen, pero no se limitan a, promotores U6 y H1. Ejemplos de promotores pol II incluyen, pero no se limitan a, el promotor LTR del virus del sarcoma de Rous (RSV) retroviral (opcionalmente con el potenciador del RSV), el promotor del citomegalovirus (CMV) (opcionalmente con el potenciador del CMV; ver, por ejemplo, Boshart, M., et al., Cell 41: 521-530 (1985)), el promotor SV40, el promotor de la dihidrofolato reductasa, el promotor de la β -actina, el promotor de la fosfoglicerol quinasa (PGK) y el promotor EF1 α . Los expertos en la materia apreciarán que el diseño de un vector de expresión puede depender de factores tales como la elección de la célula huésped que se va a transformar, el nivel de expresión deseado y similares. Se puede introducir un vector en las células huésped para producir de este modo transcripciones, proteínas o péptidos, incluidas proteínas de fusión o péptidos, codificados por ácidos nucleicos como se describe en el presente documento.

"Gen", como se usa en el presente documento, se refiere a una secuencia de polinucleótidos que comprende exón(es) y secuencias tampónas relacionadas. Un gen puede comprender además intrón(es) y/o región(es) no traducida (UTR(s)). Tal como se usa en el presente documento, el término "unido operativamente" se refiere a secuencias de polinucleótidos o secuencias de aminoácidos colocadas en una relación funcional entre sí. Por ejemplo, las secuencias tampónas (por ejemplo, un promotor o potenciador) están "unidas operativamente" a un polinucleótido que codifica un producto génico si las secuencias tampónas regulan o contribuyen a la modulación de la transcripción del polinucleótido. Los elementos tampones enlazados operativamente son típicamente contiguos con la secuencia de codificación. Sin embargo, los potenciadores pueden funcionar si están separados de un promotor por hasta varias kilobases o más. Por consiguiente, algunos elementos tampones pueden unirse operativamente a una secuencia de polinucleótidos, pero no contiguos a la secuencia de polinucleótidos. De manera similar, los elementos tampones de la traducción contribuyen a la modulación de la expresión de proteínas de un polinucleótido.

Como se usa en el presente documento, "expresión" se refiere a la transcripción de un polinucleótido de una plantilla de ADN, que resulta, por ejemplo, en un ARN mensajero (ARNm) u otro transcrito de ARN (por ejemplo, no codificante, como los ARN estructurales o de andamiaje). El término se refiere además al proceso a través del cual el ARNm transcrito se traduce en péptidos, polipéptidos o proteínas. Las transcripciones y los polipéptidos codificados se pueden denominar colectivamente como "producto(s) gen(es)". La expresión puede incluir el empalme del ARNm en una célula eucariota, si el polinucleótido se deriva de ADN genómico (ADNg).

Como se usa en el presente documento, el término "modular" se refiere a un cambio en la cantidad, el grado o la cantidad de una función. Por ejemplo, un complejo de composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda proteína de unión a ácidos nucleicos (por ejemplo, proteínas de clase 2 CRISPR-Cas), como se describe en este documento, puede modular la actividad de una secuencia promotora mediante la unión a dos o más secuencias diana de ácidos nucleicos en o cerca del promotor. Dependiendo de la acción que ocurra después de la unión, la composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda proteína de unión a ácidos nucleicos puede inducir, potenciar, suprimir o inhibir la transcripción de un gen unido operativamente a la secuencia promotora. Por lo tanto, la "modulación" de la expresión génica incluye tanto la activación génica como la represión genética.

La modulación puede analizarse determinando cualquier característica directa o indirectamente afectada por la expresión del gen objetivo. Tales características incluyen, por ejemplo, cambios en los niveles de ARN o proteínas,

actividad de proteínas, niveles de productos, expresión del gen o nivel de actividad de genes informadores. Por consiguiente, los términos "expresión de modulación", "expresión de inhibición" y "expresión de activación" de un gen pueden referirse a la capacidad de un complejo de proteína de unión de ácidos nucleicos/proteína de unión a ácidos nucleicos para cambiar, activar o inhibir la transcripción de un gen.

"Vector" y "plásmido", como se usa en el presente documento, se refieren a un vehículo polinucleótido para introducir material genético en una célula. Los vectores pueden ser lineales o circulares. Los vectores pueden contener una secuencia de replicación capaz de efectuar la replicación del vector en una célula huésped adecuada (es decir, un origen de replicación). Tras la transformación de un huésped adecuado, el vector puede replicarse y funcionar independientemente del genoma del huésped o integrarse en el genoma del huésped. El diseño del vector depende, entre otras cosas, del uso previsto y la célula huésped para el vector, y el diseño de un vector para un uso particular y la célula huésped está dentro del nivel de habilidad en la técnica. Los cuatro tipos principales de vectores son plásmidos, vectores víricos, cósmidos y cromosomas artificiales. Típicamente, los vectores comprenden un origen de replicación, un sitio de multiclonación y/o un marcador seleccionable. Un vector de expresión comprende típicamente un casete de expresión.

Como se usa en este documento, "casete de expresión" se refiere a una construcción polinucleotídica generada usando procedimientos recombinantes o por medios sintéticos y que comprende secuencias tampónas unidas operativamente a un polinucleótido seleccionado para facilitar la expresión del polinucleótido seleccionado en una célula huésped. Por ejemplo, las secuencias tampónas pueden facilitar la transcripción del polinucleótido seleccionado en una célula huésped, o la transcripción y traducción del polinucleótido seleccionado en una célula huésped. Un casete de expresión puede, por ejemplo, integrarse en el genoma de una célula huésped o estar presente en un vector para formar un vector de expresión.

Como se usa en el presente documento, un "vector de direccionamiento" es un constructo de ADN recombinante que típicamente comprende brazos de ADN adaptados, homólogos al ADN_g, que flanquean los elementos de un gen diana o secuencia diana de ácidos nucleicos (por ejemplo, un DSB). Un vector de direccionamiento puede comprender un polinucleótido donante. Los elementos del gen diana se pueden modificar de varias maneras, incluyendo eliminaciones y/o inserciones. Un gen objetivo defectuoso puede ser reemplazado por un gen objetivo funcional, o en la alternativa, un gen funcional puede ser eliminado. Opcionalmente, el polinucleótido donante de un vector de direccionamiento comprende un casete de selección que comprende un marcador seleccionable que se introduce en el gen diana. Las regiones de direccionamiento (es decir, las secuencias diana de ácidos nucleicos) adyacentes o dentro de un gen diana se pueden usar para afectar la regulación de la expresión génica.

Como se usa en el presente documento, los términos "ácido nucleico", "secuencia de ácidos nucleicos", "secuencia de nucleótidos", "oligonucleótido" y "polinucleótido" son intercambiables y se refieren a una forma polimérica de nucleótidos. Los nucleótidos pueden ser desoxirribonucleótidos (ADN), ribonucleótidos (ARN), análogos de los mismos o combinaciones de los mismos, y pueden ser de cualquier longitud. Los polinucleótidos pueden realizar cualquier función y pueden tener estructuras secundarias y terciarias. Los términos abarcan análogos conocidos de nucleótidos naturales y nucleótidos que se modifican en las unidades estructurales de base, azúcar y/o fosfato. Los análogos de un nucleótido particular tienen la misma especificidad de apareamiento de bases (por ejemplo, un análogo de los pares de bases A con T). Un polinucleótido puede comprender un nucleótido manipulado o múltiples nucleótidos manipulados. Ejemplos de nucleótidos manipulados incluyen nucleótidos fluorados, nucleótidos metilados y análogos de nucleótidos. La estructura del nucleótido puede modificarse antes o después de ensamblar un polímero. Después de la polimerización, los polinucleótidos pueden modificarse adicionalmente, por ejemplo, mediante conjugación con un componente de marcado o un componente de unión a diana. Una secuencia de nucleótidos puede incorporar componentes no nucleotídicos. Los términos también abarcan ácidos nucleicos que comprenden unidades estructurales o enlaces del esqueleto manipulado, que son sintéticos, naturales y no naturales, y tienen propiedades de unión similares a las de un polinucleótido de referencia (por ejemplo, ADN o ARN). Ejemplos de dichos análogos incluyen, pero no se limitan a, fosforotioatos, fosforamidatos, fosfonatos de metilo, fosfonatos de metilo quirales, ribonucleótidos 2-O-metilo, ácidos peptídicos nucleicos (PNA), ácido nucleico bloqueado (LNATM) (Exiqon, Inc., Woburn, MA) nucleósidos, ácido nucleico de glicol, ácidos nucleicos puenteados y estructuras de morfolino.

Los ácidos nucleicos peptídicos (PNA) son homólogos sintéticos de ácidos nucleicos en los que el esqueleto de polinucleótido fosfato-azúcar se reemplaza por un polímero pseudo-peptídico flexible. Las nucleobases están unidas al polímero. Los PNA tienen la capacidad de hibridar con alta afinidad y especificidad a secuencias complementarias de ARN y ADN.

En los ácidos nucleicos de fosforotioato, el enlace de fosforotioato (PS) sustituye a un átomo de azufre por un oxígeno sin puente en el esqueleto de fosfato de polinucleótido. Esta modificación hace que el enlace internucleotídico sea resistente a la degradación de las nucleasas. En algunas realizaciones, los enlaces fosforotioato se introducen entre los últimos 3 a 5 nucleótidos en el extremo 5' o 3' de una secuencia polinucleotídica para inhibir la degradación de la exonucleasa. La colocación de enlaces fosforotioato en todo un oligonucleótido también ayuda a reducir la degradación por las endonucleasas.

El ácido nucleico de treosa (TNA) es un polímero genético artificial. La estructura de el esqueleto de TNA comprende la repetición de los azúcares de treosa unidos por enlaces fosfodiéster. Los polímeros de TNA son resistentes a la degradación de las nucleasas. La TNA puede autoensamblarse mediante uniones de hidrógeno de pares de bases en estructuras dúplex.

Las inversiones de enlace pueden introducirse en polinucleótidos mediante el uso de "fosforamiditas invertidas" (véase, por ejemplo, www.ucalgary.ca/dnalab/synthesis/-modificaciones/enlaces). Un enlace 3'-3' en un extremo de un polinucleótido estabiliza el polinucleótido a la degradación de la exonucleasa creando un oligonucleótido que tiene dos extremos 5'-OH y ningún extremo 3'-OH. Típicamente, tales polinucleótidos tienen grupos fosforamida en la posición 5'-OH y un grupo protector de dimetoxitritilo (DMT) en la posición 3'-OH. Normalmente, el grupo protector de DMT está en el 5'-OH y la fosforamida está en el 3'-OH.

Las secuencias de polinucleótidos se muestran aquí en la orientación de 5' a 3' convencional, a menos que se indique lo contrario.

Como se usa en este documento, "identidad de secuencia" generalmente se refiere al porcentaje de identidad de las bases de nucleótidos o aminoácidos que comparan un primer polinucleótido o polipéptido con un segundo polinucleótido o polipéptido usando algoritmos que tienen varios parámetros de ponderación. La identidad de secuencia entre dos polinucleótidos o dos polipéptidos se puede determinar mediante el alineamiento de secuencias mediante diversos procedimientos y programas de computadora (por ejemplo, BLAST, CS-BLAST, FASTA, HMMER, LALIGN y similares) disponibles a través de la web mundial en sitios como, pero no limitado a, GENBANK (www.ncbi.nlm.nih.gov/genbank/) y EMBL-EBI (www.ebi.ac.uk). La identidad de secuencia entre dos polinucleótidos o dos secuencias polipeptídicas generalmente se calcula utilizando los parámetros estándar por defecto de los diversos procedimientos o programas de computadora. Un alto grado de identidad de secuencia, como se usa en este documento, entre dos polinucleótidos o dos polipéptidos está típicamente entre aproximadamente el 90% de identidad y el 100% de identidad, por ejemplo, aproximadamente el 90% de identidad o superior, preferiblemente aproximadamente el 95% de identidad o superior, más preferiblemente aproximadamente 98% de identidad o superior. Un grado moderado de identidad de secuencia, como se usa en este documento, entre dos polinucleótidos o dos polipéptidos está típicamente entre aproximadamente 80% de identidad y aproximadamente 85% de identidad, por ejemplo, aproximadamente 80% de identidad o superior, preferiblemente aproximadamente 85% de identidad. Un bajo grado de identidad de secuencia, como se usa en este documento, entre dos polinucleótidos o dos polipéptidos está típicamente entre aproximadamente 50% de identidad y 75% de identidad, por ejemplo, aproximadamente 50% de identidad, preferiblemente aproximadamente 60% de identidad, más preferiblemente aproximadamente 75% de identidad. Por ejemplo, una proteína Cas (por ejemplo, un Cas9 que comprende sustituciones de aminoácidos) puede tener un bajo grado de identidad de secuencia, un grado moderado de identidad de secuencia o un alto grado de identidad de secuencia, a lo largo de su longitud hasta una proteína Cas de referencia (por ejemplo, un Cas9 de tipo salvaje). Como otro ejemplo, una NATNA puede tener un bajo grado de identidad de secuencia, un grado moderado de identidad de secuencia o un alto grado de identidad de secuencia en su longitud en comparación con un polinucleótido de tipo salvaje de referencia que se compleja con la proteína Cas de referencia (por ejemplo, un ARNsg que forma un complejo con Cas9).

Como se usa en este documento, "hibridación" o "hibridar" o "que hibrida" es el proceso de combinar dos moléculas de ADN o ARN complementarias de cadena sencilla para formar una sola molécula de cadena doble (ADN/ADN, ADN/ARN, ARN/ARN) a través del apareamiento de bases de hidrógeno. La restricción de la hibridación se determina típicamente por la temperatura de hibridación y la concentración de sal del tampón de hibridación; por ejemplo, alta temperatura y baja sal proporcionan condiciones de hibridación de alta restricción. Ejemplos de los rangos de concentración de sal y los rangos de temperatura para diferentes condiciones de hibridación son los siguientes: alta restricción, aproximadamente de 0,01 M a aproximadamente 0,05 M de sal, temperatura de hibridación de 5°C a 10°C por debajo de T_m ; restricción moderada, aproximadamente 0,16 M a aproximadamente 0,33 M de sal, temperatura de hibridación 20°C a 29°C por debajo de T_m ; y baja restricción, aproximadamente de 0,33 M a aproximadamente 0,82 M de sal, temperatura de hibridación de 40°C a 48°C por debajo de T_m . La T_m de los ácidos nucleicos dúplex se calcula mediante procedimientos estándar bien conocidos en la técnica (véase, por ejemplo, Maniatis, T., et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press: Nueva York (1982); Casey, J., et al., *Nucleic Acids Research* 4: 1539-1552 (1977); Bodkin, D.K., et al., *Journal of Virological Methods* 10(1): 45-52 (1985); Wallace, R.B., et al., *Nucleic Acids Research* 9(4): 879-894 (1981)). Las herramientas de predicción de algoritmos para estimar T_m también están ampliamente disponibles. Las condiciones de alta restricción para la hibridación se refieren típicamente a condiciones en las que un polinucleótido complementario de una secuencia diana se hibrida predominantemente con la secuencia diana, y sustancialmente no hibrida con secuencias no diana. Típicamente, las condiciones de hibridación son de restricción moderada, preferiblemente de alta restricción.

Como se usa en este documento, un "elemento de tallo" o "estructura del tallo" se refiere a un polinucleótido que comprende dos cadenas que se sabe o se predice que forman una región de doble hebra (el "elemento de tallo"). Un "elemento de tallo-bucle" o "estructura de tallo-bucle" se refiere a una estructura de tallo en la que el extremo 3' de una hebra está unido covalentemente al extremo 5' de la segunda hebra por una secuencia de nucleótidos de nucleótidos típicamente monocatenarios ("una secuencia de nucleótidos del elemento de tallo-bucle"). En algunas realizaciones, el elemento de bucle comprende una secuencia de nucleótidos de elemento de bucle de entre

aproximadamente 3 y aproximadamente 20 nucleótidos de longitud, preferiblemente entre aproximadamente 4 y aproximadamente 10 nucleótidos de longitud. En realizaciones preferidas, una secuencia de nucleótidos del elemento de bucle es una secuencia de nucleótidos monocatenaria de bases de ácidos nucleicos no apareadas que no interactúan a través de la formación de enlaces de hidrógeno para crear un elemento de tallo dentro de la secuencia de nucleótidos del elemento de bucle. El término "elemento de horquilla" también se usa en este documento para referirse a estructuras de tallo-bucle. Tales estructuras son bien conocidas en la técnica. El emparejamiento de la base puede ser exacto; sin embargo, como se conoce en la técnica, un elemento de tallo no requiere un emparejamiento de bases exacto. Por lo tanto, el elemento de tallo puede incluir una o más no coincidencias de bases o bases no emparejadas.

Una "secuencia de nucleótidos del elemento enlazador" y "secuencia de nucleótidos enlazante" se usan indistintamente en el presente documento y se refieren a una secuencia monocatenaria de uno o más nucleótidos unidos covalentemente a un extremo 5', un extremo 3', o a ambos. Los extremos 5' y 3' de una primera secuencia de polinucleótidos, y típicamente se refieren a una secuencia de ácidos nucleicos monocatenario que conecta una primera secuencia de polinucleótidos con una segunda secuencia de polinucleótidos. En realizaciones preferidas, la secuencia de nucleótidos del elemento enlazador es una secuencia de nucleótidos monocatenaria de bases de ácidos nucleicos no apareadas que no interactúan a través de la formación de enlaces de hidrógeno para crear un elemento troncal dentro de la secuencia de nucleótidos del elemento enlazador. En algunas realizaciones, una secuencia de nucleótidos del elemento enlazador tiene una longitud de entre aproximadamente 1 y aproximadamente 20 nucleótidos, preferiblemente entre aproximadamente 2 y aproximadamente 10 nucleótidos.

Como se usa en el presente documento, el término "aminoácido" se refiere a aminoácidos naturales y sintéticos (no naturales), que incluyen análogos de aminoácidos, aminoácidos manipulados, peptidomiméticos, glicina e isómeros ópticos D o L.

Como se usa en el presente documento, los términos "péptido", "polipéptido" y "proteína" son intercambiables y se refieren a polímeros de aminoácidos. Un polipéptido puede ser de cualquier longitud. Puede ser ramificado o lineal, puede estar interrumpido por no aminoácidos y puede comprender aminoácidos manipulados. Los términos también se refieren a un polímero de aminoácido que se ha manipulado mediante, por ejemplo, acetilación, formación de enlaces disulfuro, glicosilación, lipidación, fosforilación, pegilación, biotilación, entrecruzamiento y/o conjugación (por ejemplo, con un componente de etiquetado o ligando). Las secuencias de polipéptidos se muestran en este documento en la orientación N-terminal convencional a C-terminal, a menos que se indique lo contrario.

Los polipéptidos y polinucleótidos pueden prepararse usando técnicas de rutina en el campo de la biología molecular (véase, por ejemplo, los textos estándar discutidos anteriormente). Además, esencialmente cualquier polipéptido o polinucleótido está disponible de fuentes comerciales.

Los términos "proteína de fusión" y "proteína quimérica", como se usan en este documento, se refieren a una proteína única creada al unir dos o más proteínas, dominios de proteínas o fragmentos de proteínas que no se producen de forma natural en una sola proteína. Por ejemplo, una proteína de fusión puede contener un primer dominio de una proteína Cas9 y un segundo dominio una proteína Csy4. La modificación para incluir dichos dominios en la proteína de fusión puede conferir actividad adicional sobre los polipéptidos dirigidos al sitio manipulados. Estas actividades pueden incluir actividad de nucleasa, actividad de metiltransferasa, actividad de desmetilasa, actividad de reparación del ADN, actividad de daño del ADN, actividad de desaminación, actividad de dismutasa, actividad de alquilación, actividad de depuración, actividad de oxidación, actividad formadora del dímero de pirimidina, actividad de integrasa, actividad de transposasa, actividad de recombinasa, actividad de la polimerasa, actividad ligasa, actividad helicasa, actividad fotolítica, actividad glucosilasa, actividad acetiltransferasa, actividad desacetilasa, actividad quinasa, actividad fosfatasa, actividad ligasa de ubiquitina, actividad desubiquitinante, actividad adenilación, actividad de la muertos de la, actividad de SUMOilación, actividad de deSUMOilación, actividad de ribosilación, actividad de deribosilación, actividad de miristoilación o actividad de desmiristoilación) que modifica un polipéptido asociado con la secuencia diana de ácidos nucleicos (por ejemplo, una histona). Una proteína de fusión también puede comprender etiquetas de epítipo (por ejemplo, etiquetas de histidina, etiquetas FLAG® (Sigma Aldrich, St. Louis, MO), etiquetas de Myc), secuencias de proteínas informadoras (por ejemplo, glutatión-S-transferasa, beta-galactosidasa, luciferasa, proteína fluorescente verde, proteína fluorescente cian, proteína fluorescente amarilla) y/o dominios de unión a ácidos nucleicos (por ejemplo, un dominio de unión a ADN, un dominio de unión a ARN). Una proteína de fusión también puede comprender dominios activadores (por ejemplo, factores de transcripción de choque térmico, activadores NFκB) o dominios represores (por ejemplo, un dominio KRAB). Según lo descrito por Lupo, A., et al., *Current Genomics* 14(4): 268-278 (2013), el dominio KRAB es un potente módulo de represión transcripcional y está ubicado en la secuencia amino-terminal de la mayoría de las proteínas de dedos de zinc C2H2 (véase, por ejemplo, Margolin, J., et al., *Proceedings of the National Academy of Sciences of the United States of America* 91: 4509-4513 (1994); Witzgall, R., et al., *Proceedings of the National Academy of Sciences of the United States of America* 91: 4514-4518 (1994)). El dominio KRAB generalmente se une a proteínas correpressoras y/o factores de transcripción a través de interacciones proteína-proteína, causando la represión transcripcional de los genes a los de unión las proteínas KRAB (KRAB-ZFP) (véase, por ejemplo, Friedman J.R., et al., *Genes & Development* 10: 2067-2678 (1996)). En algunas realizaciones, las secuencias de ácidos nucleicos enlazador se usan para unir las dos o más proteínas, dominios de proteínas o fragmentos de proteínas.

Una "unidad estructural", como se usa en el presente documento, se refiere a una porción de una molécula. Una unidad estructural puede ser un grupo funcional o describir una porción de una molécula con múltiples grupos funcionales (por ejemplo, que comparten aspectos estructurales comunes). Los términos "unidad estructural" y "grupo funcional" se usan típicamente de manera intercambiable; sin embargo, un "grupo funcional" puede referirse más específicamente a una porción de una molécula que comprende algún comportamiento químico común. "Unidad estructural" se utiliza a menudo como una descripción estructural. En algunas realizaciones, un término 5', un término 3' o un término 3' (por ejemplo, un término 5' no nativo y/o un término 3' no nativo en un primer elemento de tallo).

El término "etiqueta de afinidad", como se usa en el presente documento, se refiere típicamente a una o más unidades estructurales que aumentan la afinidad de unión de un componente polinucleótido de una composición de polinucleótido NASC, por ejemplo, para facilitar la formación de un complejo NASC. Algunas realizaciones de la presente invención usan una "secuencia de afinidad", que es una secuencia de polinucleótidos que comprende una o más etiquetas de afinidad. En algunas realizaciones de la presente invención, un componente polinucleotídico comprende además una secuencia de afinidad localizada en el extremo 5', el extremo 3', o situada entre el extremo 5' y el extremo 3'. Algunas realizaciones de la presente invención introducen una o más etiquetas de afinidad en el extremo N-terminal de una secuencia de la proteína Cas (por ejemplo, una secuencia de la proteína Cas9), en el extremo C-terminal de una secuencia de la proteína Cas, en una posición ubicada entre el N-terminal y C-terminal de una secuencia de la proteína Cas, o combinaciones de las mismas. En algunas realizaciones de la invención, el polipéptido Cas se modifica con una etiqueta de afinidad o una secuencia de afinidad. Una amplia variedad de etiquetas de afinidad se describe en la Solicitud de Patente Publicada de los Estados Unidos No. 2014-0315985, publicada el 23 de octubre de 2014.

Como se usa en este documento, un "entrecruzamiento" es un enlace que une una cadena de polímero (por ejemplo, un polinucleótido o polipéptido) a otra. Tales enlaces pueden ser enlaces covalentes o enlaces iónicos. En algunas realizaciones, un polinucleótido puede unirse a otro polinucleótido mediante la entrecruzamiento de los polinucleótidos. En otras realizaciones, un polinucleótido puede estar entrecruzado con un polipéptido. En realizaciones adicionales, un polipéptido puede estar entrecruzado con un polipéptido.

El término "unidad estructural de entrecruzamiento", como se usa en el presente documento, se refiere típicamente a una unidad estructural adecuada para proporcionar entrecruzamiento entre componentes de polinucleótido de una composición de polinucleótido de NASC. Una unidad estructural de entrecruzamiento es otro ejemplo de una etiqueta de afinidad.

Los términos "ligando" y "unidad estructural de unión a ligando", como se usan en este documento, se refieren a unidades estructurales que facilitan la unión de componentes polinucleotídicos para formar una composición de polinucleótido NASC. Los ligandos y los unidades estructurales de unión a ligando son etiquetas de afinidad apareadas.

Como se usa en este documento, una "célula huésped" generalmente se refiere a una célula biológica. Una célula es la unidad básica estructural, funcional y/o biológica de un organismo. Una célula puede originarse de cualquier organismo que tenga una o más células. Ejemplos de células huésped incluyen, pero no se limitan a, una célula procariota, una célula eucariota, una célula bacteriana, una célula arqueal, una célula de un organismo eucariótico de una sola célula, una célula protozoaria, una célula de una planta (por ejemplo, células de cultivos de plantas (como soja, tomates, remolacha azucarera, calabaza, heno, cannabis, tabaco, plátanos, ñames, batatas, yuca, papas, trigo, sorgo, soja, arroz, cereal, maíz, Brassica productora de aceite (por ejemplo, colza y canola que produce aceite), algodón, caña de azúcar, girasol, mijo y alfalfa), frutas, vegetales, granos, semillas, plantas con flores, coníferas, gimnospermas, helechos, licopodios, milhojas, hepáticas, musgos), algas células (por ejemplo, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh* y similares), algas marinas (por ejemplo, kelp), una célula fúngica (por ejemplo, una célula de levadura o una célula de un hongo), una célula animal, una célula de un animal invertebrado (por ejemplo, mosca de la fruta, cnidaria, equinodermo, nematodo, y similares), una célula de un animal vertebrado (por ejemplo, peces, anfibios, reptiles, aves o mamíferos), una célula de un mamífero (por ejemplo, un cerdo, una vaca, una cabra, una oveja, un roedor, una rata, un ratón, un primate no humano, un humano y similares). Además, una célula puede ser una célula madre o una célula progenitora.

Como se usa en este documento, "célula madre" se refiere a una célula que tiene la capacidad de autorrenovación, es decir, la capacidad de pasar por numerosos ciclos de división celular mientras se mantiene el estado indiferenciado. Las células madre pueden ser totipotentes, pluripotentes, multipotentes, oligopotentes o unipotentes. Las células madre pueden ser células madre pluripotentes embrionarias, fetales, amnióticas, adultas o inducidas.

Como se usa en este documento, "células madre pluripotentes inducidas" se refiere a un tipo de célula madre pluripotente que se deriva artificialmente de una célula no pluripotente, típicamente una célula somática adulta, mediante la inducción de la expresión de genes específicos.

"Planta", como se usa en el presente documento, se refiere a plantas completas, órganos de plantas, tejidos de plantas, germoplasma, semillas, células de plantas y progenie de las mismas. Las células vegetales incluyen, sin

limitación, células de semillas, cultivos en suspensión, embriones, regiones meristemáticas, tejido de callo, hojas, raíces, brotes, gametofitos, esporofitos, polen y microesporas. Las partes de la planta incluyen tejidos diferenciados e indiferenciados que incluyen, entre otros, raíces, tallos, brotes, hojas, pólenes, semillas, tejido tumoral y diversas formas de células y cultivos (por ejemplo, células individuales, protoplastos, embriones y tejido calloso). El tejido de la planta puede estar en la planta o en un órgano, tejido o cultivo celular de la planta. "Órgano vegetal" se refiere a tejido vegetal o un grupo de tejidos que constituyen una parte morfológica y funcionalmente distinta de una planta.

"Sujeto", como se usa en el presente documento, se refiere a cualquier miembro del phylum Chordata, incluidos, entre otros, los seres humanos y otros primates, incluidos los primates no humanos, como los macacos rhesus, los chimpancés y otras especies de monos y simios; animales de granja, como vacas, ovejas, cerdos, cabras y caballos; mamíferos domésticos, tales como perros y gatos; animales de laboratorio, incluidos conejos, ratones, ratas y cobayas; aves, incluyendo aves domésticas, salvajes y de caza, como pollos, pavos y otras aves gallináceas, patos y gansos; y similares. El término no denota una edad o género en particular. Por lo tanto, el término incluye a individuos adultos, jóvenes y recién nacidos, así como a hombres y mujeres. En algunas realizaciones, una célula huésped se deriva de un sujeto (por ejemplo, células madre, células progenitoras o células específicas de tejido). En algunas realizaciones, el sujeto no es un sujeto humano.

Como se usa en el presente documento, "organismo transgénico" se refiere a un organismo cuyo genoma está genéticamente manipulado. El término incluye la progenie (cualquier generación) de un organismo transgénico, siempre que la progenie tenga la modificación genética.

Como se usa en este documento, "aislado" puede referirse a un ácido nucleico o polipéptido que, por intervención humana, existe aparte de su entorno nativo y, por lo tanto, no es un producto de la naturaleza. Un ácido nucleico o polipéptido aislado puede existir en una forma purificada y/o puede existir en un entorno no nativo tal como, por ejemplo, en una célula recombinante.

En un aspecto general de la presente invención, una composición de polinucleótido NASC comprende un elemento de repetición 1 conectado con un elemento de repetición 2, un elemento de unión a proteína de unión a ácidos nucleicos 1 y un elemento de unión de proteína de unión a ácidos nucleicos 2 y un elemento espaciador 1 y un elemento espaciador 2.

El elemento de repetición 1 y el elemento de repetición dos están típicamente conectados mediante enlaces covalentes, enlaces no covalentes o una combinación de enlaces covalentes y no covalentes. En algunas realizaciones, el elemento de repetición 1 y el elemento de repetición 2 están conectados por pares de bases con enlaces de hidrógeno.

La composición de polinucleótidos de NASC es capaz de asociarse con proteínas de unión a ácidos nucleicos para formar un complejo de nucleoproteínas. En algunas realizaciones, se usan dos o más proteínas de unión a ácidos nucleicos que tienen motivos estructurales y motivos funcionales similares para formar complejos de nucleoproteínas con composiciones de polinucleótidos de NASC. En realizaciones preferidas, las proteínas de unión a ácidos nucleicos son proteínas CRISPR-Cas9 de Clase 2 Tipo II. En algunas realizaciones, la proteína CRISPR-Cas9 Clase 2 Tipo II se une a una secuencia de unión a la proteína de unión a ácidos nucleicos de doble cadena ("una proteína de unión a ácidos nucleicos de doble cadena").

Una composición de polinucleótido NASC/proteína de unión a ácidos nucleicos 1/complejo de proteína de unión a ácidos nucleicos 2 es capaz de unirse preferentemente a una secuencia diana de ácidos nucleicos en un polinucleótido (en relación con un polinucleótido que no comprende la secuencia diana de ácidos nucleicos).

La Figura 4A, Figura 4B, Figura 4C, Figura 4D, Figura 4E, Figura 4F, Figura 4G, Figura 4H, Figura 4I, Figura 4J, Figura 4K, Figura 4L, Figura 4M, Figura 4N, y Figura 4O ilustran ejemplos genéricos de diferentes tipos de andamios de ácidos nucleicos. Estas Figuras presentan ubicaciones relativas de diferentes elementos en secuencias de ácidos nucleicos manipuladas para formar andamios.

En algunos aspectos, un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio comprende:

un primer ácido nucleico manipulado que comprende (i) un elemento de unión a proteína de unión a ácidos nucleicos 1 que comprende una secuencia de unión a proteína de unión a ácidos nucleicos 1 (por ejemplo, una secuencia de unión a proteína de unión a ácidos nucleicos bicatenarios 1) que tiene un primer extremo y un segundo extremo, (ii) un elemento de repetición 1 que comprende una secuencia 1 de ácidos nucleicos repetida que tiene un primer extremo y un segundo extremo, y (iii) un elemento espaciador 1 que comprende una secuencia 1 de unión a diana de ácidos nucleicos; y

un segundo ácido nucleico manipulado que comprende (i) un elemento de unión a la proteína de unión a ácidos nucleicos 2 que comprende una secuencia 2 de unión a la proteína de unión a ácidos nucleicos (por ejemplo, una secuencia de unión a la proteína de unión a ácidos nucleicos bicatenarios 2) que tiene un primer extremo y un segundo extremo, (ii) un elemento de repetición 2 que comprende una secuencia 1C de ácidos nucleicos repetida

que tiene un primer extremo y un segundo extremo, y (iii) un elemento espaciador 2 que comprende una secuencia 2 de unión a diana de ácidos nucleicos.

La secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida son complementarias. La secuencia 1C de ácidos nucleicos repetida también se conoce como una secuencia 2 de ácidos nucleicos repetida. La secuencia 1 de ácidos nucleicos repetida está conectada con la secuencia 1C de ácidos nucleicos repetida a través de pares de bases con enlaces de hidrógeno.

5

La Tabla 2 presenta una serie de indicadores utilizados consistentemente en la Figura 4A a 4O.

Tabla 2

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio			
Indicador y región correspondiente			
Secuencia de ácidos nucleicos	Elemento general	El elemento comprende	Ejemplo de componente elemento
un primer ácido nucleico manipulado			
1-1	un elemento de repetición 1	una la secuencia 1 de ácidos nucleicos repetida	una repetición de la secuencia de ácidos nucleicos
1-2	un elemento de unión a la proteína de unión a ácidos nucleicos 1	una secuencia 1 de unión a la proteína de unión a ácidos nucleicos	una secuencia 1 de unión a la proteína de unión a ácidos nucleicos de doble cadena
Indicador y región correspondiente			
1-3	un elemento espaciador 1	una secuencia 1 de ácidos nucleicos	una secuencia 1 de unión a diana de ácidos nucleicos
un segundo ácido nucleico manipulado			
2-1	un elemento de repetición 2	una repetición de la secuencia 2 de ácidos nucleicos	una repetición de la secuencia de ácidos nucleicos
2-2	un elemento de unión a la proteína de unión a ácidos nucleicos 2	una secuencia 2 de unión a la proteína de unión a ácidos nucleicos	una secuencia 2 de unión a la proteína de unión a ácidos nucleicos de doble cadena
2-3	un elemento espaciador 2	una secuencia 2 de ácidos nucleicos	una secuencia 2 de unión a diana de ácidos nucleicos
un tercer ácido nucleico manipulado			
3-1	un elemento de repetición 3	una repetición de la secuencia 3 de ácidos nucleicos	una repetición de la secuencia de ácidos nucleicos

(continuación)

Indicador y región correspondiente			
3-2	un elemento de unión a la proteína de unión a ácidos nucleicos 3	una secuencia 3 de unión a la proteína de unión a ácidos nucleicos	una secuencia 3 de unión a la proteína de unión a ácidos nucleicos de doble cadena
3-3	un elemento espaciador 3	una secuencia 3 de ácidos nucleicos	una secuencia 3 de unión a diana de ácidos nucleicos
ácido nucleico manipulados adicionales			
#-1	un elemento de repetición #	una secuencia # de ácidos nucleicos repetida	una secuencia de ácidos nucleicos repetida
#-2	un elemento de unión a la proteína de unión a ácidos nucleicos #	una secuencia # de unión a la proteína de unión a ácidos nucleicos	una secuencia # de unión a la proteína de unión a ácidos nucleicos de doble cadena
#-3	un elemento espaciador #	una secuencia # de ácidos nucleicos	una secuencia # de unión a un ácido nucleico
Una flecha en la figura corresponde a un sitio que puede comprender secuencias de ácidos nucleicos adicionales, como una secuencia de ácidos nucleicos de elemento enlazador, y un par de flechas ilustra los límites de un elemento en particular (por ejemplo, las flechas que flanquean la región 1-1 en Figura 4A).).			
# = para ácidos nucleicos manipulados adicionales, numeración secuencial después del número 3			

Figura 4A, Figura 4C, Figura 4E, Figura 4G, Figura 4I, y Figura 4K, cada una presenta un ejemplo de una colección de seis disposiciones diferentes de la región 1-1, la región 1-2 y la región 1-3 dentro de un primer ácido nucleico manipulado, y la región 2-1, la región 2-2 y la región 2-3 dentro de un segundo ácido nucleico manipulado, en el que la secuencia 1-1 de ácidos nucleicos repetida se asocia con la secuencia 2-1 de ácidos nucleicos repetida a través de enlaces de hidrógeno entre la secuencia 1-1 de ácidos nucleicos repetida y la secuencia 2-1 de ácidos nucleicos repetida. En estas Figuras, el primer ácido nucleico manipulado es un polinucleótido único y el segundo ácido nucleico manipulado es un solo polinucleótido. Cada polinucleótido tiene un primer extremo y un segundo extremo. En algunos aspectos, el primer extremo es un extremo 5' y el segundo extremo es un extremo 3'. En otros aspectos, el primer extremo es un extremo 3' y el segundo extremo es un extremo 5'.

Figura 4B, Figura 4D, Figura 4F, Figura 4H, Figura 4J, y Figura 4L cada una presenta la misma disposición que la Figura 4A, Figura 4C, Figura 4E, Figura 4G, Figura 4I, y Figura 4K, respectivamente, en el que el primer ácido nucleico manipulado comprende múltiples polinucleótidos asociados por enlace de hidrógeno (indicado en estas Figuras como múltiples líneas rectas entre secuencias de ácidos nucleicos) y el segundo ácido nucleico manipulado comprende múltiples polinucleótidos asociados mediante enlace de hidrógeno. Cada polinucleótido tiene un primer extremo y un segundo extremo. En algunos aspectos, el primer extremo es un extremo 5' y el segundo extremo es un extremo 3', en el que se mantiene una orientación estándar de 5' a 3' entre los polinucleótidos. En otros aspectos, el primer extremo es un extremo 3' y el segundo extremo es un extremo 5', en el que se mantiene una orientación estándar de 5' a 3' entre los polinucleótidos.

La Figura 4M ilustra un ejemplo de un complejo de tres secuencias de ácidos nucleicos manipuladas que forman un andamio. En esta Figura, el primer, segundo y tercer ácidos nucleicos diseñados son cada uno un solo polinucleótido. El primer y segundo ácidos nucleicos manipulados se corresponden con el primer y segundo ácidos nucleicos manipulados genéticamente presentados en la Figura 4A, y el tercer ácido nucleico manipulado corresponde al segundo ácido nucleico manipulado de la Figura 4G.

La Figura 4N ilustra un ejemplo de un complejo de tres secuencias de ácidos nucleicos manipuladas que forman un andamio. En esta Figura, el primer y tercer ácidos nucleicos diseñados son cada uno un polinucleótido único. El segundo ácido nucleico manipulado comprende múltiples polinucleótidos asociados a través de enlaces de hidrógeno. El primer ácido nucleico manipulado corresponde al primer ácido nucleico manipulado presentado en la Figura 4C. El segundo ácido nucleico manipulado corresponde al segundo ácido nucleico manipulado presentado en

la Figura 4D. El tercer ácido nucleico manipulado corresponde al segundo ácido nucleico manipulado de la Figura 4E.

La Figura 4O ilustra un ejemplo de un complejo de tres secuencias de ácidos nucleicos manipuladas que forman un andamio. En esta Figura, el primer y tercer ácidos nucleicos diseñados son cada uno un polinucleótido único. El segundo ácido nucleico manipulado comprende múltiples polinucleótidos asociados a través de enlaces de hidrógeno. El primer ácido nucleico manipulado corresponde al primer ácido nucleico manipulado presentado en la Figura 4I. El segundo ácido nucleico manipulado corresponde al segundo ácido nucleico manipulado presentado en la Figura 4F. El tercer ácido nucleico manipulado corresponde al segundo ácido nucleico manipulado de la Figura 4K.

La presente divulgación comprende una amplia variedad de andamios basados en ácido nucleico que están compuestos por un complejo de dos o más secuencias de ácidos nucleicos manipuladas. En realizaciones preferidas, las secuencias de ácidos nucleicos manipuladas comprenden elementos de ácidos nucleicos de clase 2 CRISPR que se dirigen a ácidos nucleicos, por ejemplo, elementos que codifican secuencias de ácidos nucleicos basadas en las secuencias de ARN de guía simple ARNcr tipo 2, CRISPR-traARN-crs tipo 2 y CRISPR tipo 2. Ejemplos de elementos asociados a CRISPR de Clase 2 incluyen, entre otros, los elementos presentados en Figura 1A, Figura 1B, Figura 1C, Figura 1D, Figura 2A, Figura 2B, Figura 3A, y Figura 3B.

En algunos aspectos, los andamios de ácidos nucleicos comprenden secuencias de unión a proteínas de ácidos nucleicos que incluyen, pero no se limitan a, aquellas asociadas con los sistemas de edición del genoma (por ejemplo, nucleasas de dedos de zinc (ZFN)), nucleasas basadas en efector similares a activadores de transcripción (TALEN), meganucleasas, y CRISPR-Cas). Ejemplos de secuencias de unión a proteínas de ácidos nucleicos incluyen, pero no se limitan a, las asociadas con las siguientes proteínas de unión a ácidos nucleicos: proteínas de unión a ácidos nucleicos CRISPR de tipo 2 (por ejemplo, proteína Cpf1, proteína dCpf1 (catalíticamente inactiva), proteína Cas9 y/o proteína dCas9 (catalíticamente inactiva)); proteínas argonautas; proteínas de unión a ácidos nucleicos de doble cadena (por ejemplo, proteína Csy4 y/o proteína Csy4* (inactiva catalíticamente); ver, por ejemplo, Haurwitz, R., et al., Science 329 (5997): 1355-1358 (2010); Sternberg, S., et al., ARN 18(4): 661-672 (2012); Patente de los Estados Unidos No. 9,115,348); proteínas de unión a ARN monocatenarias (por ejemplo, proteína de unión a ARNsi p19); proteínas de unión a ADN monocatenarias (por ejemplo, DBP de adenovirus, SSB termoestable extrema (proteína de unión a ADN monocatenaria); proteínas de unión a ARN bicatenarias (por ejemplo, DICER); proteínas de unión a ADN bicatenarias (por ejemplo, ZFNs); e híbridos de ARN/ADN de doble cadena (por ejemplo, ribonucleasa H), así como versiones catalíticamente inactivas de los mismos. En realizaciones adicionales, los andamios de ácidos nucleicos y las proteínas de unión a ácidos nucleicos asociadas están en complejos de proteínas de unión de andamio de ácidos nucleicos/ácido nucleico, por ejemplo, complejos de nucleoproteínas y complejos de ribonucleoproteínas.

En algunos aspectos, cada una de las secuencias de unión a la proteína de unión a ácidos nucleicos de 1-2, 2-2 y/o 3-2 es, por ejemplo, una secuencia de unión a la proteína de unión a ADN bicatenaria, una unión de ADN de cadena sencilla secuencia de unión a proteínas, una secuencia de unión a proteína de unión a ARN bicatenaria, una secuencia de unión a proteína de unión a ARN monocatenaria, o una secuencia de unión a proteína de unión híbrida a ADN/ARN bicatenaria. En aspectos preferidos, la proteína de unión a ácidos nucleicos de unión a la secuencia de unión a proteína de unión a ácidos nucleicos es una proteína Cas9 o una proteína Cpf1.

En aspectos particulares, cada una de la secuencia 1-1 de ácidos nucleicos, secuencia 1-2 de ácidos nucleicos y/o secuencia 1-3 de ácidos nucleicos comprende una secuencia de ácidos nucleicos de unión a una secuencia de ácidos nucleicos diana (por ejemplo, un elemento espaciador).

En un primer aspecto de la presente divulgación, una composición de polinucleótido NASC comprende una NASC-PC1 y NASC-PC2. Un complejo NASC-PC1/NASC-PC2 comprende un elemento de repetición 1 conectado con un elemento de repetición 2, un elemento de unión a la proteína de unión a ácidos nucleicos bicatenarios 1 y un elemento de unión a la proteína de unión a ácidos nucleicos bicatenarios 2, y un elemento espaciador 1 y un elemento espaciador 2. Las proteínas de unión a ácidos nucleicos de doble cadena que son capaces de unirse a la NASC son una o más proteínas de Clase 2 Tipo V CRISPR-Cpf1.

La composición de polinucleótidos de NASC es capaz de asociarse con dos proteínas de Clase 2 Tipo V CRISPR-Cpf1 para formar un complejo de nucleoproteínas. En algunos aspectos, cada una de las composiciones de polinucleótido NASC-PC1 y NASC-PC2 de NASC es capaz de asociarse con dos proteínas de Clase 2 Tipo V de CRISPR-Cpf1 para formar un complejo de nucleoproteínas (por ejemplo, Figura 5A, Figura 5B, Figura 5C, Figura 5D, Figura 5E, Figura 5F y Figura 5G).

En el primer aspecto de la presente descripción, el elemento de repetición 1 comprende una secuencia 1 de ácidos nucleicos repetida, el elemento de repetición 2 comprende una secuencia 1C de ácidos nucleicos repetida, el elemento de unión a la proteína de unión a ácidos nucleicos 1 comprende una proteína de unión a ácidos nucleicos de doble cadena secuencia de unión 1, el elemento de unión a la proteína de unión al ácido nucleico 2 comprende una secuencia de unión a la proteína de unión al ácido nucleico bicatenario 2, el elemento espaciador 1 comprende

una secuencia de unión a la diana del ácido nucleico 1, y el elemento espaciador 2 comprende una secuencia de unión a la diana del ácido nucleico 2.

Las disposiciones de los elementos son típicamente las siguientes: (i) el elemento de repetición 1 es 5' del elemento de unión a la proteína de unión a ácidos nucleicos 1, el elemento de unión a la proteína de unión a ácidos nucleicos 1 es 5' del elemento espaciador 1, y el elemento de repetición 2 es 5' del elemento de unión a la proteína de unión a ácidos nucleicos 2, y el elemento de unión a la proteína de unión a ácidos nucleicos es 5' del elemento espaciador 2; o (ii) el elemento de unión a la proteína de unión al ácido nucleico 1 es 5' del elemento de repetición 1, la repetición del elemento 1 es a 5' del elemento espaciador 1, el elemento de unión a la proteína de unión al ácido nucleico 2 es 5' de la repetición del elemento 2 y el elemento de repetición 2 es 5' del elemento espaciador 2; o (iii) el elemento de unión a la proteína de unión al ácido nucleico 1 es 5' del elemento espaciador 1, el elemento de separación 1 es 5' a la repetición del elemento 1, el elemento de unión a la proteína de unión al ácido nucleico 2 es 5' del elemento espaciador 2, y el elemento espaciador 2 es 5' del elemento de repetición 2.

En algunos aspectos, (i) el elemento de unión a la proteína de unión a ácidos nucleicos 1 comprende una primera secuencia 1-1 de ácidos nucleicos del elemento de tallo y una primera secuencia 1-2 de ácidos nucleicos del elemento de tallo, y la primera secuencia 1-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo forma un primer elemento de tallo 1 a través de pares de bases unidos por hidrógeno, y/o (ii) el elemento de unión a la proteína de unión al ácido nucleico 2 comprende una primera secuencia 2-1 de ácidos nucleicos del elemento de tallo y una la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo, y la primera secuencia 2-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo forman un primer elemento de tallo 1 a través de pares de bases con enlaces de hidrógeno (por ejemplo, Figura 5B, Figura 5F, Figura 5G). En realizaciones adicionales, la primera secuencia 1-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo están conectadas por una secuencia de ácidos nucleicos del elemento del bucle 1 para formar un primer elemento 1 del tallo-bucle, y/o la primera la secuencia 2-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo están conectadas por una secuencia de ácidos nucleicos del elemento del bucle 2 para formar un primer elemento de tallo-bucle 2 (por ejemplo, Figura 5A, Figura 5C, Figura 5D, Figura 5E).

En aspectos adicionales, la secuencia 1 de ácidos nucleicos repetida está conectada con la secuencia 1C de ácidos nucleicos repetida a través de pares de bases con enlaces de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida.

En otro aspecto, la secuencia 1C de ácidos nucleicos repetida comprende además una etiqueta de afinidad 1 y la secuencia 2 de ácidos nucleicos repetida comprende además una etiqueta de afinidad 2, y la etiqueta de afinidad 1 está conectada con la etiqueta de afinidad 2. Por ejemplo, la secuencia 1C de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos del sitio de unión a la proteína efectora y la secuencia 2 de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos del sitio de unión a la proteína efectora, y un sitio de unión del efector 1 está formado por un enlace de par de bases a través de hidrógeno entre la unión de la proteína efectora la secuencia 1 de ácidos nucleicos del sitio y la secuencia de ácidos nucleicos del sitio de unión a la proteína efectora. Un ejemplo de un sitio de unión del efector es un sitio de unión a la proteína Csy4.

La Tabla 3 presenta una serie de indicadores usados consistentemente en Figura 5A, Figura 5B, Figura 5C, Figura 5D, Figura 5E, Figura 5F, Figura 5G, Figura 5H, y Figura 5I.

Tabla 3

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
500-507	corresponde a una primera secuencia de ácidos nucleicos manipulada
500	primer componente de ácidos nucleicos manipulados
501	un elemento de unión a la proteína de unión a ácidos nucleicos 1
502	un elemento de unión a la proteína de unión a ácidos nucleicos de doble cadena 1

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
501-502 corresponde a una primera secuencia 1-1 de ácidos nucleicos del elemento de tallo
501-520 corresponde a una secuencia de la mitad del tallo 1-1a del sitio de unión a la proteína CRISPR Tipo V Clase 2
520-502 corresponde a una secuencia de la mitad del tallo 1-1b del sitio de unión a la proteína CRISPR Tipo V Clase 2
502-503 corresponde a una secuencia 1 de ácidos nucleicos de elemento de bucle
503-504 corresponde a una primera secuencia 1-2 de ácidos nucleicos del elemento de tallo
un elemento de repetición 1 ¹
504-507 corresponde a una secuencia 1 de ácidos nucleicos repetida
504-505 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento enlazador
505-516 corresponde a una secuencia 1a de ácidos nucleicos repetida
516-517 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento enlazador
517-506 corresponde a una secuencia 1b de ácidos nucleicos repetida
506-507 corresponde a una secuencia 1-3 de ácidos nucleicos del elemento enlazador
un elemento espaciador 1
501-500 corresponde a una secuencia 1 de unión a diana de ácidos nucleicos
508-515 corresponde a una segunda secuencia de ácidos nucleicos manipulada
un segundo componente de ácidos nucleicos manipulado
un elemento de unión a la proteína de unión a ácidos nucleicos 2
un elemento 2 de unión a la proteína de unión a ácidos nucleicos bicatenario
514-513 corresponde a una primera secuencia 2-1 de ácidos nucleicos del elemento de tallo
521-513 corresponde a una secuencia de la mitad del tallo del sitio de unión a la proteína CRISPR de Tipo 2 Clase V 2-1c
514-521 corresponde a una secuencia de la mitad del mensaje del sitio de unión a la proteína CRISPR de Tipo 2 Clase V 2-1b

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
513-512	corresponde a una secuencia 2 de ácidos nucleicos de elemento de bucle
512-511	corresponde a una secuencia 2-2 de ácidos nucleicos del primer elemento de tallo
un elemento de repetición 2	
511-508	corresponde a una secuencia 1C de ácidos nucleicos repetida ²
510-511	corresponde a una secuencia 1 de ácidos nucleicos del elemento enlazador C-1
519-510	corresponde a una secuencia 1bC de ácidos nucleicos repetida
518-519	corresponde a una secuencia 1 de ácidos nucleicos del elemento enlazador C-2
509-518	corresponde a una secuencia 2 de ácidos nucleicos repetida
508-509	corresponde a una secuencia 1 de ácidos nucleicos del elemento enlazador C-3
un elemento espaciador 2	
514-515	corresponde a una secuencia 2 de unión a diana de ácidos nucleicos
¹ = elemento de repetición puede incluir un sitio de unión a la proteína efectora ² = "C" indica una secuencia complementaria	

La Figura 5A presenta un ejemplo de dos ácidos nucleicos diseñados que forman un andamio (NASC-PC1 y NASC-PC2). En algunos aspectos, los ácidos nucleicos manipulados son ácidos nucleicos que se dirigen a ácidos nucleicos de clase 2 tipo V CRISPR, por ejemplo, un ácido nucleico que se dirige a ácido nucleico CpfI que comprende una secuencia de ácidos nucleicos repetida unida covalentemente al extremo 5' del ácido nucleico que se dirige a CpfI ácido. La Figura 5A, 500 a 507, ilustra un primer ácido nucleico manipulado que comprende una primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos (Figura 5A, 504 a 501), una secuencia 1 de unión a diana de ácidos nucleicos (Figura 5A, 500 a 501) que se encuentra en 3' de la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos, y una primera secuencia repetida 1 (Figura 5A, 504-507) que está ubicada en 5' de la primera Clase 2 de unión a ácidos nucleicos Tipo V secuencia de unión a la proteína CRISPR. La Figura 5A, 508 a 515, ilustra un segundo ácido nucleico manipulado que comprende una segunda secuencia de unión a la proteína CRISPR de Tipo V de unión a ácidos nucleicos (Figura 5A, 514 a 511), una secuencia 2 de unión a diana de ácidos nucleicos (Figura 5A, 515 a 514) que se localiza en 3' de la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos, y una primera secuencia repetida 2 (Figura 5A, 511-508) que está ubicada en 5' de la segunda Clase de unión a ácidos nucleicos 2 Tipo V secuencia de unión a la proteína CRISPR.

El primer ácido nucleico manipulado y el segundo ácido nucleico manipulado pueden comprender elementos adicionales tales como secuencias de unión a la proteína efectora, por ejemplo, un sitio de unión a la proteína de unión al ácido nucleico bicatenario (por ejemplo, un sitio de unión a la proteína Csy4) creado por asociación de la secuencia 1 de ácidos nucleicos repetida (Figura 5A, 505-506) y la secuencia 1C de ácidos nucleicos repetida (Figura 5A, 509-510) a través de las interacciones de enlace de hidrógeno.

La Figura 5B ilustra una modificación del ejemplo mostrado en la Figura 5A, en el que la secuencia 1 de ácidos nucleicos de elemento de bucle (Figura 5A, 502 a 503) del primer ácido nucleico manipulado y la secuencia 2 de

ácidos nucleicos de elemento de bucle (Figura 5A, 513 a 512) del segundo ácido nucleico manipulado están ausentes.

La Figura 5C presenta un ejemplo de dos ácidos nucleicos diseñados que forman un andamio (NASC-PC1 y NASC-PC2). En algunos aspectos, los ácidos nucleicos manipulados son ácidos nucleicos que se dirigen a ácidos nucleicos tipo V CRISPR de Clase 2, por ejemplo, un ácido nucleico que se dirige a un ácido nucleico Cpfl que comprende una secuencia de ácidos nucleicos repetida unida covalentemente al extremo 3' del ácido nucleico Cpfl que se dirige ácido. La Figura 5C ilustra una modificación de los ácidos nucleicos manipulados, en el que se agrega una secuencia repetida al extremo 3' (Figura 5C, 500) de un primer ácido nucleico manipulado (Figura 5C, 500 a 507) y se agrega una secuencia repetida complementaria al extremo 3' (Figura 5C, 515) del segundo ácido nucleico manipulado (Figura 5C, 508 a 515). La secuencia repetida del primer ácido nucleico manipulado y la secuencia repetida complementaria del segundo ácido nucleico interactúan a través del enlace del par de bases a través de hidrógeno.

La Figura 5D presenta una modificación de NASC-PC1 y NASC-PC2 representada en la Figura 5A. En la Figura 5D, la secuencia de ácidos nucleicos repetida unida covalentemente al extremo 5' de cada ácido nucleico dirigido a ácido nucleico Cpfl comprende dos elementos de repeticiones separados por una secuencia de ácidos nucleicos elemento enlazador, en la que solo uno de los dos elementos de repeticiones de la Figura 5D, I, es complementario y capaz de formar enlaces de hidrógeno con uno de los elementos de repeticiones de la Figura 5D, II. La Figura 5D ilustra una versión de dos ácidos nucleicos manipulados genéticamente que forman un andamio, en el que la secuencia repetida 1b (Figura 5D, 506-517) del primer ácido nucleico manipulado (Figura 5D, 500 a 507) es capaz de formar enlaces de par de bases a través de hidrógeno con la secuencia 1b repetida complementaria C (Figura 5D, 519 a 510) del segundo ácido nucleico manipulado (Figura 5D, 515 a 508), en el que la secuencia repetida 1b del primer ácido nucleico manipulado y la secuencia 1b repetida complementaria C del segundo ácido nucleico manipulado interactúa a través de enlaces de par de bases a través de hidrógeno.

La Figura 5E presenta un ejemplo de cuatro ácidos nucleicos manipulados genéticamente que forman un andamio basado en dos conjuntos de los dos ácidos nucleicos manipulados genéticamente de la Figura 5D. En esta Figura, 5E, I, y Figura 5E, II, proporcionan puntos de referencia para facilitar la comparación con los dos ácidos nucleicos diseñados que se muestran en la Figura 5D (es decir, las Figuras 5D, I y II). La Figura 5E ilustra una versión manipulada del ejemplo mostrado en la Figura 5D en el que un elemento de repetición del primer ácido nucleico manipulado (Figura 5E, I; NASC-PC-1) interactúa con un elemento de repetición del segundo ácido nucleico manipulado (Figura 5E, II; NASC-PC-2) a través de hidrógeno enlace de par de bases, y un elemento de repetición del segundo ácido nucleico manipulado (Figura 5E, II) interactúa con un elemento de repetición del tercer ácido nucleico manipulado (Figura 5E, III; NASC-PC-3) a través de apareamiento por enlace de hidrógeno y un elemento de repetición del tercer ácido nucleico manipulado (Figura 5E, III) interactúa con un elemento de repetición de un cuarto ácido nucleico manipulado (Figura 5E, IV; NASC-PC-4) a través del enlace de par de bases a través de hidrógeno y un elemento de repetición del cuarto ácido nucleico manipulado (Figura 5E, IV) interactúa con un elemento de repetición del primer ácido nucleico manipulado (Figura 5E, I) a través del enlace de par de bases a través de hidrógeno.

La Figura 5F ilustra una versión manipulada del ejemplo mostrado en la Figura 5D, en el que las secuencias de ácidos nucleicos del elemento de bucle (Figura 5D, 502 a 503 y Figura 5D, 513 a 512) no están presentes en la primera secuencia de ácidos nucleicos manipulada (Figura 5F, VIII y Figura 5F, V), la segunda secuencia de ácidos nucleicos manipulada (Figura 5G, VI), la tercera secuencia de ácidos nucleicos manipulada (Figura 5G, VI) y la cuarta secuencia de ácidos nucleicos manipulada (Figura 5G, VII). La Figura 5G presenta un ejemplo de cuatro ácidos nucleicos manipulados genéticamente que forman un andamio con base en dos conjuntos de los dos ácidos nucleicos manipulados genéticamente de la Figura 5F. En esta Figura, 5G, V, y Figura 5G, VIII, proporcionan puntos de referencia para facilitar la comparación con los dos ácidos nucleicos diseñados que se muestran en la Figura 5F (es decir, Figura 5F, V y VIII). La Figura 5G ilustra una versión manipulada del ejemplo mostrado en la Figura 5F en el que un elemento de repetición del primer ácido nucleico manipulado (Figura 5G, V; NASC-PC-1) interactúa con un elemento de repetición del segundo ácido nucleico manipulado (Figura 5G, VI; NASC-PC-2) a través del hidrógeno enlace de par de bases, y un elemento de repetición del segundo ácido nucleico manipulado (Figura 5G, VI; NASC-PC-2) interactúa con un elemento de repetición del tercer ácido nucleico manipulado (Figura 5G, VII; NASC-PC) 3) a través de la unión de par de bases a través de hidrógeno, y un elemento de repetición del tercer ácido nucleico manipulado (Figura 5G, VII) interactúa con un elemento de repetición de un cuarto ácido nucleico manipulado (Figura 5G, VIII; NASC-PC-4) a través de enlaces de par de bases a través de hidrógeno, y una repetición del cuarto ácido nucleico manipulado (Figura 5G, VIII) interactúa con un elemento de repetición del primer ácido nucleico manipulado (Figuras 5G, V) a través de enlaces de par de bases a través de hidrógeno.

En otros aspectos, un complejo de nucleoproteínas puede formarse por una proteína de unión a ácidos nucleicos a una macromolécula que comprende una secuencia 1 de unión a diana de ácidos nucleicos, una secuencia 1 de ácidos nucleicos repetida, una secuencia 2 de ácidos nucleicos repetida y una secuencia 1 de ácidos nucleicos de unión a diana (por ejemplo, Figura 5H, Figura 5I). La Figura 5H ilustra una versión de dos ácidos nucleicos manipulados genéticamente que forman un andamio, en el que una secuencia 1-1b de medio tallo del sitio de unión a la proteína del Tipo V CRISPR de Clase 2 (Figura 5H, 520-502) de un primer ácido nucleico IX (Figura 5H, 500 a 502; NASC-PC1) es capaz de unir un par de bases a través de hidrógeno con la secuencia complementaria de la

mitad del tallo del sitio de unión a la proteína CRISPR de Tipo 2 complementaria 2-1b (Figura 5H, 514 a 521) de un segundo ácido nucleico X (Figura 5H, 515 a 513; NASC-PC2), y en la que la secuencia del tallo medio del sitio de unión a la proteína CRISPR de Clase 2 Tipo V 1-1b del primer ácido nucleico y la secuencia 2-1b del medio tallo del sitio de unión a la proteína CRISPR de Clase 2 complementaria del segundo ácido nucleico manipulado a través de la unión de par de bases a través de hidrógeno. La variación de secuencia entre las secuencias de la mitad del tallo que es suficiente para proporcionar una hibridación específica de la secuencia entre pares específicos de secuencias de la mitad del tallo es posible porque el reconocimiento del sitio de unión a la proteína CRISPR de Clase 2 Tipo V es tolerante de dicha variación de secuencia, siempre que se mantenga la estructura secundaria.

La Figura 5I ilustra una versión manipulada del ejemplo mostrado en la Figura 5H, en el que una secuencia de mitad de tallo del sitio de unión a la proteína CRISPR de Tipo 2 del primer ácido nucleico manipulado (Figura 5I, IX) interactúa con una secuencia de mitad de tallo del sitio de unión a la proteína CRISPR Tipo V del Tipo 2 del segundo ácido nucleico manipulado (Figura 5I, X); una secuencia de medio tallo del sitio de unión a la proteína CRISPR de Tipo 2 clase V del segundo ácido nucleico manipulado (Figura 5I, X) interactúa con una secuencia de medio tallo del sitio de unión a la proteína CRISPR de Tipo 2 del tercer ácido nucleico manipulado (Figura 5I, XI); una secuencia de mitad tallo del sitio de unión a la proteína CRISPR de Tipo 2 del tercer ácido nucleico manipulado (Figura 5I, XI) interactúa con una secuencia de mitad de tallo del sitio de unión a la proteína CRISPR Tipo 2 del cuarto ácido nucleico manipulado (Figura 5I, XII); y una secuencia de la mitad del tallo del sitio de unión a la proteína CRISPR del Tipo 2 del cuarto ácido nucleico manipulado (Figura 5I, XII) interactúa con la secuencia de la mitad del tallo del sitio de unión a la proteína CRISPR del Tipo V del primer ácido nucleico manipulado (Figura 5I, IX).

En relación con la presente invención, una composición de polinucleótido NASC comprende al menos NASC-PC1 y NASC-PC2. Un complejo NASC-PC1/NASC-PC2 comprende un elemento de repetición 1 conectado a un elemento de repetición 2, un elemento de unión a la proteína de unión a ácidos nucleicos bicatenarios 1 y un elemento de unión a la proteína de unión a ácidos nucleicos bicatenarios 2, y un elemento espaciador 1 y un elemento espaciador 2. Las realizaciones de la presente invención incluyen una composición de polinucleótido NASC que comprende elementos de unión a proteínas de unión a ácidos nucleicos bicatenarios correspondientes a una o más proteínas CRISPR-Cas9 de Clase 2 Tipo II.

En algunas realizaciones, la composición de polinucleótido de NASC es capaz de asociarse con dos proteínas de clase 2 CRISPR-Cas9 de tipo II para formar un complejo de nucleoproteínas. La Figura 6A, Figura 6B, Figura 6C, Figura 6D, Figura 6E, Figura 6F, Figura 6G, Figura 6H, Figura 6I, Figura 6K, Figura 6L, y Figura 6M ilustran elementos y ejemplos de estructuras de ácidos nucleicos manipuladas de la presente invención que comprenden típicamente una secuencia de unión a proteína CRISPR de clase 2 de unión a ácido nucleico.

La Tabla 4 presenta una serie de indicadores utilizados consistentemente en la Figura 6A, Figura 6B, Figura 6C, Figura 6D, Figura 6E, Figura 6F, y Figura 6G.

Tabla 4

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
un primer componente de ácidos nucleicos manipulado	
una primera secuencia de ácidos nucleicos manipulada (corresponde a 601-611)	
un elemento de unión a la proteína de unión a ácidos nucleicos 1	
un elemento de unión a la proteína de unión a ácidos nucleicos de doble cadena 1	
601-602 corresponde a una secuencia de ácidos nucleicos elemento enlazador	
602-603 corresponde a una secuencia 1-2 de ácidos nucleicos de horquilla	
603-604 corresponde a una secuencia de ácidos nucleicos del elemento enlazador	

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
604-605 corresponde a una secuencia 1-1 de ácidos nucleicos de horquilla
605-606 corresponde a una secuencia de ácidos nucleicos del elemento enlazador
606-607 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento nexo
607-608 corresponde a una secuencia de ácidos nucleicos del elemento enlazador
un elemento de repetición 1 ¹
608-609 corresponde a una secuencia 1 de ácidos nucleicos repetida
608-623 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento enlazador
623-624 corresponde a una secuencia 1a de ácidos nucleicos repetida
623-631 corresponde a una secuencia 1a1 de ácidos nucleicos repetida
631-632 corresponde a una secuencia 1a1 de ácidos nucleicos de protuberancia
632-624 corresponde a una secuencia 1a2 de ácidos nucleicos repetida
624-625 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento enlazador
624-647 corresponde a una secuencia 1-2-1 de ácidos nucleicos del elemento enlazador
647-648 corresponde a una secuencia 1-2a de ácidos nucleicos repetida
648-625 corresponde a una secuencia 1-2-2 de ácidos nucleicos del elemento enlazador
625-626 corresponde a una secuencia 1b de ácidos nucleicos repetida
625-633 corresponde a una secuencia 1b1 de ácidos nucleicos repetida
633-634 corresponde a una secuencia 1b1 de ácidos nucleicos de protuberancia
634-626 corresponde a una secuencia 1b2 de ácidos nucleicos repetida
626-609 corresponde a una secuencia 1-3 de ácidos nucleicos del elemento enlazador
609-610 corresponde a una secuencia de ácidos nucleicos del elemento enlazador

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
un elemento espaciador 1	
610-611 corresponde a una secuencia 1 de unión a diana de ácidos nucleicos	
un segundo componente de ácidos nucleicos manipulado	
una segunda secuencia de ácidos nucleicos manipulada (corresponde a 612-622)	
un elemento de unión a la proteína de unión a ácidos nucleicos 2	
un elemento 2 de unión a la proteína de unión a ácidos nucleicos bicatenario	
612-613 corresponde a una secuencia de ácidos nucleicos del elemento enlazador	
613-614 corresponde a una secuencia 2-2 de ácidos nucleicos en horquilla	
614-615 corresponde a una secuencia de ácidos nucleicos del elemento enlazador	
615-616 corresponde a una secuencia 2-1 de ácidos nucleicos en horquilla	
616-617 corresponde a una secuencia de ácidos nucleicos del elemento enlazador	
617-618 corresponde a una secuencia 2-1 de ácidos nucleicos del elemento nexo	
618-619 corresponde a una secuencia de ácidos nucleicos del elemento enlazador	
un elemento de repetición 2	
619-620 corresponde a una secuencia 1C ² de ácidos nucleicos repetida	
619-627 corresponde a una secuencia 2-3 de ácidos nucleicos del elemento enlazador	
627-628 corresponde a una secuencia 1bC de ácidos nucleicos repetida	
627-635 corresponde a una secuencia 1b2C de ácidos nucleicos repetida	
635-636 corresponde a una secuencia 2b2 de ácidos nucleicos de protuberancia	
636-628 corresponde a una secuencia 1C 1 de ácidos nucleicos repetida	

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
628-629	corresponde a una secuencia 2-2 de ácidos nucleicos del elemento enlazador
628-649	corresponde a una secuencia 2-2-2 de ácidos nucleicos del elemento enlazador
649-650	corresponde a una secuencia 12aC de ácidos nucleicos repetida
650-629	corresponde a una secuencia 2-2-1 de ácidos nucleicos del elemento enlazador
629-630	corresponde a una secuencia 1aC de ácidos nucleicos repetida
629-637	corresponde a una secuencia 1a2C de ácidos nucleicos repetida
637-638	corresponde a una secuencia 2a2 de ácidos nucleicos de protuberancia
638-630	corresponde a una secuencia 1a1C de ácidos nucleicos repetida
630-620	corresponde a una secuencia 2-1 de ácidos nucleicos del elemento enlazador
620-621	corresponde a una secuencia de ácidos nucleicos elemento enlazador
un elemento espaciador 2	
621-622	corresponde a una secuencia 2 de unión a diana de ácidos nucleicos
¹ = elemento de repetición puede incluir un sitio de unión a la proteína efectora	
² = "C" indica una secuencia complementaria	

Cada uno de un primer, un segundo y un tercer elemento pueden comprender secuencias de ácidos nucleicos adicionales, por ejemplo, 5' del elemento, 3' del elemento, o ambos 5' del elemento y 3' del elemento.

- 5 Cada uno de un primer, un segundo y un tercer elemento pueden comprender secuencias de ácidos nucleicos adicionales, por ejemplo, 5' del elemento, 3' del elemento, o ambos 5' del elemento y 3' del elemento.

La Figura 6A, 601-611, ilustra un ejemplo de primer ácido nucleico manipulado que comprende un primer elemento que comprende una secuencia de proteína de unión CRISPR Clase 2 Tipo V (Figura 6A, 601-607), un segundo elemento que comprende una secuencia 1 de ácidos nucleicos repetida (Figura 6A, 608-609), y un tercer elemento que comprende una secuencia 1 de ácidos nucleicos (Figura 6A, 610-611). Ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida se asocia con ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida para formar un elemento de tallo a través de enlaces de hidrógeno capaces de unirse a una proteína Clase 2 Tipo II CRISPR-Cas.

La Figura 6B ilustra una modificación de la Figura 6A, en el que el primer ácido nucleico manipulado (Figura 6B, 601-611) está asociado con el segundo ácido nucleico manipulado (Figura 6B, 612-622) a través de la unión de par de bases a través de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida (Figura 6A, 608-609) y la secuencia 1C de ácidos nucleicos repetida (Figura 6A, 619-620).

Una composición de polinucleótido NASC similar a la composición ilustrada en la Figura 6B puede construirse para su uso con proteínas de Clase 2 CRISPR-Cas de Tipo V para formar un complejo de nucleoproteínas. Un ejemplo de este tipo de composición de polinucleótido NASC se ilustra en la Figura 10, V.

La Figura 6C ilustra un ejemplo de un primer ácido nucleico manipulado, en el que el segundo elemento comprende además, en una orientación de 3' a 5', una secuencia 1-1 de ácidos nucleicos del elemento enlazador (Figura 6C, 608-623), una secuencia 1a de ácidos nucleicos repetida (Figura 6C, 623-624), una secuencia 1-2 de ácidos nucleicos del elemento enlazador (Figura 6C, 624-625), una secuencia 1b de ácidos nucleicos repetida (Figura 6D, 625-626) y un enlazador elemento de la secuencia 1-3 de ácidos nucleicos (Figura 6C, 626-609). Ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida se asocia con ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida para formar un elemento de tallo a través de enlaces de hidrógeno capaces de unirse a una proteína Clase 2 Tipo II CRISPR-Cas.

La Figura 6D ilustra una modificación de la Figura 6C en la que dos ácidos nucleicos diseñados forman un andamio, en el que el primer ácido nucleico manipulado está asociado con el segundo ácido nucleico creado mediante enlace de par de bases a través de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida (Figura 6D, 623-624) y la repetición secuencia 1aC de ácidos nucleicos (Figura 6D, 629-630) y mediante el enlace de par de bases a través de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida (Figura 6D, 625-626) y la secuencia 1bC de ácidos nucleicos repetida (Figura 6D, 627- 628).

La Figura 6E, ilustra un ejemplo de un primer ácido nucleico manipulado, en el que el segundo elemento comprende además, en una orientación de 3' a 5', una secuencia 1-1 de ácidos nucleicos de elemento enlazador (Figura 6C, 608-623), una repetición de una secuencia 1a1 de ácidos nucleicos (Figura 6C, 623-631), una secuencia de ácidos nucleicos de protuberancia (Figura 6E, 631-632), una secuencia 1a2 de ácidos nucleicos repetida (Figura 6E, 632-624), una secuencia 1-2 de ácidos nucleicos de elemento enlazador (Figura 6C, 624-625), una secuencia 1b1 de ácidos nucleicos repetida (Figura 6D, 625-633), una secuencia 1b1 de ácidos nucleicos de protuberancia (Figura 6E, 633-634) y una secuencia 1b2 de ácidos nucleicos repetida (Figura 6E, 634-626). Ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida se asocia con ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida para formar un elemento de tallo a través de enlaces de hidrógeno capaces de unirse a una proteína Clase 2 Tipo II CRISPR-Cas.

La Figura 6F ilustra una modificación de la Figura 6E en la que dos ácidos nucleicos manipulados genéticamente forman un andamio, en el que el primer ácido nucleico manipulado se asocia con el segundo ácido nucleico manipulado mediante la unión de par de bases a través de hidrógeno entre la secuencia 1a1 de ácidos nucleicos repetida (Figura 6F, 623-631) y la secuencia 1a1C de ácidos nucleicos repetida (Figura 6F, 638-630), a través de enlaces de par de bases a través de hidrógeno entre la secuencia 1a2 de ácidos nucleicos repetida (Figura 6F, 632-624) y la secuencia 1a2C de ácidos nucleicos repetida (Figura 6F, 629-637), a través de enlaces de par de bases a través de hidrógeno entre la secuencia 1b1 de ácidos nucleicos repetida (Figura 6F, 625-633) y la secuencia 1b1C de ácidos nucleicos repetida (Figura 6F, 636-628), y a través de enlaces de par de bases a través de hidrógeno entre la secuencia 1b2 de ácidos nucleicos repetida (Figura 6E, 634-626) y la secuencia 1b2C de ácidos nucleicos repetida (Figura 6E, 627-635).

La Figura 6G es una variación de la composición de polinucleótido NASC ilustrada en la Figura 6F. La secuencia 1-2 de ácidos nucleicos del elemento enlazador se modifica mediante la inserción de una secuencia de ácidos nucleicos del sitio de unión a la proteína efectora (Figura 6G, 647-648) y la secuencia 2-2 de ácidos nucleicos del elemento enlazador se modifica mediante la inserción de una proteína efectora sitio de unión secuencia 2 de ácidos nucleicos (Figura 6G, 649-650). La secuencia de ácidos nucleicos del sitio de unión a la proteína efectora y la secuencia de ácidos nucleicos del sitio de unión de la proteína efectora se conectan para formar un sitio de unión de la proteína efectora a través de pares de bases con enlaces de hidrógeno. En una realización, el sitio de unión a la proteína efectora es un sitio de unión a Csy4. Una forma enzimáticamente inactiva de la proteína Csy4 puede unirse al sitio para estabilizar aún más la estructura de la composición del polinucleótido NASC. Una forma enzimáticamente activa de la proteína Csy4 puede unirse al sitio para desestabilizar (por ejemplo, a través de la actividad endorribonucleasa) la estructura de la composición del polinucleótido NASC (por ejemplo, para inducir la ruptura de la composición de polinucleótido NASC/estructuras de jaula cerrada basadas en proteínas de unión a ácidos nucleicos). La Figura 6K ilustra una composición de polinucleótido NASC capaz de asociarse con una primera proteína ortóloga CRISPR-Cas9 de Clase 2 Tipo II y una segunda proteína ortóloga CRISPR-Cas9 de Clase 2 Tipo II para formar un complejo de nucleoproteínas. La Figura 6K ilustra una composición de polinucleótido NASC que comprende una NASC-PC1 y una NASC-PC2. El NASC-PC1 (Figura 6K, IX) comprende, en una dirección 5' a 3': una secuencia 1 de unión a diana de ácidos nucleicos; una secuencia 1 de ácidos nucleicos de enlace que comprende, una secuencia 1a de ácidos nucleicos repetida, una secuencia 1b de ácidos nucleicos repetida, una repetición de ácidos nucleicos 1c y una repetición de ácidos nucleicos 1d; y una secuencia 1 de ácidos nucleicos de proteína de Clase 2 S. Tipo II CRISPR-Cas9 de *S. thermophilus*. La NASC-PC2 (Figura 6K, X) comprende, en una dirección de 5' a 3': una secuencia 2 de unión a ácidos nucleicos diana; una secuencia 1 de ácidos nucleicos de enlace que comprende, una secuencia 1Cd de ácidos nucleicos repetida, una secuencia 1CC de ácidos nucleicos repetida, un ácido nucleico repetida 1bC, y una secuencia 1aC de ácidos nucleicos repetida; y una secuencia 1 de unión a la proteína de unión al ácido nucleico CRISPR-Cas9 de Clase 2 Tipo II de *S. pyogenes*. La NASC-PC1 y la NASC-PC2 conectadas a través de pares de bases con enlaces de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida/la secuencia 1aC de ácidos nucleicos repetida, la secuencia 1b repetida de ácidos nucleicos/la secuencia 1bC de ácidos nucleicos repetida, la secuencia 1C de ácidos nucleicos repetida/la repetición del ácido nucleico 1cC, y la repetición de la secuencia 1d de ácidos nucleicos/la secuencia 1dC de ácidos nucleicos repetida para formar una macromolécula. La macromolécula es capaz de unirse a una proteína CRISPR-Cas9 Clase 2 Tipo II

de *S. thermophilus* (alrededor de la secuencia repetida de ácidos nucleicos 1d/la región de la secuencia 1dC de ácidos nucleicos repetida y la secuencia repetida de ácidos nucleicos 1c/la región secuencia 1CC de ácidos nucleicos repetida) y una proteína CRISPR-Cas9 de Clase 2 Tipo II de *S. pyogenes* (alrededor de la secuencia 1b de ácidos nucleicos repetida/región de la secuencia 1bC de ácidos nucleicos repetida, la secuencia 1a de ácidos nucleicos repetida/región de la secuencia 1aC de ácidos nucleicos repetida). El uso de dichos complejos de composición de polinucleótido NASC/proteína ortóloga Cas9 proporciona, por ejemplo, un mayor número de secuencias diana disponibles en vista de la variabilidad de PAM entre las proteínas ortólogas Cas9 (en comparación con el uso de un complejo de ácidos nucleicos guía/proteína Cas9 que comprende solo ortólogo Cas9 solo). Además, la composición de polinucleótidos NASC/complejos de proteínas ortólogas Cas9 pueden mejorar la especificidad de la orientación de una región polinucleotídica al proporcionar una mayor flexibilidad en la elección de secuencias diana cercanas en vista de la variabilidad de PAM entre las proteínas ortólogas Cas9 (en comparación con el uso de un ácido nucleico de guía/complejo de proteínas Cas9 que comprende cualquiera de los ortólogos Cas9 solo). En vista de las enseñanzas de la presente especificación, un experto en la técnica puede aplicar este uso de dos o más proteínas ortólogas Cas9 diferentes mediante la combinación de diferentes componentes de las composiciones de polinucleótidos NASC descritas en este documento.

En una realización adicional, el NASC-PC1 (Figura 6K, IX) y el NASC-PC2 (Figura 6K, X) son capaces de asociarse con la misma proteína ortóloga CRISPR-Cas9 de Clase 2 Tipo II (véase, por ejemplo, Fonfara, I., et al., Nucleic Acids Research 42(4): 2577-2590 (2014)). En esta realización, la secuencia 1d de ácidos nucleicos repetida/la secuencia 1dC de ácidos nucleicos repetida, y la secuencia 1C de ácidos nucleicos repetida /la secuencia 1 CC de ácidos nucleicos repetida son capaces de asociarse con proteínas de *S. mutans* CRISPR-Cas9 Clase 2 Tipo II y también son capaces de asociarse con una proteína CRISPR-Cas9 de Clase 2 tipo II de *S. pyogenes*. La secuencia 1b de ácidos nucleicos repetida/la secuencia 1bC de ácidos nucleicos repetida, y la secuencia 1a de ácidos nucleicos repetida/la secuencia 1aC de ácidos nucleicos repetida son capaces de asociarse con una proteína CRISPR-Cas9 de Clase 2 Tipo II de *S. pyogenes* y también son capaces de asociarse con una proteína CRISPR-Cas9 de Clase 2 Tipo II de *S. mutans*. Por ejemplo, aunque las regiones repetidas de NASC-PC1 (Figura 6K, IX) y NASC-PC2 (Figura 6K, X) se derivan de diferentes especies que contienen loci CRISPR de Tipo II de Clase 2 (por ejemplo, *S. pyogenes* o *S. mutans*), solo se utiliza una proteína CRISPR-Cas9 Clase 2 Tipo II (por ejemplo, una proteína CRISPR-Cas9 Clase 2 Tipo 2 de *S. pyogenes* o una proteína CRISPR-Cas9 Clase 2 Tipo II de *S. mutans*) para formar una composición de polinucleótido NASC/complejo Cas9. Una ventaja de este tipo de composición de polinucleótidos de NASC es la flexibilidad para usar cualquiera de las dos proteínas Cas9 con la misma composición de polinucleótidos de NASC, y cada una de las proteínas Cas9 reconoce diferentes secuencias de PAM. Por lo tanto, aumenta el número de posibles sitios de unión que pueden dirigirse por la composición del polinucleótido NASC.

En algunas realizaciones de la presente invención, una composición de polinucleótido NASC comprende al menos tres polinucleótidos, en el que el complejo comprende un elemento de repetición 1 conectado a un elemento de repetición 1C, el elemento de repetición 2 conectado a un elemento de repetición 2C, el elemento de repetición 3 conectado a un elemento de repetición 3C, un elemento de unión a la proteína de unión a ácidos nucleicos de doble cadena 1, un elemento espaciador 1, un elemento de separación 2 y un elemento de separación 3, en el que la composición de polinucleótido de NASC es capaz de unirse a tres proteínas de unión a ácidos nucleicos. En algunas realizaciones, las proteínas de unión a ácidos nucleicos son proteínas de unión a ácidos nucleicos bicatenarias. En realizaciones preferidas, las proteínas de unión a ácidos nucleicos de clase 2 de las proteínas CRISPR-Cas.

La Tabla 5 presenta una serie de indicadores adicionales utilizados en la Figura 6H y la Figura 6I.

Tabla 5

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
un primer componente de ácidos nucleicos manipulado (I)	
una primera secuencia de ácidos nucleicos manipulados	
un elemento de unión a la proteína de unión a ácidos nucleicos 1	
un elemento de repetición 1 ¹	

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
608-609 corresponde a una secuencia 1 de ácidos nucleicos repetida
608-623 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento enlazador
623-624 corresponde a una secuencia 1a de ácidos nucleicos repetida
624-625 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento enlazador
625-626 corresponde a una secuencia 1b de ácidos nucleicos repetida
626-609 corresponde a una secuencia 1-3 de ácidos nucleicos del elemento enlazador
un elemento espaciador 1
un segundo componente de ácidos nucleicos manipulados (II)
una segunda secuencia de ácidos nucleicos manipulados
un elemento de unión a la proteína de unión a ácidos nucleicos 2
un elemento de repetición 2
619-653 corresponde a una secuencia 2 de ácidos nucleicos repetida
619-627 corresponde a una secuencia 2-3 de ácidos nucleicos del elemento enlazador
627-628 corresponde a una secuencia 1bC2 de ácidos nucleicos repetida
628-651 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento enlazador
651-652 corresponde a una secuencia 2a de ácidos nucleicos repetida
652-653 corresponde a una secuencia 2-4 de ácidos nucleicos del elemento enlazador
un elemento espaciador 2
un tercer componente de ácidos nucleicos manipulados (III)
una tercera secuencia de ácidos nucleicos manipulados
un elemento de unión a la proteína de unión a ácidos nucleicos 3

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
un elemento de repetición 3	
654-659 corresponde a una secuencia 1 de ácidos nucleicos repetida	
654-655 corresponde a una secuencia 3-3 de ácidos nucleicos del elemento enlazador	
655-656 corresponde a una secuencia 2aC de ácidos nucleicos repetida	
656-657 corresponde a una secuencia 3-2 de ácidos nucleicos del elemento enlazador	
657-658 corresponde a una secuencia 1aC de ácidos nucleicos repetida	
658-659 corresponde a una secuencia 3-1 de ácidos nucleicos del elemento enlazador	
un elemento espaciador 3	
1 = elemento de repetición puede incluir un sitio de unión a la proteína efectora	
2 = "C" indica una secuencia complementaria	

La Figura 6H ilustra una modificación de la Figura 6C donde tres ácidos nucleicos diseñados forman un andamio, en el que un primer ácido nucleico manipulado (Figura 6H, I) está asociado con un segundo ácido nucleico manipulado (Figura 6H, II) a través del enlace de par de bases a través de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida (Figura 6H, 625-626) y repetir la secuencia 1bC de ácidos nucleicos (Figura 6H, 627-628), y el segundo ácido nucleico manipulado genética (Figura 6H, II) está asociado con el tercer ácido nucleico manipulado genética (Figura 6H, III) a través de enlaces de par de bases a través de hidrógeno entre la repetición de la secuencia 2a de ácidos nucleicos (Figura 6H, 638-639) y la repetición de la secuencia 2aC de ácidos nucleicos (Figura 6H, 642-643), y el tercer ácido nucleico manipulado (Figura 6H, III) se asocia con el primer ácido nucleico manipulado (Figura 6H, I) a través de enlaces de par de bases a través de hidrógeno entre la secuencia 1aC de ácidos nucleicos repetida (Figura 6H, 644-645) y la secuencia 1a de ácidos nucleicos repetida (Figura 6H, 623-624).

La Figura 6I ilustra una modificación de la Figura 6E, donde tres ácidos nucleicos diseñados forman un andamio utilizando el ácido nucleico manipulado como se describe en la Figura 6E, en el que el primer ácido nucleico manipulado (Figura 6I, IV) está asociado con el segundo ácido nucleico manipulado (Figura 6I, V) a través de enlaces de par de bases a través de hidrógeno entre secuencias repetidas, y el segundo ácido nucleico manipulado (Figura 6I, V) se asocia con el tercer ácido nucleico manipulado (Figura 6I, VI) a través de enlaces de par de bases a través de hidrógeno entre secuencias repetidas, y el tercer ácido nucleico manipulado (Figura 6I, VI) se asocia con el primer ácido nucleico manipulado (Figura 6I, IV) a través de enlaces de par de bases a través de hidrógeno entre secuencias repetidas.

La Figura 6J ilustra una composición de polinucleótido NASC capaz de asociarse con una proteína Clase 2 Tipo II CRISPR-Cas y una proteína Clase 2 Tipo V CRISPR-Cpf1 para formar un complejo de nucleoproteínas.

La Figura 6J ilustra una composición de polinucleótido NASC capaz de asociarse con una proteína CRISPR-Cas9 Clase 2 Tipo II y una proteína Clase 2 Tipo V CRISPR-Cpf1 para formar un complejo de nucleoproteínas. La Figura 6J ilustra una composición de polinucleótido NASC que comprende una NASC-PC1 que comprende un elemento espaciador 1 y un elemento espaciador 2 (NASC-PC-2TS; Figura 6J, VII). El NASC-PC-2TS comprende, en una dirección de 5' a 3': una secuencia 1 de unión a diana del ácido nucleico de CRISPR-Cas9 Clase 2 Tipo II; una secuencia 1 de ácidos nucleicos de enlace que comprende una secuencia 1a de ácidos nucleicos repetida, una secuencia 1b de ácidos nucleicos repetida y una secuencia 1c de ácidos nucleicos repetida; y una secuencia de unión a la diana 1 del ácido nucleico CRISPR-Cpf1 de Clase 2 Tipo V. La composición de polinucleótidos NASC comprende además una NASC-PC2 que comprende un concatenado que comprende una secuencia de unión a la

proteína de unión al ácido nucleico CRISPR-Cas9 de Clase 2 Tipo II y un Tipo V de Clase 2 V CRISPR-Cpf1 secuencia de unión a la proteína de unión al ácido nucleico 2 (NASC-PC-CE; Figura 6J, VIII). El NASC-PC-CE comprende además una secuencia 1aC de ácidos nucleicos repetida, una secuencia 1bC de ácidos nucleicos repetida y una secuencia 1 CC de ácidos nucleicos repetida a través de la cual el NASC-PC-CE está conectado al NASC-PC-2TS a través de pares de bases con enlaces de hidrógeno para formar una macromolécula que es capaz de unirse a una proteína CRISPR-Cas9 Clase 2 Tipo II y una proteína Clase 2 Tipo V CRISPR-Cpf1. El uso de dichos complejos de composición de polinucleótidos NASC/proteína Cas9/proteína Cpf1 proporciona, por ejemplo, un mayor número de secuencias objetivo disponibles en vista de la variabilidad de PAM y las diferencias de longitud de secuencia diana entre la proteína Cas9 y la proteína Cpf1 (versus el uso de un complejo de ácidos nucleicos de guía/proteína Cas que comprende ya sea proteína Cas9 o proteína Cpf1 sola). Además, los complejos de composición de polinucleótidos NASC/proteína Cas9/proteína Cpf1 pueden mejorar la especificidad de la orientación de una región polinucleotídica al proporcionar una mayor flexibilidad en la elección de secuencias diana cercanas en vista de la variabilidad de PAM y las longitudes de secuencia diana entre la proteína Cas9 y la proteína Cpf1 (versus uso de un complejo de ácidos nucleicos de guía/proteína Cas que comprende proteína Cas9 o proteína Cpf1 sola). En vista de las enseñanzas de la presente especificación, un experto en la técnica puede aplicar este uso de dos o más proteínas Cas diferentes combinando diferentes componentes de las composiciones de polinucleótidos NASC descritas en el presente documento.

En otras realizaciones, una primera secuencia de ácidos nucleicos repetida de un par comprende además una primera etiqueta de afinidad y una segunda secuencia de ácidos nucleicos repetida de la pareja comprende además una segunda etiqueta de afinidad, y la primera etiqueta de afinidad está conectada con la segunda etiqueta de afinidad. Por ejemplo, la secuencia 1C de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de unión a la proteína efectora y la secuencia 2 de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora, y un sitio de unión 1 de efector está formado por una base de hidrógeno - enlace par entre la secuencia 1 del sitio de unión a la proteína efectora y la secuencia 2 de ácidos nucleicos del sitio de unión de la proteína efectora. Un ejemplo de un sitio de unión del efector es un sitio de unión a la proteína Csy4.

En otro aspecto de la presente divulgación, la composición de polinucleótido NASC comprende un componente de ácidos nucleicos concatenado diseñado ("NASC-PC-CT") y al menos una NASC-PC1 y una NASC-PC2.

En un aspecto, un elemento concatenado polinucleótido NASC diseñado (NASC-PC-CE) comprende, en una dirección 3' a 5': un primer elemento concatenado 1 que comprende un elemento 1 de unión a proteína de unión a ácidos nucleicos y un segundo concatenado el elemento 1 que comprende un elemento de repetición A1, en el que el elemento de repetición A1 comprende una secuencia A1 de ácidos nucleicos repetida; un primer elemento concatenado 2 que comprende un elemento 2 de unión a la proteína de unión a ácidos nucleicos; y un segundo elemento concatenado 2 que comprende una repetición del elemento 2, en el que la repetición del elemento 2 comprende una secuencia A2 de ácidos nucleicos repetida. El primer elemento concatenado 1 está conectado al segundo elemento concatenado 1, el segundo elemento concatenado 1 está conectado al primer elemento concatenado 2, y el primer elemento concatenado 2 está conectado al segundo elemento concatenado 2 para formar el NASC-PC-CE.

Un tercer elemento concatenado 1 (NASC-PC-CE3-1) comprende, en una dirección 3' a 5', un elemento de repetición A1C que comprende una secuencia A1C de ácidos nucleicos repetida, y un elemento espaciador 1 que comprende una secuencia 1 de unión a diana de ácidos nucleicos. Un tercer elemento concatenado 2 (NASC-PC-CE3-2) comprende, en una dirección 3' a 5', un elemento de repetición A2C que comprende una secuencia A2C de ácidos nucleicos repetida, y un elemento espaciador 2 que comprende una secuencia 2 de unión a diana de ácidos nucleicos. La secuencia A1 de ácidos nucleicos repetida está conectada con la secuencia A1C de ácidos nucleicos repetida, la secuencia A2 de ácidos nucleicos repetida está conectada con la secuencia A2-C de ácidos nucleicos repetida para formar el NASC-PC-CE.

Una primera proteína de unión a ácidos nucleicos es capaz de unirse al elemento de unión a proteína de unión a ácidos nucleicos 1 y una segunda proteína de unión a ácidos nucleicos es capaz de unirse al elemento de unión a proteína de unión a ácidos nucleicos 2. En algunos aspectos, la proteína de enlace de ácidos nucleicos el elemento de unión es un elemento de unión a la proteína de unión al ácido nucleico de doble cadena de unión a una proteína de unión al ácido nucleico de doble cadena.

En aspectos adicionales, una primera secuencia de ácidos nucleicos repetida de un par está conectada con la segunda secuencia de ácidos nucleicos repetida del par a través de pares de bases con enlaces de hidrógeno.

La Figura 7A, Figura 7B, Figura 7C, Figura 7D, Figura 7E, Figura 7F, Figura 7G, Figura 7H, y Figura 7I, ilustran elementos y ejemplos de andamios de ácidos nucleicos concatenados diseñados.

La Tabla 6 presenta una serie de indicadores utilizados consistentemente en la Figura 7A, Figura 7B, Figura 7C, Figura 7D, Figura 7E, Figura 7F, Figura 7G, Figura 7H, y Figura 7I.

Tabla 6

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
un elemento concatenado polinucleótido NASC diseñado (corresponde-700-717) (NASC-PC-CE)
un primer elemento concatenado 1 (NASC-PC-CE1-1)
un elemento de unión a la proteína de unión a ácidos nucleicos 1
700-701 corresponde a una secuencia 1-5 de ácidos nucleicos del elemento enlazador
701-702 corresponde a una secuencia 1-2 de ácidos nucleicos de horquilla
702-703 corresponde a una secuencia 1-4 de ácidos nucleicos del elemento enlazador
703-704 corresponde a una secuencia 1-1 de ácidos nucleicos de horquilla
704-705 corresponde a una secuencia 1-3 de ácidos nucleicos del elemento enlazador
705-706 corresponde a un elemento nexa 1
706-707 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento enlazador
un segundo elemento concatenado 1 (NASC-PC-CE1-2)
un elemento de repetición A1 ¹
707-708 corresponde a una secuencia A1 de ácidos nucleicos repetida
707-728 corresponde a una secuencia A1-1 de ácidos nucleicos del elemento enlazador
707-726 corresponde a una secuencia A1-4 de ácidos nucleicos del elemento enlazador
726-727 corresponde a una secuencia A1-1 de ácidos nucleicos repetida
727-728 corresponde a una secuencia A1-1 de ácidos nucleicos de protuberancia
728-708 corresponde a una secuencia A1-2 de ácidos nucleicos del elemento enlazador
728-729 corresponde a una secuencia A1-2 de ácidos nucleicos repetida
729-730 corresponde a una secuencia A1-3 de ácidos nucleicos del elemento enlazador
730-708 corresponde a una secuencia nucleica enlazadora A1-4 que puede comprender una secuencia A1 de ácidos nucleicos de sitio de unión a la proteína efectora
708-709 corresponde a una secuencia A1 de ácidos nucleicos del elemento enlazador

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
un primer elemento concatenado 2 (NASC-PC-CE-1-2)
un elemento de unión a la proteína de unión a ácidos nucleicos 2 (NASC-PC-CE2)
709-710 corresponde a una secuencia 2-5 de ácidos nucleicos del elemento enlazador
709-710 corresponde a una secuencia 2-5 de ácidos nucleicos del elemento enlazador
710-711 corresponde a una secuencia 2-1 de ácidos nucleicos de horquilla
711-712 corresponde a una secuencia 2-4 de ácidos nucleicos del elemento enlazador
712-713 corresponde a una secuencia 2-3 de ácidos nucleicos del elemento enlazador
713-714 corresponde a un elemento nexa 2
714-715 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento enlazador
un segundo elemento concatenado 2 (NASC-PC-CE2-2)
un elemento de repetición A2
715-716 corresponde a una secuencia A2 de ácidos nucleicos repetida
715-731 corresponde a una secuencia A2-4 de ácidos nucleicos del elemento enlazador
715-733 corresponde a una secuencia A2-1 de ácidos nucleicos del elemento enlazador
731-732 corresponde a una secuencia A2-1 de ácidos nucleicos repetida
732-733 corresponde a una secuencia A2-1 de ácidos nucleicos de protuberancia
733-734 corresponde a una secuencia A2-2 de ácidos nucleicos repetida
734-735 corresponde a un elemento enlazador ácido nucleico A2-3
716-735 corresponde a una secuencia A2-4 nucleica enlazadora que puede comprender una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora A2
716-717 corresponde a la secuencia A2 de ácidos nucleicos del elemento enlazador

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
un tercer elemento concatenado 1 (NASC-PC-CE3-1)
un elemento de repetición A1C2
718-719 corresponde a una secuencia A1C de ácidos nucleicos repetida
719-737 corresponde a una secuencia A1-1n de ácidos nucleicos del elemento enlazador
719-736 corresponde a una secuencia A1-1C ² de ácidos nucleicos repetida
736-737 corresponde a una secuencia A1-1n ³ de ácidos nucleicos de protuberancia
737-718 corresponde a una secuencia A1-2n de ácidos nucleicos del elemento enlazador
737-748 corresponde a una secuencia A1-2C de ácidos nucleicos repetida
738-748 corresponde a un ácido nucleico del elemento enlazador A1-3n
738-718 corresponde a una secuencia A1-4n nucleica enlazadora que puede comprender una secuencia A1C de ácidos nucleicos de sitio de unión a la proteína efectora
719-720 corresponde a una secuencia A1-4n de ácidos nucleicos del elemento enlazador
un elemento espaciador 1
720-721 corresponde a una secuencia 1 de unión a diana de ácidos nucleicos
un tercer elemento concatenado 2 (NASC-PC-CE3-2)
un elemento de repetición A2C
722-723 corresponde a una secuencia A2-1C2 de ácidos nucleicos repetida
723-740 corresponde a una secuencia A2-1n de ácidos nucleicos del elemento enlazador
723-739 corresponde a una secuencia A2-1C de ácidos nucleicos repetida
739-740 corresponde a una secuencia A2-1n de ácidos nucleicos de protuberancia
722-740 corresponde a una secuencia A2-2n de ácidos nucleicos del elemento enlazador
740-741 corresponde a una secuencia A2-2C de ácidos nucleicos repetida

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
741-749 corresponde a un ácido nucleico del elemento enlazador A2-3n
749-722 corresponde a una secuencia A2-4 nucleica enlazadora que puede comprender una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora A2C
723-724 corresponde a una secuencia A2-4n de ácidos nucleicos del elemento enlazador
un elemento espaciador 2
724-725 corresponde a una secuencia 2 de unión a diana de ácidos nucleicos
¹ = elemento de repetición puede incluir un sitio de unión a la proteína efectora ² = "C" indica una secuencia complementaria ³ = "n" indica una secuencia de cadena opuesta (por ejemplo, A2-1/A2-1n)

Cada uno de un primer, un segundo y un tercer elemento pueden comprender secuencias de ácidos nucleicos adicionales, por ejemplo, 5' del elemento, 3' del elemento, o ambos 5' del elemento y 3' del elemento.

- 5 La Figura 7A, 700-717, ilustra un ejemplo de una NASC-PC-CE que comprende un primer elemento concatenado 1 que comprende una secuencia de proteína de unión CRISPR de Tipo II de Clase 2 (Figura 7A, 700-706), un segundo elemento concatenado 1 (Figura 7A, 707-708) que comprende una secuencia A1 de ácidos nucleicos repetida, un primer elemento concatenado 2 (Figura 7A, 709-714) que comprende una secuencia de la proteína de unión CRISPR de Tipo II de Clase 2 (Figura 7A, 710-714), una segundo elemento concatenado 2 (Figura 7A, 715-717) que comprende una secuencia A2 de ácidos nucleicos repetida (Figura 7A, 715-716) y un tercer elemento concatenado 1 (NASC-PC-CE3-1; Figura 7A, 718-721) que comprende una secuencia A1C de ácidos nucleicos repetida (Figura 7A, 718-719) y una secuencia 1 de unión a diana de ácidos nucleicos (Figura 7A, 720-721), y un tercer elemento concatenado 2 (NASC-PC-CE3- 2; Figura 7A, 722-725) que comprende una secuencia A2C de ácidos nucleicos repetida (Figura 7A, 722-723) y una secuencia 2 de unión a diana de ácidos nucleicos (Figura 7A, 724-725). Una o más de las secuencias de ácidos nucleicos repetidas es una secuencia de unión a la proteína CRISPR de Tipo II de Clase II (por ejemplo, una secuencia de unión a la proteína Cas9). La secuencia A1 de ácidos nucleicos repetida está conectada para repetir la secuencia de ácidos nucleicos A1C. En un aspecto, la secuencia A1 de ácidos nucleicos repetida está conectada a través de pares de bases con enlaces de hidrógeno para repetir la secuencia de ácidos nucleicos.
- 20 La Figura 7B ilustra un ejemplo de la formación de un andamio a través de la asociación de NASC-PC-CE (Figura 7A, 700-717) con el tercer elemento concatenado 1 (Figura 7A, 718-721) a través de la unión de par de bases a través de hidrógeno entre la secuencia A1 de ácidos nucleicos repetida (Figura 7A, 707-708) y la secuencia A1C de ácidos nucleicos repetida (Figura 7A, 718-719), y la asociación de NASC-PC-CE (Figura 7A, 700-717) con el tercer elemento concatenado 2 (Figura 7A, 722-725) a través de enlaces de par de bases a través de hidrógeno entre la
- 25 la secuencia A2 de ácidos nucleicos repetida (Figura 7A, 715-716) y la secuencia A2C de ácidos nucleicos repetida (Figura 7A, 722-723).
- La Figura 7C ilustra una modificación de una NASC-PC-CE en el que el NASC-PC-CE (Figura 7A, 700-717) comprende además una secuencia A1-1 de ácidos nucleicos repetida (Figura 7C, 726-727), una secuencia A1-1 de ácidos nucleicos de protuberancia (Figura 7C, 727-728), una secuencia A1-2 de ácidos nucleicos repetida (Figura 7C, 728-729) y una secuencia A2-1 de ácidos nucleicos repetida (Figura 7C, 731-732) , una secuencia A2-1 de ácidos nucleicos de protuberancia (Figura 7C, 732-733), y una secuencia A2-2 de ácidos nucleicos repetida (Figura 7C, 733-734). El tercer elemento concatenado 1 (Figura 7C, 718-721) comprende además una secuencia A1-1C de ácidos nucleicos repetida (Figura 7C, 719-736), una secuencia A1-1 de ácidos nucleicos de protuberancia (Figura 7C, 736-737), y una secuencia A1-2C de ácidos nucleicos repetida (Figura 7C, 737-748). El tercer elemento concatenado 2 (Figura 7C, 722-725) comprende además una secuencia A2-1C de ácidos nucleicos repetida (Figura
- 30
- 35

7C, 723-739), una secuencia A2-1 de ácidos nucleicos de protuberancia (Figura 7C, 739-740), y una secuencia A2-2C de ácidos nucleicos repetida (Figura 7C, 740-741).

La Figura 7D ilustra un ejemplo de la formación de un andamio a través de la asociación de NASC-PC-CE (Figura 7D, 700-717) con el tercer elemento concatenado 1 (Figura 7D, 721-718) a través de la unión de par de bases a través de hidrógeno entre: la secuencia A1-1 de ácidos nucleicos repetida (Figura 7C, 726-727) y la secuencia A1-1 C de ácidos nucleicos repetida (Figura 7C, 719-736), y enlace de par de bases a través de hidrógeno entre la secuencia A1-2 de ácidos nucleicos repetida (Figura 7C, 728-729) y la secuencia A1-2C de ácidos nucleicos repetida (Figura 7C, 737-748); asociación del elemento concatenado manipulado 1 (Figura 7D, 700-717) con el tercer elemento de concatenado 2 (Figura 7C, 722-725) a través de la unión de par de bases a través de hidrógeno entre la secuencia A2-1 de ácidos nucleicos repetida (Figura 7C, 731-732) y repita la secuencia A2-1C de ácidos nucleicos (Figura 7C, 723-739), y a través de enlaces de par de bases a través de hidrógeno entre la secuencia A2-2 de ácidos nucleicos repetida (Figura 7C, 733-734) y la secuencia A2-2C de ácidos nucleicos repetida (Figura 7C, 740-741).

La Figura 7E presenta un ejemplo de un complejo formado a partir de cuatro secuencias de ácidos nucleicos manipuladas para hacer un andamio que comprende una NASC-PC-CE circular. En esta Figura, el NASC-PC-CE comprende dos copias del NASC-PC-CE que se muestra en la Figura 7D, 700-717 de unión desde el extremo 5' al extremo 3' para formar la circular NASC-PC-CE. En esta Figura, se muestran los números de referencia relativos a la Figura 7D para ayudar a ilustrar los componentes del elemento circular de ácidos nucleicos concatenado.

La Figura 7F es una ilustración de una modificación del ejemplo mostrado en la Figura 7D, en el que el NASC-PC-CE (Figura 7F, 700-717) comprende además un primer elemento concatenado 3 (Figura 7F, 717-744) unido covalentemente al extremo 5' (Figura 7F, 700-744). Un segundo elemento concatenado 3 está asociado con el primer elemento concatenado 3 (Figura 7F, 743-744). La Figura 7G ilustra un ejemplo de una modificación a la NASC-PC-CE (Figura 7F, 700-744) representada en la Figura 7F, en el que el NASC-PC-CE comprende un cuarto elemento concatenado (Figura 7G, 744-747) unido covalentemente al extremo 5' (Figura 7F, 700-747). En esta Figura, la región, Figura 7G, 744-745, se ilustra como un recuadro blanco para hacer las líneas cruzadas en la Figura 7H y la Figura 7I más aparente. Esta región también puede comprender una secuencia de ácidos nucleicos de elemento enlazador.

La Figura 7H, 700-747, ilustran un ejemplo de una NASC-PC-CE, en el que el segundo elemento concatenado 1 (Figura 7H, 707-708) está asociado con el tercer elemento concatenado 1 (Figura 7H, 744-747) a través de la unión de par de bases a través de hidrógeno.

La Figura 7I ilustra una modificación del ejemplo mostrado en la Figura 7H, en el que un tercer elemento concatenado I se asocia con el NASC-PC-CE a través del enlace del par de bases a través de hidrógeno para formar el elemento III (Figura 7I, III) y un cuarto elemento concatenado se asocia con el NASC-PC-CE a través del hidrógeno de base unión por pares para formar el elemento IV (Figura 7I, IV).

En otros aspectos de la presente divulgación, el NASC-PC-CE comprende polinucleótidos de nexo dividido.

Figura 8A, Figura 8B, Figura 8C, Figura 8D, Figura 8E, Figura 8F, Figura 8G, Figura 8H,

La Figura 8I, Figura 8J, Figura 8K, Figura 8L, Figura 8M, y Figura 8N ilustra elementos y ejemplos de andamios de ácidos nucleicos de nexo dividido concatenado diseñados.

La Tabla 7 presenta una serie de indicadores usados consistentemente en la Figura 8A a 8N.

Tabla 7

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio	
Indicador y región correspondiente	
un elemento concatenado polinucleótido de nexo dividido NASC diseñado (corresponde a 800-835) (NASC-PC-SCE)	
un primer elemento concatenado 1 (NASC-PC-SCE 1-1)	
un elemento de unión a la proteína de unión a ácidos nucleicos 1	

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
800-801 corresponde a una proteína de unión a ácidos nucleicos ácido nucleico de unión 1
801-802 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento de tallo nexo dividido
un elemento de repetición 1-1 ¹
802-816 corresponde un polinucleótido auxiliar 1-1
802-815 corresponde a una secuencia 1-1 de ácidos nucleicos repetida
814-815 corresponde a una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora de unión a ácidos nucleicos de doble cadena
un segundo elemento concatenado 1 (NASC-PC-SCE 1-2)
un elemento de unión a la proteína de unión a ácidos nucleicos 2
816-820 corresponde a un ácido nucleico de unión a la proteína de unión al ácido nucleico 2
820-821 corresponde a una secuencia 2-1 de ácidos nucleicos del elemento de tallo nexo dividido
un elemento de repetición 2-1
821-835 corresponde a un polinucleótido auxiliar 2-1
821-834 corresponde a una secuencia 2-1 de ácidos nucleicos repetida
833-834 corresponde a una secuencia 2 de ácidos nucleicos del sitio de unión a la proteína efectora de unión a ácidos nucleicos bicatenario
un segundo elemento concatenado 1 (NASC-PC-SCE2-1)
un elemento de repetición 1
808-807 corresponde a una primera secuencia 1-1 de ácidos nucleicos del elemento de tallo
808-810 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento de tallo inferior
810-811 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento de protuberancia
811-807 corresponde a una secuencia 1-1 de ácidos nucleicos del elemento de tallo superior
807-806 corresponde a una secuencia 1 de ácidos nucleicos de elemento de bucle

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
806-805 corresponde a una primera secuencia 1-2 de ácidos nucleicos del elemento de tallo
806-812 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento de tallo superior
812-813 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento de protuberancia
813-805 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento de tallo inferior
805-804 corresponde a una secuencia 1 de ácidos nucleicos conectiva
[805-806/808-807 corresponde a un primer elemento de tallo 1]
[811-807/806-812 corresponde a un elemento de tallo superior 1]
[810-811/812-813 corresponde a un elemento de protuberancia 1]
[808-810/813-805 corresponde a un elemento de tallo inferior 1]
804-803 corresponde a una secuencia 1-2 de ácidos nucleicos del elemento de tallo nexa dividido
[803-804/801-802 corresponde a un elemento nexa 1]
803-817 corresponde a un polinucleótido auxiliar 1-2
803-818 corresponde a una secuencia 1-2 de ácidos nucleicos repetida (en algunos aspectos complementarios a la secuencia 1 de ácidos nucleicos repetida-1)
818-819 corresponde a una secuencia 1C ² de ácidos nucleicos de sitio de unión a la proteína efectora de unión a ácidos nucleicos bicatenarios
[el sitio 814-815/818-819 corresponde a un sitio de unión a la proteína efectora de enlace de ácidos nucleicos bicatenarios 1]
un elemento espaciador 1
808-809 corresponde a una secuencia 1 de unión a diana de ácidos nucleicos
un segundo elemento concatenado 1 (NASC-PC-SCE2--2)
un elemento de repetición 2
827-826 corresponde a una primera secuencia 2-1 de ácidos nucleicos del elemento de tallo

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio
Indicador y región correspondiente
832-827 corresponde a una secuencia 2-1 de ácidos nucleicos del elemento de tallo inferior
831-832 corresponde a una secuencia 2-1 de ácidos nucleicos del elemento de protuberancia
831-826 corresponde a una secuencia 2-1 de ácidos nucleicos del elemento de tallo superior
825-826 corresponde a una secuencia 2 de ácidos nucleicos de elemento de bucle
824-825 corresponde a una secuencia 2-2 de ácidos nucleicos del primer elemento de tallo
825-830 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento de tallo superior
829-830 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento de protuberancia
829-824 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento de tallo inferior
823-824 corresponde a una secuencia 2 de ácidos nucleicos conectiva
[826-827/824-825 corresponde a un primer elemento de tallo 2]
[826-831/825-830 corresponde a un elemento de tallo superior 2]
[831-832/829-830 corresponde a un elemento de protuberancia 2]
[827-832/824-829 corresponde a un elemento de tallo inferior 2]
822-823 corresponde a una secuencia 2-2 de ácidos nucleicos del elemento de tallo nexa dividido
[822-823/820-821 corresponde a un elemento nexa 2]
822-836 corresponde a un polinucleótido auxiliar 2-2
822-837 corresponde a una secuencia 2-2 de ácidos nucleicos repetida (en algunos aspectos complementarios a la secuencia 2-1 de ácidos nucleicos repetida)
837-838 corresponde a una secuencia 2C de ácidos nucleicos del sitio de unión a la proteína efectora de unión a ácidos nucleicos bicatenario
[el sitio 837-838/833-834 corresponde a un sitio de unión a la proteína efectora de enlace de ácidos nucleicos bicatenarios 2]
un elemento espaciador 2
827-828 corresponde a una secuencia 2 de unión a diana de ácidos nucleicos

(continuación)

Indicadores numéricos utilizados para ilustrar regiones de complejos de dos o más secuencias de ácidos nucleicos manipuladas para formar un andamio

Indicador y región correspondiente

² = "C" indica una secuencia complementaria ¹ = un elemento de repetición puede incluir un sitio de unión a la proteína efectora

² = "C" indica una secuencia complementaria

La Figura 8A ilustra un ejemplo de polinucleótidos asociados con Cas9 de nexo dividido. La Figura 2B presenta un ejemplo de un polinucleótido de guía única asociado a Cas9. Los polinucleótidos asociados a Cas9 de nexo dividido de la Figura 8A se generan al dividir el esqueleto del polinucleótido dentro del elemento nexo (Figura 2B, 206) de un polinucleótido de guía única asociado a Cas9. La Figura 8A muestra los dos polinucleótidos de nexo dividido resultantes cuando no están asociados a través de interacciones de enlaces de hidrógeno. La Figura 8B presenta una vista de los polinucleótidos de nexo dividido cuando se asocian a través de interacciones de enlaces de hidrógeno. La región de las interacciones del enlace de hidrógeno se ilustra mediante un cuadro de guion roto en la Figura 8B.

La Figura 8C ilustra otro ejemplo de polinucleótidos asociados con Cas9 de nexo dividido. La Figura 2A presenta un ejemplo de un polinucleótido de guía única asociado a Cas9. Los polinucleótidos asociados a Cas9 de nexo dividido de la Figura 8C se generan al dividir el esqueleto del polinucleótido dentro del elemento nexo (Figura 2A, 206) de un polinucleótido de guía única asociado a Cas9. La Figura 8C muestra los dos polinucleótidos de nexo dividido resultantes cuando no están asociados a través de interacciones de enlace de hidrógeno. La Figura 8D presenta una vista de los polinucleótidos de nexo dividido cuando se asocian a través de interacciones de enlaces de hidrógeno. La región de las interacciones del enlace de hidrógeno se ilustra mediante un cuadro de guion roto en la Figura 8D.

La Figura 8E ilustra la adición de un polinucleótido auxiliar al polinucleótido asociado a Cas9 de división del nexo ilustrado en la Figura 8A. En esta Figura, el extremo 5' de un primer polinucleótido auxiliar (Figura 8E, 803-817) está unido covalentemente al extremo 3' de una mitad del elemento de nexo dividido (Figura 8E, 803) y el extremo 3' el extremo de un segundo polinucleótido auxiliar (Figura 8E, 802-816) se une covalentemente al extremo 5' de la otra mitad del elemento de nexo dividido (Figura 8E, 802). En algunos aspectos, solo se incluye un polinucleótido auxiliar. En otros aspectos, se incluyen dos polinucleótidos auxiliares de la misma o diferente longitud. La Figura 8E muestra los dos polinucleótidos de nexo dividido cuando no están asociados a través de interacciones de enlaces de hidrógeno.

La Figura 8F presenta una vista de los polinucleótidos de nexo dividido cuando se asocian a través de interacciones de enlaces de hidrógeno. La región de las interacciones del enlace de hidrógeno se ilustra mediante el cuadro de guion roto en la Figura 8D, 803-804/801-802. Un polinucleótido auxiliar puede comprender elementos adicionales tales como secuencias de unión a la proteína efectora, por ejemplo, un sitio de unión a la proteína de unión a ácidos nucleicos bicatenarios puede crearse por la asociación de dos polinucleótidos auxiliares a través de interacciones de enlaces de hidrógeno (por ejemplo, tal región del enlace de hidrógeno las interacciones se ilustran mediante el cuadro de guion roto en la Figura 8D, 818-819-804/814-815).

La Figura 8G ilustra otro ejemplo de la adición de un polinucleótido auxiliar al polinucleótido asociado a Cas9 de nexo dividido ilustrado en la Figura 8C. En esta Figura, el extremo 5' de un primer polinucleótido auxiliar (Figura 8G, 803-817) está unido covalentemente al extremo 3' de una mitad del elemento de nexo dividido (Figura 8E, 803) y el extremo 3' el extremo de un segundo polinucleótido auxiliar (Figura 8G, 802-816) está unido covalentemente al extremo 5' de la otra mitad del elemento de nexo dividido (Figura 8G, 802). En algunos aspectos, solo se incluye un polinucleótido auxiliar. En otros aspectos, se incluyen dos polinucleótidos auxiliares de la misma o diferente longitud. La Figura 8G muestra los dos polinucleótidos de nexo dividido cuando no están asociados a través de interacciones de enlaces de hidrógeno. La Figura 8H presenta una vista de los polinucleótidos de nexo dividido cuando se asocian a través de interacciones de enlace de hidrógeno. La región de las interacciones del enlace de hidrógeno se ilustra mediante el cuadro de guion roto en la Figura 8H, 803-804/801-802. Un polinucleótido auxiliar puede comprender elementos adicionales tales como secuencias de unión a la proteína efectora, por ejemplo, un sitio de unión a la proteína de unión a ácidos nucleicos bicatenarios puede crearse por la asociación de dos polinucleótidos auxiliares a través de interacciones de enlaces de hidrógeno (por ejemplo, tal región del enlace de hidrógeno las interacciones se ilustran mediante el cuadro de guion roto en la Figura 8H, 818-819-804/814-815).

En un aspecto, la divulgación se dirige a una composición de polinucleótido del elemento concatenado del polinucleótido dividido del nexo NAS (NAS-PC-SCE) que comprende, en una dirección de 3' a 5', un primer elemento del concatenado 1 (NAS-PC-SCE1- 1) que comprende un elemento de unión a la proteína de unión a ácidos nucleicos 1, una secuencia de ácidos nucleicos del elemento de tallo de nexo dividido 1-1, y un elemento de

repetición que comprende una secuencia de ácidos nucleicos repetida 1-1, y un segundo elemento concatenado 1 (NASC-PC- SCE1-2) que comprende un elemento de unión a la proteína de unión a ácidos nucleicos 2, una secuencia de ácidos nucleicos del elemento de tallo de nexo dividido 2-1 y un elemento de repetición que comprende una secuencia de ácidos nucleicos repetida 2-1. El NASC-PC-SCE1-1 y el NASC-PC-SCE1-2 están conectados para formar el NASC-PC-SCE. Un segundo elemento concatenado 1 (NASC-PC-SCE2-1) comprende un elemento de repetición 1 que comprende, en una dirección de 3' a 5', una secuencia 1-2 de ácidos nucleicos repetida, una secuencia de ácidos nucleicos de elemento de tallo de nexo dividido 1-2, y un primer elemento de tallo, y un elemento espaciador 1 que comprende una secuencia 1 de unión a diana de ácidos nucleicos. Un segundo elemento concatenado 2 (NASC-PC-SCE2-2) comprende un elemento de repetición 2 que comprende, en un 3' a 5' dirección, una secuencia 2-2 de ácidos nucleicos repetida, una secuencia 2-2 de ácidos nucleicos del elemento de tallo nexo dividido, y un primer elemento de tallo, y un elemento espaciador 2 que comprende una secuencia 2 de unión a ácidos nucleicos diana.

El elemento de repetición 1-1 está conectado al elemento de repetición 1-2, y el elemento de repetición 2-1 está conectado al elemento de repetición 2-2 para formar el NASC-PC-SCE, y el NASC-PC- SCE es capaz de unirse a dos proteínas de unión a ácidos nucleicos. En algunos aspectos, el elemento de unión a la proteína de unión a ácidos nucleicos es un elemento de unión a la proteína de unión a ácidos nucleicos bicatenarios de unión a una proteína de unión a ácidos nucleicos de doble cadena. En aspectos adicionales, una primera secuencia de ácidos nucleicos repetida de un par está conectada con la segunda secuencia de ácidos nucleicos repetida del par a través de pares de bases con enlaces de hidrógeno.

La Figura 8I presenta un ejemplo de una NASC-PC-SCE que comprende dos copias del polinucleótido de nexo dividido mostrado en la Figura 8B, 800-802 formando un andamio. En esta Figura, el NASC-PC-SCE es un primer polinucleótido nexo partido (Figura 8I, 800-802) unido covalentemente a un segundo polinucleótido nexo abierto (Figura 8I, 816-821) a través de un polinucleótido auxiliar (Figura 8I, 802-816). Cada primera mitad de un elemento nexo dividido (Figura 8I, 801-802 y 820-821) está conectada con la segunda mitad complementaria de su elemento nexo dividido (Figura 8I, 803-804 y 822-823, respectivamente), por ejemplo, a través de pares de bases con enlaces de hidrógeno.

La Figura 8J presenta un ejemplo de una NASC-PC-SCE que comprende dos copias del polinucleótido de nexo dividido mostrado en la Figura 8D, 800-802 formando un andamio. En esta Figura, el NASC-PC-SCE es un primer polinucleótido nexo partido (Figura 8J, 800-802) unido covalentemente a un segundo polinucleótido nexo abierto (Figura 8J, 816-821) a través de un polinucleótido auxiliar (Figura 8J, 802-816). Cada primera mitad de un elemento de nexo dividido (Figura 8J, 801-802 y 820-821) está conectada con la segunda mitad complementaria de su elemento de nexo dividido (Figura 8J, 803-804 y 822-823, respectivamente), por ejemplo, a través de pares de bases con enlaces de hidrógeno.

La Figura 8K presenta un ejemplo de una NASC-PC-SCE que comprende dos copias del polinucleótido de nexo dividido mostrado en la Figura 8F, 800-816, cada uno de los cuales comprende una secuencia auxiliar, que forma un andamio. En esta Figura, el NASC-PC-SCE es un primer polinucleótido de nexo dividido (Figura 8K, 800-816) unido covalentemente a un segundo polinucleótido de nexo dividido (Figura 8J, 816-835) a través de un polinucleótido auxiliar (Figura 8J, 802-816). Cada primera mitad de un elemento de nexo dividido (Figura 8K, 801-802 y 820-821) está conectada con la segunda mitad complementaria de su elemento de nexo dividido (Figura 8K, 803-804 y 822-823, respectivamente) y también está conectado por las secuencias auxiliares (Figura 8K, 802-815 y 803-818; Figura 8K, 821-834 y 822-836). Las conexiones se realizan, por ejemplo, a través de pares de bases con enlaces de hidrógeno.

La Figura 8L presenta un ejemplo de una NASC-PC-SCE que comprende dos copias del polinucleótido de nexo dividido mostrado en la Figura 8H, 800-816, cada uno de los cuales comprende una secuencia auxiliar, formando un andamio. En esta Figura, el NASC-PC-SCE es un primer polinucleótido nexo partido (Figura 8H, 800-816) unido covalentemente a un segundo polinucleótido nexo partido (Figura 8H, 816-835) a través de un polinucleótido auxiliar (Figura 8H, 802-816). Cada primera mitad de un elemento de nexo dividido (Figura 8H, 801-802 y 820-821) está conectada con la segunda mitad complementaria de su elemento de nexo dividido (Figura 8H, 803-804 y 822-823, respectivamente) y también está conectado por las secuencias auxiliares (Figura 8H, 802-815 y 803-818; Figura 8K, 821-834 y 822-836). Las conexiones se realizan, por ejemplo, a través de pares de bases con enlaces de hidrógeno. Los dos componentes del NASC-PC-SCE se indican en esta Figura como I y II.

La Figura 8M presenta un ejemplo de una NASC-PC-SCE que comprende los elementos I y II como se muestra en la Figura 8L NASC-PC-SCE comprende una NASC-PC-SCE circular. En esta Figura, dos conjuntos de elementos I y II, que se muestran en la 8L, 800-835, se unen desde el extremo 5' al extremo 3' para formar la NASC-PC-SCE circular. En esta Figura, los números de referencia relativos a la Figura 8L se muestran para ayudar a ilustrar los componentes de la NASC-PC-SCE circular.

La Figura 8N presenta un ejemplo de una NASC-PC-SCE que comprende los elementos I y II como se muestra en la Figura 8L, con la excepción de que las secuencias de ácidos nucleicos del primer elemento de tallo no están unidas por secuencias de ácidos nucleicos del elemento de bucle (véase, por ejemplo, Figura 8L, 806-807, 825-826). NASC-PC-SCE comprende una NASC-PC-SCE circular.

En otros aspectos, una primera secuencia de ácidos nucleicos repetida de un par comprende además la primera etiqueta de afinidad y la segunda secuencia de ácidos nucleicos repetida del par comprende además una segunda etiqueta de afinidad, y la primera etiqueta de afinidad está conectada con la segunda etiqueta de afinidad. Por ejemplo, la primera secuencia de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora y la segunda secuencia de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de sitio de unión a la proteína efectora. La secuencia de ácidos nucleicos de sitio de unión de proteína efectora está conectada por pares de bases con enlaces de hidrógeno al sitio de unión a la proteína efectora secuencia 2 de ácidos nucleicos para formar un sitio de unión a la proteína efectora 1. Un ejemplo de un sitio de unión al efector es un sitio de unión a la proteína Csy4.

En otro aspecto de la presente divulgación, una composición de polinucleótido NASC comprende una combinación de elementos de unión a proteínas de unión a ácidos nucleicos para dos o más proteínas de unión a ácidos nucleicos diferentes. En algunos aspectos, uno o más de los elementos de unión a la proteína de unión a ácidos nucleicos es un elemento de unión a la proteína de unión a ácidos nucleicos de doble cadena de unión a una proteína de unión a ácidos nucleicos de doble cadena (por ejemplo, una proteína de clase 2 CRISPR-Cas). Los aspectos de la presente divulgación incluyen una primera NATNA conectada covalentemente a una segunda NATNA para formar una composición de polinucleótido NASC. En algunos aspectos, un primer NATNA se conecta covalentemente a un segundo NATNA para formar un componente NASC-PC y dos o más componentes NASC-PC se conectan covalentemente o no covalentemente para formar una composición de polinucleótidos NASC. La composición de polinucleótido NASC es capaz de unirse a al menos dos proteínas de unión a ácidos nucleicos. Las conexiones no covalentes incluyen la conexión de los componentes NASC-PC a través de pares de bases con enlaces de hidrógeno.

La Figura 9A presenta un ejemplo de dos NATNA unidas para formar una composición de polinucleótido NASC que comprende (i) una copia de la secuencia de ácidos nucleicos manipulada por ingeniería mostrada en la Figura 5D, I, 500-507 (Figura 9A, I), y (ii) una copia de una secuencia de ácidos nucleicos manipulada por ingeniería genética correspondiente al polinucleótido de guía única mostrado en la Figura 2A que comprende además una secuencia de ácidos nucleicos del elemento enlazador unida covalentemente al extremo 3' del polinucleótido de guía única (Figura 9A, II), en la que la secuencia de ácidos nucleicos del elemento enlazador está unida covalentemente al extremo 5' del ácido nucleico manipulado secuencia mostrada en la Figura 5D, I, 500-507, formando un andamio.

La Figura 9B presenta un ejemplo de un complejo de dos conjuntos de los componentes mostrados en la Figura 9A. En esta Figura, los números de referencia relativos a la Figura 9A, I y II, se muestran para ayudar a ilustrar los componentes. Además, la Figura 9A, III, se proporciona para facilitar la comparación de la estructura del núcleo del complejo de la Figura 9A al complejo presentado en la Figura 5D.

La Figura 10 presenta un complejo de varias secuencias de ácidos nucleicos manipuladas diferentes que forman un andamio. En esta Figura, se proporcionan números de referencia para ayudar a ilustrar los componentes del andamio: la Figura 10, I, comparar con la Figura 7F; la Figura 10, II, comparar con la Figura 6H; la Figura 10, III, comparar con la Figura 8L; y la Figura 10, IV, comparar con la Figura 9A, en el que I y II están conectados a través de pares de bases unidos por hidrógeno en lugar de una conexión covalente; y la Figura 10, V en comparación con la Figura 5H.

Los tipos de conexiones entre uno o más componentes polinucleotídicos de una composición de polinucleótidos NASC incluyen, por ejemplo, enlaces covalentes y enlaces no covalentes.

Un ejemplo de un enlace no covalente es el enlace de hidrógeno. Los tipos de enlaces de hidrógeno se discuten anteriormente. Las realizaciones de la presente invención incluyen, pero no se limitan a, los siguientes tipos de enlaces de hidrógeno en pares de nucleótidos unidos por hidrógeno: enlace de hidrógeno de W-C, enlace de hidrógeno de W-C inverso, enlace de hidrógeno de Hoogsteen, enlace de hidrógeno de Hoogsteen inverso, enlace de hidrógeno oscilante, enlace de hidrógeno de oscilación inverso o combinaciones de los mismos.

Los componentes de polinucleótidos NASC están diseñados típicamente de modo que los elementos de repeticiones emparejados destinados a conectarse entre sí, particularmente si la conexión es a través de pares de bases con enlaces de hidrógeno, y solo forman conexiones (por ejemplo, enlaces de hidrógeno) entre los elementos de repeticiones emparejados. Formación de estructuras internas dentro de cada elemento de repetición que interfiere con los dos elementos de repeticiones que normalmente se evitan. Además, también se evitan las conexiones (por ejemplo, la formación de pares de bases con enlaces de hidrógeno) entre los elementos de repeticiones y otras regiones de los polinucleótidos del componente NASC.

Además de los enlaces covalentes y enlaces no covalentes, se pueden usar otros tipos de conexiones entre uno o más componentes polinucleótidos de una composición de polinucleótidos NASC que incluyen, pero no se limitan a, emparejamientos de unidades estructurales de unión ligando/ligando, y/o enlaces cruzados. Los pares de unidades estructurales de unión ligando/ligando incluyen, pero no se limitan a, una secuencia de ácidos nucleicos seleccionada y un aptámero correspondiente; y una estructura secundaria de ácidos nucleicos/una pequeña molécula, ion o proteína de unión a la estructura secundaria de ácidos nucleicos. Típicamente, un primer componente polinucleótido de una composición de polinucleótido NASC está adaptado para comprender un ligando

(por ejemplo, el primer componente polinucleótido de una composición de polinucleótido NASC comprende en su extremo 3' una secuencia de ácidos nucleicos seleccionada) y un segundo componente polinucleótido del polinucleótido NASC la composición está adaptada para comprender una unidad estructural de unión a ligando (por ejemplo, el segundo componente polinucleotídico de la composición polinucleotídica NASC comprende un aptámero en su extremo 5' de unión a la secuencia de ácidos nucleicos seleccionada).

Los agentes de entrecruzamiento útiles para formar conexiones entre uno o más componentes polinucleotídicos de una composición de polinucleótidos NASC incluyen, pero no se limitan a, agentes alquilantes (por ejemplo, 1,3-bis(2-cloroetil)-1-nitrosourea) y nitrógeno mostaza); cisplatino (cis-diaminodicloroplatino (II)) y sus derivados); radiación ionizante; ácido nitroso; productos químicos reactivos (por ejemplo, malondialdehído); psoralenos (activados en presencia de UV); y aldehídos (por ejemplo, acroleína y crotonaldehído).

En algunas realizaciones de la presente invención, las etiquetas de afinidad se introducen en dos o más componentes de polinucleótidos de una composición de polinucleótidos de NASC. Por ejemplo, una secuencia de ácidos nucleicos dentro de un componente polinucleotídico de una composición de polinucleótido NASC puede modificarse para comprender una secuencia de afinidad. Las proteínas efectoras de unión a ácidos nucleicos y sus correspondientes secuencias de unión a proteína efectora son ejemplos de etiquetas de afinidad. Se puede introducir una etiqueta de afinidad en los componentes de un primer polinucleótido de una composición de polinucleótido NASC. Una etiqueta de afinidad puede ser una secuencia de afinidad, como la secuencia de unión MS2, la secuencia de unión U1A, la secuencia tallo-bucle (por ejemplo, una secuencia de unión a la proteína Csy4, o la secuencia de unión a la proteína Cas6), la secuencia de unión a eIF4A, el efector de tipo activador de la transcripción (TALE) secuencia de unión (véase, por ejemplo, Valton, J., et al., Journal of Biological Chemistry 287(46): 38427-38432 (2012)), o secuencia de unión del dominio del dedo de zinc (véase, por ejemplo, Font, J., et al., Methods Molecular Biology 649: 479-491 (2010); Isalan, M., et al., Nature Biotechnology 19(7): 656-660 (2001)). Un segundo componente polinucleotídico de la composición polinucleotídica de NASC puede modificarse para comprender una etiqueta de afinidad correspondiente: una secuencia codificadora de MS2, secuencia codificadora de U1A, secuencia codificadora de la proteína de unión del tallo-bucle (por ejemplo, una proteína Csy4 enzimáticamente (endorribonucleasa) inactiva de unión a la proteína Csy4 secuencia de proteína), secuencia codificadora de eIF4A, secuencia codificadora de TALE o una secuencia codificadora de dominio de dedo de zinc, respectivamente. Típicamente, se usan proteínas de unión a ácidos nucleicos inactivas enzimáticamente que retienen la unión de ácidos nucleicos específica de secuencia (por ejemplo, una proteína Csy4 inactiva con endorribonucleasa (dCsy4)); sin embargo, en algunas realizaciones se usan proteínas de unión a ácidos nucleicos enzimáticamente activas o proteínas de ácidos nucleicos con actividad enzimática alterada. Cuando más de dos componentes de polinucleótidos de una composición de polinucleótidos de NASC se modifican con una secuencia de afinidad, en realizaciones preferidas, las dos secuencias de afinidad típicamente no son las mismas; por lo tanto, hay dos secuencias de afinidad diferentes asociadas con la proteína Cas.

El ejemplo 1 describe la producción de componentes ejemplares de composiciones polinucleotídicas NASC manipuladas. El Ejemplo 1 describe *in silico* el diseño de componentes de polinucleótidos de NASC correspondientes a una serie de realizaciones de las composiciones de polinucleótidos de NASC descritas en este documento. La Tabla 9 muestra una correlación entre los componentes de polinucleótido NASC y las estructuras ilustradas en las Figuras.

El Ejemplo 2 describe la producción de componentes polinucleotídicos de NASC de la presente invención. Los componentes de polinucleótidos de NASC descritos en este Ejemplo se usaron en ensayos de escisión de Cas *in vitro* para evaluar los porcentajes de escisión de secuencias diana de ácidos nucleicos por las composiciones de polinucleótidos de NASC. El ejemplo 5 describe el rendimiento de los ensayos de escisión mediados por la proteína Cas *in vitro*. El Ejemplo 3 y el Ejemplo 4 describen procedimientos que pueden usarse para la producción de secuencias diana de ADN de doble cadena para uso en ensayos de división de Cas *in vitro*.

El ejemplo 6 presenta un análisis de secuenciación profunda para la detección de modificaciones diana en células eucarióticas usando composiciones de polinucleótidos NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de clase 2 CRISPR-Cas) de la presente invención.

El Ejemplo 9 presenta un análisis alternativo, el ensayo T7E1, para la detección de modificaciones diana en células eucariotas usando composiciones de polinucleótidos NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, CRISPR-Cas de Clase 2 proteínas). El Ejemplo 7 describe la identificación y el cribado de los ARNcr de clase 2 que pueden diseñarse para hacer componentes de polinucleótido NASC de la presente invención.

El Ejemplo 8 describe la identificación y selección de traARN-crs de clase 2 que pueden usarse para diseñar componentes de polinucleótidos de NASC.

El Ejemplo 10 describe la generación y prueba de diversas modificaciones de los ARNrc de guía Clase 2 Tipo V y su idoneidad para su uso en la construcción de componentes de polinucleótidos de NASC.

El Ejemplo 11 describe la generación y prueba de diversas modificaciones de los ARN de guía de Clase 2 Tipo II y su idoneidad para su uso en la construcción de componentes de polinucleótido NASC.

El Ejemplo 12 describe el uso de composiciones de polinucleótidos de NASC para modificar las secuencias diana de ácidos nucleicos presentes en el ADN humano y medir el nivel de actividad de escisión y la especificidad de escisión en esos sitios. La medición del nivel de porcentaje de escisión y/o especificidad de escisión en un sitio particular puede proporcionar opciones para identificar las secuencias diana de ácidos nucleicos que tienen un porcentaje y/o especificidad de escisión deseados.

En otro aspecto, la presente invención está dirigida a composiciones de ácidos nucleicos/proteína que comprenden una composición de polinucleótido NASC complejada con una primera proteína CRISPR-Cas9 Clase 2 Tipo II y una segunda proteína CRISPR-Cas9 Clase 2 Tipo II. La primera proteína CRISPR-Cas9 Clase 2 Tipo II puede comprender una o más actividades de nucleasa, y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II puede comprender una o más actividades de nucleasa. En algunas realizaciones, la primera proteína CRISPR-Cas9 de Tipo II de Clase 2 es catalíticamente inactiva para una o más de las actividades de nucleasa, la segunda proteína CRISPR-Cas9 de Tipo II de Clase 2 es inactiva catalíticamente para una o más de las actividades de nucleasa, o ambas la primera proteína CRISPR-Cas9 Clase 2 Tipo II es inactiva catalíticamente para una o más de las actividades de nucleasa y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II es catalíticamente inactiva para una o más de las actividades de nucleasa. En otras realizaciones de la composición de polinucleótidos de NASC/primer proteína CRISPR-Cas9 Clase 2 Tipo II/segunda de la Clase 2 Tipo II de complejos de proteína CRISPR-Cas9, ya sea la primera Clase 2 de Tipo II CRISPR-Cas9 o la segunda clase 2 de Tipo II CRISPR- La proteína Cas9 es inactiva catalíticamente, y los complejos pueden además ser conectados con un polinucleótido donante a través de la proteína catalíticamente inactiva.

En algunos aspectos de la composición de polinucleótidos NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos, ya sea la proteína Cas9 o la proteína Cpf1 es catalíticamente inactiva (dCas9 o dCpf1) y la proteína/segunda unión composición de polinucleótidos NASC primera/ácido nucleico unión composición de proteína de ácidos nucleicos comprende además un polinucleótido donante en el que el polinucleótido donante comprende una secuencia de nucleótidos complementaria al elemento espaciador CPFL, o las regiones adyacentes al elemento espaciador CPFL, o una secuencia de nucleótidos complementaria al elemento espaciador, o las regiones adyacentes al elemento espaciador Cas9. El polinucleótido donante es capaz de asociarse con el elemento separador, o de las regiones adyacentes al elemento separador, a través de enlaces de hidrógeno entre la secuencia de nucleótidos polinucleótido donante complementaria al elemento espaciador, o la secuencia adyacente al elemento espaciador.

Las mutaciones de la proteína Cas9 que son enzimáticamente inactiva para la actividad de nucleasa relacionados con RuvC-1, la actividad de nucleasa relacionadas con HNH, y tanto la actividad de nucleasa relacionados con RuvC-1 y actividad de nucleasa relacionados HNH- son conocidos en la técnica. Las mutaciones de la proteína CPFL que son enzimáticamente inactiva se conocen en la técnica (véase, por ejemplo, Yamano, T., et al, Cell 165(4): 949-962 (2016)); Zetsche, B., et al., Cell 163: 1-13 (2015)).

En los sistemas CRISPR, la "biogénesis guía" (también conocida como "procesamiento guía") implica el truncamiento de endonucleasa o exonucleasa de la secuencia del ARN guía luego de la transcripción de la matriz CRISPR. El procesamiento enzimático del ARN guía puede llevarse a cabo por ARNasas codificadas por el operón Cas (por ejemplo, Cas6 de sistemas de Clase I Tipo I-E) o por ARNasas endógenas (por ejemplo, ARNasa III de sistemas de Clase 2 Tipo II-A).

En sistemas de Clase 2 Tipo V, la biogénesis de guía se realiza mediante la nucleasa de la proteína Cpf1. La proteína Cpf1 también es responsable de la escisión diana del ADN de doble cadena específica de la secuencia.

En el sistema del Tipo V, la división del pre-ARNcr se produce en una región en sentido ascendente (por ejemplo, en una dirección 5') desde la estructura secundaria del pseudonudo (véase, por ejemplo, la Figura 3A, 303) y da como resultado la generación de una guía Cpf1 ARNcr. En algunos aspectos, prevenir que la proteína Cpf1 se rompa en 5' del elemento de tallo ARNcr guía es útil, por ejemplo, para prevenir la separación de una composición de polinucleótido NASC/proteína Cas9/proteína Cpf1 por escisión mediada por Cpf1. Se ha demostrado que la secuencia de la proteína Cpf1 CRISPR de Tipo V se puede modificar para prevenir el procesamiento del ARN guía mediante la proteína Cpf1 Tipo V CRISPR (véase Fonfara, I., et al., Nature 532(7600): 517-521 (2016)).

Un procedimiento para prevenir la escisión de Cpf1 de las secuencias 5' de la guía ARNcr elemento de tallo es por modificación (por ejemplo, mutaciones de base, inserciones, eliminaciones o modificaciones químicas) de las bases en la región corriente arriba del pseudonudo o dentro de la pseudonudo. Nudo del pre-ARNcr para prevenir el procesamiento del pre-ARNcr por la proteína Cpf1. Para evaluar el efecto de tales modificaciones en el procesamiento de la guía, el pre-ARNcr manipulado puede incubarse en presencia de una proteína Cpf1 afin durante un período de tiempo en un tampón adecuado. La mezcla se puede tratar con Proteinasa K (Denville Scientific, South Plainfield, NJ) para eliminar la proteína y la mezcla se puede analizar mediante electroforesis en gel de poli(acrilamida) para evaluar si se produce la escisión del pre-ARNcr manipulado. Un pre-ARNcr no incubado en presencia de una proteína Cpf1 afin puede servir como un control positivo (es decir, un control para la ausencia de

procesamiento de la guía). Si no es suficiente una sola modificación en el pre-ARNcr para eliminar el procesamiento de la guía, entonces las combinaciones de modificaciones que muestran un procesamiento reducido del pre-ARNcr se pueden combinar en un diseño de pre-ARNcr y volver a probar la ausencia de actividad de procesamiento de la guía. Las modificaciones de pre-ARNcr que resultan en la incapacidad del pre-ARNcr manipulado para ser procesado pueden evaluarse adicionalmente para determinar la capacidad del complejo de proteína Cpf1-pre-ARNcr/Cpf1 para mantener la unión específica de secuencia y/o la escisión de un ADN de un ácido nucleico diana que comprende el elemento espaciador pre-ARNcr.

Un segundo procedimiento para prevenir la escisión de Cpf1 de las secuencias 5' del elemento de tallo ARNcr guía es mediante la modificación de la proteína Cpf1. En este procedimiento, los residuos de aminoácidos de la proteína Cpf1 se modifican para perturbar el procesamiento de la guía. La cristalografía de rayos X de los complejos de proteína ARNcr/Cpf1 de guía ha demostrado que el pseudonudo está unido por la interfaz de dos dominios de proteínas designados como el dominio de cuña (WED) y el dominio RuvC (véase Yamano, T., et al., *Cell* 165(4): 949-962 (2016)). Es probable que los residuos de aminoácidos de Cpf1 proximal a la región de unión al extremo 5' del ARN-cr guía y/o la estructura de pseudonudo estén involucrados en la catálisis por endonucleasa de los pre-ARNcr estrategias de mutagénesis, como la detección de alanina (véase, por ejemplo, Lefèvre, F., et al., *Nucleic Acids Research* 25(2): 447-448 (1997); Lee, et al., *Molecular Pharmacology* 50(1): 140-148 (1996)) se puede usar para modificar regiones dentro de los dominios WED y RuvC, u otros dominios dentro de la proteína Cpf1, para identificar residuos en la proteína responsable del procesamiento del ARNcr de la guía. En este procedimiento, las proteínas Cpf1 que comprenden mutaciones de alanina pueden expresarse e incubarse con un pre-ARNcr cognado en un tampón adecuado. Después de la incubación, se puede agregar Proteinasa K a la reacción de la mezcla para eliminar la proteína Cpf1 y la mezcla de reacción se pueden analizar mediante electroforesis en gel de poliacrilamida para evaluar si se produjo la escisión del pre-ARNcr manipulado. Un pre-ARN-cr no incubado en presencia de una proteína Cpf1 afín puede servir como un control positivo (es decir, un control para la ausencia de procesamiento de la guía). Si no es suficiente una sola mutación en la proteína Cpf1 para eliminar el procesamiento de la guía, entonces las combinaciones de mutaciones que muestran un procesamiento reducido del pre-ARNcr se pueden combinar en una sola construcción de la proteína Cpf1 y volver a probar la ausencia de actividad de procesamiento de la guía. Las mutaciones candidatas o combinaciones de mutaciones en la proteína Cpf1 pueden evaluarse adicionalmente para determinar la capacidad del complejo Cpf1-pre-ARNcr para mantener la unión específica de la secuencia y/o la escisión de un ácido nucleico diana del ADN que comprende el elemento espaciador pre-ARNcr.

En otro aspecto, la presente invención se refiere a secuencias de ácidos nucleicos que codifican uno o más componentes polinucleótidos de una composición de polinucleótido NASC, así como casetes de expresión, vectores y células recombinantes que comprenden secuencias de ácidos nucleicos que codifican uno o más componentes polinucleótidos de una composición polinucleotídica de NASC. En algunas realizaciones, tales casetes de expresión, vectores y células recombinantes comprenden además secuencias que codifican una o más proteínas CRISPR-Cas9 de Clase 2 Tipo II con las que la composición de polinucleótidos de NASC es capaz de formar un complejo.

Una realización adicional de la presente invención se refiere a vectores, que incluyen vectores de expresión, que comprenden una o más secuencias de ácidos nucleicos que codifican uno o más componentes polinucleótidos de una composición de polinucleótido de NASC, y opcionalmente una o más secuencias de ácidos nucleicos que codifican Clase 2 Tipo II Proteína CRISPR-Cas9 capaz de formar un complejo con la composición del polinucleótido NASC. Los vectores también pueden incluir secuencias que codifican marcadores seleccionables o cribables. Además, las secuencias de direccionamiento nuclear también se pueden agregar, por ejemplo, a las secuencias codificantes de la proteína Cas9. Los vectores también pueden incluir polinucleótidos que codifican etiquetas de proteínas (por ejemplo, etiquetas de poli-His, etiquetas de hemaglutinina, etiquetas de proteínas fluorescentes, etiquetas bioluminiscentes). Las secuencias de codificación para tales etiquetas de proteínas pueden fusionarse a, por ejemplo, una o más secuencias de ácidos nucleicos que codifican una proteína Cas9.

Los procedimientos generales para la construcción de vectores de expresión son conocidos en la técnica; además, los vectores de expresión para células huésped están disponibles comercialmente. Existen varios productos de software comerciales diseñados para facilitar la selección de vectores apropiados y su construcción, como los vectores de células de insectos para la transformación de células de insectos y la expresión de genes en células de insectos, los plásmidos bacterianos para la transformación bacteriana y la expresión de genes en células bacterianas, los plásmidos de levadura para la transformación de células. y la expresión de genes en levaduras y otros hongos, vectores de mamíferos para la transformación de células de mamíferos y la expresión de genes en células de mamíferos o mamíferos, y vectores virales (que incluyen lentivirus, retrovirus, adenovirus, virus del herpes simple I o II, parvovirus, virus de la reticuloendoteliosis y vectores de virus adeno asociados (AAV)) para la transformación celular y la expresión de genes y procedimientos para permitir fácilmente la clonación de tales polinucleótidos. Los vectores de transformación de plantas ilustrativos incluyen aquellos derivados de un plásmido Ti de *Agrobacterium tumefaciens* (Lee, L.Y., et al., *Plant Physiology* 146(2): 325-332 (2008)). También son útiles y conocidos en la técnica los plásmidos de *Agrobacterium rhizogenes*. Por ejemplo, SNAPGENE™ (GSL Biotech LLC, Chicago, IL; snapgene.com/resources/plasmid_files/your_time_is_valuable/) proporciona una lista extensa de vectores, secuencias de vectores individuales y mapas de vectores, así como fuentes comerciales para muchos de los vectores.

Los vectores lentivíricos son ejemplos de vectores útiles para la introducción en células de mamíferos de una o más secuencias de ácidos nucleicos que codifican uno o más componentes polinucleotídicos de una composición de polinucleótido NASC, y opcionalmente una o más secuencias de ácidos nucleicos que codifican una o más proteínas de unión a ácidos nucleicos (por ejemplo, Proteínas CRISPR-Cas de Clase 2) con las que la composición del polinucleótido NASC es capaz de formar un complejo. Lentivirus es un miembro de la familia Retroviridae y es un virus de ARN de una sola hebra, que puede infectar tanto células en división como no en división, así como proporcionar una expresión estable a través de la integración en el genoma. Para aumentar la seguridad de los lentivirus, los componentes necesarios para producir un vector viral se dividen en múltiples plásmidos. Los vectores de transferencia suelen ser incompetentes con la replicación y, además, pueden contener una eliminación en el 3'LTR, lo que hace que el virus se autoinactive después de la integración. Los plásmidos de empaquetamiento y envoltura se usan típicamente en combinación con un vector de transferencia. Por ejemplo, un plásmido de empaquetamiento puede codificar combinaciones de los genes Gag, Pol, Rev y Tat. Un plásmido de transferencia puede comprender LTR virales y la señal de empaquetamiento psi. El plásmido de la envoltura comprende una proteína de la envoltura (generalmente la glicoproteína del virus de la estomatitis vesicular, VSV-GP, debido a su amplio rango de infectividad).

Los vectores lentivíricos basados en el virus de inmunodeficiencia humana tipo 1 (VIH-1) tienen proteínas accesorias adicionales que facilitan la integración en ausencia de división celular. Los vectores de VIH-1 han sido diseñados para abordar una serie de problemas de seguridad. Estos incluyen la expresión separada de los genes virales en trans para prevenir los eventos de recombinación que conducen a la generación de virus competentes para la replicación. Además, el desarrollo de vectores autoinactivadores reduce el potencial de transactivación de genes vecinos y permite la incorporación de elementos tampones para dirigir la expresión génica a tipos celulares particulares (véase, por ejemplo, Cooray, S., et al., *Methods in Enzymology* 507: 29-57 (2012)).

Las células huésped transformadas (o células recombinantes) son células o la progenie de células que se han transformado o transfectado, utilizando técnicas de ADN recombinante, con una o más secuencias de ácidos nucleicos que codifican uno o más componentes polinucleotídicos de una composición de polinucleótido NASC, y opcionalmente uno o más secuencias de ácidos nucleicos que codifican una o más proteínas de unión a ácidos nucleicos (por ejemplo, proteínas CRISPR-Cas de Clase 2) con las que la composición de polinucleótido NASC es capaz de formar un complejo. Los procedimientos de introducción de polinucleótidos (por ejemplo, un vector de expresión) en células huésped se conocen en la técnica y se seleccionan típicamente en función del tipo de célula huésped. Tales procedimientos incluyen, por ejemplo, infección viral o bacteriófaga, transfección, conjugación, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina, transfección mediada por DEAE-dextrano, fusión de protoplastos, lipofección, transfección mediada por liposomas, tecnología de transferencia de genes balística (por ejemplo, utilizando una pistola de genes o un sistema de suministro de partículas biolísticas), microinyección directa y suministro mediado por nanopartículas.

Como alternativa a la expresión de una o más secuencias de ácidos nucleicos que codifican uno o más componentes polinucleotídicos de una composición de polinucleótido NASC, y opcionalmente una o más secuencias de ácidos nucleicos que codifican una o más proteínas de unión a ácidos nucleicos (por ejemplo, proteínas Clase 2 CRISPR-Cas) con las que la composición de polinucleótidos NASC es capaz de formar un complejo, una composición de polinucleótidos NASC y/o la una o más proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de clase 2 CRISPR-Cas) pueden introducirse directamente en una célula, por ejemplo. Alternativamente, uno o más componentes pueden ser expresados por una celda y el otro componente(s) introducido directamente. Los procedimientos para introducir los componentes en una célula incluyen electroporación, lipofección y tecnología de transferencia de genes balísticos.

En el presente documento se describe una variedad de células huésped que se pueden usar para producir células recombinantes mediante la introducción de una o más secuencias de ácidos nucleicos que codifican uno o más componentes polinucleotídicos de una composición de polinucleótido NASC, y opcionalmente una o más secuencias de ácidos nucleicos que codifican una o más secuencias nucleicas proteínas de unión a ácidos (por ejemplo, proteínas CRISPR-Cas de Clase 2) con las que la composición de polinucleótidos NASC es capaz de formar un complejo. Dichas células huésped incluyen, pero no se limitan a, una célula vegetal, una célula de levadura, una célula bacteriana, una célula de insecto, una célula de algas o una célula de mamífero.

Los procedimientos para introducir polinucleótidos (por ejemplo, un vector de expresión) en células huésped para producir células recombinantes son conocidos en la técnica y se seleccionan típicamente en función del tipo de célula huésped. Tales procedimientos incluyen, por ejemplo, infección viral o bacteriófaga, transfección, conjugación, electroporación, precipitación con fosfato de calcio, transfección mediada por polietilenimina, transfección mediada por DEAE-dextrano, fusión de protoplastos, lipofección, transfección mediada por liposomas, tecnología de transferencia de genes balística, microinyección directa, y la entrega mediada por nanopartículas. Para facilitar la discusión, la "transfección" se usa a continuación para referirse a cualquier procedimiento de introducción de polinucleótidos en una célula huésped.

Los procedimientos preferidos para introducir células de plantas de polinucleótidos incluyen el bombardeo con microproyectiles y la transformación mediada por *Agrobacterium*. Alternativamente, se pueden usar otras especies que no son *Agrobacterium* (por ejemplo, *Rhizobium*) y otras células procarióticas que pueden infectar células

vegetales e introducir polinucleótidos heterólogos en el genoma de la célula vegetal infectada. Otros procedimientos incluyen la electroporación, la transfección mediada por liposomas, la transformación con polen o virus, y los productos químicos que aumentan la captación libre de ADN, o la liberación libre de ADN mediante el bombardeo con microproyectiles. Véase, por ejemplo, Narusaka, Y., et al., Chapter 9, in *Transgenic Plants - Advances and Limitations*, edited by Yelda, O., ISBN 978-953-51-0181-9 (2012).

En algunos aspectos, una célula huésped se transfecta de forma transitoria o no transitoria. En algunos aspectos, una célula se transfecta como ocurre naturalmente en un sujeto. En algunos aspectos, una célula que se transfecta se toma de un sujeto, por ejemplo, una célula primaria o una célula progenitora. En algunos aspectos, la célula primaria o la célula progenitora se cultiva y/o se devuelve después de la transfección *ex vivo* al mismo sujeto (tratamiento autólogo) o a un sujeto diferente.

Los complejos de la composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda proteína de unión a ácidos nucleicos (que comprenden, por ejemplo, proteínas de Clase 2 CRISPR-Cas) descritos aquí se pueden usar para generar organismos transgénicos no humanos introduciendo específicamente el sitio una secuencia polinucleotídica seleccionada en un locus diana de ADN en el genoma para generar una modificación del ADN. El organismo transgénico puede ser un animal o una planta.

Un animal transgénico se genera típicamente introduciendo el sistema en una célula cigótica. Una técnica básica, descrita con referencia a la fabricación de ratones transgénicos (Cho, A., et al., *Generation of Transgenic Mice*, "Current Protocols in Cell Biology, CHAPTER.Unit-19.11 (2009)), implica cinco pasos básicos: primero, preparación de un sistema, como se describe en el presente documento, que incluye un polinucleótido donante adecuado; en segundo lugar, la recolección de cigotos donantes; tercero, la microinyección del sistema en el cigoto del ratón; cuarto, la implantación de cigotos microinyectados en ratones hembra receptores pseudoembarazadas; y quinto, realizar el genotipificado y el análisis de la modificación del ADN establecido en ratones fundadores. Los ratones fundadores pasarán la modificación genética a cualquier progenie. Los ratones fundadores son típicamente heterocigotos para el transgén. El apareamiento entre estos ratones producirá ratones que son homocigotos para el transgén el 25% del tiempo.

Los procedimientos para generar plantas transgénicas también son bien conocidos. Una planta transgénica generada, por ejemplo, utilizando procedimientos de transformación de *Agrobacterium*, típicamente contiene un transgén insertado en un cromosoma. Es posible producir una planta transgénica que sea homocigótica con respecto a un transgén al aparearse sexualmente (es decir, autofecundarse) una planta transgénica segregante independiente que contiene un solo transgén para sí misma, por ejemplo, una planta F0, para producir una semilla F1. Las plantas formadas por la germinación de las semillas F1 se pueden probar para determinar la homocigosidad. Los ensayos típicos de zigosidad incluyen, pero no se limitan a, ensayos de polimorfismo de nucleótido único y ensayos de amplificación térmica que distinguen entre homocigotos y heterocigotos.

Como alternativa al uso de un sistema descrito en el presente documento para la transformación directa de una planta, las plantas transgénicas pueden formarse cruzando una primera planta que se ha transformado con un sistema con una segunda planta que nunca ha sido expuesta al sistema. Por ejemplo, una primera línea de plantas que contiene un transgén puede cruzarse con una segunda línea de plantas para introgresar el transgén en la segunda línea de plantas, formando así una segunda línea de plantas transgénicas.

Otros aspectos de la presente divulgación se refieren a procedimientos para usar composiciones de nucleoproteínas que comprenden composiciones de polinucleótidos de NASC y una proteína de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas). Tales composiciones de nucleoproteínas se describen en el presente documento. Numerosos usos de las secuencias de ácidos nucleicos manipuladas descritas en el presente documento incluyen, entre otras, formar un andamio de un complejo de dos o más secuencias de ácidos nucleicos manipuladas que comprenden una secuencia de unión a la proteína CRISPR de Clase 2 de unión a los ácidos nucleicos y una secuencia de ácidos nucleicos espaciadores complementaria a una secuencia de ácidos nucleicos diana; edición precisa de las regiones de ADN (por ejemplo, escisión, inserción, modificación); atar un polinucleótido donante muy cerca de un sitio de corte (por ejemplo, escisión utilizando una proteína Cas9 o una proteína Cpf1); la escisión de una región de ADN y la unión simultánea de polinucleótido donante en el sitio de escisión; formando una histona artificial o introducción de estructura de heterocromatina, por ejemplo, utilizando dCas9; y un estricto control transcripcional de la expresión génica (por ejemplo, el bloqueo de la transcripción de un gen). Los usos adicionales de los andamios de secuencia de ácidos nucleicos diseñados en el presente documento incluyen, entre otros, procedimientos de uso y procedimientos de fabricación de láminas de partículas de nucleoproteína; biomateriales flexibles, por ejemplo, para uso en ingeniería de tejidos; vehículos de suministro de drogas enjaulados vehículos de administración de vacunas, por ejemplo, vacunas de ADN o ARN; membranas porosas de tamaño regulado, por ejemplo, haciendo y usando membranas que tienen agujeros de tamaño fijo; nanopartículas de tamaños seleccionados; y polímeros de ácidos nucleicos de proteínas.

En un aspecto, la presente divulgación incluye un procedimiento para unir una secuencia de ácidos nucleicos (por ejemplo, ADN) que comprende poner en contacto una primera secuencia diana de ácidos nucleicos en el ácido nucleico (por ejemplo, ADN) y una segunda secuencia diana de ácidos nucleicos en la secuencia de ácidos nucleicos (por ejemplo, ADN) con una composición de polinucleótido NASC/primer proteína de unión a ácidos

nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de Clase 2 CRISPR-Cas), facilitando así la unión de la nucleoproteína a la primera secuencia diana de ácidos nucleicos en la secuencia de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos en el ácido nucleico. La composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de Clase 2 CRISPR-Cas) comprende un primer elemento espaciador que es complementario a la primera secuencia diana de ácidos nucleicos (por ejemplo, ADN) y un segundo elemento espaciador que es complementario a la segunda secuencia diana de ácidos nucleicos (por ejemplo, ADN). En algunos aspectos, la secuencia diana de ácidos nucleicos es ADN. Dichos procedimientos de unión a una secuencia diana de ácidos nucleicos se pueden llevar *in vitro* (un ensayo bioquímico), en células (en células cultivadas), *ex vivo* (células eliminadas de un sujeto) o *in vivo* (células en un organismo).

Se conoce una variedad de procedimientos en la técnica para evaluar y/o cuantificar interacciones proteína-ácido nucleico que incluyen, entre otras, las siguientes: ensayos de inmunoprecipitación (ChIP), ensayos de desplazamiento de movilidad electroforética de ADN (EMSA), extracción de ADN -Análisis hacia abajo, y ensayos de captura y detección de microplacas. Kits, materiales y reactivos comerciales están disponibles para practicar muchos de estos procedimientos, por ejemplo, Thermo Scientific (Wilmington, DE), Signosis (Santa Clara, CA), Bio-Rad (Hercules, CA) y Promega (Madison, WISCONSIN). Un enfoque común para detectar interacciones proteína-ácido nucleico es EMSA (véase, por ejemplo, Hellman L.M., et al., Nature Protocols 2(8): 1849-1861 (2007)).

En otro aspecto, la presente divulgación incluye un procedimiento para cortar una secuencia de ácidos nucleicos (por ejemplo, ADN) que comprende poner en contacto una primera secuencia diana de ácidos nucleicos en el ácido nucleico (por ejemplo, ADN) y una segunda secuencia diana de ácidos nucleicos en la secuencia de ácidos nucleicos (por ejemplo, ADN) con una composición de nucleoproteínas que comprende una composición de polinucleótidos NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de clase 2 CRISPR-Cas), facilitando así la unión de la composición de nucleoproteína a la primera secuencia diana de ácidos nucleicos en la secuencia de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos en el ácido nucleico. La composición de nucleoproteínas comprende un primer elemento espaciador que es complementario a la primera secuencia diana de ácidos nucleicos (por ejemplo, ADN) y un segundo elemento espaciador que es complementario a la segunda secuencia diana de ácidos nucleicos (por ejemplo, ADN). La primera proteína de unión a ácidos nucleicos (por ejemplo, la proteína de Clase 2 CRISPR-Cas) de la composición de nucleoproteínas unida corta la primera secuencia diana de ácidos nucleicos, y la segunda proteína de unión a ácidos nucleicos (por ejemplo, la proteína de Clase 2 CRISPR-Cas) de la composición ácido nucleico unido /proteína corta la segunda secuencia diana de ácidos nucleicos. En algunos aspectos, la secuencia diana de ácidos nucleicos es ADN. Tales procedimientos de unión a una secuencia diana de ácidos nucleicos se pueden llevar *in vitro*, *en células*, *ex vivo* o *in vivo*.

Los procedimientos de unión y de unión y corte de secuencias diana de ácidos nucleicos usando una composición de proteína Cas9/*S. pyogenes* NASC-PC1/NASC-PC2/*S. thermophilus* se ilustran en la Figura 16A, Figura 16B, y Figura 16C. La Figura 16A ilustra una proteína Cas9 de *S. pyogenes* (Figura 16A, 1604) y una proteína *S. thermophilus* Cas9 (Figura 16A, 1603), una composición NASC-PC1/NASC-PC2 (Figura 16A, 1600) (generalmente estructura mostrada en la Figura 6K), un ácido nucleico bicatenario (Figura 16A, 1605) que comprende una primera secuencia de unión a diana del ADN complementaria al elemento espaciador Cas9 NAS de PCC/PCC de NASC-PC2 DE *S. pyogenes* (Figura 16A, 1602), y un ácido nucleico bicatenario (Figura 16A, 1607) que comprende una segunda secuencia de unión a diana de ADN complementaria al elemento espaciador Cas9 NASC-PC1/NASC-PC2 DE *S. thermophilus* Cas9 (Figura 16A, 1601). La Figura 16A, 1606, indica la ubicación del Cas9 de *S. pyogenes* PAM. La Figura 16A, 1608, indica la ubicación del *S. thermophilus* Cas9 PAM.

La Figura 16A ilustra la formación de la proteína Cas9 de *S. pyogenes* (Figura 16A, 1604; complejo 1609) y la proteína Cas9 de *S. thermophilus* (Figura 16A, 1603; complejo 1616) en complejo con la composición NASC-PC1/NASC-PC2 (Figura 11A, 1600; complejo 1610).

La Figura 16A ilustra el enlace de hidrógeno del complejo de nucleoproteínas a las secuencias diana de ADN de doble cadena (Figura 16A, 1611). La Figura 16A, 1611, ilustra la unión de la proteína NAS9-PC1/NASC-PC2/*S. thermophilus* Cas9/Cas9 de *S. pyogenes* al ácido nucleico bicatenario (Figura 16A, 1605) que comprende una primera secuencia de unión a la diana del ADN complementario al elemento espaciador Cas9 de NASC-PC1/NASC-PC2 de *S. pyogenes* (Figura 16A, 1602) y un ácido nucleico bicatenario (Figura 16A, 1607) que comprende una segunda secuencia de unión a diana de ADN complementaria al elemento espaciador NASC-PC1/NASC-PC2 DE *S. thermophilus* Cas9 (Figura 16A, 1601). Si la proteína Cas9 de *S. pyogenes* y la proteína Cas9 de *S. thermophilus* están inactivas enzimáticamente, el complejo de nucleoproteínas (Figura 16A, 1610) se puede usar, por ejemplo, para traer dos secuencias de ADN (Figura 16A, 1605, 1607) a proximidad (por ejemplo, Figura 16A, 1607).

La Figura 16B ilustra la escisión de la Figura 16B, 1605, ADN por una proteína enzimáticamente activa Cas9 de *S. pyogenes* y el anclaje de la Figura 16B, 1607, ADN que usa una proteína *S. thermophilus* Cas9 enzimáticamente inactiva para mantener el ADN (Figura 16B, 1607) cerca del sitio de escisión (Figura 16B, 1612, 1613). Dicho complejo de nucleoproteínas puede ayudar a mejorar la frecuencia de HDR utilizando un polinucleótido donante (Figura 16B, 1607).

La Figura 16C ilustra la escisión de la Figura 16C, 1605, ADN por una proteína enzimáticamente activa Cas9 de *S. pyogenes* para romper ambas cadenas de la Figura 16C, 1605, ADN (Figura 16C, 1612, 1613) y escisión de la Figura 16C, 1607, ADN por una proteína enzimáticamente activa *S. thermophilus* Cas9 para romper ambas cadenas en la Figura 16C, 1607, ADN (Figura 16C, 1614, 1615). Tal complejo de nucleoproteínas se puede usar para facilitar el reordenamiento cromosómico (por ejemplo, las translocaciones).

En otro aspecto más, la presente divulgación incluye un procedimiento para modificar el ADN en una célula que comprende poner en contacto una primera secuencia diana de ADN en el ADN y una segunda secuencia diana de ADN en el ADN con una composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de la proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas CRISPR-Cas de Clase 2, como la proteína Cas9 y/o proteína Cpf1), facilitando así la unión del complejo de nucleoproteína a la primera secuencia diana de ácidos nucleicos en la secuencia de ácidos nucleicos y la segunda secuencia diana de ácida nucleico en el ácido nucleico. La composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de Clase 2 CRISPR-Cas) comprende un primer elemento espaciador que es complementario a la primera secuencia diana de ácidos nucleicos y un segundo elemento espaciador es complementario a la segunda secuencia diana de ácidos nucleicos (por ejemplo, ADN). La primera proteína del complejo de nucleoproteínas unidas corta la primera secuencia diana de ADN, y la segunda proteína del complejo de nucleoproteínas unida corta la segunda secuencia diana de ADN. La celda repara tanto el primer sitio de corte como el segundo sitio de corte. Las vías ejemplares de reparación del ADN celular incluyen HDR, NHEJ y MMEJ. En algunos aspectos, la secuencia diana de ácidos nucleicos es ADNg. Dichos procedimientos de unión a una secuencia diana de ácidos nucleicos pueden llevarse a cabo *in vitro*, *en células*, *ex vivo* o *in vivo*. La etapa de contratación puede comprender además un polinucleótido donante que está presente, en el que al menos una porción del polinucleótido donante está incorporada entre el primer sitio de corte y el segundo sitio de corte.

En otro aspecto, la divulgación se refiere a un procedimiento para llevar un polinucleótido donante a la proximidad de un DSB en una diana de ácidos nucleicos, típicamente ADN, en una célula. El procedimiento comprende poner en contacto una primera secuencia diana de ADN en el ADN y una segunda secuencia diana de ADN en un polinucleótido donante con una composición de polinucleótido NASC/primer proteína de unión a ADN/segunda composición de proteína de unión a ADN (que comprende, por ejemplo, proteínas de Clase 2 CRISPR-Cas, tales como como proteína Cas9 y/o proteína Cpf1) que tienen una primera secuencia de unión a diana de ADN complementaria a la primera diana de ADN y una segunda secuencia de unión a diana de ADN complementaria a la segunda diana de ADN. La primera proteína de unión al ADN es catalíticamente activa y está asociada con la primera secuencia de unión al ADN. La segunda proteína de unión al ADN es enzimáticamente inactiva y está asociada con la segunda secuencia de unión a la diana del ADN. Poner en contacto el complejo de nucleoproteínas con la primera y la segunda secuencia diana de ADN facilita la unión del complejo de nucleoproteína a la primera secuencia diana de ADN en el ADN y la segunda secuencia diana de ADN en el polinucleótido donante. La proteína de unión al ADN catalíticamente activa del complejo de nucleoproteínas corta la primera secuencia diana de ADN para formar un sitio de corte. El polinucleótido donante está cerca del sitio de corte (por ejemplo, el DSB) porque la proteína de unión al ADN catalíticamente activa y la proteína de unión al ADN catalíticamente inactiva son complejadas con la composición del polinucleótido NASC, es decir, son parte del mismo complejo de nucleoproteína. En algunos aspectos, al menos una parte del polinucleótido donante se introduce en el sitio de corte en el ADN (por ejemplo, mediante un proceso de reparación de HDR) que resulta en la modificación del ADN.

La Figura 12 ilustra el uso de NASC-PC1/NASC-PC2/proteína Cas9 activa/dCas9 proteína Cas9, en la que los dominios de endonucleasa del Cas9 activo están activos y los dominios de endonucleasa del dCas9 están inactivos, para poner un polinucleótido donante en la proximidad de un DSB una secuencia diana de ácidos nucleicos. La Figura 12 ilustra la proteína Cas9 activa (Figura 12, 1211) y la proteína dCas9 (Figura 12, 1203), la composición NASC-PC1/NASC-PC2 (Figura 12, 1205; véase también la Figura 6F). Un ácido nucleico bicatenario (Figura 12, 1206/1207) que comprende una primera secuencia de unión a diana de ADN complementaria al elemento espaciador de composición Cas9-NASC-PC1/NASC-PC2 activo (Figura 12, 1210); y un polinucleótido donante (Figura 12, 1200/1201) que comprende una segunda secuencia de unión a diana de ADN complementaria a la composición dCas9-NASC-PC1/NASC-PC2 (Figura 12, 1204). La Figura 12 ilustra la proteína NASC-PC1/NASC-PC2/Cas9 activa/proteína dCas9 en el complejo y el enlace de hidrógeno de la primera secuencia de unión a la diana del ADN con la primera secuencia diana del ADN corriente arriba de una primera PAM de Cas9 (Figura 12, 1209) y la segunda secuencia de unión a la diana del ADN a la segunda secuencia diana corriente arriba de una segunda PAM de Cas9 (Figura 12, 1202) en el polinucleótido donante. La Figura 12, 1208, ilustra cortes de extremos romos de doble hebra hechos por Cas9 en la primera secuencia de unión a diana del ADN, lo que resulta en un segundo ácido nucleico de doble hebra (Figura 12, 1213/1212), y muestra el polinucleótido donante (Figura 12, 1200/1201) en las proximidades de los cortes romos de doble hebra. Tener el polinucleótido donante cerca de los cortes de doble cadena aumenta la probabilidad de integración de las secuencias de polinucleótido donante, o porciones de las mismas, en el ADN que comprende la primera diana de ácidos nucleicos.

En un aspecto adicional, la divulgación se refiere a un procedimiento que lleva un primer sitio diana de ácidos nucleicos, típicamente ADN, a la proximidad de un segundo sitio diana de ácidos nucleicos, típicamente ADN, en una célula. El procedimiento comprende poner en contacto una primera secuencia diana nucleica y una segunda secuencia diana nucleica con un complejo de nucleoproteínas que comprende la composición del polinucleótido

NASC en un complejo con una proteína de unión a ácidos nucleicos y una segunda proteína de unión a ácidos nucleicos, facilitando así la unión del complejo de nucleoproteína a primera secuencia diana de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos. La primera secuencia diana de ADN es complementaria a una primera secuencia de unión a ácidos nucleicos de la composición de polinucleótido NASC, en el que la primera proteína asociada es una proteína de unión a ácidos nucleicos catalíticamente inactiva (por ejemplo, una proteína dCpf1 o una proteína dCas9). La segunda secuencia diana de ADN es complementaria de una segunda secuencia de unión a ácidos nucleicos de la composición polinucleotídica de NASC, en el que la segunda proteína asociada es una proteína de unión a ácidos nucleicos catalíticamente inactiva (por ejemplo, una proteína dCpf1 o una proteína dCas9). El primer sitio objetivo de ácidos nucleicos se pone en proximidad de un segundo sitio objetivo de ácidos nucleicos debido a que la primera y la segunda proteínas de unión a ácidos nucleicos catalíticamente inactivas se complejan con la composición del polinucleótido NASC, es decir, son parte de la misma composición de ácidos nucleicos/proteína. En algunos aspectos, la primera secuencia diana de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos están en polinucleótidos separados (por ejemplo, cromosomas diferentes) o un polinucleótido único comprende la primera secuencia diana de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos (por ejemplo, diferentes secciones del mismo cromosoma).

La Figura 13 ilustra un ejemplo de una composición de polinucleótido NASC/una primera proteína dCas9/segunda proteína de unión a dCas9/una tercera composición de proteína dCas9 de unión a tres sitios dentro de un único polinucleótido de ADN. La composición de polinucleótido NASC también se ilustra en la Figura 6I. Este complejo de nucleoproteínas se puede usar en un procedimiento para llevar un primer sitio objetivo de ácidos nucleicos, típicamente ADN, a la proximidad de un segundo sitio objetivo de ácidos nucleicos, típicamente ADN, a la proximidad de un tercer sitio objetivo de ácidos nucleicos, típicamente ADN, en una celda. Este procedimiento también puede aplicarse, por ejemplo, a la detección de sitios diana de ácidos nucleicos en proximidad y modulación de la modulación transcripcional *in vitro* o *in vivo* de un gen adyacente a los tres sitios diana. Indicadores de los componentes ilustrados en la Figura 13 se presentan en la Tabla 8. La Figura 14 ilustra un ejemplo de una composición de polinucleótidos de NASC/una primera proteína dCas9/segunda proteína de unión a Cas9 activa/una tercera composición de proteína de Cas9 activa de unión a tres sitios de múltiples polinucleótidos de ADN. La composición de polinucleótido NASC también se ilustra en la Figura 6I. Este complejo de nucleoproteínas se puede usar, por ejemplo, en un procedimiento para acercar un polinucleótido donante a la proximidad de dos DSB en una diana de ácidos nucleicos, típicamente ADN, en una célula para facilitar la integración de HDR del polinucleótido donador o partes del polinucleótido donante en la región entre los dos sitios de escisión diana de ADN. Indicadores de los componentes ilustrados en la Figura 14 se presentan en la Tabla 8. La composición del polinucleótido NASC también se ilustra en la Figura 6I.

La Figura 15 ilustra un ejemplo de una composición de polinucleótidos de NASC/una primera proteína dCas9/segunda proteína de unión a Cas9 activa/una tercera composición de proteína de Cas9 activa de unión a tres sitios en tres polinucleótidos de ADN diferentes. La composición de polinucleótido NASC también se ilustra en la Figura 6I. Este complejo de nucleoproteínas se puede usar, por ejemplo, en un procedimiento para mejorar la frecuencia de ligamiento de dos polinucleótidos de ADN a los extremos 5' y 3' de un tercer polinucleótido de ADN. Indicadores de los componentes ilustrados en la Figura 15 se presentan en la Tabla 8. La composición del polinucleótido NASC también se ilustra en la Figura 6I.

Tabla 8

Indicadores y regiones correspondientes para Figura 13, Figura 14 y Figura 15					
Figura 13		Figura 14		Figura 15	
Indicador	Componente	Indicador	Componente	Indicador	Componente
1300	un extremo 3' de una primera hebra de un ADN	1400	un extremo 3' de una primera hebra de un primer ADN	1500	un extremo 3' de una primera hebra de un primer ADN
1301	un extremo 5' de una primera hebra de un ADN	1401	un extremo 3' de una segunda hebra de un primer ADN	1501	un extremo 3' de una segunda hebra de un primer ADN
1302	una primera proteína Cas9	1402	una primera proteína Cas9	1502	una primera proteína Cas9

(continuación)

Indicadores y regiones correspondientes para Figura 13, Figura 14 y Figura 15					
Figura 13		Figura 14		Figura 15	
Indicador	Componente	Indicador	Componente	Indicador	Componente
1303	una primera secuencia de unión a la diana del ADN	1403	una primera secuencia de unión a la diana del ADN	1503	una primera secuencia de unión a la diana del ADN
1304	Una primera secuencia de unión a la proteína de unión a ácidos nucleicos	1404	Una primera secuencia de unión a la proteína de unión a ácidos nucleicos	1504	Una primera secuencia de unión a la proteína de unión a ácidos nucleicos
1305	un primer PAM de Cas9	1405	un primer PAM de Cas9	1505	un primer PAM de Cas9
1306	una segunda proteína Cas9	1408	una segunda proteína Cas9	1508	una segunda proteína Cas9
1307	un segundo Cas9 PAM	1411	un segundo Cas9 PAM	1511	un segundo Cas9 PAM
1308	una segunda secuencia de unión a la diana del ADN	1409	una segunda secuencia de unión a la diana del ADN	1509	una segunda secuencia de unión a la diana del ADN
1309	una segunda secuencia de unión a la proteína de unión a ácidos nucleicos	1410	una segunda secuencia de unión a la diana de ADN secuencia de unión a la proteína	1510	una segunda secuencia de unión a la diana de ADN secuencia de unión a la proteína
1310	un extremo 3' de una primera hebra de un ADN	1406/1412	5'/3' extremos de una primera hebra de un segundo ADN	1507/1506	5'/3' extremos de un segundo ADN
1311	un extremo 5' de una primera hebra de un ADN	1413/1407	5'/3' extremos de una segunda hebra de un segundo ADN	1514/1513	5'/3' extremos de un tercer ADN
1312	una tercera proteína Cas9	1414	una tercera proteína Cas9	1515	una tercera proteína Cas9
1313	un tercer PAM Cas9	1417	un tercer PAM Cas9	1518	un tercer PAM Cas9
1314	Una tercera secuencia de unión a la proteína de unión a ácidos nucleicos	1416	Una tercera secuencia de unión a la proteína de unión a ácidos nucleicos	1517	Una tercera secuencia de unión a la proteína de unión a ácidos nucleicos
1315	una tercera secuencia de unión a la diana del ácido nucleico	1415	una tercera secuencia de unión a la diana del ácido nucleico	1516	una tercera secuencia de unión a la diana del ácido nucleico

(continuación)

Indicadores y regiones correspondientes para Figura 13, Figura 14 y Figura 15					
Figura 13		Figura 13		Figura 13	
Indicador	Indicador	Indicador	Indicador	Indicador	Indicador
	una primera rotura de doble cadena en el segundo ADN	1418	una rotura de doble cadena en el segundo ADN	1512	una rotura de doble cadena en el segundo ADN
	una segunda rotura de doble cadena en el segundo ADN	1419	una rotura de doble cadena en el tercer ADN	1519	una rotura de doble cadena en el tercer ADN

En otro aspecto más, la presente descripción también incluye procedimientos para modular la transcripción *in vitro* o *in vivo*, por ejemplo, la transcripción de un gen que comprende secuencias de elementos tampones. El procedimiento comprende poner en contacto al menos una primera secuencia diana nucleica y una segunda secuencia diana nucleica con una composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos (que comprende, por ejemplo, proteínas de Clase 2 catalíticamente inactivas CRISPR-Cas, tales como como dCas9 y/o dCpf1), facilitando así la unión de la composición de nucleoproteínas a la primera secuencia diana de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos. Al menos una de la primera secuencia diana de ADN y la segunda secuencia diana de ADN comprende las secuencias del elemento tampón. La primera secuencia de unión a diana del ADN de la composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos es complementaria a una primera secuencia diana de ácidos nucleicos. La segunda secuencia de unión a diana del ADN de la composición de polinucleótido NASC/primer proteína de unión a ácidos nucleicos/segunda composición de proteína de unión a ácidos nucleicos es complementaria a una segunda secuencia diana de ADN. Además, la primera y/o la segunda proteína pueden ser proteínas de fusión, por ejemplo, dCas9 fusionada a un dominio represor o activador, y/o dCpf1 fusionada a un dominio represor o activador. La unión de la composición de ácidos nucleicos/proteína a la primera secuencia diana de ADN y la segunda secuencia diana de ADN modula la transcripción del gen. En algunos aspectos, la primera secuencia diana de ADN y la segunda secuencia diana de ADN comprenden las secuencias del elemento tampón, y la primera secuencia diana de ADN comprende un promotor y la segunda secuencia diana de ADN comprende un sitio de inicio de transcripción.

La Figura 11 ilustra un procedimiento para modular la transcripción *in vitro* o *in vivo* usando composiciones de polinucleótidos NASC de la presente invención. En esta Figura, se forma una composición de polinucleótido NASC/una primera proteína dCas9/un segundo complejo de proteína dCsa9 (Figura 11, 1111, 1110, 1103) por la asociación de una composición de polinucleótido NASC (Figura 11, 1103) con una primera proteína dCas9 (Figura 11, 1100) y una segunda proteína dCsa9 (Figura 11, 1101). El complejo comprende una primera secuencia de unión a diana del ADN (Figura 11, 1102) complementaria a una primera secuencia de diana nucleica que es adyacente a una primera PAM de Cas9 (Figura 11, 1108) y una segunda secuencia de unión a diana de ADN (Figura 11, 1104) que es complementaria a una segunda secuencia diana nucleica que es adyacente a una segunda PAM de Cas9 (Figura 11, 1109) en un polinucleótido de ADN (Figura 11, 1105). La composición de polinucleótido NASC/primer proteína dCas9/segunda proteína dCsa9 se pone en contacto con el polinucleótido de ADN que comprende las secuencias diana de ADN, facilitando así la unión de la composición de nucleoproteínas a través de pares de bases con enlaces de hidrógeno (Figura 11, 1112, 1113) a la primera secuencia diana de ácidos nucleicos y la segunda secuencia diana de ácidos nucleicos. Al menos una de la primera secuencia diana de ADN y la segunda secuencia diana de ADN comprenden las secuencias del elemento tampón. La primera secuencia de unión a diana del ADN de la composición de polinucleótido NASC/primer proteína dCas9/segunda composición de proteína dCsa9 es complementaria a una primera secuencia diana de ácidos nucleicos. La segunda secuencia de unión a la diana del ADN de la composición del polinucleótido NASC/una primera proteína dCas9/una segunda composición de la proteína dCsa9 es complementaria a una segunda secuencia diana del ADN.

Las composiciones de polinucleótidos NASC de la presente invención se pueden usar para diseñar macromoléculas de ácidos nucleicos/proteína que se autoensamblan en arquitecturas complejas. Dichas macromoléculas tienen muchos usos en nanobiotecnología, incluidos, entre otros, la administración de fármacos, el diseño de nanomateriales de ácidos nucleicos/proteína y la formación de nanoestructuras como los nanotubos y las estructuras de jaula cerrada. El Ejemplo 13 ilustra el uso de composiciones de polinucleótidos de NASC de la presente invención

para la formación de composiciones de jaula cerrada de NASC (NASC-CC). Los NASC-CC se pueden usar para empaquetar moléculas pequeñas. El Ejemplo 14 describe procedimientos que se pueden usar para la caracterización de complejos de proteínas NASC-CC/dCas para verificar el ensamblaje adecuado y evaluar el tamaño y el volumen de los complejos de proteínas NASC-CC/dCas ensamblados. El NASC-CC descrito en el Ejemplo 13 y el Ejemplo 14 se ilustra en la Figura 6L. Dos composiciones de polinucleótidos NASC correspondientes a la composición de polinucleótidos NASC ilustrada en la Figura 6A (referido en el Ejemplo como NASC-PC1-triplex) puede conectarse utilizando secuencias de ácidos nucleicos de refuerzo de ADN de doble cadena. La secuencia de ácidos nucleicos de refuerzo de ADN de doble cadena puede comprender una primera secuencia diana de ADN y una segunda secuencia diana de ADN. Como se describe en el Ejemplo 13, el NASC-CC se autoensambla porque un primer complejo de nucleoproteínas de la proteína NASC-PC1-triplex/dCas9 comprende secuencias de unión al ADN que se unirán específicamente a la primera secuencia del ADN del objetivo de la secuencia de ácidos nucleicos. Una segunda proteína NASC-PC1-triplex/dCas9 que comprende secuencias de unión a diana de ADN se unirá específicamente a la segunda secuencia diana de ADN de la secuencia de ácidos nucleicos de refuerzo para formar una estructura de jaula cerrada. La Figura 6M ilustra la NASC-CC con seis proteínas Cas9 asociadas que forman una jaula de nucleoproteínas.

Una amplia variedad de moléculas son candidatas para su incorporación en composiciones de polinucleótidos NASC-CC para facilitar el suministro de las moléculas, que incluyen, entre otras, vacunas (por ejemplo, vacunas inactivadas, vacunas atenuadas, vacunas de subunidades de proteínas y vacunas de ácidos nucleicos); anticuerpos monoclonales; antibióticos medicamentos de molécula pequeña; terapias contra el cáncer; proteínas recombinantes, productos biológicos, y similares. Tales moléculas también se denominan en este documento "carga útil".

Las fusiones de proteínas de direccionamiento y proteínas de unión a ácidos nucleicos (por ejemplo, Cas9, Cpf1) se pueden usar para lograr el suministro selectivo de tipo de tejido, órgano o célula de las composiciones de polinucleótidos NASC-CC. Por ejemplo, los péptidos de fagos del paisaje específicos para tumores específicos pueden obtenerse por selección de afinidad y los péptidos purificados específicos para tumores específicos pueden fusionarse a una proteína Cas9. La proteína de fusión Cas9 se puede usar para ensamblar una composición de polinucleótido NASC-CC para obtener nanotransportadores dirigidos a tumores. La producción de péptidos fágicos específicos para tumores específicos ha sido descrita por Jayanna, P., et al., *Nanomedicine*. 5(1): 83 (2009).

Se pueden lograr modos alternativos de administración de NASC-CC a las células mediante el enlace, el empaquetamiento o la asociación de los ARN, ADN o proteínas de NASC-CC ("NASC-CC/Cas") con diversos ligandos o agentes químicos. Las técnicas de empaquetado incluyen el empaquetado NASC-CC/Cas en liposomas, micelas, dendrímeros, nanoesferas o nanocápsulas de autoensamblaje.

La unión covalente y no covalente de polietilenglicol (PEG; PEGilación) a moléculas y macroestructuras se ha empleado para el empaquetamiento de la carga útil para la entrega de diana a las células y se puede adaptar para la encapsulación de NASC-CC/Cas por un experto en la técnica. En vista de las enseñanzas de la presente especificación. Además, la proteína PEGilación es una forma ampliamente practicada de química de conjugación para el suministro de macromoléculas a tejidos, células y orgánulos. Las estructuras PEGiladas pueden modificarse aún más con la unión molecular de unidades estructurales que facilitan la captación celular (por ejemplo, una unidad estructural de folato). La selección de estas unidades estructurales se basa en las propiedades únicas de las células dirigidas a la administración dirigida de NASC-CC/Cas y la carga útil encapsulada (es decir, matriz extracelular, receptores o composición de anticuerpos). Estas unidades estructurales se pueden adjuntar al agente de empaquetado NASC-CC/Cas, NASC-CC, o al agente de empaquetado NASC-CC/Cas y NASC-CC. Las unidades estructurales que pueden usarse incluyen, pero no se limitan a, anticuerpos, ligandos, transferrinas, glicoproteínas, aptámeros, péptidos que penetran en las células, péptidos escindibles con metaloproteasas de matriz, integrinas, dominios de transducción de proteínas, epítopos, moléculas de adhesión celular y otros compuestos conocidos en la técnica (véase, por ejemplo, Steichen, S. et al., *European Journal of Pharmaceutical Sciences*. 48(3): 416-27 (2013); Dashpande, P., et al., *Nanomedicine*. 8(9): 1509-28 (2013)).

La liberación dirigida de agentes encapsulados NASC-CC/Cas puede facilitarse mediante la incorporación de distintas unidades estructurales químicas o motivos de secuencia en la composición NASC-CC/Cas o dentro del agente de empaquetado NASC-CC. La unión de composiciones poliméricas biodegradables (por ejemplo, una composición de PEG manipulada) a un agente de empaquetado NASC-CC/Cas o NASC-CC puede permitir la ruptura del agente de empaquetado NASC-CC/Cas o NASC-CC tras la captación celular. Se pueden emplear sitios sensibles diseñados (es decir, secuencias peptídicas sensibles proteolíticas, copolímeros sensibles al pH, enlaces sensibles al redox, etc.) o combinaciones de sitios sensibles diseñados para facilitar la liberación de agentes encapsulados NASC-CC. Se pueden utilizar enlaces lábiles entre el agente de empaquetado NASC-CC y NASC-CC, como los enlaces sensibles al pH, para alentar la disociación de NASC-CC y el agente de empaquetado NASC-CC en entornos de alto pH (por ejemplo, una vacuola endocítica). Los complejos NASC-CC/Cas pueden modificarse aún más con epítopos específicos de orgánulos (es decir, una señal de localización nuclear) para la entrega de la carga útil a orgánulos específicos.

Un experto en la técnica, en vista de las enseñanzas de la especificación, puede usar una variedad de composiciones de polinucleótidos NASC diferentes para formar una variedad de nanoestructuras.

Cualquiera de los componentes de las composiciones de nucleoproteínas que comprenden una composición de polinucleótido NASC de la presente invención o secuencias de ácidos nucleicos que codifican tales componentes, como se describió anteriormente, pueden incorporarse en un kit, incluyendo opcionalmente uno o más reactivos. En algunas realizaciones, un kit incluye un paquete con uno o más contenedores que contienen los elementos del kit, como una o más composiciones separadas u, opcionalmente, como una mezcla en la que la compatibilidad de los componentes lo permita. En algunas realizaciones, los kits también comprenden un tampón, un agente tampón, una sal, una solución acuosa estéril y/o conservantes. Los kits ilustrativos comprenden uno o más componentes de una composición de polinucleótido NASC y, opcionalmente, una o más proteínas relacionadas de Tipo II CRISPR-Cas9; y una o más secuencias de ácidos nucleicos que codifican uno o más componentes de una composición de polinucleótido NASC, y opcionalmente una o más secuencias de ácidos nucleicos que codifican una proteína CRISPR-Cas9 de Clase 2 Tipo II.

Además, los kits pueden comprender instrucciones adicionales para usar componentes de los complejos de nucleoproteínas que comprenden composiciones de polinucleótidos de NASC de la presente invención o secuencias de ácidos nucleicos que codifican tales componentes. Las instrucciones incluidas en los kits de la invención se pueden fijar al material de embalaje o se pueden incluir como un prospecto. Aunque las instrucciones suelen ser materiales escritos o impresos, no se limitan a los mismos. Esta invención contempla cualquier medio capaz de almacenar dichas instrucciones y comunicarlás a un usuario final. Dichos medios incluyen, pero no se limitan a, medios de almacenamiento electrónico (por ejemplo, discos magnéticos, cintas, cartuchos, chips), medios ópticos (por ejemplo, CD ROM), etiquetas de RF y similares. Las instrucciones también pueden incluir la dirección de un sitio de Internet que proporciona las instrucciones.

Otro aspecto de la descripción se refiere a procedimientos para fabricar o manufacturar una composición de polinucleótidos de NASC o una composición de ácidos nucleicos/proteína que comprende una composición de polinucleótidos de NASC de la presente invención. En un aspecto, los procedimientos de fabricación o manufactura comprenden componentes de polinucleótidos de síntesis química de una composición de polinucleótidos de NASC. En algunos aspectos, una composición de polinucleótido NASC comprende bases de ARN y puede generarse a partir de plantillas de ADN utilizando la transcripción *in vitro*.

En algunas realizaciones, los componentes de la composición polinucleotídica de NASC pueden modificarse mediante una unidad estructural (por ejemplo, una unidad estructural de ligando, una unidad estructural de unión a ligando, una etiqueta de afinidad, una unidad estructural de resistencia a exonucleasa). Los componentes polinucleotídicos pueden conectarse a, por ejemplo, la secuencia terminal 5' y/o la secuencia terminal 3' de un componente polinucleótido.

Una composición de ácidos nucleicos/proteína que comprende una composición de polinucleótido NASC puede comprender además una etiqueta detectable, que incluye una unidad estructural que puede proporcionar una señal detectable. Ejemplos de marcadores detectables incluyen, pero no se limitan a, una enzima, un radioisótopo, un miembro de un par de enlace específico, un fluoróforo (FAM), una proteína fluorescente (proteína verde fluorescente, proteína roja fluorescente, mCereza, tdTomate), y ADN o ARN aptámero junto con un fluoróforo adecuado (GFP mejorada (EGFP), "Espinaca"), un punto cuántico, un anticuerpo y similares. Un gran número y variedad de etiquetas detectables adecuadas son bien conocidas por los expertos en la técnica.

Una composición de ácidos nucleicos/proteína que comprende una composición de polinucleótido NASC o células manipuladas mediante el uso de una composición de ácidos nucleicos/proteína que comprende la composición de polinucleótido NASC, como se describe en el presente documento, puede usarse como una composición farmacéutica formulada, por ejemplo, con un excipiente farmacéuticamente aceptable. Excipientes ilustrativos incluyen vehículos, estabilizantes, diluyentes, agentes dispersantes, agentes de suspensión, agentes espesantes y similares. La composición farmacéutica puede facilitar la administración de una composición de ácidos nucleicos/proteína que comprende una composición polinucleotídica NASC diseñada por ingeniería genética a un organismo. Las composiciones farmacéuticas pueden administrarse en cantidades terapéuticamente eficaces mediante diversas formas y rutas que incluyen, por ejemplo, administración intravenosa, subcutánea, intramuscular, oral, en aerosol, parenteral, oftálmica y pulmonar.

Se pueden obtener numerosas ventajas utilizando las composiciones de polinucleótidos de NASC y los complejos de nucleoproteínas de la presente invención, que incluyen, entre otros, los siguientes:

- reducción en la unión fuera de la diana utilizando una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC que se dirige a la unión a múltiples secuencias de ácidos nucleicos diana utilizando un complejo de nucleoproteína único en relación con el uso de complejos de proteínas de unión a NATNA/ácido nucleico dirigidos de forma similar (por ejemplo, un complejo de proteína ARNc/Cas9);

- anclaje de un polinucleótido donante mediante el uso de una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC para llevar el polinucleótido donante a la proximidad de un corte en una cadena doble ácido nucleico;

- llevar dos polinucleótidos separados (por ejemplo, dos cromosomas diferentes) o dos regiones de un solo polinucleótido (por ejemplo, dos regiones de un cromosoma único) a la proximidad entre sí utilizando una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, Clase 2 proteínas CRISPR-Cas) y una composición de polinucleótidos NASC;

5 • la modulación transcripcional de un gen diana mediante la unión de una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC a múltiples secuencias tampónas operativamente unidas al gen objetivo;

10 • modulación transcripcional de un gen diana mediante la unión de una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC para traer dos polinucleótidos separados (por ejemplo, elemento tampón de acción trans) o dos regiones de un solo polinucleótido (por ejemplo, elemento tampón de acción cis) en la proximidad entre sí;

15 • selección simultánea de múltiples secuencias de ácidos nucleicos diana utilizando una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC, incluidas realizaciones en las que un polinucleótido donante también está unido al ácido nucleico/composición proteica;

- formar nanoestructuras biológicas que comprenden una composición de ácidos nucleicos/proteína que comprende proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y una composición de polinucleótido NASC, por ejemplo, para la formulación farmacéutica de moléculas pequeñas;

20 • construir arquitecturas a nanoescala con composiciones de ácidos nucleicos/proteína que comprenden proteínas de unión a ácidos nucleicos (por ejemplo, proteínas CRISPR-Cas de Clase 2) y composiciones de polinucleótidos NASC que tienen tamaños y formas predefinidos; y

25 • diseñar componentes de ácidos nucleicos/proteína, que comprenden composiciones de ácidos nucleicos/proteína que comprenden proteínas de unión a ácidos nucleicos (por ejemplo, proteínas de Clase 2 CRISPR-Cas) y composiciones de polinucleótidos NASC, que se autoensamblan en arquitecturas complejas predeterminadas.

Diversos ítems contemplados aquí incluyen, pero no están limitados a, uno o más de los siguientes. Los artículos están numerados para facilitar la referencia.

Los ítems de la presente divulgación incluyen, pero no se limitan a, los siguientes.

30 1. Un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio, que comprende: un primer ácido nucleico manipulado que comprende: un primer elemento 1 que comprende una primera secuencia de unión a la proteína de unión al ácido nucleico bicatenario que tiene un primer extremo y un segundo final: un segundo elemento 1 que comprende una secuencia 1 de ácidos nucleicos repetida, en el que la secuencia 1 de ácidos nucleicos repetida está próxima al primer extremo de la primera secuencia de unión a la proteína de unión a ácidos nucleicos bicatenario: un tercer elemento 1 que comprende una

35 secuencia 1 de ácidos nucleicos; -y un segundo ácido nucleico manipulado que comprende, un primer elemento 2 que comprende una segunda secuencia de unión a la proteína de unión a ácidos nucleicos bicatenario, que tiene un primer extremo y un segundo extremo - un segundo elemento 2 que comprende una secuencia 1C de ácidos nucleicos repetida, en el que la repetición la secuencia 1 de ácidos nucleicos está próxima al primer extremo de la primera secuencia de unión a la proteína de unión a ácidos nucleicos bicatenarios y un tercer elemento 2 que comprende una secuencia 2 de ácidos nucleicos; en el que la

40 secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 1C de ácidos nucleicos repetida a través de enlaces de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida.

45 2. El complejo del ítem 1, en el que el primer ácido nucleico manipulado genética comprende - el primer elemento 1 que comprende además - la primera secuencia de unión a la proteína de unión al ácido nucleico bicatenario, en el que el primer extremo es un extremo 5' y el segundo extremo es un extremo 3'; el segundo elemento 1 comprende además la secuencia 1 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' de la primera secuencia de unión a la proteína de unión a ácidos nucleicos de doble cadena, y el

50 tercer elemento 1 que comprende además la secuencia 1 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que se ubica el extremo 5' de la secuencia 1 de ácidos nucleicos está ubicado en 3' del extremo 3' de la primera secuencia de unión a la proteína de unión a ácidos nucleicos bicatenario; y - el segundo ácido nucleico manipulado comprende, el primer elemento 2 que comprende además la segunda secuencia de unión a la proteína de unión a ácidos nucleicos bicatenario, en el que el primer extremo es un

55 extremo 5' y el segundo extremo es un extremo 3' - el segundo elemento 2 que comprende además la secuencia 1C de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en la que el extremo 3' de la secuencia 1C de ácidos nucleicos repetida está situada en 5' del extremo 5' del extremo 5' del segundo doble secuencia de unión a la proteína de unión a ácidos nucleicos trenzado - y el tercer elemento 2 que

comprende además la secuencia 2 de ácidos nucleicos tiene un extremo 5' y un extremo 3', en el que el extremo 5' de la secuencia 2 de ácidos nucleicos está ubicado en 3' de la 3' final de la segunda secuencia de unión a la proteína de unión a ácidos nucleicos de doble cadena.

3. El complejo del ítem 1, en el que el primer ácido nucleico manipulado comprende, el primer elemento 1 que comprende además la primera secuencia de unión a la proteína de unión al ácido nucleico bicatenario, en el que el primer extremo es un extremo 5' y el segundo extremo es un extremo 3'; el segundo elemento 1 comprende además la secuencia 1 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicado en 5' de la 5' el final de la primera secuencia de unión a la proteína de unión a ácidos nucleicos bicatenarios y el tercer elemento 1 que comprende además la secuencia 1 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos está ubicado 5' del extremo 5' de la secuencia 1 de ácidos nucleicos repetida; y el segundo ácido nucleico manipulado comprende: el primer elemento 2 que comprende además la segunda secuencia de unión a la proteína de unión a ácidos nucleicos bicatenario, en el que el primer extremo es un extremo 5' y el segundo extremo es un extremo 3' - el segundo elemento 2 adicional que comprende la secuencia 1C de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1C de ácidos nucleicos repetida está ubicada en 5' del extremo 5' del extremo 5' de la segunda cadena doble secuencia de unión a la proteína de unión a ácidos nucleicos - y el tercer elemento 2 que comprende además la secuencia 2 de ácidos nucleicos tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 2 de ácidos nucleicos está ubicado en 5' del extremo 5' de la secuencia 1C de ácidos nucleicos repetida.

4. Un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio, que comprende: una primera secuencia de ácidos nucleicos manipulada por ingeniería que comprende, un primer elemento 1 que comprende una primera secuencia de unión a proteínas CRISPR de Clase 2 de unión a ácidos nucleicos, que tiene un primer extremo y un segundo extremo - un segundo elemento 1 que comprende una secuencia 1 de ácidos nucleicos repetida, en el que la secuencia 1 de ácidos nucleicos repetida es proximal al primer extremo de la primera secuencia de unión a la proteína CRISPR de Clase 2 de unión a ácido nucleico - y un tercer elemento 1 que comprende una secuencia 1 de ácidos nucleicos; - y una segunda secuencia de ácidos nucleicos manipulada que comprende, un primer elemento 2 que comprende una segunda secuencia de unión a proteína CRISPR de Clase 2 de unión a ácido nucleico, que tiene un primer extremo y un segundo extremo - un segundo elemento 2 que comprende una secuencia 1C de ácidos nucleicos repetida, en el que la repetición de la secuencia 2 de ácidos nucleicos está próxima al primer extremo de la segunda secuencia de unión a la proteína CRISPR de la Clase 2 de unión a ácidos nucleicos, y un tercer elemento 2 que comprende una secuencia 2 de ácidos nucleicos; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 1C de ácidos nucleicos repetida a través de enlaces de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 1C de ácidos nucleicos repetida.

5. El complejo del ítem 4, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 de unión a los ácidos nucleicos es una secuencia de unión a la proteína CRISPR de Tipo V de la Clase 2, en el que el primer extremo es un extremo 5' y el segundo extremo es la 3' extremo: la secuencia 1 de ácidos nucleicos repetida tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicada en el extremo 5' del primer ácido nucleico Clase 2 Tipo V secuencia de unión a la proteína CRISPR - y la secuencia 1 de ácidos nucleicos tiene un extremo 5' y un extremo 3', en el que el extremo 5' de la secuencia 1 de ácidos nucleicos está ubicado en 3' del extremo 3' de la primera clase de unión a ácidos nucleicos 2 secuencia de unión a la proteína CRISPR tipo V; - y la segunda secuencia de unión a la proteína CRISPR de clase 2 de unión a los ácidos nucleicos es una secuencia de unión a la proteína CRISPR de clase 2 del tipo V, en el que el primer extremo es un extremo 5' y la segunda es un extremo a 3' - la secuencia 2 de ácidos nucleicos repetida tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 2 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' del segundo ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Tipo V - Clase 2 - y la secuencia 2 de ácidos nucleicos tiene un extremo 5' y un extremo 3', en el que el extremo 5' de la secuencia 2 de ácidos nucleicos está ubicado en 3' del extremo 3' de la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos.

6. El complejo del ítem 5, en el que la secuencia 1C de ácidos nucleicos repetida comprende además una secuencia 1-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1a de ácidos nucleicos repetida, que tiene un 5' extremo y un extremo 3' - una secuencia 1-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1b de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y un enlazador secuencia de ácidos nucleicos del elemento 1-3, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 1-1 de ácidos nucleicos del elemento enlazador, la secuencia 1a de ácidos nucleicos repetida, el elemento enlazador secuencia 1-2 de ácidos nucleicos, la secuencia 1b de ácidos nucleicos repetida, y la secuencia 1-3 de ácidos nucleicos del elemento enlazador; - y la secuencia 2 de ácidos nucleicos repetida comprende además una secuencia 2-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1bC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - a

secuencia 2-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 2a de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 2-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 2-1 de ácidos nucleicos del elemento enlazador, la secuencia 1bC de ácidos nucleicos repetida, la secuencia 2-2 del ácido nucleico elemento enlazador, la secuencia 2a de ácidos nucleicos repetida, y la secuencia 2-3 de ácidos nucleicos del elemento enlazador; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida.

7. El complejo del ítem 6, que comprende además un tercer ácido nucleico manipulado que comprende, un primer elemento 3 que comprende un tercer ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Clase 2 Tipo V, en el que el primer extremo es un extremo 5' y el segundo el extremo es un extremo 3' y un segundo elemento 3 que comprende una secuencia 3 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 3 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' final de la tercera secuencia de unión a la proteína CRISPR del tipo V Clase 2 de unión a ácidos nucleicos, en el que la secuencia repetida de unión a ácidos nucleicos 3 comprende además una secuencia 3-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 2aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 3-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 3a de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3'- y una secuencia 3-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 3-1 de ácidos nucleicos del elemento enlazador, la secuencia 2aC de ácidos nucleicos repetida, la secuencia 3-2 de ácidos nucleicos de elemento enlazador, la secuencia 3a de ácidos nucleicos repetida y la secuencia 3-3 de ácidos nucleicos de elemento enlazador; - y un tercer elemento 3 que comprende una secuencia 3 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que el extremo 5' de la secuencia 3 de ácidos nucleicos está ubicado en el extremo 3' de la primera secuencia de unión a la proteína CRISPR Clase 2 Tipo V; - y un cuarto ácido nucleico manipulado que comprende, un primer elemento 4 que comprende un cuarto ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Tipo V Clase 2, en el que el primer extremo es un extremo 5' y el segundo extremo es un extremo 3' - un segundo elemento 4 que comprende una secuencia 4 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 3 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' del cuarto ácido nucleico de unión a la secuencia de unión a la proteína CRISPR Clase 2 Tipo V, en la que la secuencia 4 de unión a ácidos nucleicos repetida comprende además una secuencia 4-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 3aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3'- una secuencia 4-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3'- una secuencia 1aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3'- y una secuencia 4-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 4-1 de ácidos nucleicos del elemento enlazador, la secuencia 3aC de ácidos nucleicos repetida, la secuencia 4-2 de ácidos nucleicos de elemento enlazador, la secuencia 1aC de ácidos nucleicos repetida y la secuencia 4-3 de ácidos nucleicos elemento de enlazador; - y el tercer elemento 4 que comprende además la secuencia 4 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que el extremo 5' de la secuencia 4 de ácidos nucleicos está ubicado en 3' del extremo 3' del primer ácido nucleico de unión de la proteína CRISPR clase 2 del tipo V; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida, la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 4 de ácidos nucleicos repetida a enlaces de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida y la secuencia 1aC de ácidos nucleicos repetida, la secuencia 2 de ácidos nucleicos repetida está asociada con la secuencia 3 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 2a de ácidos nucleicos repetida y la secuencia 2aC de ácidos nucleicos repetida, y la secuencia 3 de ácidos nucleicos repetida está asociada con la secuencia 4 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 3a de ácidos nucleicos repetida y la secuencia 3aC de ácidos nucleicos repetida.

8. El complejo de uno cualquiera de los ítems 4 a 7, en el que la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida comprenden además un sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios y la unión al ácido nucleico de doble cadena el sitio de unión a proteínas 1 está formado por un enlace de par de bases a través de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida.

9. El complejo del ítem 8, en el que el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio de unión a la proteína Csy4.

10. El complejo de uno cualquiera de los ítems 4 a 9, en el que el primer ácido nucleico manipulado y el segundo ácido nucleico manipulado comprenden cada uno ARN, ADN, o una combinación de los mismos.

11. El complejo del ítem 7, en el que el primer ácido nucleico manipulado, el segundo ácido nucleico manipulado, el tercer ácido nucleico manipulado y el cuarto ácido nucleico manipulado comprenden cada uno ARN, ADN o una combinación de los mismos.
- 5 12. El complejo de uno cualquiera de los ítems 5 a 11, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos y la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos son cada una secuencia de unión a la proteína Cpfl.
13. El complejo del ítem 7 u 11, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a los ácidos nucleicos, la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a los ácidos nucleicos, la tercera secuencia de unión a ácidos nucleicos de la proteína CRISPR Clase 2 Tipo 5, y la cuarta secuencia de unión a la proteína CRISPR de Clase 2 Tipo V de unión a ácidos nucleicos son cada una secuencia de unión a la proteína Cpfl.
- 10 14. El complejo de uno cualquiera de los ítems 5, 6, 7, 8, 9 o 10, en el que (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 1 y la secuencia 2 de ácidos nucleicos comprende además secuencia 2 de ácidos nucleicos espaciadores, y (ii) la secuencia 1 de ácidos nucleicos espaciadores es complementaria a una secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos espaciadores es complementaria a una secuencia 2 de ácidos nucleicos diana.
- 15 15. El complejo del ítem 14, en el que la secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos diana son cada una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, una cadena bicatenaria ARN, un ADN bicatenario, un híbrido ARN/ADN monocatenario y un híbrido ARN/ADN bicatenario.
- 20 16. El complejo de uno cualquiera de los ítems 7, 11 o 13, donde (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 1, la secuencia 2 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 2, la secuencia 3 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 3, y la secuencia 4 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 4, y (ii) la secuencia de ácidos nucleicos espaciadores 1 es complementaria de una secuencia 1 de ácidos nucleicos diana, el la secuencia de ácidos nucleicos espaciadores 2 es complementaria a una secuencia 2 de ácidos nucleicos diana, la secuencia de ácidos nucleicos espaciadores 3 es complementaria a una secuencia 3 de ácidos nucleicos diana, y la secuencia de ácidos nucleicos espaciadores 4 es complementaria a una secuencia 4 de ácidos nucleicos diana.
- 25 17. El complejo del ítem 16, en el que la secuencia 1 de ácidos nucleicos diana, la secuencia 2 de ácidos nucleicos diana, el ácido nucleico diana 3 y el ácido nucleico diana 4 son cada uno una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, un ARN bicatenario, un ADN bicatenario, un híbrido ARN/ADN monocatenario y un híbrido ARN/ADN bicatenario.
- 30 18. Un complejo de las dos o más secuencias de ácidos nucleicos manipuladas que forman el andamio de uno cualquiera de los ítems 1 a 17, comprendiendo además el complejo una primera proteína CRISPR Clase 2 Tipo V se une a la primera secuencia de unión de la proteína CRISPR Clase 2 Tipo V de unión al primer ácido nucleico, y una segunda proteína CRISPR Clase 2 Tipo V se une a la segunda secuencia de unión de la proteína CRISPR Clase 2 Tipo V de unión al primer ácido nucleico, en el que la primera proteína CRISPR Clase 2 Tipo V y la segunda proteína CRISPR Clase 2 Tipo V son cada una seleccionada del grupo que consiste en una proteína Cpfl y una proteína Cpfl catalíticamente inactiva.
- 35 19. Un complejo de las dos o más secuencias de ácidos nucleicos manipuladas que forman el andamio de cualquiera de los ítems 7, 11, 13 o 16, y el complejo comprende además una primera proteína CRISPR Clase 2 Tipo V unida a la primera secuencia de unión de la proteína CRISPR Clase 2 Tipo V, una segunda proteína CRISPR de tipo V Clase 2 unida a la segunda secuencia de unión de la proteína CRISPR Clase 2 Tipo V de unión al segundo ácido nucleico, una tercera proteína CRISPR Clase 2 Tipo V unida a la tercera secuencia de unión de la proteína CRISPR Clase 2 Tipo V de unión al tercer ácido nucleico y una cuarta proteína CRISPR Clase 2 Tipo V unida a la tercera secuencia de unión de la proteína CRISPR Clase 2 Tipo V de unión al cuarto ácido nucleico, en el que la primera proteína CRISPR Clase 2 Tipo V, la segunda proteína CRISPR Clase 2 Tipo V, la tercera proteína CRISPR Clase 2 Tipo V y la cuarta proteína CRISPR Clase 2 Tipo V se seleccionan del grupo que consiste en una proteína Cpfl y una proteína Cpfl catalíticamente inactivas.
- 40 20. El complejo del ítem 4, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 de unión a los ácidos nucleicos es una secuencia de unión a la proteína CRISPR de Clase 2 del tipo II, en la que el primer extremo es un extremo 5' y el segundo extremo es un extremo 3': la secuencia 1 de ácidos nucleicos repetida tiene un extremo 5' y un extremo 3', donde el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicada en 5' del extremo 5' de la primera secuencia de unión de la proteína CRISPR Clase 2 Tipo V - y la secuencia 1 de ácidos nucleicos tiene un extremo 5' y un extremo 3', en el que el
- 45 50 55

extremo 3' de la secuencia 1 de ácidos nucleicos está ubicado en 5' del extremo 5' de la secuencia 1 de ácidos nucleicos repetida; - y la segunda secuencia de unión a la proteína CRISPR de clase 2 de unión a los ácidos nucleicos es una secuencia de unión a la proteína CRISPR de clase 2 tipo II, en el que el primer extremo es un extremo 5' y el segundo extremo es un extremo a 3' - la secuencia 1C de ácidos nucleicos repetida tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1C de ácidos nucleicos repetida está ubicado en 5' del extremo 5' del segundo ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Clase 2 Tipo II - y la secuencia 2 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en la que el extremo 3' de la secuencia 2 de ácidos nucleicos está situada en 5' del extremo 5' de la secuencia 1C de ácidos nucleicos repetida.

21. El complejo del ítem 20, en el que la secuencia 1C de ácidos nucleicos repetida comprende además una secuencia 1-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1a de ácidos nucleicos repetida, que tiene un 5' extremo y un extremo 3' - una secuencia 1-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1b de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y un enlazador secuencia de ácidos nucleicos del elemento 1-3, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 1-1 de ácidos nucleicos del elemento enlazador, la secuencia 1a de ácidos nucleicos repetida, el elemento enlazador secuencia 1-2 de ácidos nucleicos, la secuencia 1b de ácidos nucleicos repetida, y la secuencia 1-3 de ácidos nucleicos del elemento enlazador; - y la secuencia 2 de ácidos nucleicos repetida comprende además una secuencia 2-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - a secuencia 2-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1bC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 2-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 2-3 de ácidos nucleicos del elemento enlazador, la secuencia 1bC de ácidos nucleicos repetida, la secuencia 2-2 del ácido nucleico elemento enlazador, la repetición secuencia 1aC de ácidos nucleicos, y la secuencia 2-1 de ácidos nucleicos del elemento enlazador; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida y la secuencia 1aC de ácidos nucleicos repetida y a través del enlace de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida.

22. El complejo del ítem 21, en el que la secuencia 1a1 de ácidos nucleicos repetida comprende además una secuencia 1a1 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 1a1 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3' - y una secuencia 1a2 de ácidos nucleicos repetida, que tienen un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 1a1 de ácidos nucleicos repetida, la secuencia 1a1 de ácidos nucleicos de protuberancia, y la secuencia 1a2 de ácidos nucleicos repetida; - y la secuencia 1b de ácidos nucleicos repetida comprende además una secuencia 1b1 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 1b1 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3' - y una secuencia 1b2 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', dispuestas en el siguiente orden 3' a 5': la secuencia 1b1 de ácidos nucleicos repetida, la secuencia 1b1 de ácidos nucleicos de protuberancia y la secuencia 1b2 de ácidos nucleicos repetida; y la secuencia 1bC de ácidos nucleicos repetida comprende además una secuencia 1b2C de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 2b2 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3' - y una secuencia 1b1C de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', dispuestas en el siguiente orden 3'a 5': la secuencia 1b2C de ácido nucleico repetida, la secuencia 2b2 de ácidos nucleicos de protuberancia y la secuencia 1b1C de ácidos nucleicos repetida - y la secuencia 1aC de ácidos nucleicos repetida comprende además una secuencia 1a2C de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 2a2 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3' - y una secuencia 1a1C de ácidos nucleicos repetida, con un extremo 5' y un extremo 3', dispuestos en el siguiente orden 3' a 5': la secuencia 1a2C de ácidos nucleicos repetida, la secuencia 2a2 de ácidos nucleicos de protuberancia y la secuencia 1a1C de ácidos nucleicos repetida; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1a1 de ácidos nucleicos repetida y la secuencia 1a1C de ácidos nucleicos repetida, la secuencia 1a2 de ácidos nucleicos repetida y la secuencia 1a2C de ácidos nucleicos repetida, la secuencia 1b1 de ácidos nucleicos repetida y la secuencia 1b1C de ácidos nucleicos repetida, la secuencia 12b de ácidos nucleicos repetida y la secuencia 12bC de ácidos nucleicos repetida.

23. El complejo de uno cualquiera de los ítems 20, 21 o 22, en el que la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida comprenden un sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios y el núcleo nucleico bicatenario el sitio 1 de unión a la proteína de unión a ácido se forma mediante un enlace de par de bases a través de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida.

24. El complejo del ítem 23, en el que el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio de unión a la proteína Csy4.

25. El complejo de uno cualquiera de los ítems 20 a 24, en el que el primer ácido nucleico manipulado y el segundo ácido nucleico manipulado comprenden cada uno ARN, ADN, o una combinación de los mismos.

26. El complejo de uno cualquiera de los ítems 20 a 25, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a los ácidos nucleicos y la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácido nucleico son cada una secuencia de unión a la proteína Cas9.

27. El complejo de uno cualquiera de los ítems 20 a 26, en el que (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 1 y la secuencia 2 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 2, y (ii) la secuencia 1 de ácidos nucleicos espaciadores es complementaria a una secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos espaciadores es complementaria a una secuencia 2 de ácidos nucleicos diana.

28. El complejo del ítem 27, en el que la secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos diana son cada una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, un doble ARN de cadena múltiple, un ADN de doble cadena, un híbrido de ARN/ADN de una sola cadena y un híbrido de ARN/ADN de doble cadena.

29. Un complejo de las dos o más secuencias de ácidos nucleicos manipuladas que forman el andamio de cualquiera de los ítems 20 a 28, comprendiendo además el complejo una primera proteína CRISPR de Clase 2 Tipo II unida a la primera secuencia de unión a la proteína CRISPR Clase 2 Tipo II de unión al primer ácido nucleico, y una segunda proteína CRISPR de Clase 2 Tipo II unida a la segunda secuencia de unión a la proteína CRISPR Clase 2 Tipo II de unión al segundo ácido nucleico, en el que la primera proteína CRISPR de Clase 2 Tipo II y la segunda proteína CRISPR de Clase 2 Tipo II son cada una seleccionada del grupo que consiste en una proteína Cas9 y una proteína Cas9 catalíticamente inactiva.

30. Un complejo de tres o más secuencias de ácidos nucleicos manipuladas que forman un andamio, que comprende: un primer ácido nucleico manipulado que comprende, un primer elemento CRISPR 1 que comprende: una primera secuencia de unión a la proteína CRISPR de Tipo 2 Clase II que tiene un Extremo 5' y un extremo 3' - y una secuencia 1 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicado a 5' del extremo 5' de la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos, la secuencia 1 de ácidos nucleicos repetida que comprende además una secuencia 1-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1a de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 1-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1b de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 1-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestas en el siguiente orden 3' a 5': el elemento enlazador en la secuencia 1-1 de ácidos nucleicos, la secuencia 1a de ácidos nucleicos repetida, la secuencia 1-2 de ácidos nucleicos de elemento enlazador, la secuencia 1b de ácidos nucleicos repetida y la secuencia 1-3 de ácidos nucleicos elemento de enlazador; - y un segundo elemento CRISPR 1 que además comprende una secuencia 1 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que (i) el extremo 3' de la secuencia 1 de ácidos nucleicos está ubicado 5' del extremo 5' de la secuencia 1 de ácidos nucleicos repetida, y (ii) la secuencia 1 de ácidos nucleicos comprende una secuencia espaciadora de ácidos nucleicos 1; - un segundo ácido nucleico manipulado que comprende, - un primer elemento CRISPR 2 que comprende - una segunda secuencia de unión a la proteína CRISPR Tipo II de Clase 2 de unión a un ácido nucleico, que tiene un extremo 5' y un extremo 3' - y una secuencia 2 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 2 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' de la secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos, la secuencia 2 de ácidos nucleicos repetida que comprende además - una secuencia 2-3 de ácidos nucleicos de elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1bC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 2-4 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 2a de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 2-5 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestos en los siguientes orden 3' a 5': la secuencia 2-3 de ácidos nucleicos del elemento enlazador, la secuencia 1bC de ácidos nucleicos repetida, la secuencia 2-4 de ácidos nucleicos del elemento enlazador, la secuencia 2a de ácidos nucleicos repetida y la secuencia 2-5 de ácidos nucleicos del elemento enlazador; - un segundo elemento CRISPR 2 que comprende una secuencia 2 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que (i) el extremo 3' de la secuencia 1 de ácidos nucleicos está ubicado en 5' del extremo 5' de secuencia 2 de ácidos nucleicos repetida, y (ii) la secuencia 2 de ácidos nucleicos comprende una secuencia de ácidos nucleicos espaciadores 2; - y un tercer ácido nucleico manipulado que comprende, un primer elemento CRISPR 3 que comprende un tercer ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de tipo II, que tiene un extremo 5' y un extremo 3' - y una secuencia 3 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 3 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' del tercer ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Tipo II, la secuencia 3 de ácidos nucleicos repetida que además comprende una secuencia 3-1 de ácidos nucleicos del elemento enlazador, que tiene un

- extremo 5' y un extremo 3' - una secuencia 2aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 3-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1aC-1 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 3-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3',
5 dispuesto en el siguiente orden 3' a 5': la secuencia 3-1 de ácidos nucleicos del elemento enlazador, la secuencia 2aC de ácidos nucleicos repetida, la secuencia 3-2 de ácidos nucleicos del elemento enlazador, la secuencia 1aC-1 de ácidos nucleicos repetida, y la secuencia 3-3 de ácidos nucleicos del elemento enlazador; un segundo elemento CRISPR 3 que comprende una secuencia 3 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que (i) el extremo 3' de la secuencia 3 de ácidos nucleicos está ubicado en 5' del extremo 5' del núcleo nucleico de repetición secuencia ácida 3, y (ii) la secuencia 3 de ácidos nucleicos comprende una secuencia de ácidos nucleicos espaciadores 3; en el que la secuencia 1a de ácidos nucleicos repetida está asociada con la secuencia 1aC-1 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida y la secuencia 1aC-1 de ácidos nucleicos repetida, la secuencia 1b de ácidos nucleicos repetida está asociada con la secuencia 1bC de ácidos nucleicos a través del enlace de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida, y la secuencia 2a de ácidos nucleicos repetida está asociada con la secuencia 2aC de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 2a de ácidos nucleicos repetida y la una secuencia 2aC de ácidos nucleicos repetida.
31. El complejo del ítem 30, en el que la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida comprenden un sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios y se forma el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios por enlaces de par de bases a través de hidrógeno entre la secuencia 1 de ácidos nucleicos repetida y la secuencia 2 de ácidos nucleicos repetida.
32. El complejo del ítem 31, en el que el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio de unión a la proteína Csy4.
33. El complejo de uno cualquiera de los ítems 30, 31 o 32, en el que el primer ácido nucleico manipulado, el segundo ácido nucleico manipulado y el tercer ácido nucleico manipulado comprenden cada uno ARN, ADN o una combinación de los mismos.
34. El complejo de uno cualquiera de los ítems 30 a 33, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión al ácido nucleico, la segunda secuencia de unión a la proteína CRISPR de Clase 2 del Tipo II de unión al ácido nucleico y la tercera secuencia de unión a la proteína CRISPR de Clase 2 del Tipo II de unión al ácido nucleico es una secuencia de unión a la proteína Cas9.
35. El complejo de uno cualquiera de los ítems 30 a 34, en el que (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia 1 de ácidos nucleicos espaciadores, la secuencia 2 de ácidos nucleicos comprende además una secuencia 2 de ácidos nucleicos espaciadores, y la secuencia 3 de ácidos nucleicos comprende además una secuencia 3 de ácidos nucleicos espaciadores, y (ii) la secuencia de ácidos nucleicos espaciadores 1 es complementaria a una secuencia 1 de ácidos nucleicos diana, la secuencia de ácidos nucleicos espaciadores 2 es complementaria a una secuencia 2 de ácidos nucleicos diana, y la secuencia 3 de ácidos nucleicos espaciadores es complementaria a una secuencia 3 de ácidos nucleicos diana.
36. El complejo del ítem 35, en el que la secuencia 1 de ácidos nucleicos diana, la secuencia 2 de ácidos nucleicos diana y la secuencia 3 de ácidos nucleicos diana son cada una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, un ARN bicatenario, un ADN bicatenario, un híbrido ARN/ADN monocatenario y un híbrido ARN/ADN bicatenario.
37. Un complejo de las tres o más secuencias de ácidos nucleicos manipuladas que forman el andamio de uno cualquiera de los ítems 30 a 36, comprendiendo además el complejo una primera proteína CRISPR de Tipo II de Clase 2 unida a la primera secuencia de unión de ácidos nucleicos de unión a la proteína CRISPR de Clase 2 Tipo II, una segunda proteína CRISPR de tipo II clase 2 unida a la segunda secuencia de unión de ácidos nucleicos de unión a la proteína CRISPR de Clase 2 Tipo II, y una tercera proteína CRISPR de Clase 2 del Tipo II unida a la tercera secuencia de unión de ácidos nucleicos de unión a la proteína CRISPR de Clase 2 Tipo II, en la que la primera proteína CRISPR de Tipo II de Clase 2, la segunda proteína CRISPR de Tipo II de Clase 2 y la tercera proteína CRISPR de Tipo II de Clase 2 se seleccionan cada una del grupo que consiste en una proteína Cas9 y una proteína Cas9 catalíticamente inactiva.
38. Un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio, que comprende: un ácido nucleico concatenado 1 que tiene un extremo 5' y un extremo 3' que comprende un primer elemento 1 concatenado que comprende una primera secuencia de unión de ácidos nucleicos de unión a la proteína CRISPR de Clase 2 Tipo II, que tiene un extremo 5'y un extremo 3' - un segundo elemento concatenado 1 que comprende una secuencia A1 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3' en el que la primera secuencia de unión de ácidos nucleicos de unión a la proteína CRISPR de

Clase 2 Tipo II se ubica en 3' del extremo 3' de la secuencia A1 de ácidos nucleicos repetida - un primer elemento concatenado 2 que comprende una segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II que tiene un extremo 5' y un extremo 3' extremo: un segundo elemento concatenado 2 que comprende una secuencia A2 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', en el que la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II se une a 3' del extremo 3' la secuencia de ácidos nucleicos de la repetida A2, en el que el extremo 5' del primer elemento concatenado 1 está unido covalentemente al extremo 3' del primer elemento concatenado 2 para formar el ácido nucleico concatenado 1; - un tercer elemento concatenado 1 que tiene un extremo 5' y un extremo 3' que comprende una secuencia A1C de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3' y una secuencia 1 de ácidos nucleicos que tiene un extremo 5' y un extremo 3', en el que la secuencia 1 de ácidos nucleicos está ubicada en 5' del extremo 5' de la secuencia A1C de ácidos nucleicos repetida, en el que (i) la secuencia A1C de ácidos nucleicos repetida es complementaria a la secuencia A1 de ácidos nucleicos repetida, (ii) la secuencia A1C de ácidos nucleicos repetida está asociada con la secuencia A1 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia A1C de ácidos nucleicos repetida y la secuencia A1 de ácidos nucleicos repetida; - y un tercer elemento concatenado 2 que tiene un extremo 5' y un extremo 3' que comprende una secuencia A2C de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3' y una secuencia 2 de ácidos nucleicos que tiene un extremo 5' y un extremo 3', en el que la secuencia 2 de ácidos nucleicos está localizada en 5' del extremo 5' de la secuencia A2C de ácidos nucleicos repetida, en el que (i) la secuencia A2C de ácidos nucleicos repetida es complementaria a la secuencia A2 de ácidos nucleicos repetida, (ii) la secuencia A2C de ácidos nucleicos repetida está asociada a la secuencia A2 de ácidos nucleicos repetida, y (iii) la secuencia A2C de ácidos nucleicos repetida está asociada a la secuencia A2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia A2C de ácidos nucleicos repetida y la secuencia A2 de ácidos nucleicos repetida.

39. El complejo del ítem 38, en el que la secuencia A1 de ácidos nucleicos repetida comprende además una secuencia A1-1 de ácidos nucleicos del elemento enlazador que tiene un extremo 5' y un extremo 3', el extremo 3' de la secuencia de ácidos nucleicos del elemento enlazador A1-1 ubicado en 5' del extremo 5' de la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos, comprendiendo la secuencia A1-1 de ácidos nucleicos del elemento enlazador, una secuencia de ácidos nucleicos A1-1 repetida que tiene un extremo 5' y un extremo 3', y una secuencia A1-1 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3', el extremo 3' de la secuencia A1-1 de ácidos nucleicos de protuberancia adyacente al extremo 5' de la secuencia A1-1 de ácidos nucleicos repetida - y una secuencia A1-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', que comprende una secuencia A1-2 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', el extremo 3' de la secuencia A1-2 de ácidos nucleicos del elemento enlazador está ubicado en 5' del extremo 5' del ácido nucleico del elemento enlazador A1-1; - la secuencia A2 de ácidos nucleicos repetida comprende además una secuencia A2-1 de ácidos nucleicos del elemento enlazador que tiene un extremo 5' y un extremo 3', el extremo 3' de la secuencia A2-1 de ácidos nucleicos del elemento enlazador ubicada en 5' de la 5' el final de la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos, comprendiendo la secuencia A2-1 de ácidos nucleicos del elemento enlazador, una secuencia A2-1 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', y una secuencia A2-1 de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3', el extremo 3' de la secuencia A2-1 de ácidos nucleicos de protuberancia adyacente al extremo 5' de la secuencia A2-1 de ácidos nucleicos repetida - y una secuencia A2-2 de ácidos nucleicos de elemento enlazador, que tiene un extremo 5' y un extremo 3', que comprende una secuencia A2-2 de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', el extremo 3' de la una secuencia A2-2 de ácidos nucleicos de elemento enlazador ubicado en 5' del extremo 5' del ácido nucleico del elemento enlazador A2-1; - el tercer elemento concatenado 1 en el que la secuencia A1C de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de elemento enlazador A1-1C que comprende una secuencia A1-1C de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', el extremo 5' del la secuencia A1-1 C de ácidos nucleicos repetida ubicada en 3' del extremo 3' de la secuencia 1 de ácidos nucleicos, y una secuencia A1-1 C de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3', el extremo 5' de la secuencia A1-1C de ácidos nucleicos del protuberancia localizada en 3' del extremo 3' de la secuencia A1-1C de ácidos nucleicos repetida, en el que (i) la secuencia A1-1C de ácidos nucleicos repetida es complementaria de la secuencia A1-1 de ácidos nucleicos repetida, y (ii) la secuencia A1-1C de ácidos nucleicos repetida está asociada con la secuencia A1-1 de ácidos nucleicos repetida a través de enlaces de hidrógeno entre la secuencia A1-1C de ácidos nucleicos repetida y la secuencia A1-1 de ácidos nucleicos repetida - y una secuencia A1-2C de ácidos nucleicos de elemento enlazador que tiene un extremo 5' y un extremo de 3', que comprende una secuencia A1-2C de ácidos nucleicos repetida, teniendo un extremo 5' y un extremo 3', el extremo 5' de la secuencia A1-2C de ácidos nucleicos del elemento enlazador se ubica en 3' del extremo 3' de la secuencia A1-1C de ácidos nucleicos del elemento enlazador, en el que (i) la secuencia A1-2C de ácidos nucleicos repetida es complementaria a la secuencia A1-2 de ácidos nucleicos repetida, y (ii) la secuencia A1-2C de ácidos nucleicos repetida está asociada con la secuencia A1-2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia A1-2C de ácidos nucleicos repetida y la secuencia A1-2 de ácidos nucleicos repetida; - y la secuencia A2C de ácidos nucleicos repetida comprende además una secuencia de ácidos nucleicos de elemento enlazador A2-1C que comprende una secuencia A2-1C de ácidos nucleicos repetida

- que tiene un extremo 5' y un extremo 3', el extremo 5' de la secuencia A2-1C de ácidos nucleicos repetida se ubica en 3' del extremo 3' de la secuencia 2 de ácidos nucleicos, y una secuencia A2-1C de ácidos nucleicos de protuberancia, que tiene un extremo 5' y un extremo 3', el extremo 5' de la secuencia A2-1C de ácidos nucleicos de protuberancia se ubica en 3' del extremo 3' de la secuencia A2-1C de ácidos nucleicos repetida, en el que (i) la secuencia A2-1C de ácidos nucleicos repetida es complementaria a la secuencia A2-1 de ácidos nucleicos repetida, y (ii) la secuencia A2-1C de ácidos nucleicos repetida está asociada con la secuencia A2-1 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia A2-1C de ácidos nucleicos repetida y la secuencia A-12 de ácidos nucleicos repetida - y una secuencia de ácidos nucleicos del elemento enlazador A2-2C con un Extremo 5' y un extremo 3', que comprende una secuencia A2-2C de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3', el extremo 5' del enlazador e la secuencia A2-2C de ácidos nucleicos del elemento enlazador se ubica en 3' del extremo 3' de la secuencia A2-1C de ácidos nucleicos del elemento enlazador, en el que (i) la secuencia A2-2C de ácidos nucleicos repetida es complementaria de la secuencia A2-2 de ácidos nucleicos repetida, y (ii) la secuencia A2-2C de ácidos nucleicos repetida está asociada con la secuencia A2-2 de ácidos nucleicos repetida a través de enlaces de hidrógeno entre la secuencia A2-2C de ácidos nucleicos repetida y la secuencia A2-2 de ácidos nucleicos repetida.
40. El complejo del ítem 38 o 39, en el que la secuencia A1 de ácidos nucleicos repetida y la secuencia A1C de ácidos nucleicos repetida comprenden además un sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios y el sitio de unión a la proteína de unión a ácidos nucleicos bicatenarios 1 está formado por enlaces de par de bases a través de hidrógeno entre la secuencia A1 de ácidos nucleicos repetida y la secuencia A1C de ácidos nucleicos repetida.
41. El complejo de uno cualquiera de los ítems 38 a 40, en el que la secuencia A2 de ácidos nucleicos repetida y la secuencia A2C de ácidos nucleicos repetida comprenden además un sitio 2 de unión a la proteína de unión a ácidos nucleicos bicatenarios y la unión al ácido nucleico de doble cadena el sitio de unión a la proteína 2 está formado por un enlace de par de bases a través de hidrógeno entre la secuencia A2 de ácidos nucleicos repetida y la secuencia A2C de ácidos nucleicos repetida.
42. El complejo del ítem 40 o 41, en el que el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio de unión a la proteína Csy4.
43. El complejo de uno cualquiera de 38 a 42, en el que el ácido nucleico concatenado manipulado 1, el tercer elemento concatenado 1 y el tercer elemento concatenado 2 comprenden cada uno ARN, ADN o una combinación de los mismos.
44. El complejo de uno cualquiera de 38 a 43, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácido nucleico y la segunda secuencia de unión a la proteína CRISPR de Clase II Tipo II de unión a ácido nucleico son, cada una, una secuencia de unión a la proteína Cas9.
45. El complejo de uno cualquiera de 38 a 44, en el que (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 1 y la secuencia 2 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 2, y (ii) la secuencia de ácidos nucleicos espaciadores 1 es complementaria a una secuencia 1 de ácidos nucleicos diana y la secuencia de ácidos nucleicos espaciadores 2 es complementaria a una secuencia 2 de ácidos nucleicos diana.
46. El complejo del ítem 45, en el que la secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos diana son cada una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, un doble ARN de cadena múltiple, un ADN de doble cadena, un híbrido de ARN/ADN de una sola cadena y un híbrido de ARN/ADN de doble cadena.
47. Un complejo de las dos o más secuencias de ácidos nucleicos manipuladas que forman el andamio de uno cualquiera de los ítems 38 a 46, comprendiendo el complejo además una primera proteína CRISPR de Clase 2 Tipo II unida a la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II y una segunda proteína CRISPR de Tipo II Clase 2 se une a la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II, en el que la primera proteína CRISPR de Clase 2 Tipo II y la segunda proteína CRISPR de Clase 2 Tipo II se seleccionan cada una de el grupo que consiste en una proteína Cas9 y una proteína Cas9 catalíticamente inactiva.
48. Un complejo de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio, que comprende: un ácido nucleico de nexo dividido concatenado manipulado 1, que tiene un extremo 5' y un extremo 3', que comprende un primer elemento 1 de nexo dividido, que tiene un extremo 5' y un extremo 3', que comprende una primera secuencia de unión a proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos y una secuencia 1-1 de ácidos nucleicos del elemento de tallo nexo dividido que tiene un extremo 5' y un extremo 3' en el que la primera secuencia de unión a proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos se ubica en 3' al extremo 3' de la secuencia 1-1 de ácidos nucleicos del elemento de tallo nexo dividido - y un primer elemento 2 del nexo dividido que tiene un extremo 5' y un extremo 3', que

comprende una segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos y una secuencia de ácidos nucleicos del elemento de tallo de nexo dividido 2-1 que tiene un extremo 5' y un extremo 3' en el que la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos se ubica 3' al extremo 3' de la secuencia 2-1 de ácidos nucleicos del elemento de tallo del nexo dividido, y un polinucleótido auxiliar 1-1 que tiene un extremo 5' y un extremo 3', en el que el extremo 5' del primer elemento de nexo dividido 1 se une covalentemente al extremo 3' del polinucleótido auxiliar 1-1, y el extremo 5' del polinucleótido auxiliar 1-1 está unido covalentemente al extremo 3' del primer elemento de nexo dividido 2 para formar el elemento de nexo dividido concatenado; - un segundo elemento de nexo dividido 1, que tiene un extremo 5' y un extremo 3', que comprende una secuencia 1 de ácidos nucleicos que tiene un extremo 5' y un extremo 3' y una primera secuencia 1-1 de ácidos nucleicos del elemento de tallo que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos está unido covalentemente al extremo 5' de la primera secuencia 1-1 de ácidos nucleicos del elemento de tallo - una secuencia de ácidos nucleicos del elemento de bucle 1 que tiene un extremo 5' extremo y un extremo 3', en el que el extremo 3' de la primera secuencia 1-1 de ácidos nucleicos del elemento de tallo se une covalentemente al extremo 5' de la secuencia 1 de ácidos nucleicos del elemento del bucle 1 - una primera secuencia 1-2 de ácidos nucleicos del elemento de tallo con un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos del elemento de bucle está unido covalentemente al extremo 5' de la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo - una secuencia 1 de ácidos nucleicos conectiva que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo está unido covalentemente al extremo 5' de la secuencia 1 del ácido nucleico conectiva - y una secuencia 1-2 de ácidos nucleicos del elemento de tallo nexo dividido, en el que el extremo 3' de la secuencia 1 de ácidos nucleicos conectiva está unido covalentemente al extremo 5' de la secuencia de ácidos nucleicos del elemento de tallo nexo dividido 1-2, en el que (i) la primera secuencia 1-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo forman un primer elemento de tallo 1 mediante un enlace de par de bases a través de hidrógeno entre la primera secuencia de ácidos nucleicos del elemento de tallo 1-1 y la primera secuencia 1-2 de ácidos nucleicos del elemento de tallo, y (ii) la secuencia 1-1 de ácidos nucleicos del elemento de tallo nexo dividido y la secuencia de ácidos nucleicos del elemento de tallo nexo dividido 1-2 forman un elemento de tallo del nexo de la división 1 por la base de hidrógeno - enlace par entre la secuencia 1-1 de ácidos nucleicos del elemento de tallo nexo dividido y la secuencia 1-2 de ácidos nucleicos del elemento de tallo nexo dividido; - y un segundo elemento 2 del nexo dividido, que tiene un extremo 5' y un extremo 3' extremo, que comprende una secuencia 2 de ácidos nucleicos que tiene un extremo 5' y un extremo 3' y una secuencia 2-1 de ácidos nucleicos del primer elemento de tallo que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 2 de ácidos nucleicos está covalentemente unido al extremo 5' de la primera secuencia 2-1 de ácidos nucleicos del elemento de tallo - una secuencia de ácidos nucleicos del elemento de bucle 2 que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la primera secuencia 2-1 de ácidos nucleicos del elemento de tallo se une covalentemente al extremo 5' de la secuencia de ácidos nucleicos del elemento del bucle 2 - una primera secuencia 2-2 de ácidos nucleicos del elemento de tallo que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 2 de ácidos nucleicos del elemento de bucle está unido covalentemente al extremo 5' de la secuencia 2-2 de ácidos nucleicos del primer elemento de tallo - una secuencia 2 de ácidos nucleicos conectiva que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo está unida covalentemente al extremo 5' de la secuencia de ácidos nucleicos conectivo 2 - y una secuencia a 2-2 de ácidos nucleicos del elemento de tallo de nexo dividido, en el que el extremo 3' de la secuencia 1 de ácidos nucleicos conectiva se une covalentemente al extremo 5' de la secuencia 2-2 de ácidos nucleicos del elemento de tallo de nexo dividido, en el que (i) la primera secuencia 2-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo forman un primer elemento de tallo 2 mediante el enlace de par de bases a través de hidrógeno entre la primera secuencia 2-1 de ácidos nucleicos del elemento de tallo y la primera secuencia 2-2 de ácidos nucleicos del elemento de tallo, y (ii) la secuencia 2-1 de ácidos nucleicos del elemento de tallo de nexo dividido 2-1 y la secuencia 2-2 de ácidos nucleicos del elemento de tallo de nexo dividido forman un elemento de tallo 2 de nexo dividido por el enlace de un par de bases a través de hidrógeno entre la secuencia 2-1 de ácidos nucleicos del elemento de tallo de nexo dividido y la secuencia 2-2 de ácidos nucleicos del elemento de tallo de nexo dividido.

49. El complejo del ítem 48, en el que el primer elemento de tallo 1 comprende además en una dirección 5' a 3' una secuencia 1-1 de ácidos nucleicos del elemento de tallo inferior, una secuencia 1-1 de ácidos nucleicos del elemento de protuberancia, una secuencia 1-1 de ácidos nucleicos del elemento de tallo superior, la secuencia de ácidos nucleicos del elemento de bucle 1, una secuencia 1-2 de ácidos nucleicos del elemento de tallo superior, una secuencia 1-2 de ácidos nucleicos del elemento de protuberancia, y una secuencia 1-2 de ácidos nucleicos del elemento de tallo inferior, donde la secuencia 1-1 de ácidos nucleicos del elemento de tallo superior y la secuencia 1-2 de ácidos nucleicos del elemento de tallo superior forman un elemento de tallo superior 1 mediante el enlace de par de bases a través de hidrógeno entre la secuencia 1-1 de nucleótidos del elemento de tallo superior y la secuencia 1-2 de nucleótidos del elemento de tallo superior, y la secuencia 1-1 de ácidos nucleicos del elemento de tallo inferior y la secuencia 1-2 de ácidos nucleicos del elemento de tallo inferior forman un elemento de tallo inferior 1 mediante la unión de par de bases a través de

hidrógeno entre la secuencia 1-1 de ácidos nucleicos del elemento de tallo inferior y la secuencia 1-2 de nucleótidos del elemento de tallo inferior.

50. El complejo del artículo 48 o 49, en el que el primer elemento de tallo 2 comprende además en una dirección 5' a 3' una secuencia 2-1 de ácidos nucleicos del elemento de tallo inferior, una secuencia 2-1 de ácidos nucleicos del elemento de protuberancia, y una secuencia 2-1 de ácidos nucleicos del elemento de tallo superior, la secuencia 2 de ácidos nucleicos del elemento de bucle, una secuencia 2-2 de ácidos nucleicos del elemento de tallo superior, una secuencia 2-2 de ácidos nucleicos del elemento de protuberancia 2-2 y una secuencia 2-2 de ácidos nucleicos del elemento de tallo inferior, en el que la secuencia 2-1 de ácidos nucleicos del elemento de tallo superior y la secuencia 2-2 de ácidos nucleicos del elemento de tallo superior forman un elemento de tallo superior 2 por el enlace de par de bases a través de hidrógeno entre la secuencia 2-1 de nucleótidos del elemento de tallo superior y la secuencia 2-2 de ácidos nucleicos del elemento de tallo superior, y la secuencia 2-1 de ácidos nucleicos del elemento de tallo inferior y la secuencia 2-2 de ácidos nucleicos del elemento de tallo inferior forman un elemento de tallo inferior 2 por enlace de par de bases a través de hidrógeno entre la secuencia 2-1 de ácidos nucleicos del elemento de tallo inferior y la secuencia 2-2 de nucleótidos del elemento de tallo inferior.

51. El complejo de uno cualquiera de los ítems 48 a 50, en el que el segundo elemento 1 del nexo dividido comprende además un polinucleótido auxiliar 1-2, que tiene un extremo 5' y un extremo 3', en el que el extremo 5' del polinucleótido auxiliar 1-2 es 3' del extremo 3' de la secuencia 1-2 de ácidos nucleicos del elemento de tallo nexo dividido, en el que el polinucleótido auxiliar 1-2 está asociado con el polinucleótido auxiliar 1-1 a través del enlace de par de bases a través de hidrógeno.

52. El complejo del ítem 51, en el que el polinucleótido auxiliar 1-1 y el polinucleótido auxiliar 1-2 comprenden además un sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios y el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es formado por enlaces de par de bases a través de hidrógeno entre el polinucleótido auxiliar 1-1 y el polinucleótido auxiliar 1-2.

53. El complejo del ítem 52, en el que el sitio 1 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio 1 de unión a la proteína Csy4.

54. El complejo de uno cualquiera de los ítems 48 a 53, en el que el segundo elemento de nexo dividido 2 comprende además un polinucleótido auxiliar 2-2, que tiene un extremo 5' y un extremo 3', en el que el extremo 5' del polinucleótido auxiliar 2-2 es 3' del extremo 3' de la secuencia 2-2 de ácidos nucleicos del elemento de tallo nexo dividido - y el primer elemento de nexo dividido 2 comprende además un polinucleótido auxiliar 2-1 que tiene un extremo 5' y un extremo 3' extremo, en el que (i) el extremo 5' del primer elemento de nexo dividido 2 se une covalentemente al extremo 3' del polinucleótido auxiliar 2-1, y (ii) el polinucleótido auxiliar 2-2 está asociado con el polinucleótido auxiliar 2-1 a través de enlaces de bases a través de hidrógeno.

55. El complejo del ítem 54, en el que el polinucleótido auxiliar 2-1 y el polinucleótido auxiliar 2-2 comprenden además un sitio 2 de unión a la proteína de unión a ácidos nucleicos bicatenarios y el sitio 2 de unión a la proteína de unión a ácidos nucleicos bicatenarios es formado por enlaces de par de bases a través de hidrógeno entre el polinucleótido auxiliar 2-1 y el polinucleótido auxiliar 2-2.

56. El complejo del ítem 55, en el que el sitio 2 de unión a la proteína de unión a ácidos nucleicos bicatenarios es un sitio 2 de unión a la proteína Csy4.

57. El complejo de uno cualquiera de 48 a 56, en el que el ácido nucleico 1 de nexo dividido concatenado manipulado, el tercer elemento 1 concatenado y el tercer elemento 2 concatenado comprenden cada uno ARN, ADN o una combinación de los mismos.

58. El complejo de uno cualquiera de 48 a 57, en el que la primera secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácido nucleico y la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos son cada una secuencia de unión a la proteína Cas9.

59. El complejo de uno cualquiera de 48 a 58, en el que (i) la secuencia 1 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 1 y la secuencia 2 de ácidos nucleicos comprende además una secuencia de ácidos nucleicos espaciadores 2, y (ii) la secuencia de ácidos nucleicos espaciadores 1 es complementaria a una secuencia 1 de ácidos nucleicos diana y la secuencia de ácidos nucleicos espaciadores 2 es complementaria a una secuencia 2 de ácidos nucleicos diana.

60. El complejo del ítem 59, en el que la secuencia 1 de ácidos nucleicos diana y la secuencia 2 de ácidos nucleicos diana son cada una secuencia de ácidos nucleicos seleccionada del grupo que consiste en un ARN monocatenario, un ADN monocatenario, un doble ARN de cadena múltiple, un ADN de doble cadena, un híbrido de ARN/ADN de una sola cadena y un híbrido de ARN/ADN de doble cadena.

61. Un complejo de las dos o más secuencias de ácidos nucleicos manipuladas que forman el andamio de uno cualquiera de los ítems 48 a 60, comprendiendo además el complejo una primera proteína CRISPR de Clase 2 Tipo II unida a la primera secuencia de unión a la proteína CRISPR de Tipo II Clase 2 de unión a ácidos nucleicos y una segunda proteína CRISPR de Tipo II Clase 2 se unen a la segunda secuencia de unión a la proteína CRISPR de Tipo II Clase 2 de unión a ácidos nucleicos, en el que se seleccionan la primera proteína CRISPR de Clase 2 Tipo II y la segunda proteína CRISPR de Clase 2 Tipo II del grupo que consiste en una proteína Cas9 y una proteína Cas9 catalíticamente inactiva.

62. Un andamio de ácidos nucleicos manipulados, que comprende; un primer ácido nucleico manipulado que comprende - un primer elemento 1 que comprende una primera secuencia de unión a la proteína CRISPR de Tipo II Clase 2 de unión a ácidos nucleicos, que tiene un extremo 5' y un extremo 3' - y un segundo elemento 1 que comprende una secuencia 1 de ácidos nucleicos repetida que tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1 de ácidos nucleicos repetida está ubicado en 5' del extremo 5' de la primera secuencia de unión a la proteína CRISPR de Tipo II Clase 2 de unión a ácidos nucleicos, comprendiendo la secuencia 1 de ácidos nucleicos repetida además una secuencia 1-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1a de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia de ácidos nucleicos del elemento enlazador 1 -2, que tiene un extremo 5' y un extremo 3' - una secuencia 1b de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 1-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuesto en el siguiente orden 3' a 5': la secuencia 1-1 de ácidos nucleicos del elemento enlazador, la secuencia 1a de ácidos nucleicos repetida, la secuencia 1-2 de ácidos nucleicos del elemento enlazador, la secuencia 1b de ácidos nucleicos repetida y la secuencia 1-3 de ácidos nucleicos del elemento enlazador; en el que ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida se asocia con ninguna secuencia de ácidos nucleicos dentro de la secuencia 1 de ácidos nucleicos repetida para formar un elemento de tallo a través de enlaces de hidrógeno capaces de unirse a una proteína Clase 2 Tipo II CRISPR-Cas.

63. El andamio de ácidos nucleicos manipulado según el ítem 62, que además comprende un tercer elemento 1 que comprende una secuencia 1 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que (i) el extremo 3' de la secuencia 1 de ácidos nucleicos está covalentemente unido al extremo 5' de la secuencia 1 de ácidos nucleicos repetida, y (ii) la secuencia 1 de ácidos nucleicos comprende una secuencia espaciadora de ácidos nucleicos 1.

64. El andamio de ácidos nucleicos manipulado según el ítem 62 o 63, que además comprende: un segundo ácido nucleico manipulado que comprende - un primer elemento 2 que comprende un segundo ácido nucleico de unión a la secuencia de unión a la proteína CRISPR de Tipo II Clase 2, que tiene un extremo 5' y un extremo 3'- y un segundo elemento 2 que comprende la secuencia 1C de ácidos nucleicos repetida tiene un extremo 5' y un extremo 3', en el que el extremo 3' de la secuencia 1C de ácidos nucleicos repetida está ubicado en 5' del extremo 5' de la segunda secuencia de unión a la proteína CRISPR de Clase 2 Tipo II de unión a ácidos nucleicos, comprendiendo la secuencia 2 de ácidos nucleicos repetida además, una secuencia 2-1 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1bC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - una secuencia 2-2 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3' - una secuencia 1aC de ácidos nucleicos repetida, que tiene un extremo 5' y un extremo 3' - y una secuencia 2-3 de ácidos nucleicos del elemento enlazador, que tiene un extremo 5' y un extremo 3', dispuestas en el siguiente orden 3' a 5': la secuencia de ácidos nucleicos del elemento enlazador 2-3, la secuencia 1bC de ácidos nucleicos repetida, la secuencia 2-2 de ácidos nucleicos del elemento enlazador, la secuencia 1aC de ácidos nucleicos repetida y la secuencia 2-1 de ácidos nucleicos del elemento enlazador; en el que la secuencia 1 de ácidos nucleicos repetida está asociada con la secuencia 2 de ácidos nucleicos repetida a través del enlace de hidrógeno entre la secuencia 1a de ácidos nucleicos repetida y la secuencia 1aC de ácidos nucleicos repetida y a través del enlace de hidrógeno entre la secuencia 1b de ácidos nucleicos repetida y la secuencia 1bC de ácidos nucleicos repetida.

65. El andamio de ácidos nucleicos manipulado según el artículo 64, que además comprende un tercer elemento 2 que comprende una secuencia 2 de ácidos nucleicos, que tiene un extremo 5' y un extremo 3', en el que (i) el extremo 3' de la secuencia 2 de ácidos nucleicos está unida covalentemente al extremo 5' de la secuencia 1C de ácidos nucleicos repetida, y (ii) la secuencia 2 de ácidos nucleicos comprende una secuencia 2 de ácidos nucleicos espaciadores.

Parte experimental

Los aspectos de la presente invención se ilustran en los siguientes ejemplos. Se han realizado esfuerzos para garantizar la precisión con respecto a los números utilizados (por ejemplo, cantidades, concentraciones, cambios porcentuales y similares), pero se deben tener en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, la temperatura está en grados centígrados y la presión está en o cerca de la atmosférica.

Ejemplo 1

Los ejemplos que no pertenecen a la invención son solo para fines ilustrativos. Diseño *in silico* de componentes polinucleotídicos NASC.

El presente ejemplo proporciona una descripción del diseño de los componentes polinucleotídicos de NASC para una serie de realizaciones de las NASC descritas en este documento.

- 5 La Tabla 9 muestra una correlación entre los componentes de polinucleótido de NASC y las estructuras ilustradas en las Figuras. La columna "Proteína CAS asociada" enumera las proteínas Cas con las que se puede usar el componente polinucleótido NASC. A menos que se indique lo contrario, la proteína Cas9 es una proteína Cas9 de *S. pyogenes* (proteína Cas9 de *S. pyogenes*, SEQ ID NO. 100 o proteína dCas9 de *S. pyogenes* (SEQ ID NO. 101)). La proteína Cpf1 es un *Acidaminococcus* sp. proteína Cpf1 (dCpf SEQ ID NO. 105) a menos que se indique lo contrario.
- 10 Las secuencias que hibridan entre los componentes polinucleotídicos están subrayadas. Las secuencias de unión a los ácidos nucleicos están indicadas por una serie de veinte Ns, en el que N es cualquier nucleótido. Un experto en la técnica puede diseñar una secuencia de unión a diana de ácidos nucleicos.

Tabla 9

Ejemplos de secuencias de componentes polinucleotídicos de NASC					
Figura	Indicador de Figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
4A	1-1, 1-2, 1-3	NASC-PC1	Cpf1	<u>CUCCGGCGAUGUCACAC</u> <u>CGAACUGAUAAUUUCUA</u> CUCUUGUAGAUNNNNNN NNNNNNNNNNNNNNNN	SEQ ID NO. 64
4A	2-3, 2-2, 2-1	NASC-PC2	Cpf1	<u>UCGGUGUGACAUCGCCG</u> <u>GAGUUGAUAAAUUUCUA</u> CUCUUGUAGAUNNNNNN NNNNNNNNNNNNNNNN	SEQ ID NO. 65
4B	1-1	NASC-PC1-1	Cpf1	<u>CUCCGGCGAUGUCACAC</u> <u>CGAACUGAUAAUUUCUA</u> C	SEQ ID NO. 66
4B	1-3	NASC-PC1-2	Cpf1	<u>GUAGAUNNNNNNNNNNN</u> NNNNNNNNNNNN	SEQ ID NO. 67
4B	2-1	NASC-PC2-1	Cpf1	<u>UCGGUGUGACAUCGCCG</u> <u>GAGUUGAUAAAUUUGU</u> AG	SEQ ID NO. 68
4B	2-3	NASC-PC2-2	Cpf1	<u>CUACAUNNNNNNNNNNN</u> NNNNNNNNNN	SEQ ID NO. 69

(continuación)

Figura	Indicador de Figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
4C	1-1, 1-3, 1-2	NASC-PC1	Cpf1	AAUUUCUACUCUUGUAG AUNNNNNNNNNNNNNNN NNNNNNACUGAUCUCCG <u>GCGAUGUCACACCGA</u>	SEQ ID NO. 70
4C	2-2, 2-3, 2-2	NASC-PC2	Cpf1	AAUUUCUACUCUUGUAG AUNNNNNNNNNNNNNNN NNNNNNUUGAUUAUCGGU <u>GUGACAUCGCCGGAG</u>	SEQ ID NO. 71
4M	1-1, 1-2, 1-3, 1-4	NASC-PC1	Cpf1	<u>CUCCGGCGAUGUCACAC</u> <u>CGAACUGAUAAUUUCUA</u> CUCUUGUAGAUNNNNNN NNNNNNNNNNNNNNNUU GAUAGUCUAAGGCAGCU <u>AGGGUCU</u>	SEQ ID NO. 72
4M	2-1, 2-2, 2-3	NASC-PC2	Cpf1	<u>UCGGUGUGACAUCGCCG</u> <u>GAGUUGAUAAAUUUCUA</u> CUCUUGUAGAUNNNNNN NNNNNNNNNNNNNNNN	SEQ ID NO. 73
4M	3-2, 3-3, 3-1	NASC-PC3	Cpf1	AAUUUCUACUCUUGUAG AUNNNNNNNNNNNNNNN NNNNNNUACCAAAGACC <u>CUAGCUGCCUUAGAC</u>	SEQ ID NO. 74
5D	507-500 I	NASC-PC1	Cpf1	<u>CUCCGGCGAUGUCACAC</u> <u>CGAACUGAUGUCUAAGG</u> CAGCUAGGGUCUUUGAU AAAUUUCUACUCUUGUA GAUNNNNNNNNNNNNN NNNNNNN	SEQ ID NO. 75

(continuación)

FIGURA	Indicador de figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
5D	508-515 II	NASC-PC2	Cpf1	<u>UGCGAACCACUGUGAGC</u> <u>CAGUACCAAUCGGUGUG</u> <u>ACAUCGCCGGAGUUGAU</u> <u>AAAUUUCUACUCUUGUA</u> <u>GAUNNNNNNNNNNNNN</u> <u>NNNNNNNN</u>	SEQ ID NO. 76
5F	VIII 507-503	NASC-PC1-1	Cpf1	<u>CUCCGGCGAUGUCACAC</u> <u>CGAACUGAUGUCUAAGG</u> <u>CAGCUAGGGUCUUUGAU</u> <u>AAAUUUCUAC</u>	SEQ ID NO. 77
5F	VIII 502-500	NASC-PC1-2	Cpf1	<u>GUAGAUNNNNNNNNNNN</u> <u>NNNNNNNNNNNN</u>	SEQ ID NO. 78
5F	V 508-512	NASC-PC2-1	Cpf1	<u>UGCGAACCACUGUGAGC</u> <u>CAGUACCAAUCGGUGUG</u> <u>ACAUCGCCGGAGUUGAU</u> <u>AAAUUUGUAG</u>	SEQ ID NO. 79
5F	V 513-515	NASC-PC2-2	Cpf1	<u>CUACAUNNNNNNNNNNN</u> <u>NNNNNNNNNNNN</u>	SEQ ID NO. 80
6D	611-601	NASC-PC1	Cas9	<u>NNNNNNNNNNNNNNNNNN</u> <u>NNNNGUUUUAGUCCCUA</u> <u>AUUAAAUUUCUUGAAAU</u> <u>UGGUUAUAUAAGGAGGG</u> <u>ACUACAACAAAGAGUUU</u> <u>GCGGGACUCUGCGGGGU</u> <u>UACAAUCCCCUAAAACC</u> <u>GCUUUUAAAAUCAAU</u> <u>AAAUUUUGCUUU</u>	SEQ ID NO. 81

(continuación)

FIGURA	Indicador de figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
6D	612-622	NASC-PC2	Cas9	<u>NNNNNNNNNNNNNNNNNNNNNN</u> <u>NNNNGUUGUAGUCCCUC</u> <u>CUUAUAUACCAAGAAAA</u> <u>AGAAAUUUAUUAGGG</u> <u>ACUAAAACAAAGAGUUU</u> <u>GCGGGACUCUGCGGGGU</u> <u>UACAAUCCCCUAAAACC</u> <u>GCUUUUAAAAUUCAAAU</u> <u>AAAUUUUGCUUU</u>	SEQ ID NO. 82
6G	611-601	NASC-PC1	Cas9	<u>NNNNNNNNNNNNNNNNNNNNNN</u> <u>NNNNGUCUCAGAGCUAU</u> <u>GCAGUCCUGGACAACUG</u> <u>CCGAACCUCAUGAGAAU</u> <u>CCAAGUAUGUGUAAGGC</u> <u>UAGUCCGUUAUCAACUU</u> <u>GAAAAAGUGGCACCGAG</u> <u>UCGGUGCUU</u>	SEQ ID NO. 83
6G	622-612	NASC-PC2	Cas9	<u>NNNNNNNNNNNNNNNNNNNNNN</u> <u>NNNNGCACAUGAGGAUU</u> <u>CUCAUGAGGGACGGCAG</u> <u>AAGAACAGGACUGCAUA</u> <u>GCAAGUUGAGAUAAAGGC</u> <u>UAGUCCGUUAUCAACUU</u> <u>GAAAAAGUGGCACCGAG</u> <u>UCGGUGCUU</u>	SEQ ID NO. 84
6I	IV	NASC-PC1	Cas9	<u>NNNNNNNNNNNNNNNNNNNNNN</u> <u>NNNNGUUUUAGAGCUAU</u> <u>GCUGUUUUGGAAAGGUC</u> <u>AUGUCCUUCAAAGUUGU</u> <u>AAUAAGGCUAGUCCGUU</u> <u>AUCAACUUGAAAAAGUG</u> <u>GCACCGAGUCGGUGCUU</u>	SEQ ID NO. 85

(continuación)

FIGURA	Indicador de figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
61	V	NASC-PC2	Cas9	NNNNNNNNNNNNNNNNN NNNNGUAAUAGAAUCGU <u>GCUGAAAAGGAAACAAA</u> <u>ACAGCAUAGCAAGUUAA</u> <u>AAUAAGGCUAGUCCGUU</u> AUCAACUUGAAAAAGUG GCACCGAGUCGGUGCUU	SEQ ID NO. 86
61	VI	NASC-PC3	Cas9	NNNNNNNNNNNNNNNNN NNNNGUUACAGAUGAAG <u>GACAUGACCGAAACUUU</u> <u>UCAGCACGAUAAGUUAAU</u> <u>UAUAAGGCUAGUCCGUU</u> AUCAACUUGAAAAAGUG GCACCGAGUCGGUGCUU	SEQ ID NO. 87
7B	725-722	NASC-PC1	Cas9	NNNNNNNNNNNNNNNNN NNNNGUUUUAGUCCCUA <u>AUUAAAUUUCUU</u>	SEQ ID NO. 88
7B	721-718	NASC-PC2	Cas9	NNNNNNNNNNNNNNNNN NNNNGUUGUAGUCCUC <u>CUUAUAUACCAA</u>	SEQ ID NO. 89
7B	717-700	NASC-PC-NTS	Cas9	<u>AAGAAAUUUAAUUAGG</u> <u>GACUAAAACAAAGAGUU</u> UGCGGGACUCUGCGGGG UUACAAUCCCCUAAAAC CGCUUUUAAAAUUCAAA UAAAUUUUGCUUUAGUU <u>GAUAAAUUUGGUAAU</u> <u>AAGGAGGGACUACAACA</u> AAGAGUUUGCGGGACUC UGCGGGGUUACAAUCCC CUAAAACCGCUUUUAAA AUUCAAAUAAAUUUUGC UUU	SEQ ID NO. 90

(continuación)

FIGURA	Indicador de figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
8L	I 809-817	NASC-PC1	Cas9	NNNNNNNNNNNNNNNNNNNNNN NNNNGUUUUAGAGCTAT <u>GCTGTGAAAACAGCATA</u> <u>GCAAGUUAAAAUAAGGC</u> UACUGCCG	SEQ ID NO. 91
8L	II 828-836	NASC-PC2	Cas9	NNNNNNNNNNNNNNNNNNNNNN NNNNGUUUUAGAGCTAT <u>GCTGTGAAAACAGCATA</u> <u>GCAAGUUAAAAUAAGGC</u> UAGUCACG	SEQ ID NO. 92
8L	835-800	NASC-PC-NTS	Cas9	<u>CGGCAGUCCGUUAUCAA</u> CUUGAAAAAGUGGCACC GAGUCGGUGCUUAGUUG AUAAU <u>CGUGACGUCCG</u> UUAUCAACUUGAAAAAG UGGCACCGAGUCGGUGC UUU	SEQ ID NO. 93
9A	II+I	NASC-PC1-2TS	Cas9 Cpf1	NNNNNNNNNNNNNNNNNNNNNN NNNNGUUUUAGAGCUAU GCUGUGAAAACAGCAUA GCAAGUUAAAAUAAGGC UAGUCCGUUAUCAACUU GAAAAAGUGGCACCGAG UCGGUGCUUACUGAUAA UUUCUACUCUUGUAGAU NNNNNNNNNNNNNNNNNNNNNN NNNN	SEQ ID NO. 94
6J	VII	NASC-PC1-2TS	Cas9 Cpf1	NNNNNNNNNNNNNNNNNNNNNN NNNNGUUUUAGAGCUAU <u>GCUGU</u> UACCAAUUGAUA <u>GUAGA</u> NNNNNNNNNNNNNNNN NNNNNNNNNNNN	SEQ ID NO. 97

(continuación)

FIGURA	Indicador de figura	Designación genérica del componente NASC	Proteína Cas asociada	Secuencia	SEQ ID NO.
6J	VIII	NASC-PC-NTS	Cpf1 Cas9	AAUUUCUACUUGAUAAAC AGCAUAGCAAGUUAAAA UAAGGCUAUCCGUUAUC AACUUGAAAAAGUGGCA CCGAGUCGGUGCUUU	SEQ ID NO. 98
6K	X	NASC-PC1	Cas9	NNNNNNNNNNNNNNNNNN NNNNGUUUUUGUACUCU CAAGAUUCAAAUAACAG CAUAGCAAGUUAAAAUA AGGCUAUCCGUUAUCAA CUUGAAAAAGUGGCACC GAGUCGGUGCUUU	SEQ ID NO. 95
6K	IX	NASC-PC2	Cas9*	NNNNNNNNNNNNNNNNNN NNNNGUUUUUAGAGCUAU GCUGUUAACGUAAAUCUU GCAGAAGCUACAAAGAU AAGGCUUCAUGCCGAAA UCAACACCCUGUCAUUU UAUGGCAGGGUGUUU	SEQ ID NO. 96
* proteína CRISPR-I Cas9 de <i>S. Thermophilus</i> , SEQ ID NO. 108 o proteína CRISPR-I dCas9 de <i>S. thermophilus</i> , SEQ ID NO. 109					

5 Siguiendo los lineamientos de la presente especificación, un experto en la técnica puede diseñar componentes de polinucleótidos NASC (por ejemplo, basados en otros componentes de polinucleótidos NASC descritos aquí) para diferentes proteínas Cas relacionadas (por ejemplo, proteína Cas9 de *C. jejuni* (SEQ ID NO. 103), proteína dCas9 *C. jejuni* (SEQ ID NO. 56), Cas9 *S. aureus* (SEQ ID NO. 99), dCas9 *S. aureus* (SEQ ID NO. 102), proteína Cpf1 de la bacteria *Lachnospiraceae* (SEQ ID NO. 106)), proteína dCpf1 de la bacteria *Lachnospiraceae* (SEQ ID NO. 107) o Cpf1 *Acidaminococcus* sp. (SEQ ID NO. 104).

Ejemplo 2

10 Producción de ARNs y componentes polinucleótidos NASC

Este ejemplo describe la producción de ARNs y componentes de polinucleótidos NASC NASC-PC1 (Tabla 9, secuencia diana genérica SEQ ID NO. 83) y NASC-PC2 (Tabla 9, secuencia diana genérica SEQ ID NO. 84), como se ilustra en Figura 6G. Los ARNs y los componentes polinucleotídicos de NASC descritos en este Ejemplo se usaron en ensayos de escisión de Cas (Ejemplo 5).

- 15 NASC-PC1 y NASC-PC2 comprendieron diferentes secuencias de ácidos nucleicos del primer elemento de tallo (ilustradas en las Figuras 6E, 608-609 y 619-620) para limitar la formación de estructuras secundarias dentro de cada NASC-PC que pueden interferir con la formación de estructura estable entre las primeras secuencias de ácidos nucleicos del elemento de tallo NASC-PC1 y las secuencias de ácidos nucleicos del primer elemento de tallo NASC-PC2 complementarias a la primera secuencia de ácidos nucleicos del elemento de tallo.

Se utilizaron dos cadenas principales de ARNsg (ARNsg-1 y ARNsg-2), cada una de las cuales comprende diferentes secuencias de ácidos nucleicos de tallo superior e inferior (ilustradas en la Figura 2C, 221-222/227-228 y 223-224/225). 226, respectivamente); las secuencias de protuberancia fueron las mismas.

5 Se seleccionaron cuatro secuencias de unión a diana de ácidos nucleicos, cada una de 20 nucleótidos de longitud. Una de las cuatro secuencias de unión a los ácidos nucleicos se incorporó en el extremo 5' de un ARNsg-1 y un esqueleto ARNsg-2, y el extremo 5' de una NASC-PC1 y una NASC-PC2. Las cuatro secuencias diana de ADN de doble cadena eran las siguientes: la diana 1 (AAVST1) correspondía a una secuencia diana de AAVS-1 humana. Diana 2 (VT2), Diana 3 (VT3) y Diana 4 (VT4) fueron secuencias diana de ADN presentes en la secuencia del vector (SEQ ID NO. 20).

10 Los componentes de ARN se produjeron mediante transcripción *in vitro* usando un kit de síntesis de ARN de alto rendimiento T7 (New England Biolabs, Ipswich, MA) de una plantilla de ADN de doble cadena que incorpora un promotor T7 en el extremo 5' de las secuencias de ADN.

Se montó una plantilla de ADN de doble cadena para cada ARNsg, NASC-PC 1 y NASC-PC2 mediante PCR utilizando cebadores de oligonucleótidos 3' superpuestos que contienen secuencias de ADN correspondientes a cada ARNsg, NASC-PC1 y NASC-PC2. Los cebadores oligonucleotídicos se presentan en la Tabla 10.

15

Tabla 10

Cebadores superpuestos para la generación de plantillas de codificación de ARNsg, NASC-PC1 y NASC-PC2		
Designación de constructo	Diana	Oligonucleótido
ARNsg-1-AA V ST1	Diana-1	SEQ ID NO. 1, 3, 11, 12, 2
ARNsg -1-VT2	Diana-2	SEQ ID NO. 1, 4, 11, 12, 2
ARNsg -1-VT3	Diana-3	SEQ ID NO. 1, 5, 11, 12, 2
ARNsg -1-VT4	Diana-4	SEQ ID NO. 1, 6, 11, 12, 2
ARNsg -2-AAVST1	Diana-1	SEQ ID NO. 1, 7, 13, 14, 2
ARNsg -2-VT2	Diana-2	SEQ ID NO. 1, 8, 13, 14, 2
ARNsg -2-VT3	Diana-3	SEQ ID NO. 1, 9, 13, 14, 2
ARNsg -2-VT4	Diana-4	SEQ ID NO. 1, 10, 13, 14, 2
NASC-PC1-AAVST1	Diana-1	SEQ ID NO. 1, 3, 15, 16, 2
NASC-PC1-VT2	Diana-2	SEQ ID NO. 1, 4, 15, 16, 2
NASC-PC1-VT3	Diana-3	SEQ ID NO. 1, 5, 15, 16, 2
NASC-PC1-VT4	Diana-4	SEQ ID NO. 1, 6, 15, 16, 2
NASC-PC2-AAVST1	Diana-1	SEQ ID NO. 1, 7, 17, 18, 2
NASC-PC2-VT2	Diana-2	SEQ ID NO. 1, 8, 17, 18, 2
NASC-PC2-VT3	Diana-3	SEQ ID NO. 1, 9, 17, 18, 2

(continuación)

Cebadores superpuestos para la generación de plantillas de codificación de ARNsg, NASC-PC1 y NASC-PC2		
Designación de constructo	Designación de constructo	Designación de constructo
NASC-PC2-VT4	Diana-4	SEQ ID NO. 1, 10, 17, 18, 2

Los cebadores de ADN estaban presentes a una concentración de 2 nM cada uno. Un cebador de ADN correspondió al promotor T7 (SEQ ID NO. 1) y el otro al extremo 3' de la secuencia de ARN (SEQ ID NO. 2) y se usaron a una concentración de 640 nM para dirigir la reacción de amplificación. Las reacciones de PCR se realizaron utilizando la mezcla maestra 2X Hot Start High-Fidelity 2X (New England Biolabs, Ipswich, MA) siguiendo las instrucciones del fabricante. Las reacciones de ensamblaje de PCR se llevaron a cabo utilizando las siguientes condiciones de ciclos térmicos: 98°C durante 2 minutos; 2 ciclos de 20 segundos a 98°C, 20 segundos a 52.5°C, 20 segundos a 72°C; seguido de 32 ciclos de 20 segundos a 98°C, 20 segundos a 57°C, 20 segundos a 72°C; y una extensión final a 72°C durante 2 minutos. La calidad del producto de ADN se evaluó después de la reacción de PCR mediante electroforesis en gel de agarosa (1.5%, SYBR® Safe, Life Technologies, Grand Island, NY).

Entre 0,25-05 µg de la plantilla de ADN para cada ARNsg, NASC-PC1 y NASC-PC2 se usó como plantilla para la transcripción usando el kit de síntesis de ARN de alto rendimiento T7 (New England Biolabs, Ipswich, MA) durante aproximadamente 16 horas a 37°C. Las reacciones de transcripción se trataron con ADNasa I (New England Biolabs, Ipswich, MA) y se purificaron utilizando el kit de concentración y concentración de ARN GeneJet (Life Technologies, Grand Island, NY). El rendimiento de ARN se cuantificó utilizando el sistema Nanodrop™ 2000 (Thermo Scientific, Wilmington, DE). La calidad del ARN transcrito se verificó mediante electroforesis en gel de agarosa (2%, SYBR® Safe; Life Technologies, Grand Island, NY). Las secuencias de componentes de ARNsg y polinucleótido NASC se muestran en la Tabla 11.

Tabla 11

ARNsg, NASC-PC1, y NASC-PC2 Secuencias		
Nombre	Secuencia*	SEQ ID NO.
ARNsg-1-AAVST1	GGGGCCACUAGGGACAGGAUGUCUCAGAGCUA UGCAGUCCUGGACAACUGCCGAACAGGACUGC AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGGCACCGAGUCGGUGCUU	SEQ ID NO. 25
ARNsg-1-VT2	GUAGGCUAUAGUGUAGAUCUGUCUCAGAGCUA UGCAGUCCUGGACAACUGCCGAACAGGACUGC AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGGCACCGAGUCGGUGCUU	SEQ ID NO. 26
ARNsg-1-VT3	GGAAAAAGUGGAAGCGGCGAGUCUCAGAGCUA UGCAGUCCUGGACAACUGCCGAACAGGACUGC AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGGCACCGAGUCGGUGCUU	SEQ ID NO. 27

(continuación)

ARNsg, NASC-PC1, y NASC-PC2 Secuencias		
Nombre	Nombre	Nombre
ARNsg-1-VT4	GGCGAUAAGUCGUGUCUUACGUCUCAGAGCUA UGCAGUCCUGGACAACUGCCGAACAGGACUGC AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 28
ARNsg-2-AAVST1	GGGGCCACUAGGGACAGGAUGCACAUGAGGAU UCUCAUGAGGGACGGCAGAAGAACCUCAUGAG AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 29
ARNsg-2-VT2	GUAGGCUAUAGUGUAGAUCUGCACAUGAGGAU UCUCAUGAGGGACGGCAGAAGAACCUCAUGAG AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 30
ARNsg-2-VT3	GGAAAAAGUGGAAGCGGCGAGCACAUGAGGAU UCUCAUGAGGGACGGCAGAAGAACCUCAUGAG AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 31
ARNsg-2-VT4	GGCGAUAAGUCGUGUCUUACGCACAUGAGGAU UCUCAUGAGGGACGGCAGAAGAACCUCAUGAG AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 32
NASC-PC1- AAVST1	GGGGCCACUAGGGACAGGAUGUCUCAGAGCUA <u>UGCAGUCCUGGACAACUGCCGAACCUCAUGAG</u> <u>AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC</u> AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 33
NASC-PC1-VT2	GUAGGCUAUAGUGUAGAUCUGUCUCAGAGCUA <u>UGCAGUCCUGGACAACUGCCGAACCUCAUGAG</u> <u>AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC</u> AACUUGAAAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 34

NASC-PC1-VT3	<u>GGAAAAAGUGGAAGCGGCGAGUCUCAGAGCUA</u> <u>UGCAGUCCUGGACAACUGCCGAACCUCAUGAG</u> <u>AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 35
NASC-PC1-VT4	<u>GGCGAUAAGUCGUGUCUACGUCUCAGAGCUA</u> <u>UGCAGUCCUGGACAACUGCCGAACCUCAUGAG</u> <u>AAUCCAAGUAUGUGUAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 36
NASC-PC2-AAVST1	<u>GGGGCCACUAGGGACAGGAUGCACAUGAGGAU</u> <u>UCUCAUGAGGGACGGCAGAAGAACAGGACUGC</u> <u>AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 37
NASC-PC2-VT2	<u>GUAGGCUAUAGUGUAGAUCUGCACAUGAGGAU</u> <u>UCUCAUGAGGGACGGCAGAAGAACAGGACUGC</u> <u>AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 38
NASC-PC2-VT3	<u>GGAAAAAGUGGAAGCGGCGAGCACAUGAGGAU</u> <u>UCUCAUGAGGGACGGCAGAAGAACAGGACUGC</u> <u>AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 39
NASC-PC2-VT4	<u>GGCGAUAAGUCGUGUCUACGCACAUGAGGAU</u> <u>UCUCAUGAGGGACGGCAGAAGAACAGGACUGC</u> <u>AUAGCAAGUUGAGAUAAAGGCUAGUCCGUUAUC</u> <u>AACUUGAAAAAGUGGCACCGAGUCGGUGCUU</u>	SEQ ID NO. 40
*Las regiones de hibridación de NASC-PC están subrayadas		

Este procedimiento para la producción de componentes de polinucleótido ARNsg y NASC puede aplicarse a la producción de otros componentes de polinucleótido ARNsg y NASC por un experto en la técnica en vista de las enseñanzas de la especificación.

5 Ejemplo 3

Producción de secuencias diana de ADN de doble cadena para uso en ensayos de escisión mediante la clonación de secuencias diana de ADN de doble cadena en plásmidos

Las secuencias diana de ADN de doble cadena para uso en los ensayos de escisión de proteína Cas *in vitro* se produjeron a través de la unión de una secuencia diana de ácidos nucleicos de doble cadena (por ejemplo, la secuencia de destino AAVS-1) en un esqueleto de vector de clonación. Cada vector se transformó en una cepa adecuada de *E. coli* para la producción de secuencias diana de ADN de doble cadena.

Se adjuntó una secuencia diana de ADN monocatenario de 25 nucleótidos correspondiente al Sitio 1 de Integración de Virus Adenoasociado (AAVS-1) humano con una secuencia de ácidos nucleicos aleatorizada de 47 nucleótidos en el extremo 5' y una secuencia de ácidos nucleicos aleatorizada de 53 nucleótidos en el extremo 3'. Los cebadores

oligonucleotídicos directos e inversos compatibles con el sistema de vectores Electra™ (DNA2.0, Newark, CA) se incorporaron en el extremo 5' y el extremo 3' de la secuencia diana del ADN, produciendo una secuencia de ADN monocatenaria de 237 pb. Se proporcionó una secuencia de ácidos nucleicos de 237 pb de ADN monocatenario ("fragmento de clonación de ADN"), así como secuencias de ácidos nucleicos de cebadores de oligonucleótidos de amplificación directa e inversa, a un fabricante comercial para la síntesis. Estas secuencias de ADN monocatenarias se muestran en la Tabla 12.

Tabla 12

Secuencias de ADN monocatenario		
Descripción	Secuencia*	SEQ ID NO.
Fragmento de clonación de ADN	TACACGTACTTAGTCGCTGAAGCTCTTCTAT GCAAGCAGAAGACGGCATAACGAGATCGAG TAATGTGACTGGAGTTCAGACGTGTGCTCT TCCGATCTGCTACTGGGGCCACTAGGGACA GGATNNGGTGCTAGCTCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGTAGGCTATAGTG TAGATCTCGGTGGTCGCCGTATCATTGGTA GAAGAGCCGTCAATCGAGTTCGTACCT	SEQ ID NO. 19
Cebador de avance	TACACGTACTTAGTCGCTGAAGCTCTTCTAT GCAAGCAGAAGACGGCATAACGAGAT	SEQ ID NO. 21
Cebador reverso	AGGTACGAACCTCGATTGACGGCTCTTCTAC CAATGATACGGCGACCACCGAGATCT	SEQ ID NO. 22
*La secuencia diana de ADN AAVS-1 que comprende un PAM está subrayada		

El fragmento de clonación de ADN monocatenario se amplificó mediante PCR para generar ADN bicatenario para su uso con el Sistema de vectores Electra™ (ADN2.0, Newark, CA). La mezcla de reacción de PCR fue la siguiente: 0,5 unidades de KAPA HiFi Hot Start DNA Polymerase (Kapa Biosystems, Wilmington, MA), 1x tampón de reacción, dNTPs 0,3 mM, cebador directo 200 nM (SEQ ID NO. 21), cebador reverso 200nM (SEQ ID NO. 22) y 80 nM del fragmento de clonación de ADN (SEQ ID NO. 19) en un volumen total de 25 µL. El fragmento de clonación de ADN se amplificó utilizando las siguientes condiciones: 95°C durante 4 minutos, 30 ciclos de 20 segundos a 98°C, 20 segundos a 60°C y 30 segundos a 72°C, seguido de una extensión final a 72°C durante 5 minutos. Los productos de PCR se purificaron utilizando tubos de purificación de PCR Spin Smart™ (Denville Scientific, South Plainfield, NJ) y se cuantificaron utilizando un espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE).

El fragmento de clonación de ADN bicatenario se clonó en el vector "pD441-SR: T5-sRBS-ORF, Ecoli-Elec D" disponible comercialmente (un vector de ADN bacteriano Electra™ (DNA2.0, Newark, CA)) utilizando el protocolo de clonación del fabricante. Se preparó la siguiente mezcla de reacción de clonación: 20 ng del fragmento de clonación amplificado por PCR, 20 ng del vector de ADN bacteriano, 2 µL de la mezcla de tampón Electra™ (ADN2.0, Newark, CA) y 1 µL de la mezcla de enzimas Electra™ (ADN2.0, Newark, CA) en un volumen final de 20 µL. La mezcla de reacción de clonación se agitó brevemente a continuación, se sometió a centrifugación usando una centrifuga de sobremesa y se incubó a temperatura ambiente durante 20 minutos.

Después de la incubación, se mezclaron 1 µL de One Shot® Mach1™ T1R (Thermo Scientific, Wilmington, DE) células de *E. coli* químicamente competentes con 2 µL de la mezcla de reacción de clonación para formar una mezcla de transformación que se incubó en hielo durante 30 minutos. La mezcla de transformación se sometió a choque térmico durante 30 segundos a 42°C y se incubó en hielo durante 2 minutos. Se añadieron 250 µL de medio S.O.C. a temperatura ambiente (Thermo Scientific, Wilmington, DE) a la mezcla de transformación, y la mezcla se incubó a 37°C durante 1 hora con agitación. Después de esta incubación, 50 µL de la mezcla celular se extendieron en una placa de agar LB con 50 µg/ml de kanamicina, y la placa se incubó durante la noche a 37°C para la formación de colonias bacterianas.

Se recogieron cinco colonias bacterianas y se transfirieron a tubos de cultivo separados de 15 ml que contenían 5 ml de LB suplementado con 50 µg/ml de medio de cultivo de kanamicina y los tubos se incubaron durante 8 horas con agitación. Las células se sedimentaron mediante centrifugación a 4000 RPM durante 15 minutos, se aspiró el medio de cultivo y las células se resuspendieron en 200 µl de medio de cultivo LB sin antibióticos. Los vectores de ADN se extrajeron de las bacterias de cada una de las cinco colonias bacterianas utilizando el kit QIAprep Spin Miniprep Kit (Qiagen, Venlo, Países Bajos) siguiendo las instrucciones del fabricante. Los rendimientos del vector de ADN se cuantificaron utilizando un espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE). Se secuenciaron 250 ng de cada vector de ADN para verificar la incorporación del fragmento de clonación de ADN correspondiente a la SEQ ID NO. 19. La secuencia completa del vector de ADN, incluida la secuencia diana AAVS-1, se proporciona como SEQ ID NO. 20.

Un clon bacteriano identificado con contenido de un vector de ADN que comprende el fragmento de clonación de ADN se cultivó en 100 ml de LB suplementado con 50 µg/ml de medio de cultivo de kanamicina y se cultivó durante una noche a 37°C con agitación. Las células se sedimentaron mediante centrifugación a 4000 RPM durante 15 minutos, se aspiró el medio de cultivo y el vector de ADN se purificó utilizando un kit QIAprep Spin Maxiprep (Qiagen, Venlo, Países Bajos) siguiendo las instrucciones del fabricante. Los rendimientos del vector de ADN se cuantificaron utilizando un espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE).

El vector de ADN se preparó para usarse en el ensayo de escisión de Cas por linealización del vector circular con una endonucleasa de restricción de tipo II de *Ascl*. Para linealizar el vector de ADN circular, se montó la siguiente mezcla de reacción: 1 unidad de endonucleasa de restricción *Ascl* (New England Biolabs, Ipswich, MA) por 1 µg de vector de ADN circular y 1x tampón CutSmart® (New England Biolabs, Ipswich, MA) En un volumen final de 50 µL. La mezcla de reacción se incubó durante 1 hora a 37°C, y la reacción se detuvo por incubación a 80°C durante 20 minutos. El vector de ADN lineal se purificó utilizando el kit de purificación de PCR QIAquick (Qiagen, Venlo, Países Bajos) siguiendo las instrucciones del fabricante. Los rendimientos del vector de ADN lineal se cuantificaron usando un espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE). Se pueden usar otros procedimientos de clonación y vectores de ADN adecuados para la incorporación de secuencias diana de ADN de doble cadena siguiendo esencialmente el procedimiento descrito en este Ejemplo. Si la linealización del vector de ADN es indeseable o innecesaria, se puede usar un vector de ADN circular en los ensayos de división de Cas.

Ejemplo 4

Producción de secuencias diana de ADN de doble cadena para su uso en ensayos de escisión mediante PCR

Las secuencias diana de ADN de doble cadena para uso en ensayos de escisión de proteína Cas *in vitro* pueden producirse usando amplificación por PCR de secuencias diana de ácidos nucleicos seleccionadas de ADN genómico humano.

El ADN genómico humano que comprende el sitio de integración de virus adenoasociado 1 (AAVS-1) se puede preparar mediante extracción con fenol-cloroformo de la línea celular humana K562 (American Type Culture Collection (ATCC), Manassas, VA). Las reacciones de PCR pueden llevarse a cabo con la mezcla maestra Q5 Hot Start High-Fidelity 2X (New England Biolabs, Ipswich, MA) siguiendo las instrucciones del fabricante. Se puede usar 20 ng/µL de ADN_g en un volumen final de 25 µl para amplificar la secuencia diana de ácidos nucleicos seleccionada en las siguientes condiciones: 98°C durante 2 minutos, 35 ciclos de 20 segundos a 98°C, 20 segundos a 60°C, 20 segundos a 72°C y una extensión final a 72°C durante 2 minutos. Los productos de PCR se pueden purificar utilizando tubos de purificación de PCR Spin Smart™ (Denville Scientific, South Plainfield, NJ) y se pueden cuantificar utilizando un espectrofotómetro Nanodrop™ 2000 UV-Vis (Thermo Scientific, Wilmington, DE).

En la Tabla 13 se presentan ejemplos de cebadores de avance y reversos que pueden usarse para la amplificación de las secuencias diana de ADN de AAVS-1 a partir de ADN_g.

Tabla 13

Cebadores oligonucleótidos de secuencias diana ADN AAVS-1	
SEQ ID NO.	Secuencia
SEQ ID NO. 23	CCCCGTTCTCCTGTGGATTC
SEQ ID NO. 24	ATCCTCTCTGGCTCCATCGT

La secuencia diana de ADN de AAVS-1 se puede amplificar utilizando la SEQ ID NO. 23 y SEQ ID NO. 24 para producir una secuencia diana de ADN AAVS-1 de doble cadena de 495 pb.

Se pueden obtener otras secuencias diana de ADN de doble cadena adecuadas usando esencialmente el mismo procedimiento eligiendo cebadores de oligonucleótidos adecuados. Se puede usar el ADNg de cualquier organismo (por ejemplo, plantas, bacterias, levaduras, algas y similares) en lugar de ADN derivado de células humanas. Además, las secuencias diana de ADN pueden amplificarse mediante PCR a partir de polinucleótidos distintos de ADNg (por ejemplo, vectores y fragmentos de ADN aislados en gel).

Ejemplo 5

Ensayos de escisión de Cas

Este ejemplo ilustra el uso de composiciones de polinucleótidos de NASC y una proteína Cas9 en un ensayo *in vitro* para evaluar los porcentajes de escisión de secuencias diana de ácidos nucleicos mediante las composiciones de polinucleótidos de NASC.

NASC-PC1 y NASC-PC2 comprendieron diferentes secuencias de ácidos nucleicos del primer elemento de tallo para limitar la formación de estructuras secundarias dentro de cada NASC-PC que pueden interferir con la formación de una estructura secundaria estable a través de la formación de enlaces de hidrógeno entre el primer tallo del NASC-PC1 las secuencias de ácidos nucleicos elemental y las secuencias de ácidos nucleicos del primer elemento de tallo NASC-PC2 complementarias a la secuencia de ácidos nucleicos del primer elemento de tallo NASC-PC2.

Los componentes genéricos de la composición de polinucleótidos NASC usados en este Ejemplo fueron NASC-PC1 (Tabla 9, secuencia diana genérica SEQ ID NO. 83) y NASC-PC2 (Tabla 9, secuencia diana genérica SEQ ID NO. 84). La estructura general de este par NASC-P1/NASC-P2 se ilustra en la Figura 6G.

Los complejos de ribonucleoproteínas de la proteína ARNsg/Cas9 y la proteína NASC-PC1/NASC-PC2/Cas9 se usaron en ensayos de escisión de Cas9 *in vitro* para evaluar el porcentaje de escisión de cada complejo en relación con las secuencias diana de ADN de doble cadena correspondientes en un vector de ADN.

Los complejos de ribonucleoproteínas de la proteína ARNsg/Cas9 y NASC-PC1/NASC-PC2/Cas9 usaron las construcciones de ARNsg y NASC-PC1/NASC-P2 establecidas en el Ejemplo 2, Tabla 11. LA diana 1 (AAVST1) correspondió a un AAVS-1 humano de secuencia diana. Diana 2 (VT2), Diana 3 (VT3) y Diana 4 (VT4) fueron secuencias diana de ADN presentes en la secuencia del vector (SEQ ID NO. 20). En una reacción de escisión con un solo ARNsg o un solo componente polinucleótido NASC, se usó el vector linealizado. En una reacción de escisión con dos componentes de ARNsg o NASC-PC1/NASC-PC2, se utilizó el vector circular. La escisión del plásmido lineal con un ARNsg produjo dos fragmentos de ADN, la escisión del plásmido circular con dos ARNsg o componentes NASC-PC1/NASC-PC2 produjo dos fragmentos diana de ADN. El tamaño de las secuencias diana de ADN de doble cadena y los tamaños de los fragmentos de escisión predichos se presentan en la Tabla 14.

Tabla 14

Diana y Tamaños de fragmentos de escisión			
Objetivo	Vector diana de ADN	Fragmento 1 (bp)	Fragmento 2 (bp)
AAVST1	Lineal	1706	2469
VT2	Lineal	1769	2406
VT3	Lineal	3214	961
VT4	Lineal	350	3825
AAVST1/VT3	Circular	1509	2666
AAVST1/VT4	Circular	1357	2818
VT2/VT3	Circular	1446	2729
VT2/VT4	Circular	1420	2755

5 Los componentes de ARNsg y NASC-PC1/NASC-PC2 se diluyeron a una concentración de trabajo adecuada. Los componentes de ARNsg y NASC-PC se dividieron en alícuotas en tubos separados hasta una concentración final de 50 nM. Los pares de ARNsg y componentes NASC-PC1/NASC-PC2 se dividieron en alícuotas en tubos separados a una concentración final de 50 nM para cada componente. Todos los ARN se incubaron durante 2 minutos a 95°C, se retiraron del termociclador y se dejaron equilibrar a temperatura ambiente. Las combinaciones de los componentes de ARNsg y NASC-PC1/NASC-PC2 que se usaron en las reacciones de escisión se presentan en la Tabla 15.

Tabla 15

Componentes de la mezcla de reacción ARNsg y NASC-PC1/NASC-PC2				
Reacción	Tipo ARN-1	SEQ ID NO.	Tipo ARN-2	SEQ ID NO.
1	ARNsg-1-AAVST1	SEQ ID NO. 25	-	-
2	ARNsg-1-VT2	SEQ ID NO. 26	-	-
3	ARNsg-1-VT3	SEQ ID NO. 27	-	-
4	ARNsg-1-VT4	SEQ ID NO. 28	-	-
5	ARNsg-2-AAVST1	SEQ ID NO. 29	-	-
6	ARNsg-2-VT2	SEQ ID NO. 30	-	-
7	ARNsg-2-VT3	SEQ ID NO. 31	-	-
8	ARNsg-2-VT4	SEQ ID NO. 32	-	-
9	NASC-PC1-AAVST1	SEQ ID NO. 33	-	-
10	NASC-PC1-VT2	SEQ ID NO. 34	-	-
11	NASC-PC1-VT3	SEQ ID NO. 35	-	-
12	NASC-PC1-VT4	SEQ ID NO. 36	-	-
13	NASC-PC2-AAVST1	SEQ ID NO. 37	-	-
14	NASC-PC2-VT2	SEQ ID NO. 38	-	-
15	NASC-PC2-VT3	SEQ ID NO. 39	-	-
16	NASC-PC2-VST4	SEQ ID NO. 40	-	-
17	ARNsg-1-AAVST1	SEQ ID NO. 25	ARNsg-2-VT3	SEQ ID NO. 31
18	NASC-PC1-AAVST1	SEQ ID NO. 33	NASC-PC2-VT3	SEQ ID NO. 39
19	ARNsg-1-AAVST1	SEQ ID NO. 25	ARNsg-2-VT4	SEQ ID NO. 32

(continuación)

Componentes de la mezcla de reacción ARNsg y NASC-PC1/NASC-PC2				
Reacción	Reacción	Reacción	Reacción	Reacción
20	NASC-PC1- AAVST1	SEQ ID NO. 33	NASC-PC2-VT4	SEQ ID NO. 40
21	ARNsg-1-VT2	SEQ ID NO. 26	ARNsg-2-VT3	SEQ ID NO.31
22	NASC-PC1-VT2	SEQ ID NO. 34	NASC-PC2-VT3	SEQ ID NO. 39
23	ARNsg-1-VT2	SEQ ID NO. 26	ARNsg-2-VT4	SEQ ID NO. 32
24	NASC-PC1-VT2	SEQ ID NO. 34	NASC-PC2-VT4	SEQ ID NO. 40

5 Cada componente de la mezcla de reacción de ARNsg y el componente o componentes de la mezcla de reacción NASC-PC1/NASC-PC2 se agregaron a una mezcla de reacción Cas9. La proteína Cas9 de *S. pyogenes* se expresó de forma recombinante en *E. coli* y se purificó para su uso en el ensayo de escisión bioquímica *in vitro*. La mezcla de reacción Cas9 comprendía la proteína Cas9 diluida hasta una concentración final de 200 nM en tampón de reacción (HEPES 20 mM, KCl 100 mM, MgCl₂ 5 mM y glicerol al 5% a pH 7,4). Cada mezcla de reacción Cas9 se incubó a 37°C durante 10 minutos. La escisión en cada mezcla de reacción Cas9 se inició mediante la adición del vector diana de ADN a una concentración final de 5 nM. Cada mezcla de reacción Cas9 se mezcló, se centrifugó brevemente y se incubó durante 15 minutos a 37°C. La reacción de escisión se terminó mediante la adición de Proteinasa K (Denville Scientific, South Plainfield, NJ) a una concentración final de 0,2 µg/µL y 0,44 µmg/µL de Solución de ARNasa A (SigmaAldrich, St. Louis, MO) a cada mezcla de reacción Cas9.

15 Cada mezcla de reacción Cas9 se incubó luego durante 25 minutos a 37°C y 25 minutos a 55°C. Cada mezcla de reacción Cas9 se evaluó para determinar la actividad de escisión mediante el uso del sistema Fragment Analyzer™ (Advanced Analytical Technologies, Ames, IA) y el kit de reactivos NGS de alta sensibilidad DNF-474-05000 (Advanced Analytical Technologies, Ames, IA). Los datos del sistema Fragment Analyzer™ proporcionaron la concentración de cada fragmento de escisión y del vector diana de ADN que permaneció después de la escisión para cada mezcla de reacción Cas9. Para cada mezcla de reacción Cas9, el porcentaje de escisión se calculó dividiendo la suma de los fragmentos de escisión por la suma de ambos fragmentos de escisión y el vector diana de ADN que permaneció después de la escisión.

20 La Tabla 16 presenta los datos de escisión para cada uno de los complejos de ribonucleoproteínas de la proteína ARNsg/Cas9 y NASC-PC1/NASC-PC2/Cas9.

Tabla 16

Escisión bioquímica de secuencias diana de ADN con complejos de proteína ARNsg/Cas9 y complejos de proteína NASC-PC1/NASC-PC2/Cas9					
Reacción	SEQ ID NO.	Tipo ARN-1	SEQ ID NO.	Tipo ARN-2	Porcentaje de escisión
1	SEQ ID NO. 25	ARNsg-1-AAVST1	-	-	100%
2	SEQ ID NO. 29	ARNsg-1-VT2	-	-	95%
3	SEQ ID NO. 26	ARNsg-1-VT3	-	-	92%
4	SEQ ID NO. 30	ARNsg-1-VT4	-	-	94%

(continuación)

Escisión bioquímica de secuencias diana de ADN con complejos de proteína ARNsg/Cas9 y complejos de proteína NASC-PC1/NASC-PC2/Cas9

Reacción	Reacción	Reacción	Reacción	Reacción	Reacción
5	SEQ ID NO. 27	ARNsg-2-AAVST1	-	-	100%
6	SEQ ID NO. 31	ARNsg-2-VT2	-	-	100%
7	SEQ ID NO. 28	ARNsg-2-VT3	-	-	96%
8	SEQ ID NO. 32	ARNsg-2-VT4	-	-	95%
9	SEQ ID NO. 33	NASC-PC 1-AAVST1	-	-	LOD*
10	SEQ ID NO. 37	NASC-PC1-VT2	-	-	LOD
11	SEQ ID NO. 34	NASC-PC1-VT3	-	-	LOD
12	SEQ ID NO. 38	NASC-PC1-VT4	-	-	LOD
13	SEQ ID NO. 35	NASC-PC2- AAVST1	-	-	LOD
14	SEQ ID NO. 39	NASC-PC2-VT2	-	-	LOD
15	SEQ ID NO. 36	NASC-PC2-VT3	-	-	LOD
16	SEQ ID NO. 40	NASC-PC2-VT4	-	-	LOD
17	SEQ ID NO. 25	ARNsg-1-AAVST1	SEQ ID NO. 31	ARNsg-2-VT3	88%
18	SEQ ID NO. 33	NASC-PC 1-AAVST1	SEQ ID NO. 39	NASC-PC2-VT3	89%
19	SEQ ID NO. 25	ARNsg-1-AAVST1	SEQ ID NO. 32	ARNsg-2-VT4	91%
20	SEQ ID NO. 33	NASC-PC1- AAVST1	SEQ ID NO. 40	NASC-PC2-VT4	92%
21	SEQ ID NO. 26	ARNsg-1-VT2	SEQ ID NO. 31	ARNsg-2-VT3	81%
22	SEQ ID NO. 34	NASC-PC1-VT2	SEQ ID NO. 39	NASC-PC2-VT3	86%
23	SEQ ID NO. 26	ARNsg-1-VT2	SEQ ID NO. 32	ARNsg-2-VT4	94%
24	SEQ ID NO. 34	NASC-PC1-VT2	SEQ ID NO. 40	NASC-PC2-VT4	96%

*LOD indica valores de escisión por debajo del límite de detección

Los datos presentados en la Tabla 16 demuestran que cada complejo de proteína NASC-PC1/NASC-PC2/Cas9 de las Reacciones 18, 20, 22 y 24 (Tabla 16) facilitó la escisión específica de sitio mediada por la proteína Cas de las dos secuencias diana de ADN correspondientes a las dos secuencias de unión de ácidos nucleicos del complejo de proteínas NASC-PC1/NASC-PC2/Cas9. Además, el porcentaje de escisión específica del sitio por cada complejo de proteína NASC-PC1/NASC-PC2/Cas9 fue esencialmente equivalente a la escisión específica del sitio de las mismas dos secuencias diana de ADN por dos complejos de proteína ARNsg/Cas9, en el que un ARNsg/Cas9 división proteica dirigida al complejo de proteínas en una primera secuencia diana de ADN y una segunda escisión dirigida a un complejo de proteína ARNsg/Cas9 en una segunda secuencia diana de ADN (compare el porcentaje de escisión de las reacciones 17 con 18; 19 con 20; 21 con 22; y 23 con 24). Los datos presentados en la Tabla 16 también demuestran que se requirió que cada NASC-PC1 se emparejara con una NASC-PC2 complementaria para apuntar a la escisión específica del sitio por las proteínas Cas9 asociadas (véase Tabla 16, Reacciones 9-16); es decir, un componente polinucleótido individual de una composición de polinucleótido NASC era incapaz de soportar la escisión específica del sitio mediada por la proteína Cas.

Siguiendo la guía de la presente especificación y los Ejemplos, el ensayo de escisión bioquímica descrito en este Ejemplo puede ser practicado por un experto en la técnica con otras composiciones de polinucleótidos de NASC y proteínas Cas relacionadas (por ejemplo, proteínas Cas9 y proteínas Cpf1).

Ejemplo 6

Análisis de secuenciación profunda para la detección de modificaciones de secuencia diana en células eucarióticas

Este ejemplo ilustra el uso del análisis de secuenciación profunda para evaluar y comparar el porcentaje de escisión en células usando complejos de proteína polinucleótido NASC/proteína Cas en relación con las secuencias diana de ADN de doble cadena seleccionadas.

A. Selección de la secuencia diana del genoma

Se pueden seleccionar dos secuencias de ácidos nucleicos diana de regiones exónicas en el genoma humano (por ejemplo, la secuencia del gen de Reparación de Rayos X Complementando 5 (XRCC5)). Las secuencias de ácidos nucleicos de veinte nucleótidos de longitud que son adyacentes en 5' a las secuencias de PAM (por ejemplo, un Cas9 de *S. pyogenes* PAM 5' - NGG) pueden seleccionarse, por ejemplo, las secuencias de ADN diana de XRCC5 presentadas en la Tabla 17.

Tabla 17

XRCC5 secuencias de ADN diana			
Nombre Diana	Secuencia diana	coordenada cromosómica hg38	SEQ ID NO.
XRCC5T1	GGTGGACAAGCGGCAGATAG	chr2:216109346-216109365	SEQ ID NO. 41
XRCC5T3	GCACCATGTTGCCGGTCCTC	chr2:216109421-216109440	SEQ ID NO. 42

B. Construcción de composiciones de polinucleótidos NASC

Se puede usar una composición de polinucleótido NASC que comprende NASC-PC1 y NASC-PC2. Las secuencias de unión al ácido nucleico correspondientes a XRCC5T1 pueden incorporarse en el extremo 5' de NASC-PC1, y las secuencias de unión al ácido nucleico correspondientes a XRCC5T3 pueden incorporarse al extremo 5' de NASC-PC2. Como controles positivos, se puede incorporar una secuencia de unión a diana de ácidos nucleicos correspondiente a XRCC5T1 en el extremo 5' de un ARNsg, y una secuencia de unión a diana de ácidos nucleicos correspondiente a XRCC5T3 puede incorporarse en el extremo 5' de ARNsg. NASC-PC1, NASC-PC2 y ARNsgs se pueden producir como se describe en el Ejemplo 2. En la Tabla 18 se dan ejemplos de secuencias para NASC-PC1, NASC-PC2 y ARNsgs.

Tabla 18

Secuencias de componentes polinucleotídicos de ARNsg y NASC			
Designación del componente	Tipo ARN	Secuencia de ARN *	SEQ ID NO.
ARNsg-XRCC5-T1	ARNsg	GGUGGACAAGCGGCAGAUAGGUU UUAGAGCUAUGCUGUUUUGGAAA CAAAACAGCAUAGCAAGUUA UAAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCU U	SEQ ID NO. 43
ARNsg-XRCC5-T3	ARNsg	GCACCAUGUUGCCGGUCCUCGUUU UAGAGCUAUGCUGUUUUGGAAAC AAAACAGCAUAGCAAGUUA AAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 44
NASC-PC1-XRCC5-T1	NASC ₁	GGUGGACAAGCGGCAGAUAGGUU UUAGAGCUAUGCUGUUUUGGAAA CUUUUCAGCACGAUAAGUUAUUA UAAAGGCUAGUCCGUUAUCAACUUG AAAAAGUGGCACCGAGUCGGUGCU U	SEQ ID NO. 45
NASC-PC2-XRCC5-T3	NASC _{1c}	GCACCAUGUUGCCGGUCCUCGUAA UAGAAUCGUGCUGAAAAGGAAAC AAAACAGCAUAGCAAGUUA AAGGCUAGUCCGUUAUCAACUUGA AAAAGUGGCACCGAGUCGGUGCUU	SEQ ID NO. 46
*Las regiones de hibridación de NASC-PC están subrayadas			

C. Formación de complejos de nucleoproteínas de proteína NASC/Cas9

5 Cas9 de *S. pyogenes* se puede etiquetar en el extremo C con dos secuencias de localización nuclear (NLS) y se puede expresar de forma recombinante en *E. coli*, y se puede purificar utilizando procedimientos cromatográficos.

10 Los complejos de ribonucleoproteínas se pueden formar a una concentración de 80 pmol de proteína Cas9: 120 pmol de NASC-PC1: 120 pmoles de NASC-PC2. Los componentes de ARNsg de control se pueden ensamblar individualmente en complejos de ribonucleoproteínas con la proteína Cas9 de una manera similar. Antes del ensamblaje con la proteína Cas9, NASC-PC1, NASC-PC2 y los ARNsg pueden diluirse hasta la concentración deseada (120 pmol) en un volumen final de 2 µL, incubarse durante 2 minutos a 95°C, retirarse del termociclador, y se dejan equilibrar a temperatura ambiente. La proteína Cas9 se puede diluir a una concentración apropiada en el tampón de unión (HEPES 20 mM, KCl 100 mM, MgCl₂ 5 mM y glicerol al 5% a pH 7,4) hasta un volumen final de 3 µL y se puede mezclar con los 2 µL de cada NASC-PC1, NASC-PC2, y los ARNsg seguidos de incubación a 37°C durante 30 minutos.

D. Transfecciones celulares utilizando los complejos de ribonucleoproteínas NASC/Cas9

Los complejos de ribonucleoproteínas pueden transfectarse en células HEK293 (ATCC, Manassas VA), utilizando el Shuttle System de 96 pozos Nucleofector® (Lonza, Allendale, NJ) siguiendo el protocolo del fabricante. Los complejos de ribonucleoproteínas se pueden dispensar en un volumen final de 5 µL en pozos individuales de una placa de 96 pozos, en el que los pozos contienen células HEK293 en medio de cultivo. El medio de cultivo celular se puede eliminar de los pozos de la placa y las células se pueden separar con la enzima TrypLE™ (Thermo Scientific, Wilmington, DE). Las células HEK293 suspendidas se pueden sedimentar por centrifugación durante 3 minutos a 200 x g, los reactivos TrypLE se pueden aspirar y las células se pueden lavar con solución salina tamponada con fosfato (PBS) libre de calcio y magnesio. Las células se pueden sedimentar por centrifugación durante 3 minutos a 200 X g del PBS aspirado y el sedimento celular se puede volver a suspender en 10 ml de PBS libre de calcio y magnesio.

Las células se pueden contar utilizando el Contador de células automatizado Countess® II (Life Technologies, Grand Island, NY). Se pueden transferir $2,2 \times 10^7$ células a un tubo de microcentrifuga de 1,5 ml y sedimentar. El PBS se puede aspirar y las células se pueden volver a suspender en la solución Nucleofector™ SF (Lonza, Allendale, NJ) a una densidad de 1×10^7 células/ml. Se pueden agregar 20 µl de la suspensión celular a cada pozo individual que contiene 5 µL de complejos de ribonucleoproteínas, y el volumen completo de cada pozo se puede transferir a un pozo de una placa de Nucleocuvette™ de 96 pozos (Lonza, Allendale, NJ). La placa se puede cargar en el Nucleofector™ de 96 pozos Shuttle™ (Lonza, Allendale, NJ) y las células se pueden nucleofectar utilizando el programa 96-CM-130 Nucleofector™ (Lonza, Allendale, NJ). Post-nucleofección, 70 µL de medio de Eagle manipulado por Dulbecco (DMEM; Thermo Scientific, Wilmington, DE) suplementado con un 10% de suero fetal bovino (FBS; Thermo Scientific, Wilmington, DE), penicilina y estreptomina (Life Technologies, Grand Island, NY) se pueden agregar a cada pozo y luego se pueden transferir 50 µl de la suspensión celular a una placa de cultivo celular de 96 pozos que contiene 150 µl de medio de cultivo completo DMEM precalentado. La placa puede luego transferirse a una incubadora de cultivo de tejidos y mantenerse a 37°C en CO₂ al 5% durante 48 horas.

E. Generación de secuencia diana de ADN de doble cadena para secuenciación profunda

El ADNg puede aislarse de las células HEK293 48 horas después de la transfección de los complejos de ribonucleoproteínas utilizando 50 µl de solución de extracción de ADN QuickExtract (Epicenter, Madison, WI) por pozo, seguido de incubación a 37°C durante 10 minutos, 65°C durante 6 minutos y 95°C durante 3 minutos para detener la reacción. El ADNg aislado se puede diluir con 50 µL de agua estéril y las muestras se pueden almacenar a -80°C.

Usando el ADNg aislado, se puede realizar una primera PCR usando la mezcla maestra 2X Hot Start High-Fidelity (New England Biolabs, Ipswich, MA) a una concentración de 1x, cebadores a 0,5 µM cada uno, 3,75 µl de ADNg en un volumen final de 10 µl y amplificado a 98°C durante 1 minuto, 35 ciclos de 10 segundos a 98°C, 20 segundos a 60°C, 30 segundos a 72°C y una extensión final a 72°C durante 2 minutos. Los cebadores pueden diseñarse para amplificar la región XRCC5_T1 (por ejemplo, SEQ ID NO. 47 y SEQ ID NO. 48) o XRCC5_T3 (SEQ ID NO. 49 y SEQ ID NO. 50). El ADNg preparado a partir de las muestras nucleofectadas NASC-PC1/NASC-PC2/Cas9 y las muestras nucleofectadas ARNsg/Cas9 se pueden amplificar con ambos pares de cebadores, por separado, para evaluar la edición de cada sitio objetivo por las ribonucleoproteínas. Cada reacción de PCR se puede diluir 1:100 en agua.

Se puede configurar un PCR de "código de barras" utilizando cebadores de índice únicos para cada muestra para facilitar la secuenciación múltiplex. Ejemplos de tales pares de cebadores se muestran en la Tabla 19.

Tabla 19

Cebadores de códigos de barras		
ID	Muestra	Cebador
CEBADOR DE CÓDIGO DE BARRAS conjunto-1	XRCC5_T1-ARNsg	SEQ ID NO. 51, 52
CEBADOR DE CÓDIGO DE BARRAS conjunto-2	XRCC5_T3-ARNsg	SEQ ID NO. 51,
CEBADOR DE CÓDIGO DE BARRAS conjunto-3	XRCC5-T1 NASC-PC1	SEQ ID NO. 51, 54
CEBADOR DE CÓDIGO DE BARRAS conjunto-4	XRCC5-T13 NASC-PC1	SEQ ID NO. 51, 55

La PCR de código de barras se puede realizar utilizando la mezcla maestra 2X Hot Start High-Fidelity (New England Biolabs, Ipswich, MA) a una concentración 1x, cebadores a 0,5 µM cada uno (Tabla 19), 1 µL de 1:100 primera PCR

diluida, en un volumen final de 10 µL y se puede amplificar a 98°C durante 1 minuto, 12 ciclos de 10 segundos a 98°C, 20 segundos a 60°C, 30 segundos a 72°C y una extensión final a 72°C para 2 minutos.

F. Depuración por SPRIselect

- 5 Todas las reacciones de PCR del código de barras se pueden agrupar y transferir a un solo tubo de microcentrífuga ("biblioteca de amplicones") para la depuración basada en perlas SPRIselect (Beckman Coulter, Pasadena, CA) de amplicones para secuenciación.

10 A cada tubo, se pueden agregar, mezclar e incubar a temperatura ambiente durante 10 minutos volúmenes de 0.9x de perlas SPRIselect. El tubo de microcentrífuga se puede colocar en el soporte del tubo magnético (Beckman Coulter, Pasadena, CA) hasta que la solución se aclare. El sobrenadante se puede eliminar y desechar, y las perlas residuales se pueden lavar con 1 volumen de etanol al 85% e incubar a temperatura ambiente durante 30 segundos. Después de la incubación, el etanol se puede aspirar y las perlas se pueden secar al aire a temperatura ambiente durante 10 minutos. Cada tubo de microcentrífuga se puede retirar del soporte magnético y se pueden agregar a las perlas volúmenes de 0,25x de tampón Qiagen EB (Qiagen, Venlo, Países Bajos), mezclar vigorosamente e incubar durante 2 minutos a temperatura ambiente. Cada tubo de microcentrífuga puede devolverse al imán, incubarse hasta 15 que la solución se haya aclarado, y luego el sobrenadante que contiene los amplicones purificados se puede dispensar en un tubo limpio de microcentrífuga. La biblioteca de amplicones purificados se puede cuantificar utilizando el sistema Nanodrop™ 2000 (Thermo Scientific, Wilmington, DE) y la calidad de la biblioteca se puede analizar utilizando el sistema Analizador de fragmentos™ (Advanced Analytical Technologies, Ames, IA) y el DNF-910 de doble cadena Kit de reactivos de ADN (Advanced Analytical Technologies, Ames, IA).

20 G. Montaje de secuenciación profunda

La biblioteca de amplicones agrupados puede normalizarse a una concentración de 4 nM calculada a partir de los valores cuantificados y el tamaño promedio de los amplicones. La biblioteca de amplicones se puede analizar en MiSeq Sequencer (Illumina, San Diego, CA) con MiSeq Reagent Kit v2 (Illumina, San Diego, CA) durante 300 ciclos con dos series de 151 pares de extremos pareados más dos lecturas de índice de ocho ciclos.

25 H. Análisis de datos de secuenciación profunda

Las identidades de los productos en los datos de secuenciación pueden determinarse en función de las secuencias de códigos de barras de índice adaptadas a los amplicones en la PCR de código de barras. Se puede usar un script computacional para procesar los datos de MiSeq que se ejecutan, por ejemplo, las siguientes tareas:• Las lecturas se pueden alinear con el genoma humano (compilación GRCh38/38) usando Bowtie (bowtie-bio.sourceforge.net/index.shtml) para ftware.

- Las lecturas alineadas se pueden comparar con la región del locus de tipo salvaje esperada (por ejemplo, XRCC5_T1 o XRCC5_T3)
 - La secuencia de locus y las lecturas que no se alinean con ninguna parte del locus de destino se pueden descartar.
 - Las lecturas coincidentes de las locus de destino de tipo salvaje pueden ser contabilizadas.
- 35 • Las lecturas con *inserción-eliminación* (inserción o eliminación de bases) se pueden clasificar por tipo de *inserción-eliminación* y se pueden contar.
- Las lecturas de *inserción-eliminación* totales se pueden dividir por la suma de las lecturas de tipo salvaje y las lecturas de inserción-eliminación para dar lecturas con un porcentaje de mutación.

40 Mediante la identificación de secuencias inserción-eliminación en las regiones seleccionadas por los complejos de ribonucleoproteínas de la proteína NASC-PC1/NASC-PC2/Cas9 y los complejos de ribonucleoproteínas de la proteína ARNsg/Cas9, se puede determinar el direccionamiento específico de la secuencia en una línea celular humana. La edición en muestras NASC-PC1/NASC-PC2 se puede comparar con las eficiencias de edición de los controles de ARNsg.

45 Siguiendo la guía de la presente especificación y los Ejemplos, la edición *en célula* de una secuencia genómica puede ser practicada por un experto en la técnica con otras proteínas Cas y sus afines a las composiciones de polinucleótidos NASC.

Ejemplo 7

Identificación y cribado de ARNcr

50 En este Ejemplo, se describe un procedimiento a través del cual se pueden identificar los ARNcr de especies que tienen un sistema CRISPR de Clase 2. El procedimiento presentado aquí está adaptado de Chylinski, K., et al., RNA Biology 10(5): 726-737 (2013). No todos los pasos siguientes son necesarios para la evaluación ni el orden de los pasos debe ser el presentado.

A. Identificar una especie que contiene un locus CRISPR de clase 2

Usando la herramienta básica de búsqueda de alineación local (BLAST, blast.ncbi.nlm.nih.gov/Blast.cgi), se puede realizar una búsqueda de los genomas de diversas especies para identificar las nucleasas de CAS CRISPR de clase 2, (por ejemplo, proteína Cas9, proteína Cpf1, proteínas similares a Cas9, proteínas similares a Cpf1, etc.). Los sistemas CRISPR de clase 2 exhiben una alta diversidad en la secuencia a través de las especies, sin embargo, los ortólogos de nucleasas CRISPR de Clase 2 tienen dominios conservados, por ejemplo, un dominio de endonucleasa HNH y/o un dominio RuvC/RNase H. Los resultados primarios de BLAST se pueden filtrar para los dominios identificados, las secuencias incompletas o truncadas se pueden descartar y se pueden identificar las especies que tienen ortólogos de nucleasa CRISPR de Clase 2. Si se puede identificar un ortólogo de nucleasa CRISPR de Clase 2 en una especie, se pueden sondear secuencias adyacentes a la secuencia codificadora del ortólogo de la proteína Cas (por ejemplo, una proteína Cas9 o una proteína Cpf1) para otras proteínas Cas y una matriz asociada repetida-espaciador para identificar todas las secuencias pertenecientes al locus CRISPR-Cas. Esto se puede hacer alineando con otros loci CRISPR de Clase 2 conocidos.

Una vez que se puede identificar la secuencia del locus CRISPR de Clase 2 para el ortólogo de la nucleasa para la especie, se puede usar la selección predictiva *in silico* para extraer la secuencia de ARNcr. La secuencia de ARNcr está contenida en la matriz repetida CRISPR y puede identificarse por sus secuencias repetidas distintivas interpuestas por secuencias espaciadoras extrañas.

B. Preparación de la biblioteca ARN-Seq

La matriz CRISPR putativa que contiene el ARNcr individual identificado *in silico* puede validarse adicionalmente utilizando la secuenciación de ARN (ARN-seq).

Las células de especies identificadas por comprender ARNcr putativo pueden obtenerse de un depósito comercial (por ejemplo, ATCC, Manassas, VA; German Collection of Microorganisms and Cell Cultures GmbH (DSMZ), Braunschweig, Alemania).

Las células se pueden cultivar hasta la fase media del tronco y se puede preparar el ARN total utilizando el reactivo Trizol (Sigma Aldrich, St. Louis, MO) y tratar con DNaseI (Fermentas, Vilnius, Lituania).

10 µg del ARN total se pueden tratar con el kit de eliminación de ARNr Ribo-cero (Illumina, San Diego, CA) y el ARN restante se purifica utilizando RNA Clean and Concentrators (Zymo Research, Irvine, CA).

Luego se puede preparar una biblioteca utilizando un kit de preparación de biblioteca de ARN TruSeq Small (Illumina, San Diego, CA) siguiendo las instrucciones del fabricante. Esto da como resultado que los ADNc tengan secuencias adaptadoras.

La biblioteca de ADNc resultante puede secuenciarse utilizando el secuenciador MiSeq (Illumina, San Diego, CA).

C. Procesamiento de datos de secuenciación

Las lecturas de secuenciación de la biblioteca de ADNc pueden procesarse, por ejemplo, utilizando el siguiente procedimiento.

Las secuencias del adaptador se pueden eliminar con Cutadapt 1.1 (pypi.python.org/pypi/cutadapt/1.1) y se pueden recortar aproximadamente 15 nt desde el extremo 3' de la lectura para mejorar la calidad de la lectura.

Las lecturas pueden alinearse con el genoma de las especies respectivas (es decir, a partir de las cuales se identificó el ARNcr putativo) utilizando Bowtie 2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>).

El archivo de alineación/mapa de secuencia (SAM), generado por Bowtie 2, se puede convertir en un archivo de alineación/mapa binario (BAM) utilizando SAMTools (samtools.sourceforge.net/) para los siguientes pasos de análisis de secuenciación.

El mapa de cobertura de lectura al locus o loci CRISPR, se puede calcular a partir del archivo BAM utilizando BedTools (bedtools.readthedocs.org/en/latest/).

El archivo BED, generado en el paso anterior, se puede cargar en el Integrative Genomics Viewer (IGV; www.broadinstitute.org/igv/) para visualizar la secuenciación de lectura de pila. Se pueden utilizar las acumulaciones de lectura para identificar los extremos 5' y 3' de la secuencia de ARNcr putativa transcrita.

Los datos de ARN-seq se pueden usar para validar que una secuencia de elemento ARNcr putativo se transcriba activamente *in vivo*. Los aciertos confirmados de la comparación de las cribas *in silico* y ARN-seq pueden validarse para la capacidad funcional de soportar la escisión por nucleasa CRISPR de Clase 2 de una secuencia de ácidos nucleicos diana de ADN de doble cadena usando los procedimientos descritos aquí (por ejemplo, los Ejemplos 2, 3 y 5). Se sabe en la técnica que los sistemas CRISPR Clase 2 Tipo V solo requieren un ARNcr para facilitar la escisión de la nucleasa Cpf1 de una secuencia diana de ADN de doble cadena, mientras que los sistemas CRISPR de Tipo II

de Clase 2 requieren un ARNcr y un TraARNcr para facilitar la escisión de la nucleasa Cas9 de una secuencia diana de ADN de doble cadena.

Siguiendo los lineamientos de la presente especificación y los Ejemplos, un experto en la técnica puede practicar la identificación de secuencias de ARNcr asociadas a proteínas Cas9.

5 Ejemplo 8

Identificación y cribado de TraARNcr

Este ejemplo ilustra un procedimiento por el cual se pueden identificar traARN-cr de especies que tienen, por ejemplo, un sistema de CRISPR-Cas9 Clase 2 Tipo II. Esto se adaptó de Chylinski, K., et al., RNA Biology 10(5): 726-737 (2013). No todos los pasos siguientes son necesarios para la evaluación ni el orden de los pasos debe ser el presentado.

A. Identificar una especie que contenga un sistema CRISPR-Cas9 Tipo II

Usando la herramienta básica de búsqueda de alineación local (BLAST, blast.ncbi.nlm.nih.gov/Blast.cgi), se puede realizar una búsqueda de los genomas de varias especies para identificar una proteína Cas9. Los sistemas CRISPR-Cas9 Clase 2 Tipo II exhiben una alta diversidad en la secuencia a través de las especies, sin embargo, los ortólogos de Cas9 exhiben arquitecturas de dominio conservadas de un dominio de endonucleasa HNH central y un dominio de RuvC/ARNasa dividido. Los resultados primarios de BLAST se pueden filtrar para los dominios identificados; Las secuencias incompletas o truncadas se pueden descartar y los ortólogos de Cas9 se pueden identificar.

Si se puede identificar un ortólogo de Cas9 en una especie, las secuencias adyacentes a la secuencia codificadora del ortólogo de Cas9 se pueden sondear para detectar otras proteínas Cas y una matriz repetida-espaciador asociada a Cas para identificar todas las secuencias que pertenecen al locus CRISPR-Cas9. Esto se puede hacer mediante la alineación con otros loci CRISPR-Cas9 de Clase 2 Tipo II conocidos, con el conocimiento de que las especies estrechamente relacionadas exhiben una arquitectura de locus CRISPR-Cas9 similar (por ejemplo, composición de proteína Cas, tamaño, orientación, ubicación de la matriz, ubicación de traARN-cr, etc.). El elemento traARN-cr está típicamente contenido dentro del locus CRISPR-Cas9 de Clase 2 Tipo II y puede identificarse fácilmente por su complementariedad de secuencia con los elementos de repeticiones en la matriz repetida espaciadora. Las secuencias de tracr complementarias a los elementos de repeticiones se denominan secuencias de antirrepetición de tracr.

Una vez que se identifica la secuencia del locus CRISPR-Cas9 correspondiente al ortólogo de Cas9 para una especie, se puede usar la selección predictiva *in silico* para extraer la secuencia antirrepetición tracr para identificar el traARN-cr asociado a Cas. Las supuestas repeticiones se pueden detectar, por ejemplo, de la siguiente manera.

Si la secuencia repetida es de una especie conocida, la secuencia repetida se puede identificar y recuperar de la base de datos CRISPRdb (crispr.u-psud.fr/crispr/). Si la secuencia repetida no es de una especie conocida, la secuencia repetida se puede predecir empleando el software CRISPRfinder (crispr.u-psud.fr/Server/) utilizando el locus CRISPR-Cas9 de Clase 2 Tipo II para la especie como se describe anteriormente.

La secuencia repetida identificada para la especie puede usarse para sondear el locus CRISPR-Cas9 para la secuencia repetida (por ejemplo, usando el algoritmo BLASTp o similar). La búsqueda está típicamente restringida a regiones intergénicas del locus CRISPR-Cas9.

Se puede validar una región antirrepetición tracr identificada para complementar la secuencia repetida identificada.

Se puede sondear una región antirrepetida putativa en las regiones 5' y 3' de la región antirrepetida putativa para la presencia de un terminador transcripcional independiente de Rho (TransTerm HP, transterm.cbcb.umd.edu/).

Combinando la secuencia identificada que comprende el elemento antirrepetición y el terminador transcripcional independiente de Rho, se puede determinar que la secuencia es el traARN-cr putativo de la especie dada.

B. Preparación de la biblioteca ARN-Seq

El traARN-cr putativo identificado *in silico* se puede validar adicionalmente utilizando la secuenciación de ARN (ARN-seq). Las células de especies que comprenden el traARN-cr putativo pueden obtenerse de un repositorio comercial (por ejemplo, ATCC, Manassas VA; DSMZ, Braunschweig, Alemania).

Las células se pueden cultivar hasta la fase media del tronco y el ARN total se puede preparar utilizando el reactivo Trizol (SigmaAldrich, St. Louis, MO) y se pueden tratar con DNaseI (Fermentas, Vilnius, Lituania).

10 µg del ARN total pueden tratarse utilizando un kit de eliminación de ARNr Ribo-cero (Illumina, San Diego, CA) y el ARN restante se purifica utilizando ARN limpio y concentradores (Zymo Research, Irvine, CA).

Se puede preparar una biblioteca utilizando un kit de preparación de biblioteca de ARN pequeño TruSeq (Illumina, San Diego, CA) siguiendo las instrucciones del fabricante. Esto da como resultado que los ADNc tengan secuencias adaptadoras.

La biblioteca de ADNc resultante se puede secuenciar utilizando un secuenciador MiSeq (Illumina, San Diego, CA).

5 C. Procesamiento de datos de secuenciación

Las lecturas de secuenciación de la biblioteca de ADNc pueden procesarse, por ejemplo, usando el siguiente procedimiento.

Las secuencias adaptadoras pueden eliminarse utilizando cutadapt 1.1 (pypi.python.org/pypi/cutadapt/1.1) y se pueden recortar aproximadamente 15 nt desde el extremo 3' de la lectura para mejorar la calidad de lectura.

10 Las lecturas se pueden alinear con el genoma de las especies respectivas (es decir, a partir de las cuales se identificó el ARNcr putativo) utilizando Bowtie 2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). El archivo de alineación/mapa de secuencia (SAM) generado por Bowtie 2 se puede convertir en un archivo de alineación/mapa binario (BAM) utilizando SAMTools (<http://samtools.sourceforge.net/>) para los siguientes pasos de análisis de secuenciación.

15 El mapa de cobertura de lectura para el locus o loci CRISPR se puede calcular a partir del archivo BAM utilizando BedTools (bedtools.readthedocs.org/es/latest/).

El archivo BED, generado en el paso anterior, se puede cargar en el Integrative Genomics Viewer (IGV; www.broadinstitute.org/igv/) para visualizar la secuenciación de lectura de pila. Se pueden utilizar las acumulaciones de lectura para identificar los extremos 5' y 3' de la secuencia de traARN-cr putativa transcrita.

20 Los datos de ARN-seq se pueden usar para validar que una secuencia de elemento traARN-cr putativa se transcribe activamente *in vivo*. Los aciertos confirmados de la comparación de las cribas *in silico* y ARN-seq pueden validarse para la capacidad funcional de la secuencia de traARN-cr identificada y su ARNcr análogo para apoyar la escisión mediada por Cas9 de una secuencia diana de ADN de doble cadena utilizando los procedimientos descritos aquí (por ejemplo, Ejemplos 2, 3 y 5).

25 Siguiendo la guía de la presente especificación y los Ejemplos, la identificación de las secuencias de traARN-cr relacionadas con las proteínas Cas9 puede ser realizada por un experto en la técnica.

Ejemplo 9

Ensayo T7E1 para la detección de modificaciones de secuencia diana en células eucarióticas

30 Este ejemplo ilustra el uso de ensayos de T7E1 para evaluar y comparar el porcentaje de escisión *in vivo* de complejos de proteínas NASC/Cas9 (por ejemplo, complejos de proteínas NASC-PC1/NASC-PC2/Cas9) en relación con secuencias diana de ADN de doble cadena seleccionadas.

A. Transfecciones celulares que utilizan componentes de polinucleótidos de Cas

35 NASC-PC1 y NASC-PC2 pueden transfectarse en células HEK293 que expresan de manera constitutiva Cas9 de *S. pyogenes* (HEK293-Cas9), utilizando el Sistema de Transferencia de 96 pozos Nucleofector® (Lonza, Allendale, NJ) y el siguiente protocolo. NASC-PC1 y NASC-PC2 pueden diluirse individualmente a la concentración apropiada (por ejemplo, 120 pmol), mezclarse, incubarse durante 2 minutos a 95°C, retirarse del termociclador, dejar que se equilibre a temperatura ambiente y dispensar en un volumen final de 5 µL en una placa de 96 pozos. El medio de cultivo se puede aspirar de las células HEK293-Cas9, las células se pueden lavar una vez con PBS libre de calcio y magnesio, y se pueden tripsinizar mediante la adición de TrypLE (Life Technologies, Grand Island, NY) seguido de

40 una incubación a 37°C durante 3-5 minutos. Las células tripsinizadas pueden pipetearse suavemente hacia arriba y hacia abajo para formar una suspensión unicelular y agregarse al medio de cultivo completo DMEM compuesto por medio de cultivo DMEM (Life Technologies, Grand Island, NY) que contiene 10% de suero fetal bovino (FBS; Thermo Scientific, Wilmington, DE) y suplementado con penicilina y estreptomycin (Life Technologies, Grand Island, NY).

45 Las células se pueden sedimentar mediante centrifugación durante 3 minutos a 200 X g, el medio de cultivo se puede aspirar y las células se pueden volver a suspender en PBS. Las células se pueden contar utilizando el contador de células automatizado Countess® II (Life Technologies, Grand Island, NY). Se pueden transferir $2,2 \times 10^7$ células a un tubo de microcentrifuga de 1,5 ml y sedimentar. El PBS se puede aspirar y las células se pueden volver a suspender en la solución Nucleofector™ SF (Lonza, Allendale, NJ) a una densidad de 1×10^7 células/ml. Se pueden agregar 20 µL de la suspensión celular a los pozos individuales que contienen 5 µL de NASC-PC1/NASC-PC2 y se puede transferir todo el volumen a los pozos de una placa Nucleocuvette™ de 96 pozos (Lonza, Allendale, NJ). La placa se puede cargar en el Nucleofector™ de 96 pozos Shuttle™ (Lonza, Allendale, NJ) y las células se pueden nucleofectar utilizando el programa 96-CM-130 Nucleofector™ (Lonza, Allendale, NJ). Después de la nucleofección, se pueden agregar 70 µl de medio de cultivo completo DMEM a cada pozo, y 50 µl de la suspensión celular se pueden transferir a una placa de cultivo celular de 96 pozos recubierta con colágeno que contiene 150 µl

50

de medio de cultivo completo DMEM precalentado. La placa se puede transferir a una incubadora de cultivo de tejidos y se mantiene a 37°C en CO₂ al 5% durante 48 horas.

B. Generación de secuencia diana de ADN de doble cadena para el ensayo T7E1E

ADNg se puede aislar de células HEK293-Cas9 48 horas después de la transfección con NASC-PC1/NASC-PC2 utilizando 50 µl de solución de extracción de ADN QuickExtract (Epicenter, Madison, WI) por pozo, seguido de incubación a 37°C durante 10 minutos, 65°C durante 6 minutos y 95°C durante 3 minutos para detener la reacción. El ADNg puede diluirse con 150 µl de agua y las muestras se pueden almacenar a -80°C.

El ADN para T7E1 puede generarse mediante amplificación por PCR de secuencias diana de ADN de doble cadena (por ejemplo, XRCC5_T1 y XRCC5_T3) a partir de ADNg aislado. Las reacciones de PCR se pueden configurar utilizando 8 µL de ADNg como plantilla con la polimerasa KAPA HiFi Hot Start y contienen 0,5U de polimerasa, 1x tampón de reacción, dNTP 0,4 mM y cebadores de avance y reversos 300 nM dirigidos a una de las secuencias diana de ADN de doble cadena (por ejemplo, SEQ ID NO. 47/SEQ ID NO. 48 y SEQ ID NO. 49/SEQ ID NO. 50) en un volumen total de 25 µL. Las secuencias diana de ADN pueden amplificarse utilizando las siguientes condiciones: 95°C durante 5 minutos, 4 ciclos de 20 segundos a 98°C, 20 segundos a 70°C, menos 2°C/ciclo, 30 segundos a 72°C, seguido por 30 ciclos de 15 segundos a 98°C, 20 segundos a 62°C, 20 segundos a 72°C y una extensión final a 72°C durante 1 minuto.

C. Ensayo T7E1

Las secuencias diana de ADN de doble cadena amplificadas por PCR para los ensayos de T7E1 se pueden desnaturalizar a 95°C durante 10 minutos y luego se les permite volver a fusionar enfriando a 25°C a -5°C/segundo en un ciclador térmico. El ADN recocido se puede incubar con 0,5 µL de T7 Endonucleasa I en 1x tampón NEBuffer 2 (New England Biolabs, Ipswich, MA) en un volumen total de 15 µL durante 25 minutos a 37°C. Las reacciones de T7E1 se pueden analizar utilizando el sistema Fragment Analyzer™ (Advanced Analytical Technologies, Ames, IA) y el kit de reactivos de ADN de doble cadena DNF-910 (Advanced Analytical Technologies, Ames, IA). El sistema Fragment Analyzer™ proporciona la concentración de cada fragmento de escisión y de la secuencia diana de ADN de doble cadena que permanece después de la escisión.

Los porcentajes de escisión de las secuencias diana de ADN de doble cadena se pueden calcular a partir de la concentración de cada fragmento de escisión y la secuencia diana de ADN de doble cadena que permanece después de que se haya producido la escisión, utilizando la siguiente fórmula:

$$\% \text{ de escisión} = \left(1 - \sqrt[3]{\frac{(frag1 + frag2)}{(frag1 + frag2 + progenitora)}} \right)$$

ECUACIÓN 1

En la Ecuación 1, las concentraciones de frag1 y frag2 corresponden a la concentración de fragmentos de escisión de Cas9 de la secuencia diana de ADN de doble cadena y la progenitora corresponde a la secuencia diana de ADN de doble hebra que permanece después de que haya tenido lugar la escisión.

El ensayo T7E1 para la detección de modificaciones de secuencia diana en células eucarióticas proporciona datos para demostrar que las composiciones de polinucleótido NASC descritas en el presente documento facilitan la escisión *in vivo* específica de sitio mediada por Cas9 de múltiples secuencias diana de ADN de doble cadena. Los polinucleótidos ARNsg, ARNcr y/o ARNcr/traARN-cr que tienen la misma secuencia de unión a diana de ADN que la composición de polinucleótido NASC también pueden incluirse en el ensayo para comparar los porcentajes de escisión específicos del sitio mediados por Cas entre los constructos.

Siguiendo la guía de la presente especificación y los Ejemplos, el ensayo T7E1 descrito en este Ejemplo puede ser practicado por un experto en la técnica con otras proteínas Cas y sus composiciones de polinucleótidos NASC afines.

Ejemplo 10

Sondeo de sitios tolerantes a la modificación en el esqueleto de ARN guía de clase 2 tipo V Cpf1

Este ejemplo describe la generación y prueba de diversas modificaciones de los ARNcr de guía Clase 2 Tipo V y su idoneidad para su uso en la construcción de componentes de polinucleótidos NASC. El procedimiento descrito a continuación está adaptado de Briner, A., et al., Molecular Cell 56(2): 333-339 (2014). No todos los pasos siguientes son necesarios para la evaluación ni el orden de los pasos debe ser el presentado.

En este Ejemplo, pueden introducirse modificaciones en el esqueleto de ARNcr, y el ARNcr manipulado se puede probar con una nucleasa Cpf1 afin para facilitar la identificación de regiones o posiciones en el esqueleto de ARNcr de cpf1-cr, en el que se pueden diseñar los enlaces para los componentes polinucleotídicos de NASC.

Se puede seleccionar un ARNcr de un sistema CRISPR Tipo V de Clase 2 (por ejemplo, *Acidaminococcus* sp. Cpf1) para modificación. La secuencia de ARNcr puede modificarse *in silico* para introducir una o más sustituciones, eliminaciones o inserciones de bases en secuencias de ácidos nucleicos en regiones seleccionadas de una o más de las siguientes regiones: secuencias de ácidos nucleicos 5' del pseudonudo, la secuencia 1 de ARN Cpf1-stem, el bucle de pseudonudo (secuencia de ácidos nucleicos del elemento de bucle), la secuencia 1C de ARN de Cpf1-stem, o el elemento espaciador.

La secuencia de ARNcr puede modificarse *in silico* para introducir una o más rupturas en el esqueleto de fosfodiéster en una o más regiones seleccionadas de entre las siguientes: secuencias de ácidos nucleicos 5' de la secuencia 1 de pseudonudo, Cpf1-tall ARN, el bucle de pseudonudo (secuencia de ácidos nucleicos del elemento de bucle), secuencia 1C de ARN de Cpf1-stem, o el elemento espaciador.

La modificación de la base también se puede usar para introducir no coincidencias en las interacciones del par de bases a través de hidrógeno de cualquiera de las regiones ARNcr, o la mutación del par de bases que introduce una interacción alternativa del par de bases a través de hidrógeno mediante la sustitución de dos bases, en el que la base de hidrógeno alternativa La interacción entre pares difiere de la interacción original entre el par de bases a través de hidrógeno (por ejemplo, la interacción original entre el par de bases a través de hidrógeno es el emparejamiento de bases de Watson-Crick y la sustitución de las dos bases forman un emparejamiento de bases de Hoogsteen inverso). La sustitución de bases también se puede usar para introducir la interacción del par de bases a través de hidrógeno dentro del esqueleto de ARNcr (por ejemplo, dentro de la secuencia de bucle de pseudonudo).

Las regiones del ARNcr pueden diseñarse independientemente para introducir elementos de estructura secundaria en el esqueleto de ARNcr. Dichos elementos de estructura secundaria incluyen, entre otros, los siguientes: elementos de tallo-bucle, elementos de tallo, pseudonudos y ribozimas. Además, el esqueleto de ARN guía de ARNcr puede modificarse para eliminar partes del esqueleto de ARNcr mediante eliminación en el extremo 5', extremo 3' o interno al ARNcr. También se pueden introducir estructuras de esqueleto alternativas.

Se pueden proporcionar secuencias de ARNcr diseñadas *in silico* a un fabricante comercial para la síntesis

Los ARNcr manipulados pueden evaluarse por su capacidad para soportar la escisión de una secuencia diana de ADN de doble cadena mediada por la proteína Cpf1 relacionada al ARNcr que dio lugar al ARNcr manipulado. La amplificación de secuencias diana de ADN de doble cadena y el ensayo de escisión bioquímica se pueden llevar a cabo de una manera similar a las descritas en el Ejemplo 4 y el Ejemplo 5, respectivamente. El ARNcr manipulado que es capaz de mediar la escisión de una secuencia diana de ADN con sus proteínas Cpf1 afines puede validarse en cuanto a la actividad en células usando el procedimiento descrito en el Ejemplo 6.

Siguiendo la guía de la presente especificación y Ejemplos, la modificación de un ARNcr de Cpf1 (por ejemplo, la introducción o eliminación de diversas secuencias, y/o la introducción o eliminación de modificaciones estructurales secundarias) se puede usar para sondear ubicaciones para la inserción o enlaces para facilitar la realización de composiciones polinucleótidas de NASC. Este ejemplo puede ser practicado por un experto en la técnica con otras proteínas Cpf1 CRISPR de Tipo V y otros ARNcr CRISPR de Tipo V en vista de las enseñanzas de la presente especificación.

Ejemplo 11

Sondeo de sitios tolerantes a la modificación en el esqueleto de ARN de guía Cas9 de clase 2 tipo II

Este ejemplo describe la generación y el ensayo de diversas modificaciones de los ARN guía de Tipo II de Clase 2 y su idoneidad para su uso en la construcción de composiciones de polinucleótidos NASC.

En este ejemplo, pueden introducirse modificaciones en el esqueleto de ARN del (los) ARN(s) de guía CRISPR de Clase 2 Tipo II (por ejemplo, ARN de guía dual o ARN de guía única) para identificar ubicaciones para la manipulación o la unión de diversas secuencias de ácidos nucleicos. El procedimiento descrito a continuación está adaptado de Briner, A., et al., *Molecular Cell* 56(2): 333-339 (2014). No todos los pasos siguientes son necesarios para la evaluación ni el orden de los pasos debe ser el presentado.

Se puede seleccionar ARNsg, ARNcr, traARN-cr o ARNcr y ARNcr CRISPR de tipo 2 de Tipo II (denominado colectivamente un "ARN de guía Cas9") para manipulación.

La secuencia de ARN guía de Cas9 se puede modificar *in silico* para introducir una o más sustituciones, eliminaciones o inserciones de bases en regiones seleccionadas de una o más de las siguientes: una secuencia de unión a un ácido nucleico, una secuencia de ácidos nucleicos de tallo inferior, una secuencia de ácidos nucleicos de protuberancia, una secuencia de ácidos nucleicos del tallo superior, una primera secuencia de ácidos nucleicos del elemento de tallo-bucle, una secuencia de ácidos nucleicos nexa, una secuencia de ácidos nucleicos de enlace y/u horquillas 3'. La secuencia de ARN guía de Cas9 se puede modificar *in silico* para introducir una o más rupturas en el esqueleto de fosfodiéster en una o más regiones seleccionadas de entre las siguientes: una secuencia de unión a diana de ácidos nucleicos, una secuencia de ácidos nucleicos de tallo inferior, una secuencia de ácidos nucleicos de protuberancia, una secuencia de ácidos nucleicos del tallo superior, una primera secuencia de ácidos nucleicos del

elemento de tallo-bucle, una secuencia de ácidos nucleicos nexos, una secuencia de ácidos nucleicos de enlace y horquillas 3'.

La modificación de la base se puede usar para introducir no coincidencias en las interacciones del par de bases a través de hidrógeno de cualquiera de las regiones de ARN guía de Cas9. La mutación del par de bases se puede usar para introducir una interacción alternativa del par de bases a través de hidrógeno mediante la sustitución de dos bases, en el que la interacción del par de bases a través de hidrógeno alternativa difiere de la interacción del par de bases a través de hidrógeno original (por ejemplo, la interacción original entre el par de bases a través de hidrógeno es el emparejamiento de bases de Watson-Crick y la sustitución de las dos bases forman un emparejamiento de bases de Hoogsteen inverso). La sustitución de bases también se puede usar para introducir la interacción del par de bases a través de hidrógeno dentro del esqueleto de ARN guía Cas9 (por ejemplo, dentro de la secuencia de protuberancia). Las regiones del ARN guía de Cas9 pueden diseñarse de manera independiente para introducir elementos de estructura secundaria en el esqueleto de ARN guía de Cas9. Dichos elementos de estructura secundaria incluyen, entre otros, los siguientes: elementos de tallo-bucle, elementos de tallo, pseudonudos y ribozimas. Además, el esqueleto del ARN guía de Cas9 se puede modificar para eliminar partes del esqueleto del ARN guía de Cas9 mediante la eliminación en el extremo 5', el extremo 3' y/o interno al ARN guía de Cas9. También se pueden introducir estructuras de esqueleto alternativas.

Se pueden proporcionar secuencias de ARN guía CRISPR Cas9 de Clase 2 Tipo II diseñadas *in silico* a un fabricante comercial para la síntesis.

Los ARN de guía CRISPR Cas9 Clase II Tipo II manipulados pueden evaluarse por su capacidad para soportar la escisión de una secuencia diana de ADN de doble cadena mediada por la proteína Cas9 relacionada al ARN guía de Cas9 que dio lugar al ARN guía de Cas9 manipulado. La amplificación de una secuencia diana de ADN de doble cadena y el ensayo de escisión bioquímica se pueden llevar a cabo de una manera similar a las descritas en el Ejemplo 4 y el Ejemplo 5, respectivamente. Los ARN de guía de Cas9 manipulados capaces de mediar la escisión de una secuencia diana de ADN con sus proteínas Cas9 afines pueden validarse para la actividad en células usando el procedimiento descrito en el Ejemplo 6.

Siguiendo la guía de la presente especificación y Ejemplos, la modificación de uno o más ARN de la guía de Cas9 (por ejemplo, introducción o eliminación de varias secuencias, y/o introducción o eliminación de modificaciones estructurales secundarias) se puede utilizar para probar ubicaciones de inserción o enlaces para facilitar la composición de polinucleótidos NASC. Este ejemplo puede ser practicado por un experto en la técnica con otras proteínas CRISPR Cas9 de Tipo II y otros ARN de guía CRISPR Cas9 de Tipo II en vista de las enseñanzas de la presente especificación.

Ejemplo 12

Detección de composiciones de polinucleótidos NASC que comprenden secuencias de unión a diana de ADN

Este ejemplo ilustra el uso de composiciones de polinucleótidos de NASC de la presente invención para modificar secuencias diana de ADN presentes en ADN humano y medir el nivel de actividad de escisión en esos sitios.

Los sitios diana (secuencias diana de ADN) pueden seleccionarse primero a partir de ADN humano. Los componentes individuales de las composiciones de polinucleótidos NASC pueden diseñarse para dirigirse a las secuencias seleccionadas. Se pueden realizar ensayos (por ejemplo, como se describe en el Ejemplo 5) para determinar el nivel de escisión de la secuencia diana del ADN.

No todos los pasos siguientes se requieren para cada selección, ni el orden de los pasos debe ser como se presenta, y la selección se puede acoplar a otros experimentos, o formar parte de un experimento mayor.

A. Selección de regiones diana de ADN (secuencias diana de ADN) de ADN humano

Las secuencias PAM (es decir, NGG, TTN, etc.) para una proteína Cas (por ejemplo, Cas9 de *S. pyogenes* o *Acidaminococcus* sp. Cpf1) pueden identificarse dentro de la región genómica seleccionada.

Se pueden identificar y seleccionar una o más secuencias diana de ADN de Cas9 (20 nucleótidos de longitud) que son adyacentes en 5' a una secuencia de PAM o una o más secuencias diana de ADN de Cpf1 (de 20 a 24 nucleótidos de longitud) que son de 3' adyacente a una secuencia PAM.

Los criterios para la selección de secuencias diana de ácidos nucleicos pueden incluir, entre otros, los siguientes: homología con otras regiones en el genoma, porcentaje de contenido de G-C, temperatura de fusión, presencia de homopolímero dentro del espaciador, distancia entre las dos secuencias y otros criterios conocidos por los expertos en la técnica.

Si se desea usar una composición de polinucleótido NASC CRISPR de Tipo II, la secuencia de unión a la diana del ADN puede incorporarse en el extremo 5'. Si se desea usar una composición de polinucleótido NASC CRISPR de Tipo V, la secuencia de unión a diana del ADN puede incorporarse en el extremo 3'. Un fabricante comercial

típicamente sintetiza composiciones de polinucleótidos NASC basadas en secuencias proporcionadas. Alternativamente, las composiciones de polinucleótidos de NASC pueden producirse como se describe en el Ejemplo 2 mediante transcripción *in vitro*.

Las composiciones de polinucleótidos de NASC como se describen en el presente documento se pueden usar con nucleasa CRISPR de Tipo II Clase 2 (por ejemplo, una nucleasa Cas9), una nucleasa CRISPR de Tipo V Clase 2 (por ejemplo, una nucleasa Cpf1), o ambas o nucleasa cognada Tipo II Clase 2 CRISPR y una nucleasa CRISPR de Tipo 2 Clase V para formar complejos de proteína NASC/Cas.

B. Determinación de porcentajes de escisión y especificidad

Los porcentajes y la especificidad de la escisión *in vitro* (por ejemplo, la cantidad de unión fuera de la diana) relacionados con las composiciones de polinucleótidos de NASC se pueden determinar, por ejemplo, utilizando los ensayos de escisión descritos en el Ejemplo 5 y se pueden comparar de la siguiente manera:

(1) Si solo se puede identificar o seleccionar un solo par de secuencias diana de ADN para un NASC, se puede determinar el porcentaje de escisión y la especificidad para cada una de las secuencias diana de ADN. Si así se desea, el porcentaje de escisión y/o la especificidad se pueden alterar en experimentos adicionales utilizando procedimientos que incluyen, entre otros, la modificación del NASC; o la introducción de proteínas efectoras/secuencias de unión a la proteína efectora para modificar el NASC, un componente polinucleótido NASC, o la proteína Cas; o introducir unidades estructurales de unión ligando/ligando para modificar el polinucleótido NASC o la proteína Cas.

(2) Si pueden identificarse o seleccionarse múltiples pares de secuencias diana de ADN para un NASC, los datos de porcentaje de escisión y los datos de especificidad de sitio obtenidos de los ensayos de escisión se pueden comparar entre diferentes ADN que comprenden la secuencia de unión de objetivo para identificar las secuencias diana de ADN que tienen el porcentaje de escisión deseado y la especificidad. Los datos de porcentaje de escisión y los datos de especificidad proporcionan criterios en los cuales basar las elecciones para una variedad de aplicaciones. Por ejemplo, en algunas situaciones, la actividad de la composición del polinucleótido NASC puede ser el factor más importante. En otras situaciones, la especificidad del sitio de escisión puede ser relativamente más importante que el porcentaje de escisión. Si así se desea, el porcentaje de escisión y/o la especificidad se pueden alterar en experimentos adicionales utilizando procedimientos que incluyen, entre otros, la modificación del NASC; o la introducción de proteínas efectoras/secuencias de unión a la proteína efectora para modificar el NASC, un componente polinucleótido NASC, o la proteína Cas; o introduciendo unidades estructurales de unión ligando/ligando para modificar el componente polinucleotídico de NASC o la proteína Cas.

Alternativamente, o además del análisis *in vitro*, los porcentajes de división celular y las especificidades asociadas con las composiciones de polinucleótidos de NASC se pueden obtener utilizando, por ejemplo, el procedimiento descrito en el Ejemplo 6, que se puede comparar de la siguiente manera:

(1) Si solo se puede identificar o seleccionar un solo par de secuencias diana de ADN para un NASC, se puede determinar el porcentaje de escisión y la especificidad para cada una de las secuencias diana de ADN. Si así se desea, el porcentaje de escisión y/o la especificidad se pueden alterar en experimentos adicionales utilizando procedimientos que incluyen, entre otros, la modificación del NASC; o la introducción de proteínas efectoras/secuencias de unión a la proteína efectora para modificar el NASC, un componente polinucleótido NASC, o la proteína Cas; o introduciendo unidades estructurales de unión ligando/ligando para modificar el componente polinucleotídico de NASC o la proteína Cas.

(2) Si pueden identificarse o seleccionarse múltiples pares de secuencias diana de ADN para un NASC, los datos de porcentaje de escisión y los datos de especificidad de sitio obtenidos de los ensayos de escisión se pueden comparar entre diferentes ADN que comprenden la secuencia de unión de objetivo para identificar las secuencias diana de ADN que tienen el porcentaje de escisión deseado y la especificidad. Los datos de porcentaje de escisión y los datos de especificidad proporcionan criterios en los que basar las elecciones para una variedad de aplicaciones. Por ejemplo, en algunas situaciones, la actividad de la composición del polinucleótido NASC puede ser el factor más importante. En otras situaciones, la especificidad del sitio de escisión puede ser relativamente más importante que el porcentaje de escisión. Si así se desea, el porcentaje de escisión y/o la especificidad se pueden alterar en experimentos adicionales utilizando procedimientos que incluyen, entre otros, la modificación del NASC; o la introducción de proteínas efectoras/secuencias de unión a la proteína efectora para modificar el NASC, un componente polinucleótido NASC, o la proteína Cas; o introduciendo unidades estructurales de unión ligando/ligando para modificar el componente polinucleotídico de NASC o la proteína Cas.

Siguiendo los lineamientos de la presente especificación y Ejemplos, un experto en la técnica puede practicar la criba descrita en este Ejemplo con otras composiciones de polinucleótidos de NASC para uso con proteínas relacionadas CRISPR de Clase II Tipo II, relacionadas con Proteínas CRISPR Cpf1 Clase 2 de Tipo V, o ambas proteínas CRISPR Cas9 de Clase 2 Tipo V análogas afines y proteínas CRISPR Cpf1 de Clase 2 análogas afines.

Ejemplo 13Manipulación de complejos de jaula cerrada de ribonucleoproteína que comprenden composiciones de polinucleótidos NASC

Este ejemplo ilustra el uso de composiciones de polinucleótidos NASC de la presente invención para la formación de complejos de jaula cerrada NASC-CC para empaquetar moléculas pequeñas.

Se puede diseñar una NASC-CC, por ejemplo, utilizando una primera composición de polinucleótidos de NASC y una segunda composición de polinucleótidos de NASC, cada una con la estructura general mostrada en la Figura 6H (Figura 6H, I, NASC-PC1; Figura 6H, II, NASC-PC2; y Figura 6H, III, NASC-PC-3). La primera composición de polinucleótidos de NASC y la segunda composición de polinucleótidos de NASC se pueden usar en combinación con tres secuencias de ADN de doble cadena. Cada secuencia de ADN de doble cadena ("una secuencia de refuerzo de ADN de doble cadena") puede comprender dos secuencias diana de ADN únicas, en el que la primera secuencia de ADN es complementaria a una primera secuencia de unión a ácidos nucleicos de la primera composición de polinucleótidos de NASC, y la segunda secuencia diana de ADN es complementaria a una segunda secuencia de unión a ácidos nucleicos de la segunda composición de polinucleótido NASC.

NASC-CC y las proteínas Cas asociadas pueden usarse para crear complejos de jaula cerrada adecuados para el empaquetamiento de moléculas. El tamaño de la jaula puede variarse cambiando el diseño de los componentes de NASC-CC o uniendo secuencias diana de ADN de diferente longitud.

A. Diseño de componentes NASC-CC

Una primera composición de polinucleótido NASC (referida en este Ejemplo como un "NASC-triplex1") puede diseñarse incluyendo una NASC-PC1, una NASC-PC2 y una NASC-PC3, que son similares en estructura a los representados en la Figura 6A (referido en este Ejemplo como un "NASC-PC1-triplex1", un "NASC-PC2-triplex1" y un "NASC-PC3-triplex1"). Se puede agregar una primera secuencia diana de ADN de 20 nucleótidos al extremo 5' (véase, por ejemplo, la Figura 6A, 610-611) de cada uno de NASC-PC1-triplex1, NASC-PC2-triplex1 y NASC-PC3-triplex1. La secuencia diana de ADN típicamente se seleccionará para que no tenga homología limitada o no con las secuencias de ADN nativas en un organismo en el que se introducirá el NASC-CC (por ejemplo, ADN humano o ADN vegetal).

Una segunda composición de polinucleótido NASC (denominada en este Ejemplo como "NASC-triplex2") puede diseñarse y comprender una NASC-PC1-triplex2, una NASC-PC2-triplex2 y una NASC-PC3-triplex2, que son similares en estructura a las representadas en la Figura 6A. Se puede agregar una segunda secuencia diana de ADN de 20 nucleótidos al extremo 5' (véase, por ejemplo, la Figura 6A, 610-611) de cada uno de NASC-PC1-triplex2, NASC-PC2-triplex2 y NASC-PC3-triplex2. La secuencia diana de ADN típicamente se seleccionará para que no tenga homología limitada o no con las secuencias de ADN nativas en un organismo en el que se introducirá el NASC-CC (por ejemplo, ADN humano o ADN vegetal). Además, las secuencias diana de ADN de 20 nucleótidos deben ser distintas de (es decir, no complementarias) de las secuencias diana de ADN diseñadas en el triplex1 de NASC.

Los componentes ilustrativos de NASC-triplex1 y NASC-triplex2 se presentan en la Tabla 20. En la tabla, la columna "Secuencia diana" indica la secuencia diana de ADN de 20 pb que es complementaria a la secuencia de unión a ácidos nucleicos en el componente polinucleótido NASC correspondiente.

Tabla 20

Componentes NASC-triplex1 y NASC-triplex2				
NASC-triplex	Componente NASC	Secuencia diana	Secuencia*	SEQ ID NO.
NASC-triplex1	NASC-PC1-triplex1	1	AUCUUGUUGACACGAGGAAUG UUUUAGUCCCUAAUUAAAUU CUUGAAAUUGGUAUAUAAGGA GGGACUACAACAAAGAGUUUG CGGGACUCUGCGGGGUUACAA UCCCCUAAAACCGCUUUUAAAA UUCAAAUAAAUUUUGCUUU	SEQ ID NO. 57

(continuación)

Componentes NASC-triplex1 y NASC-triplex2				
NASC-triplex	NASC-triplex	NASC-triplex	NASC-triplex	NASC-triplex
NASC-triplex1	NASC-PC2-triplex1	1	<u>AUCUUGUUGACACGAGGAAUG</u> <u>UUGUAGUCCCUCCUUAUUAUACC</u> <u>AAGAAAAAGAAAUUUAAAACU</u> <u>GAACUCCAACAAGAGUUUGC</u> GGGACUCUGCGGGGUUACAAU CCCCUAAAACCGCUUUUAAAAU UCAAUAAAUUUUGCUUU	SEQ ID NO. 58
NASC-triplex1	NASC-PC3-triplex1	1	<u>AUCUUGUUGACACGAGGAAUG</u> <u>UUGGAGUUCAGUUUUAUUUU</u> <u>CUUGAAAAAGAAAUUUAAUUA</u> <u>GGGACUAAAACAAGAGUUUG</u> CGGGACUCUGCGGGGUUACAA UCCCCUAAAACCGCUUUUAAAA UUCAAAUAAAUUUUGCUUU	SEQ ID NO. 59
NASC-triplex2	NASC-PC1-triplex2	2	<u>CGAUUAUUACAGCAAGGUGG</u> <u>UUUUAGACCCCUCCAUUUC</u> <u>GCGAAAGCGUUUUGAGAGAGU</u> <u>GAACUACAACAAGAGUUUGC</u> GGGACUCUGCGGGGUUACAAU CCCCUAAAACCGCUUUUAAAAU UCAAUAAAUUUUGCUUU	SEQ ID NO. 60
NASC-triplex2	NASC-PC2-triplex2	2	<u>CGAUUAUUACAGCAAGGUGG</u> <u>UUGUAGUUCACUCUCUAAAA</u> <u>CGCGAAAAAGAAAUUUAAUAA</u> <u>GGAACUACAACAAGAGUUUG</u> CGGGACUCUGCGGGGUUACAA UCCCCUAAAACCGCUUUUAAAA UUCAAAUAAAUUUUGCUUU	SEQ ID NO. 61

(continuación)

Componentes NASC-triplex1 y NASC-triplex2				
NASC-triplex	NASC-triplex	NASC-triplex	NASC-triplex	NASC-triplex
NASC-triplex2	NASC-PC3-triplex2	2	<u>CGAUAAUACAGCAAGGUGG</u> <u>UUGUAGUCCUUAUAAAUUU</u> <u>CUUGAAAGCGAAAUGGAAGAG</u> <u>GGGUCUAAAACAAAGAGUUUG</u> <u>CGGGACUCUGCGGGGUUACAA</u> <u>UCCCCUAAAACCGCUUUUAAAA</u> <u>UUCAAAUAAAUUUUGCUUU</u>	SEQ ID NO. 62
*Las regiones de hibridación NASC-triplex están subrayadas				

Una secuencia de refuerzo de ADN de doble cadena se puede diseñar para incorporar, en la dirección 5' a 3', una secuencia aleatoria de 20 nucleótidos en el extremo 5', secuencia diana 1, la secuencia PAM de *C. jejuni* 5'-NNNACA-3' (donde "N" es cualquier nucleótido), 50 nucleótidos de secuencia aleatoria, el complemento inverso de la secuencia PAM de *C. jejuni*, el complemento inverso de la secuencia diana 2 y una secuencia aleatoria de 20 nucleótidos en el extremo 3'. La secuencia de refuerzo de ADN de doble cadena se podrá dirigir a las proteínas dCas9 de *C. jejuni* cuando se unan tanto a la NASC-PC1-triplex1 como a la NASC-PC1-TRP2 y llevarán las dos NASC una cerca de la otra. La secuencia de la secuencia de refuerzo de ADN de doble cadena se puede proporcionar a un fabricante comercial para la síntesis del ADN de doble cadena. Alternativamente, la secuencia de la secuencia de refuerzo de ADN de doble cadena se puede construir utilizando oligonucleótidos de ADN de cadena simple, similar a la construcción de la plantilla de ADN de doble cadena presentada en el Ejemplo 2.

Una secuencia ilustrativa para una secuencia de refuerzo de ADN de doble cadena se muestra en la Tabla 21.

Tabla 21

Secuencia de refuerzo de ADN de doble cadena	
Secuencia*	SEQ ID NO.
CGTCGCTATGATTTGCCTATATCTTGTTGACACGAGGAAT GTAACAACGAGTTCCGCTATTGGGATGGAGTTTAACTGT CGCAACTCTCATCGCAATGTCAGTCACCTTGCTGTATTAT ATCGCGCATGATAAAGTACGCCAT	SEQ ID NO. 63
*Las secuencias de destino y PAM están en negrita	

B. Manipulación y producción de la proteína dCas9 de *C. jejuni*

Una secuencia de aminoácidos Cas9 de *C. jejuni* (por ejemplo, *C. jejuni* NCTC 1168; SEQ ID NO. 103) puede mutarse a partir de un ácido aspártico en la posición del aminoácido 8 a una alanina (D8A) e histidina en la posición 559 a alanina (D8A/H559A) para generar una forma inactiva por nucleasa de la proteína Cas9 de *C. jejuni* (proteína dCas9 de *C. jejuni*; SEQ ID NO. 56). La proteína dCas9 de *C. jejuni* seguirá siendo capaz de unirse a un triplex1 NASC. Tres proteínas dCas9 de *C. jejuni* son capaces de unirse al NASC-triplex1 y dirigir al NASC-triplex1 para que se una a las secuencias diana complementarias de las secuencias de unión a los ácidos nucleicos. La proteína dCas9 de *C. jejuni* se puede etiquetar en el extremo C con dos secuencias de localización nuclear (NLS) y se puede expresar de forma recombinante en *E. coli*, y se puede purificar utilizando procedimientos cromatográficos.

C. Formación de NASC-CCs.

NASC-triplex1 puede formarse mezclando NASC-PC1-triplex1, NASC-PC2-triplex1 y NASC-PC3-triplex1 (Tabla 20) en igual concentración molar, incubando durante 2 minutos a 95°C, recocado por enfriamiento a 25°C a -0,5°C/segundo en un termociclador, y luego permitir que la mezcla se equilibre a temperatura ambiente. NASC-triplex2 puede formarse mezclando NASC-PC1-triplex2, NASC-PC2-triplex2 y NASC-PC3-triplex2 (Tabla 20) en una concentración molar igual, incubando durante 2 minutos a 95°C, fusionando por enfriamiento a 25°C a -0,5°C/segundo en un termociclador, y luego permitiendo que la mezcla se equilibre a temperatura ambiente.

Los complejos de jaula cerrada de ribonucleoproteína pueden formarse mezclando NASC-triplex1 en presencia de una concentración en exceso de la proteína dCas9 de *C. jejuni* en un tampón de unión (HEPES 20 mM, KCl 100 mM, MgCl₂ 5 mM y glicerol al 5% a pH 7,4) y se incubaba a 37°C durante 20 minutos. La secuencia de refuerzo de ADN de doble cadena se puede agregar a una concentración limitante a la mezcla que comprende el complejo de proteínas NASC-triplex1/dCas9, y se incubaba a 37°C durante 20 minutos. NASC-triplex2 se puede agregar a la mezcla de NASC-triplex1/dCas9 proteína/secuencias de refuerzo de ADN de doble cadena a una concentración equivalente de NASC-triplex1. La mezcla se puede incubar durante 1 hora a 37°C. NASC-triplex1/dCas9 proteína/secuencias de refuerzo de ADN de doble cadena/NASC-triplex2/dCas9 complejos de jaula cerrada pueden congelarse a -80°C para el almacenamiento a largo plazo.

La Figura 6L ilustra un ejemplo de una estructura de andamio de ácidos nucleicos subyacente (NASC-CC), con las proteínas Cas omitidas por claridad. NASC-triplex1 y NASC-triplex2 son las estructuras en esta Figura que corresponden a la estructura mostrada en la Figura 6G. Las líneas discontinuas en esta Figura dan una indicación de los tipos de conexiones creadas por las secuencias de refuerzo de ADN de doble cadena entre NASC-triplex1 y NASC-triplex2. La Figura 6M ilustra el NASC-CC en complejo con las proteínas de la proteína dCas9. Las proteínas dCas9 están representadas en esta Figura por los círculos grises.

Siguiendo la guía de la presente especificación y el Ejemplo, la formulación de otros complejos de jaula cerrada de ribonucleoproteína NASC-CC (por ejemplo, que comprenden varias combinaciones de las composiciones de NASC descritas en este documento) puede ser practicada por un experto en la técnica con otras composiciones de la NASC y las proteínas Cas relacionadas.

Ejemplo 14

Análisis estructural de complejos de jaula cerrada de ribonucleoproteína NASC

El siguiente ejemplo describe la caracterización de los complejos de jaula cerrada de proteína NASC-CC/dCas para verificar el ensamblaje adecuado y evaluar el tamaño y el volumen de los complejos de proteína NASC-CC/dCas ensamblados. El procedimiento descrito a continuación está adaptado de Andersen, F., et al., *Nucleic Acids Research* 36(4): 1113-1119 (2008) y Lapinaite, A., et al., *Nature* 502(7472): 519-523 (2013). No todos los pasos siguientes son necesarios para la evaluación ni el orden de los pasos debe ser el presentado.

A. Ensayo de cambio de movilidad electroforética de complejos de proteínas NASC-CC/dCas

Los complejos de proteínas NASC-CC/dCas pueden formularse como se describe en el Ejemplo 13, manipulados para que se pueda usar una secuencia de refuerzo de ADN de doble cadena radiomarcada. La secuencia de refuerzo de ADN de doble cadena puede marcarse radiactivamente preparando la siguiente mezcla de reacción: secuencias de refuerzo de ADN de doble cadena en presencia de polinucleótido quinasa T4 (New England Biolabs, Ipswich, MA), α -³²P ATP (Promega, Madison, WI), y el tampón de reacción polinucleótido quinasa IX T4. La mezcla de reacción se puede incubar y luego inactivar por calor a 65°C durante 20 minutos. El ADN radiomarcado se puede purificar utilizando una columna Illustra MicroSpin G-25 (GE Healthcare, Pittsburgh, PA).

Alternativamente, uno o más de los componentes de NASC-CC pueden ser radiomarcados de una manera similar.

Los complejos de proteína NASC-CC/dCas9 radiomarcados se pueden dividir en partes alícuotas en un volumen de 10 μ l, y se resuelven a 4°C por electroforesis en un gel de poliácridamida nativa al 8% que contiene IX Tris/Borato/EDTA (ácido 90 mM, ácido bórico 90 mM, EDTA 2 mM en pH 8,3) y MgCl₂ 5 mM. Posteriormente, el gel se puede secar y obtener imágenes utilizando el sistema PMI™ (Bio-Rad Laboratories, Hercules, CA). Componentes polinucleotídicos individuales de NASC-CC (por ejemplo, NASC-PC1-triplex1, NASC-PC2-triplex1, NASC-PC3-triplex1, NASC-PC1-triplex2, NASC-PC2-triplex2, NASC-PC3-triplex2, se pueden usar secuencias de refuerzo de ADN de doble cadena, y/o componentes individuales complejados con proteína dCas9) como controles para la comparación para identificar el cambio de movilidad electroforética de los complejos de proteínas NASC-CC/dCas9 completamente formados.

B. Dispersión de rayos X de ángulo pequeño de los complejos de proteínas NASC-CC/dCas9

Los complejos de proteína NASC-CC/dCas9, descritos en el Ejemplo 13, pueden dializarse a 4°C en un tampón de HEPES 20 mM, KCl 100 mM, MgCl₂ 5 mM y glicerol al 5% a pH 7,4. La preparación dializada de los complejos de

proteínas NASC-CC/dCas9 se puede dispensar en los pozos de una placa de 96 pozos utilizando una serie de concentración de 1 mg/mL a 5 mg/mL en un volumen final de 40 µL.

Las mediciones de dispersión de rayos X de ángulo pequeño (SAXS) se pueden recopilar en un proveedor de servicios, como la Fuente de luz avanzada (Berkeley, CA), utilizando una línea de luz de biología estructural para ciencias de la vida (SIBYLS) con un detector Mar165CCD. Los datos se pueden recopilar en múltiples marcos con rangos de tiempo de exposición de 0,5 segundos a 10 segundos y distancias de detección de 1,5 metros a 5 metros. Se pueden evaluar las condiciones óptimas de recolección para determinar el daño mínimo de la radiación a la muestra, así como las relaciones óptimas de señal a ruido. Del mismo modo, la energía de la línea de luz kiloelectronvoltios (keV) se puede sintonizar desde un rango de 7keV a 15 keV. El control de solo tampón se puede usar como fondo y se resta de las mediciones.

El procesamiento y análisis de los datos se puede realizar utilizando un software de línea de haz estándar y PRIMUS (Konarev, P., et al., Journal of Applall Crystallography 36: 1277-1282 (2003)). El modelado de datos se puede realizar utilizando programas de análisis SAXS, como un paquete de software de código abierto (por ejemplo, ATSAS 2.7.2, Petoukhov, M., et al., Journal of Applall Crystallography 45: 342-350 (2012)). Las coordenadas atómicas de la proteína Cas9 y ARN de guía única en diferentes estados de unión a nucleótidos (por ejemplo, ARNsg solo, ARNsg más cadena diana, ARNsg más diana y no diana), así como estructuras (por ejemplo, nucleasas, proteínas, bicatenarias ADN y ARN) están disponibles en la Protein Database (PDB, www.rcsb.org/pdb/home/home.do) o en el Electron Microscopy Data Bank (EMDB, www.ebi.ac.uk/pdbe/emdb/). Estas coordenadas atómicas se pueden usar para calcular el volumen interno, el tamaño de los poros y los tamaños de caja cerrada de los complejos de proteínas NASC-CC/dCas9 mediante el modelado, combinado con datos SAXS.

Los complejos de proteínas NASC-CC/dCas9 pueden modificarse para aumentar o disminuir el volumen interno, el tamaño de los poros o los tamaños de jaula cerrada según sea necesario para el empaquetado y el suministro de biomoléculas, proteínas u otras cargas útiles. Dichas modificaciones pueden incluir, pero no se limitan a, alargar o acortar la primera secuencia de ácidos nucleicos del elemento de tallo (Figura 6A, 608-609; Figura 6H, 623-624/658-657, 626-625/627-628, 652-651/655-646), y/o la secuencia de refuerzo de ADN de doble cadena (Tabla 21).

Siguiendo la guía de la presente especificación y los Ejemplos, el análisis de las características estructurales de los complejos de proteínas NASC-CC/Cas, incluidos volúmenes internos, tamaños de poros y tamaños de jaula cerrada, puede ser practicado por un experto en la técnica.

Listado de secuencias

<110> CARIBOU BIOSCIENCES, INC

<120> Ácidos nucleicos manipulados dirigidos a ácidos nucleicos

<130> CBI020,30

<150> 62/263,232

<151> 2015-12-04

<160> 109

<170>

<210> 1

<211> 23

<212> ADN

<213> Secuencia Artificial

<220>

<223> Cebador Oligonucleotídico

<400> 1

agtaataata cgactcacta tag 23

<210> 2

<211> 26

<212> ADN

<213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 2
 5 aagcaccgac tcggtgccac ttttc 26
 <210> 3
 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 10 <220>
 <223> Cebador Oligonucleotídico
 <400> 3
 taatacgact cactataggg gccactaggg acaggatgtc tcagagctat gcagt 55
 <210> 4
 15 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 20 <400> 4
 taatacgact cactatagta ggctatagt tagatctgtc tcagagctat gcagt 55
 <210> 5
 <211> 55
 <212> ADN
 25 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 5
 taatacgact cactatagga aaaagtggaa gcggcgagtc tcagagctat gcagt 55
 30 <210> 6
 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 <220>
 35 <223> Cebador Oligonucleotídico
 <400> 6
 taatacgact cactataggg gataagtcgt gtcttacgtc tcagagctat gcagt 55

<210> 7
 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> Cebador Oligonucleotídico
 <400> 7
 taatacgact cactataggg gccactaggg acaggatgca catgaggatt ctcac 55
 <210> 8
 10 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 15 <400> 8
 taatacgact cactatagta ggctatagtg tagatctgca catgaggatt ctcac 55
 <210> 9
 <211> 55
 <212> ADN
 20 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 9
 taatacgact cactatagga aaaagtggaa gcggcgagca catgaggatt ctcac 55
 25 <210> 10
 <211> 55
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> Cebador Oligonucleotídico
 <400> 10
 taatacgact cactataggc gataagtcgt gtcttacgca catgaggatt ctcac 55
 <210> 11
 <211> 60
 35 <212> ADN
 <213> Secuencia Artificial
 <220>

<223> Cebador Oligonucleotídico
 <400> 11
 gtctcagagc tatgcagtc tggacaactg ccgaacagga ctgcatagca agttgagata 60
 <210> 12
 5 <211> 60
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 10 <400> 12
 gactcgggtgc cacttttca agttgataac ggactagcct tatctcaact tgctatgcag 60
 <210> 13
 <211> 60
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 13
 gcacatgagg attctcatga gggacggcag aagaacctca tgagaatcca agtatgtgta 60
 20 <210> 14
 <211> 60
 <212> ADN
 <213> Secuencia Artificial
 <220>
 25 <223> Cebador Oligonucleotídico
 <400> 14
 gactcgggtgc cacttttca agttgataac ggactagcct tacacatact tggattctca 60
 <210> 15
 <211> 60
 30 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 15gtctcagagc tatgcagtc tggacaactg ccgaacctca tgagaatcca agtatgtgta 60
 35 <210> 16
 <211> 60
 <212> ADN

<213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 <400> 16
 5 gactcgggtgc cactttttca agttgataac ggactagcct tacacatact tggattctca 60
 <210> 17
 <211> 60
 <212> ADN
 <213> Secuencia Artificial
 10 <220>
 <223> Cebador Oligonucleotídico
 <400> 17
 gccatgagg attctcatga gggacggcag aagaacagga ctgcatagca agttgagata 60
 <210> 18
 15 <211> 60
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador Oligonucleotídico
 20 <400> 18
 gactcgggtgc cactttttca agttgataac ggactagcct tatctcaact tgctatgcag 60
 <210> 19
 <211> 237
 <212> ADN
 25 <213> Secuencia Artificial
 <220>
 <223> oligonucleótidos de clonación
 <220>
 <221>característica_misc
 30 <222> (125)..(125)
 <223> n es a, c, g, o t
 <400> 19
 tacacgtact tagtcgctga agctcttcta tgcaagcaga agacggcata cgagatcgag 60
 taatgtgact ggagttcaga cgtgtgctct tccgatctgc tactggggcc actagggaca 120
 ggatnnggtgc tagctcagat cggaagagcg tcgtgtaggg aaagagtgtg ggctatagtg 180
 tagatctcgg tggtcgccgt atcattggta gaagagccgt caatcgagtt cgtacct 237

<210> 20

<211> 4175

<212> ADN

<213> Secuencia Artificial

5 <220>

<223> secuencia diana AAVS-1

<400> 20

ctcatgacca aaatccctta acgtgagtta cgcgcgcgtc gttccactga gcgtcagacc	60
ccgtagaaaa gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct	120
tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa	180
ctctttttcc gaaggtaact ggcttcagca gagcgcagat accaaatact gttcttctag	240
tgtagccgta gttagccac cacttcaaga actctgtagc accgcctaca tacctcgctc	300
tgctaatacct gttaccagtg gctgctgccg gtggcgataa gtctgtctctt accgggttg	360
actcaagacg atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggctcgtgca	420
cacagcccag cttggagcga acgacctaca ccgaactgag atacctacag cgtgagctat	480
gagaaagcgc cacgcttccc gaagggagaa aggcggacag gtatccggta agcggcagg	540
tcggaacagg agagcgcacg agggagcttc cagggggaaa cgcctggtat ctttatagtc	600
ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc	660
ggagcctatg gaaaaacgcc agcaacgcgg ccttttttacg gttcctggcc ttttgctggc	720
cttttgctca catgttcttt cctgcgttat cccctgattc tgtggataac cgtattaccg	780
cctttgagtg agctgatacc gctcgcgcga gccgaacgac cgagcgcagc gagtcaagtga	840
gcgaggaagc ggaaggcgag agtagggaac tgccaggcat caaactaagc agaaggcccc	900
tgacggatgg cctttttgcg tttctacaaa ctctttctgt gttgtaaaac gacggccagt	960
cttaagctcg ggccccctgg gcggttctga taacgagtaa tcgttaatcc gcaaataacg	1020
taaaaacccg cttcggcggg tttttttatg gggggagttt agggaaagag catttgtcag	1080
aatattttaag ggcgcctgtc actttgcttg atatatgaga attatttaac cttataaatg	1140
agaaaaaagc aacgcacttt aaataagata cgttgctttt tcgattgatg aacacctata	1200
attaaactat tcatctatta tttatgattt tttgtatata caatatttct agtttgtaa	1260
agagaattaa gaaaataaat ctcgaaaata ataaaggga aatcagtttt tgatatcaaa	1320
attatacatg tcaacgataa tacaaaatat aatacaaaact ataagatgtt atcagtattt	1380
attatgcatt tagaataaat tttgtgtcgc ccttaattgt gagcggataa caattacgag	1440
cttcatgcac agtgaaatca tgaaaaattt atttgctttg tgagcggata acaattataa	1500
tatgtggaat tgtgagcgtc cacaattcca caacggtttc cctctagaaa taattttgtt	1560

ES 2 735 773 T3

taacttttaa	ggaggtaaaa	aatgtacacg	tacttagtcg	ctgaagctct	tctatgcaag	1620
cagaagacgg	catacgagat	cgagtaatgt	gactggagtt	cagacgtgtg	ctcttccgat	1680
ctgctactgg	ggccactagg	gacaggattg	gtgctagctc	agatcggaag	agcgtcgtgt	1740
agggaaagag	tgtaggctat	agtgtagatc	tcggtggtcg	ccgtatcatt	ggtagaagag	1800
ccgtcaatcg	agttcgtacc	tggttgaccc	caagggcgac	acccctaata	tagcccgggc	1860
gaaaggccca	gtctttcgac	tgagcctttc	gttttatttg	atgcctggca	gttccctact	1920
ctcgcatggg	gagtccccac	actaccatcg	gcgctacggc	gtttcacttc	tgagttcggc	1980
atggggtcag	gtgggaccac	cgcgctactg	ccgccaggca	aacaaggggt	gttatgagcc	2040
atattcaggt	ataaatgggc	tcgcgataat	gttcagaatt	ggtaattgg	ttgtaacact	2100
gaccctatt	tgtttatttt	tctaaataca	ttcaaataatg	tatccgctca	tgagacaata	2160
accctgataa	atgcttcaat	aatattgaaa	aaggaagaat	atgagccata	ttcaacggga	2220
aacgtcgagg	ccgcgattaa	attccaacat	ggatgctgat	ttatatgggt	ataaatgggc	2280
tcgcgataat	gtcgggcaat	caggtgcgac	aatctatcgc	ttgtatggga	agcccgatgc	2340
gccagagttg	tttctgaaac	atggcaaagg	tagcgttgcc	aatgatgtta	cagatgagat	2400
ggtcagacta	aactggctga	cggaaatttat	gccacttccg	accatcaagc	attttatccg	2460
tactcctgat	gatgcatggg	tactcaccac	tcgatcccc	ggaaaaacag	cgttccaggt	2520
attagaagaa	tatcctgatt	caggtgaaaa	tattgttgat	gcgctggcag	tgttcctgcg	2580
ccggttgcac	tcgattcctg	tttgtaattg	tccttttaac	agcgatcgcg	tatttcgcct	2640
cgcgcaggcg	caatcacgaa	tgaataacgg	tttggttgat	gcgagtgatt	ttgatgacga	2700
gcgtaatggc	tggcctgttg	aacaagtctg	gaaagaaatg	cataaaacttt	tgccattctc	2760
accggattca	gtcgtcactc	atgggtgattt	ctcacttgat	aaccttattt	ttgacgaggg	2820
gaaattaata	ggttgatttg	atgttggacg	agtcggaatc	gcagaccgat	accaggatct	2880
tgccatccta	tggaaactgcc	tcggtgagtt	ttctccttca	ttacagaaac	ggctttttca	2940
aaaatatggg	attgataatc	ctgatatgaa	taaattgcag	tttcatttga	tgctcgatga	3000
gtttttctaa	gcggcgcgcc	atcgaatggc	gcaaaacctt	tcgcggtatg	gcatgatagc	3060
gcccgaaga	gagtcaattc	aggggtggtga	atatgaaacc	agtaacgtta	tacgatgtcg	3120
cagagtatgc	cgggtgtctct	tatcagaccg	tttccgcgct	ggtgaaccag	gccagccacg	3180
tttctgcgaa	aacgcgggaa	aaagtggaag	cggcgatggc	ggagctgaat	tacattccca	3240
accgcgtggc	acaacaactg	gcggggcaaac	agtcgttgct	gattggcggt	gccacctcca	3300
gtctggccct	gcacgcgccg	tcgcaaattg	tcgcggcgat	taaatactcg	gccgatcaac	3360
tgggtgccag	cgtggtggtg	tcgatggtag	aacgaagcgg	cgtcgaagcc	tgtaaagcgg	3420

cggtgcacaa tcttctcgcg caacgcgtca gtgggctgat cattaactat ccgctggatg 3480
accaggatgc cattgctgtg gaagctgcct gcactaatgt tccggcgta tttcttgatg 3540
tctctgacca gacacccatc aacagtatta ttttctcca tgaggacggg acgcgactgg 3600
gcgtggagca tctggtcgca ttgggtcacc agcaaatcg cgtgttagcg ggcccattaa 3660
gttctgtctc ggcggtctg cgtctggctg gctggcataa atatctcact cgcaatcaaa 3720
ttcagccgat agcggaacgg gaagggcact ggagtgccat gtccgggtttt caacaaacca 3780
tgcaaatgct gaatgagggc atcgttccca ctgcgatgct gggtgccaac gatcagatgg 3840
cgctggggcg aatgcgcgcc attaccgagt ccgggctgcg cgttggtgcg gatattctcg 3900
tagtgggata cgacgatacc gaagatagct catgttatat ccggccgta accaccatca 3960
aacaggattt tcgcctgctg gggcaaacca gcgtggaccg cttgctgcaa ctctctcagg 4020
gccaggcggt gaagggcaat cagctgttgc cagtctcact ggtgaaaaga aaaaccaccc 4080
tggcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg 4140
cacgacaggt ttcccgactg gaaagcgggc agtga 4175

<210> 21

<211> 56

<212> ADN

5 <213> Secuencia Artificial

<220>

<223> oligonucleótidos de clonación

<400> 21

tacacgtact tagtcgctga agctcttcta tgcaagcaga agacggcata cgagat 56

10 <210> 22

<211> 56

<212> ADN

<213> Secuencia Artificial

<220>

15 <223> oligonucleótidos de clonación

<400> 22

aggtacgaac tcgattgacg gctcttctac caatgatacg gcgaccacg agatct 56

<210> 23

<211> 20

20 <212> ADN

<213> Secuencia Artificial

<220>

<223> cebador

<400> 23
 ccccggtctc ctgtggattc 20
 <210> 24
 <211> 20
 5 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> cebador
 <400> 24
 10 atcctctctg gctccatcgt 20
 <210> 25
 <211> 127
 <212> ARN
 <213> Secuencia Artificial
 15 <220>
 <223> ARNsg-1-AAVST1
 <400> 25
 gggggccacua gggacaggau gucucagagc uaucgagucc uggacaacug ccgaacagga 60
 cugcauagca aguugagaua aggcuaagucc guuaucacu ugaaaaagug gcaccgaguc 120
 ggugcuu 127
 <210> 26
 20 <211> 127
 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> ARNsg-1-VT2
 25 <400> 26
 guaggcuaua guguaagauu gucucagagc uaucgagucc uggacaacug ccgaacagga 60
 cugcauagca aguugagaua aggcuaagucc guuaucacu ugaaaaagug gcaccgaguc 120
 ggugcuu 127
 <210> 27
 <211> 127
 <212> ARN
 30 <213> Secuencia Artificial
 <220>
 <223> ARNsg-1-VT3

	<400> 27	
	ggaaaaagug gaagcggcga gucucagagc uaugcagucc uggacaacug ccgaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 28	
	<211> 127	
5	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> ARNsg-1-VT4	
	<400> 28	
	ggcgauaagu cgugucuuaac gucucagagc uaugcagucc uggacaacug ccgaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
10	ggugcuu	127
	<210> 29	
	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
15	<220>	
	<223> ARNsg-2-AAVST1	
	<400> 29	
	ggggccacua gggacaggau gcacaugagg auucucauga gggacggcag aagaaccuca	60
	ugagaaucca aguauugua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 30	
20	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> ARNsg-2-VT2	
25	<400> 30	
	guaggcuaua guguagauu gcacaugagg auucucauga gggacggcag aagaaccuca	60
	ugagaaucca aguauugua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 31	

	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
5	<223> ARNsg-2-VT3	
	<400> 31	
	ggaaaaagug gaagcggcga gcacaugagg auucucauga gggacggcag aagaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 32	
	<211> 127	
10	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> ARNsg-2-VT4	
	<400> 32	
	ggcgauaagu cgugucuuaac gcacaugagg auucucauga gggacggcag aagaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guaucaacu ugaaaaagug gcaccgaguc	120
15	ggugcuu	127
	<210> 33	
	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
20	<220>	
	<223> NASC-PC1-AAVST1	
	<400> 33	
	ggggccacua gggacaggau gucucagagc uaucgagucc uggacaacug ccgaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 34	
25	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> NASC-PC1-VT2	
30	<400> 34	

	guagggcuaua guguagauca gucucagagc uaugcagucc uggacaacug ccgaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 35	
	<211> 127	
	<212> ARN	
5	<213> Secuencia Artificial	
	<220>	
	<223> NASC-PC1-VT3	
	<400> 35	
	ggaaaaagug gaagcggcga gucucagagc uaugcagucc uggacaacug ccgaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
10	<210> 36	
	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
15	<223> NASC-PC1-VT4	
	<400> 36	
	ggcgauaagu cgugucuauac gucucagagc uaugcagucc uggacaacug ccgaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 37	
	<211> 127	
20	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> NASC-PC2- AAVST1	
	<400> 37	
	ggggccacua gggacaggau gcacaugagg auucucauga gggacggcag aagaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
25	ggugcuu	127
	<210> 38	
	<211> 127	

	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> NASC-PC2-VT2	
5	<400> 38	
	guaggcuaua guguagauca gcacaugagg auucucauga gggacggcag aagaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 39	
	<211> 127	
	<212> ARN	
10	<213> Secuencia Artificial	
	<220>	
	<223> NASC-PC2-VT3	
	<400> 39	
	ggaaaaagug gaagcggcga gcacaugagg auucucauga gggacggcag aagaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
15	<210> 40	
	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
20	<223> NASC-PC2-VT4	
	<400> 40	
	ggcgauaagu cgugucuuaa gcacaugagg auucucauga gggacggcag aagaacagga	60
	cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 41	
	<211> 20	
25	<212> ADN	
	<213> Homo sapiens	
	<400> 41	
	ggtggacaag cggcagatag	20
	<210> 42	
30	<211> 20	

<212> ADN

<213> Homo sapiens

<400> 42

gcaccatgtt gccggtcctc 20

5 <210> 43

<211> 118

<212> ARN

<213> Secuencia Artificial

<220>

10 <223> ARNsg- XRCC5-T1

<400> 43

gguggacaag cggcagauag guuuuagagc uaugcuguuu uggaaacaaa acagcauagc 60

aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 44

<211> 118

15 <212> ARN

<213> Secuencia Artificial

<220>

<223> ARNsg- XRCC5-T3

<400> 44

gcaccauguu gccgguccuc guuuuagagc uaugcuguuu uggaaacaaa acagcauagc 60

20 aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 45

<211> 118

<212> ARN

<213> Secuencia Artificial

25 <220>

<223> NASCA-PC1- XRCC5-T1

<400> 45

gguggacaag cggcagauag guuuuagagc uaugcuguuu uggaaacuuu ucagcacgau 60

aaguuaauau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 46

30 <211> 118

<212> ARN

<213> Secuencia Artificial

<220>

<223> NASCA-PC2- XRCC5-T3

<400> 46

gcaccauguu gccgguccuc gaaauagaau cgugcugaaa aggaacaaa acagcauagc 60

aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 47

5 <211> 50

<212> ADN

<213> Secuencia Artificial

<220>

<223>Región diana 1 XRCC5

10 <400> 47

cactcttcc ctacacgacg ctctccgat cttgcgatg ctcagagttc 50

<210> 48

<211> 48

<212> ADN

15 <213> Secuencia Artificial

<220>

<223>Región diana 1 XRCC5

<400> 48

ggagttcaga cgtgtgctct tccgatctcc aagtccatgg ctttctt 48

20 <210> 49

<211> 52

<212> ADN

<213> Secuencia Artificial

<220>

25 <223> Región diana 1 XRCC5

<400> 49

cactcttcc ctacacgacg ctctccgat ctttcaggc ctagcaggaa ac 52

<210> 50

<211> 48

30 <212> ADN

<213> Secuencia Artificial

<220>

<223> Región diana 1 XRCC5

<400> 50

35 ggagttcaga cgtgtgctct tccgatctcc cattcttctg cttgaccg 48

<210> 51
 <211> 57
 <212> ADN
 <213> Secuencia Artificial
 5 <220>
 <223> Cebador
 <400> 51
 caagcagaag acggcatacg agattacgtg atgtgactgg agttcagacg tgtgctc 57
 <210> 52
 10 <211> 58
 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador
 15 <400> 52
 aatgatacgg cgaccaccga gatctacacc gtctaataca ctcttcct acacgacg 58
 <210> 53
 <211> 58
 <212> ADN
 20 <213> Secuencia Artificial
 <220>
 <223> Cebador
 <400> 53
 aatgatacgg cgaccaccga gatctacact ctctccgaca ctcttcct acacgacg 58
 25 <210> 54
 <211> 58
 <212> ADN
 <213> Secuencia Artificial
 <220>
 30 <223> Cebador
 <400> 54
 aatgatacgg cgaccaccga gatctacact cgactagaca ctcttcct acacgacg 58
 <210> 55
 <211> 58
 35 <212> ADN
 <213> Secuencia Artificial
 <220>

<223> Cebador

<400> 55aatgatacgg cgaccaccga gatctacact tctagctaca ctcttcct acacgacg 58

<210> 56

<211> 984

5 <212> PRT

<213> Campylobacter jejuni

<400> 56

ES 2 735 773 T3

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

ES 2 735 773 T3

Arg	Cys	Ile	Ala	Gln	Ser	Phe	Leu	Lys	Asp	Glu	Leu	Lys	Leu	Ile	Phe	195	200	205	
Lys	Lys	Gln	Arg	Glu	Phe	Gly	Phe	Ser	Phe	Ser	Lys	Lys	Phe	Glu	Glu	210	215	220	
Glu	Val	Leu	Ser	Val	Ala	Phe	Tyr	Lys	Arg	Ala	Leu	Lys	Asp	Phe	Ser	225	230	235	240
His	Leu	Val	Gly	Asn	Cys	Ser	Phe	Phe	Thr	Asp	Glu	Lys	Arg	Ala	Pro	245	250	255	
Lys	Asn	Ser	Pro	Leu	Ala	Phe	Met	Phe	Val	Ala	Leu	Thr	Arg	Ile	Ile	260	265	270	
Asn	Leu	Leu	Asn	Asn	Leu	Lys	Asn	Thr	Glu	Gly	Ile	Leu	Tyr	Thr	Lys	275	280	285	
Asp	Asp	Leu	Asn	Ala	Leu	Leu	Asn	Glu	Val	Leu	Lys	Asn	Gly	Thr	Leu	290	295	300	
Thr	Tyr	Lys	Gln	Thr	Lys	Lys	Leu	Leu	Gly	Leu	Ser	Asp	Asp	Tyr	Glu	305	310	315	320
Phe	Lys	Gly	Glu	Lys	Gly	Thr	Tyr	Phe	Ile	Glu	Phe	Lys	Lys	Tyr	Lys	325	330	335	
Glu	Phe	Ile	Lys	Ala	Leu	Gly	Glu	His	Asn	Leu	Ser	Gln	Asp	Asp	Leu	340	345	350	
Asn	Glu	Ile	Ala	Lys	Asp	Ile	Thr	Leu	Ile	Lys	Asp	Glu	Ile	Lys	Leu	355	360	365	
Lys	Lys	Ala	Leu	Ala	Lys	Tyr	Asp	Leu	Asn	Gln	Asn	Gln	Ile	Asp	Ser	370	375	380	
Leu	Ser	Lys	Leu	Glu	Phe	Lys	Asp	His	Leu	Asn	Ile	Ser	Phe	Lys	Ala	385	390	395	400
Leu	Lys	Leu	Val	Thr	Pro	Leu	Met	Leu	Glu	Gly	Lys	Lys	Tyr	Asp	Glu	405	410	415	
Ala	Cys	Asn	Glu	Leu	Asn	Leu	Lys	Val	Ala	Ile	Asn	Glu	Asp	Lys	Lys	420	425	430	
Asp	Phe	Leu	Pro	Ala	Phe	Asn	Glu	Thr	Tyr	Tyr	Lys	Asp	Glu	Val	Thr	435	440	445	

ES 2 735 773 T3

Asn	Pro	Val	Val	Leu	Arg	Ala	Ile	Lys	Glu	Tyr	Arg	Lys	Val	Leu	Asn
450						455					460				
Ala	Leu	Leu	Lys	Lys	Tyr	Gly	Lys	Val	His	Lys	Ile	Asn	Ile	Glu	Leu
465					470					475					480
Ala	Arg	Glu	Val	Gly	Lys	Asn	His	Ser	Gln	Arg	Ala	Lys	Ile	Glu	Lys
				485					490					495	
Glu	Gln	Asn	Glu	Asn	Tyr	Lys	Ala	Lys	Lys	Asp	Ala	Glu	Leu	Glu	Cys
			500					505					510		
Glu	Lys	Leu	Gly	Leu	Lys	Ile	Asn	Ser	Lys	Asn	Ile	Leu	Lys	Leu	Arg
	515						520					525			
Leu	Phe	Lys	Glu	Gln	Lys	Glu	Phe	Cys	Ala	Tyr	Ser	Gly	Glu	Lys	Ile
530						535					540				
Lys	Ile	Ser	Asp	Leu	Gln	Asp	Glu	Lys	Met	Leu	Glu	Ile	Asp	Ala	Ile
545					550					555					560
Tyr	Pro	Tyr	Ser	Arg	Ser	Phe	Asp	Asp	Ser	Tyr	Met	Asn	Lys	Val	Leu
				565					570						575
Val	Phe	Thr	Lys	Gln	Asn	Gln	Glu	Lys	Leu	Asn	Gln	Thr	Pro	Phe	Glu
			580					585					590		
Ala	Phe	Gly	Asn	Asp	Ser	Ala	Lys	Trp	Gln	Lys	Ile	Glu	Val	Leu	Ala
		595					600					605			
Lys	Asn	Leu	Pro	Thr	Lys	Lys	Gln	Lys	Arg	Ile	Leu	Asp	Lys	Asn	Tyr
	610					615					620				
Lys	Asp	Lys	Glu	Gln	Lys	Asn	Phe	Lys	Asp	Arg	Asn	Leu	Asn	Asp	Thr
625					630					635					640
Arg	Tyr	Ile	Ala	Arg	Leu	Val	Leu	Asn	Tyr	Thr	Lys	Asp	Tyr	Leu	Asp
				645					650					655	
Phe	Leu	Pro	Leu	Ser	Asp	Asp	Glu	Asn	Thr	Lys	Leu	Asn	Asp	Thr	Gln
			660					665					670		
Lys	Gly	Ser	Lys	Val	His	Val	Glu	Ala	Lys	Ser	Gly	Met	Leu	Thr	Ser
	675						680					685			
Ala	Leu	Arg	His	Thr	Trp	Gly	Phe	Ser	Ala	Lys	Asp	Arg	Asn	Asn	His

ES 2 735 773 T3

690					695					700					
Leu	His	His	Ala	Ile	Asp	Ala	Val	Ile	Ile	Ala	Tyr	Ala	Asn	Asn	Ser
705					710					715					720
Ile	Val	Lys	Ala	Phe	Ser	Asp	Phe	Lys	Lys	Glu	Gln	Glu	Ser	Asn	Ser
				725					730					735	
Ala	Glu	Leu	Tyr	Ala	Lys	Lys	Ile	Ser	Glu	Leu	Asp	Tyr	Lys	Asn	Lys
			740					745						750	
Arg	Lys	Phe	Phe	Glu	Pro	Phe	Ser	Gly	Phe	Arg	Gln	Lys	Val	Leu	Asp
		755					760					765			
Lys	Ile	Asp	Glu	Ile	Phe	Val	Ser	Lys	Pro	Glu	Arg	Lys	Lys	Pro	Ser
770						775					780				
Gly	Ala	Leu	His	Glu	Glu	Thr	Phe	Arg	Lys	Glu	Glu	Glu	Phe	Tyr	Gln
785					790					795					800
Ser	Tyr	Gly	Gly	Lys	Glu	Gly	Val	Leu	Lys	Ala	Leu	Glu	Leu	Gly	Lys
				805					810					815	
Ile	Arg	Lys	Val	Asn	Gly	Lys	Ile	Val	Lys	Asn	Gly	Asp	Met	Phe	Arg
			820					825					830		
Val	Asp	Ile	Phe	Lys	His	Lys	Lys	Thr	Asn	Lys	Phe	Tyr	Ala	Val	Pro
		835					840					845			
Ile	Tyr	Thr	Met	Asp	Phe	Ala	Leu	Lys	Val	Leu	Pro	Asn	Lys	Ala	Val
850						855					860				
Ala	Arg	Ser	Lys	Lys	Gly	Glu	Ile	Lys	Asp	Trp	Ile	Leu	Met	Asp	Glu
865					870					875					880
Asn	Tyr	Glu	Phe	Cys	Phe	Ser	Leu	Tyr	Lys	Asp	Ser	Leu	Ile	Leu	Ile
				885					890					895	
Gln	Thr	Lys	Asp	Met	Gln	Glu	Pro	Glu	Phe	Val	Tyr	Tyr	Asn	Ala	Phe
			900					905					910		
Thr	Ser	Ser	Thr	Val	Ser	Leu	Ile	Val	Ser	Lys	His	Asp	Asn	Lys	Phe
		915					920					925			
Glu	Thr	Leu	Ser	Lys	Asn	Gln	Lys	Ile	Leu	Phe	Lys	Asn	Ala	Asn	Glu
930						935					940				

ES 2 735 773 T3

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

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
965 970 975

Arg Gln Arg Glu Asp Phe Lys Lys
980

<210> 57

<211> 146

<212> ARN

5 <213> Secuencia Artificial

<220>

<223> NASC-PC1-triplex1

<400> 57

aucuuguuga cacgaggaau guuuuagucc cuaauuaau uucuugaaau ugguauauaa 60

ggagggacua caacaaagag uuugcgggac ucugcggggu uacaaucucc uaaaaccgcu 120

uuuaaaauuc aaauaaauuu ugcuuu 146

10 <210> 58

<211> 146

<212> ARN

<213> Secuencia Artificial

<220>

15 <223> NASC-PC2-triplex1

<400> 58

aucuuguuga cacgaggaau guuguagucc cuccuuauau accaagaaaa agaaauuuuaa 60

aacugaacuc caacaaagag uuugcgggac ucugcggggu uacaaucucc uaaaaccgcu 120

uuuaaaauuc aaauaaauuu ugcuuu 146

<210> 59

<211> 146

20 <212> ARN

<213> Secuencia Artificial

<220>

<223> NASC-PC3-triplex1

<400> 59

	aucuuguuga cacgaggaau guuggaguuc aguuuuuuuu uucuugaaaa agaaauuuuaa	60
	uuagggacua aaacaaagag uuugcgggac ucugcggggg uacaaucucc uaaaaccgcu	120
	uuuuuuuuuuc aaauuuuuuu ugcuuu	146
	<210> 60	
	<211> 146	
	<212> ARN	
5	<213> Sequencia Artificial	
	<220>	
	<223> NASC-PC1-triplex2	
	<400> 60	
	cgauuuuuua cagcaaggug guuuuagacc ccucuuccau uucgcgaaag cguuuugaga	60
	gagugaacua caacaaagag uuugcgggac ucugcggggg uacaaucucc uaaaaccgcu	120
	uuuuuuuuuuc aaauuuuuuu ugcuuu	146
10	<210> 61	
	<211> 146	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
15	<223> NASC-PC2-triplex2	
	<400> 61	
	cgauuuuuua cagcaaggug guuguaguuc acucucucaa aacgcgaaaa agaaauuuuaa	60
	uaaggaacua caacaaagag uuugcgggac ucugcggggg uacaaucucc uaaaaccgcu	120
	uuuuuuuuuuc aaauuuuuuu ugcuuu	146
	<210> 62	
	<211> 146	
20	<212> ARN	
	<213> Sequence Artificial	
	<220>	
	<223> NASC-PC3-triplex2	
	<400> 62	
	cgauuuuuua cagcaaggug guuguaguuc cuuuuuuuuu uucuugaaag cgaaauugaa	60
	gaggggucua aaacaaagag uuugcgggac ucugcggggg uacaaucucc uaaaaccgcu	120
25	uuuuuuuuuuc aaauuuuuuu ugcuuu	146
	<210> 63	
	<211> 144	

<212> ADN

<213> Secuencia Artificial

<220>

<223>secuencia de oligonucleótidos

5 <400> 63

cgtcgctatg atttgcctat atcttgttga cgcgaggaat gtaaacaacg agttccgcta 60

ttgggatgga gtttaactgt cgcaactctc atcgcaatgt cagtcacctt gctgtattat 120

atcgcgcatg ataaagtacg ccat 144

<210> 64

<211> 65

<212> ARN

10 <213> Secuencia Artificial

<220>

<223> Componente polinucleótido NASCA

<220>

<221>característica_misc

15 <222> (46)..(65)

<223> n es a, c, g, o u

<400> 64

cuccggcgau gucacaccga acugauaaau ucuacucuug uagaunnnnn nnnnnnnnnn 60

nnnnn 65

<210> 65

20 <211> 65

<212> ARN

<213> Secuencia Artificial

<220>

<223>NASCA polynucleotide component

25 <220>

<221>característica_misc

<222> (46)..(65)

<223> n es a, c, g, o u

<400> 65

ucggugugac aucgccggag uugauaaaau ucuacucuug uagaunnnnn nnnnnnnnnn 60

30 nnnnn 65

<210> 66

<211> 35

<212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 5 <400> 66
 cuccggcgau gucacaccga acugauaaau ucuac 35
 <210> 67
 <211> 26
 <212> ARN
 10 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 15 <222> (7) .. (26)
 <223> n es a, c, g, o u
 <400> 67
 guagaunnnn nnnnnnnnnn nnnnnn 26
 <210> 68
 20 <211> 35
 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 25 <400> 68
 ucggugugac aucgccggag uugauaaau uguag 35
 <210> 69
 <211> 26
 <212> ARN
 30 <213> Secuencia Artificial
 <220>
 <223>Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 35 <222> (7) .. (26)
 <223> n es a, c, g, o u
 <400> 69

	cuacaunnnn nnnnnnnnnn nnnnnn	26	
	<210> 70		
	<211> 65		
	<212> ARN		
5	<213> Secuencia Artificial		
	<220>		
	<223> Componente polinucleótido NASCA		
	<220>		
	<221característica_misc		
10	<222> (20) .. (39)		
	<223> n es a, c, g, o u		
	<400> 70		
	aaauucuacu cuuguagaun nnnnnnnnnn nnnnnnnnna cugaucuccg gcgaugucac	60	
	accga	65	
	<210> 71		
15	<211> 65		
	<212> ARN		
	<213> Secuencia Artificial		
	<220>		
	<223> Componente polinucleótido NASCA		
20	<220>		
	<221característica_misc		
	<222> (20) .. (39)		
	<223> n es a, c, g, o u		
	<400> 71		
	aaauucuacu cuuguagaun nnnnnnnnnn nnnnnnnnnu ugauaucggu gugacaucgc	60	
25	cggag	65	
	<210> 72		
	<211> 91		
	<212> ARN		
	<213> Secuencia Artificial		
30	<220>		
	<223> Componente polinucleótido NASCA		
	<220>		
	<221característica_misc		
	<222> (46)..(65)		

	<223> n es a, c, g, o u	
	<400> 72	
	cuccggcgau gucacaccga acugauaaau ucuacucuug uagaunnnnn nnnnnnnnnn	60
	nnnnnuugau agucuaaggc agcuaggguc u	91
	<210> 73	
5	<211> 65	
	<212> ARN	
	<213> I Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	
10	<220>	
	<221característica_misc	
	<222> (46)..(65)	
	<223> n es a, c, g, o u	
	<400> 73	
	ucggugugac aucgccggag uugauaaaau ucuacucuug uagaunnnnn nnnnnnnnnn	60
15	nnnnn	65
	<210> 74	
	<211> 65	
	<212> ARN	
	<213> Secuencia Artificial	
20	<220>	
	<223> Componente polinucleótido NASCA	
	<220>	
	<221característica_misc	
	<222> (20) .. (39)	
25	<223> n es a, c, g, o u	
	<400> 74	
	aaauucuacu cuuguagaun nnnnnnnnnn nnnnnnnnnu accaaagacc cuagcugccu	60
	uagac	65
	<210> 75	
	<211> 91	
30	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	

	<220>	
	<221>característica_misc	
	<222> (72)..(91)	
	<223> n es a, c, g, o u	
5	<400> 75	
	cuccggcgau gucacaccga acugaugucu aaggcagcua gggucuuuga uaaauuucua	60
	cucuuguaga unnnnnnnnnn nnnnnnnnnn n	91
	<210> 76	
	<211> 91	
	<212> ARN	
10	<213> I Secuencia Artificial	
	<220>	
	<223>Componente polinucleótido NASCA	
	<220>	
	<221>característica_misc	
15	<222> (72)..(91)	
	<223> n es a, c, g, o u	
	<400> 76	
	ugcgaaccac ugugagccag uaccaaucgg ugugacaucg ccggaguuga uaaauuucua	60
	cucuuguaga unnnnnnnnnn nnnnnnnnnn n	91
	<210> 77	
20	<211> 61	
	<212> ARN	
	<213> I Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	
25	<400> 77	
	cuccggcgau gucacaccga acugaugucu aaggcagcua gggucuuuga uaaauuucua	60
	c	61
	<210> 78	
	<211> 26	
	<212> ARN	
30	<213> Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	
	<220>	

<221característica_misc
 <222> (7).. (26)
 <223> n es a, c, g, o u
 <400> 78
 5 guagaunnnn nnnnnnnnnn nnnnnn 26
 <210> 79
 <211> 61
 <212> ARN
 <213> Secuencia Artificial
 10 <220>
 <223> Componente polinucleótido NASCA
 <400> 79
 ugcgaccac ugugagccag uaccaaucgg ugugacaucg ccggaguuga uaaaauugua 60
 g 61
 <210> 80
 15 <211> 26
 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 20 <220>
 <221característica_misc
 <222> (7).. (26)
 <223> n es a, c, g, o u
 <400> 80
 25 cuacaunnnn nnnnnnnnnn nnnnnn 26
 <210> 81
 <211> 146
 <212> ARN
 <213> Secuencia Artificial
 30 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221característica_misc
 <222> (1)..(20)
 35 <223> n es a, c, g, o u

	<400> 81	
	nnnnnnnnnn nnnnnnnnnn guuuuagucc cuaauaaaau uucuugaaau ugguauauaa	60
	ggagggacua caacaaagag uuugcgggac ucugcggggu uacaaucucc uaaaaccgcu	120
	uuuaaaauuc aaauaaaauu ugcuuu	146
	<210> 82	
	<211> 146	
5	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	
	<220>	
10	<221característica_misc	
	<222> (1)..(20)	
	<223> n es a, c, g, o u	
	<400> 82	
	nnnnnnnnnn nnnnnnnnnn guuguagucc cuccuuauau accaagaaaa agaaauuuuaa	60
	uuagggacua aaacaaagag uuugcgggac ucugcggggu uacaaucucc uaaaaccgcu	120
	uuuaaaauuc aaauaaaauu ugcuuu	146
15	<210> 83	
	<211> 127	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
20	<223> Componente polinucleótido NASCA	
	<220>	
	<221característica_misc	
	<222> (1)..(20)	
	<223> n es a, c, g, o u	
25	<400> 83	
	nnnnnnnnnn nnnnnnnnnn gucucagagc uaugcagucc uggacaacug ccgaaccuca	60
	ugagaaucca aguaugugua aggcuaagucc guuaucacu ugaaaaagug gcaccgaguc	120
	ggugcuu	127
	<210> 84	
	<211> 127	
	<212> ARN	
30	<213> Secuencia Artificial	

<220>
 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 5 <222> (1)..(20)
 <223> n es a, c, g, o u
 <400> 84
 nnnnnnnnnn nnnnnnnnnn gcacaugagg auucucauga gggacggcag aagaacagga 60
 cugcauagca aguugagaua aggcuaagucc guuaucaacu ugaaaaagug gcaccgaguc 120
 ggugcuu 127
 <210> 85
 10 <211> 118
 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 15 <220>
 <221>característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, o u
 <400> 85
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguuu uggaaagguc auguccuua 60
 20 aaguuguaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118
 <210> 86
 <211> 118
 <212> ARN
 <213> Secuencia Artificial
 25 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 <222> (1)..(20)
 30 <223> n es a, c, g, o u
 <400> 86
 nnnnnnnnnn nnnnnnnnnn guaauagaau cgugcugaaa aggaaacaaa acagcauagc 60
 aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 87
 <211> 118
 <212> ARN
 <213> Secuencia Artificial

5 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 <222> (1) .. (20)

10 <223> n es a, c, g, o u
 <400> 87
 nnnnnnnnnn nnnnnnnnnn guuacagaug aaggacauga ccgaaacuuu ucagcacgau 60
 aaguuaauau aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuu 118

<210> 88
 <211> 45

15 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 <220>

20 <221>característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, o u
 <400> 88
 nnnnnnnnnn nnnnnnnnnn guuuuagucc cuaauaaaau uucuu 45

25 <210> 89
 <211> 45
 <212> ARN
 <213> Secuencia Artificial
 <220>

30 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, o u

35 <400> 89nnnnnnnnnn nnnnnnnnnn guuguagucc cuccuuauau accaa 45

<210> 90
 <211> 205
 <212> ARN
 <213> Secuencia Artificial

5 <220>
 <223> Componente polinucleótido NASCA
 <400> 90
 aagaaauuuu auuagggacu aaaacaaaga guuugcggga cucugcgggg uuacaauccc 60
 cuaaaaccgc uuuuaaaauu caaauaaauu uugcuuuagu ugauaaaauu gguauauaag 120
 gagggacuac aacaaagagu uugcgggacu cugcgggguu acaauccccc aaaaccgcuu 180
 uuaaaauuca aauaaaauuu gcuuu 205

<210> 91

10 <211> 75
 <212> ADN/ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA

15 <220>
 <221>característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, t o u
 <400> 91
 nnnnnnnnnn nnnnnnnnnn guuuuagagc tatgtgtgtga aaacagcata gcaaguuaaa 60

20 auaaggcuac ugccg 75

<210> 92
 <211> 75
 <212> ADN/ARN
 <213> Secuencia Artificial

25 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221>característica_misc
 <222> (1)..(20)

30 <223> n es a, c, g, t o u
 <400> 92

	nnnnnnnnnn nnnnnnnnnn guuuuagagc tatgctgtga aaacagcata gcaaguuaaa	60
	auaaggcuag ucacg	75
	<210> 93	
	<211> 105	
	<212> ARN	
5	<213> Secuencia Artificial	
	<220>	
	<223> Componente polinucleótido NASCA	
	<400> 93	
	cggcaguccg uuaucaacuu gaaaaagugg caccgagucg gugcuuaguu gauaaaucgu	60
	gacguccguu aucaacuuga aaaaguggca ccgagucggu gcuuu	105
10	<210> 94	
	<211> 155	
	<212> ARN	
	<213> Secuencia Artificial	
	<220>	
15	<223> Componente polinucleótido NASCA	
	<220>	
	<221>característica_misc	
	<222> (1) .. (20)	
	<223> n es a, c, g, o u	
20	<220>	
	<221>característica_misc	
	<222> (136)..(155)	
	<223> n es a, c, g, o u	
	<400> 94	
	nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguga aaacagcaua gcaaguuaaa	60
	auaaggcuag uccguuauca acuugaaaaa guggcaccga gucggugcuu acugauaaau	120
25	ucuacucuug uagaunnnnnn nnnnnnnnnn nnnnn	155
	<210> 95	
	<211> 114	
	<212> ARN	
	<213> Secuencia Artificial	
30	<220>	
	<223> Componente polinucleótido NASCA	
	<220>	

<221característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, o u
 <400> 95
 nnnnnnnnnn nnnnnnnnnn guuuuugua ucucaagauu caaauaacag cauagcaagu 60
 5 uaaaauaagg cuauccguua ucaacuugaa aaaguggcac cgagucggug cuuu 114
 <210> 96
 <211> 116
 <212> ARN
 <213> Secuencia Artificial
 10 <220>
 <223> Componente polinucleótido NASCA
 <220>
 <221característica_misc
 <222> (1) .. (20)
 15 <223> n es a, c, g, o u
 <400> 96
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguua cguaaaucuu gcagaagcua 60
 caaagauaag gcuucaugcc gaaaucaaca ccugucuu uuauggcagg guguuu 116
 <210> 97
 <211> 76
 20 <212> ARN
 <213> Secuencia Artificial
 <220>
 <223> Componente polinucleótido NASCA
 <220>
 25 <221característica_misc
 <222> (1)..(20)
 <223> n es a, c, g, o u
 <220>
 <221característica_misc
 30 <222> (57)..(76)
 <223> n es a, c, g, o u
 <400> 97
 nnnnnnnnnn nnnnnnnnnn guuuuagagc uaugcuguua ccaauugaua guagaunnnn 60
 nnnnnnnnnn nnnnnnn 76

<210> 98
 <211> 83
 <212> ARN
 <213> Secuencia Artificial
 5 <220>
 <223> Componente polinucleótido NASCA
 <400> 98
 aaauucuacu ugauaacagc auagcaaguu aaaauaaggc uauccguuau caacuugaaa 60
 aaguggcacc gagucggugc uuu 83
 <210> 99
 10 <211> 1053
 <212> PRT
 <213> Staphylococcus aureus
 <400> 99

ES 2 735 773 T3

Met Lys Arg Asn Tyr Ile Leu Gly Leu Asp Ile Gly Ile Thr Ser Val
1 5 10 15

Gly Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly
20 25 30

Val Arg Leu Phe Lys Glu Ala Asn Val Glu Asn Asn Glu Gly Arg Arg
35 40 45

Ser Lys Arg Gly Ala Arg Arg Leu Lys Arg Arg Arg Arg His Arg Ile
50 55 60

Gln Arg Val Lys Lys Leu Leu Phe Asp Tyr Asn Leu Leu Thr Asp His
65 70 75 80

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

ES 2 735 773 T3

Ser Gln Lys Leu Ser Glu Glu Glu Phe Ser Ala Ala Leu Leu His Leu
 100 105 110
 Ala Lys Arg Arg Gly Val His Asn Val Asn Glu Val Glu Glu Asp Thr
 115 120 125
 Gly Asn Glu Leu Ser Thr Lys Glu Gln Ile Ser Arg Asn Ser Lys Ala
 130 135 140
 Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
 145 150 155 160
 Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr
 165 170 175
 Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190
 Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg
 195 200 205
 Arg Thr Tyr Tyr Glu Gly Pro Gly Glu Gly Ser Pro Phe Gly Trp Lys
 210 215 220
 Asp Ile Lys Glu Trp Tyr Glu Met Leu Met Gly His Cys Thr Tyr Phe
 225 230 235 240
 Pro Glu Glu Leu Arg Ser Val Lys Tyr Ala Tyr Asn Ala Asp Leu Tyr
 245 250 255
 Asn Ala Leu Asn Asp Leu Asn Asn Leu Val Ile Thr Arg Asp Glu Asn
 260 265 270
 Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
 275 280 285
 Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
 290 295 300
 Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
 305 310 315 320
 Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr
 325 330 335
 Ala Arg Lys Glu Ile Ile Glu Asn Ala Glu Leu Leu Asp Gln Ile Ala

ES 2 735 773 T3

340										345					350				
Lys	Ile	Leu	Thr	Ile	Tyr	Gln	Ser	Ser	Glu	Asp	Ile	Gln	Glu	Glu	Leu				
		355					360					365							
Thr	Asn	Leu	Asn	Ser	Glu	Leu	Thr	Gln	Glu	Glu	Ile	Glu	Gln	Ile	Ser				
	370					375					380								
Asn	Leu	Lys	Gly	Tyr	Thr	Gly	Thr	His	Asn	Leu	Ser	Leu	Lys	Ala	Ile				
385					390					395					400				
Asn	Leu	Ile	Leu	Asp	Glu	Leu	Trp	His	Thr	Asn	Asp	Asn	Gln	Ile	Ala				
			405						410					415					
Ile	Phe	Asn	Arg	Leu	Lys	Leu	Val	Pro	Lys	Lys	Val	Asp	Leu	Ser	Gln				
			420					425					430						
Gln	Lys	Glu	Ile	Pro	Thr	Thr	Leu	Val	Asp	Asp	Phe	Ile	Leu	Ser	Pro				
		435					440					445							
Val	Val	Lys	Arg	Ser	Phe	Ile	Gln	Ser	Ile	Lys	Val	Ile	Asn	Ala	Ile				
	450					455					460								
Ile	Lys	Lys	Tyr	Gly	Leu	Pro	Asn	Asp	Ile	Ile	Ile	Glu	Leu	Ala	Arg				
465					470				475						480				
Glu	Lys	Asn	Ser	Lys	Asp	Ala	Gln	Lys	Met	Ile	Asn	Glu	Met	Gln	Lys				
				485					490					495					
Arg	Asn	Arg	Gln	Thr	Asn	Glu	Arg	Ile	Glu	Glu	Ile	Ile	Arg	Thr	Thr				
			500					505					510						
Gly	Lys	Glu	Asn	Ala	Lys	Tyr	Leu	Ile	Glu	Lys	Ile	Lys	Leu	His	Asp				
		515					520					525							
Met	Gln	Glu	Gly	Lys	Cys	Leu	Tyr	Ser	Leu	Glu	Ala	Ile	Pro	Leu	Glu				
	530					535					540								
Asp	Leu	Leu	Asn	Asn	Pro	Phe	Asn	Tyr	Glu	Val	Asp	His	Ile	Ile	Pro				
545					550					555					560				
Arg	Ser	Val	Ser	Phe	Asp	Asn	Ser	Phe	Asn	Asn	Lys	Val	Leu	Val	Lys				
				565				570						575					
Gln	Glu	Glu	Asn	Ser	Lys	Lys	Gly	Asn	Arg	Thr	Pro	Phe	Gln	Tyr	Leu				
			580					585					590						

ES 2 735 773 T3

Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
 595 600 605
 Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
 610 615 620
 Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
 625 630 635
 Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
 645 650 655
 Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys
 660 665 670
 Val Lys Ser Ile Asn Gly Gly Phe Thr Ser Phe Leu Arg Arg Lys Trp
 675 680 685
 Lys Phe Lys Lys Glu Arg Asn Lys Gly Tyr Lys His His Ala Glu Asp
 690 695 700
 Ala Leu Ile Ile Ala Asn Ala Asp Phe Ile Phe Lys Glu Trp Lys Lys
 705 710 715 720
 Leu Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys
 725 730 735
 Gln Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu
 740 745 750
 Ile Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp
 755 760 765
 Tyr Lys Tyr Ser His Arg Val Asp Lys Lys Pro Asn Arg Glu Leu Ile
 770 775 780
 Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
 785 790 795 800
 Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
 805 810 815
 Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
 820 825 830
 Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly
 835 840 845

ES 2 735 773 T3

Asp Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr
850 855 860

Leu Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile
865 870 875 880

Lys Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp
885 890 895

Tyr Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr
900 905 910

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
915 920 925

Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
930 935 940

Lys Ala Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945 950 955 960

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

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990

Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005

Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020

Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 100

<211> 1368

<212> PRT

5 <213> Streptococcus pyogenes

<400> 100

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

ES 2 735 773 T3

Gly	Trp	Ala	Val	Ile	Thr	Asp	Asp	Tyr	Lys	Val	Pro	Ser	Lys	Lys	Phe	20	25	30
Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys	Asn	Leu	Ile	35	40	45
Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu	50	55	60
Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys	65	70	75
Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	85	90	95
Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	100	105	110
His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	115	120	125
His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	130	135	140
Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	145	150	155
Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	165	170	175
Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln	Thr	Tyr	180	185	190
Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val	Asp	Ala	195	200	205
Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu	Glu	Asn	210	215	220
Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Ser	Leu	Phe	Gly	Asn	225	230	235
Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser	Asn	Phe	245	250	255
Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr	Tyr	Asp	260	265	270

ES 2 735 773 T3

Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr	Ala	Asp	275	280	285
Leu	Phe	Leu	Ala	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	Ser	Asp	290	295	300
Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser	Ala	Ser	305	310	315
Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu	Leu	Lys	325	330	335
Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile	Phe	Phe	340	345	350
Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly	Ala	Ser	355	360	365
Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys	Met	Asp	370	375	380
Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu	Leu	Arg	385	390	395
Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu	405	410	415
Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe	420	425	430
Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile	435	440	445
Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp	450	455	460
Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu	465	470	475
Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr	485	490	495
Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser	500	505	510
Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys			

ES 2 735 773 T3

515	520	525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln 530 535 540		
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr 545 550 555 560		
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp 565 570 575		
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly 580 585 590		
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp 595 600 605		
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr 610 615 620		
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala 625 630 635 640		
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr 645 650 655		
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp 660 665 670		
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe 675 680 685		
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe 690 695 700		
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu 705 710 715 720		
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly 725 730 735		
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly 740 745 750		
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln 755 760 765		

ES 2 735 773 T3

Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	
770						775					780					
Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	
785					790					795					800	
Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	
				805					810					815		
Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	
			820					825					830			
Leu	Ser	Asp	Tyr	Asp	Val	Asp	His	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	
		835					840					845				
Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	
	850					855					860					
Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	
865					870					875					880	
Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	
				885					890					895		
Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	
			900					905					910			
Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	
		915					920					925				
Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	
	930					935					940					
Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	
945					950					955					960	
Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	
				965					970					975		
Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	
			980					985					990			
Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	
		995					1000					1005				
Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala		
	1010					1015						1020				

ES 2 735 773 T3

Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe
1025						1030					1035			
Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala
1040						1045					1050			
Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu
1055						1060					1065			
Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val
1070						1075					1080			
Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr
1085						1090					1095			
Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys
1100						1105					1110			
Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro
1115						1120					1125			
Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val
1130						1135					1140			
Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys
1145						1150					1155			
Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser
1160						1165					1170			
Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys
1175						1180					1185			
Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu
1190						1195					1200			
Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly
1205						1210					1215			
Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val
1220						1225					1230			
Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser
1235						1240					1245			
Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys
1250						1255					1260			

ES 2 735 773 T3

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 101

<211> 1368

<212> PRT

5 <213> Streptococcus pyogenes

<400> 101

ES 2 735 773 T3

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

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

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

ES 2 735 773 T3

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
 130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175

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

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Ser Leu Phe Gly Asn
 225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300

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

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser

ES 2 735 773 T3

355		360		365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp				
370		375		380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg				
385		390		400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu				
	405		410	415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe				
	420		425	430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile				
	435		440	445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp				
450		455		460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu				
465		470		480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr				
	485		490	495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser				
	500		505	510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys				
	515		520	525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln				
530		535		540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr				
545		550		560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp				
	565		570	575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly				
	580		585	590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp				
595		600		605

ES 2 735 773 T3

Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr	610	615	620	
Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala	625	630	635	640
His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr	645	650	655	
Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp	660	665	670	
Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe	675	680	685	
Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe	690	695	700	
Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	Asp	Ser	Leu	705	710	715	720
His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro	Ala	Ile	Lys	Lys	Gly	725	730	735	
Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu	Val	Lys	Val	Met	Gly	740	745	750	
Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg	Glu	Asn	Gln	755	760	765	
Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu	Arg	Met	Lys	Arg	Ile	770	775	780	
Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile	Leu	Lys	Glu	His	Pro	785	790	795	800
Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu	Tyr	Leu	Tyr	Tyr	Leu	805	810	815	
Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu	Leu	Asp	Ile	Asn	Arg	820	825	830	
Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro	Gln	Ser	Phe	Leu	Lys	835	840	845	
Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg	Ser	Asp	Lys	Asn	Arg	850	855	860	

ES 2 735 773 T3

Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val	Val	Lys	Lys	Met	Lys	865	870	875	880
Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu	Ile	Thr	Gln	Arg	Lys	885	890	895	
Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly	Leu	Ser	Glu	Leu	Asp	900	905	910	
Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu	Thr	Arg	Gln	Ile	Thr	915	920	925	
Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met	Asn	Thr	Lys	Tyr	Asp	930	935	940	
Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val	Ile	Thr	Leu	Lys	Ser	945	950	955	960
Lys	Leu	Val	Ser	Asp	Phe	Arg	Lys	Asp	Phe	Gln	Phe	Tyr	Lys	Val	Arg	965	970	975	
Glu	Ile	Asn	Asn	Tyr	His	His	Ala	His	Asp	Ala	Tyr	Leu	Asn	Ala	Val	980	985	990	
Val	Gly	Thr	Ala	Leu	Ile	Lys	Lys	Tyr	Pro	Lys	Leu	Glu	Ser	Glu	Phe	995	1000	1005	
Val	Tyr	Gly	Asp	Tyr	Lys	Val	Tyr	Asp	Val	Arg	Lys	Met	Ile	Ala		1010	1015	1020	
Lys	Ser	Glu	Gln	Glu	Ile	Gly	Lys	Ala	Thr	Ala	Lys	Tyr	Phe	Phe		1025	1030	1035	
Tyr	Ser	Asn	Ile	Met	Asn	Phe	Phe	Lys	Thr	Glu	Ile	Thr	Leu	Ala		1040	1045	1050	
Asn	Gly	Glu	Ile	Arg	Lys	Arg	Pro	Leu	Ile	Glu	Thr	Asn	Gly	Glu		1055	1060	1065	
Thr	Gly	Glu	Ile	Val	Trp	Asp	Lys	Gly	Arg	Asp	Phe	Ala	Thr	Val		1070	1075	1080	
Arg	Lys	Val	Leu	Ser	Met	Pro	Gln	Val	Asn	Ile	Val	Lys	Lys	Thr		1085	1090	1095	
Glu	Val	Gln	Thr	Gly	Gly	Phe	Ser	Lys	Glu	Ser	Ile	Leu	Pro	Lys		1100	1105	1110	

ES 2 735 773 T3

Arg	Asn	Ser	Asp	Lys	Leu	Ile	Ala	Arg	Lys	Lys	Asp	Trp	Asp	Pro
1115						1120					1125			
Lys	Lys	Tyr	Gly	Gly	Phe	Asp	Ser	Pro	Thr	Val	Ala	Tyr	Ser	Val
1130						1135					1140			
Leu	Val	Val	Ala	Lys	Val	Glu	Lys	Gly	Lys	Ser	Lys	Lys	Leu	Lys
1145						1150					1155			
Ser	Val	Lys	Glu	Leu	Leu	Gly	Ile	Thr	Ile	Met	Glu	Arg	Ser	Ser
1160						1165					1170			
Phe	Glu	Lys	Asn	Pro	Ile	Asp	Phe	Leu	Glu	Ala	Lys	Gly	Tyr	Lys
1175						1180					1185			
Glu	Val	Lys	Lys	Asp	Leu	Ile	Ile	Lys	Leu	Pro	Lys	Tyr	Ser	Leu
1190						1195					1200			
Phe	Glu	Leu	Glu	Asn	Gly	Arg	Lys	Arg	Met	Leu	Ala	Ser	Ala	Gly
1205						1210					1215			
Glu	Leu	Gln	Lys	Gly	Asn	Glu	Leu	Ala	Leu	Pro	Ser	Lys	Tyr	Val
1220						1225					1230			
Asn	Phe	Leu	Tyr	Leu	Ala	Ser	His	Tyr	Glu	Lys	Leu	Lys	Gly	Ser
1235						1240					1245			
Pro	Glu	Asp	Asn	Glu	Gln	Lys	Gln	Leu	Phe	Val	Glu	Gln	His	Lys
1250						1255					1260			
His	Tyr	Leu	Asp	Glu	Ile	Ile	Glu	Gln	Ile	Ser	Glu	Phe	Ser	Lys
1265						1270					1275			
Arg	Val	Ile	Leu	Ala	Asp	Ala	Asn	Leu	Asp	Lys	Val	Leu	Ser	Ala
1280						1285					1290			
Tyr	Asn	Lys	His	Arg	Asp	Lys	Pro	Ile	Arg	Glu	Gln	Ala	Glu	Asn
1295						1300					1305			
Ile	Ile	His	Leu	Phe	Thr	Leu	Thr	Asn	Leu	Gly	Ala	Pro	Ala	Ala
1310						1315					1320			
Phe	Lys	Tyr	Phe	Asp	Thr	Thr	Ile	Asp	Arg	Lys	Arg	Tyr	Thr	Ser
1325						1330					1335			
Thr	Lys	Glu	Val	Leu	Asp	Ala	Thr	Leu	Ile	His	Gln	Ser	Ile	Thr

ES 2 735 773 T3

1340

1345

1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
 1355 1360 1365

<210> 102

<211> 1053

<212> PRT

5 <213> Staphylococcus aureus

<400> 102

ES 2 735 773 T3

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

ES 2 735 773 T3

195					200					205					
Arg	Thr	Tyr	Tyr	Glu	Gly	Pro	Gly	Glu	Gly	Ser	Pro	Phe	Gly	Trp	Lys
210						215					220				
Asp	Ile	Lys	Glu	Trp	Tyr	Glu	Met	Leu	Met	Gly	His	Cys	Thr	Tyr	Phe
225					230					235					240
Pro	Glu	Glu	Leu	Arg	Ser	Val	Lys	Tyr	Ala	Tyr	Asn	Ala	Asp	Leu	Tyr
				245					250					255	
Asn	Ala	Leu	Asn	Asp	Leu	Asn	Asn	Leu	Val	Ile	Thr	Arg	Asp	Glu	Asn
			260					265					270		
Glu	Lys	Leu	Glu	Tyr	Tyr	Glu	Lys	Phe	Gln	Ile	Ile	Glu	Asn	Val	Phe
		275					280					285			
Lys	Gln	Lys	Lys	Lys	Pro	Thr	Leu	Lys	Gln	Ile	Ala	Lys	Glu	Ile	Leu
	290					295					300				
Val	Asn	Glu	Glu	Asp	Ile	Lys	Gly	Tyr	Arg	Val	Thr	Ser	Thr	Gly	Lys
305					310					315					320
Pro	Glu	Phe	Thr	Asn	Leu	Lys	Val	Tyr	His	Asp	Ile	Lys	Asp	Ile	Thr
				325					330					335	
Ala	Arg	Lys	Glu	Ile	Ile	Glu	Asn	Ala	Glu	Leu	Leu	Asp	Gln	Ile	Ala
			340					345					350		
Lys	Ile	Leu	Thr	Ile	Tyr	Gln	Ser	Ser	Glu	Asp	Ile	Gln	Glu	Glu	Leu
		355					360					365			
Thr	Asn	Leu	Asn	Ser	Glu	Leu	Thr	Gln	Glu	Glu	Ile	Glu	Gln	Ile	Ser
	370					375					380				
Asn	Leu	Lys	Gly	Tyr	Thr	Gly	Thr	His	Asn	Leu	Ser	Leu	Lys	Ala	Ile
385					390					395					400
Asn	Leu	Ile	Leu	Asp	Glu	Leu	Trp	His	Thr	Asn	Asp	Asn	Gln	Ile	Ala
				405					410					415	
Ile	Phe	Asn	Arg	Leu	Lys	Leu	Val	Pro	Lys	Lys	Val	Asp	Leu	Ser	Gln
			420					425					430		
Gln	Lys	Glu	Ile	Pro	Thr	Thr	Leu	Val	Asp	Asp	Phe	Ile	Leu	Ser	Pro
		435					440					445			

ES 2 735 773 T3

Val	Val	Lys	Arg	Ser	Phe	Ile	Gln	Ser	Ile	Lys	Val	Ile	Asn	Ala	Ile	450	455	460
Ile	Lys	Lys	Tyr	Gly	Leu	Pro	Asn	Asp	Ile	Ile	Ile	Glu	Leu	Ala	Arg	465	470	475
Glu	Lys	Asn	Ser	Lys	Asp	Ala	Gln	Lys	Met	Ile	Asn	Glu	Met	Gln	Lys	485	490	495
Arg	Asn	Arg	Gln	Thr	Asn	Glu	Arg	Ile	Glu	Glu	Ile	Ile	Arg	Thr	Thr	500	505	510
Gly	Lys	Glu	Asn	Ala	Lys	Tyr	Leu	Ile	Glu	Lys	Ile	Lys	Leu	His	Asp	515	520	525
Met	Gln	Glu	Gly	Lys	Cys	Leu	Tyr	Ser	Leu	Glu	Ala	Ile	Pro	Leu	Glu	530	535	540
Asp	Leu	Leu	Asn	Asn	Pro	Phe	Asn	Tyr	Glu	Val	Asp	His	Ile	Ile	Pro	545	550	555
Arg	Ser	Val	Ser	Phe	Asp	Asn	Ser	Phe	Asn	Asn	Lys	Val	Leu	Val	Lys	565	570	575
Gln	Glu	Glu	Ala	Ser	Lys	Lys	Gly	Asn	Arg	Thr	Pro	Phe	Gln	Tyr	Leu	580	585	590
Ser	Ser	Ser	Asp	Ser	Lys	Ile	Ser	Tyr	Glu	Thr	Phe	Lys	Lys	His	Ile	595	600	605
Leu	Asn	Leu	Ala	Lys	Gly	Lys	Gly	Arg	Ile	Ser	Lys	Thr	Lys	Lys	Glu	610	615	620
Tyr	Leu	Leu	Glu	Glu	Arg	Asp	Ile	Asn	Arg	Phe	Ser	Val	Gln	Lys	Asp	625	630	635
Phe	Ile	Asn	Arg	Asn	Leu	Val	Asp	Thr	Arg	Tyr	Ala	Thr	Arg	Gly	Leu	645	650	655
Met	Asn	Leu	Leu	Arg	Ser	Tyr	Phe	Arg	Val	Asn	Asn	Leu	Asp	Val	Lys	660	665	670
Val	Lys	Ser	Ile	Asn	Gly	Gly	Phe	Thr	Ser	Phe	Leu	Arg	Arg	Lys	Trp	675	680	685
Lys	Phe	Lys	Lys	Glu	Arg	Asn	Lys	Gly	Tyr	Lys	His	His	Ala	Glu	Asp	690	695	700

ES 2 735 773 T3

Ala	Leu	Ile	Ile	Ala	Asn	Ala	Asp	Phe	Ile	Phe	Lys	Glu	Trp	Lys	Lys		
705					710					715						720	
Leu	Asp	Lys	Ala	Lys	Lys	Val	Met	Glu	Asn	Gln	Met	Phe	Glu	Glu	Lys		
				725					730						735		
Gln	Ala	Glu	Ser	Met	Pro	Glu	Ile	Glu	Thr	Glu	Gln	Glu	Tyr	Lys	Glu		
			740					745						750			
Ile	Phe	Ile	Thr	Pro	His	Gln	Ile	Lys	His	Ile	Lys	Asp	Phe	Lys	Asp		
		755					760					765					
Tyr	Lys	Tyr	Ser	His	Arg	Val	Asp	Lys	Lys	Pro	Asn	Arg	Glu	Leu	Ile		
	770					775					780						
Asn	Asp	Thr	Leu	Tyr	Ser	Thr	Arg	Lys	Asp	Asp	Lys	Gly	Asn	Thr	Leu		
785					790					795					800		
Ile	Val	Asn	Asn	Leu	Asn	Gly	Leu	Tyr	Asp	Lys	Asp	Asn	Asp	Lys	Leu		
				805					810						815		
Lys	Lys	Leu	Ile	Asn	Lys	Ser	Pro	Glu	Lys	Leu	Leu	Met	Tyr	His	His		
			820					825					830				
Asp	Pro	Gln	Thr	Tyr	Gln	Lys	Leu	Lys	Leu	Ile	Met	Glu	Gln	Tyr	Gly		
		835					840					845					
Asp	Glu	Lys	Asn	Pro	Leu	Tyr	Lys	Tyr	Tyr	Glu	Glu	Thr	Gly	Asn	Tyr		
	850					855					860						
Leu	Thr	Lys	Tyr	Ser	Lys	Lys	Asp	Asn	Gly	Pro	Val	Ile	Lys	Lys	Ile		
865					870					875					880		
Lys	Tyr	Tyr	Gly	Asn	Lys	Leu	Asn	Ala	His	Leu	Asp	Ile	Thr	Asp	Asp		
				885					890					895			
Tyr	Pro	Asn	Ser	Arg	Asn	Lys	Val	Val	Lys	Leu	Ser	Leu	Lys	Pro	Tyr		
			900					905						910			
Arg	Phe	Asp	Val	Tyr	Leu	Asp	Asn	Gly	Val	Tyr	Lys	Phe	Val	Thr	Val		
		915					920					925					
Lys	Asn	Leu	Asp	Val	Ile	Lys	Lys	Glu	Asn	Tyr	Tyr	Glu	Val	Asn	Ser		
	930					935					940						
Lys	Ala	Tyr	Glu	Glu	Ala	Lys	Lys	Leu	Lys	Lys	Ile	Ser	Asn	Gln	Ala		
945					950					955					960		

ES 2 735 773 T3

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

Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
 980 985 990

Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
 995 1000 1005

Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
 1010 1015 1020

Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
 1025 1030 1035

Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
 1040 1045 1050

<210> 103

<211> 984

<212> PRT

5 <213> Campylobacter jejuni

<400> 103

ES 2 735 773 T3

Met	Ala	Arg	Ile	Leu	Ala	Phe	Asp	Ile	Gly	Ile	Ser	Ser	Ile	Gly	Trp
1				5					10					15	
Ala	Phe	Ser	Glu	Asn	Asp	Glu	Leu	Lys	Asp	Cys	Gly	Val	Arg	Ile	Phe
			20					25					30		
Thr	Lys	Val	Glu	Asn	Pro	Lys	Thr	Gly	Glu	Ser	Leu	Ala	Leu	Pro	Arg
		35					40					45			
Arg	Leu	Ala	Arg	Ser	Ala	Arg	Lys	Arg	Leu	Ala	Arg	Arg	Lys	Ala	Arg
	50					55					60				
Leu	Asn	His	Leu	Lys	His	Leu	Ile	Ala	Asn	Glu	Phe	Lys	Leu	Asn	Tyr
65					70					75					80
Glu	Asp	Tyr	Gln	Ser	Phe	Asp	Glu	Ser	Leu	Ala	Lys	Ala	Tyr	Lys	Gly
				85					90					95	
Ser	Leu	Ile	Ser	Pro	Tyr	Glu	Leu	Arg	Phe	Arg	Ala	Leu	Asn	Glu	Leu
			100					105					110		
Leu	Ser	Lys	Gln	Asp	Phe	Ala	Arg	Val	Ile	Leu	His	Ile	Ala	Lys	Arg
		115					120					125			

ES 2 735 773 T3

Arg	Gly	Tyr	Asp	Asp	Ile	Lys	Asn	Ser	Asp	Asp	Lys	Glu	Lys	Gly	Ala	130	135	140
Ile	Leu	Lys	Ala	Ile	Lys	Gln	Asn	Glu	Glu	Lys	Leu	Ala	Asn	Tyr	Gln	145	150	155
Ser	Val	Gly	Glu	Tyr	Leu	Tyr	Lys	Glu	Tyr	Phe	Gln	Lys	Phe	Lys	Glu	165	170	175
Asn	Ser	Lys	Glu	Phe	Thr	Asn	Val	Arg	Asn	Lys	Lys	Glu	Ser	Tyr	Glu	180	185	190
Arg	Cys	Ile	Ala	Gln	Ser	Phe	Leu	Lys	Asp	Glu	Leu	Lys	Leu	Ile	Phe	195	200	205
Lys	Lys	Gln	Arg	Glu	Phe	Gly	Phe	Ser	Phe	Ser	Lys	Lys	Phe	Glu	Glu	210	215	220
Glu	Val	Leu	Ser	Val	Ala	Phe	Tyr	Lys	Arg	Ala	Leu	Lys	Asp	Phe	Ser	225	230	235
His	Leu	Val	Gly	Asn	Cys	Ser	Phe	Phe	Thr	Asp	Glu	Lys	Arg	Ala	Pro	245	250	255
Lys	Asn	Ser	Pro	Leu	Ala	Phe	Met	Phe	Val	Ala	Leu	Thr	Arg	Ile	Ile	260	265	270
Asn	Leu	Leu	Asn	Asn	Leu	Lys	Asn	Thr	Glu	Gly	Ile	Leu	Tyr	Thr	Lys	275	280	285
Asp	Asp	Leu	Asn	Ala	Leu	Leu	Asn	Glu	Val	Leu	Lys	Asn	Gly	Thr	Leu	290	295	300
Thr	Tyr	Lys	Gln	Thr	Lys	Lys	Leu	Leu	Gly	Leu	Ser	Asp	Asp	Tyr	Glu	305	310	315
Phe	Lys	Gly	Glu	Lys	Gly	Thr	Tyr	Phe	Ile	Glu	Phe	Lys	Lys	Tyr	Lys	325	330	335
Glu	Phe	Ile	Lys	Ala	Leu	Gly	Glu	His	Asn	Leu	Ser	Gln	Asp	Asp	Leu	340	345	350
Asn	Glu	Ile	Ala	Lys	Asp	Ile	Thr	Leu	Ile	Lys	Asp	Glu	Ile	Lys	Leu	355	360	365
Lys	Lys	Ala	Leu	Ala	Lys	Tyr	Asp	Leu	Asn	Gln	Asn	Gln	Ile	Asp	Ser			

ES 2 735 773 T3

370		375		380
Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala				
385		390		395
				400
Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu				
		405		410
				415
Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys				
		420		425
				430
Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr				
		435		440
				445
Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn				
		450		455
				460
Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu				
		465		470
				475
Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys				
		485		490
				495
Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys				
		500		505
				510
Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg				
		515		520
				525
Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile				
		530		535
				540
Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile				
		545		550
				555
Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu				
		565		570
				575
Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu				
		580		585
				590
Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala				
		595		600
				605
Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr				
		610		615
				620

ES 2 735 773 T3

Lys	Asp	Lys	Glu	Gln	Lys	Asn	Phe	Lys	Asp	Arg	Asn	Leu	Asn	Asp	Thr	625	630	635	640
Arg	Tyr	Ile	Ala	Arg	Leu	Val	Leu	Asn	Tyr	Thr	Lys	Asp	Tyr	Leu	Asp	645	650	655	
Phe	Leu	Pro	Leu	Ser	Asp	Asp	Glu	Asn	Thr	Lys	Leu	Asn	Asp	Thr	Gln	660	665	670	
Lys	Gly	Ser	Lys	Val	His	Val	Glu	Ala	Lys	Ser	Gly	Met	Leu	Thr	Ser	675	680	685	
Ala	Leu	Arg	His	Thr	Trp	Gly	Phe	Ser	Ala	Lys	Asp	Arg	Asn	Asn	His	690	695	700	
Leu	His	His	Ala	Ile	Asp	Ala	Val	Ile	Ile	Ala	Tyr	Ala	Asn	Asn	Ser	705	710	715	720
Ile	Val	Lys	Ala	Phe	Ser	Asp	Phe	Lys	Lys	Glu	Gln	Glu	Ser	Asn	Ser	725	730	735	
Ala	Glu	Leu	Tyr	Ala	Lys	Lys	Ile	Ser	Glu	Leu	Asp	Tyr	Lys	Asn	Lys	740	745	750	
Arg	Lys	Phe	Phe	Glu	Pro	Phe	Ser	Gly	Phe	Arg	Gln	Lys	Val	Leu	Asp	755	760	765	
Lys	Ile	Asp	Glu	Ile	Phe	Val	Ser	Lys	Pro	Glu	Arg	Lys	Lys	Pro	Ser	770	775	780	
Gly	Ala	Leu	His	Glu	Glu	Thr	Phe	Arg	Lys	Glu	Glu	Glu	Phe	Tyr	Gln	785	790	795	800
Ser	Tyr	Gly	Gly	Lys	Glu	Gly	Val	Leu	Lys	Ala	Leu	Glu	Leu	Gly	Lys	805	810	815	
Ile	Arg	Lys	Val	Asn	Gly	Lys	Ile	Val	Lys	Asn	Gly	Asp	Met	Phe	Arg	820	825	830	
Val	Asp	Ile	Phe	Lys	His	Lys	Lys	Thr	Asn	Lys	Phe	Tyr	Ala	Val	Pro	835	840	845	
Ile	Tyr	Thr	Met	Asp	Phe	Ala	Leu	Lys	Val	Leu	Pro	Asn	Lys	Ala	Val	850	855	860	
Ala	Arg	Ser	Lys	Lys	Gly	Glu	Ile	Lys	Asp	Trp	Ile	Leu	Met	Asp	Glu	865	870	875	880

ES 2 735 773 T3

Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
885 890 895

Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
900 905 910

Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
915 920 925

Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
930 935 940

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

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
965 970 975

Arg Gln Arg Glu Asp Phe Lys Lys
980

<210> 104

<211> 1307

<212> PRT

5 <213> Acidaminococcus sp.

<400> 104

ES 2 735 773 T3

Met Thr Gln Phe Glu Gly Phe Thr Asn Leu Tyr Gln Val Ser Lys Thr
1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Lys His Ile Gln
20 25 30

Glu Gln Gly Phe Ile Glu Glu Asp Lys Ala Arg Asn Asp His Tyr Lys
35 40 45

Glu Leu Lys Pro Ile Ile Asp Arg Ile Tyr Lys Thr Tyr Ala Asp Gln
50 55 60

Cys Leu Gln Leu Val Gln Leu Asp Trp Glu Asn Leu Ser Ala Ala Ile
65 70 75 80

Asp Ser Tyr Arg Lys Glu Lys Thr Glu Glu Thr Arg Asn Ala Leu Ile
85 90 95

Glu Glu Gln Ala Thr Tyr Arg Asn Ala Ile His Asp Tyr Phe Ile Gly
100 105 110

ES 2 735 773 T3

Arg Thr Asp Asn Leu Thr Asp Ala Ile Asn Lys Arg His Ala Glu Ile
 115 120 125
 Tyr Lys Gly Leu Phe Lys Ala Glu Leu Phe Asn Gly Lys Val Leu Lys
 130 135 140
 Gln Leu Gly Thr Val Thr Thr Thr Glu His Glu Asn Ala Leu Leu Arg
 145 150 155 160
 Ser Phe Asp Lys Phe Thr Thr Tyr Phe Ser Gly Phe Tyr Glu Asn Arg
 165 170 175
 Lys Asn Val Phe Ser Ala Glu Asp Ile Ser Thr Ala Ile Pro His Arg
 180 185 190
 Ile Val Gln Asp Asn Phe Pro Lys Phe Lys Glu Asn Cys His Ile Phe
 195 200 205
 Thr Arg Leu Ile Thr Ala Val Pro Ser Leu Arg Glu His Phe Glu Asn
 210 215 220
 Val Lys Lys Ala Ile Gly Ile Phe Val Ser Thr Ser Ile Glu Glu Val
 225 230 235 240
 Phe Ser Phe Pro Phe Tyr Asn Gln Leu Leu Thr Gln Thr Gln Ile Asp
 245 250 255
 Leu Tyr Asn Gln Leu Leu Gly Gly Ile Ser Arg Glu Ala Gly Thr Glu
 260 265 270
 Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
 275 280 285
 Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
 290 295 300
 Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
 305 310 315 320
 Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
 325 330 335
 Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
 340 345 350
 Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
 355 360 365

ES 2 735 773 T3

Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
370 375 380

Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
385 390 395 400

Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
405 410 415

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

Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
435 440 445

Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
450 455 460

Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
465 470 475 480

Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
485 490 495

Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
500 505 510

Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val
515 520 525

Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp
530 535 540

Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn
545 550 555 560

Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys
565 570 575

Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys
580 585 590

Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys
595 600 605

Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr

ES 2 735 773 T3

610					615					620						
Pro	Ile	Leu	Leu	Ser	Asn	Asn	Phe	Ile	Glu	Pro	Leu	Glu	Ile	Thr	Lys	
625					630					635					640	
Glu	Ile	Tyr	Asp	Leu	Asn	Asn	Pro	Glu	Lys	Glu	Pro	Lys	Lys	Phe	Gln	
				645					650					655		
Thr	Ala	Tyr	Ala	Lys	Lys	Thr	Gly	Asp	Gln	Lys	Gly	Tyr	Arg	Glu	Ala	
			660					665						670		
Leu	Cys	Lys	Trp	Ile	Asp	Phe	Thr	Arg	Asp	Phe	Leu	Ser	Lys	Tyr	Thr	
		675						680					685			
Lys	Thr	Thr	Ser	Ile	Asp	Leu	Ser	Ser	Leu	Arg	Pro	Ser	Ser	Gln	Tyr	
	690					695					700					
Lys	Asp	Leu	Gly	Glu	Tyr	Tyr	Ala	Glu	Leu	Asn	Pro	Leu	Leu	Tyr	His	
705					710					715					720	
Ile	Ser	Phe	Gln	Arg	Ile	Ala	Glu	Lys	Glu	Ile	Met	Asp	Ala	Val	Glu	
				725					730					735		
Thr	Gly	Lys	Leu	Tyr	Leu	Phe	Gln	Ile	Tyr	Asn	Lys	Asp	Phe	Ala	Lys	
			740					745						750		
Gly	His	His	Gly	Lys	Pro	Asn	Leu	His	Thr	Leu	Tyr	Trp	Thr	Gly	Leu	
		755					760					765				
Phe	Ser	Pro	Glu	Asn	Leu	Ala	Lys	Thr	Ser	Ile	Lys	Leu	Asn	Gly	Gln	
	770					775					780					
Ala	Glu	Leu	Phe	Tyr	Arg	Pro	Lys	Ser	Arg	Met	Lys	Arg	Met	Ala	His	
785						790					795				800	
Arg	Leu	Gly	Glu	Lys	Met	Leu	Asn	Lys	Lys	Leu	Lys	Asp	Gln	Lys	Thr	
				805					810					815		
Pro	Ile	Pro	Asp	Thr	Leu	Tyr	Gln	Glu	Leu	Tyr	Asp	Tyr	Val	Asn	His	
			820					825					830			
Arg	Leu	Ser	His	Asp	Leu	Ser	Asp	Glu	Ala	Arg	Ala	Leu	Leu	Pro	Asn	
			835				840					845				
Val	Ile	Thr	Lys	Glu	Val	Ser	His	Glu	Ile	Ile	Lys	Asp	Arg	Arg	Phe	
	850					855					860					

ES 2 735 773 T3

Thr	Ser	Asp	Lys	Phe	Phe	Phe	His	Val	Pro	Ile	Thr	Leu	Asn	Tyr	Gln	865	870	875	880
Ala	Ala	Asn	Ser	Pro	Ser	Lys	Phe	Asn	Gln	Arg	Val	Asn	Ala	Tyr	Leu	885	890	895	
Lys	Glu	His	Pro	Glu	Thr	Pro	Ile	Ile	Gly	Ile	Asp	Arg	Gly	Glu	Arg	900	905	910	
Asn	Leu	Ile	Tyr	Ile	Thr	Val	Ile	Asp	Ser	Thr	Gly	Lys	Ile	Leu	Glu	915	920	925	
Gln	Arg	Ser	Leu	Asn	Thr	Ile	Gln	Gln	Phe	Asp	Tyr	Gln	Lys	Lys	Leu	930	935	940	
Asp	Asn	Arg	Glu	Lys	Glu	Arg	Val	Ala	Ala	Arg	Gln	Ala	Trp	Ser	Val	945	950	955	960
Val	Gly	Thr	Ile	Lys	Asp	Leu	Lys	Gln	Gly	Tyr	Leu	Ser	Gln	Val	Ile	965	970	975	
His	Glu	Ile	Val	Asp	Leu	Met	Ile	His	Tyr	Gln	Ala	Val	Val	Val	Leu	980	985	990	
Glu	Asn	Leu	Asn	Phe	Gly	Phe	Lys	Ser	Lys	Arg	Thr	Gly	Ile	Ala	Glu	995	1000	1005	
Lys	Ala	Val	Tyr	Gln	Gln	Phe	Glu	Lys	Met	Leu	Ile	Asp	Lys	Leu		1010	1015	1020	
Asn	Cys	Leu	Val	Leu	Lys	Asp	Tyr	Pro	Ala	Glu	Lys	Val	Gly	Gly		1025	1030	1035	
Val	Leu	Asn	Pro	Tyr	Gln	Leu	Thr	Asp	Gln	Phe	Thr	Ser	Phe	Ala		1040	1045	1050	
Lys	Met	Gly	Thr	Gln	Ser	Gly	Phe	Leu	Phe	Tyr	Val	Pro	Ala	Pro		1055	1060	1065	
Tyr	Thr	Ser	Lys	Ile	Asp	Pro	Leu	Thr	Gly	Phe	Val	Asp	Pro	Phe		1070	1075	1080	
Val	Trp	Lys	Thr	Ile	Lys	Asn	His	Glu	Ser	Arg	Lys	His	Phe	Leu		1085	1090	1095	
Glu	Gly	Phe	Asp	Phe	Leu	His	Tyr	Asp	Val	Lys	Thr	Gly	Asp	Phe		1100	1105	1110	

ES 2 735 773 T3

Ile Leu His Phe Lys Met Asn Arg Asn Leu Ser Phe Gln Arg Gly
1115 1120 1125

Leu Pro Gly Phe Met Pro Ala Trp Asp Ile Val Phe Glu Lys Asn
1130 1135 1140

Glu Thr Gln Phe Asp Ala Lys Gly Thr Pro Phe Ile Ala Gly Lys
1145 1150 1155

Arg Ile Val Pro Val Ile Glu Asn His Arg Phe Thr Gly Arg Tyr
1160 1165 1170

Arg Asp Leu Tyr Pro Ala Asn Glu Leu Ile Ala Leu Leu Glu Glu
1175 1180 1185

Lys Gly Ile Val Phe Arg Asp Gly Ser Asn Ile Leu Pro Lys Leu
1190 1195 1200

Leu Glu Asn Asp Asp Ser His Ala Ile Asp Thr Met Val Ala Leu
1205 1210 1215

Ile Arg Ser Val Leu Gln Met Arg Asn Ser Asn Ala Ala Thr Gly
1220 1225 1230

Glu Asp Tyr Ile Asn Ser Pro Val Arg Asp Leu Asn Gly Val Cys
1235 1240 1245

Phe Asp Ser Arg Phe Gln Asn Pro Glu Trp Pro Met Asp Ala Asp
1250 1255 1260

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 105

<211> 1307

<212> PRT

5 <213> Acidaminococcus sp.

<400> 105

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

ES 2 735 773 T3

Lys Ile Lys Gly Leu Asn Glu Val Leu Asn Leu Ala Ile Gln Lys Asn
 275 280 285
 Asp Glu Thr Ala His Ile Ile Ala Ser Leu Pro His Arg Phe Ile Pro
 290 295 300
 Leu Phe Lys Gln Ile Leu Ser Asp Arg Asn Thr Leu Ser Phe Ile Leu
 305 310 315 320
 Glu Glu Phe Lys Ser Asp Glu Glu Val Ile Gln Ser Phe Cys Lys Tyr
 325 330 335
 Lys Thr Leu Leu Arg Asn Glu Asn Val Leu Glu Thr Ala Glu Ala Leu
 340 345 350
 Phe Asn Glu Leu Asn Ser Ile Asp Leu Thr His Ile Phe Ile Ser His
 355 360 365
 Lys Lys Leu Glu Thr Ile Ser Ser Ala Leu Cys Asp His Trp Asp Thr
 370 375 380
 Leu Arg Asn Ala Leu Tyr Glu Arg Arg Ile Ser Glu Leu Thr Gly Lys
 385 390 395 400
 Ile Thr Lys Ser Ala Lys Glu Lys Val Gln Arg Ser Leu Lys His Glu
 405 410 415
 Asp Ile Asn Leu Gln Glu Ile Ile Ser Ala Ala Gly Lys Glu Leu Ser
 420 425 430
 Glu Ala Phe Lys Gln Lys Thr Ser Glu Ile Leu Ser His Ala His Ala
 435 440 445
 Ala Leu Asp Gln Pro Leu Pro Thr Thr Leu Lys Lys Gln Glu Glu Lys
 450 455 460
 Glu Ile Leu Lys Ser Gln Leu Asp Ser Leu Leu Gly Leu Tyr His Leu
 465 470 475 480
 Leu Asp Trp Phe Ala Val Asp Glu Ser Asn Glu Val Asp Pro Glu Phe
 485 490 495
 Ser Ala Arg Leu Thr Gly Ile Lys Leu Glu Met Glu Pro Ser Leu Ser
 500 505 510
 Phe Tyr Asn Lys Ala Arg Asn Tyr Ala Thr Lys Lys Pro Tyr Ser Val

ES 2 735 773 T3

515		520		525
Glu Lys Phe Lys Leu Asn Phe Gln Met Pro Thr Leu Ala Ser Gly Trp				
530		535		540
Asp Val Asn Lys Glu Lys Asn Asn Gly Ala Ile Leu Phe Val Lys Asn				
545		550		555
Gly Leu Tyr Tyr Leu Gly Ile Met Pro Lys Gln Lys Gly Arg Tyr Lys				
	565		570	575
Ala Leu Ser Phe Glu Pro Thr Glu Lys Thr Ser Glu Gly Phe Asp Lys				
	580		585	590
Met Tyr Tyr Asp Tyr Phe Pro Asp Ala Ala Lys Met Ile Pro Lys Cys				
	595		600	605
Ser Thr Gln Leu Lys Ala Val Thr Ala His Phe Gln Thr His Thr Thr				
	610		615	620
Pro Ile Leu Leu Ser Asn Asn Phe Ile Glu Pro Leu Glu Ile Thr Lys				
	625		630	635
Glu Ile Tyr Asp Leu Asn Asn Pro Glu Lys Glu Pro Lys Lys Phe Gln				
	645		650	655
Thr Ala Tyr Ala Lys Lys Thr Gly Asp Gln Lys Gly Tyr Arg Glu Ala				
	660		665	670
Leu Cys Lys Trp Ile Asp Phe Thr Arg Asp Phe Leu Ser Lys Tyr Thr				
	675		680	685
Lys Thr Thr Ser Ile Asp Leu Ser Ser Leu Arg Pro Ser Ser Gln Tyr				
	690		695	700
Lys Asp Leu Gly Glu Tyr Tyr Ala Glu Leu Asn Pro Leu Leu Tyr His				
	705		710	715
Ile Ser Phe Gln Arg Ile Ala Glu Lys Glu Ile Met Asp Ala Val Glu				
	725		730	735
Thr Gly Lys Leu Tyr Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ala Lys				
	740		745	750
Gly His His Gly Lys Pro Asn Leu His Thr Leu Tyr Trp Thr Gly Leu				
	755		760	765

ES 2 735 773 T3

Phe Ser Pro Glu Asn Leu Ala Lys Thr Ser Ile Lys Leu Asn Gly Gln
 770 775 780
 Ala Glu Leu Phe Tyr Arg Pro Lys Ser Arg Met Lys Arg Met Ala His
 785 790 795 800
 Arg Leu Gly Glu Lys Met Leu Asn Lys Lys Leu Lys Asp Gln Lys Thr
 805 810 815
 Pro Ile Pro Asp Thr Leu Tyr Gln Glu Leu Tyr Asp Tyr Val Asn His
 820 825 830
 Arg Leu Ser His Asp Leu Ser Asp Glu Ala Arg Ala Leu Leu Pro Asn
 835 840 845
 Val Ile Thr Lys Glu Val Ser His Glu Ile Ile Lys Asp Arg Arg Phe
 850 855 860
 Thr Ser Asp Lys Phe Phe Phe His Val Pro Ile Thr Leu Asn Tyr Gln
 865 870 875 880
 Ala Ala Asn Ser Pro Ser Lys Phe Asn Gln Arg Val Asn Ala Tyr Leu
 885 890 895
 Lys Glu His Pro Glu Thr Pro Ile Ile Gly Ile Ala Arg Gly Glu Arg
 900 905 910
 Asn Leu Ile Tyr Ile Thr Val Ile Asp Ser Thr Gly Lys Ile Leu Glu
 915 920 925
 Gln Arg Ser Leu Asn Thr Ile Gln Gln Phe Asp Tyr Gln Lys Lys Leu
 930 935 940
 Asp Asn Arg Glu Lys Glu Arg Val Ala Ala Arg Gln Ala Trp Ser Val
 945 950 955 960
 Val Gly Thr Ile Lys Asp Leu Lys Gln Gly Tyr Leu Ser Gln Val Ile
 965 970 975
 His Glu Ile Val Asp Leu Met Ile His Tyr Gln Ala Val Val Val Leu
 980 985 990
 Ala Asn Leu Asn Phe Gly Phe Lys Ser Lys Arg Thr Gly Ile Ala Glu
 995 1000 1005
 Lys Ala Val Tyr Gln Gln Phe Glu Lys Met Leu Ile Asp Lys Leu
 1010 1015 1020

ES 2 735 773 T3

Asn	Cys	Leu	Val	Leu	Lys	Asp	Tyr	Pro	Ala	Glu	Lys	Val	Gly	Gly
1025						1030					1035			
Val	Leu	Asn	Pro	Tyr	Gln	Leu	Thr	Asp	Gln	Phe	Thr	Ser	Phe	Ala
1040						1045					1050			
Lys	Met	Gly	Thr	Gln	Ser	Gly	Phe	Leu	Phe	Tyr	Val	Pro	Ala	Pro
1055						1060					1065			
Tyr	Thr	Ser	Lys	Ile	Asp	Pro	Leu	Thr	Gly	Phe	Val	Asp	Pro	Phe
1070						1075					1080			
Val	Trp	Lys	Thr	Ile	Lys	Asn	His	Glu	Ser	Arg	Lys	His	Phe	Leu
1085						1090					1095			
Glu	Gly	Phe	Asp	Phe	Leu	His	Tyr	Asp	Val	Lys	Thr	Gly	Asp	Phe
1100						1105					1110			
Ile	Leu	His	Phe	Lys	Met	Asn	Arg	Asn	Leu	Ser	Phe	Gln	Arg	Gly
1115						1120					1125			
Leu	Pro	Gly	Phe	Met	Pro	Ala	Trp	Asp	Ile	Val	Phe	Glu	Lys	Asn
1130						1135					1140			
Glu	Thr	Gln	Phe	Asp	Ala	Lys	Gly	Thr	Pro	Phe	Ile	Ala	Gly	Lys
1145						1150					1155			
Arg	Ile	Val	Pro	Val	Ile	Glu	Asn	His	Arg	Phe	Thr	Gly	Arg	Tyr
1160						1165					1170			
Arg	Asp	Leu	Tyr	Pro	Ala	Asn	Glu	Leu	Ile	Ala	Leu	Leu	Glu	Glu
1175						1180					1185			
Lys	Gly	Ile	Val	Phe	Arg	Asp	Gly	Ser	Asn	Ile	Leu	Pro	Lys	Leu
1190						1195					1200			
Leu	Glu	Asn	Asp	Asp	Ser	His	Ala	Ile	Asp	Thr	Met	Val	Ala	Leu
1205						1210					1215			
Ile	Arg	Ser	Val	Leu	Gln	Met	Ala	Asn	Ser	Asn	Ala	Ala	Thr	Gly
1220						1225					1230			
Glu	Asp	Tyr	Ile	Asn	Ser	Pro	Val	Arg	Asp	Leu	Asn	Gly	Val	Cys
1235						1240					1245			
Phe	Asp	Ser	Arg	Phe	Gln	Asn	Pro	Glu	Trp	Pro	Met	Asp	Ala	Asp
1250						1255					1260			

ES 2 735 773 T3

Ala Asn Gly Ala Tyr His Ile Ala Leu Lys Gly Gln Leu Leu Leu
1265 1270 1275

Asn His Leu Lys Glu Ser Lys Asp Leu Lys Leu Gln Asn Gly Ile
1280 1285 1290

Ser Asn Gln Asp Trp Leu Ala Tyr Ile Gln Glu Leu Arg Asn
1295 1300 1305

<210> 106

<211> 1245

<212> PRT

5 <213> Lachnospiraceae bacterium

<400> 106

ES 2 735 773 T3

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

ES 2 735 773 T3

Glu Asn Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg
 180 185 190
 Cys Ile Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe
 195 200 205
 Glu Lys Val Asp Ala Ile Phe Asp Lys His Glu Val Gln Glu Ile Lys
 210 215 220
 Glu Lys Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly
 225 230 235 240
 Glu Phe Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn
 245 250 255
 Ala Ile Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly
 260 265 270
 Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu
 275 280 285
 Pro Lys Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser
 290 295 300
 Leu Ser Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu
 305 310 315 320
 Val Phe Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile
 325 330 335
 Lys Lys Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala
 340 345 350
 Gly Ile Phe Val Lys Asn Gly Pro Ala Ile Ser Thr Ile Ser Lys Asp
 355 360 365
 Ile Phe Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr
 370 375 380
 Asp Asp Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu
 385 390 395 400
 Asp Asp Arg Arg Lys Ser Phe Lys Lys Ile Gly Ser Phe Ser Leu Glu
 405 410 415
 Gln Leu Gln Glu Tyr Ala Asp Ala Asp Leu Ser Val Val Glu Lys Leu

ES 2 735 773 T3

420					425					430					
Lys	Glu	Ile	Ile	Ile	Gln	Lys	Val	Asp	Glu	Ile	Tyr	Lys	Val	Tyr	Gly
435					440					445					
Ser	Ser	Glu	Lys	Leu	Phe	Asp	Ala	Asp	Phe	Val	Leu	Glu	Lys	Ser	Leu
450			455					460							
Lys	Lys	Asn	Asp	Ala	Val	Val	Ala	Ile	Met	Lys	Asp	Leu	Leu	Asp	Ser
465		470					475			480					
Val	Lys	Ser	Phe	Glu	Asn	Tyr	Ile	Lys	Ala	Phe	Phe	Gly	Glu	Gly	Lys
485					490					495					
Glu	Thr	Asn	Arg	Asp	Glu	Ser	Phe	Tyr	Gly	Asp	Phe	Val	Leu	Ala	Tyr
500			505					510							
Asp	Ile	Leu	Leu	Lys	Val	Asp	His	Ile	Tyr	Asp	Ala	Ile	Arg	Asn	Tyr
515		520					525								
Val	Thr	Gln	Lys	Pro	Tyr	Ser	Lys	Asp	Lys	Phe	Lys	Leu	Tyr	Phe	Gln
530		535					540								
Asn	Pro	Gln	Phe	Met	Gly	Gly	Trp	Asp	Lys	Asp	Lys	Glu	Thr	Asp	Tyr
545		550					555			560					
Arg	Ala	Thr	Ile	Leu	Arg	Tyr	Gly	Ser	Lys	Tyr	Tyr	Leu	Ala	Ile	Met
565					570					575					
Asp	Lys	Lys	Tyr	Ala	Lys	Cys	Leu	Gln	Lys	Ile	Asp	Lys	Asp	Asp	Val
580			585					590							
Asn	Gly	Asn	Tyr	Glu	Lys	Ile	Asn	Tyr	Lys	Leu	Leu	Pro	Gly	Pro	Asn
595		600					605								
Lys	Met	Leu	Pro	Lys	Val	Phe	Phe	Ser	Lys	Lys	Trp	Met	Ala	Tyr	Tyr
610		615					620								
Asn	Pro	Ser	Glu	Asp	Ile	Gln	Lys	Ile	Tyr	Lys	Asn	Gly	Thr	Phe	Lys
625		630					635			640					
Lys	Gly	Asp	Met	Phe	Asn	Leu	Asn	Asp	Cys	His	Lys	Leu	Ile	Asp	Phe
645					650					655					
Phe	Lys	Asp	Ser	Ile	Ser	Arg	Tyr	Pro	Lys	Trp	Ser	Asn	Ala	Tyr	Asp
660			665					670							

ES 2 735 773 T3

Phe Asn Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr
 675 680 685
 Arg Glu Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser
 690 695 700
 Lys Lys Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe
 705 710 715 720
 Gln Ile Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn
 725 730 735
 Leu His Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly
 740 745 750
 Gln Ile Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser
 755 760 765
 Leu Lys Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala
 770 775 780
 Asn Lys Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp
 785 790 795 800
 Val Tyr Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile
 805 810 815
 Pro Ile Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr
 820 825 830
 Glu Val Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly
 835 840 845
 Ile Asp Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly
 850 855 860
 Lys Gly Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn
 865 870 875 880
 Phe Asn Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys
 885 890 895
 Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu
 900 905 910
 Asn Ile Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys
 915 920 925

ES 2 735 773 T3

Ile Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Glu Asp
930 935 940

Leu Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val
945 950 955 960

Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val
965 970 975

Asp Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr
980 985 990

Gln Ile Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn
995 1000 1005

Gly Phe Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp
1010 1015 1020

Pro Ser Thr Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser
1025 1030 1035

Ile Ala Asp Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met
1040 1045 1050

Tyr Val Pro Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys
1055 1060 1065

Asn Phe Ser Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu
1070 1075 1080

Tyr Ser Tyr Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys
1085 1090 1095

Asn Asn Val Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr
1100 1105 1110

Lys Glu Leu Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp
1115 1120 1125

Ile Arg Ala Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser
1130 1135 1140

Ser Phe Met Ala Leu Met Ser Leu Met Leu Gln Met Arg Asn Ser
1145 1150 1155

Ile Thr Gly Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys
1160 1165 1170

ES 2 735 773 T3

Asn Ser Asp Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln
1175 1180 1185

Glu Asn Ala Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr
1190 1195 1200

Asn Ile Ala Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys
1205 1210 1215

Ala Glu Asp Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn
1220 1225 1230

Lys Glu Trp Leu Glu Tyr Ala Gln Thr Ser Val Lys
1235 1240 1245

<210> 107

<211> 1246

<212> PRT

5 <213> Lachnospiraceae bacterium

<400> 107

ES 2 735 773 T3

Met	Leu	Lys	Asn	Val	Gly	Ile	Asp	Arg	Leu	Asp	Val	Glu	Lys	Gly	Arg	1	5	10	15
Lys	Asn	Met	Ser	Lys	Leu	Glu	Lys	Phe	Thr	Asn	Cys	Tyr	Ser	Leu	Ser	20	25	30	
Lys	Thr	Leu	Arg	Phe	Lys	Ala	Ile	Pro	Val	Gly	Lys	Thr	Gln	Glu	Asn	35	40	45	
Ile	Asp	Asn	Lys	Arg	Leu	Leu	Val	Glu	Asp	Glu	Lys	Arg	Ala	Glu	Asp	50	55	60	
Tyr	Lys	Gly	Val	Lys	Lys	Leu	Leu	Asp	Arg	Tyr	Tyr	Leu	Ser	Phe	Ile	65	70	75	80
Asn	Asp	Val	Leu	His	Ser	Ile	Lys	Leu	Lys	Asn	Leu	Asn	Asn	Tyr	Ile	85	90	95	
Ser	Leu	Phe	Arg	Lys	Lys	Thr	Arg	Thr	Glu	Lys	Glu	Asn	Lys	Glu	Leu	100	105	110	
Glu	Asn	Leu	Glu	Ile	Asn	Leu	Arg	Lys	Glu	Ile	Ala	Lys	Ala	Phe	Lys	115	120	125	
Gly	Asn	Glu	Gly	Tyr	Lys	Ser	Leu	Phe	Lys	Lys	Asp	Ile	Ile	Glu	Thr	130	135	140	

ES 2 735 773 T3

Ile Leu Pro Glu Phe Leu Asp Asp Lys Asp Glu Ile Ala Leu Val Asn
145 150 155 160

Ser Phe Asn Gly Phe Thr Thr Ala Phe Thr Gly Phe Phe Asp Asn Arg
165 170 175

Glu Asn Met Phe Ser Glu Glu Ala Lys Ser Thr Ser Ile Ala Phe Arg
180 185 190

Cys Ile Asn Glu Asn Leu Thr Arg Tyr Ile Ser Asn Met Asp Ile Phe
195 200 205

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

Glu Lys Ile Leu Asn Ser Asp Tyr Asp Val Glu Asp Phe Phe Glu Gly
225 230 235 240

Glu Phe Phe Asn Phe Val Leu Thr Gln Glu Gly Ile Asp Val Tyr Asn
245 250 255

Ala Ile Ile Gly Gly Phe Val Thr Glu Ser Gly Glu Lys Ile Lys Gly
260 265 270

Leu Asn Glu Tyr Ile Asn Leu Tyr Asn Gln Lys Thr Lys Gln Lys Leu
275 280 285

Pro Lys Phe Lys Pro Leu Tyr Lys Gln Val Leu Ser Asp Arg Glu Ser
290 295 300

Leu Ser Phe Tyr Gly Glu Gly Tyr Thr Ser Asp Glu Glu Val Leu Glu
305 310 315 320

Val Phe Arg Asn Thr Leu Asn Lys Asn Ser Glu Ile Phe Ser Ser Ile
325 330 335

Lys Lys Leu Glu Lys Leu Phe Lys Asn Phe Asp Glu Tyr Ser Ser Ala
340 345 350

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

Ile Phe Gly Glu Trp Asn Val Ile Arg Asp Lys Trp Asn Ala Glu Tyr
370 375 380

Asp Asp Ile His Leu Lys Lys Lys Ala Val Val Thr Glu Lys Tyr Glu

ES 2 735 773 T3

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

ES 2 735 773 T3

Lys Gly Asp Met Phe Asn Leu Asn Asp Cys His Lys Leu Ile Asp Phe
 645 650 655
 Phe Lys Asp Ser Ile Ser Arg Tyr Pro Lys Trp Ser Asn Ala Tyr Asp
 660 665 670
 Phe Asn Phe Ser Glu Thr Glu Lys Tyr Lys Asp Ile Ala Gly Phe Tyr
 675 680 685
 Arg Glu Val Glu Glu Gln Gly Tyr Lys Val Ser Phe Glu Ser Ala Ser
 690 695 700
 Lys Lys Glu Val Asp Lys Leu Val Glu Glu Gly Lys Leu Tyr Met Phe
 705 710 715 720
 Gln Ile Tyr Asn Lys Asp Phe Ser Asp Lys Ser His Gly Thr Pro Asn
 725 730 735
 Leu His Thr Met Tyr Phe Lys Leu Leu Phe Asp Glu Asn Asn His Gly
 740 745 750
 Gln Ile Arg Leu Ser Gly Gly Ala Glu Leu Phe Met Arg Arg Ala Ser
 755 760 765
 Leu Lys Lys Glu Glu Leu Val Val His Pro Ala Asn Ser Pro Ile Ala
 770 775 780
 Asn Lys Asn Pro Asp Asn Pro Lys Lys Thr Thr Thr Leu Ser Tyr Asp
 785 790 795 800
 Val Tyr Lys Asp Lys Arg Phe Ser Glu Asp Gln Tyr Glu Leu His Ile
 805 810 815
 Pro Ile Ala Ile Asn Lys Cys Pro Lys Asn Ile Phe Lys Ile Asn Thr
 820 825 830
 Glu Val Arg Val Leu Leu Lys His Asp Asp Asn Pro Tyr Val Ile Gly
 835 840 845
 Ile Ala Arg Gly Glu Arg Asn Leu Leu Tyr Ile Val Val Val Asp Gly
 850 855 860
 Lys Gly Asn Ile Val Glu Gln Tyr Ser Leu Asn Glu Ile Ile Asn Asn
 865 870 875 880
 Phe Asn Gly Ile Arg Ile Lys Thr Asp Tyr His Ser Leu Leu Asp Lys
 885 890 895

ES 2 735 773 T3

Lys Glu Lys Glu Arg Phe Glu Ala Arg Gln Asn Trp Thr Ser Ile Glu
 900 905 910
 Asn Ile Lys Glu Leu Lys Ala Gly Tyr Ile Ser Gln Val Val His Lys
 915 920 925
 Ile Cys Glu Leu Val Glu Lys Tyr Asp Ala Val Ile Ala Leu Ala Asp
 930 935 940
 Leu Asn Ser Gly Phe Lys Asn Ser Arg Val Lys Val Glu Lys Gln Val
 945 950 955 960
 Tyr Gln Lys Phe Glu Lys Met Leu Ile Asp Lys Leu Asn Tyr Met Val
 965 970 975
 Asp Lys Lys Ser Asn Pro Cys Ala Thr Gly Gly Ala Leu Lys Gly Tyr
 980 985 990
 Gln Ile Thr Asn Lys Phe Glu Ser Phe Lys Ser Met Ser Thr Gln Asn
 995 1000 1005
 Gly Phe Ile Phe Tyr Ile Pro Ala Trp Leu Thr Ser Lys Ile Asp
 1010 1015 1020
 Pro Ser Thr Gly Phe Val Asn Leu Leu Lys Thr Lys Tyr Thr Ser
 1025 1030 1035
 Ile Ala Asp Ser Lys Lys Phe Ile Ser Ser Phe Asp Arg Ile Met
 1040 1045 1050
 Tyr Val Pro Glu Glu Asp Leu Phe Glu Phe Ala Leu Asp Tyr Lys
 1055 1060 1065
 Asn Phe Ser Arg Thr Asp Ala Asp Tyr Ile Lys Lys Trp Lys Leu
 1070 1075 1080
 Tyr Ser Tyr Gly Asn Arg Ile Arg Ile Phe Arg Asn Pro Lys Lys
 1085 1090 1095
 Asn Asn Val Phe Asp Trp Glu Glu Val Cys Leu Thr Ser Ala Tyr
 1100 1105 1110
 Lys Glu Leu Phe Asn Lys Tyr Gly Ile Asn Tyr Gln Gln Gly Asp
 1115 1120 1125
 Ile Arg Ala Leu Leu Cys Glu Gln Ser Asp Lys Ala Phe Tyr Ser
 1130 1135 1140

ES 2 735 773 T3

Ser Phe Met Ala Leu Met Ser Leu Met Leu Gln Met Ala Asn Ser
1145 1150 1155

Ile Thr Gly Arg Thr Asp Val Asp Phe Leu Ile Ser Pro Val Lys
1160 1165 1170

Asn Ser Asp Gly Ile Phe Tyr Asp Ser Arg Asn Tyr Glu Ala Gln
1175 1180 1185

Glu Asn Ala Ile Leu Pro Lys Asn Ala Asp Ala Asn Gly Ala Tyr
1190 1195 1200

Asn Ile Ala Arg Lys Val Leu Trp Ala Ile Gly Gln Phe Lys Lys
1205 1210 1215

Ala Glu Asp Glu Lys Leu Asp Lys Val Lys Ile Ala Ile Ser Asn
1220 1225 1230

Lys Glu Trp Leu Glu Tyr Ala Gln Thr Ser Val Lys His
1235 1240 1245

<210> 108

<211> 1121

<212> PRT

5 <213> Streptococcus thermophilus

<400> 108

ES 2 735 773 T3

Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1 5 10 15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
20 25 30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
35 40 45

Asn Arg Gln Gly Arg Arg Leu Thr Arg Arg Lys Lys His Arg Arg Val
50 55 60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
65 70 75 80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
85 90 95

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

ES 2 735 773 T3

Val	Lys	His	Arg	Gly	Ile	Ser	Tyr	Leu	Asp	Asp	Ala	Ser	Asp	Asp	Gly
		115					120					125			
Asn	Ser	Ser	Ile	Gly	Asp	Tyr	Ala	Gln	Ile	Val	Lys	Glu	Asn	Ser	Lys
	130					135					140				
Gln	Leu	Glu	Thr	Lys	Thr	Pro	Gly	Gln	Ile	Gln	Leu	Glu	Arg	Tyr	Gln
145					150					155					160
Thr	Tyr	Gly	Gln	Leu	Arg	Gly	Asp	Phe	Thr	Val	Glu	Lys	Asp	Gly	Lys
				165					170					175	
Lys	His	Arg	Leu	Ile	Asn	Val	Phe	Pro	Thr	Ser	Ala	Tyr	Arg	Ser	Glu
			180					185					190		
Ala	Leu	Arg	Ile	Leu	Gln	Thr	Gln	Gln	Glu	Phe	Asn	Pro	Gln	Ile	Thr
		195					200					205			
Asp	Glu	Phe	Ile	Asn	Arg	Tyr	Leu	Glu	Ile	Leu	Thr	Gly	Lys	Arg	Lys
	210					215					220				
Tyr	Tyr	His	Gly	Pro	Gly	Asn	Glu	Lys	Ser	Arg	Thr	Asp	Tyr	Gly	Arg
225					230					235					240
Tyr	Arg	Thr	Ser	Gly	Glu	Thr	Leu	Asp	Asn	Ile	Phe	Gly	Ile	Leu	Ile
				245					250					255	
Gly	Lys	Cys	Thr	Phe	Tyr	Pro	Asp	Glu	Phe	Arg	Ala	Ala	Lys	Ala	Ser
			260					265					270		
Tyr	Thr	Ala	Gln	Glu	Phe	Asn	Leu	Leu	Asn	Asp	Leu	Asn	Asn	Leu	Thr
		275					280					285			
Val	Pro	Thr	Glu	Thr	Lys	Lys	Leu	Ser	Lys	Glu	Gln	Lys	Asn	Gln	Ile
	290					295					300				
Ile	Asn	Tyr	Val	Lys	Asn	Glu	Lys	Ala	Met	Gly	Pro	Ala	Lys	Leu	Phe
305					310					315					320
Lys	Tyr	Ile	Ala	Lys	Leu	Leu	Ser	Cys	Asp	Val	Ala	Asp	Ile	Lys	Gly
				325					330					335	
Tyr	Arg	Ile	Asp	Lys	Ser	Gly	Lys	Ala	Glu	Ile	His	Thr	Phe	Glu	Ala
			340					345					350		
Tyr	Arg	Lys	Met	Lys	Thr	Leu	Glu	Thr	Leu	Asp	Ile	Glu	Gln	Met	Asp

ES 2 735 773 T3

355		360		365
Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu	370	375	380	
Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser	385	390	395	400
Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn	405	410	415	
Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met	420	425	430	
Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr	435	440	445	
Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys	450	455	460	
Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro	465	470	475	480
Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala	485	490	495	
Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg	500	505	510	
Glu Thr Asn Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys	515	520	525	
Ala Asn Lys Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln	530	535	540	
Tyr Asn Gly Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys	545	550	555	560
Gln Leu Ala Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys	565	570	575	
Leu Tyr Thr Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser	580	585	590	
Asn Gln Phe Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp	595	600	605	

ES 2 735 773 T3

Asp	Ser	Leu	Ala	Asn	Lys	Val	Leu	Val	Tyr	Ala	Thr	Ala	Asn	Gln	Glu	610	615	620	
Lys	Gly	Gln	Arg	Thr	Pro	Tyr	Gln	Ala	Leu	Asp	Ser	Met	Asp	Asp	Ala	625	630	635	640
Trp	Ser	Phe	Arg	Glu	Leu	Lys	Ala	Phe	Val	Arg	Glu	Ser	Lys	Thr	Leu	645	650	655	
Ser	Asn	Lys	Lys	Lys	Glu	Tyr	Leu	Leu	Thr	Glu	Glu	Asp	Ile	Ser	Lys	660	665	670	
Phe	Asp	Val	Arg	Lys	Lys	Phe	Ile	Glu	Arg	Asn	Leu	Val	Asp	Thr	Arg	675	680	685	
Tyr	Ala	Ser	Arg	Val	Val	Leu	Asn	Ala	Leu	Gln	Glu	His	Phe	Arg	Ala	690	695	700	
His	Lys	Ile	Asp	Thr	Lys	Val	Ser	Val	Val	Arg	Gly	Gln	Phe	Thr	Ser	705	710	715	720
Gln	Leu	Arg	Arg	His	Trp	Gly	Ile	Glu	Lys	Thr	Arg	Asp	Thr	Tyr	His	725	730	735	
His	His	Ala	Val	Asp	Ala	Leu	Ile	Ile	Ala	Ala	Ser	Ser	Gln	Leu	Asn	740	745	750	
Leu	Trp	Lys	Lys	Gln	Lys	Asn	Thr	Leu	Val	Ser	Tyr	Ser	Glu	Asp	Gln	755	760	765	
Leu	Leu	Asp	Ile	Glu	Thr	Gly	Glu	Leu	Ile	Ser	Asp	Asp	Glu	Tyr	Lys	770	775	780	
Glu	Ser	Val	Phe	Lys	Ala	Pro	Tyr	Gln	His	Phe	Val	Asp	Thr	Leu	Lys	785	790	795	800
Ser	Lys	Glu	Phe	Glu	Asp	Ser	Ile	Leu	Phe	Ser	Tyr	Gln	Val	Asp	Ser	805	810	815	
Lys	Phe	Asn	Arg	Lys	Ile	Ser	Asp	Ala	Thr	Ile	Tyr	Ala	Thr	Arg	Gln	820	825	830	
Ala	Lys	Val	Gly	Lys	Asp	Lys	Ala	Asp	Glu	Thr	Tyr	Val	Leu	Gly	Lys	835	840	845	
Ile	Lys	Asp	Ile	Tyr	Thr	Gln	Asp	Gly	Tyr	Asp	Ala	Phe	Met	Lys	Ile	850	855	860	

ES 2 735 773 T3

Tyr	Lys	Lys	Asp	Lys	Ser	Lys	Phe	Leu	Met	Tyr	Arg	His	Asp	Pro	Gln	
865					870					875					880	
Thr	Phe	Glu	Lys	Val	Ile	Glu	Pro	Ile	Leu	Glu	Asn	Tyr	Pro	Asn	Lys	
				885					890					895		
Gln	Ile	Asn	Glu	Lys	Gly	Lys	Glu	Val	Pro	Cys	Asn	Pro	Phe	Leu	Lys	
			900					905					910			
Tyr	Lys	Glu	Glu	His	Gly	Tyr	Ile	Arg	Lys	Tyr	Ser	Lys	Lys	Gly	Asn	
		915					920					925				
Gly	Pro	Glu	Ile	Lys	Ser	Leu	Lys	Tyr	Tyr	Asp	Ser	Lys	Leu	Gly	Asn	
	930					935					940					
His	Ile	Asp	Ile	Thr	Pro	Lys	Asp	Ser	Asn	Asn	Lys	Val	Val	Leu	Gln	
945					950					955					960	
Ser	Val	Ser	Pro	Trp	Arg	Ala	Asp	Val	Tyr	Phe	Asn	Lys	Thr	Thr	Gly	
				965					970					975		
Lys	Tyr	Glu	Ile	Leu	Gly	Leu	Lys	Tyr	Ala	Asp	Leu	Gln	Phe	Glu	Lys	
			980					985					990			
Gly	Thr	Gly	Thr	Tyr	Lys	Ile	Ser	Gln	Glu	Lys	Tyr	Asn	Asp	Ile	Lys	
		995					1000					1005				
Lys	Lys	Glu	Gly	Val	Asp	Ser	Asp	Ser	Glu	Phe	Lys	Phe	Thr	Leu		
	1010					1015					1020					
Tyr	Lys	Asn	Asp	Leu	Leu	Leu	Val	Lys	Asp	Thr	Glu	Thr	Lys	Glu		
	1025					1030					1035					
Gln	Gln	Leu	Phe	Arg	Phe	Leu	Ser	Arg	Thr	Met	Pro	Lys	Gln	Lys		
	1040					1045					1050					
His	Tyr	Val	Glu	Leu	Lys	Pro	Tyr	Asp	Lys	Gln	Lys	Phe	Glu	Gly		
	1055					1060					1065					
Gly	Glu	Ala	Leu	Ile	Lys	Val	Leu	Gly	Asn	Val	Ala	Asn	Ser	Gly		
	1070					1075					1080					
Gln	Cys	Lys	Lys	Gly	Leu	Gly	Lys	Ser	Asn	Ile	Ser	Ile	Tyr	Lys		
	1085					1090					1095					
Val	Arg	Thr	Asp	Val	Leu	Gly	Asn	Gln	His	Ile	Ile	Lys	Asn	Glu		
	1100					1105					1110					
Gly	Asp	Lys	Pro	Lys	Leu	Asp	Phe									
	1115					1120										

<210> 109

<211> 1121

<212> PRT

<213> Streptococcus thermophilus

5 <400> 109

ES 2 735 773 T3

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

ES 2 735 773 T3

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
 210 215 220
 Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
 225 230 235 240
 Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
 245 250 255
 Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
 260 265 270
 Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
 275 280 285
 Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
 290 295 300
 Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
 305 310 315 320
 Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
 325 330 335
 Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
 340 345 350
 Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
 355 360 365
 Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
 370 375 380
 Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
 385 390 395 400
 Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
 405 410 415
 Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
 420 425 430
 Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
 435 440 445
 Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys

ES 2 735 773 T3

450					455					460					
Thr	Lys	Tyr	Ile	Asp	Glu	Lys	Leu	Leu	Thr	Glu	Glu	Ile	Tyr	Asn	Pro
465					470					475					480
Val	Val	Ala	Lys	Ser	Val	Arg	Gln	Ala	Ile	Lys	Ile	Val	Asn	Ala	Ala
				485					490					495	
Ile	Lys	Glu	Tyr	Gly	Asp	Phe	Asp	Asn	Ile	Val	Ile	Glu	Met	Ala	Arg
			500					505					510		
Glu	Thr	Asn	Glu	Asp	Asp	Glu	Lys	Lys	Ala	Ile	Gln	Lys	Ile	Gln	Lys
		515					520					525			
Ala	Asn	Lys	Asp	Glu	Lys	Asp	Ala	Ala	Met	Leu	Lys	Ala	Ala	Asn	Gln
	530					535					540				
Tyr	Asn	Gly	Lys	Ala	Glu	Leu	Pro	His	Ser	Val	Phe	His	Gly	His	Lys
545					550					555					560
Gln	Leu	Ala	Thr	Lys	Ile	Arg	Leu	Trp	His	Gln	Gln	Gly	Glu	Arg	Cys
				565					570					575	
Leu	Tyr	Thr	Gly	Lys	Thr	Ile	Ser	Ile	His	Asp	Leu	Ile	Asn	Asn	Ser
			580					585					590		
Asn	Gln	Phe	Glu	Val	Asp	Ala	Ile	Leu	Pro	Leu	Ser	Ile	Thr	Phe	Asp
		595					600					605			
Asp	Ser	Leu	Ala	Asn	Lys	Val	Leu	Val	Tyr	Ala	Thr	Ala	Asn	Gln	Glu
	610					615					620				
Lys	Gly	Gln	Arg	Thr	Pro	Tyr	Gln	Ala	Leu	Asp	Ser	Met	Asp	Asp	Ala
625					630					635					640
Trp	Ser	Phe	Arg	Glu	Leu	Lys	Ala	Phe	Val	Arg	Glu	Ser	Lys	Thr	Leu
				645					650					655	
Ser	Asn	Lys	Lys	Lys	Glu	Tyr	Leu	Leu	Thr	Glu	Glu	Asp	Ile	Ser	Lys
			660					665					670		
Phe	Asp	Val	Arg	Lys	Lys	Phe	Ile	Glu	Arg	Asn	Leu	Val	Asp	Thr	Arg
		675					680					685			
Tyr	Ala	Ser	Arg	Val	Val	Leu	Asn	Ala	Leu	Gln	Glu	His	Phe	Arg	Ala
	690					695					700				

ES 2 735 773 T3

His 705	Lys	Ile	Asp	Thr	Lys 710	Val	Ser	Val	Val	Arg 715	Gly	Gln	Phe	Thr	Ser 720
Gln	Leu	Arg	Arg	His 725	Trp	Gly	Ile	Glu	Lys 730	Thr	Arg	Asp	Thr	Tyr 735	His
His	His	Ala	Val	Asp 740	Ala	Leu	Ile	Ile	Ala 745	Ala	Ser	Ser	Gln 750	Leu	Asn
Leu	Trp	Lys 755	Lys	Gln	Lys	Asn	Thr 760	Leu	Val	Ser	Tyr	Ser 765	Glu	Asp	Gln
Leu 770	Leu	Asp	Ile	Glu	Thr	Gly 775	Glu	Leu	Ile	Ser	Asp 780	Asp	Glu	Tyr	Lys
Glu 785	Ser	Val	Phe	Lys 790	Ala	Pro	Tyr	Gln	His	Phe 795	Val	Asp	Thr	Leu	Lys 800
Ser	Lys	Glu	Phe	Glu 805	Asp	Ser	Ile	Leu	Phe 810	Ser	Tyr	Gln	Val	Asp 815	Ser
Lys	Phe	Asn 820	Arg	Lys	Ile	Ser	Asp	Ala 825	Thr	Ile	Tyr	Ala	Thr 830	Arg	Gln
Ala	Lys	Val 835	Gly	Lys	Asp	Lys	Ala 840	Asp	Glu	Thr	Tyr	Val 845	Leu	Gly	Lys
Ile 850	Lys	Asp	Ile	Tyr	Thr	Gln 855	Asp	Gly	Tyr	Asp	Ala 860	Phe	Met	Lys	Ile
Tyr 865	Lys	Lys	Asp	Lys	Ser 870	Lys	Phe	Leu	Met	Tyr 875	Arg	His	Asp	Pro	Gln 880
Thr	Phe	Glu	Lys	Val 885	Ile	Glu	Pro	Ile	Leu 890	Glu	Asn	Tyr	Pro	Asn 895	Lys
Gln	Ile	Asn 900	Glu	Lys	Gly	Lys	Glu	Val 905	Pro	Cys	Asn	Pro	Phe 910	Leu	Lys
Tyr	Lys	Glu 915	Glu	His	Gly	Tyr	Ile 920	Arg	Lys	Tyr	Ser	Lys 925	Lys	Gly	Asn
Gly 930	Pro	Glu	Ile	Lys	Ser	Leu 935	Lys	Tyr	Tyr	Asp	Ser 940	Lys	Leu	Gly	Asn
His 945	Ile	Asp	Ile	Thr 950	Pro	Lys	Asp	Ser	Asn	Asn 955	Lys	Val	Val	Leu	Gln 960

ES 2 735 773 T3

Ser Val Ser Pro Trp Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly
965 970 975

Lys Tyr Glu Ile Leu Gly Leu Lys Tyr Ala Asp Leu Gln Phe Glu Lys
980 985 990

Gly Thr Gly Thr Tyr Lys Ile Ser Gln Glu Lys Tyr Asn Asp Ile Lys
995 1000 1005

Lys Lys Glu Gly Val Asp Ser Asp Ser Glu Phe Lys Phe Thr Leu
1010 1015 1020

Tyr Lys Asn Asp Leu Leu Leu Val Lys Asp Thr Glu Thr Lys Glu
1025 1030 1035

Gln Gln Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln Lys
1040 1045 1050

His Tyr Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu Gly
1055 1060 1065

Gly Glu Ala Leu Ile Lys Val Leu Gly Asn Val Ala Asn Ser Gly
1070 1075 1080

Gln Cys Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr Lys
1085 1090 1095

Val Arg Thr Asp Val Leu Gly Asn Gln His Ile Ile Lys Asn Glu
1100 1105 1110

Gly Asp Lys Pro Lys Leu Asp Phe
1115 1120

REIVINDICACIONES

1. Una composición de dos o más secuencias de ácidos nucleicos manipuladas que forman un andamio ("NASC"), comprendiendo la composición NASC:

- 5 un primer componente de ácidos nucleicos manipulado ("NASC-PC1") que comprende, en una dirección de 5' a 3',
- un elemento espaciador 1 que comprende una secuencia 1 de unión a diana de ácidos nucleicos,
- un elemento de repetición 1 que comprende una secuencia 1 de ácidos nucleicos de repetición, y
- un elemento de unión a la proteína de unión a ácidos nucleicos 1 que comprende una secuencia 1 de unión a la proteína de unión a ácidos nucleicos de doble cadena,
- 10 en la que el elemento espaciador 1 está conectado covalentemente con el elemento de repetición 1, y el elemento de repetición 1 está conectado covalentemente con el elemento de unión 1 a la proteína de unión a ácidos nucleicos; y
- un segundo componente de ácidos nucleicos manipulado ("NASC-PC2") que comprende, en una dirección de 5' a 3',
- 15 un elemento espaciador 2 que comprende una secuencia 2 de unión a diana de ácidos nucleicos,
- un elemento de repetición 2 que comprende una secuencia 1C de ácidos nucleicos de repetición, y
- un elemento de unión 2 a la proteína de unión a ácidos nucleicos que comprende una secuencia 2 de unión a la proteína de unión a ácidos nucleicos de doble cadena,
- 20 en la que el elemento espaciador 2 está conectado covalentemente con el elemento de repetición 2, y el elemento de repetición 2 está conectado covalentemente con el elemento 2 de unión a la proteína de unión a ácidos nucleicos;
- en la que hay una conexión entre la secuencia 1 de ácidos nucleicos de repetición y la secuencia 1C de ácidos nucleicos de repetición a través de pares de bases con enlaces de hidrógeno, la conexión forma la composición NASC, y la composición NASC es capaz de unirse a una primera proteína CRISPR-Cas9 Clase 2 Tipo II y una segunda proteína CRISPR-Cas9 Clase 2 Tipo II.
- 25

2. La composición NASC de la reivindicación 1,

- en la que la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II son la misma proteína CRISPR-Cas9 Clase 2 Tipo II, o
- 30 en la que la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II son proteínas ortólogas CRISPR-Cas9 Clase 2 Tipo II.

3. La composición NASC de cualquier reivindicación precedente, en la que

- el elemento espaciador 1 comprende además una secuencia 3' de ácidos nucleicos del elemento enlazador de la secuencia de unión 1 a la diana del ácido nucleico y 5' del elemento 1 de repetición; y
- 35 el elemento espaciador 2 comprende además una secuencia 3' de ácidos nucleicos del elemento enlazador de la secuencia de unión 2 a la diana de ácidos nucleicos y 5' del elemento 2 de repetición.

4. La composición NASC de la reivindicación 3, en la que

- la secuencia 1 de ácidos nucleicos de repetición comprende, además, en una dirección de 5' a 3', una secuencia 1b de ácidos nucleicos de repetición, una secuencia 1-2 de ácidos nucleicos del elemento enlazador y una secuencia 1a de ácidos nucleicos de repetición; y
- 40 la secuencia 1C de ácidos nucleicos de repetición comprende, además, en una dirección de 5' a 3', una secuencia 1aC de ácidos nucleicos de repetición, una secuencia 2-2 de ácidos nucleicos del elemento enlazador y una secuencia 1bC de ácidos nucleicos de repetición;
- en el que la secuencia 1b de ácidos nucleicos de repetición y la secuencia 1bC de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno, y la secuencia 1a de ácidos nucleicos de repetición y la secuencia 1aC de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno.
- 45

5. La composición NASC de la reivindicación 4, en la que

la secuencia 1b de repetición de ácidos nucleicos comprende, además, en una dirección de 5' a 3',
 una secuencia 1b2 de ácidos nucleicos de repetición,
 una secuencia 1b1 de ácidos nucleicos con una protuberancia, y
 una secuencia 1b1 de ácidos nucleicos de repetición;

5 la secuencia 1a de ácidos nucleicos de repetición comprende, además, en una dirección de 5' a 3',
 una secuencia 1a2 de ácidos nucleicos de repetición,
 una secuencia 1a1 de ácidos nucleicos con una protuberancia, y
 una secuencia 1a1 de ácidos nucleicos de repetición;

la secuencia 1aC de ácidos nucleicos de repetición comprende, además, en una dirección de 5' a 3',
 10 una secuencia 1a1C de ácidos nucleicos de repetición,
 una secuencia 2a2 de ácidos nucleicos con una protuberancia, y
 una secuencia 1a2C de ácidos nucleicos de repetición; y

la secuencia 1bC de ácidos nucleicos de repetición comprende, además, en una dirección de 5' a 3',
 una secuencia 1b1C de ácidos nucleicos de repetición,
 15 una secuencia 2b2 de ácidos nucleicos con una protuberancia, y
 una secuencia 1b2C de ácidos nucleicos de repetición;

20 en la que la secuencia 1a1 de ácidos nucleicos de repetición y la secuencia 1a1C de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno, la secuencia 1a2 de ácidos nucleicos de repetición y la secuencia 1a2C de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno, la secuencia 1b1 de ácidos nucleicos de repetición y la secuencia 1b1C de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno, y la secuencia 1b2 de ácidos nucleicos de repetición y la secuencia 1b2C de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno.

6. La composición NASC de la reivindicación 5, en la que

25 la secuencia 1-2 de ácidos nucleicos del elemento enlazador comprende, además, en una dirección de 5' a 3',
 una secuencia 1-2-2 de ácidos nucleicos del elemento enlazador,
 una secuencia 1-2a de ácidos nucleicos de repetición, y
 una secuencia 1-2 de ácidos nucleicos del elemento enlazador-1;

la secuencia 2-2 de ácidos nucleicos del elemento enlazador comprende, además, en una dirección de 5' a 3',
 30 una secuencia 2-2-1 de ácidos nucleicos de elemento enlazador
 una secuencia 1-2aC de ácidos nucleicos de repetición, y
 una secuencia 2-2-2 de ácidos nucleicos del elemento enlazador;

35 en la que la secuencia 1-2a de ácidos nucleicos de repetición y la secuencia 1-2aC de ácidos nucleicos de repetición están conectadas a través de pares de bases con enlaces de hidrógeno y forman una región 1-2 de ácidos nucleicos de doble cadena.

7. La composición NASC de la reivindicación 6, en la que

la región 1-2 de ácidos nucleicos de doble cadena comprende además un sitio de unión a la proteína efectora;
 40 la secuencia 1-2a de ácidos nucleicos de repetición comprende además una secuencia 1-2a de ácidos nucleicos del sitio de unión a la proteína efectora; y

la secuencia 1-2aC de ácidos nucleicos de repetición comprende además una secuencia 1-2aC de ácidos nucleicos del sitio de unión a la proteína efectora;

en la que un sitio de unión del efector se forma mediante una unión de par de bases de hidrógeno entre la secuencia 1-2a de ácidos nucleicos del sitio de unión de la proteína efectora y la secuencia 1-2aC de ácidos nucleicos del sitio de unión de la proteína efectora, y opcionalmente

en la que el sitio de unión a la proteína efectora es un sitio de unión a la proteína Csy4.

8. La composición NASC de una cualquiera de las reivindicaciones 1 a 7, en la que

la secuencia 1 de ácidos nucleicos de repetición comprende además una etiqueta de afinidad 1; y

la secuencia 1C de ácidos nucleicos de repetición comprende además una etiqueta de afinidad 2;

en la que la etiqueta de afinidad 1 está conectada con la etiqueta de afinidad 2.

9. La composición NASC de cualquier reivindicación precedente, en la que el NASC-PC1 comprende ARN, ADN o ARN y ADN; y/o

en la que el NASC-PC2 comprende ARN, ADN o ARN y ADN.

10. La composición NASC de cualquier reivindicación precedente, que comprende además un polinucleótido donante.

11. Una composición de ácidos nucleicos/proteína, que comprende:

la composición NASC de cualquier reivindicación precedente; y

una primera proteína CRISPR-Cas9 Clase 2 Tipo II y una segunda proteína CRISPR-Cas9 Clase 2 Tipo II.

12. La composición de ácidos nucleicos/proteína de la reivindicación 11,

en la que la primera proteína CRISPR-Cas9 Clase 2 Tipo II es la misma que la segunda proteína CRISPR-Cas9 Clase 2 Tipo II, y la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II son seleccionadas del grupo que consiste en una proteína Cas9 de *Streptococcus pyogenes*, una proteína Cas9 de *Streptococcus thermophilus*, una proteína Cas9 de *Staphylococcus aureus* y una proteína Cas9 de *Campylobacter jejuni*; o

en la que la primera proteína CRISPR-Cas9 Clase 2 Tipo II es diferente de la segunda proteína CRISPR-Cas9 Clase 2 Tipo II, y la primera proteína CRISPR-Cas9 Clase 2 Tipo II y la segunda proteína CRISPR-Cas9 Clase 2 Tipo II se seleccionan de el grupo que consiste en una proteína Cas9 de *Streptococcus pyogenes*, una proteína Cas9 de *Streptococcus thermophilus*, una proteína Cas9 de *Staphylococcus aureus* y una proteína Cas9 de *Campylobacter jejuni*.

13. La composición de ácidos nucleicos/proteína de cualquiera de las reivindicaciones 11 o 12, en la que la primera proteína CRISPR-Cas9 Tipo II de Clase 2 y la segunda proteína CRISPR-Cas9 Clase II de Tipo 2 se seleccionan del grupo que consiste en la proteína Cas9/proteína Cas9, Proteína Cas9/proteína dCas9, proteína dCas9/proteína Cas9 y proteína dCas9/proteína dCas9, respectivamente.

14. Un kit, que comprende:

la composición NASC de una cualquiera de las reivindicaciones 1-10, o una o más secuencias de ácidos nucleicos que codifican la composición NASC de una cualquiera de las reivindicaciones 1-10; y

un tampón.

15. El kit de la reivindicación 14, que comprende, además:

una o más proteínas CRISPR-Cas9 Clase 2 Tipo II o una o más secuencias de ácidos nucleicos que codifican la una o más proteínas CRISPR-Cas9 Clase 2 Tipo II, o

complejos de nucleoproteínas que comprenden la composición NASC y una o más proteínas CRISPR-Cas9 Clase 2 Tipo II.

FIG. 1A

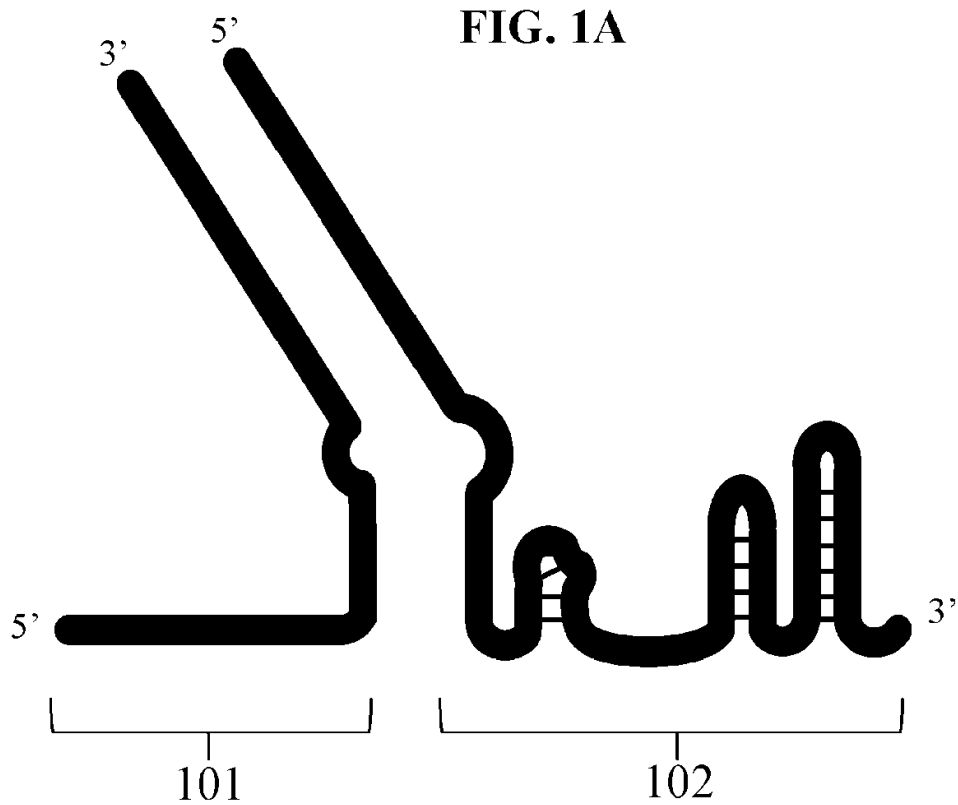


FIG. 1B

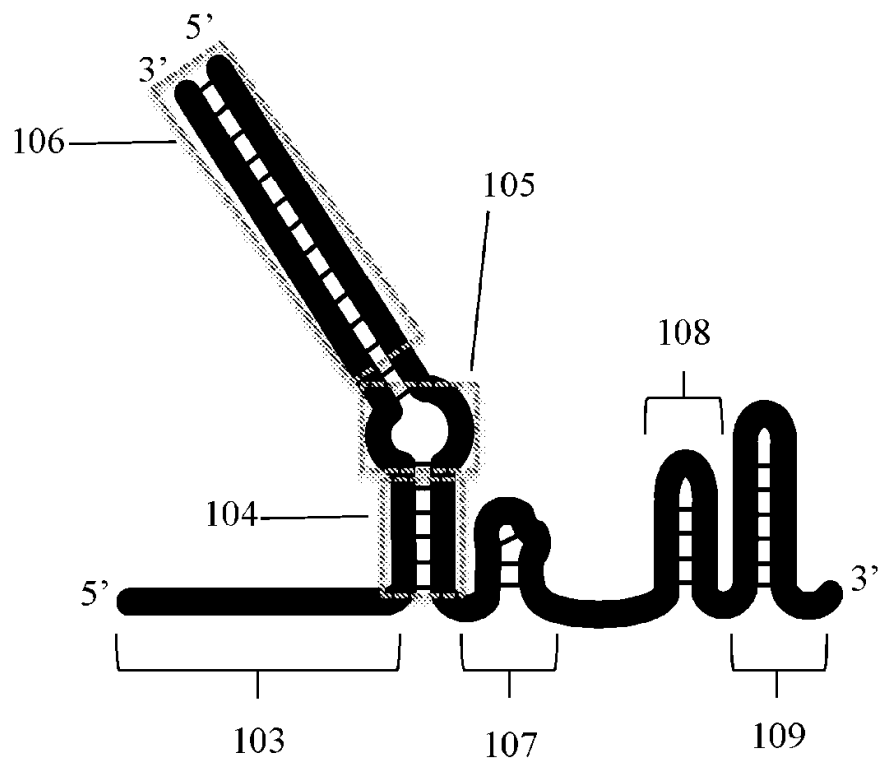


FIG. 1C

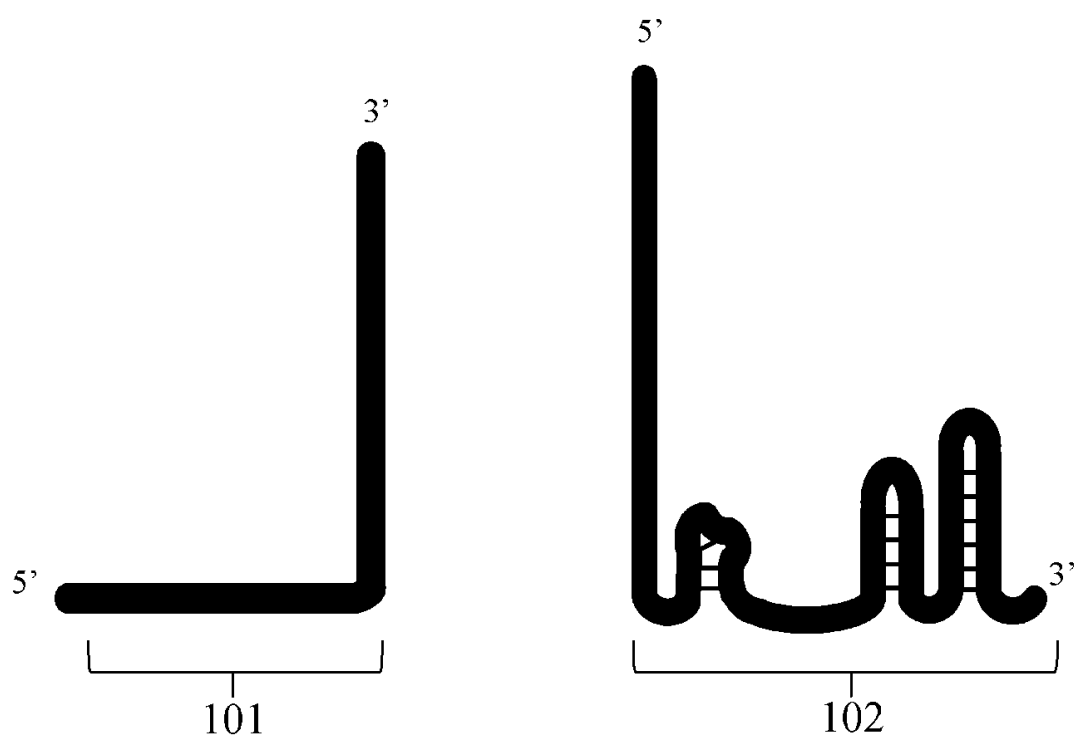
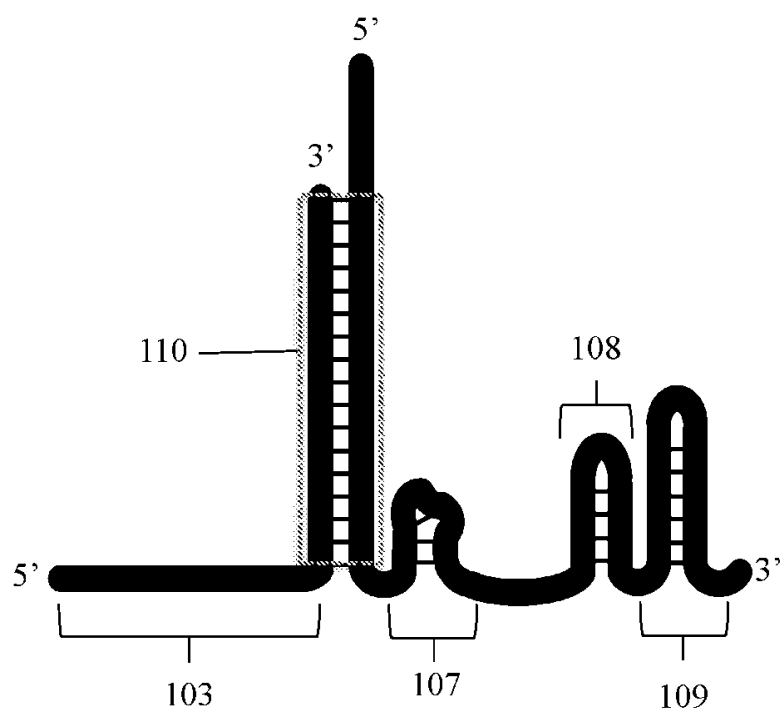


FIG. 1D



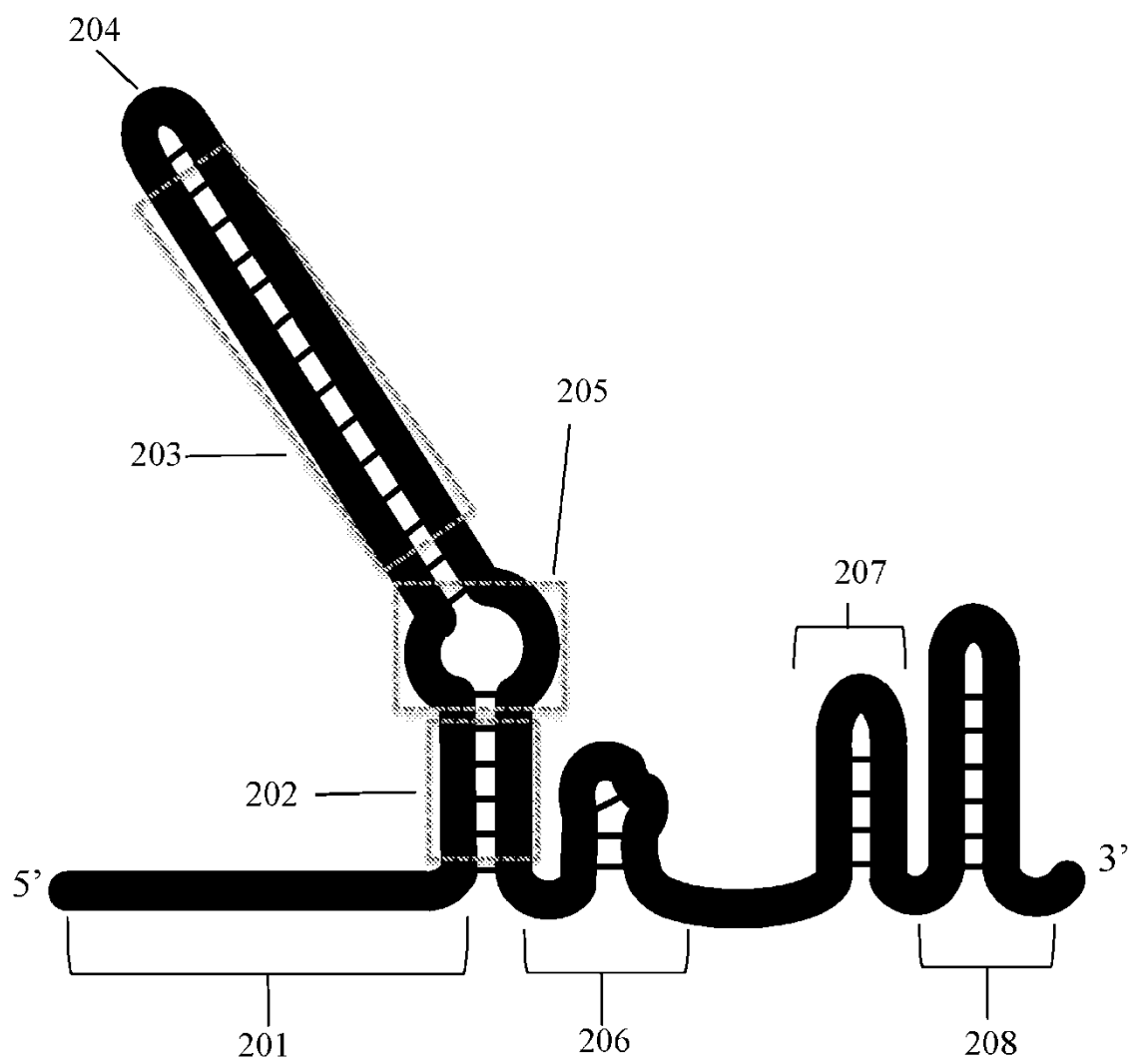


FIG. 2A

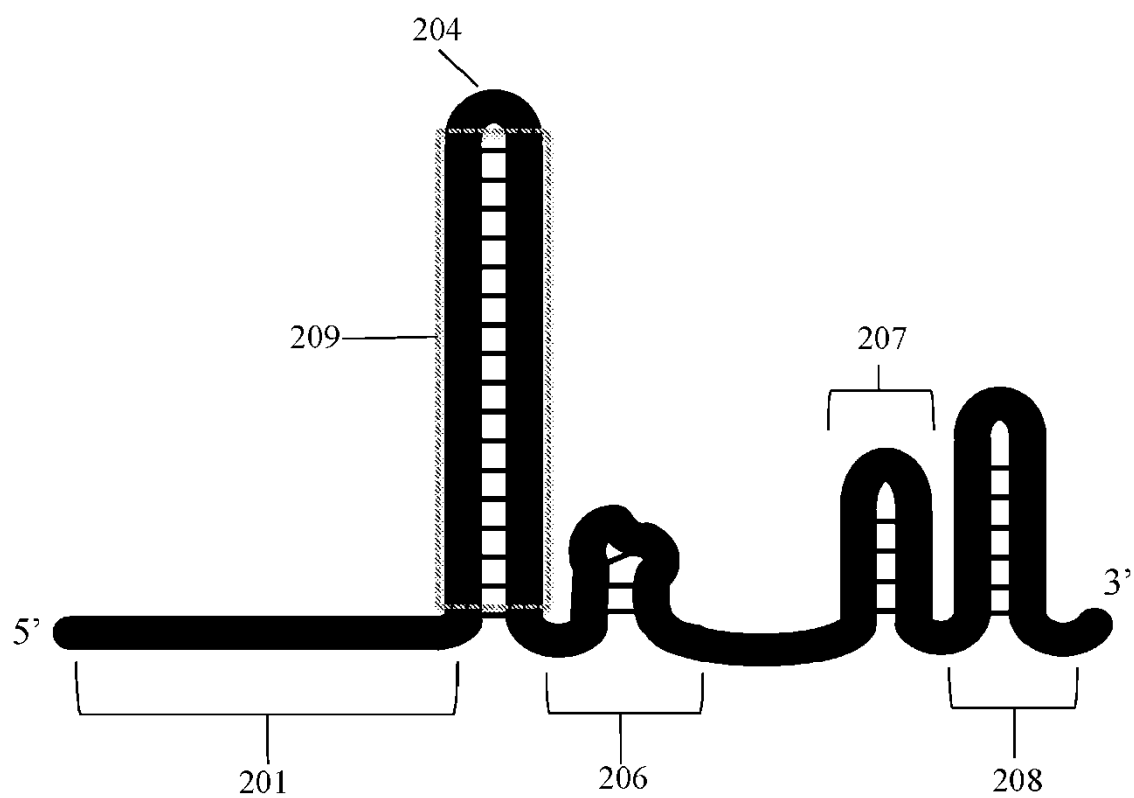


FIG. 2B

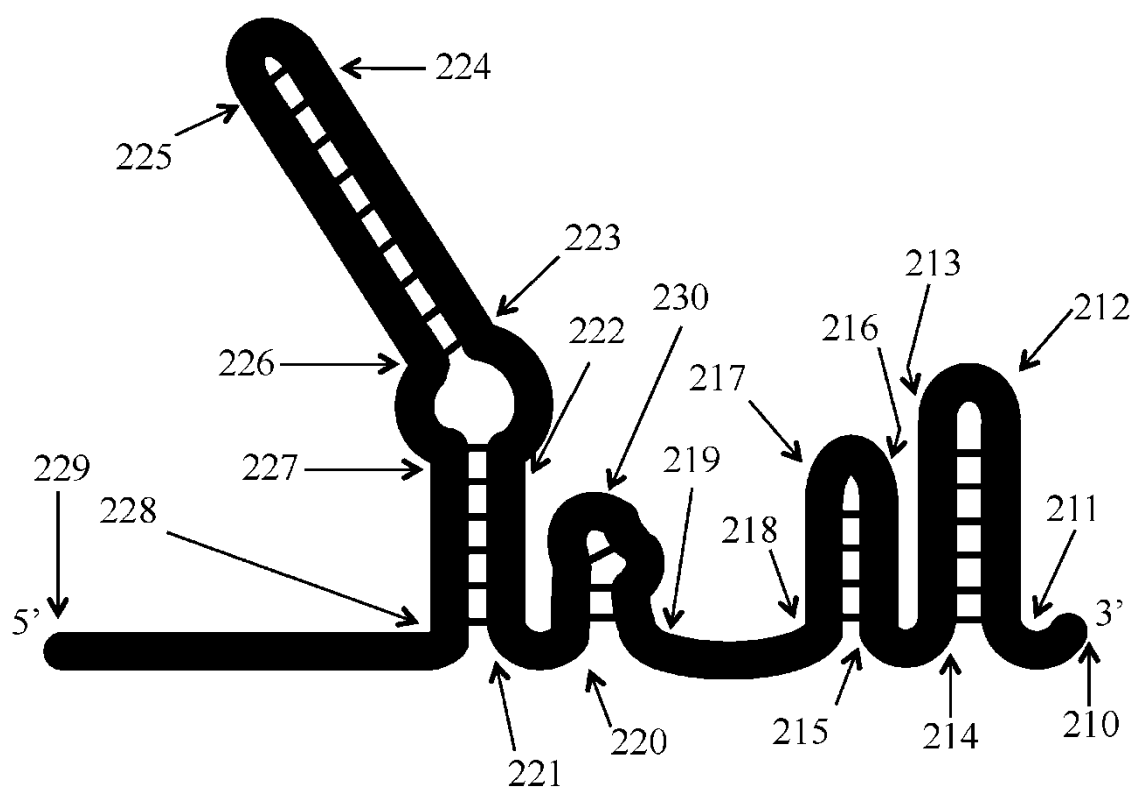


FIG. 2C

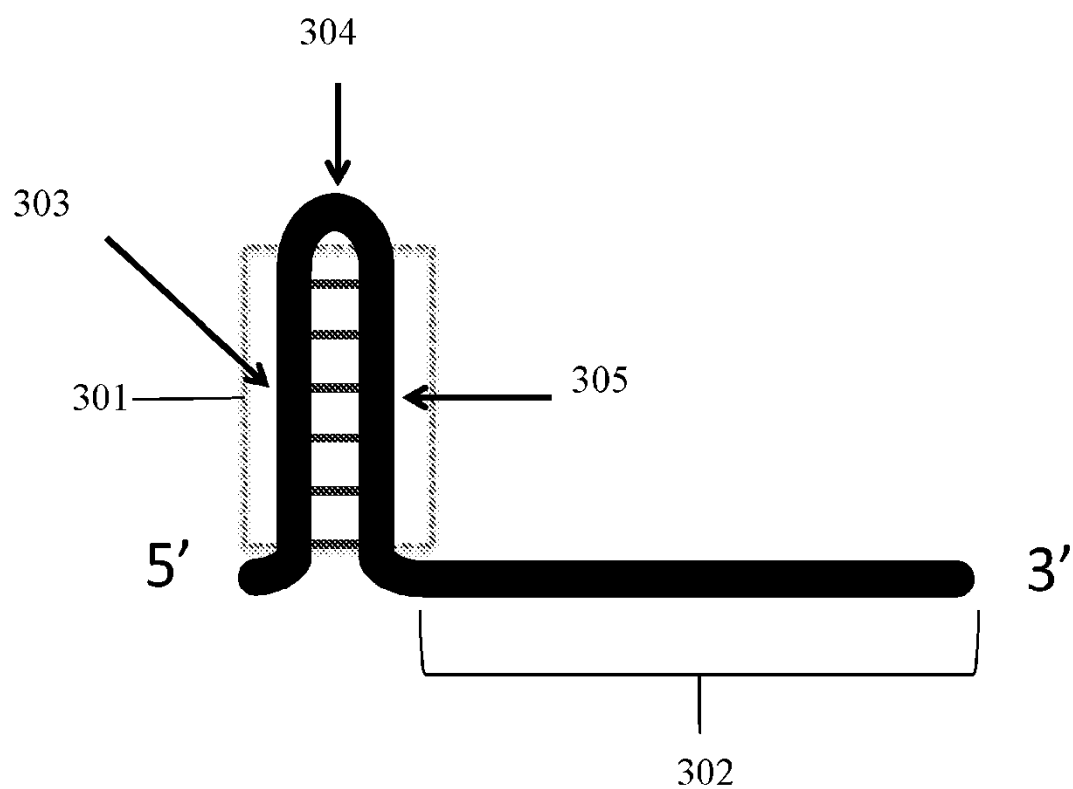


FIG. 3A

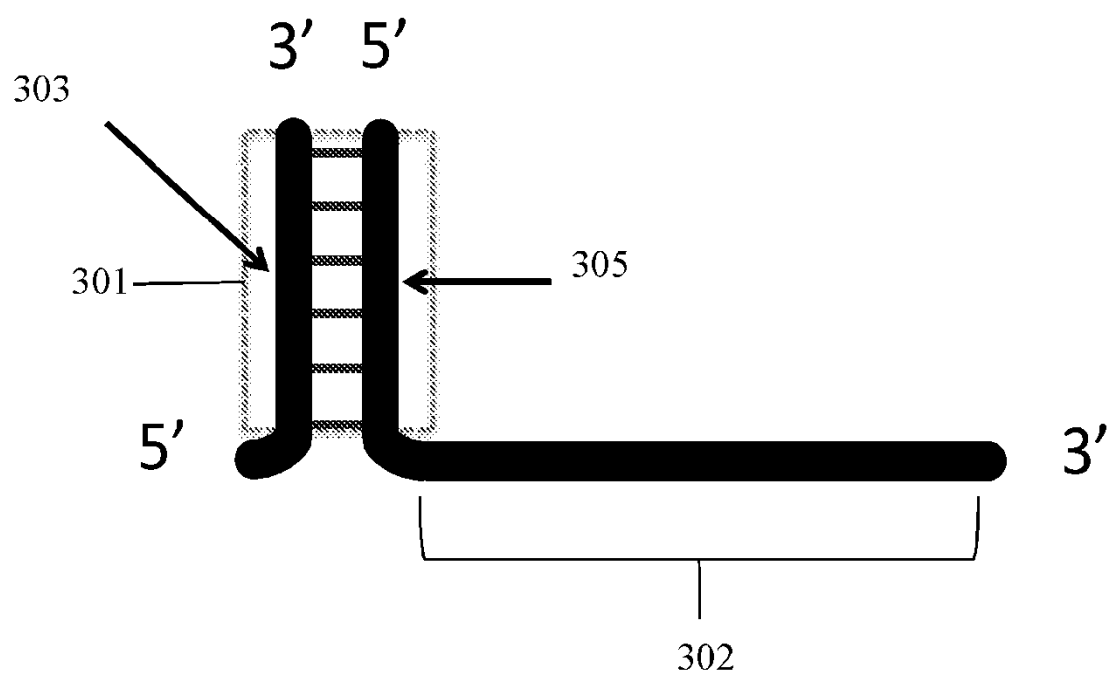


FIG. 3B

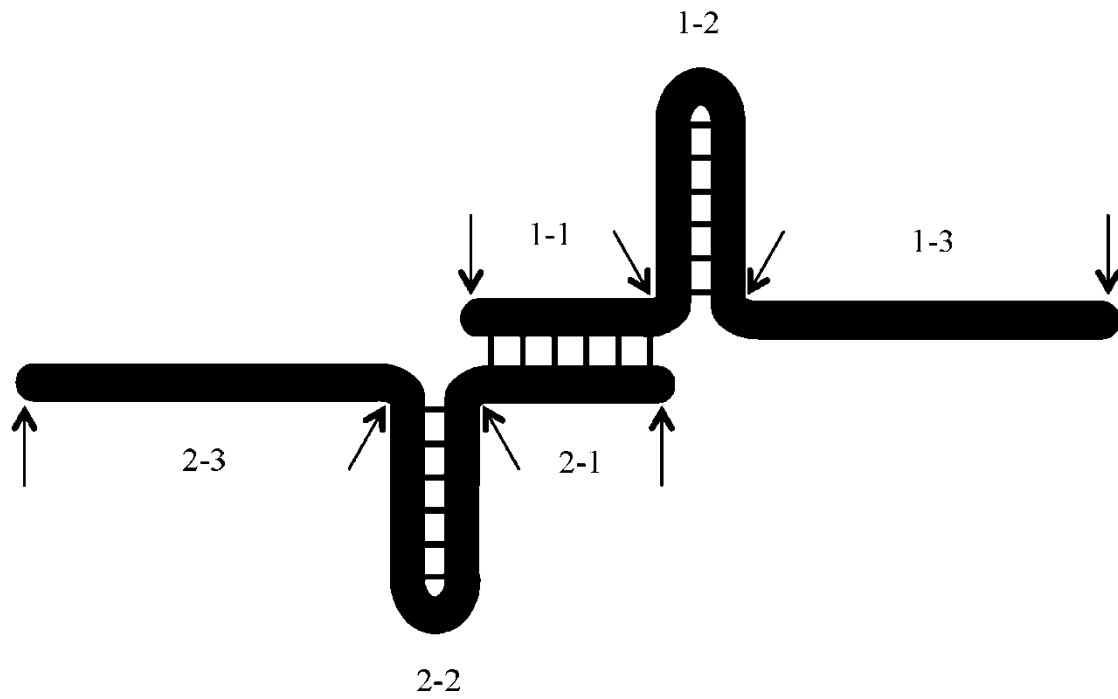


FIG. 4A

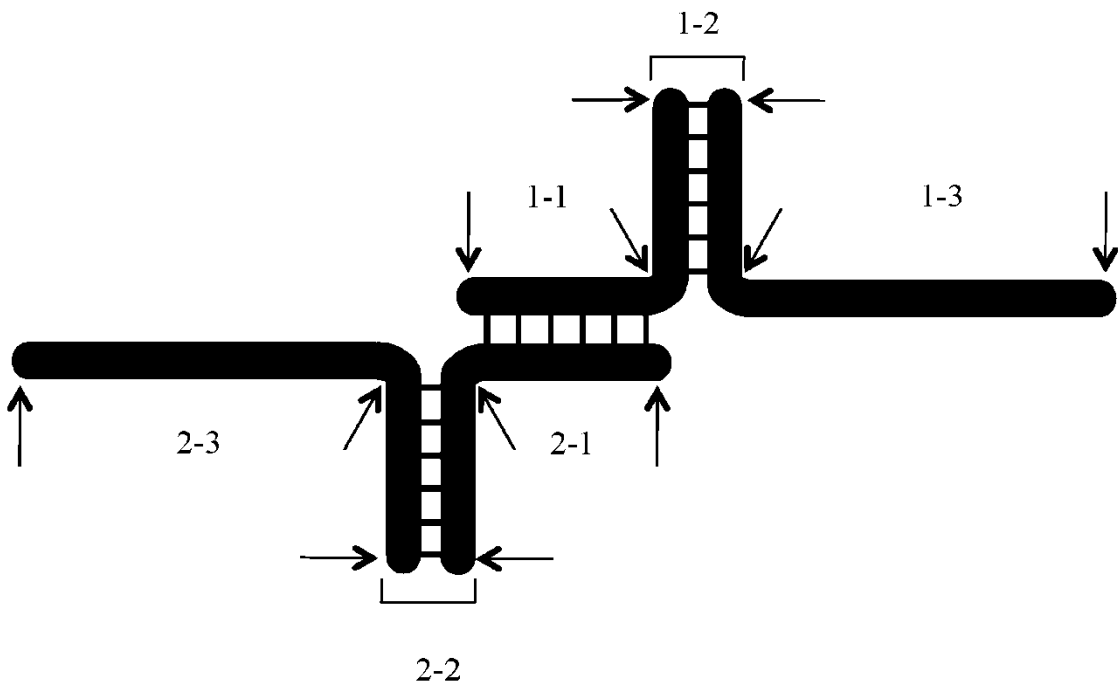
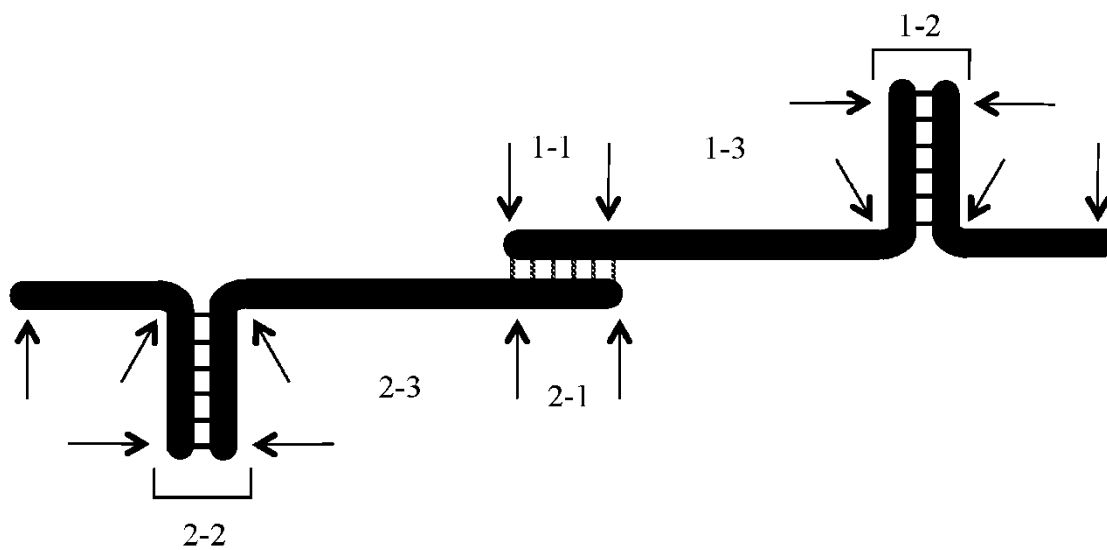
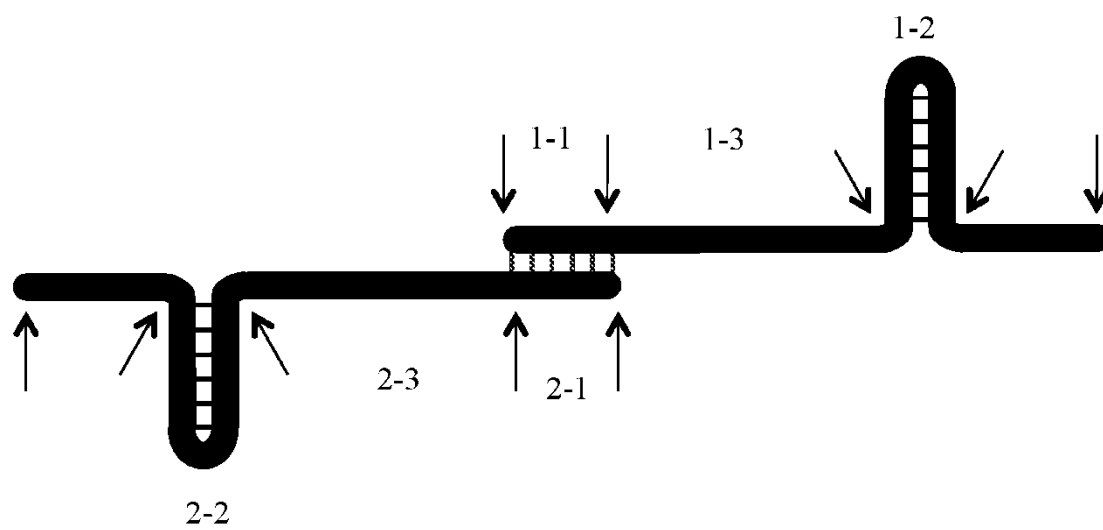


FIG. 4B



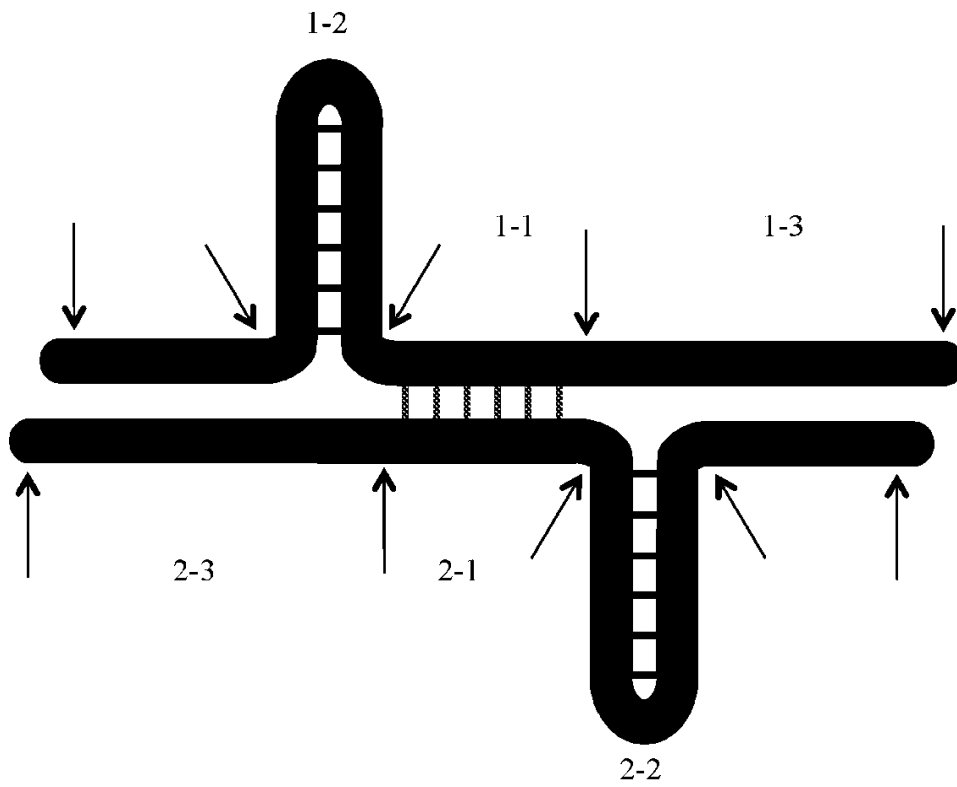


FIG. 4E

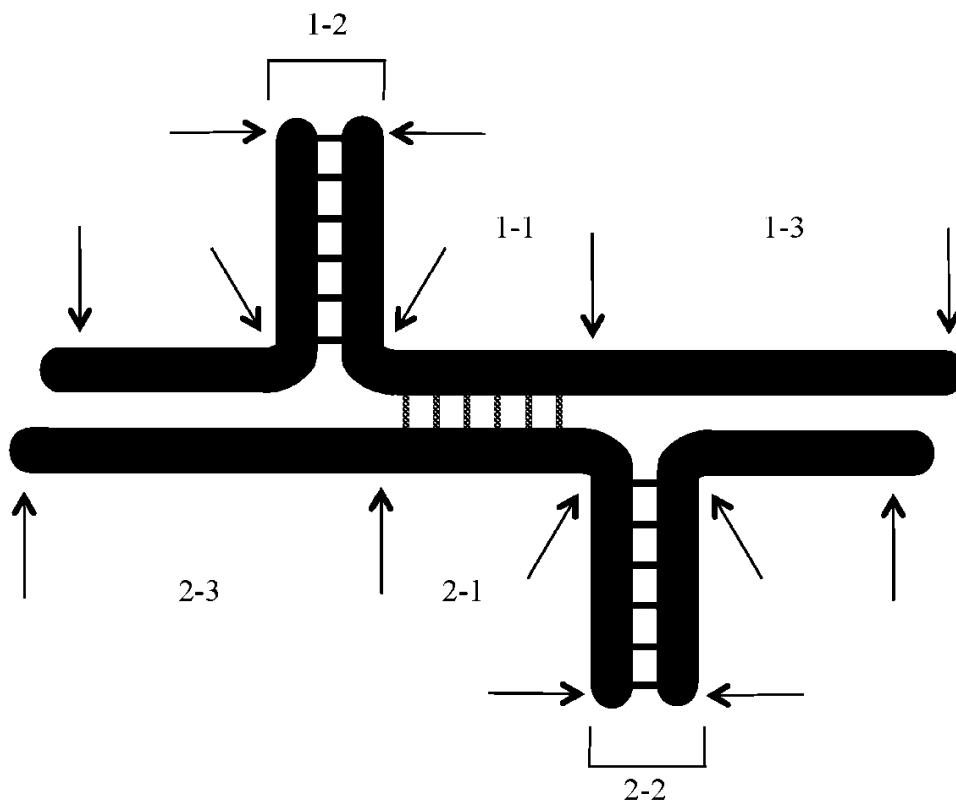


FIG. 4F

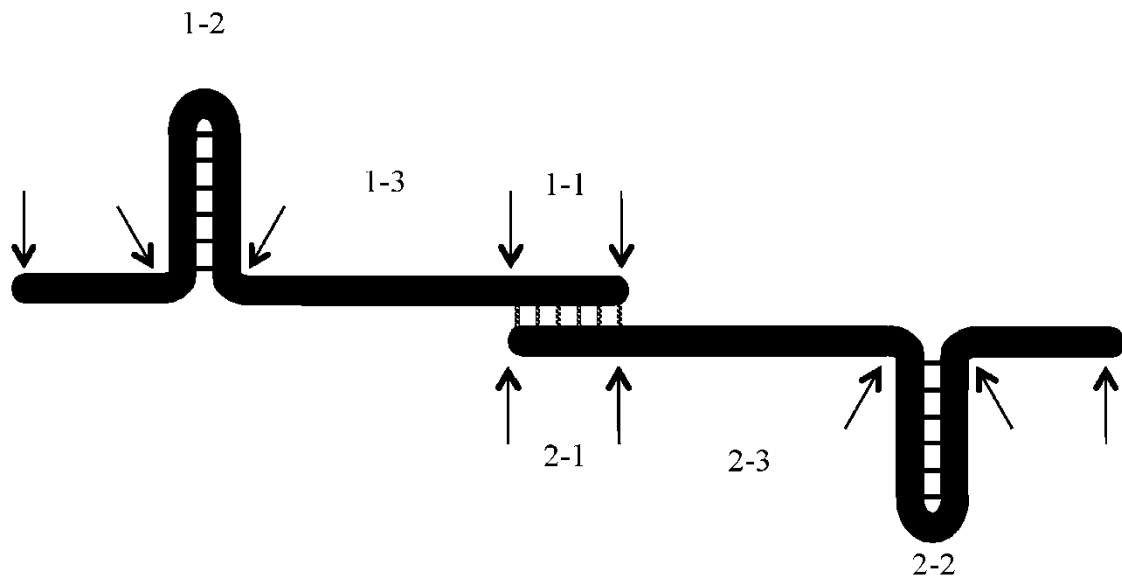


FIG. 4G

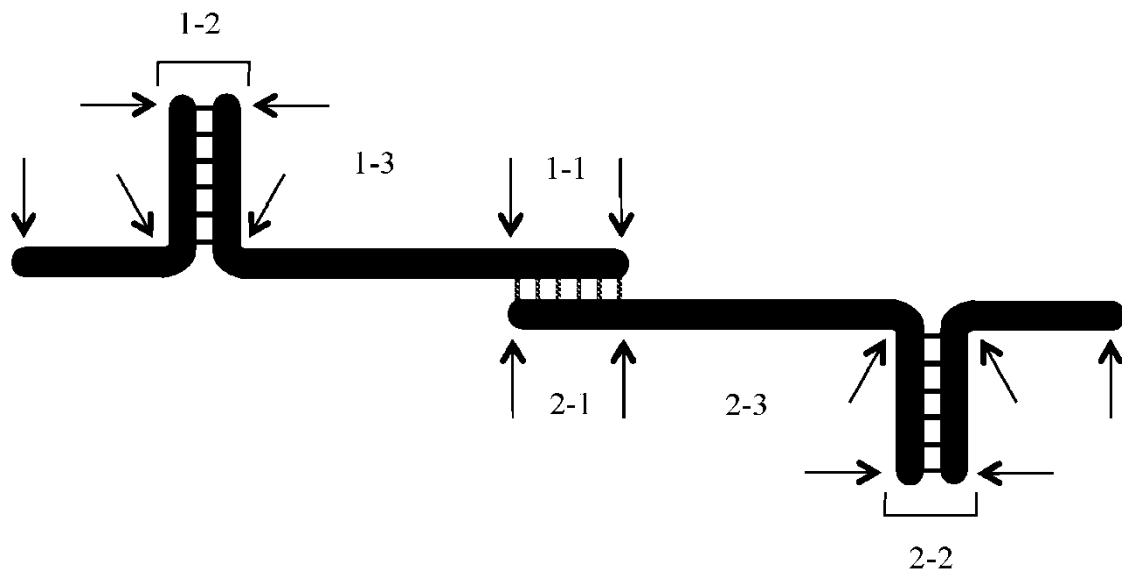


FIG. 4H

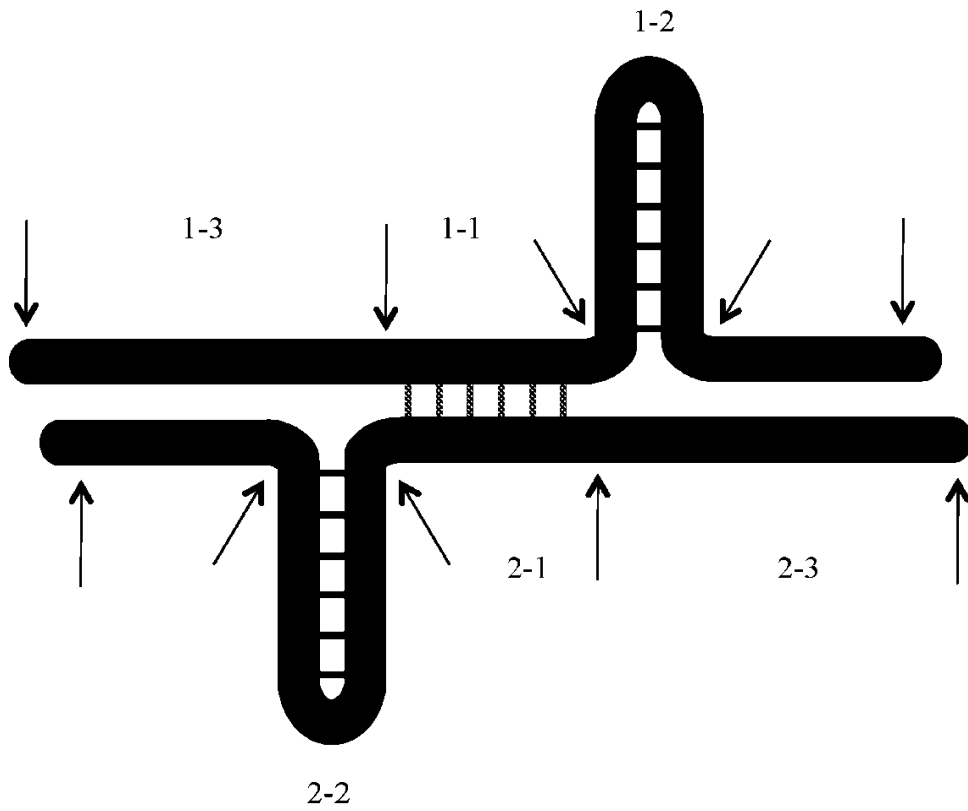


FIG. 4I

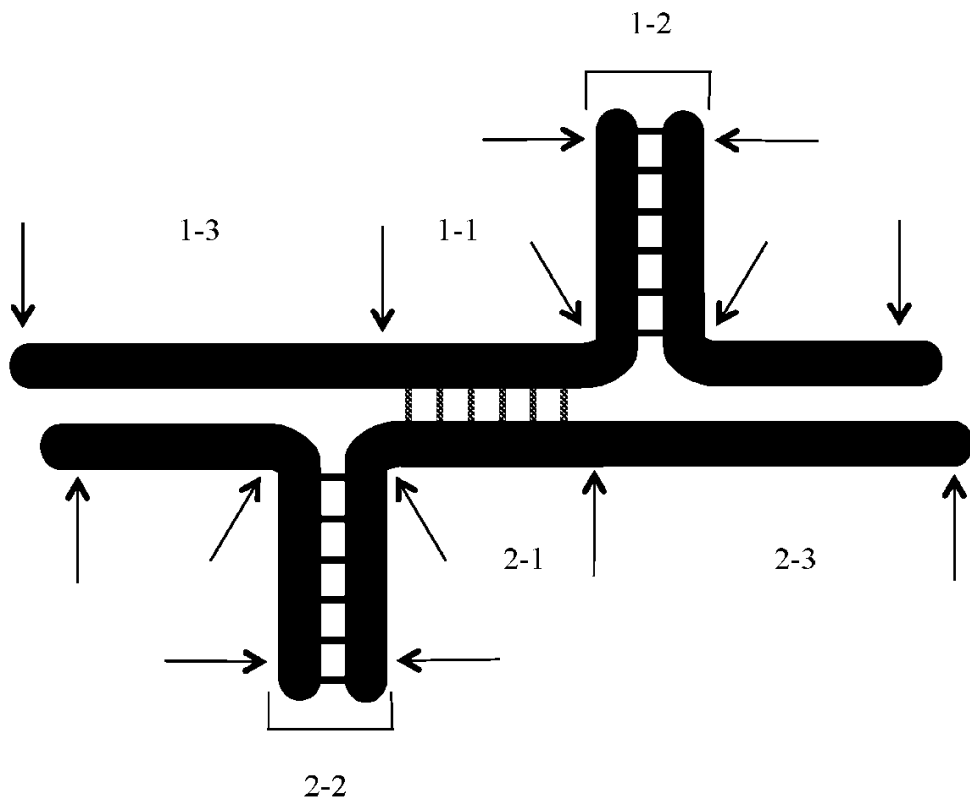


FIG. 4J

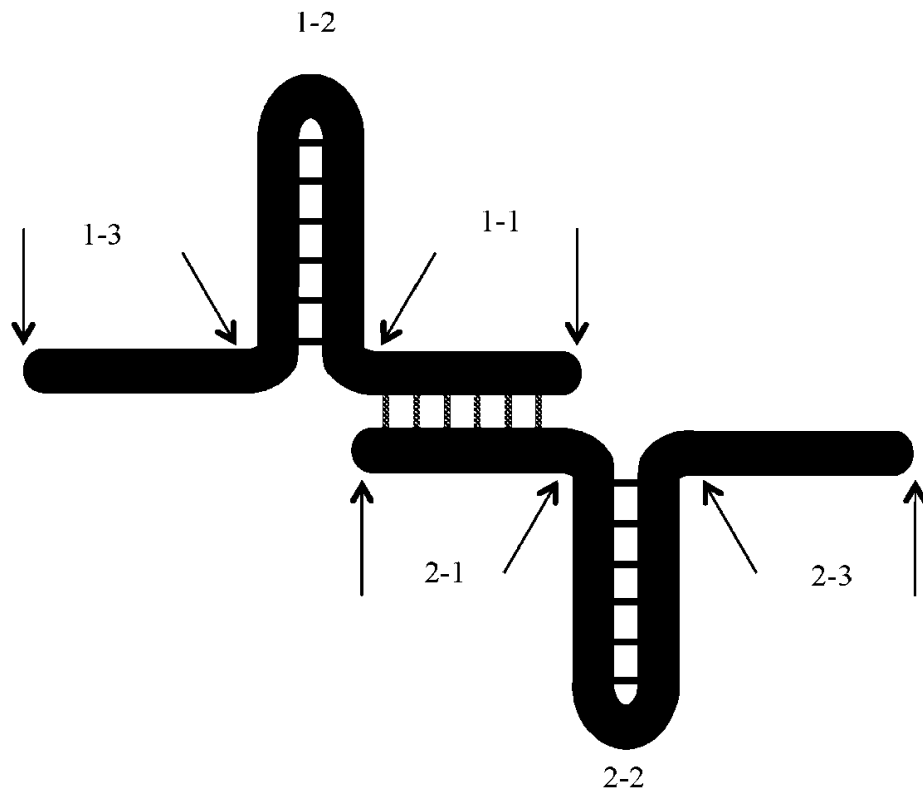


FIG. 4K

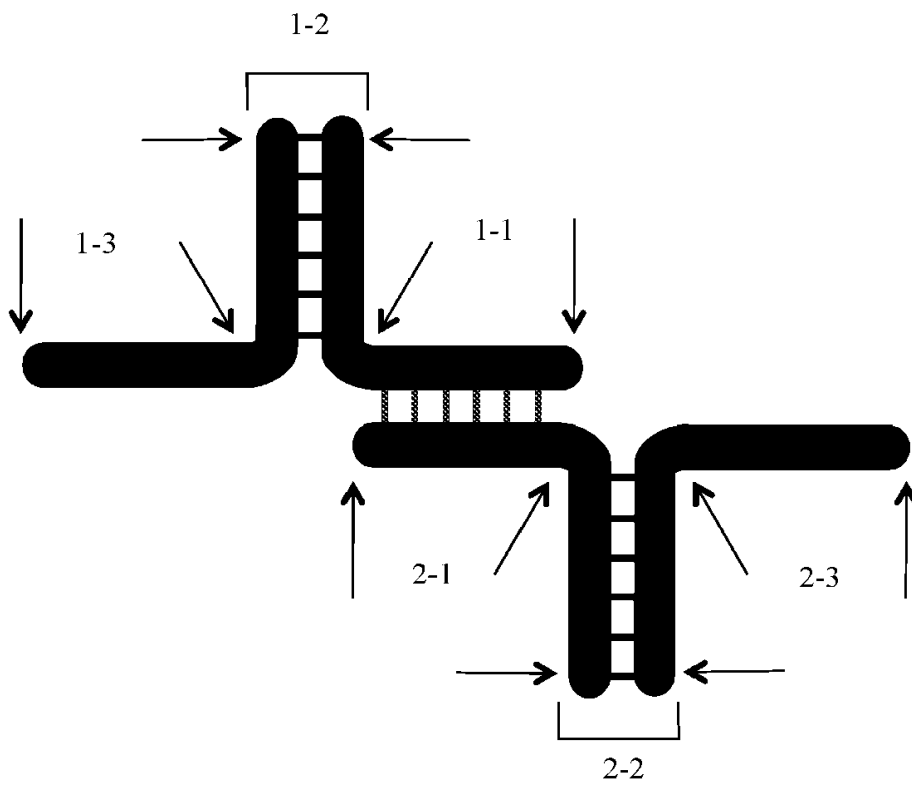


FIG. 4L

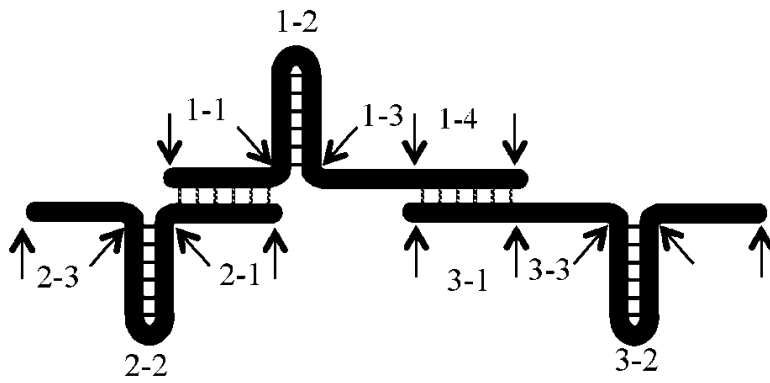


FIG. 4M

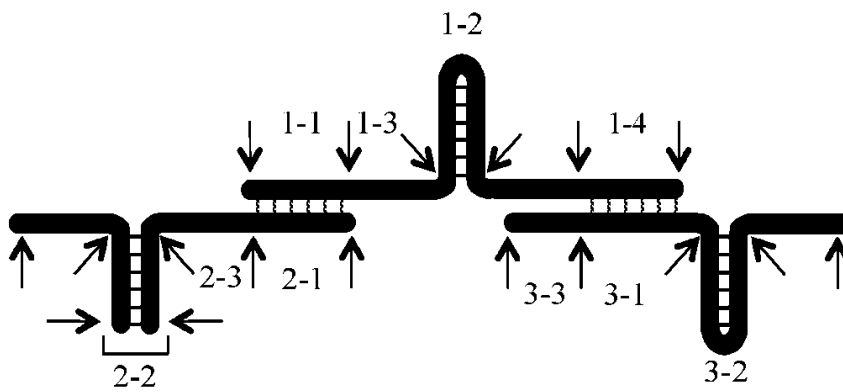


FIG. 4N

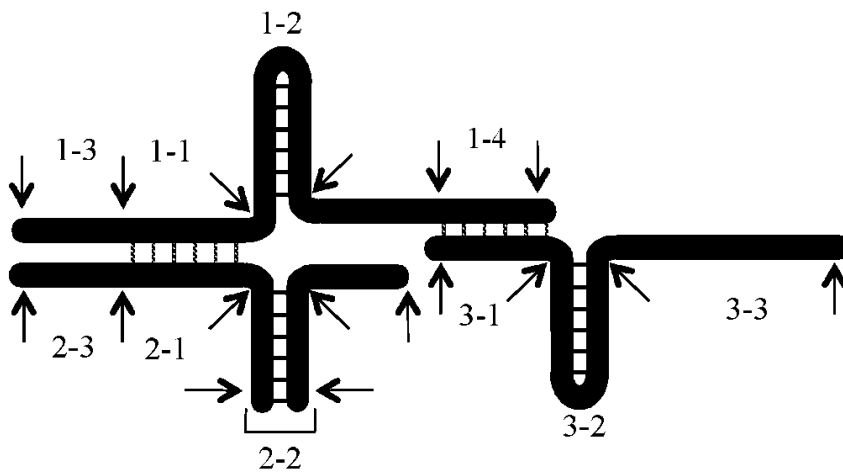


FIG. 4O

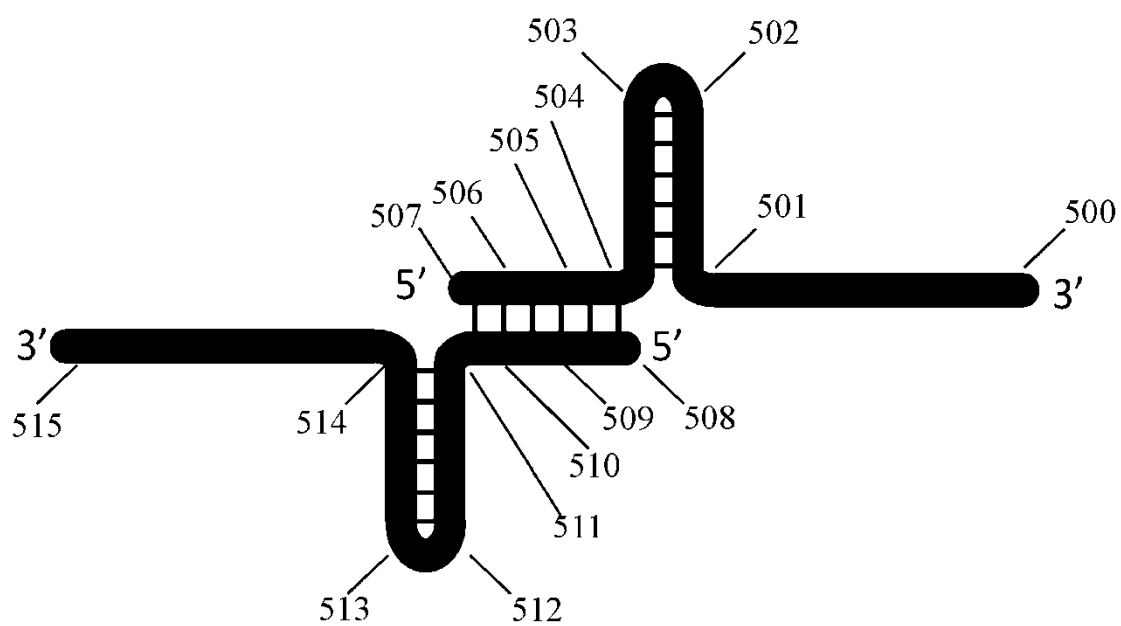


FIG. 5A

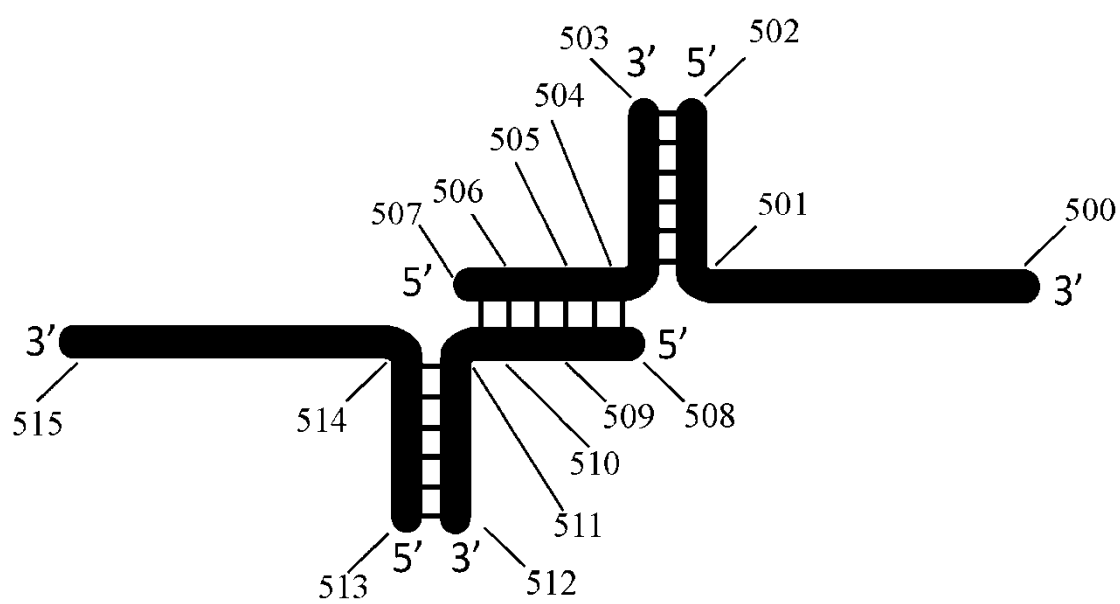


FIG. 5B

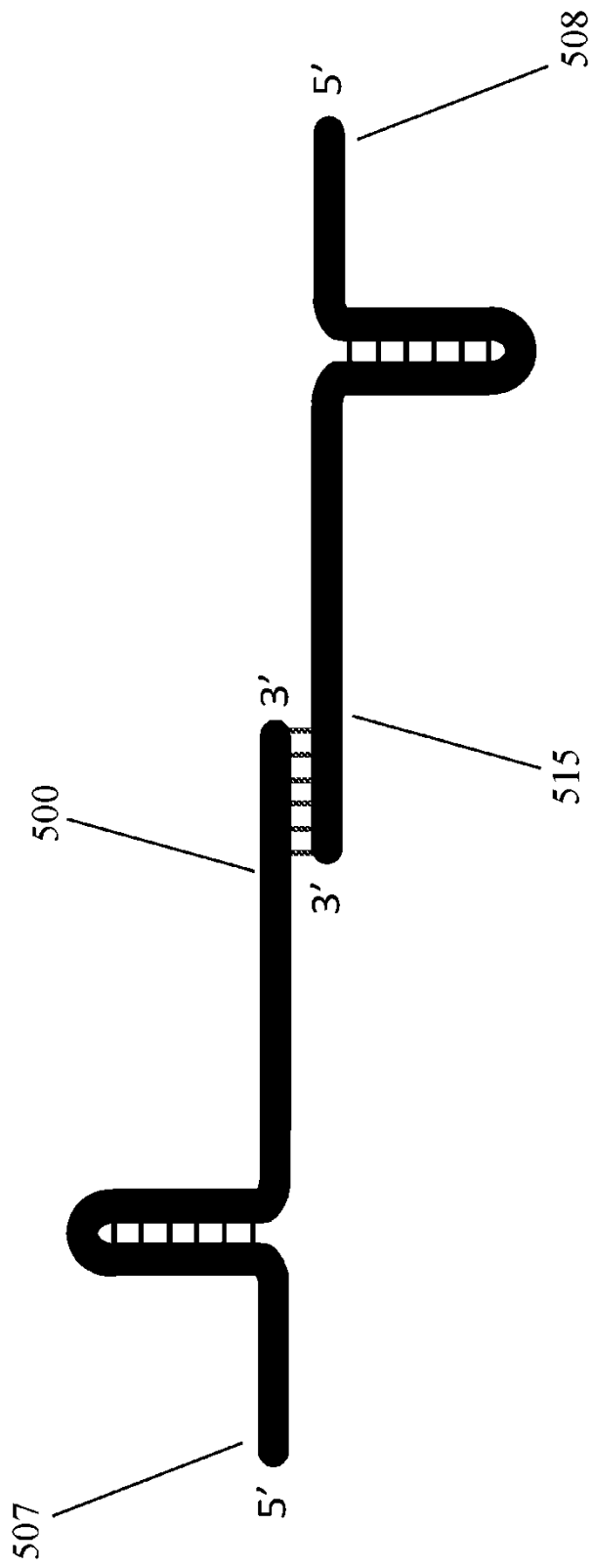


FIG. 5C

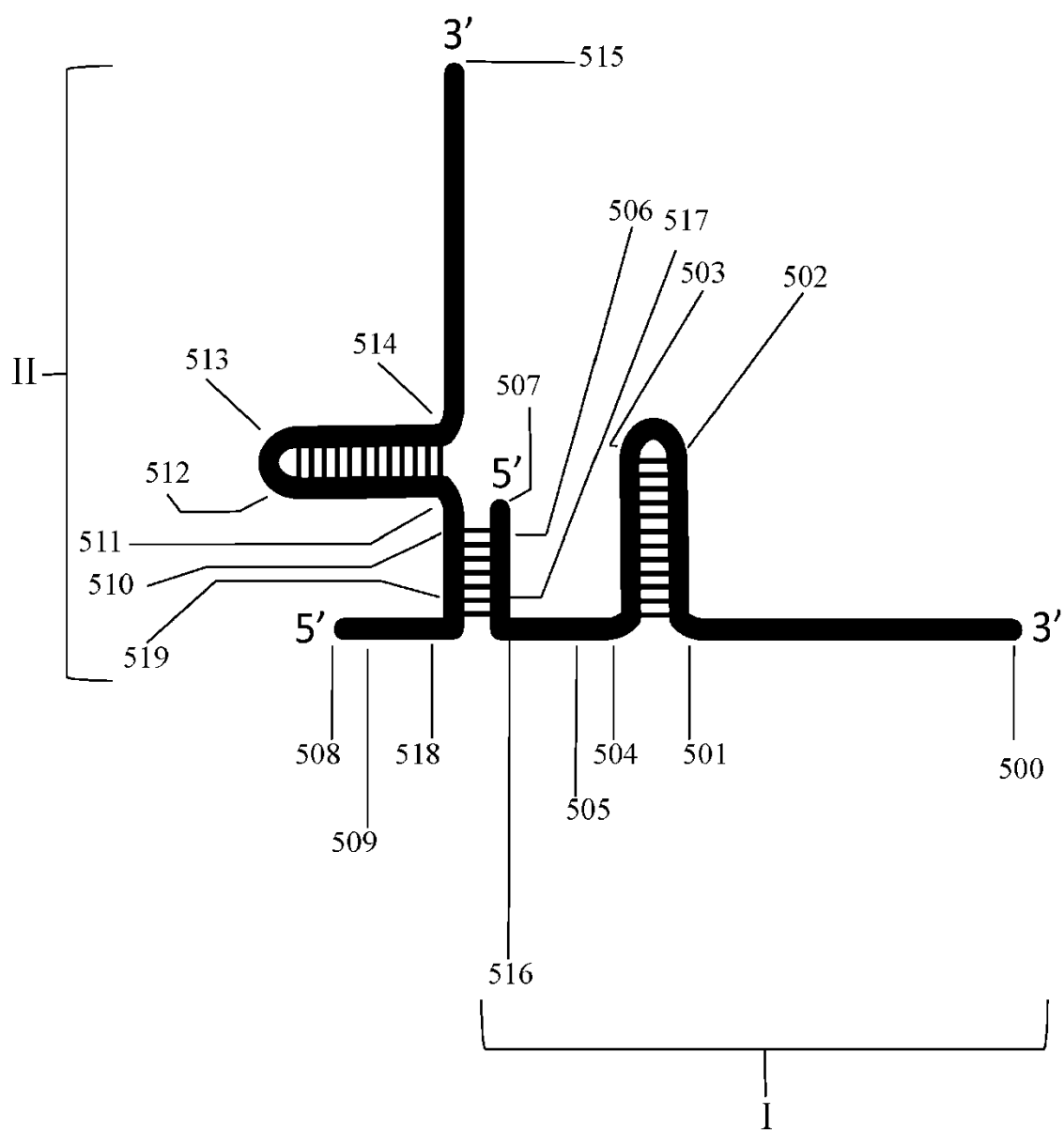


FIG. 5D

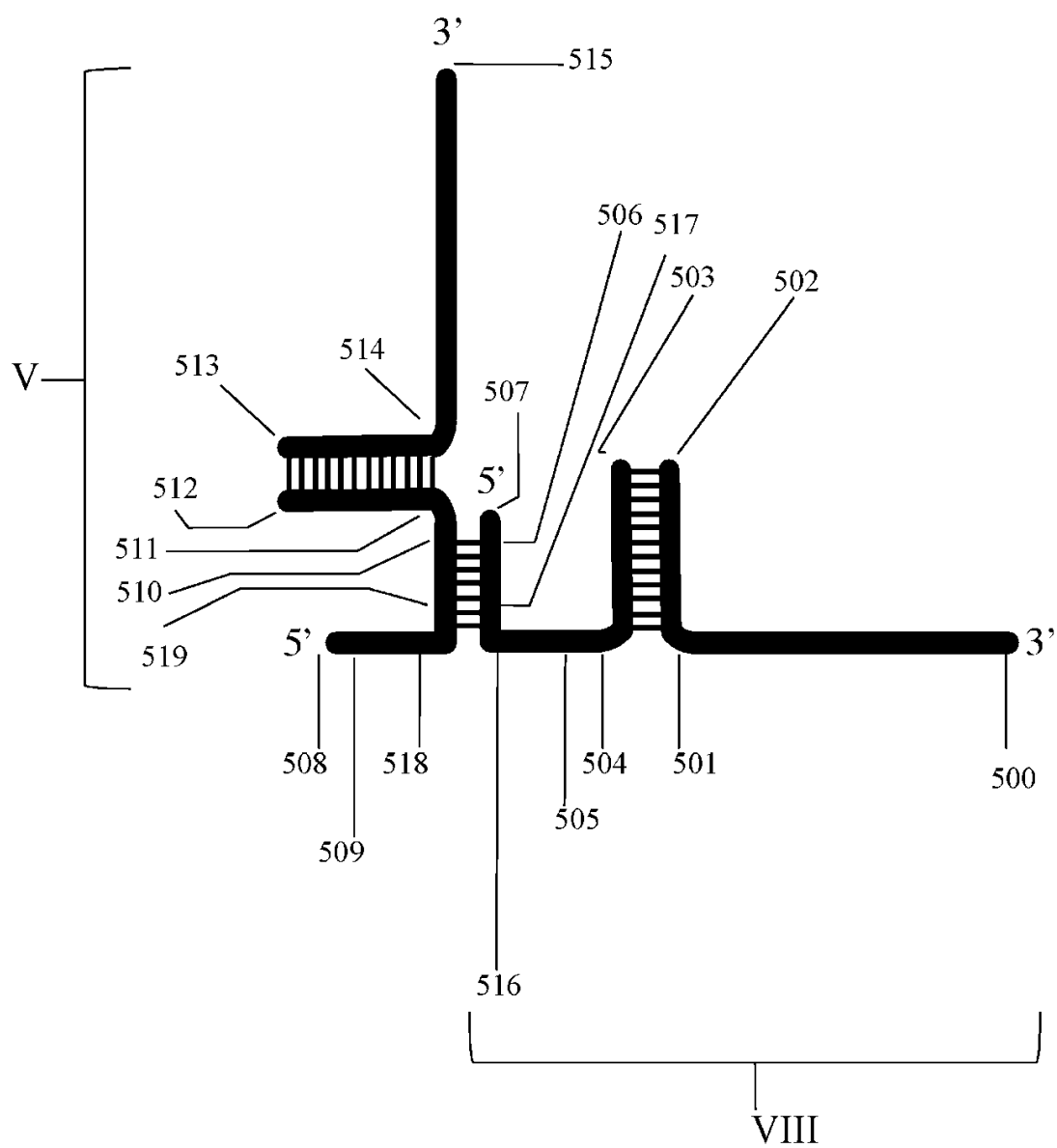


FIG. 5F

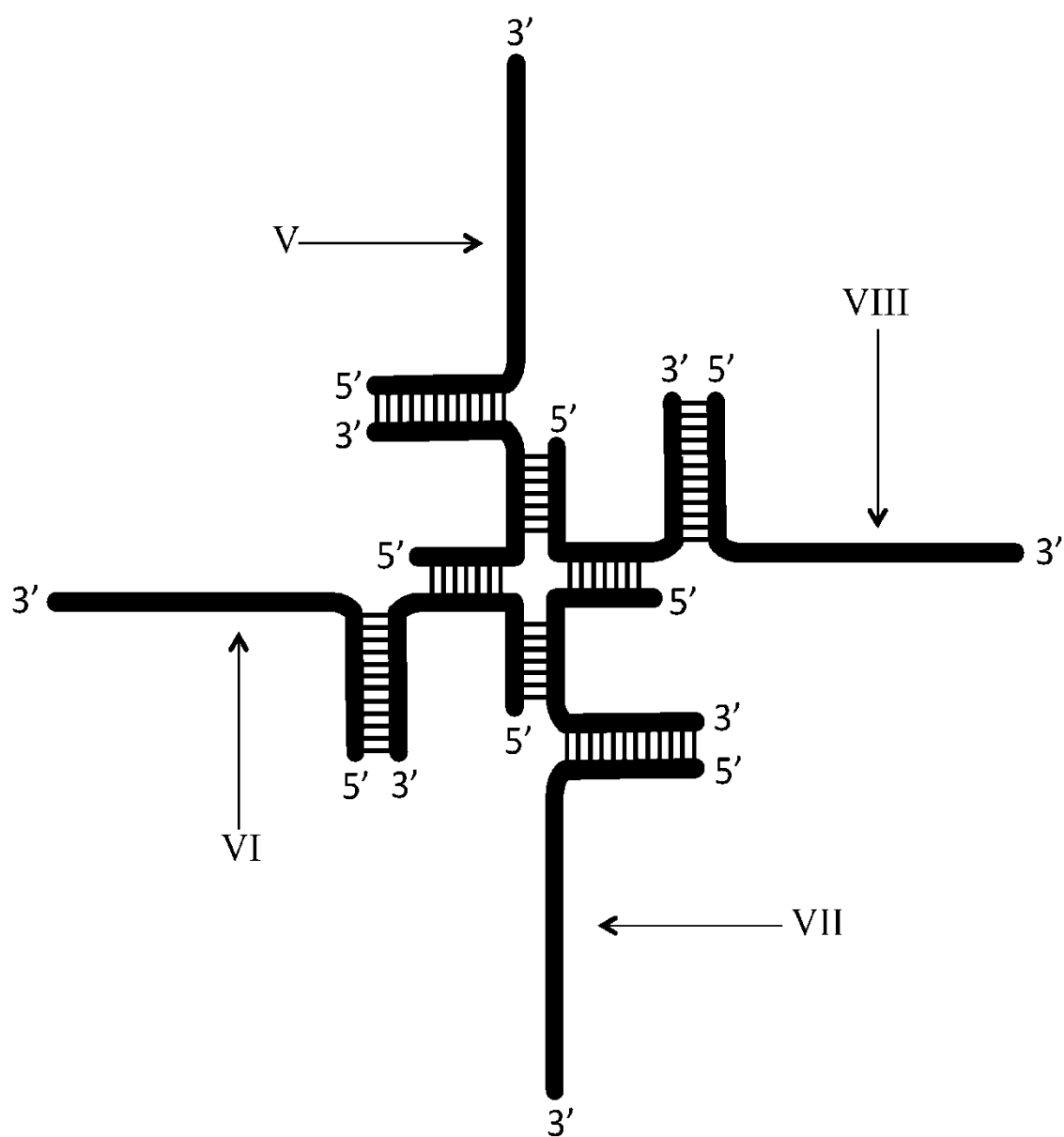


FIG. 5G

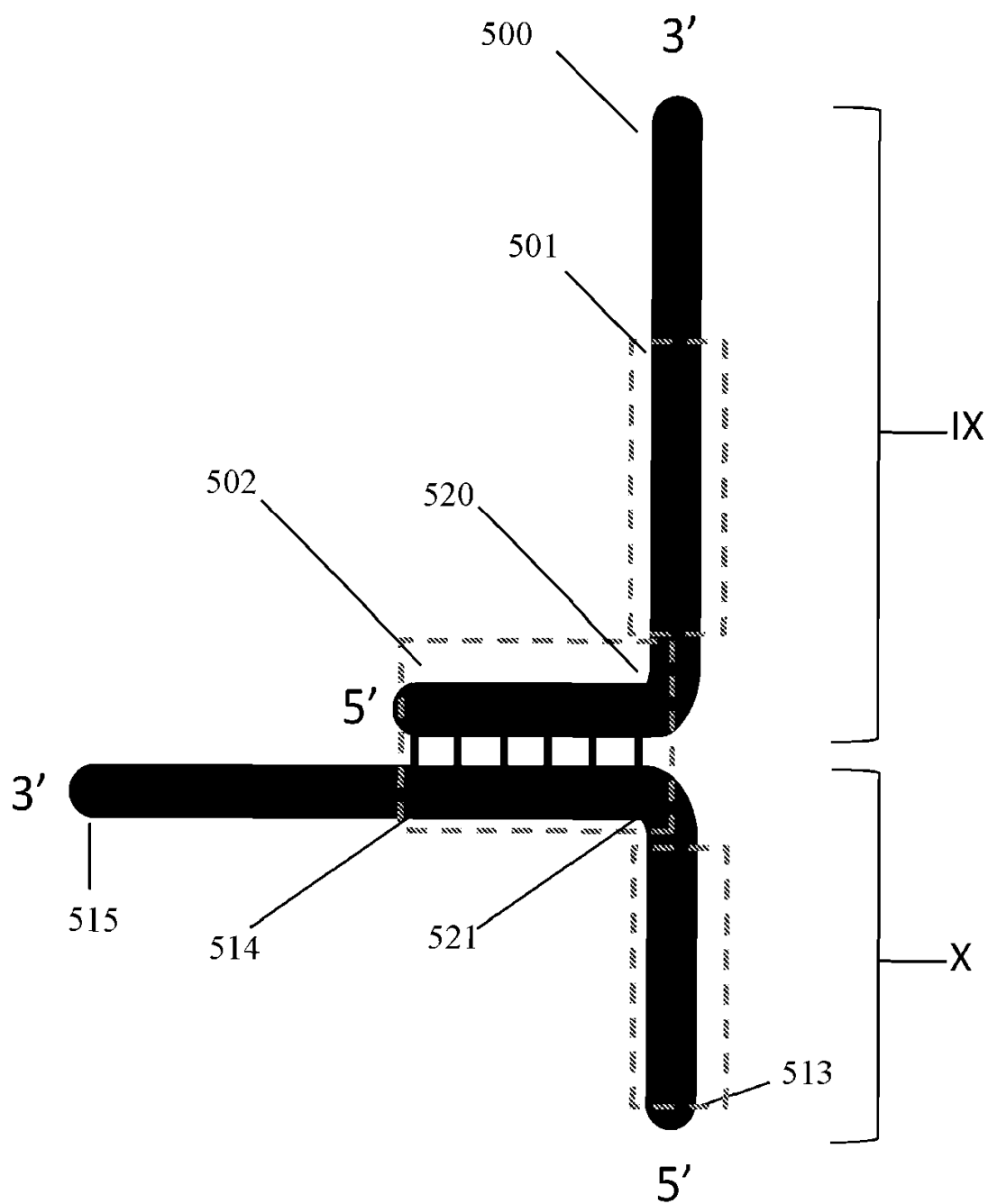


FIG. 5H

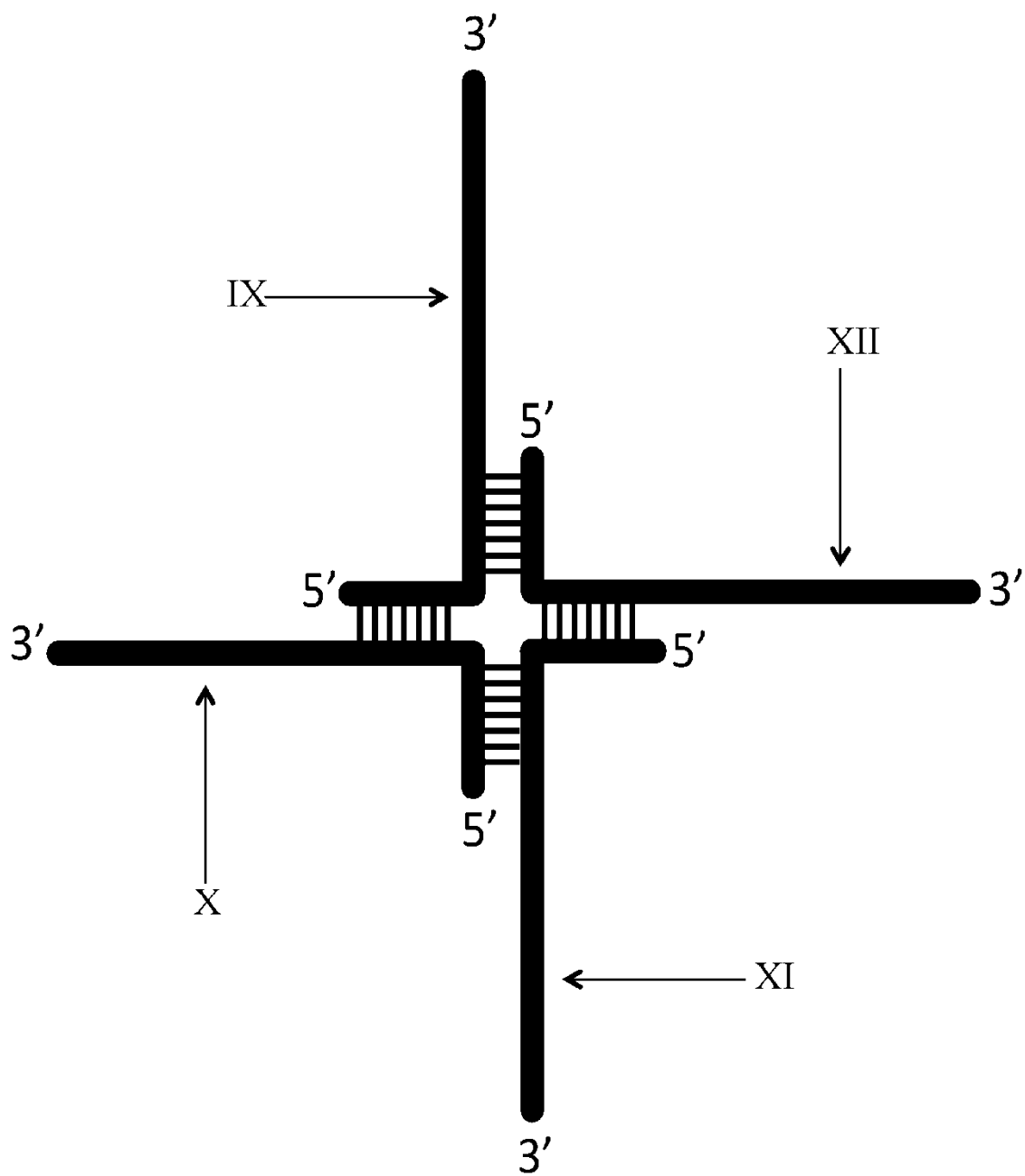


FIG. 5I

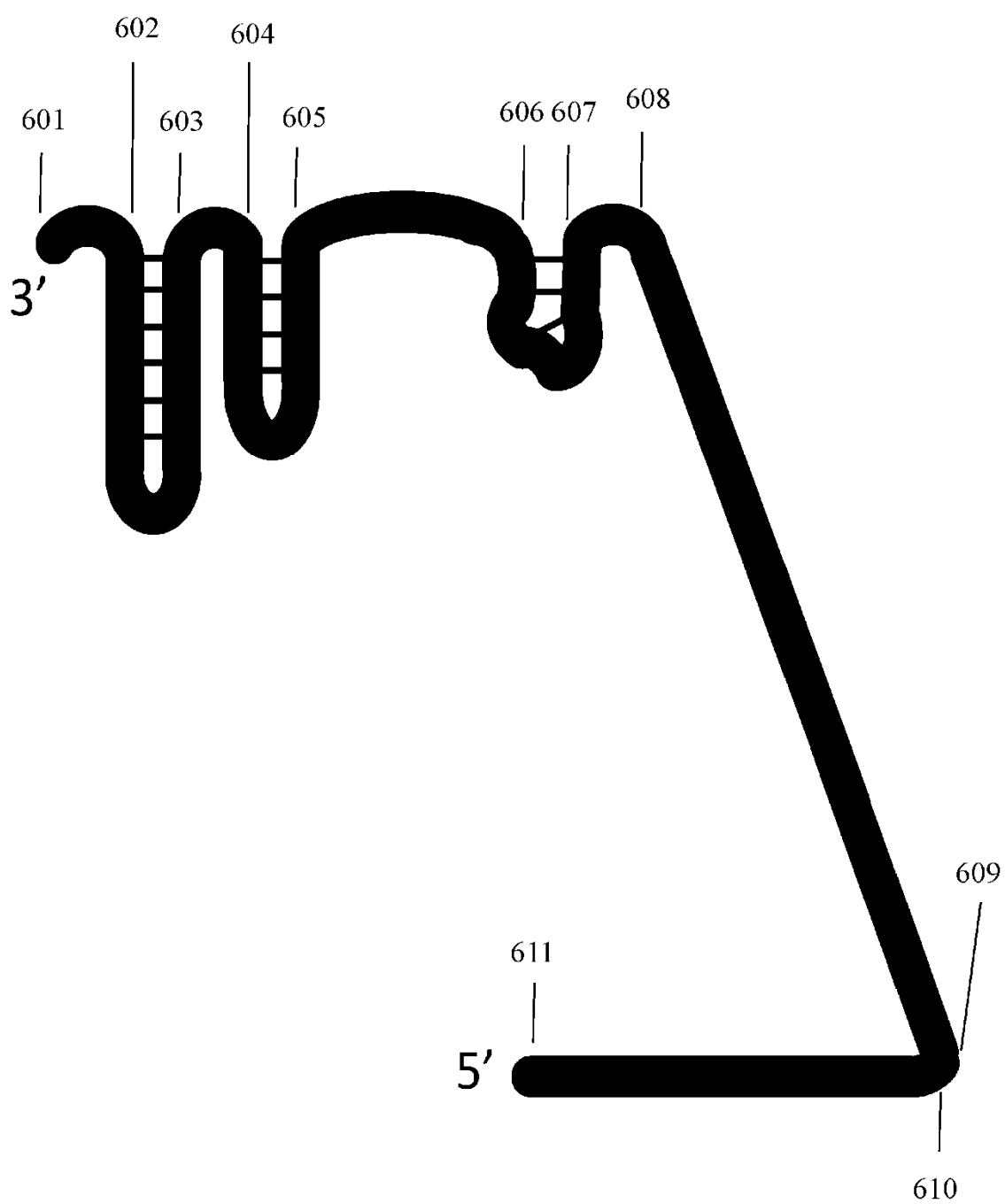


FIG. 6A

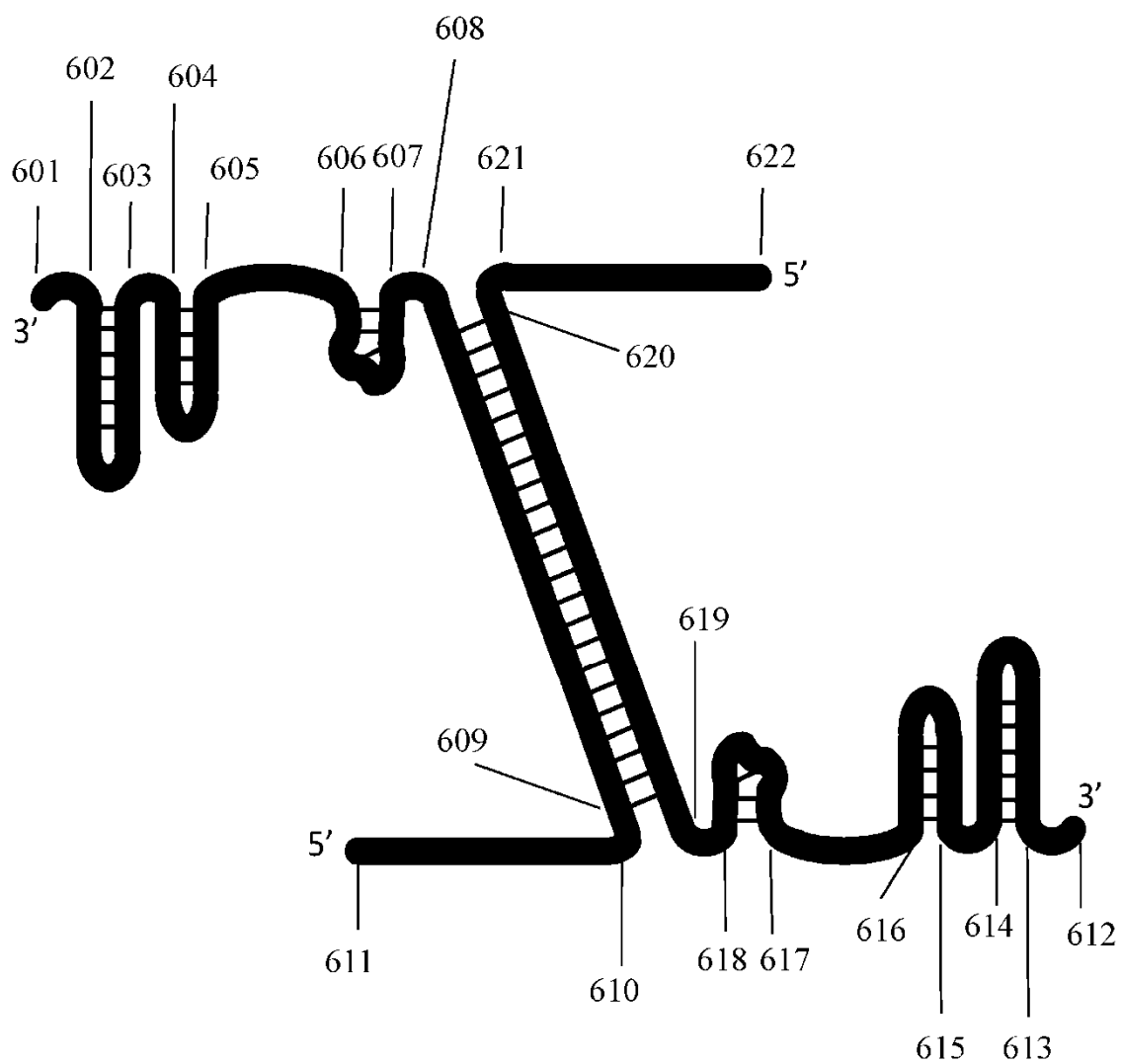


FIG. 6B

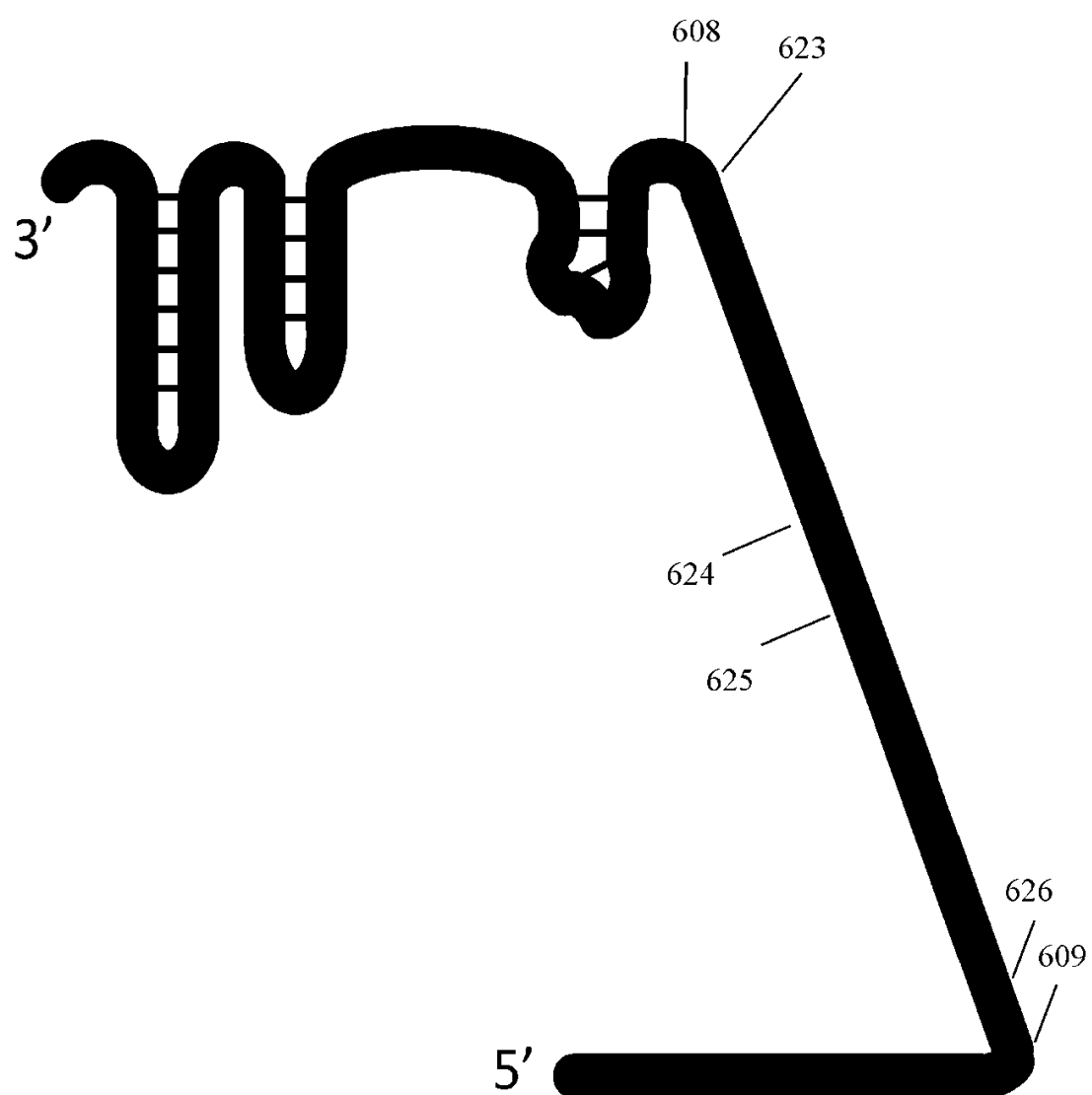


FIG. 6C

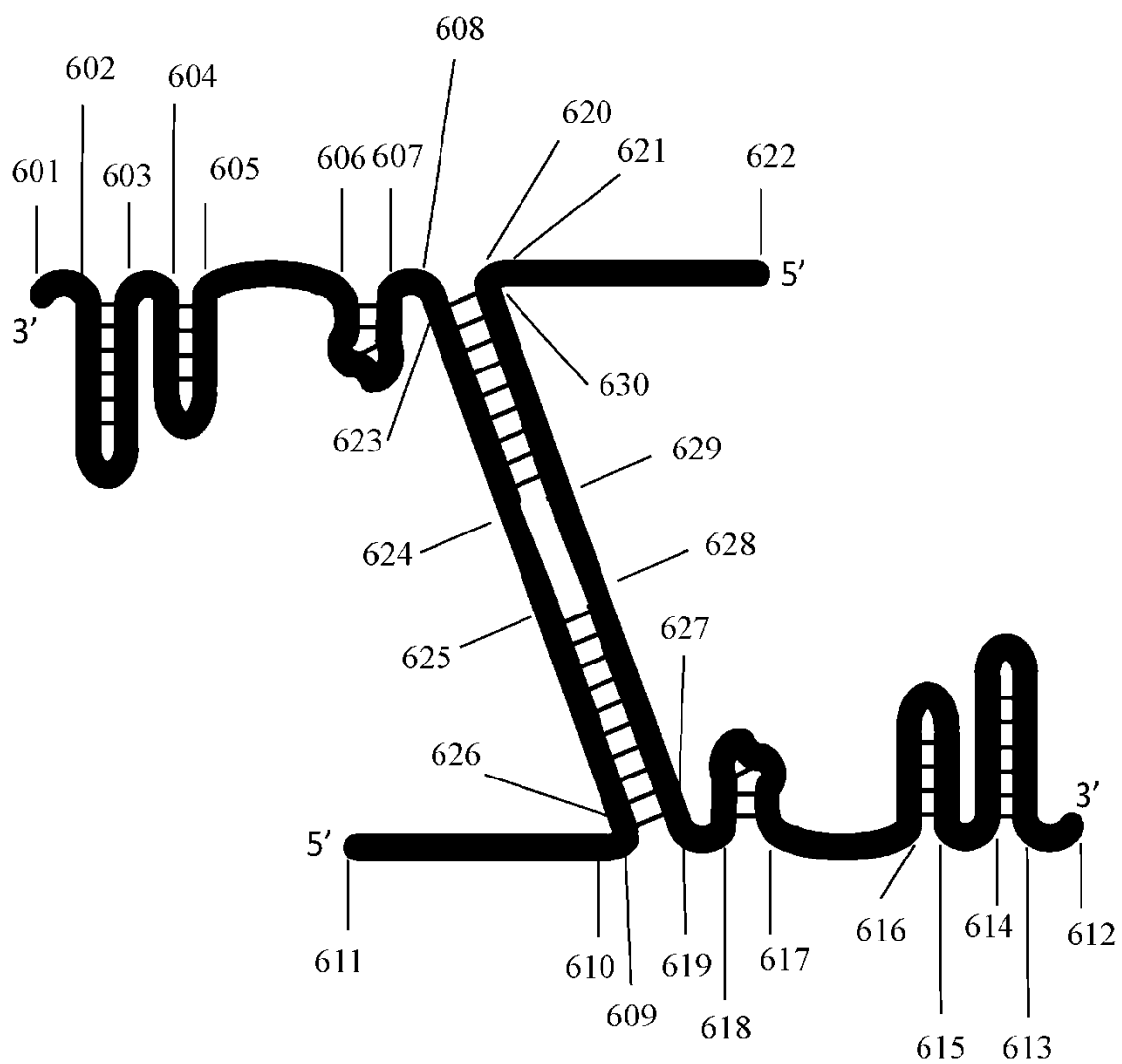


FIG. 6D

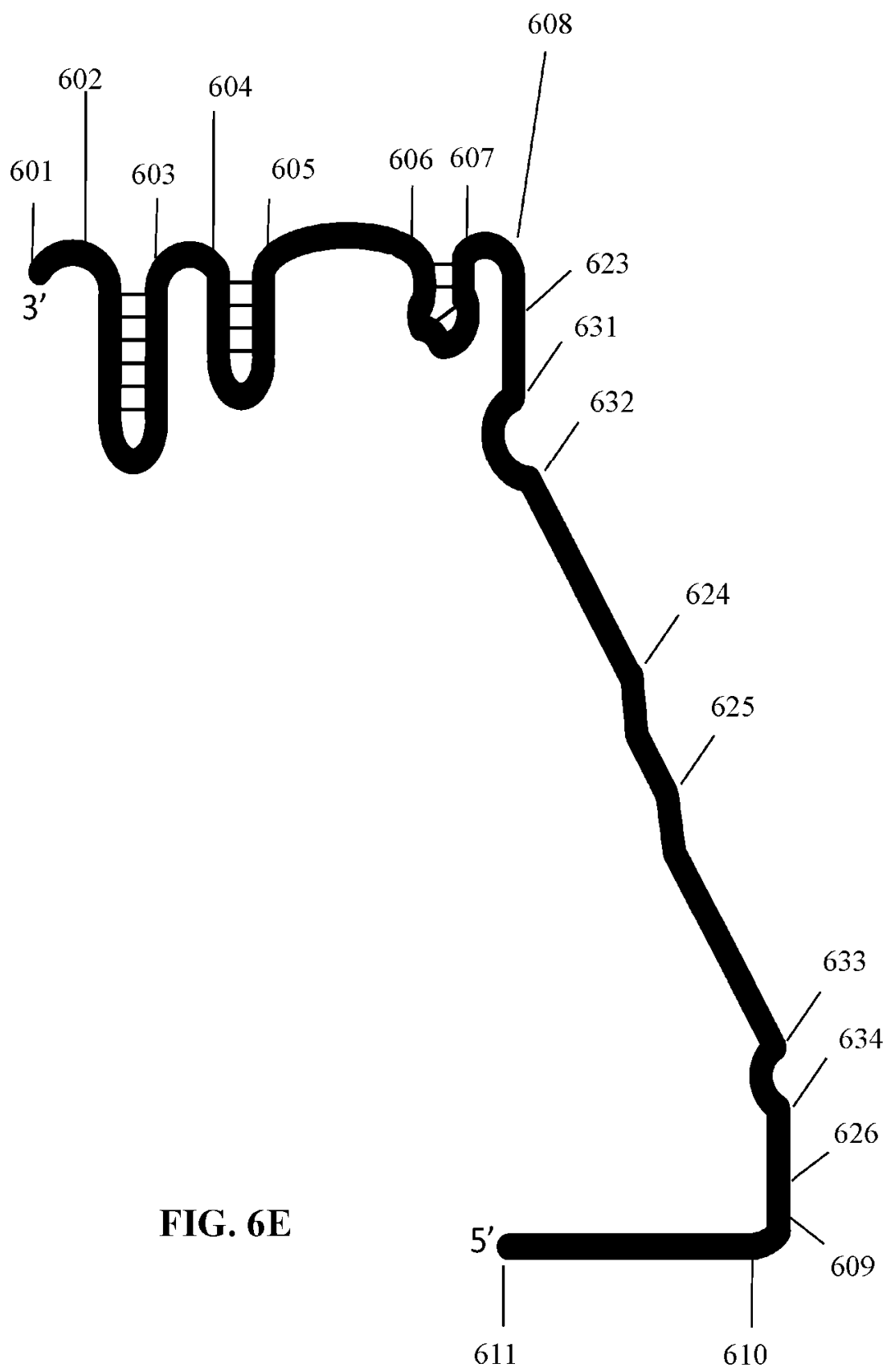


FIG. 6E

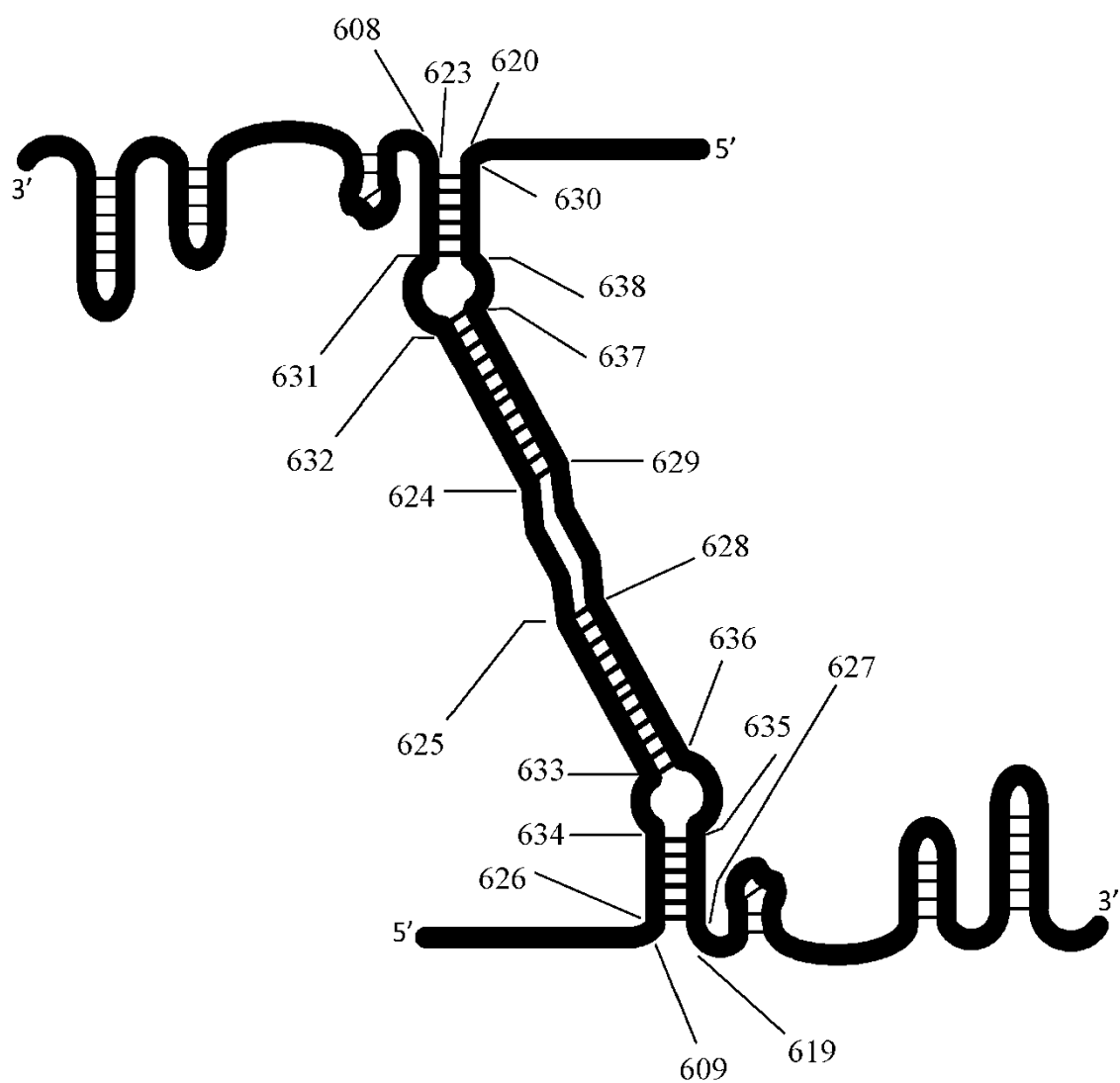


FIG. 6F

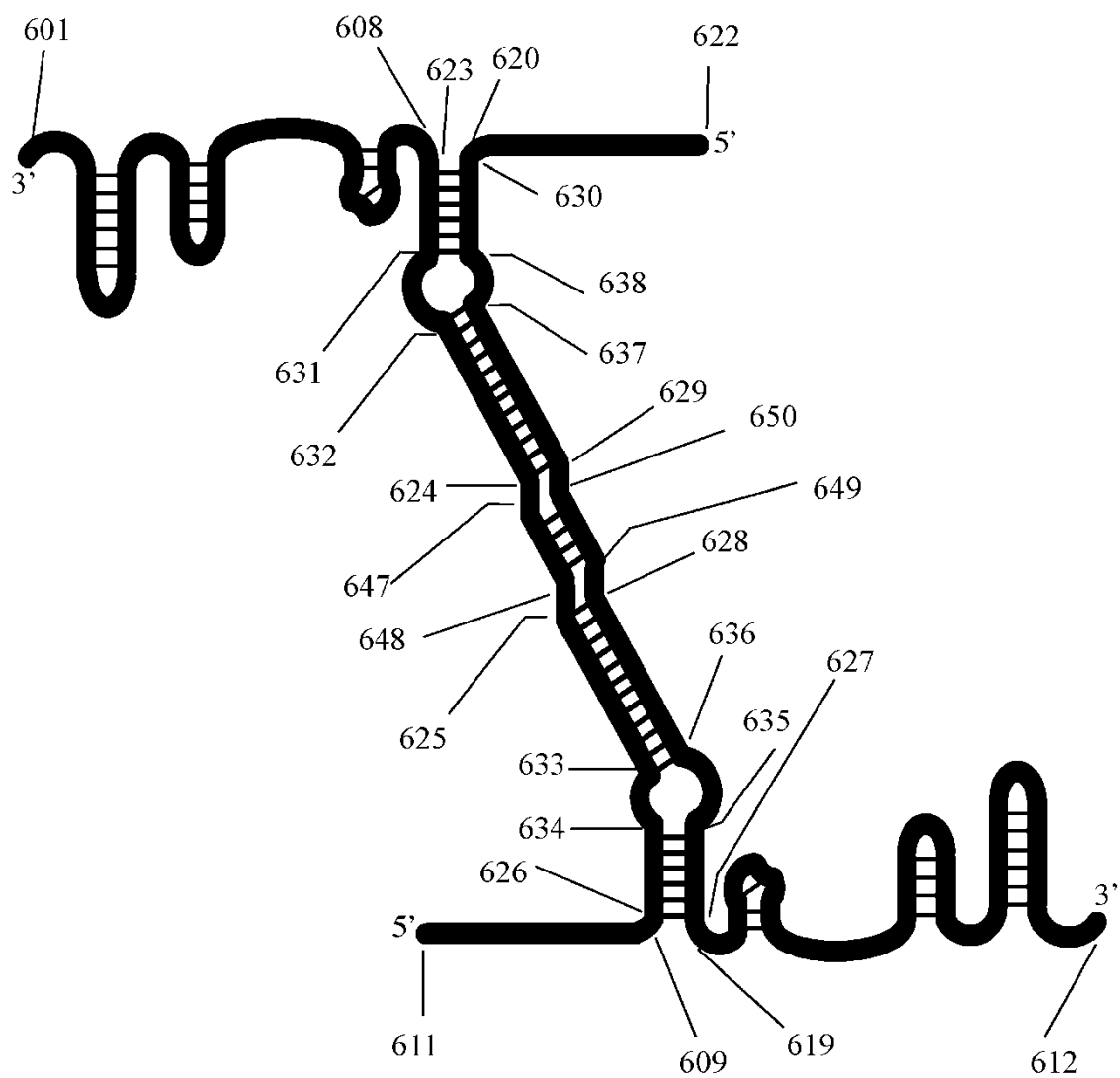


FIG. 6G

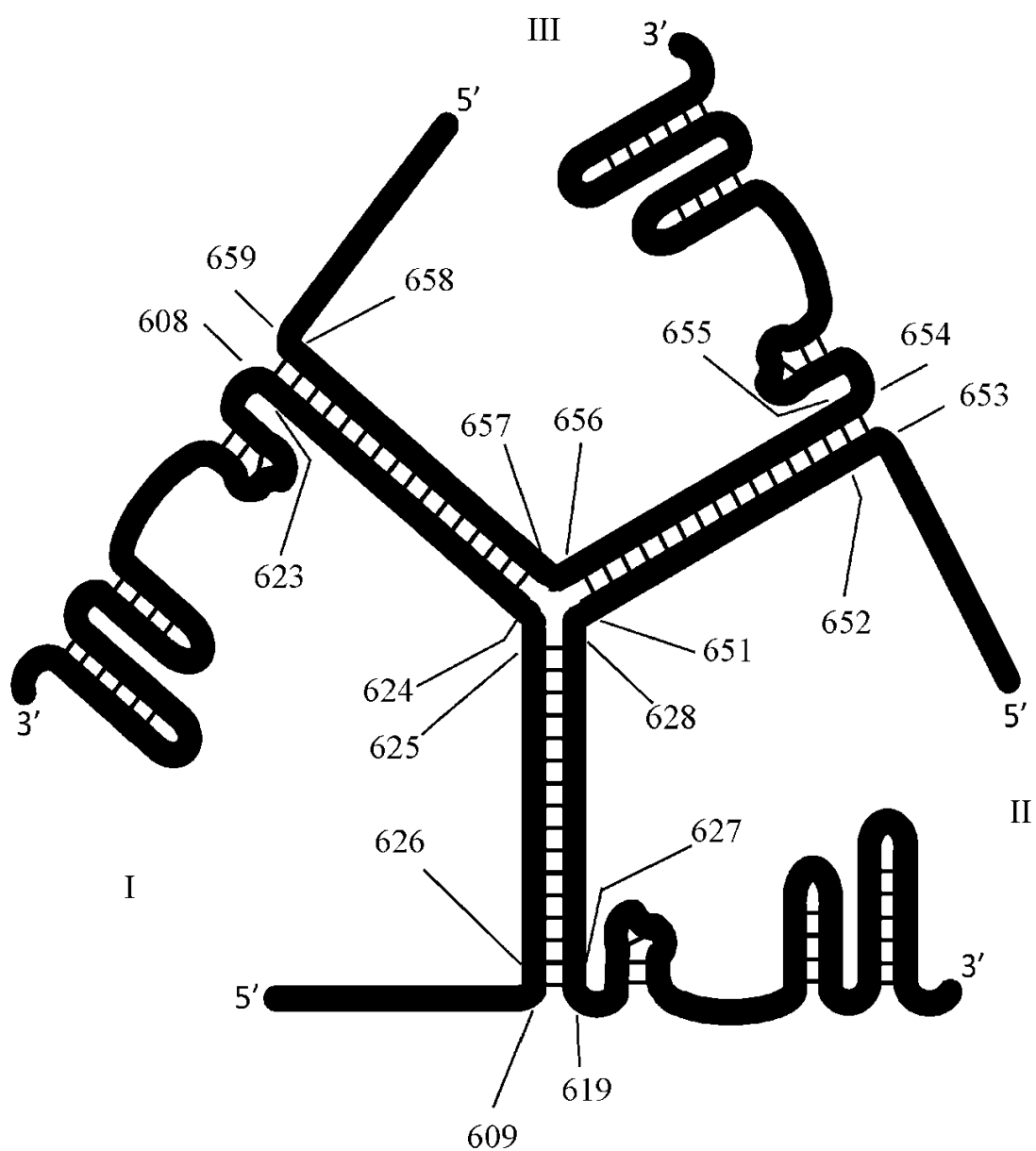


FIG. 6H

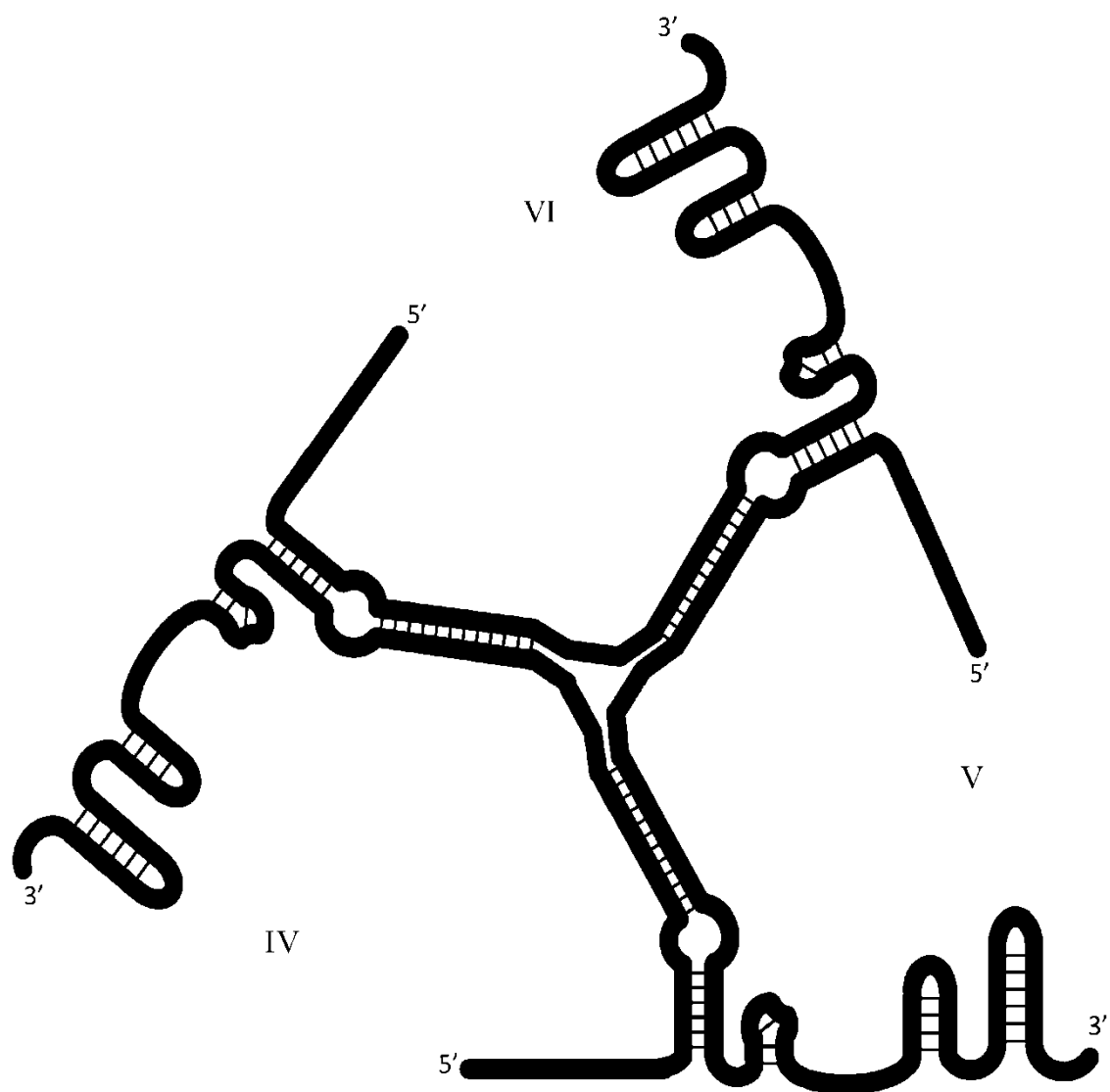


FIG. 6I

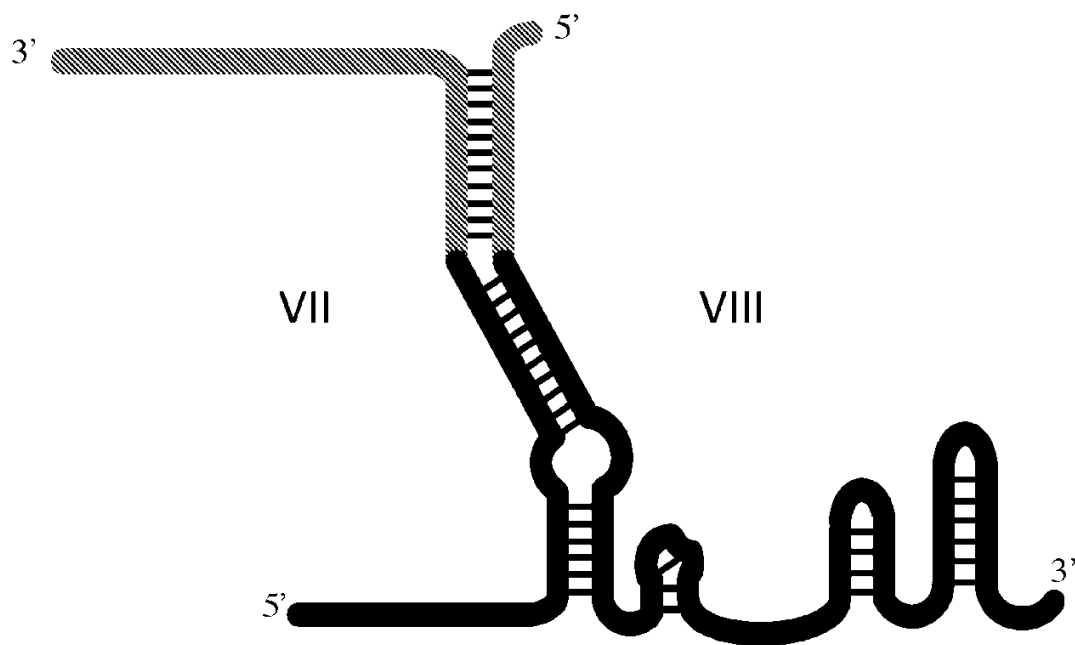


FIG. 6J

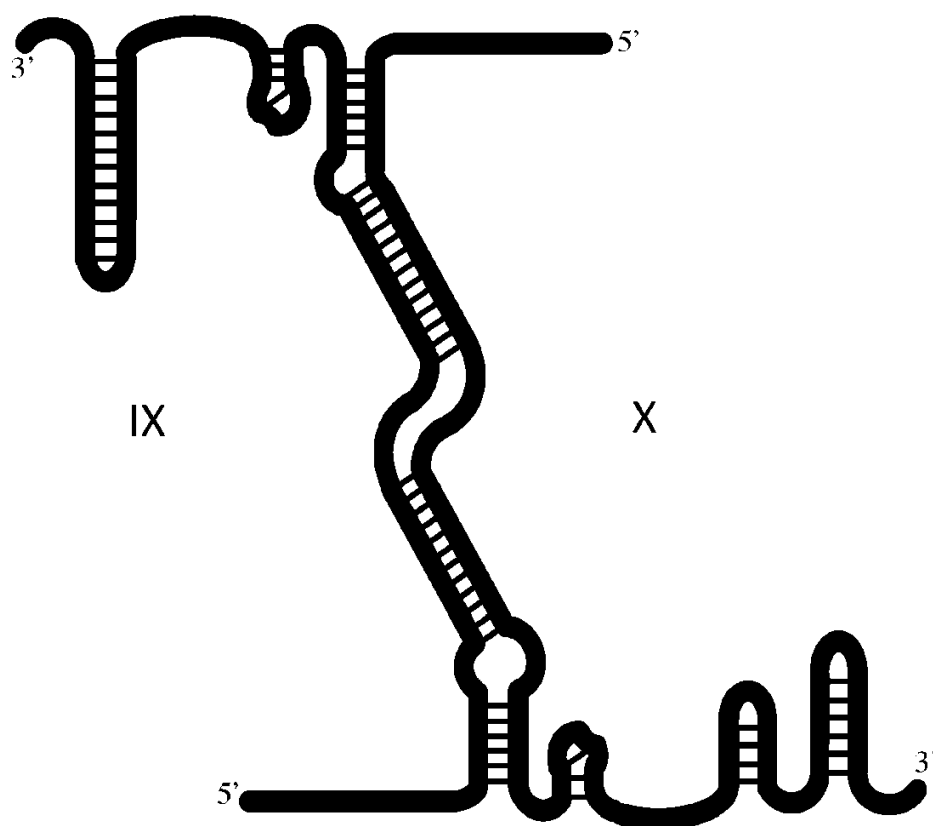


FIG. 6K

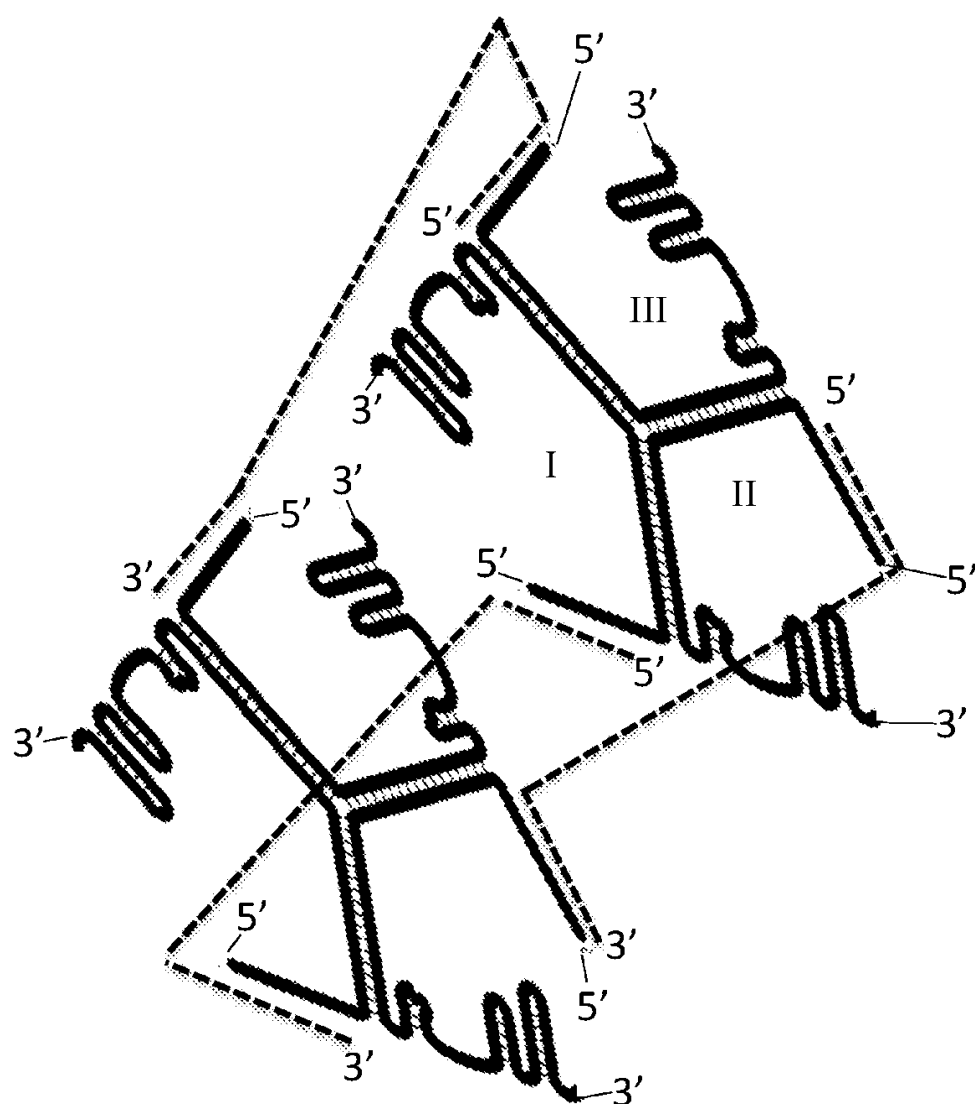


FIG. 6L

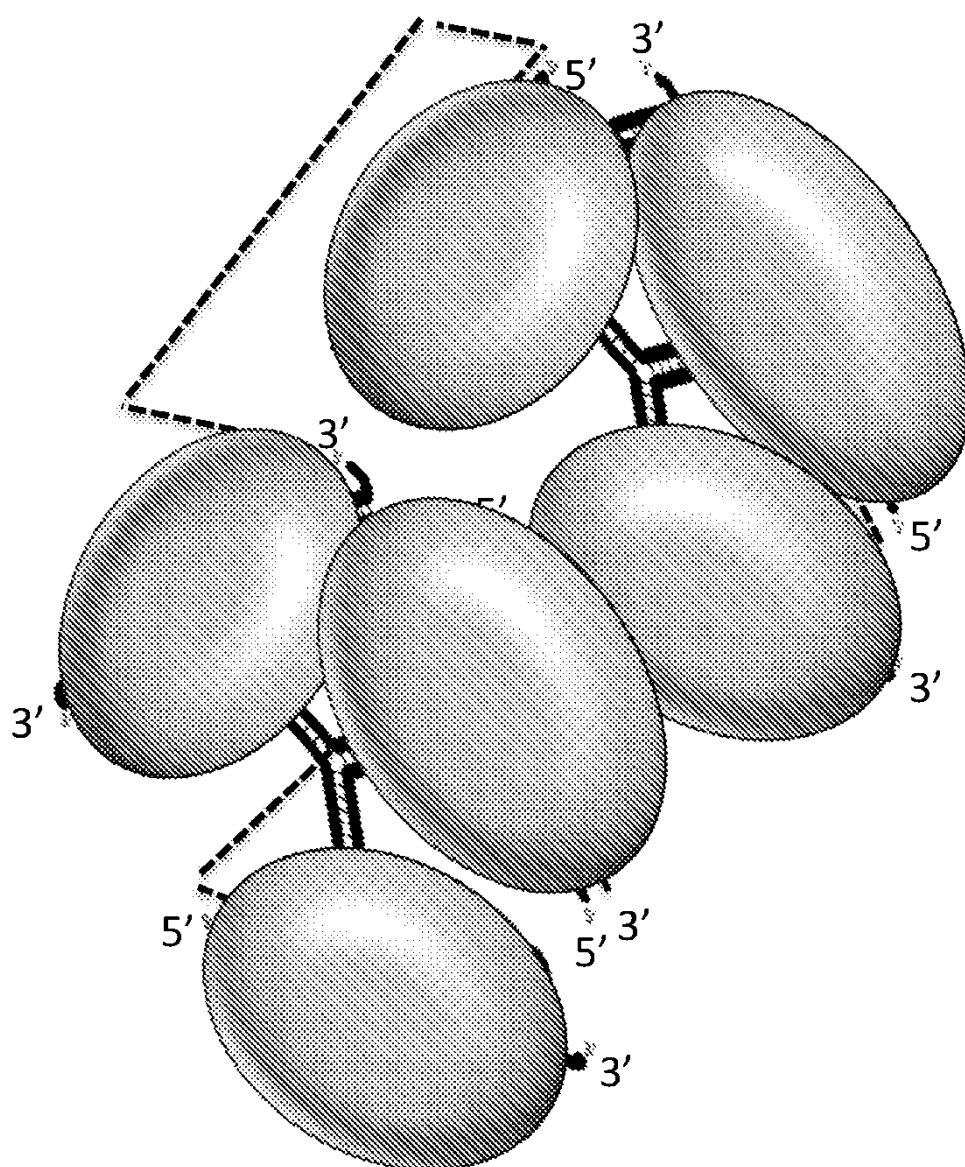


FIG. 6M

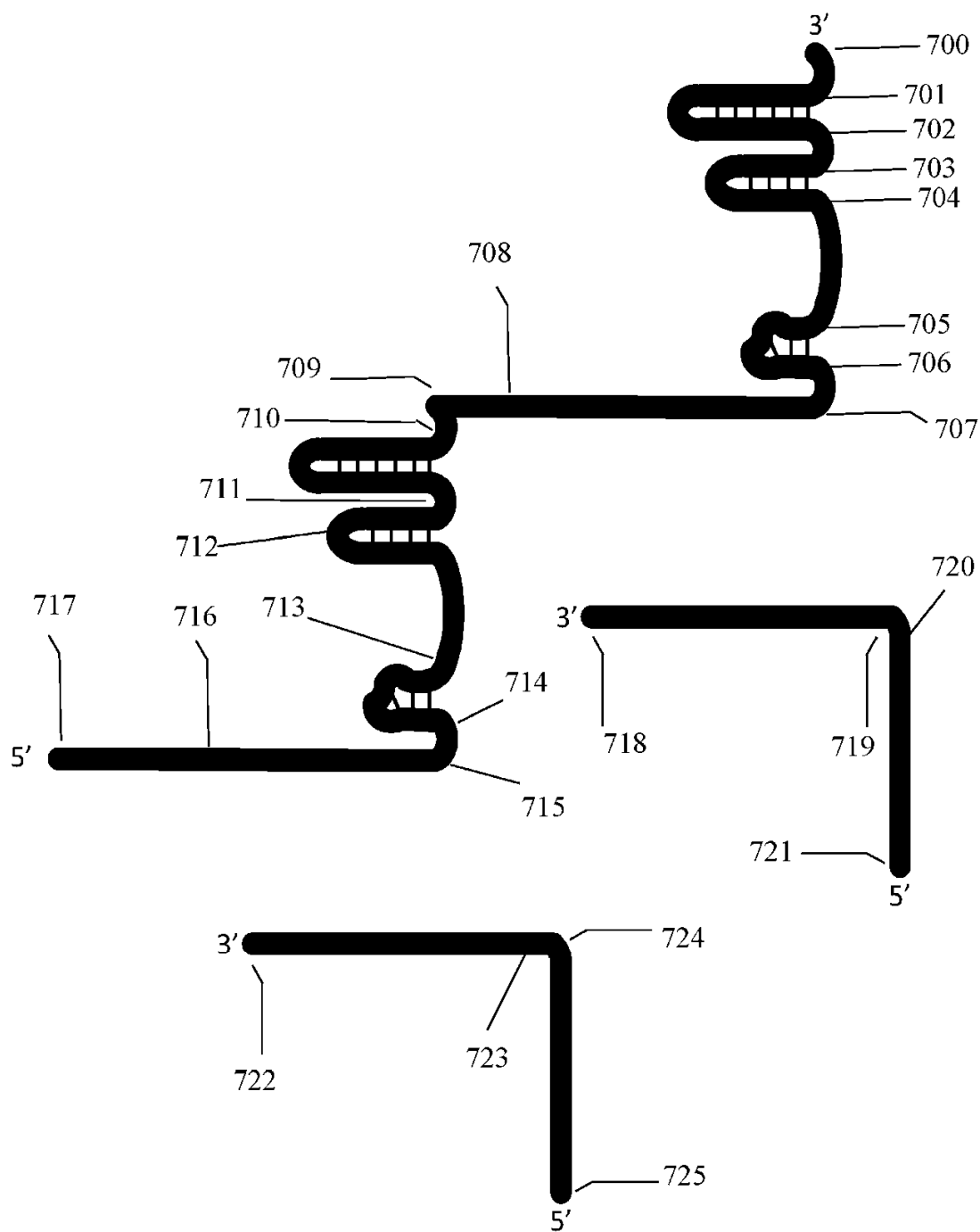


FIG. 7A

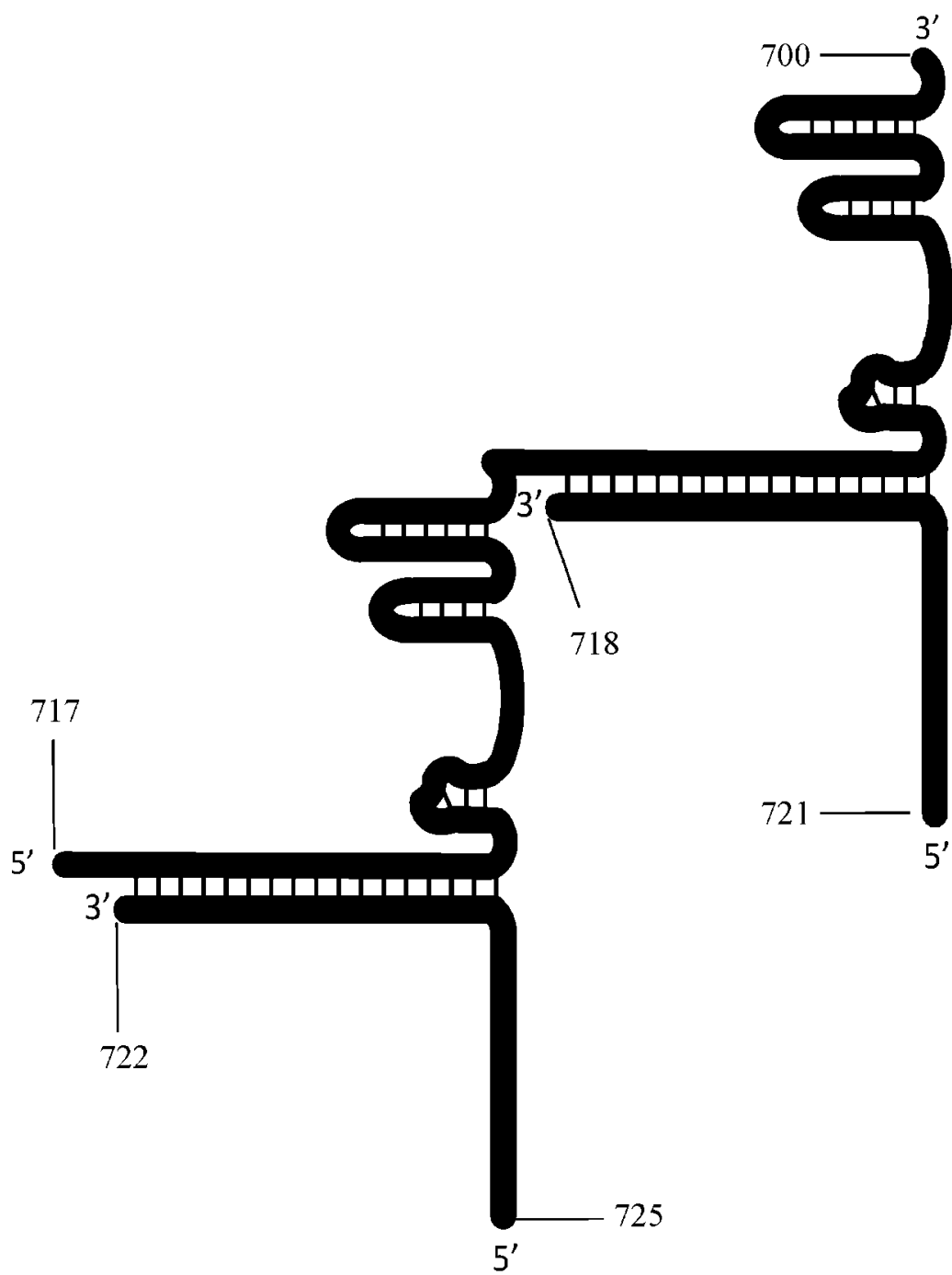


FIG. 7B

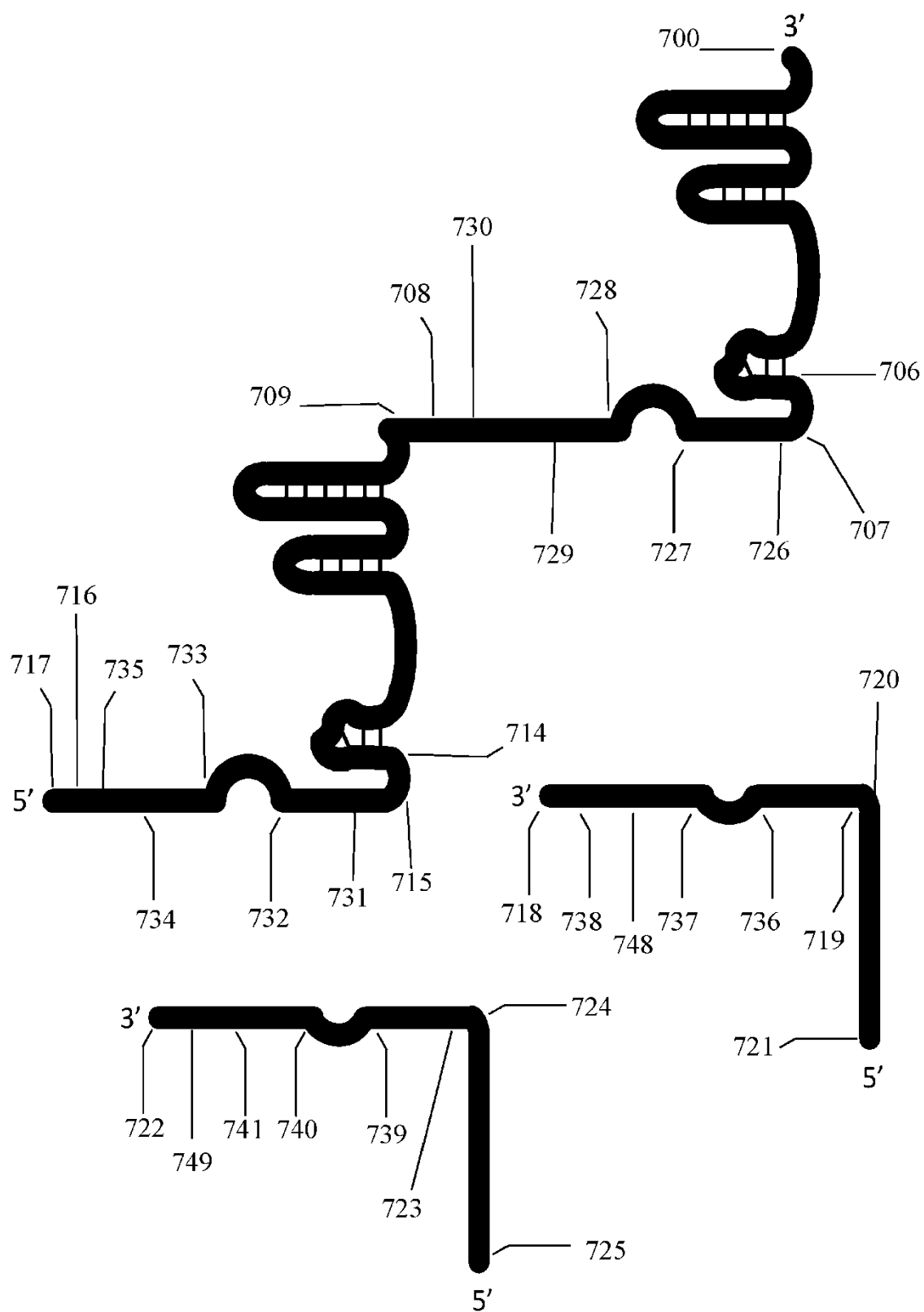


FIG. 7C

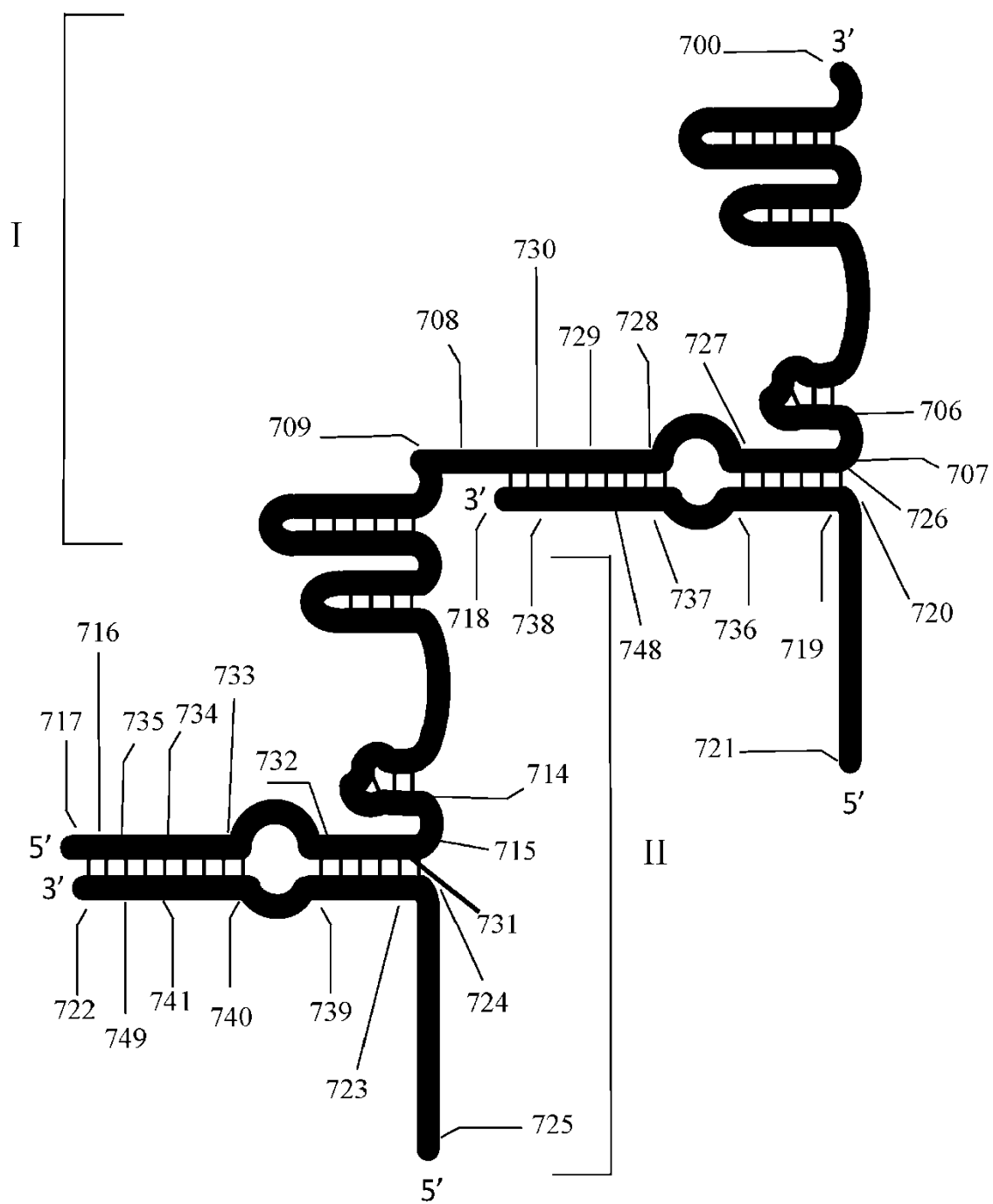


FIG. 7D

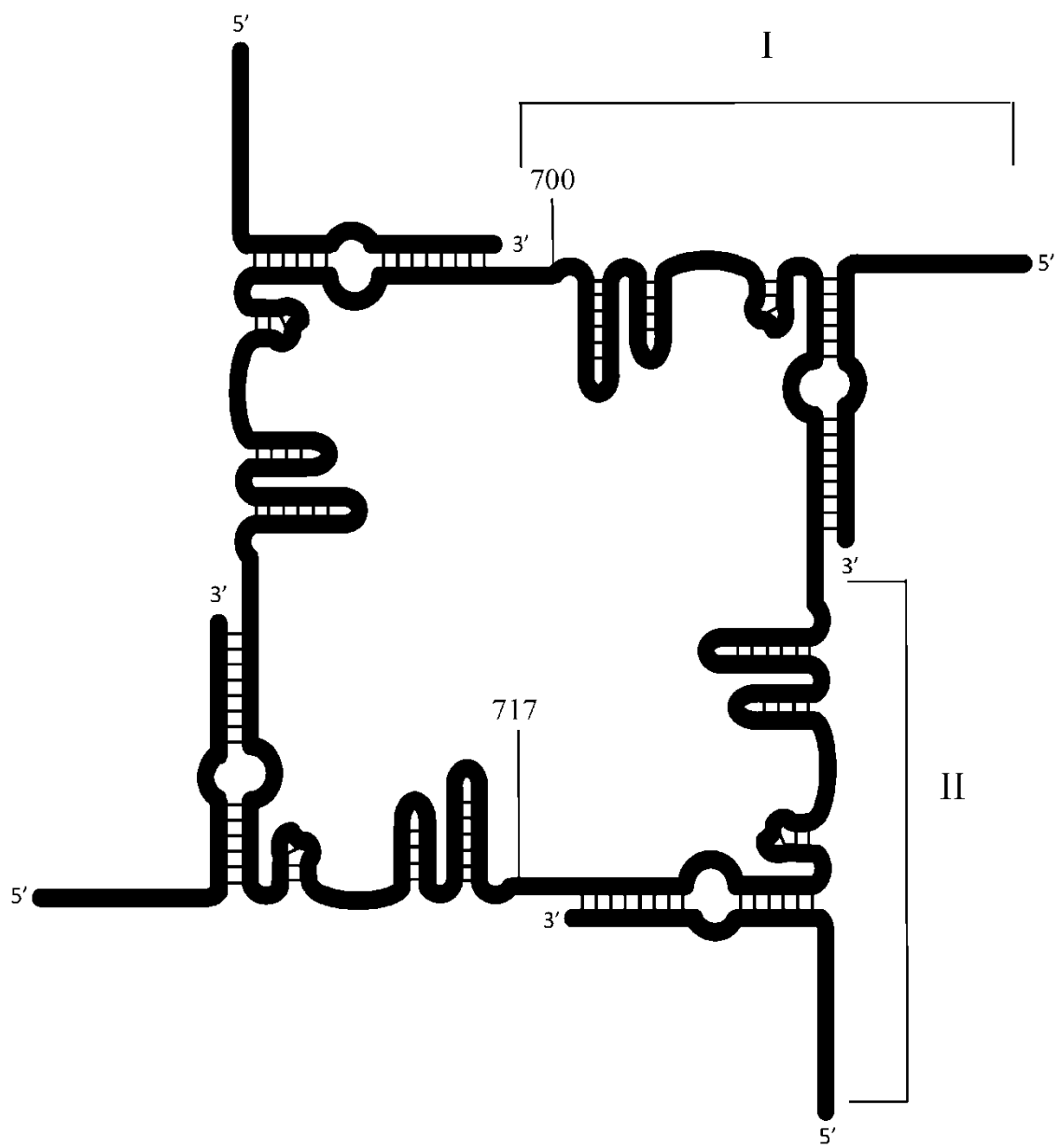


FIG. 7E

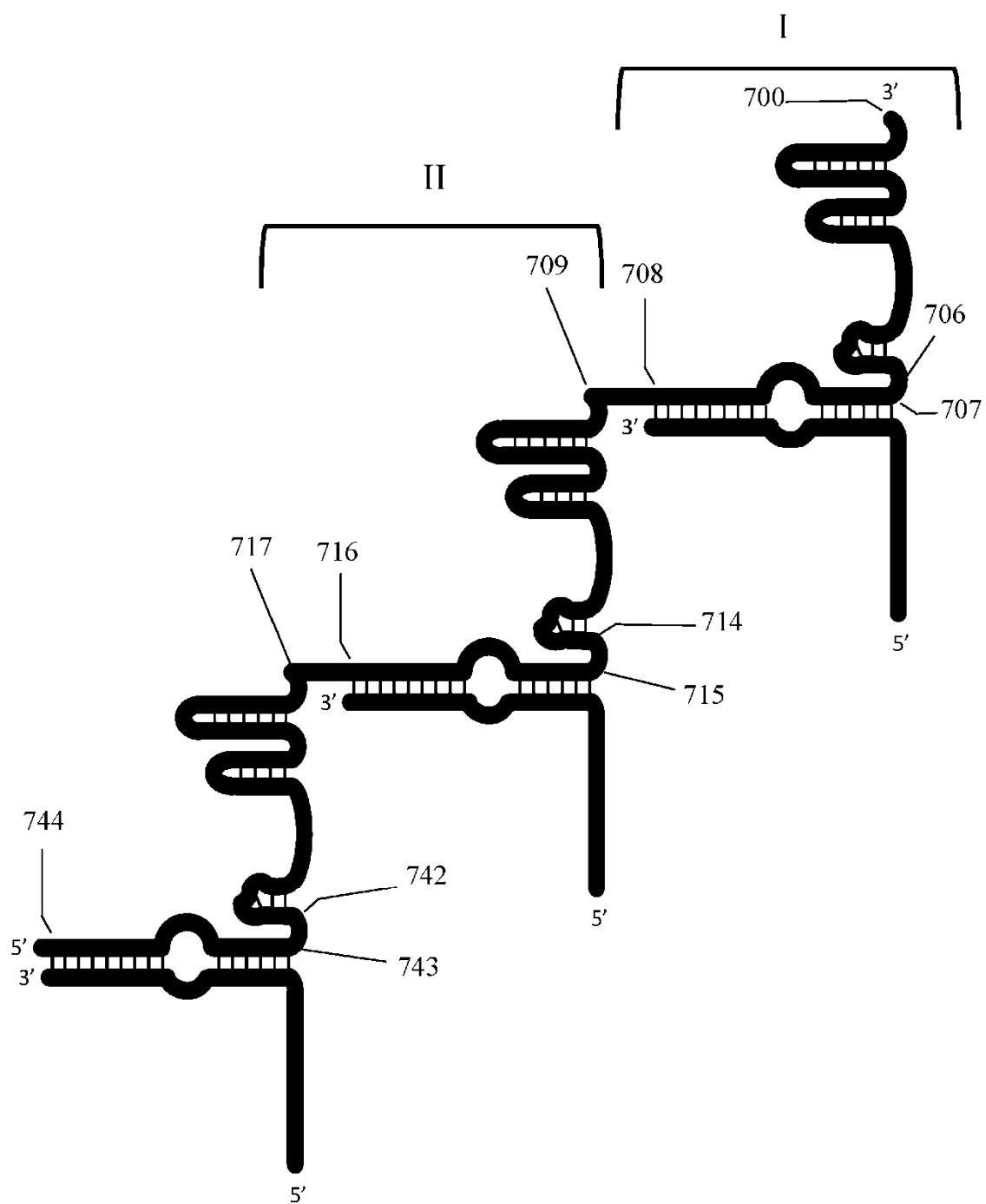
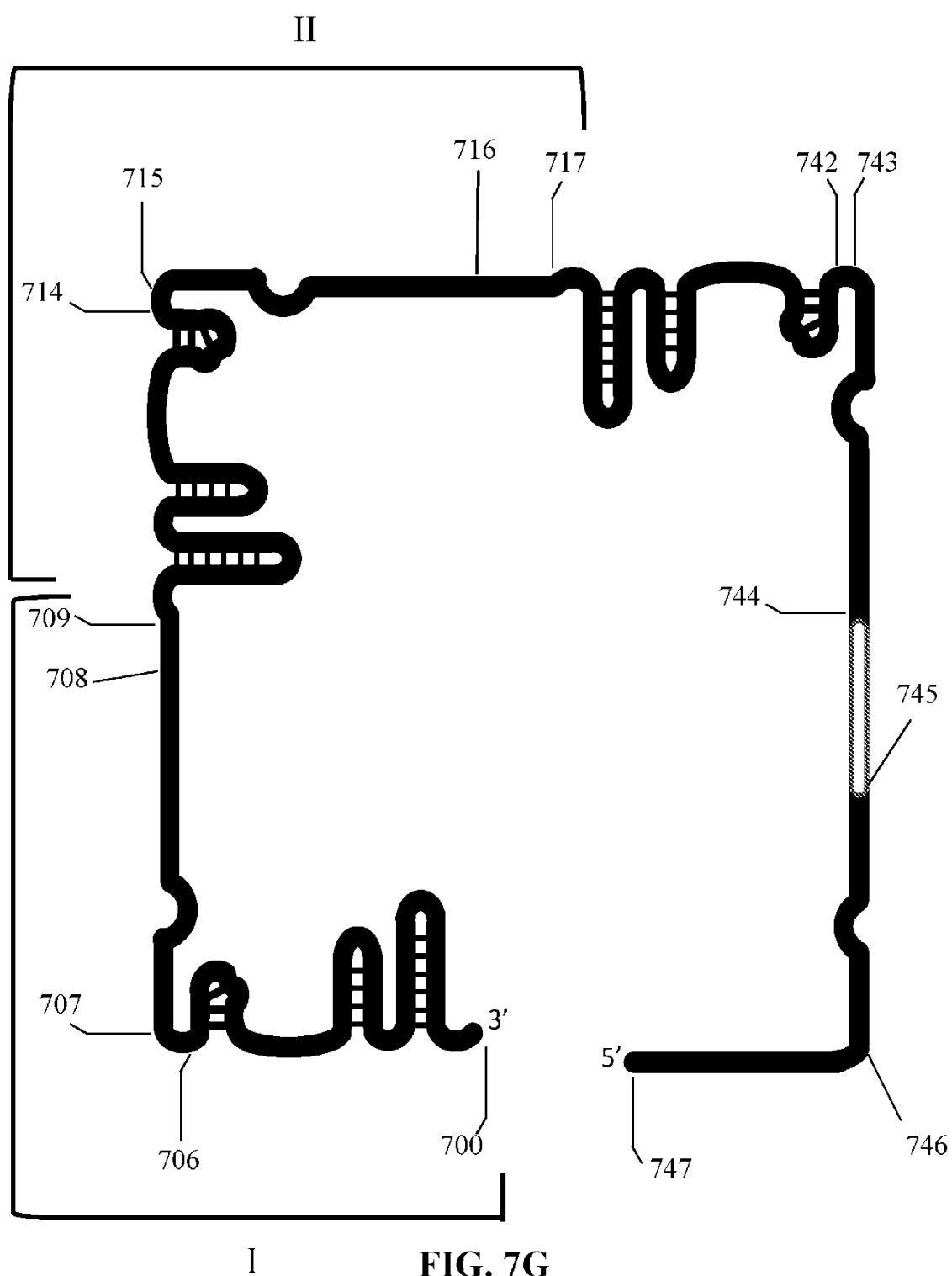


FIG. 7F



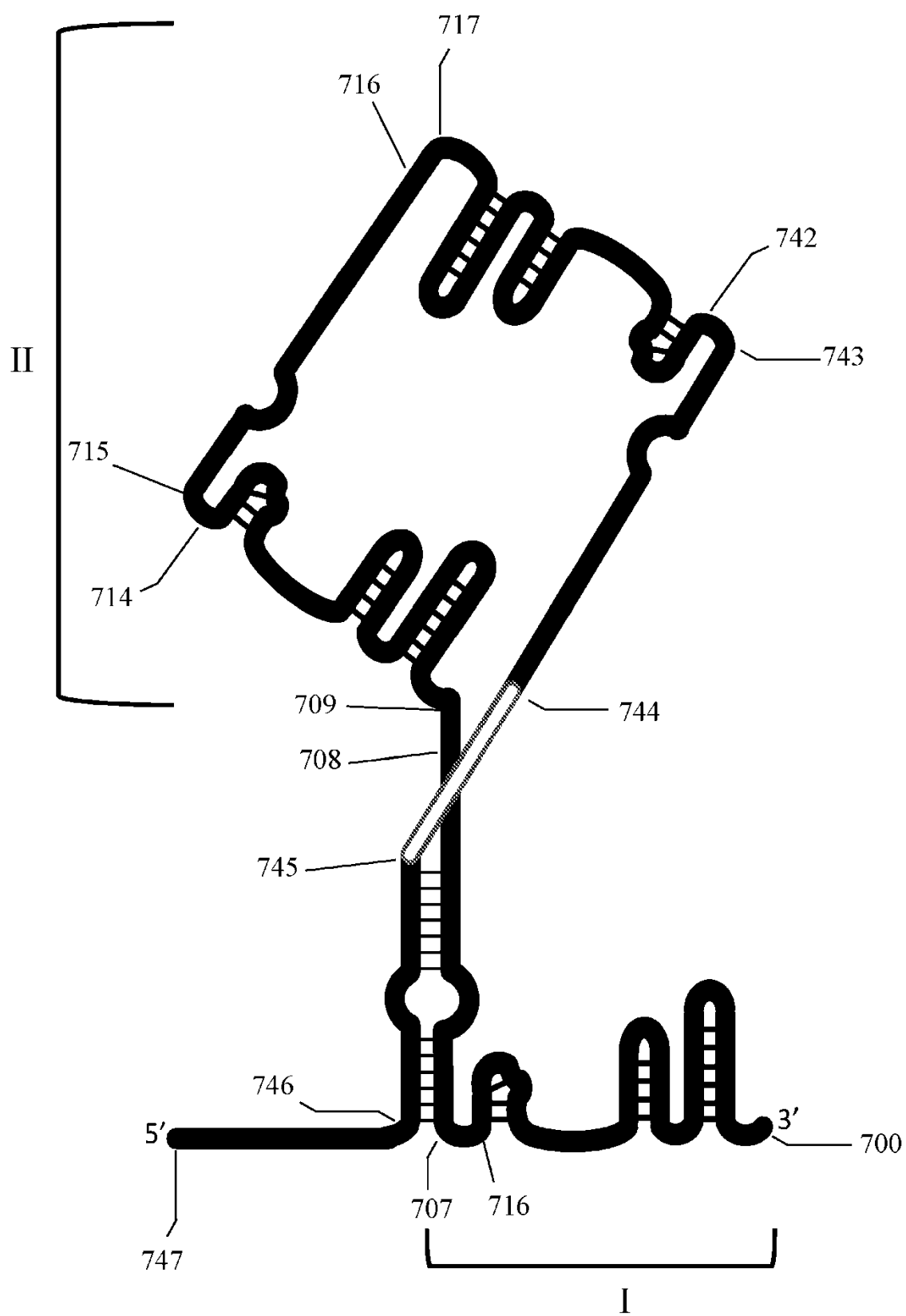


FIG. 7H

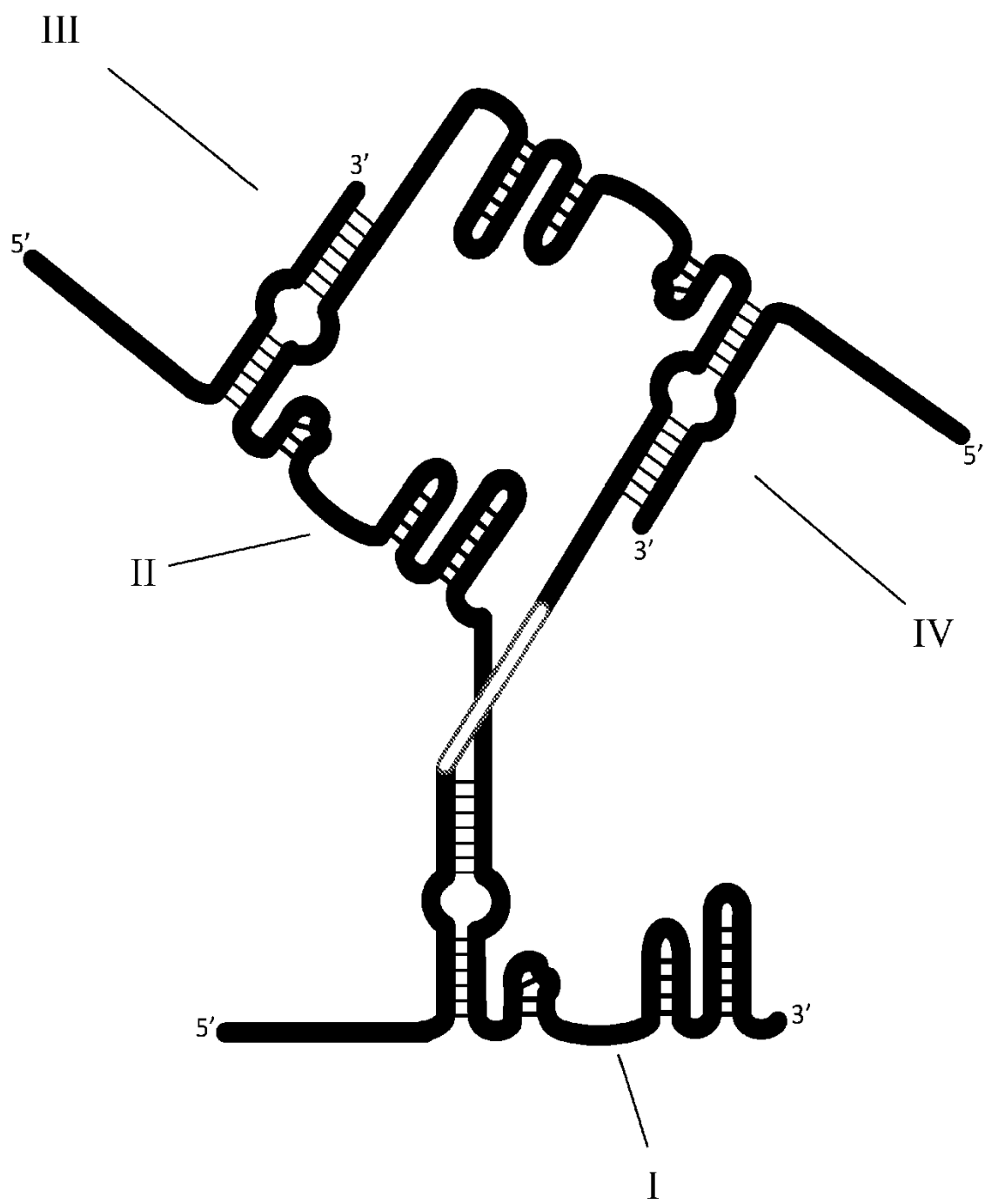


FIG. 71

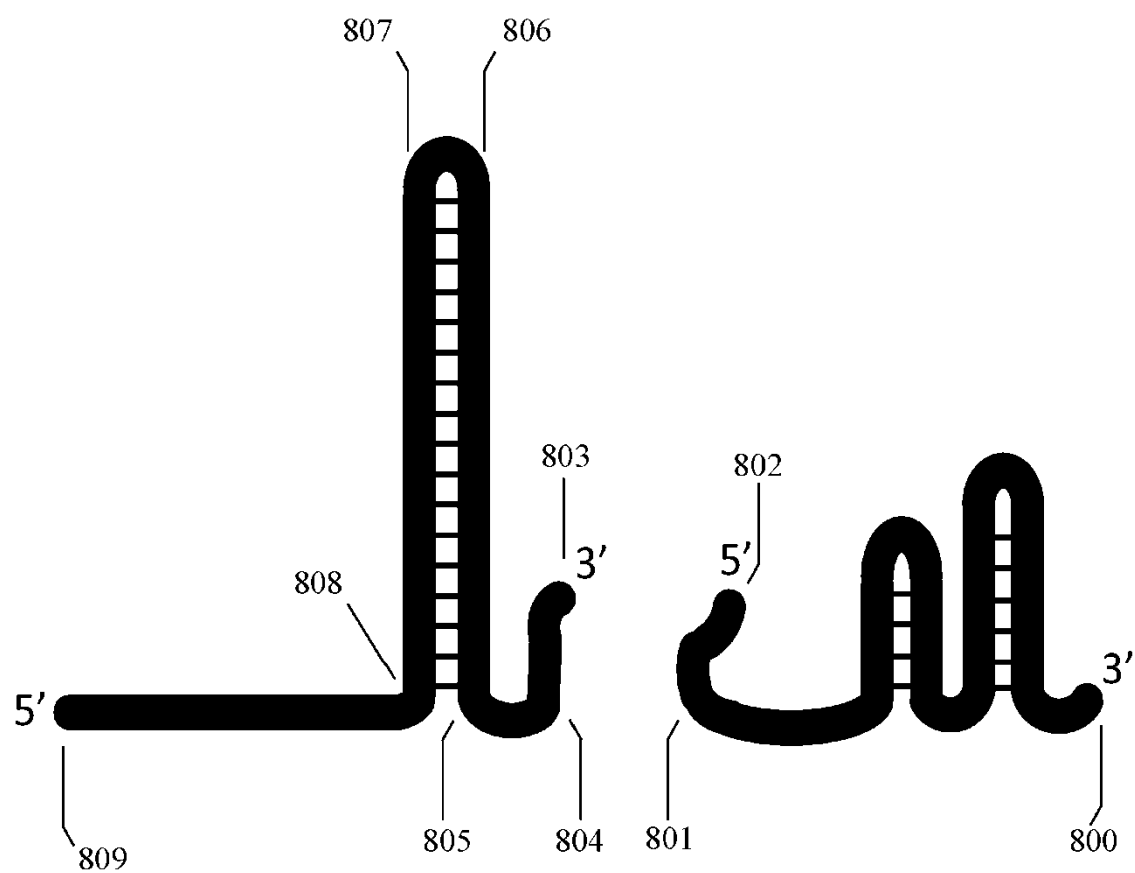


FIG. 8A

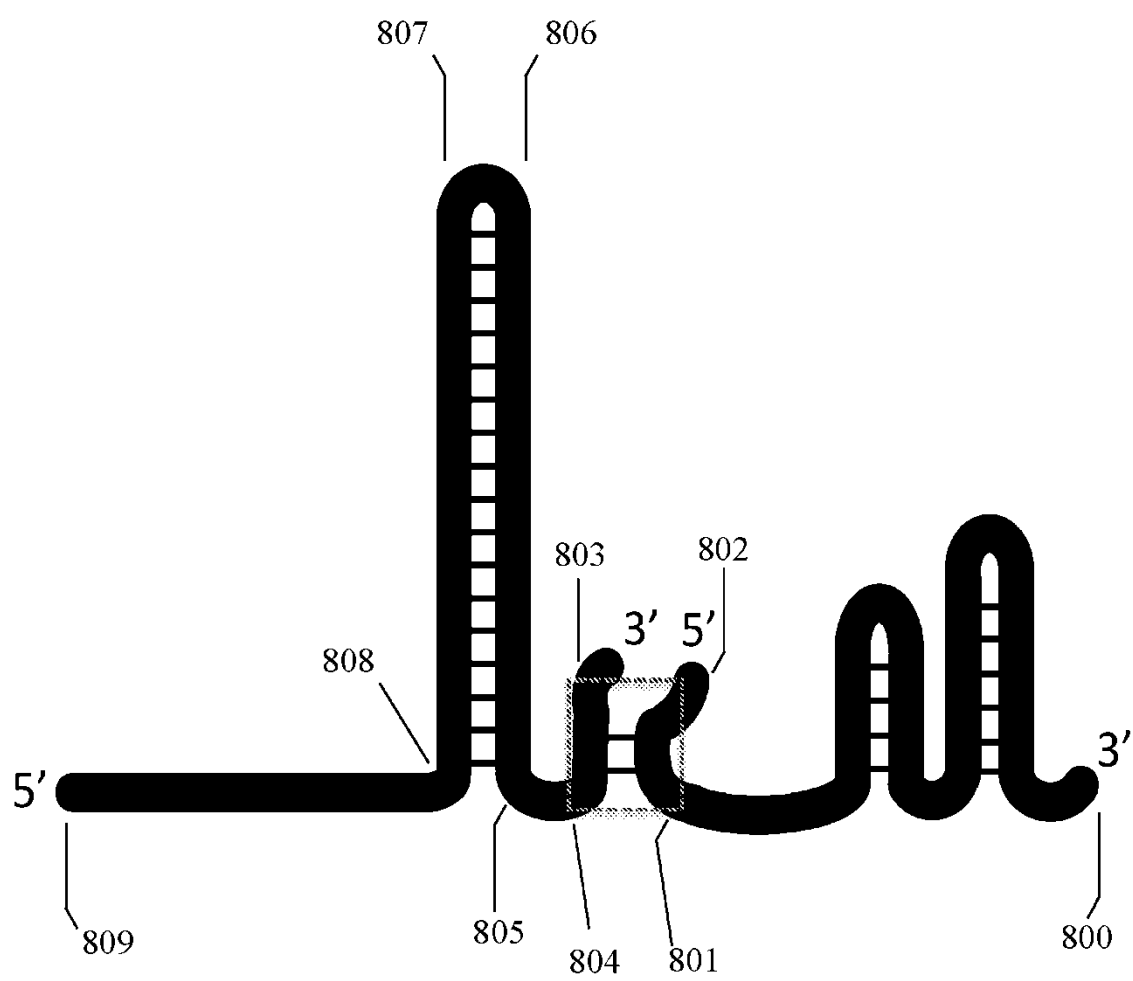


FIG. 8B

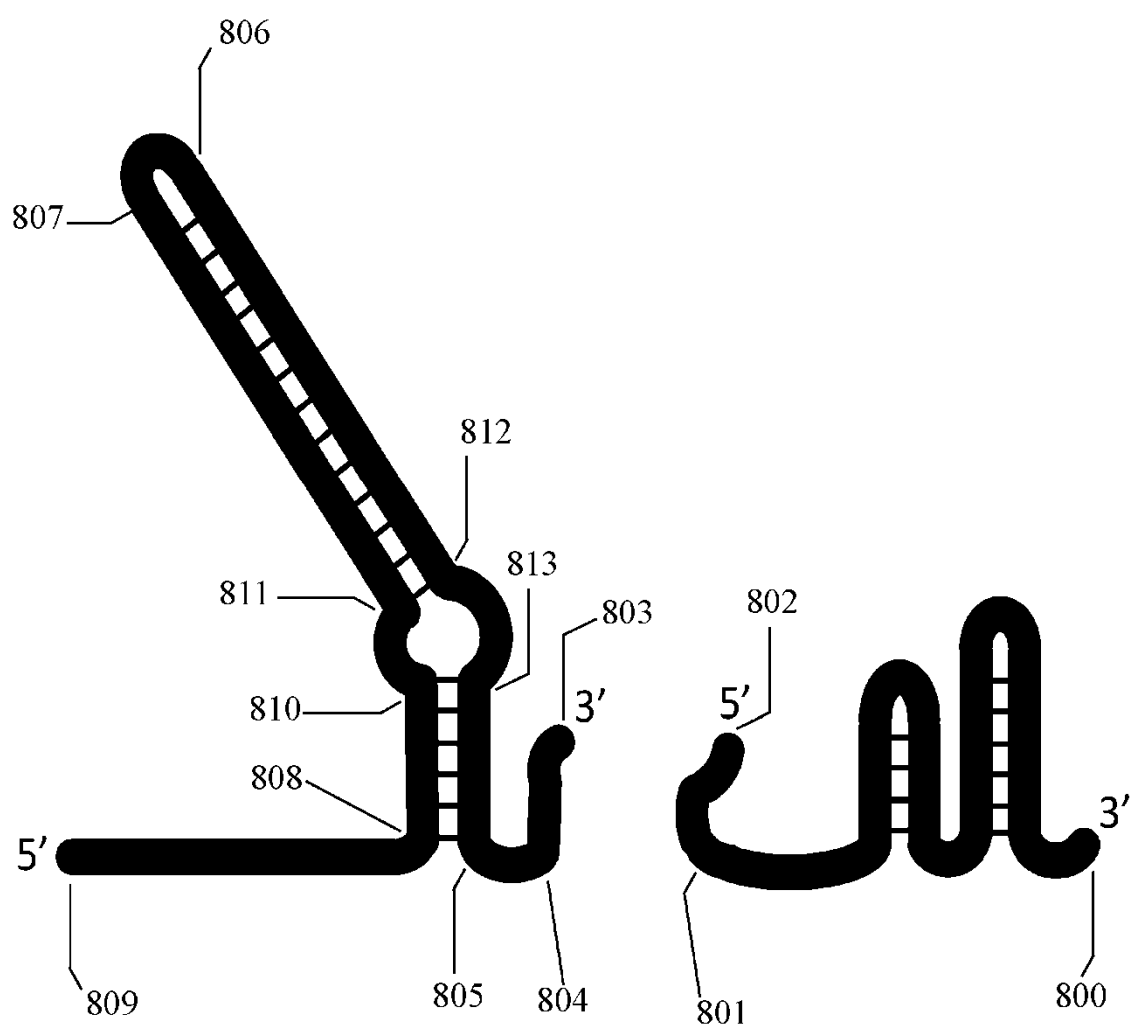


FIG. 8C

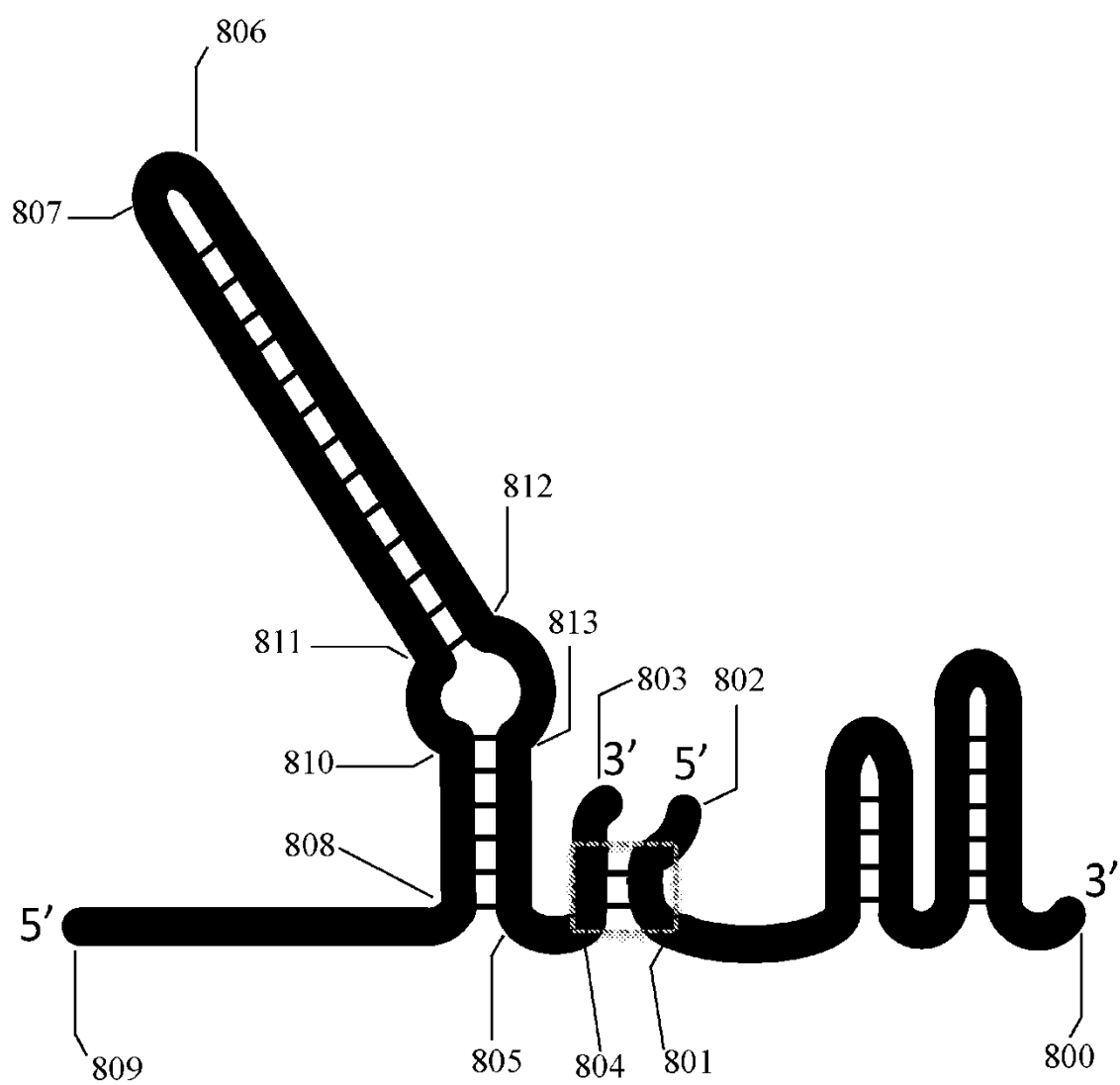


FIG. 8D

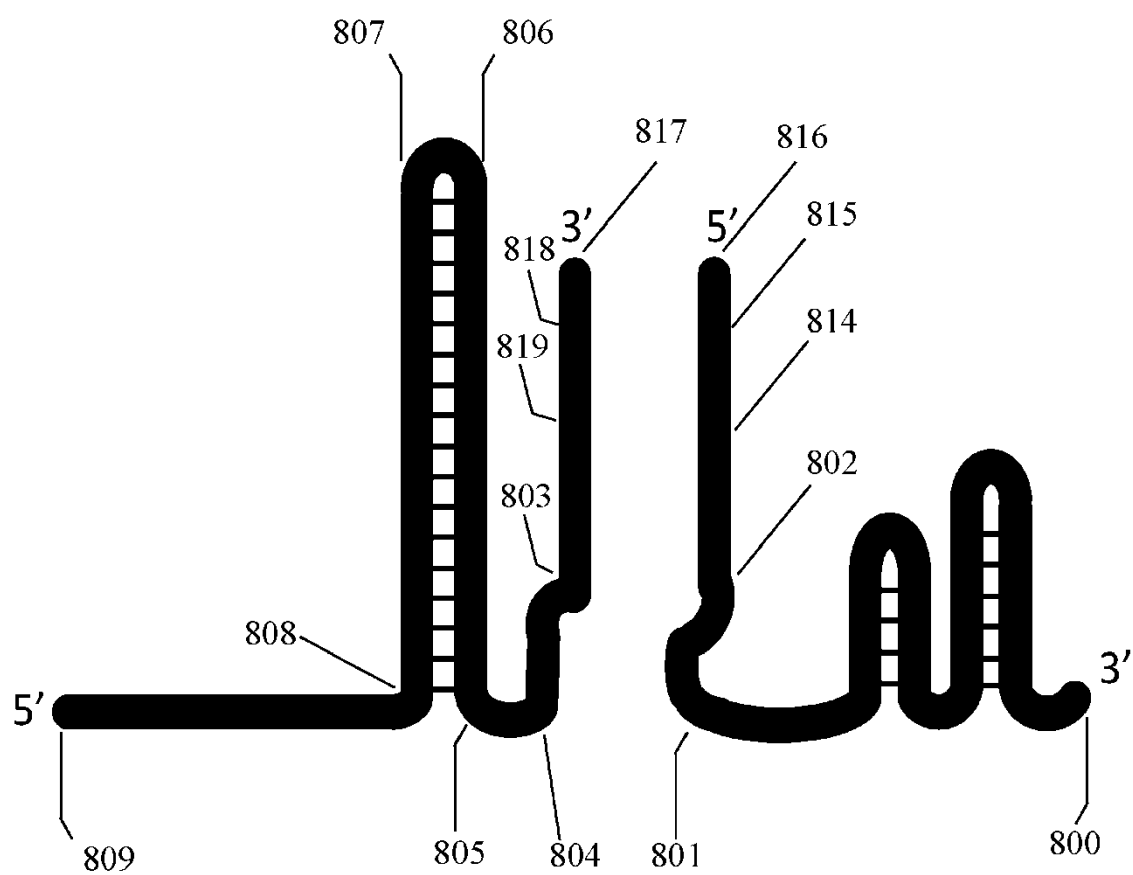


FIG. 8E

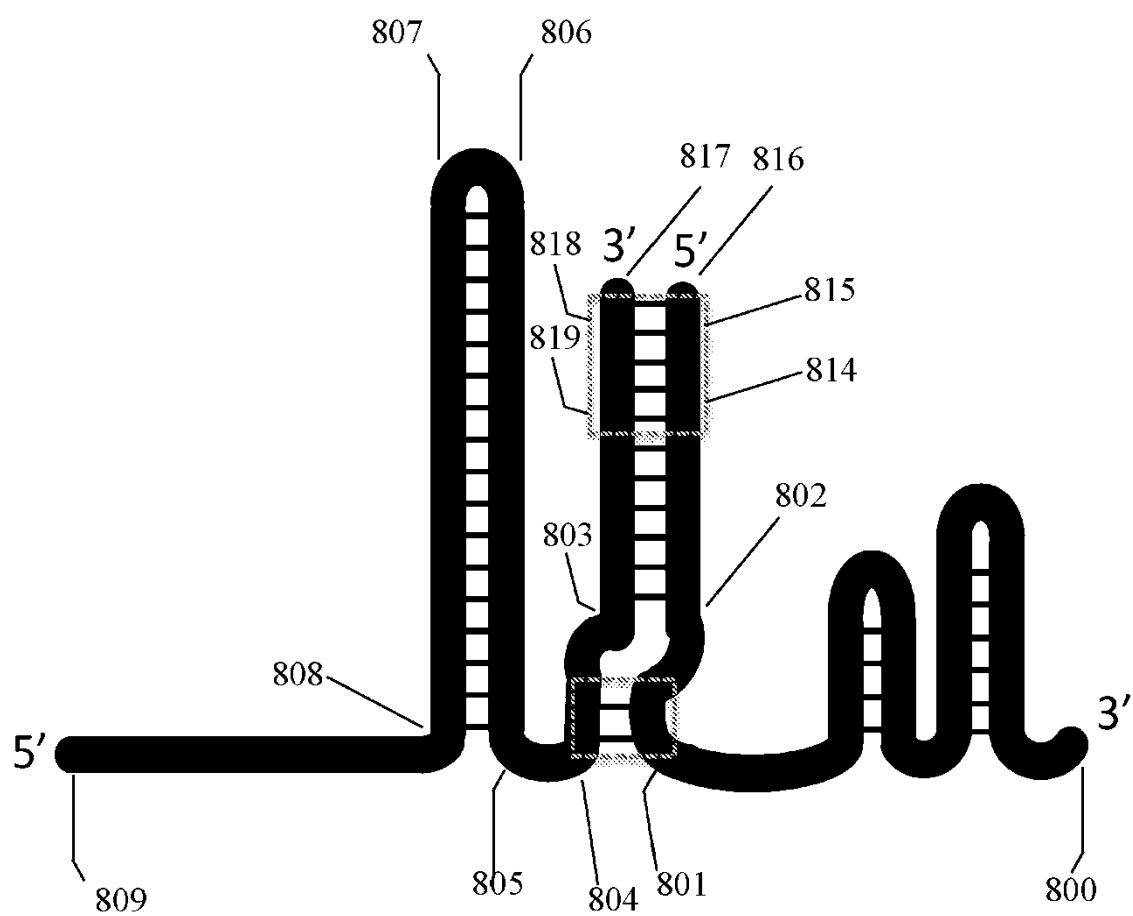


FIG. 8F

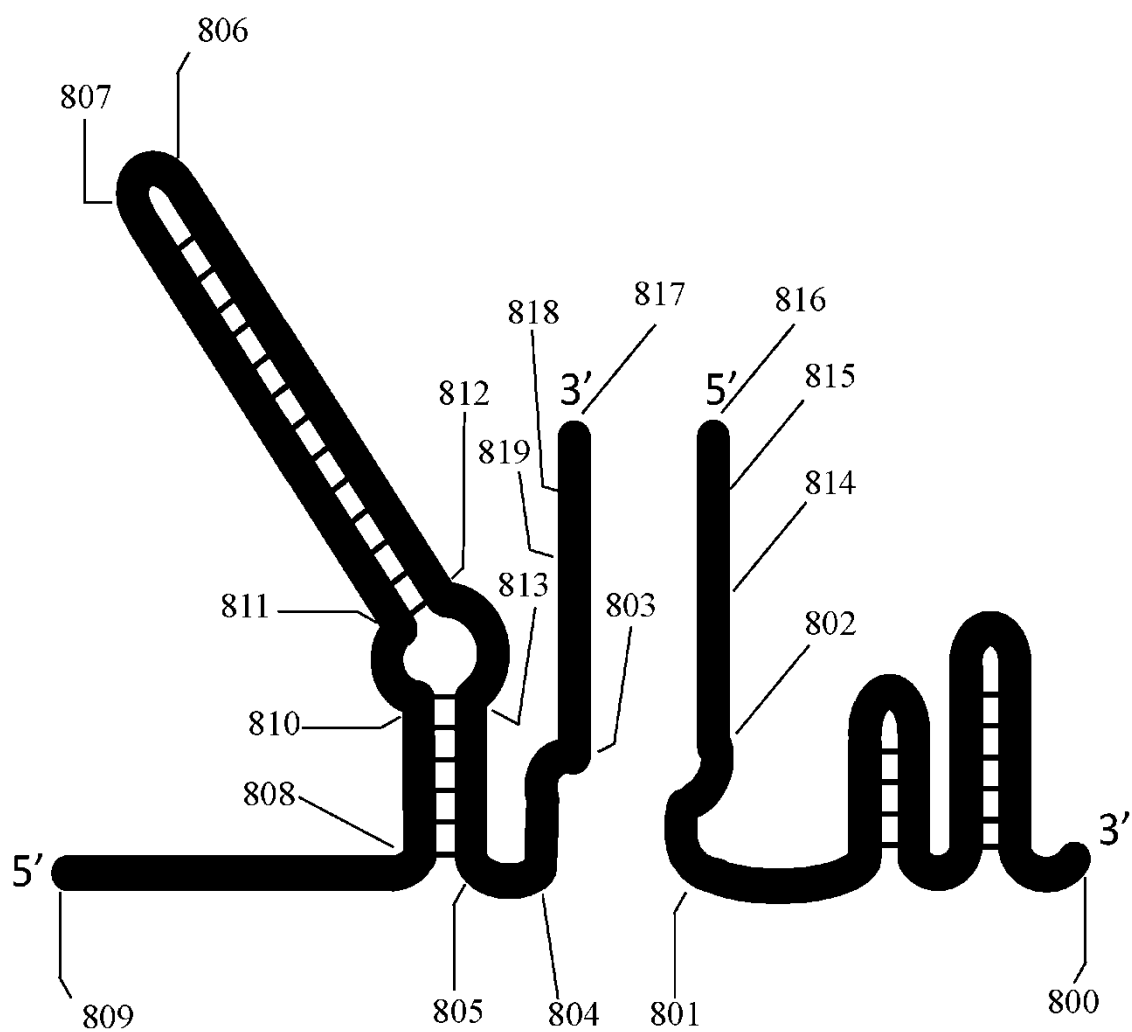


FIG. 8G

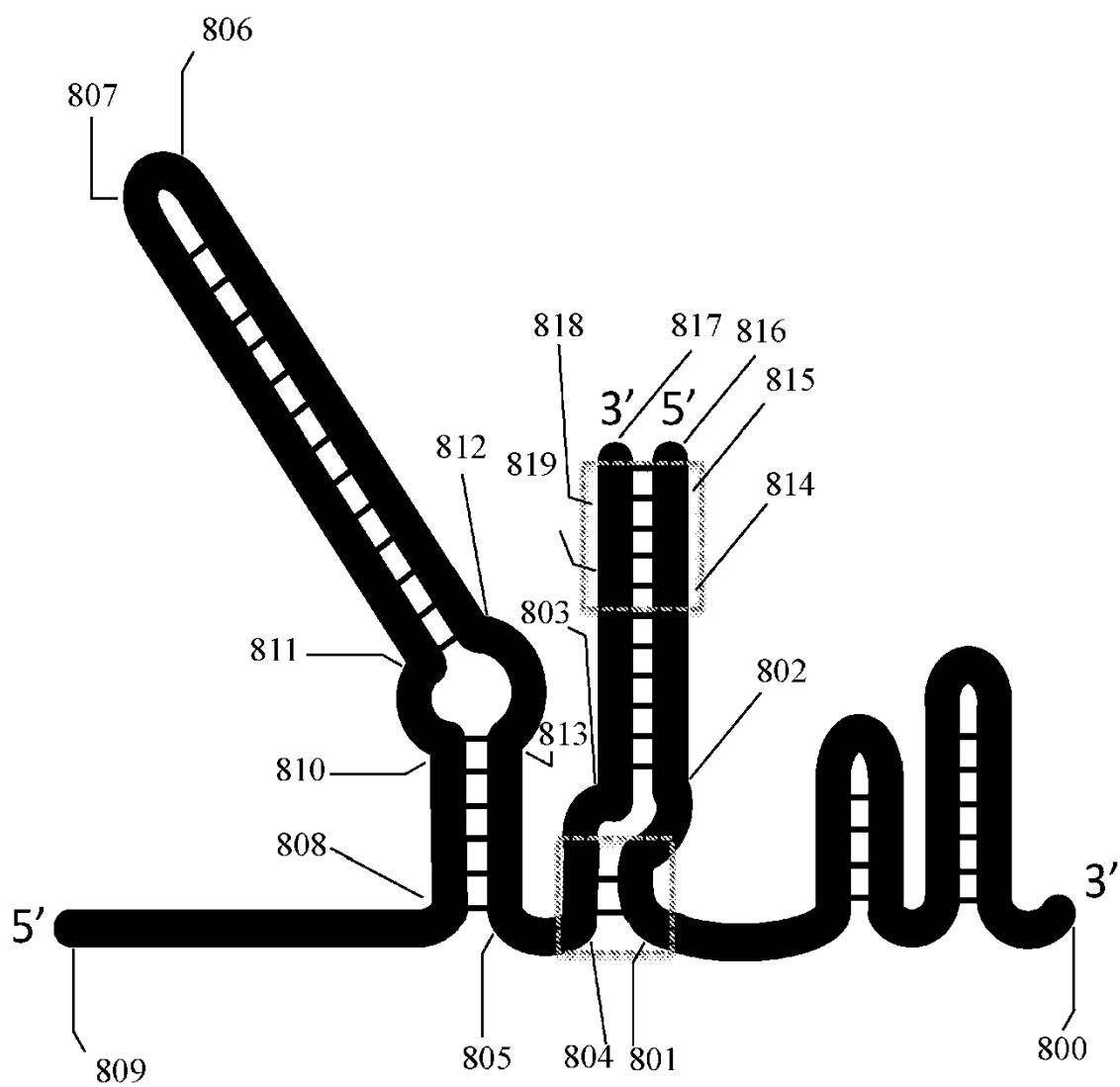


FIG. 8H

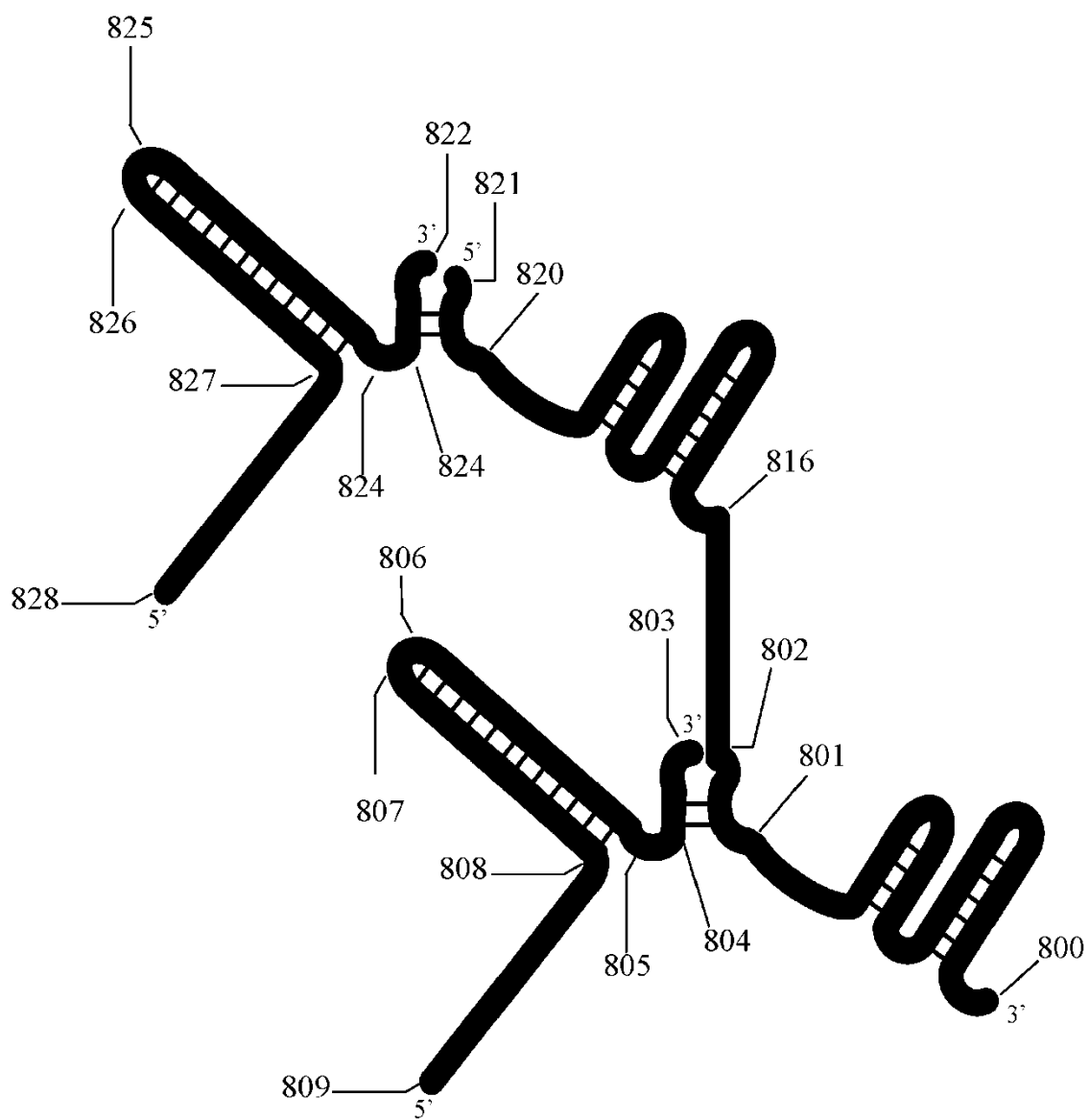


FIG. 8I

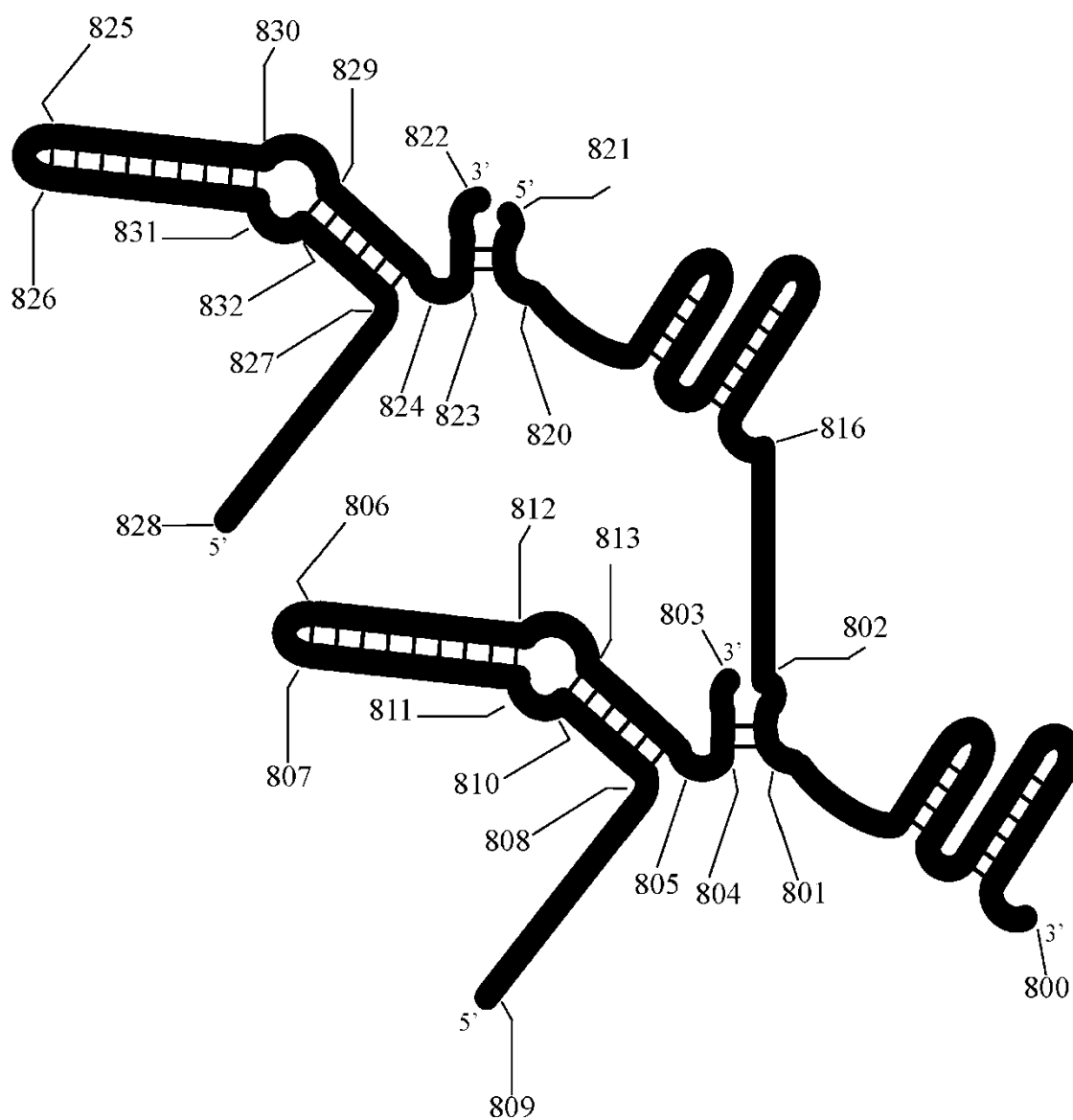


FIG. 8J

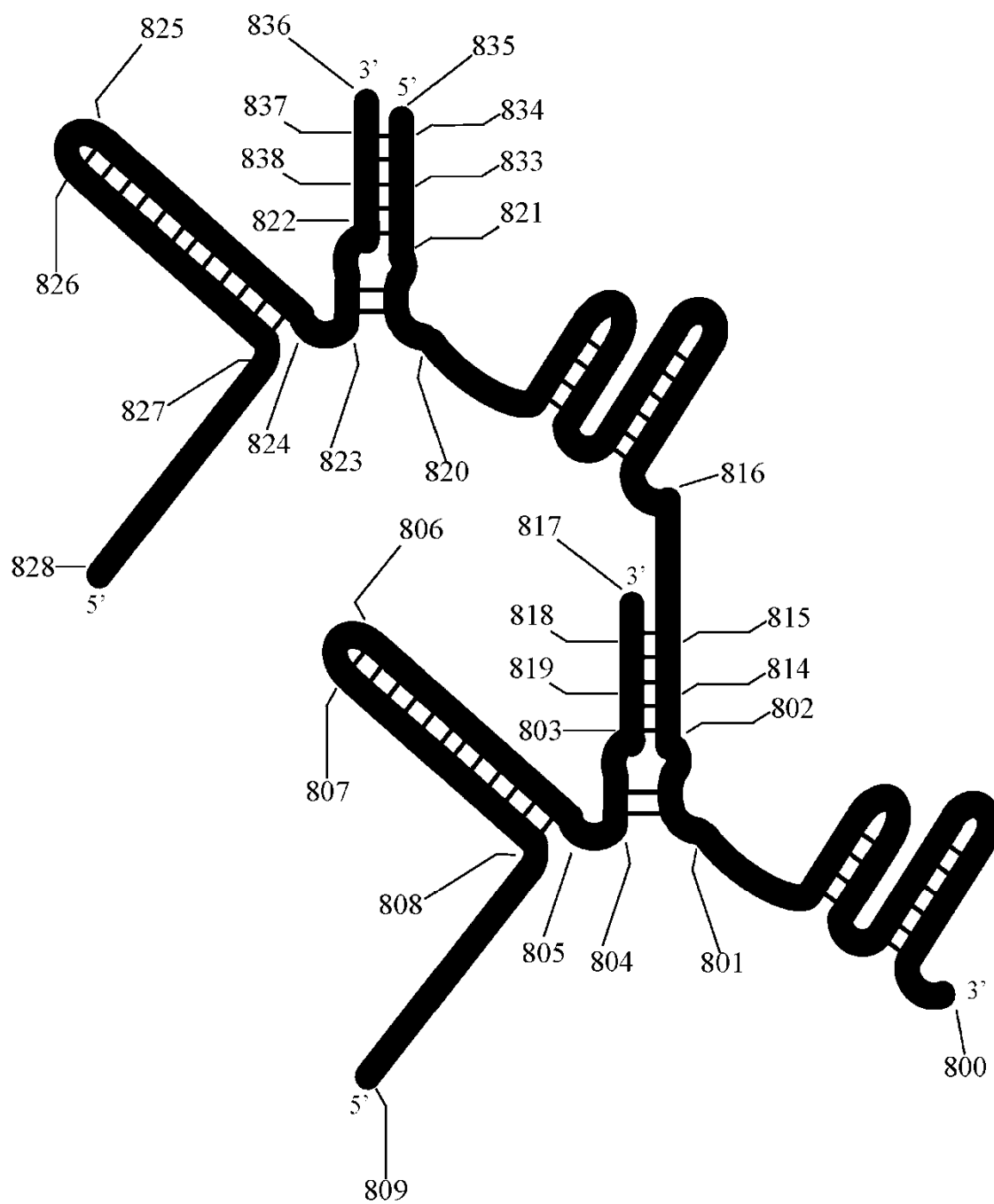


FIG. 8K

II

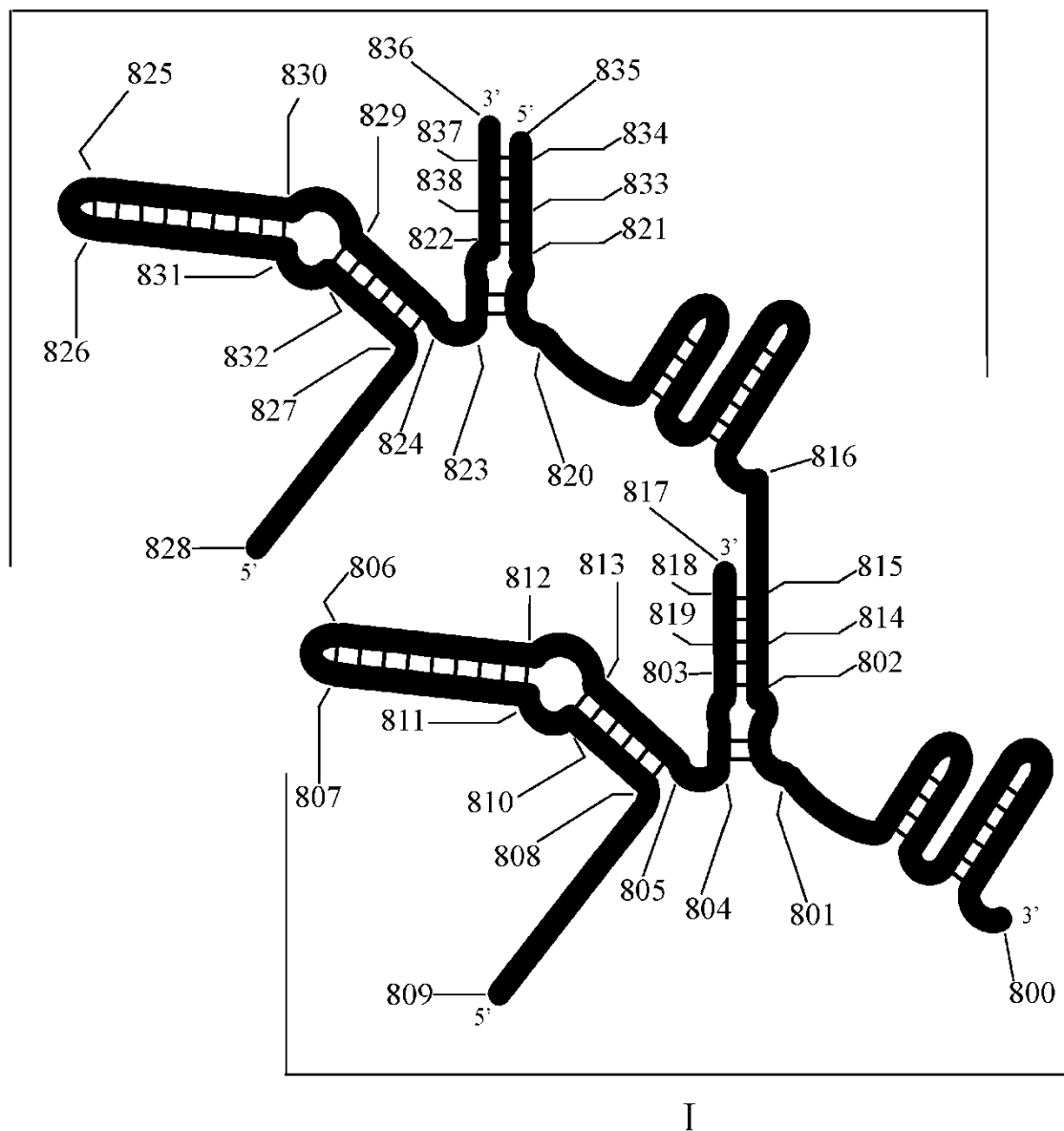


FIG. 8L

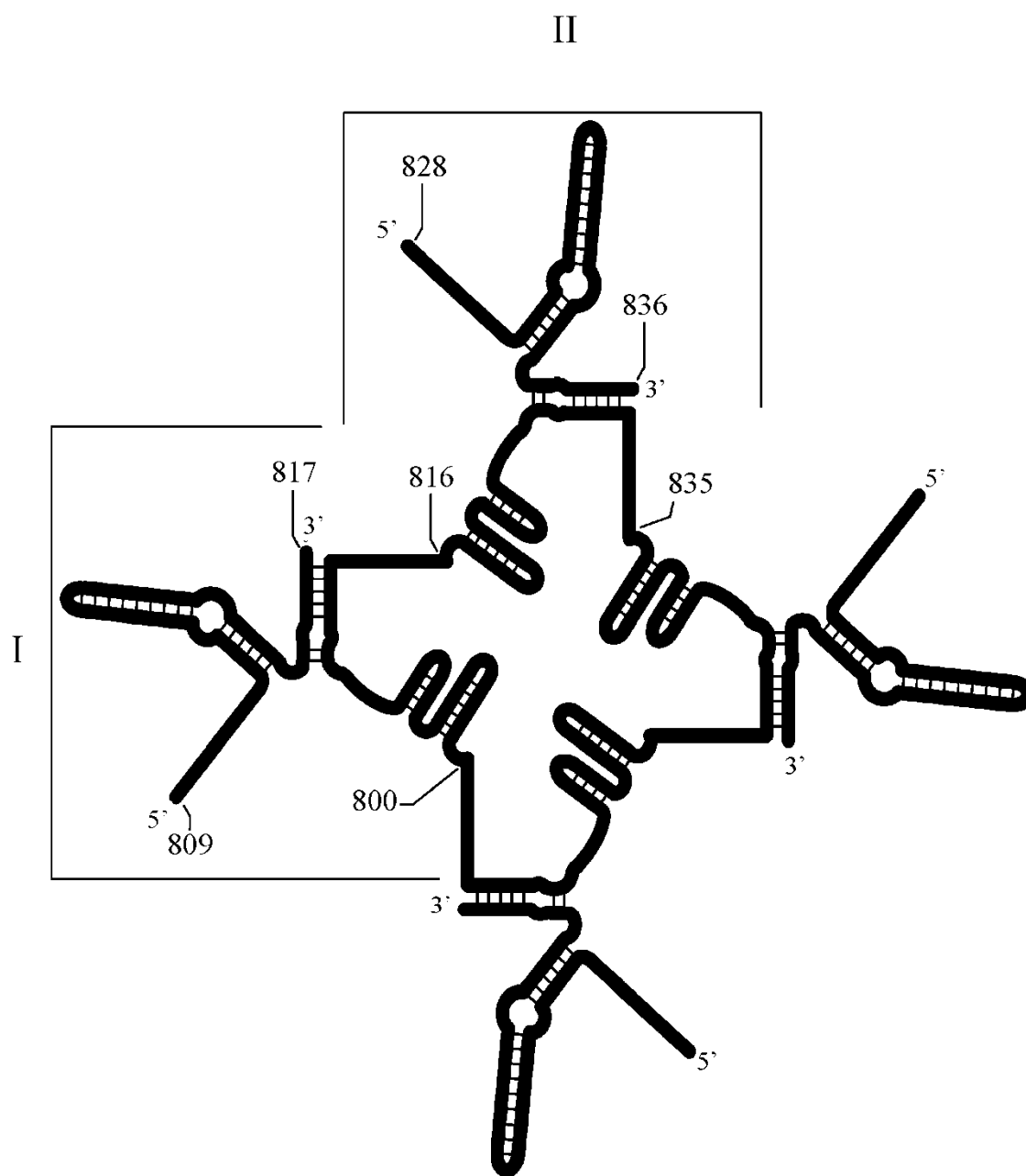


FIG. 8M

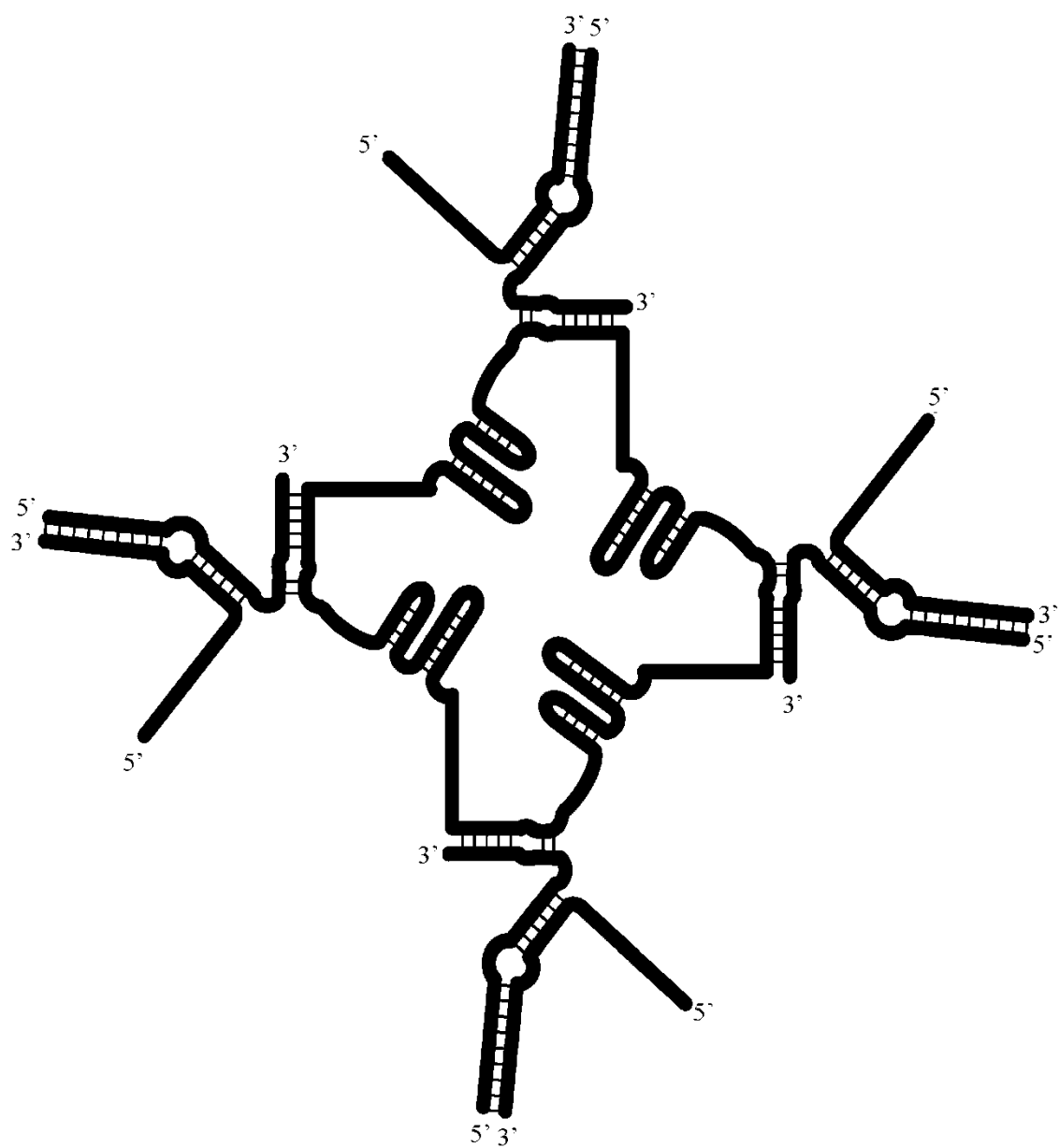


FIG. 8N

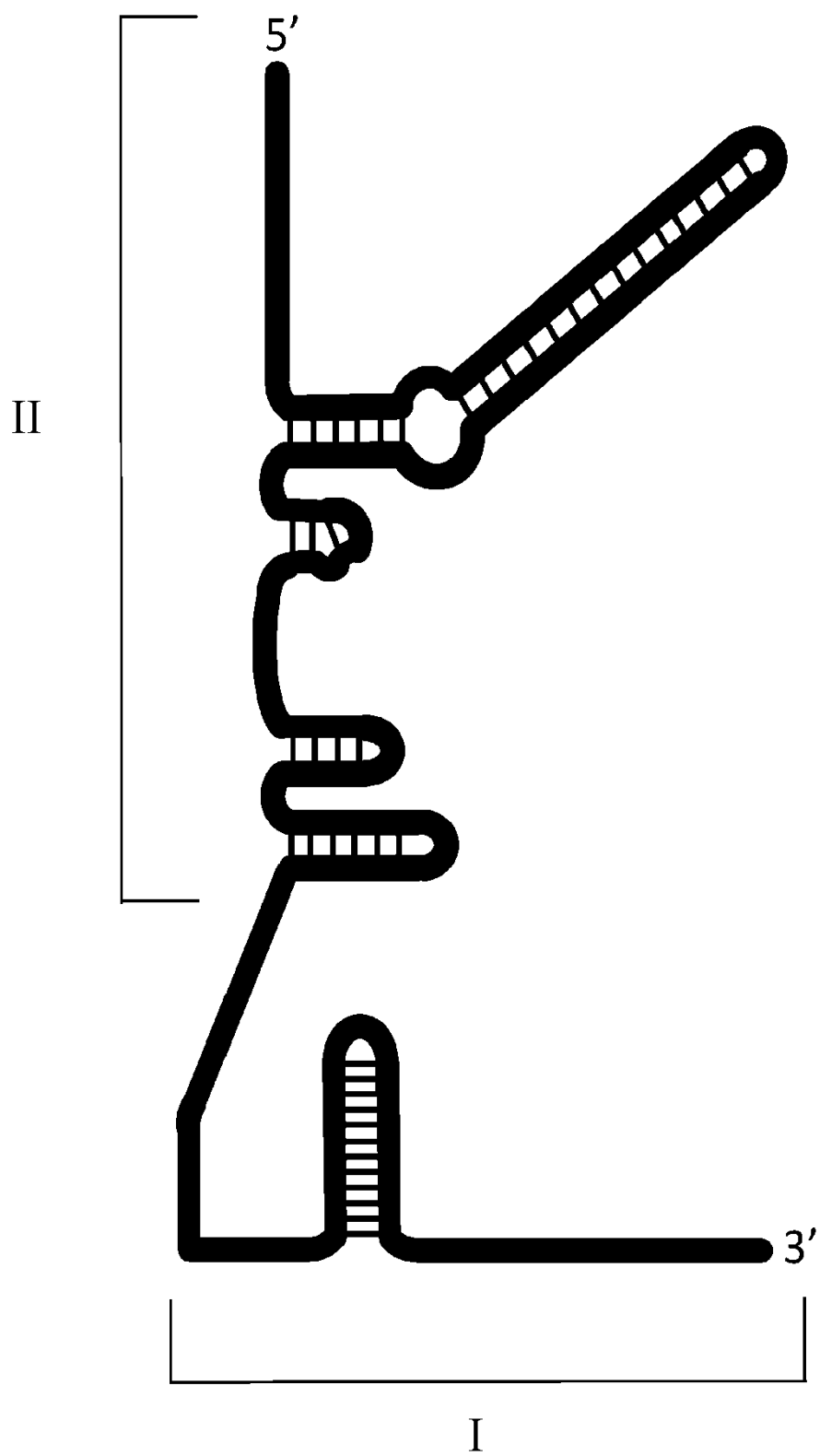


FIG. 9A

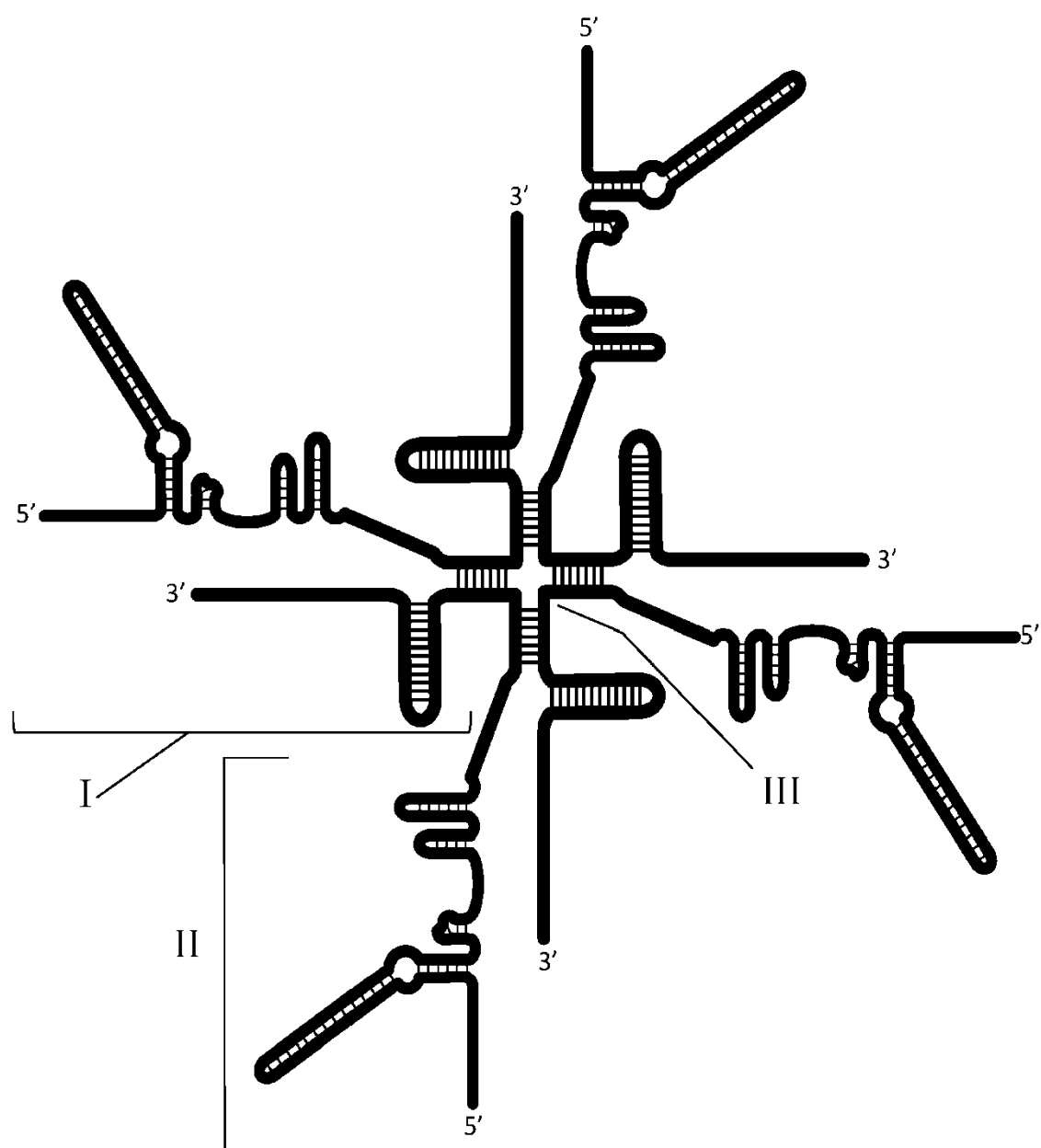


FIG. 9B

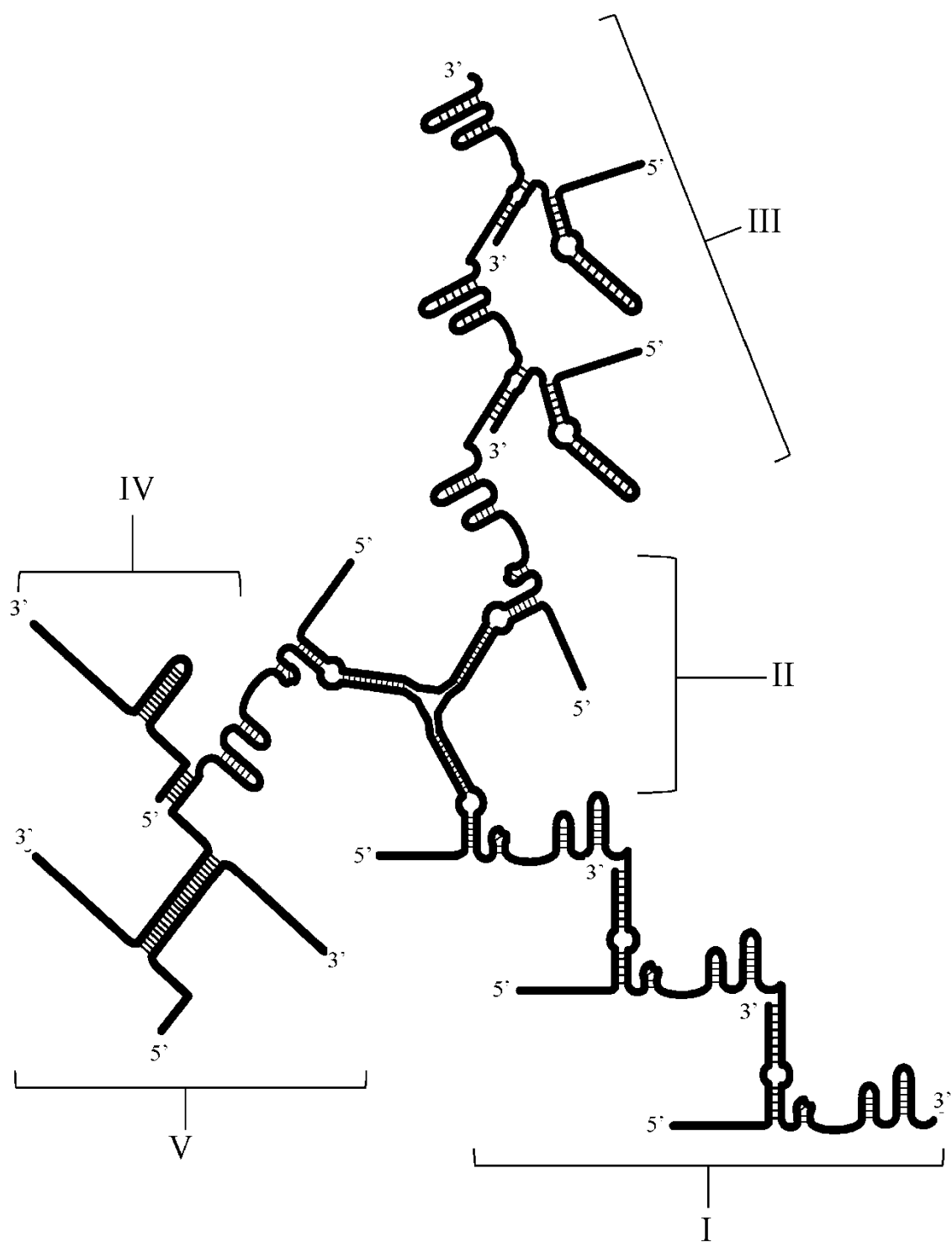


FIG. 10

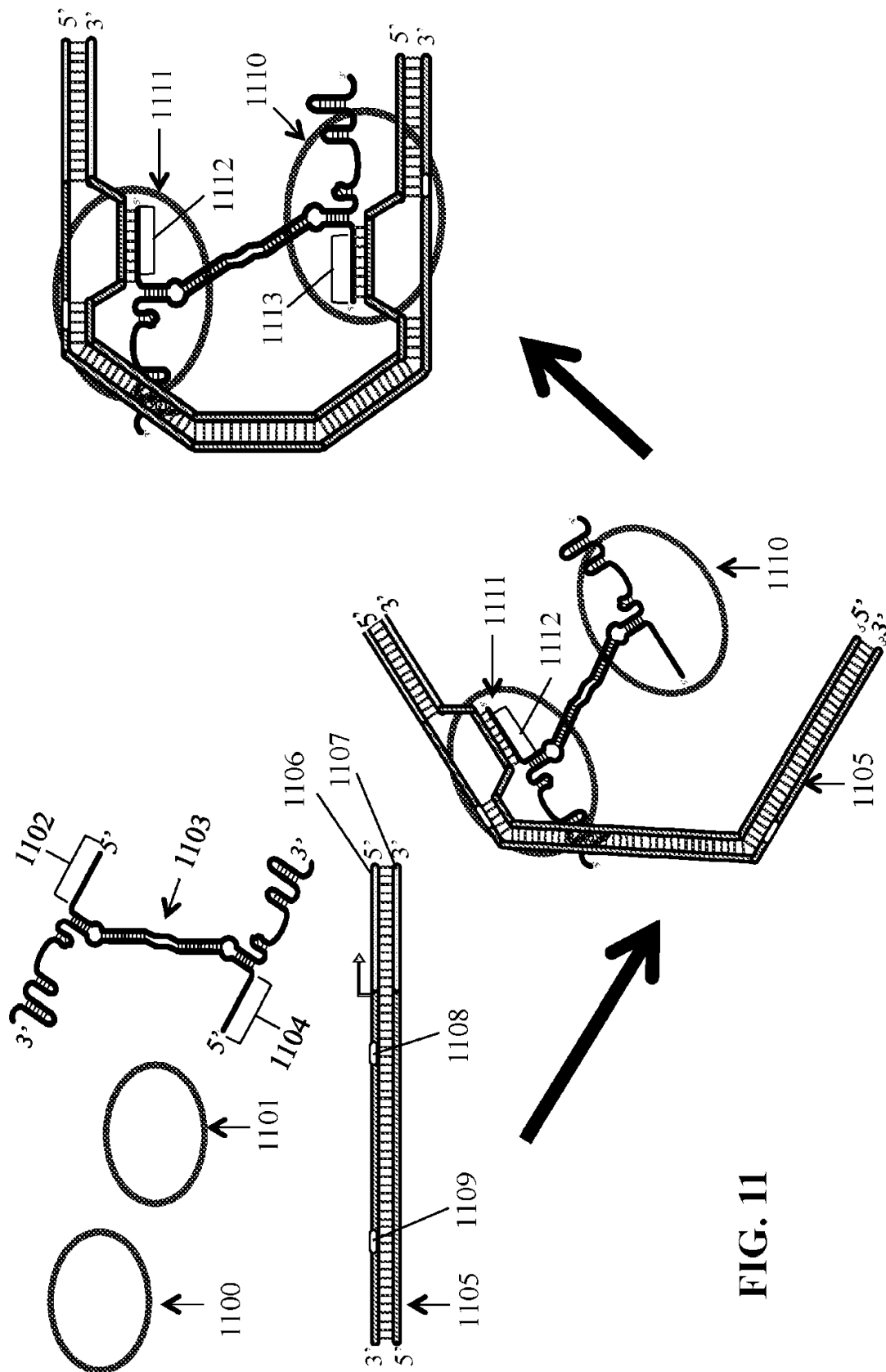


FIG. 11

FIG. 12

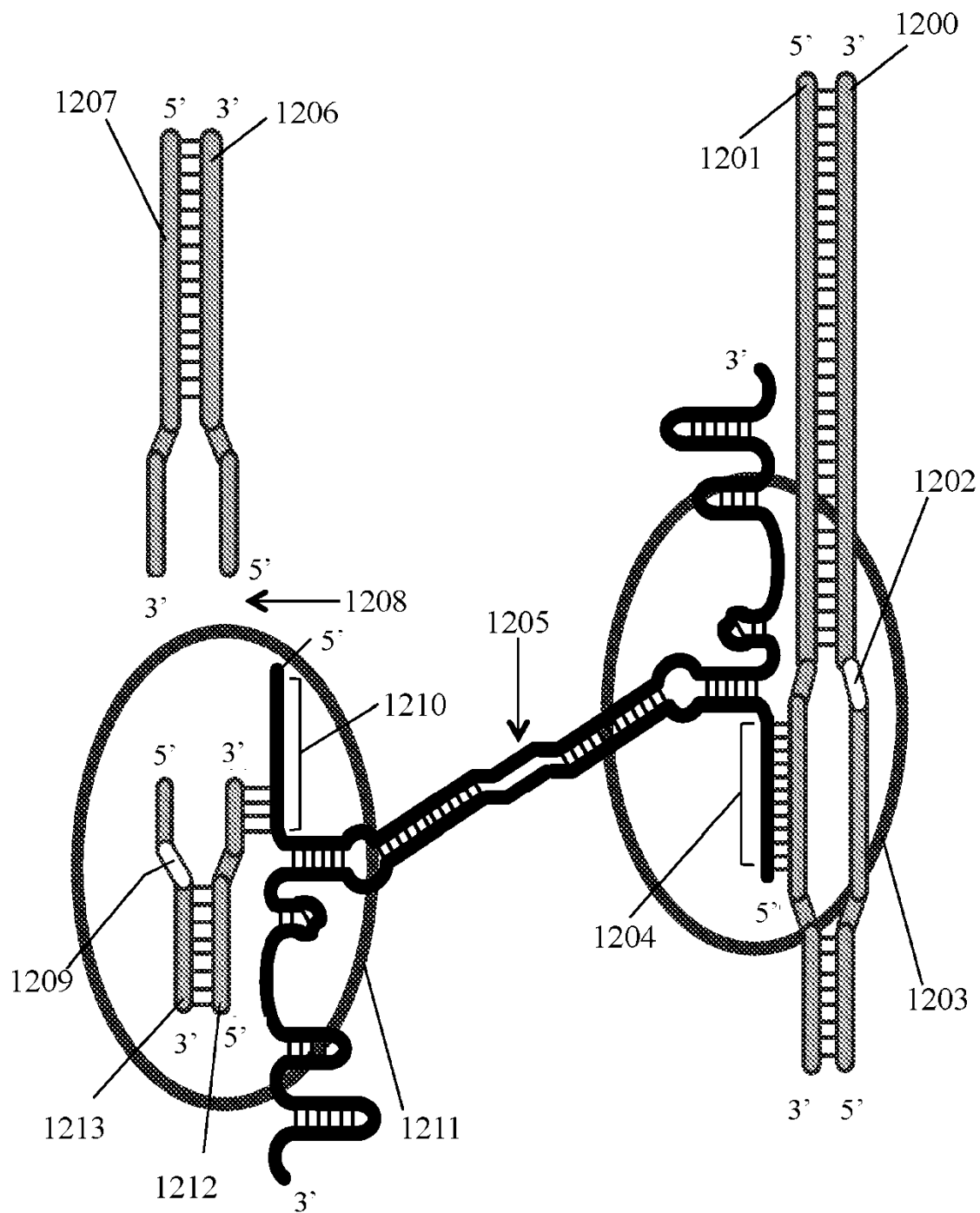


FIG. 13

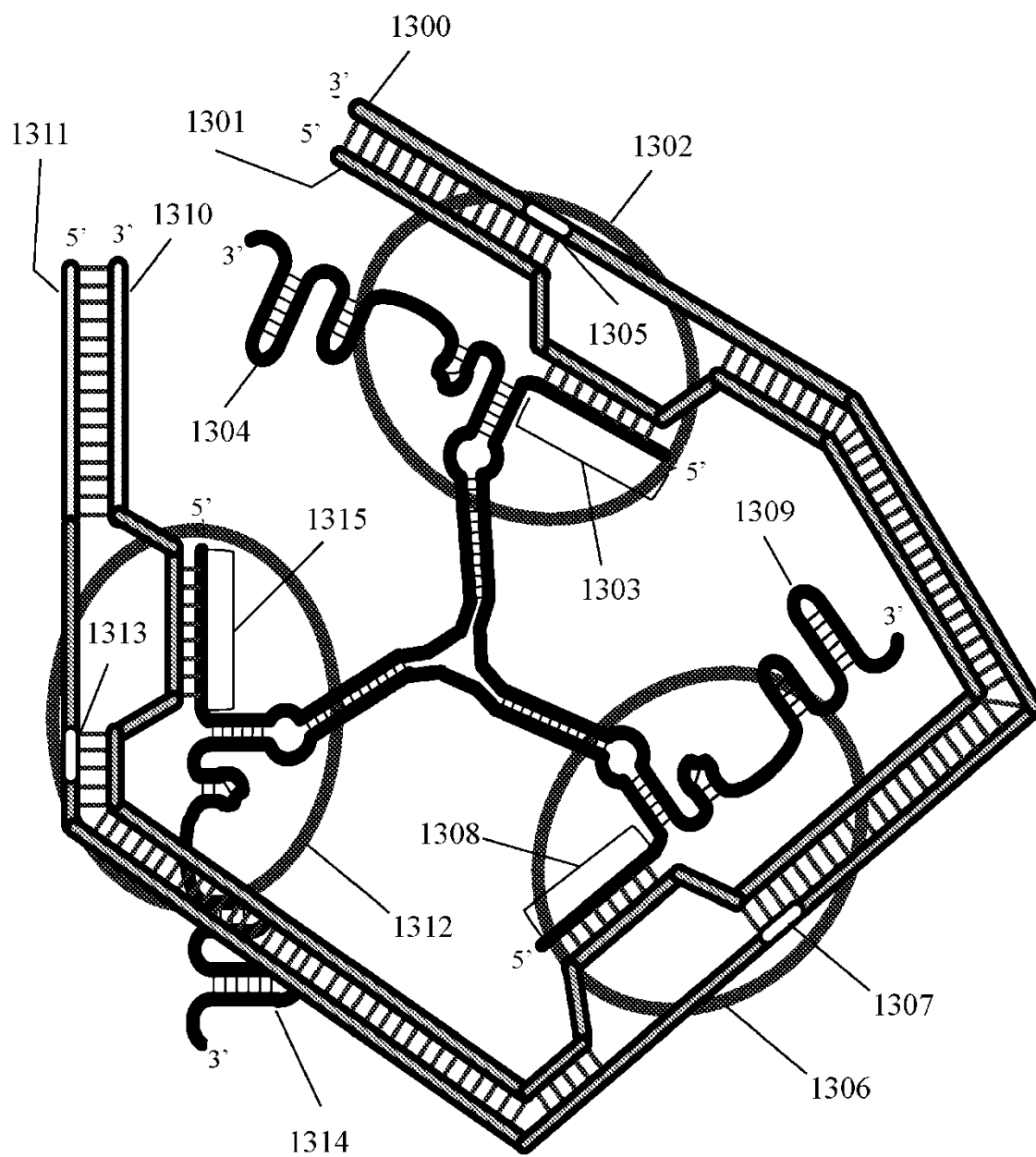


FIG. 14

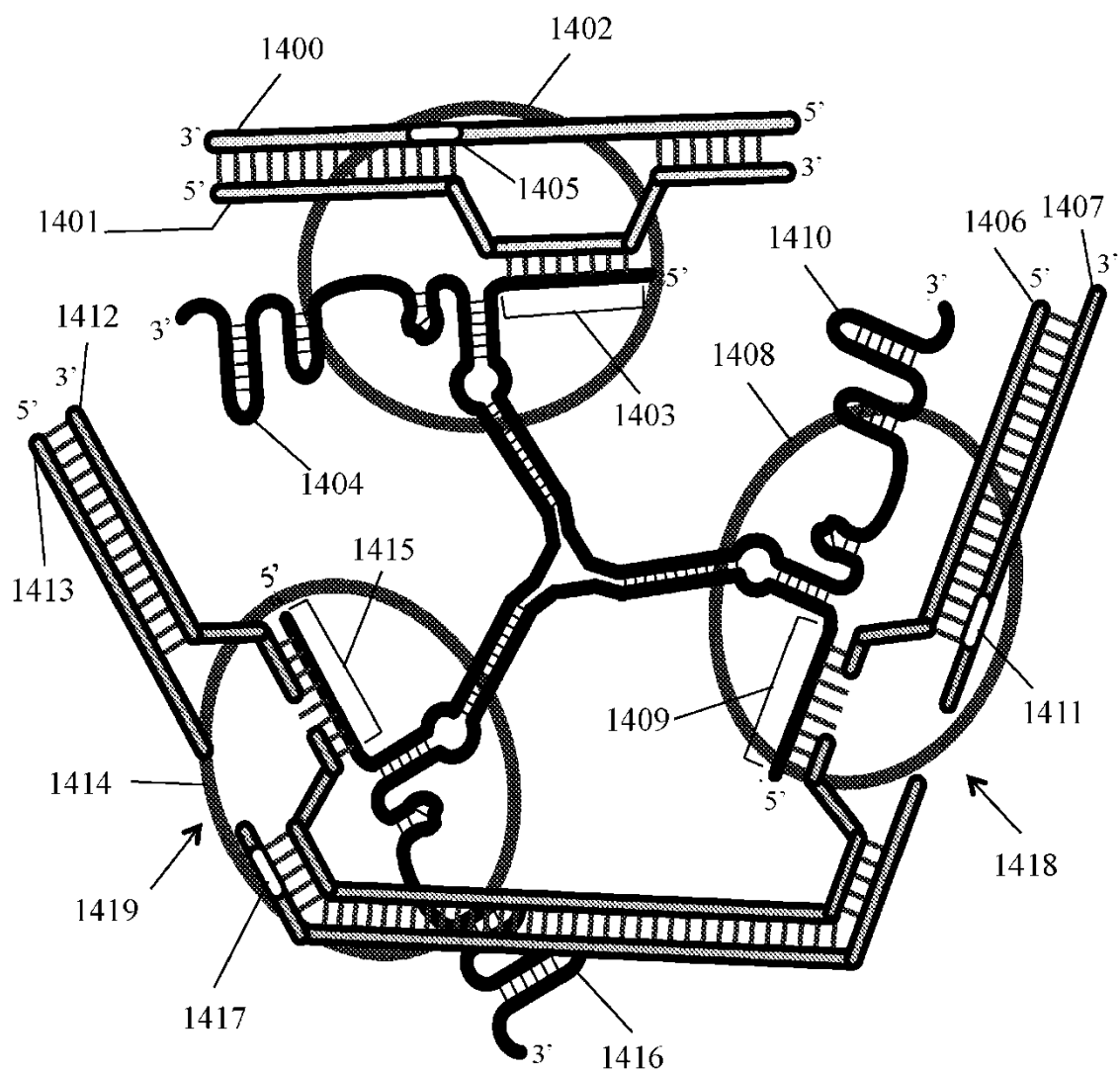
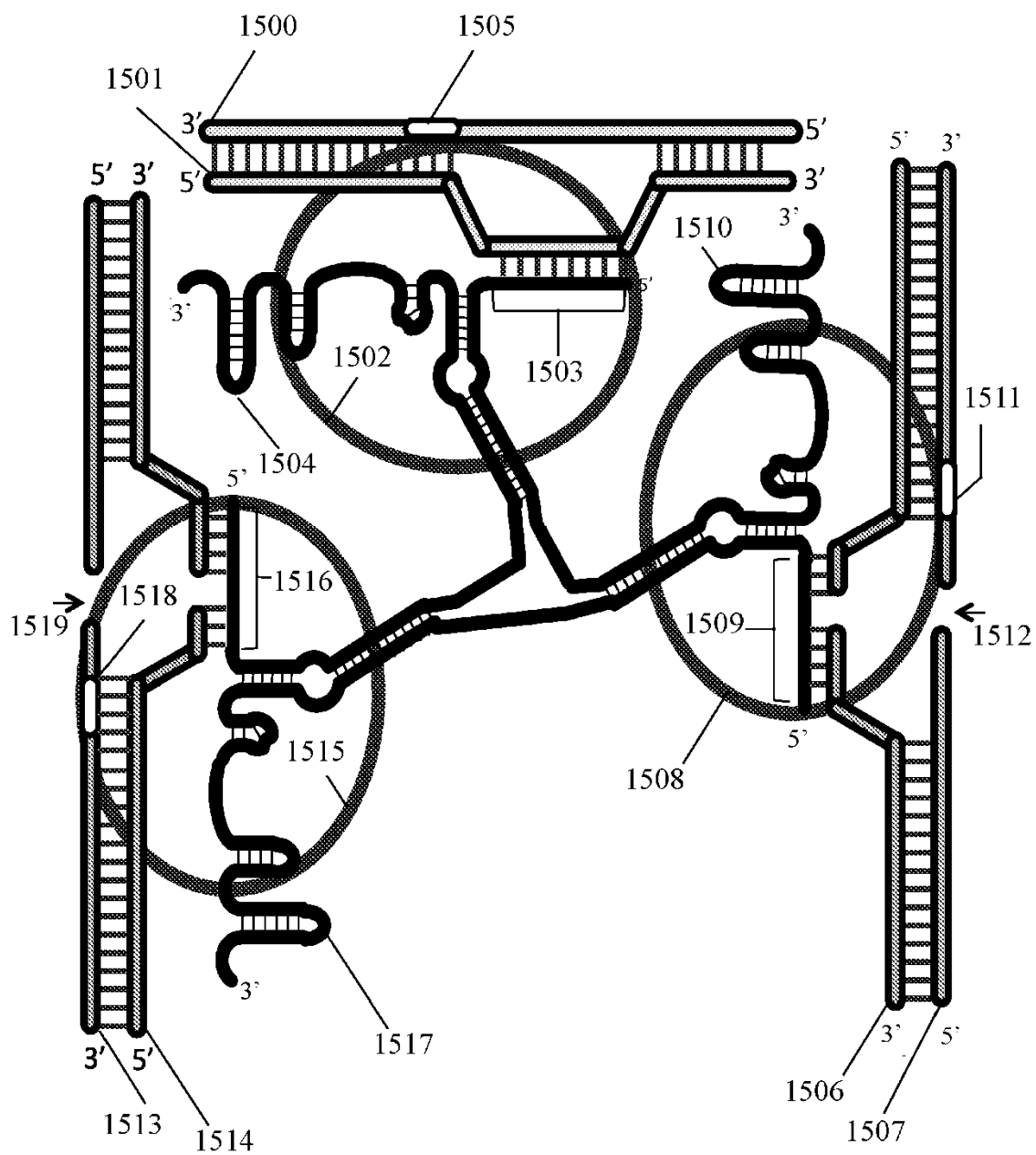


FIG. 15



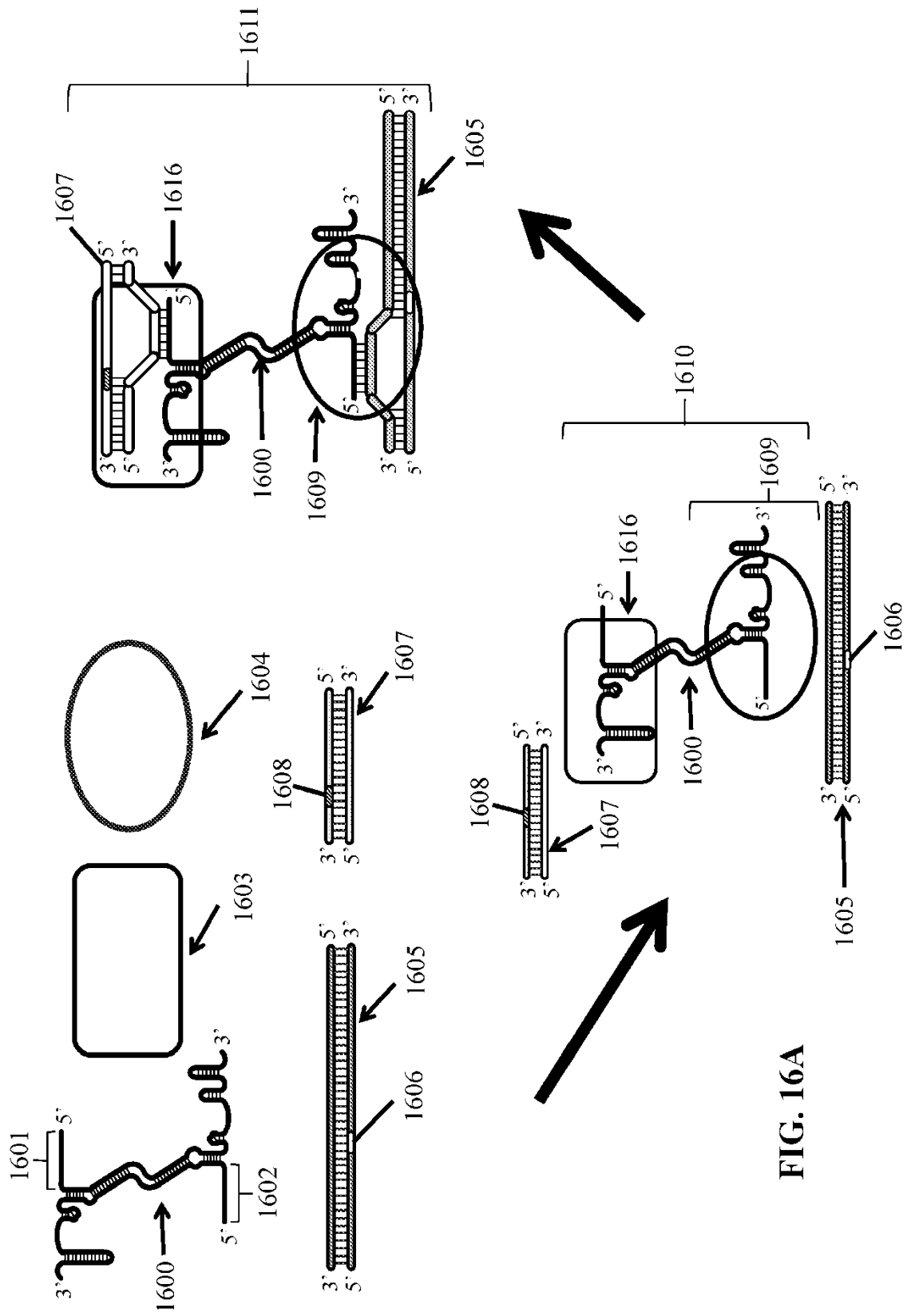


FIG. 16A

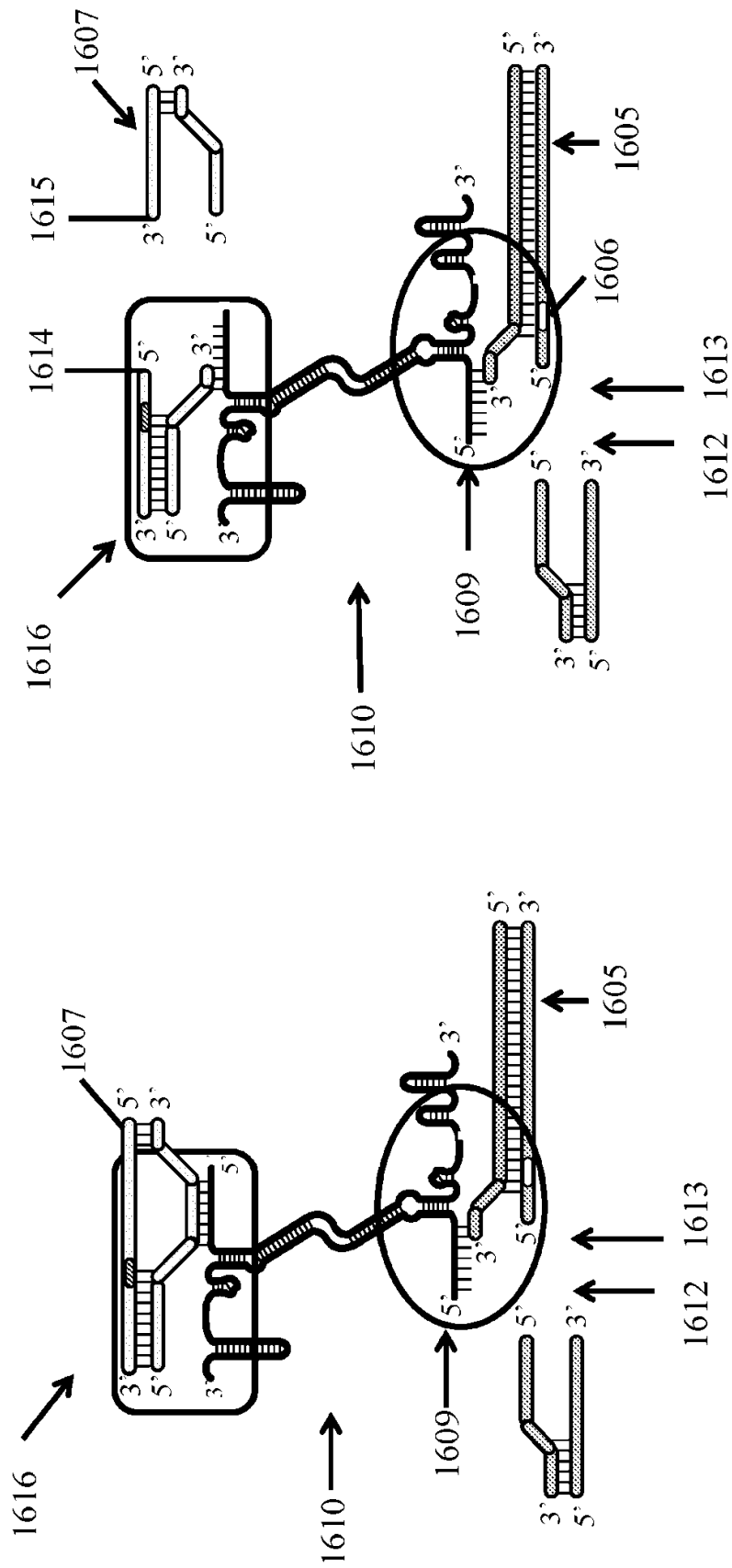


FIG. 16C

FIG. 16B