

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 773 052**

51 Int. Cl.:

C12P 21/06 (2006.01)

C07H 21/04 (2006.01)

A61K 38/04 (2006.01)

C12N 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.05.2011 PCT/US2011/000758**

87 Fecha y número de publicación internacional: **10.11.2011 WO11139349**

96 Fecha de presentación y número de la solicitud europea: **02.05.2011 E 11777690 (6)**

97 Fecha y número de publicación de la concesión europea: **15.01.2020 EP 2566972**

54 Título: **Composiciones para unir módulos de dedos de cinc**

30 Prioridad:

03.05.2010 US 343729 P

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

09.07.2020

73 Titular/es:

**SANGAMO THERAPEUTICS, INC. (100.0%)
Point Richmond Tech Center, 501 Canal Blvd.,
Suite A100
Richmond, CA 94804, US**

72 Inventor/es:

**PASCHON, DAVID y
REBAR, EDWARD J.**

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 773 052 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

Composiciones para unir módulos de dedos de cinc

5 **Campo técnico**

La presente divulgación pertenece a los campos del genoma y de la ingeniería de proteínas.

10 **Antecedentes**

10 Las proteínas de dedos de cinc con regiones de reconocimiento que se modifican mediante ingeniería genérica para su unión a sitios diana seleccionados se unen regularmente a otras proteínas de dedos de cinc, así como a dominios reguladores, y se usan para modificar la expresión de genes y sitios diana genómicos. Por ejemplo, se han utilizado nucleasas artificiales que comprenden dominios de unión a ADN unidos operativamente a dominios de escisión para la alteración dirigida de secuencias genómicas, incluyendo, la inserción de secuencias exógenas, la inactivación de 15 uno o más genes endógenos, la creación de organismos (por ejemplo, animales o cultivos) y estirpes celulares con patrones de expresión génica alterados y similares. Véanse, por ejemplo, las Publicaciones de Patente de los EE.UU. N.º 20050064474; 20060063231; 20070134796; 20080015164 y la Publicación internacional N.º 2007/139982.

20 Los módulos de proteínas de dedos de cinc (por ejemplo, dedos de cinc modificados mediante ingeniería genética de uno o más dedos) se unen normalmente entre sí usando secuencias de enlazador "canónicas" de 5 aminoácidos, tales como TGEKP (SEQ ID NO: 1) o enlazadores flexibles más largos. Véanse, las Patentes de los EE.UU. N.º 6.479.626; 6.903.185; 7.153.949 y la Publicación de Patente de los EE.UU. N.º 20030119023. Sin embargo, los 25 módulos de proteínas de dedos de cinc unidos a través de estos enlazadores canónicos se unen más eficazmente solo cuando no hay ningún hueco entre los subsitios diana del módulo unido en la molécula de ácido nucleico diana. Asimismo, los enlazadores flexibles y largos descritos anteriormente, diseñados para permitir que los módulos unidos se unan a sitios diana con huecos de 1, 2 o 3 pares de bases no distinguen entre estos huecos de pares de bases diferentes en términos de unión. Véanse, las Patentes de los EE.UU. N.º 6.479.626; 6.903.185; 7.153.949 y la 30 Publicación de Patente de los EE.UU. N.º 20030119023. Por tanto, sigue existiendo la necesidad de métodos y composiciones para unir módulos de dedos de cinc entre sí que mejoren tanto la afinidad de las proteínas que abarcan un hueco entre módulos de 1, 2 o 3 pb y también mejoren la selectividad de estas proteínas para unir dianas que abarcan un hueco de una longitud deseada y no se unen de forma no selectiva a otras dianas sin el hueco de esa longitud deseada. Los enlazadores para módulos de dedos de cinc que distinguen entre huecos de 0, 35 1, 2, 3 o incluso más pares de bases entre subsitios del módulo adyacentes permitirían una mayor capacidad de diseño de cualquier proteína de fusión de dedos de cinc, incluyendo los factores de transcripción de dedos de cinc (ZFP-TF, por sus siglas en inglés) y las nucleasas de dedos de cinc (ZFN, por sus siglas en inglés).

40 **Sumario**

40 En el presente documento se desvelan enlazadores para su uso en la unión de módulos de unión de ADN (por ejemplo, módulos de dedos de cinc) entre sí. También se describen proteínas de fusión, por ejemplo, proteínas de dedos de cinc que comprenden estos enlazadores que a su vez se fusionan con dominios reguladores tales como dominios reguladores de la transcripción o a nucleasas. La divulgación también proporciona métodos de uso de 45 estas proteínas de fusión y composiciones de las mismas para la modulación de la expresión génica, la escisión dirigida de ADN celular (por ejemplo, cromatina celular endógena) en una región de interés y/o la recombinación homóloga en una región predeterminada de interés en las células.

50 Por tanto, en un aspecto, en el presente documento se describen enlazadores que comprenden 5 o más aminoácidos entre el último resto del dedo amino (N)-terminal (normalmente el resto coordinador de cinc carboxi (C)-terminal) y el primer resto del dedo C-terminal (normalmente el primer resto aromático conservado en (N-terminal)), por ejemplo, 7-17 aminoácidos. En determinadas realizaciones, el enlazador comprende un resto N-terminal, un resto C-terminal y restos internos con respecto a los restos terminales, y adicionalmente en donde el resto N-terminal o los restos internos comprenden al menos un resto de prolina, por ejemplo, un enlazador que comprende la 55 secuencia de aminoácidos $X^{N-term}-X_n-X^{C-term}$, en donde X es cualquier resto de aminoácido, X_n comprende al menos 3 restos de aminoácidos y al menos uno de entre X^{N-term} y X_n comprende un resto de prolina. En determinadas realizaciones, el enlazador comprende al menos dos restos de prolina (por ejemplo, 2, 3, 4 o más). En otras realizaciones, donde el enlazador comprende al menos un resto de prolina y al menos un resto básico (por ejemplo, Arg, His o Lys). En otras realizaciones, donde el enlazador comprende al menos dos restos básicos (por ejemplo, Arg, His o Lys). En determinadas realizaciones, el enlazador es uno que se muestra en cualquiera de las Tablas 4, 5, 60 6, 9, 10, 11 o 13.

65 En otro aspecto, se proporcionan polipéptidos de fusión que comprenden un enlazador como se describe en el presente documento. Son materia objeto de la invención las proteínas de fusión que son proteínas de dedos de cinc de múltiples dedos con enlazadores como se reivindican.

En otro aspecto, se proporcionan polinucleótidos que codifican cualquiera de los enlazadores o proteínas de fusión como se reivindican.

5 En otro aspecto más, también se proporcionan células que comprenden cualquiera de los polipéptidos (por ejemplo, polipéptidos de fusión) y/o polinucleótidos como se reivindican.

También se describen organismos (por ejemplo, mamíferos, hongos y plantas) que comprenden los polipéptidos (por ejemplo, polipéptidos de fusión) y/o polinucleótidos como se describen en el presente documento.

10 Una proteína de fusión puede expresarse en una célula, por ejemplo, entregando la proteína de fusión a la célula o entregando un polinucleótido que codifica la proteína de fusión a una célula. Si el polinucleótido es ADN, entonces se transcribe y se traduce para generar la proteína de fusión. Si se entrega como una molécula de ARN, entonces se traduce inmediatamente, generando de este modo la proteína de fusión. Se presentan métodos para la entrega de polinucleótidos y polipéptidos a las células en otra parte en la presente divulgación.

15 Estos y otros aspectos serán fácilmente evidentes para el experto en la materia a la luz de la divulgación en su totalidad.

20 Breve descripción de los dibujos

La **Figura 1**, los **paneles A y B**, muestran secuencias de amino de enlazadores y proteínas de dedos de cinc de ejemplo. La **Figura 1A** muestra la secuencia de aminoácidos de cada ZFP hospedadora (F1-F4 de ZFP 8196 que se muestra en la SEQ ID NO: 130; la SEQ ID NO: 131; la SEQ ID NO: 132 y la SEQ ID NO: 133; de F1 a F4 de ZFP 7263 se muestra en la SEQ ID NO: 134; la SEQ ID NO: 135; la SEQ ID NO: 136 y la SEQ ID NO: 137; de F1 a F4 de ZFP 7264 se muestra en la SEQ ID NO: 138; la SEQ ID NO: 139; la SEQ ID NO: 140 y la SEQ ID NO: 141) utilizada para estos estudios. Los aminoácidos se designan mediante un código de una sola letra. Cada secuencia se enumera en la dirección amino terminal → carboxi terminal, de manera que el extremo amino de cada proteína es la primera metionina del dedo 1 y el extremo carboxi es la serina final del dedo 4. "F1", "F2", "F3" y "F4" designan los dedos primero, segundo, tercero y cuarto, respectivamente, de cada proteína. El subrayado indica restos de aminoácidos en las confluencias de los dedos que convencionalmente se considera que son una secuencia de enlazador. Las hélices de reconocimiento están en un recuadro. La **Figura 1B** muestra diseños de biblioteca de enlazadores en los que cada biblioteca se generó reemplazando codones para dos o tres restos en el enlazador central con una mezcla de dos a doce codones completamente aleatorizados. Los codones de la biblioteca se indican mediante (NNS)₂₋₁₂.

35 La **Figura 2**, los **paneles A a D**, son gráficos que representan la selectividad de hueco de grupos de fagos con los enlazadores y proteínas de dedos de cinc indicados. La Figura 2A muestra el grupo de fagos seleccionado entre la biblioteca de ZFP8196 usando una diana con una base insertada de 1 pb (ATAAACTGdCAAAAGGC (SEQ ID NO: 33) (Tabla 2A)) que se sometió a ensayo para determinar la unión a cada diana de ZFP8196 en la Tabla 2C. La Figura 2B muestra el grupo de fagos seleccionado entre la biblioteca de ZFP7263 usando una diana con una base insertada de 1 pb (CCACTCTGhTGGAAAGTG (SEQ ID NO: 43) (Tabla 2A)) que se sometió a ensayo para determinar la unión a cada diana de ZFP7263 en la Tabla 2C. La Figura 2C muestra el grupo de fagos seleccionado entre la biblioteca de ZFP7264 usando una diana con una base insertada de 1 pb (TTAAAGCGhGCTCCGAA (SEQ ID NO: 38) (Tabla 2A)) sometida a ensayo para determinar la unión a cada diana de ZFP7264 en la Tabla 2C. La Figura 2D muestra el grupo de fagos seleccionado entre la biblioteca de ZFP8196 usando una diana con una base insertada de 2 pb (ATAAACTGdbCAAAAGGC (SEQ ID NO: 34) (Tabla 2A)) sometida a ensayo para determinar la unión a cada diana de ZFP8196 en la Tabla 2C. Cada ensayo también incluyó dos dianas de control para las otras dos ZFP hospedadoras para descartar la unión no específica al ADN, así como una muestra de control negativo que no incluía un sitio diana. Se indica el % de fagos que se unió exitosamente a cada diana. Cada grupo de fagos era de la quinta ronda de selección. La eficiencia de retención se determinó esencialmente como se describió anteriormente (Rebar, et al. *Methods in Enzymology*, 1996 (267): 129-149).

50 La **Figura 3**, los **paneles A y B**, muestran enlazadores seleccionados para sitios diana que contienen el hueco indicado. La Figura 3A muestra secuencias de enlazador seleccionadas para omitir un hueco de 1 pb en el contexto de ZFP8196, ZFP7263 y ZFP7264 (SEQ ID NO: 142 a 166). La Figura 3B muestra secuencias de enlazador (SEQ ID NO: 167 a 174) para omitir un hueco de 2 pb en el contexto de ZFP8196. Los enlazadores seleccionados están enriquecidos con prolina y arginina (sombreados). Las preferencias de longitud también son evidentes y dependen del número de bases omitidas.

60 La **Figura 4**, los **paneles A a E**, son gráficos que representan la selectividad de hueco para los enlazadores seleccionados para omitir 1 par de bases en la proteína de dedos de cinc designada ZFP8196. En cada panel, las puntuaciones de ELISA se normalizaron con el enlazador parental, no de omisión, en su sitio diana sin huecos. La "secuencia de hueco" se refiere a la identidad de la base o bases entre los subsitios del módulo donde (-) indica la diana sin huecos. Las Figuras 4A-4C representan resultados de tres de los enlazadores de omisión de hueco de 1 pb, (enlazadores denominados 1f (SEQ ID NO: 54), 1d (SEQ ID NO: 56) y 1c (SEQ ID NO: 55)). La Figura 4D muestra resultados con un enlazador flexible convencional que anteriormente se ha demostrado que permite la

modificación de un locus endógeno en células humanas (TGGGGSQKP, SEQ ID NO: 2) (Véase Hockemeyer *et al.* (2009) *Nature Biotechnology* 27: 851-857) y la Figura 4E representa los resultados para un enlazador flexible publicado anteriormente (LRQKDERP, SEQ ID NO: 3) (Véase Kim JS y Pabo CO (1998) *Proc Natl Acad Sci USA* 95 (6): 2812-2817). Los enlazadores seleccionados 1c, 1d y 1f (Figuras 4A-4C) muestran preferencias claras para los cuatro sitios diana con un hueco de un solo par de bases, mientras que los enlazadores de control en las Figuras 4D y 4E muestran una unión global menos eficaz y poca selectividad de hueco.

La **Figura 5**, los **paneles A a D**, son gráficos que representan la selectividad de hueco para enlazadores seleccionados para omitir 1 par de bases en ZFP7264. Las Figuras 5A a 5C representan los resultados de un ensayo ELISA del enlazador 1e (SEQ ID NO: 12) en el fondo ZFP7264. La Figura 5A muestra los resultados para el enlazador 1e, seleccionado para omitir un hueco de 1 pb entre los subsitios del módulo. La Figura 5B muestra los resultados para un enlazador flexible convencional (TGGGGSQKP, SEQ ID NO: 2) y la Figura 5C muestra los resultados para un enlazador flexible LRQKDERP (SEQ ID NO: 3). Las puntuaciones de ELISA se normalizan con la ZFP7264 parental, que de omisión, en su diana sin huecos. La "secuencia de hueco" es la identidad de la base o bases omitidas entre los subsitios del módulo donde (-) indica la diana sin huecos. La Figura 5D muestra una versión ampliada de los datos de la Figura 5B donde el intervalo de puntuaciones de ELISA es 0-0,6 en comparación con 0-5 en los otros paneles.

La **Figura 6**, los **paneles A a F**, son gráficos que representan la selectividad de hueco para los enlazadores seleccionados para omitir 2 pb en ZFP8196. Las Figuras 6A a 6E muestran los resultados de un ELISA que somete a ensayo los enlazadores seleccionados para omitir un hueco de 2 pb entre los subsitios del módulo en el fondo ZFP8196. Las figuras 6A a 6C muestran los resultados para los enlazadores seleccionados 2f (SEQ ID NO: 69), 2d (SEQ ID NO: 70) y 2e (SEQ ID NO: 71), mientras que la Figura 6D muestra los resultados para un enlazador flexible anteriormente publicado (LRQKDGGSERP (SEQ ID NO: 68)) y la Figura 6E muestra los resultados para un enlazador flexible convencional (TGGGSGGSQKP (SEQ ID NO: 14)). La Figura 6F muestra una versión ampliada de los datos que se muestran en la Figura 6E, donde el intervalo de puntuaciones de ELISA es de 0-0,1 en comparación con 0-1 en los otros paneles. La "secuencia de hueco" es la identidad de la base o bases entre los subsitios del módulo donde (-) indica la diana sin huecos. Los enlazadores seleccionados (Figuras 6A-6C) muestran una clara preferencia por un hueco de 2 pb en comparación con un hueco de 1 pb o sin hueco, mientras que los enlazadores de control en las Figuras 6D y 6E muestran una unión global menos eficaz y poca selectividad de hueco.

La **Figura 7**, los **paneles A y B**, representan un resumen de los datos de ELISA de un estudio diseñado para analizar la portabilidad de los enlazadores de omisión de 1 pb a diferentes fondos de ZFP. Se sometieron a ensayo doce ZFP diferentes (indicadas como ZFP1, ZFP2, etc.). La Figura 7A muestra puntuaciones de ELISA normalizadas a ZFP de control positivo convencionales que se ha demostrado que modifican eficientemente un locus de IL2R γ endógeno cuando se usan como ZFN (Urnov *et al.* (2005) *Nature* 435 (7042): 646-651). La Figura 7B muestra todas las puntuaciones normalizadas adicionalmente para cada ZFP parental que lleva el enlazador flexible convencional TGGGGSQKP (SEQ ID NO: 2). Los valores subrayados en la Figura 7B indican una mejora de >4 veces en la puntuación de ELISA para las ZFP con los enlazadores seleccionados (1e (SEQ ID NO: 12), 1f (SEQ ID NO: 54), 1d (SEQ ID NO: 56) y 1c (SEQ ID NO: 55)) en comparación con la misma ZFP hospedadora con el enlazador flexible TGGGGSQKP (SEQ ID NO: 2). En general, los enlazadores 1e, 1f, 1d y 1c conducen a un aumento general en la puntuación de ELISA de 3-5 veces sobre el enlazador flexible.

La **Figura 8**, los **paneles A y B**, representan un resumen de los datos de ELISA de un estudio diseñado para analizar la portabilidad de los enlazadores de omisión de 2 pb a diferentes fondos de ZFP. Se sometieron a ensayo seis ZFP diferentes (indicadas como ZFP13, ZFP14, etc.). La Figura 8A muestra puntuaciones de ELISA normalizadas a ZFP de control positivo convencionales que se ha demostrado que modifican eficientemente un locus de IL2R γ endógeno cuando se usan como ZFN (Urnov *et al.* (2005) *Nature* 435 (7042): 646-651). La Figura 8B muestra todas las puntuaciones normalizadas adicionalmente para cada ZFP parental que lleva el enlazador flexible convencional TGGGSGGSQKP (SEQ ID NO: 14). Los valores subrayados en la Figura 8B indican una mejora de >2 veces en la puntuación de ELISA para las ZFP con los enlazadores seleccionados (2f (SEQ ID NO: 69), 2d (SEQ ID NO: 70) y 2e (SEQ ID NO: 71)) en comparación con la misma ZFP hospedadora con el enlazador flexible TGGGSGGSQKP (SEQ ID NO: 14). En general, los enlazadores 2f (SEQ ID NO: 69), 2d (SEQ ID NO: 70) y 2e (SEQ ID NO: 71) condujeron a un aumento general en la puntuación de ELISA de 1,9-2,4 veces sobre el enlazador flexible.

La **Figura 9**, los **paneles A y B**, muestran resultados de estudios de modificación génica endógena, determinados mediante ensayos CEL-I, con ZFN que contienen enlazadores seleccionados. Las Figuras 9A y 9B representan geles de ejemplo utilizados para determinar la actividad nucleasa de ZFN en loci endógenos mediante el ensayo CEL-I (que mide la actividad de unión de extremos no homólogos (NHEJ, por sus siglas en inglés), Surveyor™, Transkaryotic) para determinar si enlazadores como se describen en el presente documento pueden usarse en el contexto de diferentes ZFN. El gel que se muestra en la Figura 9A representa los resultados de los enlazadores 1e (SEQ ID NO: 12), 1f (SEQ ID NO: 54), 1d (SEQ ID NO: 56) y 1c (SEQ ID NO: 55) en los fondos ZFN3 y ZFN4. El gel que se muestra en la Figura 9B representa los resultados de los enlazadores 2f, 2d y 2e en el fondo ZFN14. El porcentaje de modificación génica mediante NHEJ, "Mod. Gén. (%)", se indica en la parte inferior de las calles. El

control negativo, "neg", es una muestra transfectada con un plásmido que lleva GFP. Los resultados de las ZFN que usan un enlazador flexible convencional (TGGGGSQKP (SEQ ID NO: 2) para la Figura 9A y TGGGSGGSQKP (SEQ ID NO: 14) para la Figura 9B) se muestran en las calles etiquetadas "C". Las calles sin etiquetar contienen muestras de ZFN que llevan otros enlazadores que no se desarrollaron adicionalmente en estos estudios. Los datos en los geles demuestran que los enlazadores como se describen en el presente documento aumentan significativamente los niveles de modificación génica en comparación con los enlazadores flexibles.

La **Figura 10**, los **paneles A y B**, representan un resumen de los estudios de modificación génica para las ZFN como se han descrito anteriormente para la Figura 9 que contienen los enlazadores indicados seleccionados para omitir 1 pb. La Figura 10A es la cuantificación del porcentaje de modificación génica para cada ZFN con el conjunto de cinco enlazadores sometidos a ensayo (flexible, 1e (SEQ ID NO: 12), 1f (SEQ ID NO: 54), 1d (SEQ ID NO: 56) y 1c (SEQ ID NO: 55)). La Figura 10B muestra estos mismos datos normalizados al enlazador flexible (TGGGGSQKP, SEQ ID NO: 2) y también muestra el aumento promedio en la modificación génica en todos los pares de ZFN activas. Las muestras producidas usando condiciones de alta expresión (véase el Ejemplo 3) están resaltadas en color gris. Las ZFN que llevan enlazadores de ejemplo que mejoraron el nivel de modificación génica en >2 veces están subrayadas en la Figura 10B. En general, las ZFN que llevan enlazadores 1e (SEQ ID NO: 12), 1f (SEQ ID NO: 54), 1d (SEQ ID NO: 56) y 1c (SEQ ID NO: 55) conducen a un aumento promedio en la modificación génica de 1,8 a 2,8 veces sobre sus respectivas ZFN hospedadoras que llevan el enlazador flexible.

La **Figura 11**, los **paneles A y B**, representan un resumen de estudios de modificación génica como se describen para la Figura 9, para las ZFN que contienen los enlazadores indicados seleccionados para omitir un hueco de 2 pb entre los subsitios del módulo de las 6 ZFN hospedadoras. La Figura 11A es la cuantificación del porcentaje de modificación génica para cada ZFN con el conjunto de cuatro enlazadores sometidos a ensayo (flexible, 2f (SEQ ID NO: 69), 2d (SEQ ID NO: 70) y 2e (SEQ ID NO: 71)). La Figura 11B muestra estos mismos datos normalizados con respecto al enlazador flexible (TGGGSGGSQKP, SEQ ID NO: 14) y también muestra el aumento promedio en todos los pares de ZFN activas. Las muestras producidas usando condiciones de alta expresión (véase el Ejemplo 3) están resaltadas en color gris. Las ZFN que llevan enlazadores de ejemplo que mejoraron el nivel de modificación génica en >2 veces están subrayadas en la Figura 11B. Las ZFN que llevan enlazadores 2f (SEQ ID NO: 69), 2d (SEQ ID NO: 70) y 2e (SEQ ID NO: 71) condujeron a un aumento promedio en la modificación génica de 1,5-2,0 veces sobre sus respectivas ZFN hospedadoras que llevan el flexible enlazador.

La **Figura 12**, muestra la secuencia de aminoácidos de la ZFP8196 hospedadora utilizada para la selección secundaria para los enlazadores que abarcan un hueco de 2 pb. Los aminoácidos se designan mediante un código de una sola letra. La secuencia se enumera en la dirección amino terminal → carboxi terminal, de manera que el extremo amino de la proteína es la primera metionina del dedo 1 y el extremo carboxi es la serina final del dedo 4. "F1" (SEQ ID NO: 130), "F2" (SEQ ID NO: 131), "F3" (SEQ ID NO: 132) y "F4" (SEQ ID NO: 133) designan los dedos primero, segundo, tercero y cuarto, respectivamente, de la proteína. Las hélices de reconocimiento están en un recuadro. La biblioteca de enlazadores se generó reemplazando los codones para dos restos en el enlazador central con una mezcla de cinco a siete codones completamente aleatorizados, seguido de un codón aleatorizado para obtener restos de fenilalanina (F), lisina (L) o tirosina (Y), y los tres últimos codones se fijaron para que fueran arginina (R), prolina (P) y prolina (P). Los codones de la biblioteca se indican mediante (NNS)₅₋₇ y (F/I/Y).

La **Figura 13**, los **paneles A y B**, representan la selectividad de hueco del grupo de fagos de la selección secundaria para enlazadores que abarcan un hueco de 2 pb y las secuencias de aminoácidos resultantes de los clones obtenidos en la selección. La Figura 13A muestra el grupo de fagos seleccionado entre la biblioteca de ZFP8196 usando una diana con un hueco insertado de 2 pb (ATAAACTGdbCAAAAGGC (SEQ ID NO: 34) (Tabla 2A)) sometido a ensayo para la unión a cada diana de ZFP8196 en la Tabla 2C. Cada ensayo también incluyó una diana de control para otra ZFP hospedadora para descartar la unión no específica a ADN, así como una muestra de control negativo que no incluía un sitio diana. Se indica el % de fagos que se unió exitosamente a cada diana. El grupo de fagos era de la sexta ronda de selección. La eficiencia de retención se determinó esencialmente como se describió anteriormente (Rebar, *et al. Methods in Enzymology*, 1996 (267): 129-149). La Figura 13B muestra secuencias de aminoácidos (SEQ ID NO: 175 a 210) de enlazadores seleccionados para omitir un hueco de 2 pb de la selección secundaria en el contexto de ZFP8196. Los enlazadores seleccionados están enriquecidos con prolina y arginina (sombreados).

Descripción detallada

En el presente documento se describen composiciones para unir dominios de unión a ADN, en particular módulos de dedos de cinc, a otros módulos de dedos de cinc. A diferencia de los enlazadores descritos anteriormente, los enlazadores que se describen en el presente documento permiten la unión preferente y/o selectiva de dianas que llevan huecos entre subsitios del módulo de 1 o 2 pb. Los enlazadores también son capaces de unir dianas que llevan huecos de 1 o 2 pb a afinidades más altas que los diseños de enlazadores actuales. Se muestran enlazadores de ejemplo en las Tablas 11 y 13. Por tanto, determinados enlazadores que se describen en el presente documento aumentan significativamente la capacidad de diseñar proteínas de dedos de cinc que se unen a sitios diana específicos, aumentando de este modo la actividad de las proteínas de fusión (por ejemplo, ZFP-TF o ZFN) que comprenden estos enlazadores.

General

La puesta en práctica de los métodos, así como la preparación y el uso de las composiciones que se desvelan en el presente documento emplean, a menos que se indique lo contrario, técnicas convencionales en biología molecular, bioquímica, estructura y análisis de cromatina, química computacional, cultivo celular, ADN recombinante y campos relacionados, están dentro de la experiencia de la técnica. Estas técnicas se explican por completo en la bibliografía. Véase, por ejemplo, Sambrook *et al.* *MOLECULAR CLONING: A LABORATORY MANUAL*, Segunda edición, Cold Spring Harbor Laboratory Press, 1989 y Tercera edición, 2001; Ausubel *et al.*, *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, John Wiley & Sons, Nueva York, 1987 y actualizaciones periódicas; la serie *METHODS IN ENZYMOLOGY*, Academic Press, San Diego; Wolffe, *CHROMATIN STRUCTURE AND FUNCTION*, Tercera edición, Academic Press, San Diego, 1998; *METHODS IN ENZYMOLOGY*, vol. 304, "Chromatin" (P.M. Wassarman y A. P. Wolffe, ed.), Academic Press, San Diego, 1999; y *METHODS IN MOLECULAR BIOLOGY*, vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definiciones

Las expresiones "ácido nucleico", "polinucleótido" y "oligonucleótido" se usan indistintamente y se refieren a un desoxirribonucleótido o polímero de ribonucleótido, en conformación lineal o circular y en forma monocatenaria o bicatenaria. Para los fines de la presente divulgación, estas expresiones no deben interpretarse como limitantes con respecto a la longitud de un polímero. Las expresiones pueden abarcar análogos conocidos de nucleótidos naturales, así como nucleótidos que se modifican en los restos de base, azúcar y/o fosfato (por ejemplo, cadenas principales de fosforotioato). En general, un análogo de un nucleótido particular tiene la misma especificidad de apareamiento de bases; es decir, un análogo de A se apareará con T.

Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente para referirse a un polímero de restos de aminoácidos. Los términos también se aplican a polímeros de aminoácidos en los que uno o más aminoácidos son análogos químicos o derivados modificados de un aminoácido de origen natural correspondiente.

Un polipéptido es normalmente sustancialmente idéntico a un segundo polipéptido, por ejemplo, donde los dos péptidos difieren solo en sustituciones conservadoras. Una "sustitución conservadora", cuando describe una proteína, se refiere a un cambio en la composición de aminoácidos de la proteína que no altera sustancialmente la actividad de la proteína. Por tanto, "variaciones modificadas conservadoramente" de una secuencia de aminoácidos particular se refiere a sustituciones de aminoácidos de los aminoácidos que no son críticos para la actividad de la proteína o a la sustitución de aminoácidos con otros aminoácidos que tienen propiedades similares (por ejemplo, ácidos, básicos, con carga positiva o negativa, polar o no polar, etc.) de manera que las sustituciones de aminoácidos incluso críticos no alteren sustancialmente la actividad. Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares son bien conocidas en la técnica. Véanse, por ejemplo, Creighton (1984) *Proteins*, W. H. Freeman and Company. Además, las sustituciones, supresiones o adiciones individuales que alteran, añaden o eliminan un único aminoácido o un pequeño porcentaje de aminoácidos en una secuencia codificada son también "variaciones modificadas conservadoramente".

"Unión" se refiere a una interacción no covalente específica de secuencia entre macromoléculas (por ejemplo, entre una proteína y un ácido nucleico). No todos los componentes de una interacción de unión deben ser específicos de secuencia (por ejemplo, contactos con grupos fosfato en una cadena principal de ADN), siempre que la interacción en su conjunto sea específica de secuencia. Dichas interacciones se caracterizan generalmente por una constante de disociación (K_d) de 10^{-6} M⁻¹ o inferior. "Afinidad" se refiere a la fuerza de la unión: definiéndose el aumento de la afinidad de unión por una K_d más baja.

Una "proteína de unión" es una proteína que puede unirse no covalentemente a otra molécula. Una proteína de unión puede unirse a, por ejemplo, una molécula de ADN (una proteína de unión a ADN), una molécula de ARN (una proteína de unión a ARN) y/o una molécula de proteína (una proteína de unión a proteína). En el caso de una proteína de unión a proteína, puede unirse a sí misma (para formar homodímeros, homotrímeros, etc.) y/o puede unirse a una o más moléculas de una proteína o proteínas diferentes. Una proteína de unión puede tener más de un tipo de actividad de unión. Por ejemplo, las proteínas de dedos de cinc tienen actividad de unión a ADN, unión a ARN y unión a proteína.

Una "proteína de unión a ADN de dedos de cinc" (o dominio de unión) es una proteína o un dominio dentro de una proteína más grande, que se une a ADN de una manera específica de secuencia a través de uno o más dedos de cinc, que son regiones de secuencia de aminoácidos dentro del dominio de unión cuya estructura se estabiliza a través de la coordinación de un ion de cinc. La expresión proteína de unión a ADN de dedos de cinc con frecuencia se abrevia como proteína de dedos de cinc o ZFP (por sus siglas en inglés).

Los dominios de unión de dedos de cinc (por ejemplo, regiones de reconocimiento de dedos de cinc) pueden "modificarse mediante ingeniería genética" para unirse a una secuencia de nucleótidos predeterminada. Son ejemplos no limitantes de métodos para modificar proteínas de dedos de cinc mediante ingeniería genética, el diseño

y la selección. Una proteína de dedos de cinc diseñada es una proteína que no aparece en la naturaleza cuyo diseño/composición es resultado principalmente de criterios racionales. Los criterios racionales para el diseño incluyen la aplicación de reglas de sustitución y algoritmos informatizados para procesar información en una base de datos que almacena información de diseños de ZFP existentes y datos de unión. Véanse, por ejemplo, las Patentes de los EE.UU. 6.140.081; 6.453.242; y 6.534.261; véanse también los documentos WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 y WO 03/016496.

Una proteína de dedos de cinc "seleccionada" es una proteína que no se encuentra en la naturaleza cuya producción es resultado principalmente de un proceso empírico tal como la presentación de fagos, la trampa de interacción o la selección híbrida. Véanse, por ejemplo, los documentos US 5.789.538; US 5.925.523; US 6.007.988; US 6.013.453; US 6.200.759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970, WO 01/88197 y WO 02/099084.

Un "dominio regulador" o "dominio funcional" se refiere a una proteína o un dominio de proteína que tiene actividad de modulación de la transcripción cuando está conectada a un dominio de unión a ADN, es decir, una ZFP. Normalmente, un dominio regulador se une covalentemente o no covalentemente a una ZFP (por ejemplo, para formar una molécula de fusión) para efectuar la modulación de la transcripción. Los dominios reguladores pueden ser dominios de activación o dominios de represión. Los dominios de activación incluyen, pero sin limitación, VP16, VP64 y la subunidad p65 del factor nuclear Kappa B. Los dominios de represión incluyen, pero sin limitación, KOX, KRAB MBD2B y v-ErbA. Los dominios reguladores adicionales incluyen, por ejemplo, factores de transcripción y cofactores (por ejemplo, MAD, ERD, SID, factor de respuesta de crecimiento temprano 1 y receptores de hormonas nucleares), endonucleasas, integrasas, recombinasas, metiltransferasas, histona acetiltransferasas, histona desacetilasas, etc. Los activadores y los represores incluyen coactivadores y correpresores (véase, por ejemplo, Utley *et al.*, *Nature* 394: 498-502 (1998)). Como alternativa, una ZFP puede actuar sola, sin un dominio regulador, para efectuar la modulación de la transcripción. Los dominios reguladores también pueden ser nucleasas, tales como dominios de escisión o semidominios de escisión.

"Escisión" se refiere a la ruptura de la cadena principal covalente de una molécula de ADN. La escisión puede iniciarse mediante una diversidad de métodos incluyendo, pero sin limitación, la hidrólisis enzimática o química de un enlace fosfodiéster. Es posible la escisión de cadena sencilla y la división de cadena doble, y la división de cadena doble puede producirse como resultado de dos eventos de escisión de cadena sencilla. La escisión de ADN puede dar como resultado la producción de extremos romos o extremos escalonados. En determinadas realizaciones, se usan polipéptidos de fusión para la escisión dirigida de ADN bicatenario.

Un "semidominio de escisión" es una secuencia polipeptídica que, junto con un segundo polipéptido (ya sea idéntico o diferente) forma un complejo que tiene actividad de escisión (preferentemente actividad de escisión de cadena doble). Las expresiones "semidominios de escisión primero y segundo"; "semidominios de escisión + y -" y "semidominios de escisión derecho e izquierdo" se usan indistintamente para referirse a pares de semidominios de escisión que se dimerizan.

Un "semidominio de escisión modificado mediante ingeniería genética" es un semidominio de escisión que se ha modificado de manera de formar heterodímeros obligados con otro semidominio de escisión (por ejemplo, otro semidominio de escisión modificado mediante ingeniería). Véase, *también*, la Publicación de Patente de los EE.UU. N.º 20050064474; y el documento WO 2007/13989.

La "cromatina" es la estructura nucleoproteínica que comprende el genoma celular. La cromatina celular comprende ácido nucleico, principalmente ADN y proteínas, incluyendo histonas y proteínas cromosómicas no histonas. La mayor parte de la cromatina celular eucariota existe en forma de nucleosomas, en donde un núcleo de nucleosoma comprende aproximadamente 150 pares de bases de ADN asociados a un octámero que comprende dos de cada una de las histonas H2A, H2B, H3 y H4; y ADN enlazador (de longitud variable dependiendo del organismo) y se extiende entre núcleos de nucleosomas. Generalmente se asocia una molécula de histona HI al ADN enlazador. Para los fines de la presente divulgación, el término "cromatina" pretende abarcar todos los tipos de nucleoproteína celular, tanto procariótica como eucariótica. La cromatina celular incluye cromatina tanto cromosómica como episómica.

Un "cromosoma" es un complejo de cromatina que comprende todo o una parte del genoma de una célula. El genoma de una célula con frecuencia se caracteriza por su cariotipo, que es la colección de todos los cromosomas que comprenden el genoma de la célula. El genoma de una célula puede comprender uno o más cromosomas.

Un "episoma" es un ácido nucleico replicativo, un complejo nucleoproteínico u otra estructura que comprende un ácido nucleico que no forma parte del cariotipo cromosómico de una célula. Los ejemplos de episomas incluyen plásmidos y determinados genomas víricos.

Una "región accesible" es un sitio en la cromatina celular en el que un sitio diana presente en el ácido nucleico puede estar unido por una molécula exógena que reconoce el sitio diana. Sin desear quedar ligado a ninguna teoría en particular, se cree que una región accesible es una que no está empaquetada en una estructura nucleosómica.

La estructura distintiva de una región accesible con frecuencia puede detectarse por su sensibilidad a sondas químicas y enzimáticas, por ejemplo, nucleasas.

5 Un "sitio diana" o "secuencia diana" es una secuencia de ácido nucleico que define una parte de un ácido nucleico al que se unirá una molécula de unión, a condición de que existan condiciones suficientes para la unión. Por ejemplo, la secuencia 5'-GAATTC-3' es un sitio diana para la endonucleasa de restricción Eco RI.

10 Un "subsitio de módulo" es una secuencia de ácido nucleico que define una parte de un ácido nucleico al que se unirá un módulo de dedos de cinc (por ejemplo, 1, 2, 3 o más dedos de cinc) dentro de una proteína de unión a ADN de dedos de cinc más grande, a condición de que existan condiciones suficientes para la unión.

15 Una molécula "exógena" es una molécula que normalmente no está presente en una célula, pero puede introducirse en una célula mediante uno o más métodos genéticos, bioquímicos u otros. La "presencia normal en la célula" se determina con respecto a la etapa de desarrollo particular y las condiciones ambientales de la célula. Por tanto, por ejemplo, una molécula que está presente solo durante el desarrollo embrionario del músculo es una molécula exógena con respecto a una célula muscular adulta. De manera similar, una molécula inducida por choque térmico es una molécula exógena con respecto a una célula no sometida a choque térmico. Una molécula exógena puede comprender, por ejemplo, una versión funcional de una molécula endógena defectuosa, una versión defectuosa de una molécula endógena que funciona normalmente o un ortólogo (versión funcional de una molécula endógena de una especie diferente).

20 Una molécula exógena puede ser, entre otras cosas, una molécula pequeña, tal como la que se genera mediante un proceso químico combinatorio o una macromolécula tal como una proteína, ácido nucleico, hidrato de carbono, lípido, glucoproteína, lipoproteína, polisacárido, cualquier derivado modificado de las moléculas anteriores o cualquier complejo que comprenda una o más de las moléculas anteriores. Los ácidos nucleicos incluyen ADN y ARN, pueden ser monocatenarios o bicatenarios; pueden ser lineales, ramificados o circulares; y pueden ser de cualquier longitud. Los ácidos nucleicos incluyen aquellos capaces de formar dúplex, así como ácidos nucleicos que forman tríplex. Véanse, por ejemplo, las Patentes de los EE.UU. N.º 5.176.996 y 5.422.251. Las proteínas incluyen, pero sin limitación, proteínas de unión a ADN, factores de transcripción, factores de remodelación de la cromatina, proteínas de unión a ADN metiladas, polimerasas, metilasas, desmetilasas, acetilasas, desacetilasas, cinasas, fosfatasas, integrasas, recombinasas, ligasas, topoisomerasas, girasas y helicasas.

25 Una molécula exógena puede ser el mismo tipo de molécula que una molécula endógena, por ejemplo, una proteína o un ácido nucleico exógenos. Por ejemplo, un ácido nucleico exógeno puede comprender un genoma vírico infectante, un plásmido o episoma introducido en una célula o un cromosoma que normalmente no está presente en la célula. Los expertos en la materia conocen métodos para la introducción de moléculas exógenas en las células e incluyen, pero sin limitación, transferencia mediada por lípidos (es decir, liposomas, incluyendo lípidos neutros y catiónicos), electroporación, inyección directa, fusión celular, bombardeo de partículas, coprecipitación con fosfato de calcio, transferencia mediada por DEAE-dextrano y transferencia mediada por vectores víricos.

30 Por el contrario, una molécula "endógena" es una que normalmente está presente en una célula particular en una etapa de desarrollo particular en condiciones ambientales particulares. Por ejemplo, un ácido nucleico endógeno puede comprender un cromosoma, el genoma de una mitocondria, cloroplasto u otro orgánulo, o un ácido nucleico episómico natural. Las moléculas endógenas adicionales pueden incluir proteínas, por ejemplo, factores de transcripción y enzimas.

35 Una molécula de "fusión" es una molécula en la que dos o más moléculas de subunidad están unidas, preferentemente de forma covalente. Las moléculas de subunidad pueden ser del mismo tipo químico de molécula o pueden ser de diferentes tipos químicos de moléculas. Los ejemplos del primer tipo de molécula de fusión incluyen, pero sin limitación, proteínas de fusión (por ejemplo, una fusión entre un dominio de unión a ADN de ZFP y un dominio de escisión) y ácidos nucleicos de fusión (por ejemplo, un ácido nucleico que codifica la proteína de fusión descrita anteriormente). Los ejemplos del segundo tipo de molécula de fusión incluyen, pero sin limitación, una fusión entre un ácido nucleico formador de tríplex y un polipéptido, y una fusión entre una proteína fijadora de surco menor y un ácido nucleico.

40 La expresión de una proteína de fusión en una célula puede ser resultado de la entrega de la proteína de fusión a la célula o mediante la entrega de un polinucleótido que codifica la proteína de fusión a una célula, en donde el polinucleótido se transcribe y el transcrito se traduce, para generar la proteína de fusión. El corte y empalme trans, la escisión de polipéptidos y la ligadura de polipéptidos también pueden estar implicados en la expresión de una proteína en una célula. Se presentan métodos para la entrega de polinucleótidos y polipéptidos a las células en otra parte en la presente divulgación.

45 Un "gen", para los fines de la presente divulgación, incluye una región de ADN que codifica un producto génico (véase más adelante), así como todas las regiones de ADN que regulan la producción del producto génico, ya sea que dichas secuencias reguladoras estén adyacentes o no a secuencias codificantes y/o transcritas. En consecuencia, un gen incluye, pero no se limita necesariamente a, secuencias promotoras, terminadores,

secuencias reguladoras de la traducción tales como sitios de unión a ribosomas y sitios internos de entrada al ribosoma, potenciadores, silenciadores, aisladores, elementos de contorno, orígenes de replicación, sitios de unión a matriz y regiones de control de locus.

5 La "expresión génica" se refiere a la conversión de la información, contenida en un gen, en un producto génico. Un producto génico puede ser el producto directo de la transcripción de un gen (por ejemplo, ARNm, ARNt, ARNr, ARN antisentido, ribozima, ARN estructural o cualquier otro tipo de ARN) o una proteína producida mediante traducción de un ARNm. Los productos génicos también incluyen ARN que se modifican, mediante procesos tales como protección con capuchón, poliadenilación, metilación y edición, y proteínas modificadas mediante, por ejemplo, metilación, acetilación, fosforilación, ubiquitinación, ribosilación de ADP, miristilación y glucosilación.

10 La "modulación" de la expresión génica se refiere a un cambio en la actividad de un gen. La modulación de la expresión puede incluir, pero sin limitación, la activación génica y la represión génica. La inactivación génica se refiere a cualquier reducción en la expresión génica en comparación con una célula que no incluye una ZFP como se describe en el presente documento. Por tanto, la inactivación de genes puede ser parcial o completa.

Las células "eucariotas" incluyen, pero sin limitación, células fúngicas (tales como levadura), células vegetales, células animales, células de mamíferos y células humanas (por ejemplo, linfocitos T).

20 Una "región de interés" es cualquier región de cromatina celular, tal como, por ejemplo, un gen o una secuencia no codificante dentro o adyacente a un gen, en la que es deseable unir una molécula exógena. La unión puede ser con fines de escisión de ADN dirigida y/o recombinación dirigida. Una región de interés puede estar presente en un cromosoma, un episoma, un genoma organular (por ejemplo, mitocondrial, cloroplasto) o un genoma vírico infectante, por ejemplo. Una región de interés puede estar dentro de la región codificante de un gen, dentro de regiones no codificantes transcritas tales como, por ejemplo, secuencias líder, secuencias de arrastre o intrones, o dentro de regiones no transcritas, ya sea aguas arriba o aguas abajo de la región codificante. Una región de interés puede ser tan pequeña como de un solo par de nucleótidos o de hasta 20.000 pares de nucleótidos de longitud, o cualquier valor integral de pares de nucleótidos, o de hasta la longitud de un cromosoma. No es necesario que una región de interés comprenda solo secuencias contiguas de ácido nucleico.

30 Las expresiones "unión operativa" y "unido operativamente" (u "unido operablemente") se usan indistintamente con referencia a una yuxtaposición de dos o más componentes (tales como elementos de secuencia), en la que los componentes se disponen de manera que ambos componentes funcionen normalmente y permitan la posibilidad de que al menos uno de los componentes pueda mediar una función que se ejerce sobre al menos uno de los otros componentes. A modo de ilustración, una secuencia reguladora de la transcripción, tal como un promotor, se une operativamente a una secuencia codificante si la secuencia reguladora de la transcripción controla el nivel de transcripción de la secuencia codificante en respuesta a la presencia o ausencia de uno o más factores reguladores de la transcripción. Una secuencia reguladora de la transcripción generalmente se une operativamente en *cis* con una secuencia codificante, pero no es necesario que esté directamente adyacente a ella. Por ejemplo, un potenciador es una secuencia reguladora de la transcripción que se une operativamente a una secuencia codificante, aunque no son contiguas.

45 Con respecto a los polipéptidos de fusión, la expresión "unido operativamente" puede referirse al hecho de que cada uno de los componentes realiza la misma función en unión con el otro componente como lo haría si no estuviera unido de este modo. Por ejemplo, con respecto a un polipéptido de fusión en el que un dominio de unión a ADN de ZFP se fusiona con un dominio de escisión, el dominio de unión a ADN de ZFP y el dominio de escisión están en unión operativa si, en el polipéptido de fusión, la parte de dominio de unión a ADN de ZFP es capaz de unir su sitio diana y/o su sitio de unión, mientras que el dominio de escisión es capaz de escindir el ADN en las proximidades del sitio diana.

50 Un "fragmento funcional" de una proteína, polipéptido o ácido nucleico es una proteína, polipéptido o ácido nucleico cuya secuencia no es idéntica a la proteína, polipéptido o ácido nucleico de longitud completa, pero conserva una o más de las funciones de la proteína, polipéptido o ácido nucleico de longitud completa. Un fragmento funcional puede poseer más, menos, o el mismo número de restos que la molécula nativa correspondiente, y/o puede contener una o más sustituciones de aminoácidos o nucleótidos. Se conocen bien en la técnica métodos para determinar la función de un ácido nucleico (por ejemplo, la función codificante, la capacidad de hibridarse con otro ácido nucleico). De manera similar, se conocen bien métodos para determinar la función de la proteína. Por ejemplo, la función de unión a ADN de un polipéptido puede determinarse, por ejemplo, mediante unión a filtro, cambio de movilidad electroforética o ensayos de inmunoprecipitación. La escisión del ADN puede someterse a ensayo mediante electroforesis en gel. Véase Ausubel *et al.*, citado anteriormente. La capacidad de una proteína para interactuar con otra proteína puede determinarse, por ejemplo, mediante coimmunoprecipitación, ensayos de dos híbridos o complementación, tanto genética como bioquímica. Véanse, por ejemplo, Fields *et al.* (1989) *Nature* **340**: 245-246; la Patente de los EE.UU. N.º 5.585.245 y el documento PCT WO 98/44350.

65 **Enlazadores**

En el presente documento se describen secuencias de aminoácidos que fusionan (unen) módulos de unión a ADN (por ejemplo, módulos de dedos de cinc) entre sí. Los módulos de dedos de cinc fusionados usando los enlazadores que se describen en el presente documento pueden contener 1, 2, 3, 4 o incluso más dedos de cinc. En determinadas realizaciones, los módulos de dedos de cinc contienen 1, 2 o 3 dedos de cinc, que cuando se unen entre sí forman una proteína de dedos de cinc de 3 dedos o más.

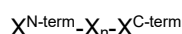
Las secuencias de enlazador que se describen en el presente documento se extienden entre el último resto de la hélice α en un dedo de cinc y el primer resto de la lámina β en el siguiente dedo de cinc. La secuencia de enlazador, por tanto, une dos dedos de cinc. Normalmente, el último aminoácido (C-terminal) en un dedo de cinc es el resto de coordinación de cinc C-terminal, mientras que un resto aromático (por ejemplo, Phe) es normalmente el primer aminoácido del siguiente dedo de cinc. En consecuencia, en un dedo de cinc de "tipo silvestre", la treonina es el primer resto en el enlazador y la prolina es el último resto del enlazador. Por tanto, por ejemplo, la secuencia de enlazador canónica para Zif268 es TG(E/Q)(K/R)P (SEQ ID NO: 129). Véanse, por ejemplo, las Patentes de los EE.UU. N.º 6.479.626; 6.903.185 y 7.153.949.

Se describen enlazadores adicionales, por ejemplo, en la Publicación de Patente de los EE.UU. 20030119023, que describe enlazadores que incluyen múltiples restos de glicina (por ejemplo, TGGGGSQKP (SEQ ID NO: 2), TGGGSGGSQKP (SEQ ID NO: 14) y TGGGSGGSGGSQKP (SEQ ID NO: 15), TGGEKP (SEQ ID NO: 16), TGGQKP (SEQ ID NO: 17), TGGSGEKP (SEQ ID NO: 18), TGGSGQKP (SEQ ID NO: 19), TGGSGGSQKP (SEQ ID NO: 20) y TGGSGGSQKP (SEQ ID NO: 21).

Normalmente, los enlazadores se preparan usando ácidos nucleicos recombinantes que codifican el enlazador y los módulos de unión a ácido nucleico, que se fusionan a través de la secuencia de aminoácidos del enlazador. Los enlazadores también pueden prepararse usando síntesis peptídica y después se uniéndolos a los módulos de unión a ácido nucleico. Se conocen en la técnica métodos de manipulación de ácidos nucleicos y métodos de síntesis peptídica (véase, por ejemplo, Maniatis, *et al.*, 1991. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press).

Los enlazadores que se describen en el presente documento son más rígidos que los enlazadores utilizados anteriormente y permitir la unión eficiente de cada módulo de dedos de cinc a su sitio diana solo cuando los subsitios están separados por un número específico de pares de bases.

Por tanto, a diferencia de los enlazadores anteriores, los enlazadores que se describen en el presente documento incluyen al menos un resto de prolina interno o N-terminal, es decir, un resto de prolina que no está en el extremo C-terminal del enlazador. Los enlazadores que se describen en el presente documento tienen la siguiente estructura general de aminoácidos:



donde X es cualquier resto de aminoácido, X_n comprende al menos 3 restos de aminoácidos y al menos uno de entre $X^{N\text{-term}}$ y X_n comprende un resto de prolina. Se muestran ejemplos no limitantes de dichos enlazadores en las Tablas 4, 5, 6, 9, 10, 11 o 13. Asimismo, los enlazadores que se describen en el presente documento también incluyen normalmente al menos dos restos básicos, por ejemplo, uno o más restos de arginina, uno o más restos de histidina, uno o más restos de lisina o combinaciones de los mismos.

Los enlazadores de la invención pueden tener de 5 a 20 aminoácidos de longitud. En determinadas realizaciones, los enlazadores tienen 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 aminoácidos de longitud. La materia objeto de la invención son secuencias de enlazador como se definen en las proteínas de dedos de cinc de múltiples dedos de las reivindicaciones.

Módulos de unión a ADN

Las secuencias de enlazador que se describen en el presente documento se usan ventajosamente para unir módulos de unión a ADN.

En los métodos que se desvelan en el presente documento puede usarse cualquier dominio de unión a ADN. En determinadas realizaciones, el dominio de unión a ADN comprende una proteína de dedos de cinc. Preferentemente, la proteína de dedos de cinc no es de origen natural ya que se modifica mediante ingeniería genética para su unión al sitio diana de elección. Véanse, por ejemplo, Beerli *et al.* (2002) *Nature Biotechnol.* **20**: 135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* **70**: 313-340; Isalan *et al.* (2001) *Nature Biotechnol.* **19**: 656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* **12**: 632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* **10**: 411-416. Un dominio de unión de dedos de cinc modificado mediante ingeniería genética puede tener una especificidad de unión novedosa, en comparación con una proteína de dedos de cinc de origen natural. Los métodos de ingeniería genética incluyen, pero sin limitación, el diseño racional y diversos tipos de selección. El diseño racional incluye, por ejemplo, el uso de bases de datos que comprenden secuencias de nucleótidos de triplete (o cuadruplete) y secuencias de aminoácidos de dedos de cinc individuales, en el que cada secuencia de nucleótidos de triplete o cuadruplete se asocia a una o más secuencias de

aminoácidos de dedos de cinc que unen la secuencia de triplete o cuadruplete particular. Véanse, por ejemplo, las Patentes de los EE.UU. 6.453.242 y 6.534.261 del mismo solicitante.

5 Se desvelan ejemplos de métodos de selección, incluyendo la presentación en fagos y sistemas de dos híbridos, en las Patentes de los EE.UU. 5.789.538; 5.925.523; 6.007.988; 6.013.453; 6.410.248; 6.140.466; 6.200.759; y 6.242.568; así como en los documentos WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 y GB 2.338.237. Además, se ha descrito la potenciación de la especificidad de unión para dominios de unión de dedos de cinc, por ejemplo, en el documento WO 02/077227 del mismo solicitante.

10 Los expertos en la materia conocen la selección de sitios diana; ZFP y métodos para el diseño y la construcción de proteínas de fusión (y polinucleótidos que las codifican) y éstos se describen en detalle en las Publicaciones de Solicitud de Patente de los EE.UU. N.º 20050064474 y 20060188987.

15 Además, como se desvela en estas y otras referencias, los dominios de dedos de cinc y/o las proteínas de dedos de cinc de múltiples dedos pueden unirse entre sí usando cualquier secuencia de enlazador adecuada, incluyendo, por ejemplo, enlazadores de 5 a 20 aminoácidos de longitud. Véanse, también, las Patentes de los EE.UU. N.º 6.479.626; 6.903.185; y 7.153.949 para secuencias de enlazador de ejemplo de 6 o más aminoácidos de longitud. Las proteínas que se describen en el presente documento pueden incluir cualquier combinación de enlazadores adecuados entre los dedos de cinc individuales de la proteína.

20 Como alternativa, el dominio de unión a ADN puede derivar de una nucleasa. Por ejemplo, las secuencias de reconocimiento de endonucleasas de asentamiento y meganucleasas tales como I-Scel, I-CeuI, PI-PspI, PI-Sce, I-ScelV, I-Csml, I-PanI, I-Scell, I-Ppol, I-ScellI, I-Crel, I-TevI, I-TevII y I-TevIII, son conocidas. Véanse también la Patente de los EE.UU. N.º 5.420.032; la Patente de los EE.UU. N.º 6.833.252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**: 3379-3388; Dujon *et al.* (1989) *Gene* **82**: 115-118; Perler *et al.* (1994) *Nucleic Acids Res.* **22**, 1125-1127; Jasin (1996) *Trends Genet.* **12**: 224-228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**: 163-180; Argast *et al.* (1998) *J. Mol. Biol.* **280**: 345-353 y el catálogo de New England Biolabs. Además, la especificidad de unión a ADN de las endonucleasas de asentamiento y de las meganucleasas puede modificarse mediante ingeniería genética para la unión a sitios diana no naturales. Véanse, por ejemplo, Chevalier *et al.* (2002) *Molec. Cell* **10**: 895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* **31**: 2952-2962; Ashworth *et al.* (2006) *Nature* **441**: 656-659; Paques *et al.* (2007) *Current Gene Therapy* **7**: 49-66; la Publicación de Patente de los EE.UU. N.º 20070117128.

35 En algunas realizaciones, el dominio de unión a ADN es un dominio modificado mediante ingeniería genética de un efector TAL (TALE) derivado del patógeno vegetal *Xanthomonas* (véase, Miller *et al.* (2010) *Nature Biotechnology*, **22** de diciembre [Epub antes de impresión]; Boch *et al.* (2009) *Science* **29** de octubre de 2009 (10.1126/science.117881) y Moscou y Bogdanove, (2009) *Science* **29** de octubre de 2009 (10.1126/science.1178817); véanse, también, las Solicitudes Provisionales de los EE.UU. N.º 61/395.836, presentada el 17 de mayo de 2010; 61/409.421, presentada el sábado, 21 de agosto de 2010; 61/45.121, presentada el miércoles, 13 de octubre de 2010; 61/459.891, presentada el lunes, 20 de diciembre de 2010; la Solicitud N.º 61/462.482, presentada el miércoles, 2 de febrero de 2011; la Solicitud N.º 61/465.869, presentada el 40 24 de marzo de 2011.

Dominios reguladores

45 Los módulos de dedos de cinc unidos como se describen en el presente documento con frecuencia se expresan con un dominio exógeno (o fragmento funcional del mismo) como proteínas de fusión. Los dominios reguladores comunes para la adición a la ZFP incluyen, por ejemplo, dominios de factores de transcripción (activadores, represores, coactivadores, corepresores), silenciadores, oncogenes (por ejemplo, miembros de la familia myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos, etc.); enzimas de reparación de ADN y sus factores y modificadores asociados; enzimas de reordenamiento de ADN y sus factores y modificadores asociados; proteínas asociadas a cromatina y sus modificadores (por ejemplo, cinasas, acetilasas y desacetilasas); y enzimas modificadoras de ADN (por ejemplo, metiltransferasas, topoisomerasas, helicasas, ligasas, cinasas, fosfatasas, polimerasas, endonucleasas) y sus factores y modificadores asociados.

55 Un dominio funcional de ejemplo para la fusión con un dominio de unión a ADN tal como, por ejemplo, una ZFP, que ha de usarse para reprimir la expresión de un gen es un dominio de represión KRAB de la proteína KOX-1 humana (véase, por ejemplo, Thiesen *et al.*, *New Biologist* **2**, 363-374 (1990); Margolin *et al.*, *Proc. Natl. Acad. Sci. USA* **91**, 4509-4513 (1994); Pengue *et al.*, *Nucl. Acids Res.* **22**: 2908-2914 (1994); Witzgall *et al.*, *Proc. Natl. Acad. Sci. USA* **91**, 4514-4518 (1994). Otro dominio de represión adecuado es la proteína de dominio de unión a metilo 2B (MBD-2B) (véase también Hendrich *et al.* (1999) *Mamm. Genome* **10**: 906-912 para una descripción de proteínas MBD). Otro dominio de represión útil es el asociado a la proteína v-ErbA. Véanse, por ejemplo, Damm, *et al.* (1989) *Nature* **339**: 593-597; Evans (1989) *Int. J. Cancer Suppl.* **4**: 26-28; Pain *et al.* (1990) *New Biol.* **2**: 284-294; Sap *et al.* (1989) *Nature* **340**: 242-244; Zenke *et al.* (1988) *Cell* **52**: 107-119; y Zenke *et al.* (1990) *Cell* **61**: 1035-1049.

65 Los dominios de represión de ejemplo adicionales incluyen, pero sin limitación, KRAB (también denominado "KOX"), SID, MBD2, MBD3, miembros de la familia DNMT (por ejemplo, DNMT1, DNMT3A, DNMT3B), Rb y MeCP2.

Véanse, por ejemplo, Bird *et al.* (1999) *Cell* 99: 451-454; Tyler *et al.* (1999) *Cell* 99: 443-446; Knoepfler *et al.* (1999) *Cell* 99: 447-450; y Robertson *et al.* (2000) *Nature Genet.* 25: 338-342. Los dominios de represión de ejemplo adicionales incluyen, pero sin limitación, ROM2 y AtHD2A. Véanse, por ejemplo, Chem *et al.* (1996) *Plant Cell* 8: 305-321; y Wu *et al.* (2000) *Plant J.* 22: 19-27.

5 Los dominios adecuados para conseguir la activación incluyen el dominio de activación VP16 del VHS (véase, por ejemplo, Hagmann *et al.*, *J. Virol.* 71, 5952-5962 (1997)), receptores de hormonas nucleares (véase, por ejemplo, Torchia *et al.*, *Curr. Opin. Cell. Biol.* 10:373-383 (1998)); la subunidad p65 del factor nuclear kappa B (Bitko y Barik, *J. Virol.* 72: 5610-5618 (1998) y Doyle y Hunt, *Neuroreport* 8: 2937-2942 (1997)); Liu *et al.*, *Cancer Gene Ther.* 5: 3-28 (1998)) o dominios funcionales quiméricos artificiales tales como VP64 (Seifpal *et al.*, *EMBO J.* 11, 4961-4968 (1992)). Los dominios de activación de ejemplo adicionales incluyen, pero sin limitación, VP16, VP64, p300, CBP, PCAF, SRC1 PVALF, AtHD2A y ERF-2. Véanse, por ejemplo, Robyr *et al.* (2000) *Mol. Endocrinol.* 14: 329-347; Collingwood *et al.* (1999) *J. Mol. Endocrinol.* 23: 255-275; Leo *et al.* (2000) *Gene* 245:1-11; Manteuffel-Cymborowska (1999) *Acta Biochim. Pol.* 46: 77-89; McKenna *et al.* (1999) *J. Steroid Biochem. Mol. Biol.* 69: 3-12; Malik *et al.* (2000) *Trends Biochem. Sci.* 25: 277-283; y Lemon *et al.* (1999) *Curr. Opin. Genet. Dev.* 9: 499-504. Los dominios de activación de ejemplo adicionales incluyen, pero sin limitación, OsGAL, HALF-1, C1, API, ARF-5, -6, -7 y -8, CPRF1, CPRF4, MYC-RP/GP y TRAB1. Véanse, por ejemplo, Ogawa *et al.* (2000) *Gene* 245: 21-29; Okanami *et al.* (1996) *Genes Cells* 1: 87-99; Goff *et al.* (1991) *Genes Dev.* 5: 298-309; Cho *et al.* (1999) *Plant Mol. Biol.* 40: 419-429; Ulmason *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849; Sprenger-Hausseis *et al.* (2000) *Plant J.* 22: 1-8; Gong *et al.* (1999) *Plant Mol. Biol.* 41: 33-44; y Hobo *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96: 15,348-15,353.

En determinadas realizaciones, el dominio regulador comprende una nucleasa (por ejemplo, dominio de escisión). Dichas nucleasas modificadas mediante ingeniería genética pueden usarse para crear una ruptura bicatenaria (DSB, por sus siglas en inglés) en una secuencia de nucleótidos diana, lo que aumenta la frecuencia de introducción de ácido nucleico donante a través de recombinación homóloga en el locus diana (integración dirigida) más de 1000 veces. Además, la reparación inexacta de una DSB específica del sitio mediante unión de extremos no homólogos (NHEJ) también puede provocar la alteración del gen. Las nucleasas pueden usarse para una amplia diversidad de fines, tales como la modificación mediante ingeniería genética de estirpes celulares, así como para aplicaciones terapéuticas.

30 Los dominios de escisión de las proteínas de fusión que se desvelan en el presente documento pueden obtenerse de cualquier endonucleasa o exonucleasa. Las endonucleasas de ejemplo de las que puede derivar un dominio de escisión incluyen, pero sin limitación, endonucleasas de restricción y endonucleasas de asentamiento. Véanse, por ejemplo, *2002-2003 Catalogue*, New England Biolabs, Beverly, MA; y Belfort *et al.* (1997) *Nucleic Acids Res.* 25: 3379-3388. Se conocen enzimas adicionales que escinden el ADN (por ejemplo, Nucleasa S1; nucleasa de judía mungo; DNasa pancreática I; nucleasa micrococcica; endonucleasa HO de levadura; véase también Linn *et al.* (ed.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). Una o más de estas enzimas (o fragmentos funcionales de las mismas) pueden usarse como fuente de dominios de escisión y semidominios de escisión.

40 De manera similar, un semidominio de escisión puede derivar de cualquier nucleasa o parte de la misma, como se ha expuesto anteriormente, que requiere dimerización para la actividad de escisión. En general, se necesitan dos proteínas de fusión para la escisión si las proteínas de fusión comprenden semidominios de escisión. Como alternativa, puede usarse una única proteína que comprenda dos semidominios de escisión. Los dos semidominios de escisión pueden derivar de la misma endonucleasa (o fragmentos funcionales de la misma) o cada semidominio de escisión puede derivar de una endonucleasa diferente (o fragmentos funcionales de la misma).

Además, los sitios diana para las dos proteínas de fusión se disponen, preferentemente, uno con respecto al otro, de manera que la unión de las dos proteínas de fusión a sus respectivos sitios diana coloque los semidominios de escisión en una orientación espacial entre sí que permita que los semidominios de escisión formen un dominio de escisión funcional, por ejemplo, mediante dimerización. Por tanto, en determinadas realizaciones, los bordes cercanos de los sitios diana están separados por 5-8 nucleótidos o por 15-18 nucleótidos. Sin embargo, puede intervenir cualquier número integral de nucleótidos o pares de nucleótidos entre dos sitios diana (por ejemplo, de 2 a 50 pares de nucleótidos o más). En general, el sitio de escisión se encuentra entre los sitios diana.

55 Hay presentes endonucleasas de restricción (enzimas de restricción) en muchas especies y son capaces de unirse de manera específica de secuencia al ADN (en un sitio de reconocimiento) y escindir el ADN en el sitio de unión o cerca del mismo. Determinadas enzimas de restricción (por ejemplo, Tipo IIS) escinden el ADN en sitios eliminados del sitio de reconocimiento y tienen dominios de unión y escisión separables. Por ejemplo, la enzima de Tipo IIS *Fok I* cataliza la escisión bicatenaria de ADN, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y 13 nucleótidos desde su sitio de reconocimiento en la otra. Véanse, por ejemplo, las Patentes de los EE.UU. 5.356.802; 5.436.150 y 5.487.994; así como Li *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89: 4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90: 2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* 91: 883-887; Kim *et al.* (1994b) *J. Biol. Chem.* 269: 31.978-31.982. Por tanto, en una realización, las proteínas de fusión comprenden el dominio de escisión (o semidominio de escisión) de al menos una enzima de restricción de Tipo IIS y uno o más dominios de unión de dedos de cinc, que pueden o no estar modificados mediante ingeniería genética.

Una enzima de restricción de Tipo IIS de ejemplo, cuyo dominio de escisión es separable del dominio de unión, es *Fok I*. Esta enzima particular es activa en forma de dímero. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95**: 10.570-10.575. En consecuencia, para los fines de la presente divulgación, la parte de la enzima *Fok I* utilizada en las proteínas de fusión que se desvelan se considera un semidominio de escisión. Por tanto, para la escisión bicatenaria dirigida y/o la sustitución dirigida de secuencias celulares usando fusiones de dedos de cinc-*Fok I* pueden usarse dos proteínas de fusión, comprendiendo cada una un semidominio de escisión *FokI*, para reconstituir un dominio de escisión catalíticamente activo. Como alternativa, también puede usarse una única molécula de polipéptido que contenga un dominio de unión de dedos de cinc y dos semidominios de escisión *Fok I*. En otra parte de la presente divulgación se proporcionan parámetros para la escisión dirigida y la alteración de secuencia dirigida usando fusiones de dedos de cinc-*Fok I*.

Un dominio de escisión o semidominio de escisión puede ser cualquier parte de una proteína que conserve la actividad de escisión o que conserve la capacidad de multimerizarse (por ejemplo, dimerizarse) para formar un dominio de escisión funcional.

Se describen ejemplos de enzimas de restricción de Tipo IIS en la Publicación Internacional WO 07/014275. Enzimas de restricción adicionales también contienen dominios de unión y escisión separables, y éstas se contemplan en la presente divulgación. Véase, por ejemplo, Roberts *et al.* (2003) *Nucleic Acids Res.* **31**: 418-420.

En determinadas realizaciones, el dominio de escisión comprende uno o más semidominios de escisión modificados mediante ingeniería genética (también denominados mutantes de dominio de dimerización) que minimizan o evitan la homodimerización, como se describe, por ejemplo, en las Publicaciones de Patentes de los EE.UU. N.º 20050064474 y 20060188987, y en la Solicitud de los EE.UU. N.º 11/805.850 (presentada el 23 de mayo de 2007), Los restos de aminoácidos en las posiciones 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537 y 538 de *Fok I* son todos dianas para influir en la dimerización de los semidominios de escisión de *Fok I*.

Los semidominios de escisión modificados mediante ingeniería genética de ejemplo de *Fok I* que forman heterodímeros obligados incluyen un par en el que un primer semidominio de escisión incluye mutaciones en los restos de aminoácidos en las posiciones 490 y 538 de *Fok I* y un segundo semidominio de escisión incluye mutaciones en los restos de aminoácidos 486 y 499.

Por tanto, en una realización, una mutación en 490 reemplaza Glu (E) con Lys (K); la mutación en 538 reemplaza Iso (I) con Lys (K); la mutación en 486 reemplazó Gln (Q) con Glu (E); y la mutación en la posición 499 reemplaza Iso (I) con Lys (K). Específicamente, los semidominios de escisión modificados mediante ingeniería genética que se describen en el presente documento se prepararon mutando las posiciones 490 (E→K) y 538 (I→K) en un semidominio de escisión para producir un semidominio de escisión modificado mediante ingeniería genética designado "E490K:I538K" y mutando las posiciones 486 (Q→E) y 499 (I→L) en otro semidominio de escisión para producir un semidominio de escisión modificado mediante ingeniería genética designado "Q486E:I499L". Los semidominios de escisión modificados mediante ingeniería genética que se describen en el presente documento son mutantes heterodiméricos obligados en los que la escisión anormal se minimiza o se suprime. Véase, por ejemplo, el Ejemplo 1 del documento WO 07/139898. En determinadas realizaciones, el semidominio de escisión modificado mediante ingeniería genética comprende mutaciones en las posiciones 486, 499 y 496 (numeradas con respecto a *FokI* de tipo silvestre), por ejemplo, mutaciones que reemplazan el resto de Gln (Q) de tipo silvestre en la posición 486 con un resto de Glu (E), el resto de Iso (I) de tipo silvestre en la posición 499 con un resto de Leu (L) y el resto de Asn (N) de tipo silvestre en la posición 496 con un resto de Asp (D) o Glu (E) (también denominados dominios "ELD" y "ELE", respectivamente). En otras realizaciones, el semidominio de escisión modificado mediante ingeniería genética comprende mutaciones en las posiciones 490, 538 y 537 (numeradas con respecto a *FokI* de tipo silvestre), por ejemplo, mutaciones que reemplazan el resto de Glu (E) de tipo silvestre en la posición 490 con un resto de Lys (K), el resto de Iso (I) de tipo silvestre en la posición 538 con un resto de Lys (K) y el resto de His (H) de tipo silvestre en la posición 537 con un resto de Lys (K) o un resto de Arg (R) (también denominados dominios "KKK" y "KKR", respectivamente). En otras realizaciones, el semidominio de escisión modificado mediante ingeniería genética comprende mutaciones en las posiciones 490 y 537 (numeradas con respecto a *FokI* de tipo silvestre), por ejemplo, mutaciones que reemplazan el resto de Glu (E) de tipo silvestre en la posición 490 con un resto de Lys (K) y el resto de His (H) de tipo silvestre en la posición 537 con un resto de Lys (K) o un resto de Arg (R) (también denominados dominios "KIK" y "KIR", respectivamente). (Véase la Solicitud de Patente de los EE.UU. N.º: 12/931.660).

Los semidominios de escisión modificados mediante ingeniería genética que se describen en el presente documento pueden prepararse usando cualquier método adecuado, por ejemplo, mediante mutagénesis dirigida al sitio de semidominios de escisión de tipo silvestre (*FokI*) como se describe en la Publicación de Patente de los EE.UU. N.º 20050064474 (véase, por ejemplo, el Ejemplo 5); y el documento WO 07/139898.

Como alternativa, las nucleasas pueden ensamblarse *in vivo* en el sitio diana del ácido nucleico usando la denominada tecnología de "enzima dividida" (véase, por ejemplo, la Publicación de Patente de los EE.UU. N.º 20090068164). Los componentes de dichas enzimas vézamas divididas pueden expresarse en construcciones de expresión separadas o pueden unirse en un marco de lectura abierto donde los componentes individuales están separados, por ejemplo, por una secuencia IRES o péptido 2A autoescindible. Los componentes pueden ser dominios de unión

de dedos de cinc individuales o dominios de un dominio de unión de ácido nucleico de meganucleasa.

Las moléculas de fusión se construyen mediante métodos de clonación y conjugación bioquímica que son bien conocidos por los expertos en la materia. Las moléculas de fusión comprenden un dominio de unión a ADN y un dominio funcional (por ejemplo, un dominio de activación o de represión de la transcripción). Las moléculas de fusión también comprenden opcionalmente señales de localización nuclear (tal como, por ejemplo, la del antígeno T de medio de SV40) y marcadores de epítipo (tales como, por ejemplo, FLAG y hemaglutinina). Las proteínas de fusión (y los ácidos nucleicos que las codifican) se diseñan de manera que se conserve el marco de lectura de la traducción entre los componentes de la fusión.

Para dichas aplicaciones, la molécula de fusión normalmente se formula con un vehículo farmacéuticamente aceptable, como saben los expertos en la materia. Véase, por ejemplo, *Remington's Pharmaceutical Sciences*, 17ª edición, 1985; y el documento WO 00/42219 del mismo solicitante.

Kits

En el presente documento se describen kits que comprenden cualquiera de los enlazadores y/o para realizar cualquiera de los métodos anteriores. Los kits normalmente contienen una secuencia de enlazador como se describe en el presente documento (o un polinucleótido que codifica un enlazador como se describe en el presente documento). El kit puede suministrar el enlazador solo o puede proporcionar vectores en los que puede insertarse fácilmente un dominio de unión a ADN y/o una nucleasa de elección. Los kits también pueden contener células, tampones para la transformación de células, medios de cultivo para células y/o tampones para realizar ensayos. Normalmente, los kits también contienen una etiqueta, que incluye cualquier material tal como instrucciones, embalaje o un folleto publicitario que se adjunta a, o que acompaña de otro modo, los otros componentes del kit.

Aplicaciones

Los enlazadores que se desvelan se usan ventajosamente para potenciar el repertorio de sitios diana para proteínas de dedos de cinc modificadas mediante ingeniería genética. Por ejemplo, los enlazadores que se describen en el presente documento facilitan la unión a los sitios diana deseados cuando los subsitios del módulo no son adyacentes. Por tanto, habría un aumento eficaz en el número de ZFP que podrían construirse para el direccionamiento a una secuencia de ácido nucleico dada para un tamaño de repertorio dado. Asimismo, debido a que los enlazadores que se describen distinguen entre diversas separaciones de subsitios del módulo (por ejemplo, huecos de 0, 1 y 2 pares de bases), reducen la unión de las ZFP a sitios diana inadecuados. Por ejemplo, una ZFP con un enlazador flexible diseñado para omitir 2 pares de bases (por ejemplo, TGGGGSGGSQKP (SEQ ID NO: 14)) puede unirse a sitios diana con 0, 1 o 2 pares de bases entre los subsitios del módulo. Esta misma ZFP con un enlazador de omisión de 2 pb como se describe en el presente documento debería unirse bien a una diana con 2 pares de bases entre los subsitios del módulo, pero no debería ser capaz de unirse eficientemente a dianas con 0 o 1 par de bases entre los subsitios del módulo (sitios diana inadecuados o no intencionados).

Por tanto, pueden usarse enlazadores que se describen en el presente documento en cualquier aplicación para la que se usen actualmente proteínas de dedos de cinc, incluyendo, pero sin limitación, factores de transcripción de dedos de cinc (ZFP-TF) para la modulación de la expresión génica y/o en nucleasas de dedos de cinc (ZFN) para la escisión. Véanse, por ejemplo, las Patentes de los EE.UU. N.º 6.534.261; 6.599.692; 6.689.558; 7.067.317; 7.262.054 y 7.253.273; las Publicaciones de Patente de los EE.UU. N.º 20050064474; 2006/0063231; 2007/0134796; 2007/0218528; 2008/0015164; 2008/0188000; 2008/0299580 y 2008/0159996.

En consecuencia, los enlazadores que se desvelan pueden usarse en cualquier ZFP o ZFN para cualquier método en el que sea deseable una modulación o escisión específicamente dirigida. Por ejemplo, pueden usarse ZFP-TF y ZFN para tratar enfermedades genéticas, infecciones (víricas o bacterianas), para generar estirpes celulares, animales y/o plantas en los que los genes deseados se activan, se reprimen, son diana mediante recombinación homóloga y/o se desactivan o activan. En consecuencia, los enlazadores que se describen en el presente documento también pueden usarse para clonar más eficientemente ADN y en modificaciones genómicas facilitadas por ZFN, que es ampliamente aplicable en muchas áreas de la biotecnología y la ciencia básica.

Ejemplos

Ejemplo 1: Selección de enlazadores

Las selecciones de enlazadores se realizaron en el contexto de tres ZFP hospedadoras diferentes: "ZFP7263", "ZFP7264" y "ZFP8196" (véanse las Publicaciones de Patente de los EE.UU. N.º 20050064474 para 7263 y 7264 y 20080159996 para 8196), que contenía cada una cuatro dedos. Se proporcionan hélices de reconocimiento de cada dedo para cada ZFP hospedadora en la Tabla 1, mientras que la secuencia completa de cada ZFP hospedadora se proporciona en la Figura 1A. Las selecciones se realizaron de la siguiente manera: (i) en primer lugar, se generó una biblioteca dentro de cada ZFP hospedadora que reemplazó los codones en el enlazador central con una mezcla de dos a doce codones completamente aleatorizados (Figura 1B); la secuenciación de bibliotecas no tratadas mostró

una buena diversidad de secuencias sin ningún clon representado más de una vez; (ii) a continuación, las bibliotecas se expresaron en la superficie del bacteriófago filamentosos; (iii) después, las bibliotecas de ZFP expresadas en fagos se seleccionaron para su unión a variantes de diana biotiniladas que contenían una inserción de 1 o 2 pb en el centro del sitio de unión a la proteína hospedadora (es decir, en la región abarcada por el enlazador aleatorizado) (Tabla 2A). Cada inserción comprendía un hueco entre las secuencias de unión para los dedos segundo y tercero de la proteína hospedadora que debe conectarse mediante un enlazador más largo para permitir la unión eficiente (Tabla 2A). Las inserciones consistían en una mezcla de bases con el fin de favorecer la selección de enlazadores sin especificidad de base intrínseca. Se realizaron cinco ciclos de selección. Durante los últimos cuatro ciclos, se empleó una contraselección con un exceso molar de 1000 veces de sitios de unión que no estaban biotinilados y que contenían longitudes de hueco no diana (es decir, si se seleccionaron fagos usando una secuencia diana con una longitud de hueco de 1 pb, la contraselección comprendía dianas con huecos de 0, 2, 3 y 4 pb; si se seleccionaron fagos usando una secuencia diana con una longitud de hueco de 2 pb, la contraselección comprendía dianas con huecos de 0, 1 y 3 y 4 pb; véase la Tabla 2B).

Se detectaron grupos de fagos de la quinta ronda de selección para determinar la capacidad de unir selectivamente secuencias que llevaban la longitud de hueco diana y estos estudios revelaron una unión selectiva de hueco (Figura 2). En particular, los grupos de fagos seleccionados para omitir un hueco de 1 pb en el contexto de ZFP8196 mostraron una preferencia de 25 veces por dianas que llevaban un hueco de 1 pb en comparación con las que no tenían huecos. Los grupos de fagos seleccionados para omitir un hueco de 1 pb en el contexto de ZFP7263 mostraron una preferencia de 26 veces por dianas que llevaban un hueco de 1 pb en comparación con las que no tenían huecos. Los grupos de fagos seleccionados para omitir un hueco de 1 pb en el contexto de ZFP7264 mostraron una preferencia de 5,5 veces por dianas que llevaban un hueco de 1 pb en comparación con las que no tenían huecos. Cada uno de estos grupos también presentó poca o ninguna unión a dianas que llevaban distancias de separación más largas (2, 3 o 4 pb).

Los grupos de fagos seleccionados para omitir un hueco de 2 pb en el contexto de ZFP8196 mostraron una preferencia de 7 veces por dianas que llevaban un hueco de 2 pb en comparación con las de un hueco de 1 pb, así como una preferencia >30 veces por encima de dianas que llevaban huecos de 0,3 y 4 pb.

Tabla 1: Hélices de reconocimiento de ZFP hospedadora

ZFP	Dedo 1	Dedo 2	Dedo 3	Dedo 4
8196	RSDNLSV (SEQ ID NO: 22)	QKINLQV (SEQ ID NO: 23)	RSDVLSE (SEQ ID NO: 24)	QRNHRTT (SEQ ID NO: 25)
7264	RSDLSE (SEQ ID NO: 26)	ARSTRTT (SEQ ID NO: 27)	RSDSLSK (SEQ ID NO: 28)	QRSNLKV (SEQ ID NO: 29)
7263	RSDNLSV (SEQ ID NO: 22)	RNAHRIN (SEQ ID NO: 30)	RSDLSE (SEQ ID NO: 26)	ARSTRTN (SEQ ID NO: 31)

Tabla 2A: Sitios diana utilizados para la selección

ZFP con enlazador aleatorizado	Sitios diana
8196	ATAAACTGdCAAAAGGC (SEQ ID NO:33) ATAAACTGdbCAAAAGGC (SEQ ID NO:34)
7264	TTAAAGCGhGCTCCGAA (SEQ ID NO:38) TTAAAGCGhdGCTCCGAA (SEQ ID NO:39)
7263	CCACTCTGhTGGAAGTG (SEQ ID NO:43) CCACTCTGhhTGGAAGTG (SEQ ID NO:44)

Tabla 2A. Sitios diana utilizados para las selecciones. Los sitios diana de ADN bicatenario utilizados en los estudios de fagos tenían la forma general: TATAAT(X)₁₇₋₁₈TTACAGTCAGTCCACACGTC, (SEQ ID NO: 67) donde (X)₁₇₋₁₈ se reemplazó con las secuencias enumeradas en la tabla. Los dúplex de ADN se preparando prolongando un cebador que se hibridó con la secuencia en cursiva y que estaba biotinilado en su extremo 5'. Las bases subrayadas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las bases en minúsculas indican nucleótidos insertados (o bases "hueco") que deben ser abarcados por los enlazadores seleccionados. Los códigos de degeneración para las bases de hueco son los siguientes: "d" indica una mezcla de A, G y T; "b" indica una mezcla de C, G y T; "h" indica una mezcla de A, C y T; y "v" indica una mezcla de A, C y G.

Tabla 2B: Sitios competidores utilizados durante la selección

ZFP con enlazador aleatorizado	Sitios competidores
8196	<p>ATAAACTGCAAAAAGGC (SEQ ID NO:32)</p> <p>ATAAACTGdCAAAAAGGC (SEQ ID NO:33)</p> <p>ATAAACTGdbCAAAAAGGC (SEQ ID NO:34)</p> <p>ATAAACTGdbbCAAAAAGGC (SEQ ID NO:35)</p> <p>ATAAACTGdbbbCAAAAAGGC (SEQ ID NO:36)</p>
7264	<p>TTAAAGCGGCTCCGAA (SEQ ID NO:37)</p> <p>TTAAAGCGhGCTCCGAA (SEQ ID NO:38)</p> <p>TTAAAGCGhdGCTCCGAA (SEQ ID NO:39)</p> <p>TTAAAGCGhdvGCTCCGAA (SEQ ID NO:40)</p> <p>TTAAAGCGhdvdGCTCCGAA (SEQ ID NO:41)</p>
7263	<p>CCACTCTGTGGAAGTG (SEQ ID NO:42)</p> <p>CCACTCTGhTGAAGTG (SEQ ID NO:43)</p> <p>CCACTCTGhhTGAAGTG (SEQ ID NO:44)</p> <p>CCACTCTGhhhTGAAGTG (SEQ ID NO:45)</p> <p>CCACTCTGhhhbTGAAGTG (SEQ ID NO:46)</p>

5 Tabla 2B. Sitios competidores utilizados durante las selecciones. Los sitios competidores de ADN bicatenario tenían la forma general: TATAAT(X)₁₆₋₂₀TTCACAGTCAGTCCACACGTC, (SEQ ID NO: 67) donde (X)₁₆₋₂₀ se reemplazó con las secuencias enumeradas en la tabla. Los dúplex de ADN se prepararon prolongando un cebador (no biotinilado) que se hibridó con la secuencia en cursiva. Las bases subrayadas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las bases en minúsculas indican nucleótidos insertados (o bases "hueco"). Los códigos de degeneración para las bases de hueco son los siguientes: "d" indica una mezcla de A, G y T; "b" indica una mezcla de C, G y T; "h" indica una mezcla de A, C y T; y "v" indica una mezcla de A, C y G.

Tabla 2C: Dianas utilizadas para estudios de selectividad de hueco de grupo de fagos

ZFP con enlazador aleatorizado	Hueco	Sitios diana
8196	0 huecos	ATAAACTGCAAAAAGGC (SEQ ID NO:32)
	1 hueco	ATAAACTGdCAAAAAGGC (SEQ ID NO:33)
	2 huecos	ATAAACTGdbCAAAAAGGC (SEQ ID NO:34)
	3 huecos	ATAAACTGdbbCAAAAAGGC (SEQ ID NO:35)
	4 huecos	ATAAACTGdbbbCAAAAAGGC (SEQ ID NO:36)
7264	0 huecos	TTAAAGCGGCTCCGAA (SEQ ID NO:37)
	1 hueco	TTAAAGCGhGCTCCGAA (SEQ ID NO:38)
	2 huecos	TTAAAGCGhdGCTCCGAA (SEQ ID NO:39)
	3 huecos	TTAAAGCGhdvGCTCCGAA (SEQ ID NO:40)
	4 huecos	TTAAAGCGhdvdGCTCCGAA (SEQ ID NO:41)
7263	0 huecos	CCACTCTGTGGAAGTG (SEQ ID NO:42)
	1 hueco	CCACTCTGhTGAAGTG (SEQ ID NO:43)
	2 huecos	CCACTCTGhhTGAAGTG (SEQ ID NO:44)
	3 huecos	CCACTCTGhhhTGAAGTG (SEQ ID NO:45)
	4 huecos	CCACTCTGhhhbTGAAGTG (SEQ ID NO:46)

10
15
20 Tabla 2C. Dianas utilizadas para estudios de selectividad de hueco de grupo de fagos. Los sitios de ADN bicatenario utilizados en los estudios de selectividad de hueco de grupo de fagos tenían la forma general: TATAAT(X)₁₆₋₂₀TTCACAGTCAGTCCACACGTC, (SEQ ID NO: 67) donde (X)₁₆₋₂₀ se reemplazó con las secuencias enumeradas en la tabla. Los dúplex de ADN se prepararon prolongando un cebador biotinilado que se hibridó con la secuencia en cursiva. Las bases subrayadas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las bases en minúsculas indican nucleótidos insertados (o bases "hueco"). Los códigos de degeneración para las bases de hueco son los siguientes: "d" indica una mezcla de A, G y T; "b" indica una mezcla de C, G y T; "h" indica una mezcla de A, C y T; y "v" indica una mezcla de A, C y G.

Secuenciación

Se subclonaron y secuenciaron genes que codifican las ZFP seleccionadas. La Figura 3A presenta enlazadores

seleccionados para omitir huecos de 1 pb en el contexto de las tres proteínas hospedadoras, mientras que la Figura 3B muestra enlazadores seleccionados para omitir 2 pb en el contexto del hospedador "ZFP8196". Los resultados de secuenciación revelaron un fuerte sesgo de composición en los enlazadores seleccionados hacia secuencias ricas en prolina y arginina. También se observaron tendencias claras en la longitud de los enlazadores: aunque las bibliotecas iniciales codificaron proporciones aproximadamente iguales de longitudes de 11 enlazadores diferentes (2-12 restos), los enlazadores seleccionados presentaban distribuciones más estrechas de 5-8 restos (para el hueco de 1 pb) o 9-11 restos (para el hueco de 2 pb).

Ejemplo 2: Caracterización inicial de las ZFP seleccionadas

Como evaluación funcional inicial de los enlazadores seleccionados para omitir 1 pb, se subclonaron ZFP que llevaban los enlazadores enumerados en la Figura 3A, se expresaron como proteína libre usando un kit de transcripción-traducción *in vitro* y se evaluaron mediante ELISA para la unión a dianas que llevaban inserciones de 0, 1 o 2 pb opuestas al enlazador seleccionado. Las dianas para estos estudios se enumeran en la Tabla 3. Se generaron nueve proteínas de control adicionales reemplazando el enlazador central de cada ZFP hospedadora con tres secuencias de enlazador alternativas, caracterizadas anteriormente, que colectivamente representaban el estado de la técnica para abarcar 1 pb. Las secuencias de estos enlazadores de control fueron LRQKDERP (SEQ ID NO: 3) (véase, la Patente de los EE.UU. N.º 6.479.626), TGEKGK (SEQ ID NO: 48), TGGGGSQKP (SEQ ID NO: 2). Estas proteínas de control, así como las ZFP hospedadoras, también se incluyeron en los estudios de ELISA.

La Tabla 3 muestra las dianas utilizadas para estudios de ELISA de ZFP seleccionadas para omitir un hueco de 1 pb. Los sitios de ADN bicatenario utilizados en estos estudios tenían la forma general TTAG(X)₁₆₋₁₈TATC, (SEQ ID NO: 94) donde (X)₁₆₋₁₈ se reemplazó con las secuencias enumeradas en la tabla. Cada diana de ADN bicatenario se preparó hibridando un oligonucleótido complementario que llevaba una biotina en su extremo 5'. Los subrayados indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las letras en minúsculas indican nucleótidos insertados (o bases "hueco").

Los resultados de estos estudios se proporcionan en las Tablas 4, 5 y 6, enumerando cada tabla datos para proteínas derivadas de una ZFP hospedadora diferente. La Tabla 4 proporciona datos para proteínas derivadas de ZFP8196; La Tabla 5 proporciona datos para proteínas derivadas de ZFP7263; y la Tabla 6 proporciona datos para proteínas derivadas de ZFP7264. En cada tabla, se enumeran datos de unión para la ZFP hospedadora en la fila superior, seguidos de datos de unión para tres proteínas de control en las filas 2-4, seguidos de datos para las ZFP seleccionadas entre las bibliotecas de presentación de fagos. Los valores se normalizan a la señal de ELISA obtenida de la unión de la ZFP parental a su diana no modificada.

Cada conjunto de proteínas presentó un patrón similar de comportamiento de unión, en tres aspectos clave: En primer lugar, cada ZFP parental se unió bien a su diana no modificada (la diana de "hueco de 0 pb" en las Tablas 4, 5 y 6) pero no a ninguna variante que llevase inserciones de 1 o 2 pb. Esto se esperaba puesto que los enlazadores parentales (TGEKP (SEQ ID NO: 1) (para ZFP8196) o TGSQKP (SEQ ID NO: 72) (para ZFP7263 y ZFP7264) son demasiado cortos para abarcar cualquier base insertada adicional.

En segundo lugar, en casi todos los casos, las proteínas de control se unieron muy mal a dianas con una inserción de 1 pb (los valores de ELISA normalizados fueron de 0,10 o menos para 31 de 36 de dichas mediciones). Esto indica el mal rendimiento de los enlazadores disponibles antes de estos estudios. Además, los enlazadores utilizados por estas proteínas no mostraron una preferencia uniforme por las dianas con una inserción de 1 pb (frente a una inserción de 0 pb).

En tercer lugar, a diferencia del comportamiento de las proteínas de control, las ZFP seleccionadas por fagos se unieron con una afinidad mucho mayor a dianas que llevaban una inserción de 1 pb, así como con un nivel mucho más alto de discriminación frente a las dianas de unión que no contenían base insertada. Estas proteínas también fueron muy selectivas para dianas de unión con una inserción de 1 pb frente a dianas que llevaban una inserción de 2 pb.

Tabla 3: Dianas utilizadas para estudios de ELISA de ZFP seleccionadas para omitir un huecode 1 pb

ZFP con enlazador aleatorizado	Secuencia de hueco	Sitios diana
8196	-	ATAAACTGCAAAAAGGC (SEQ ID NO: 32)
	A	ATAAACTGaCAAAAAGGC (SEQ ID NO: 73)
	C	ATAAACTGcCAAAAAGGC (SEQ ID NO: 74)
	G	ATAAACTGgCAAAAAGGC (SEQ ID NO: 75)
	T	ATAAACTGtCAAAAAGGC (SEQ ID NO: 76)
	TC	ATAAACTGtcCAAAAAGGC (SEQ ID NO: 77)
	AC	ATAAACTGacCAAAAAGGC (SEQ ID NO: 78)
	TG	ATAAACTGtgCAAAAAGGC (SEQ ID NO: 79)

ZFP con enlazador aleatorizado	Secuencia de hueco	Sitios diana
7264	-	TTAAAGCGGCTCCGAA (SEQ ID NO: 37)
	A	TTAAAGCGaGCTCCGAA (SEQ ID NO: 80)
	C	TTAAAGCGcGCTCCGAA (SEQ ID NO: 81)
	G	TTAAAGCGgGCTCCGAA (SEQ ID NO: 82)
	T	TTAAAGCGtGCTCCGAA (SEQ ID NO: 83)
	TT	TTAAAGCGtGCTCCGAA (SEQ ID NO: 84)
	TA	TTAAAGCGtaGCTCCGAA (SEQ ID NO: 85)
	CT	TTAAAGCGctGCTCCGAA (SEQ ID NO: 86)
7263	-	CCACTCTGTGGAAGTG (SEQ ID NO: 42)
	A	CCACTCTGaTGAAGTG (SEQ ID NO: 87)
	C	CCACTCTGcTGAAGTG (SEQ ID NO: 88)
	G	CCACTCTGgTGAAGTG (SEQ ID NO: 89)
	T	CCACTCTGtTGAAGTG (SEQ ID NO: 90)
	AC	CCACTCTGacTGAAGTG (SEQ ID NO: 91)
	AT	CCACTCTGatTGAAGTG (SEQ ID NO: 92)
	CT	CCACTCTGctTGAAGTG (SEQ ID NO: 93)

Tabla 4: Resultados de ELISA para variantes de la ZFP "8196" con diferentes enlazadores centrales

Secuencia del enlazador central	Puntuación de ELISA para la unión a dianas que tienen el hueco indicado [la puntuación se normaliza a 8196 unida a su diana sin hueco (entrada subrayada)]								
	hueco de 0 pb	hueco de 1 pb				relación promedio de 1 pb: puntuación de 0 pb	hueco de 2 pb		
	-	A	C	G	T		TC	AC	TG
TGEKP (SEQ ID NO: 1)	<u>1,00</u>	0,01	0,01	0,03	0,01	0,02	0,00	0,00	0,00
TGGGGSQKP (SEQ ID NO: 2)	0,00	0,00	0,00	0,00	0,00	1,20	0,00	0,00	0,00
LRQKDERP (SEQ ID NO: 3)	0,01	0,01	0,01	0,04	0,08	3,49	0,00	0,00	0,00
TGEGGKP (SEQ ID NO: 48)	0,10	0,00	0,00	0,03	0,03	0,15	0,00	0,00	0,00
TPDAPKPKP (SEQ ID NO: 49)	0,02	0,16	0,13	0,68	0,95	23,75	0,01	0,00	0,01
TPGLHRPKP (SEQ ID NO: 50)	0,04	0,19	0,10	0,65	0,81	10,94	0,01	0,00	0,01
TEPRAKPPKP (SEQ ID NO: 51)	0,01	0,39	0,17	0,78	0,93	70,72	0,02	0,01	0,01
TPSHTPRPKP (SEQ ID NO: 52)	0,02	0,30	0,13	0,84	0,80	25,10	0,02	0,01	0,01
TGYSIPRPKP (SEQ ID NO: 53)	0,01	0,13	0,06	0,43	0,55	44,57	0,01	0,00	0,01
TYPRPIAAKP (SEQ ID NO: 54) (designado 1f)	0,01	0,41	0,14	0,65	0,64	82,25	0,01	0,00	0,01
THPRAPIPKP (SEQ ID NO: 55) (designado 1c)	0,00	0,20	0,09	0,57	0,60	78,86	0,01	0,00	0,00
TPNRRP APKP (SEQ ID NO: 56) (designado 1d)	0,00	0,23	0,09	0,52	0,52	90,27	0,01	0,01	0,01
TSPRLP APKP (SEQ ID NO: 57)	0,01	0,26	0,14	0,62	0,81	67,95	0,01	0,00	0,01
TCPRPPTKPK (SEQ ID NO: 58)	0,00	0,18	0,05	0,48	0,62	70,16	0,01	0,00	0,01
TSSPRSNAPK (SEQ ID NO: 59)	0,01	0,05	0,02	0,20	0,25	20,85	0,01	0,00	0,01
TVSPAPCRSKP (SEQ ID NO: 60)	0,01	0,03	0,01	0,14	0,19	11,52	0,02	0,00	0,01
TPDRPISTCKP (SEQ ID NO: 61)	0,01	0,11	0,05	0,29	0,47	15,41	0,03	0,01	0,02

Tabla 5: Resultados de ELISA para variantes de la ZFP "7263" con diferentes enlazadores centrales

Secuencia del enlazador central	Puntuación de ELISA para la unión a dianas que tienen el hueco indicado [la puntuación se normaliza a 7263 unida a su diana sin hueco (entrada subrayada)]								
	hueco de 0 pb	hueco de 1 pb					hueco de 2 pb		
	-	A	C	G	T	relación promedio de 1 pb: puntuación de 0 pb	AC	AT	CT
TGSQKP (SEQ ID NO: 72)	<u>1,00</u>	0,01	0,01	0,03	0,02	0,02	0,01	0,01	0,00
TGGGGSQKP (SEQ ID NO: 2)	0,51	0,06	0,05	0,41	0,39	0,44	0,01	0,01	0,02
LRQKDERP (SEQ ID NO: 3)	0,25	0,03	0,02	0,18	0,13	0,36	0,01	0,01	0,01
TGEGGKP (SEQ ID NO: 48)	1,30	0,02	0,02	0,05	0,04	0,03	0,01	0,01	0,01
TPRPPPKP (SEQ ID NO: 4)	0,14	0,97	0,67	1,85	2,09	10,20	0,02	0,01	0,01
TQRPQIPPKP (SEQ ID NO: 62)	0,15	1,66	1,00	2,86	3,05	14,68	0,03	0,02	0,01
TPNRCPPTKP (SEQ ID NO: 63)	0,31	1,68	1,13	2,62	3,16	7,53	0,03	0,02	0,01
TYPRPLL (SEQ ID NO: 7)	0,29	1,95	1,27	3,88	3,97	10,08	0,03	0,01	0,01
TPLCQRPMKQK P (SEQ ID NO: 8)	0,28	1,82	1,28	3,44	4,00	10,88	0,08	0,05	0,02

Tabla 6: Resultados de ELISA para variantes de la ZFP "7264" con diferentes enlazadores centrales

Secuencia del enlazador central	Puntuación de ELISA para la unión a dianas que tienen el hueco indicado [la puntuación se normaliza a 7264 unida a su diana sin hueco (entrada subrayada)]								
	hueco de 0 pb	hueco de 1 pb					hueco de 2 pb		
	-	A	C	G	T	relación promedio de 1 pb: puntuación de 0 pb	TT	TA	CT
TGSQKP (SEQ ID NO: 72)	<u>1,00</u>	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
TGGGGSQKP (SEQ ID NO: 2)	0,46	0,07	0,04	0,08	0,17	0,19	0,03	0,03	0,07
LRQKDERP (SEQ ID NO: 3)	0,26	0,05	0,03	0,06	0,10	0,22	0,02	0,02	0,02
TGEGGKP (SEQ ID NO: 48)	1,39	0,02	0,03	0,05	0,08	0,03	0,03	0,02	0,03
TGLPKPKP (SEQ ID NO: 64)	0,14	0,19	0,11	0,38	0,86	2,84	0,03	0,02	0,02
TSRPRPKP (SEQ ID NO: 11)	0,18	0,52	0,22	0,77	2,07	4,93	0,03	0,03	0,03
TLPLPRPKP (SEQ ID NO: 65)	0,25	0,58	0,25	0,85	1,36	3,01	0,04	0,03	0,03
TVPRTPPKP (SEQ ID NO: 12) (designado 1e)	0,16	2,35	1,02	1,58	2,55	11,71	0,05	0,05	0,06
TLPPCFRPKP (SEQ ID NO: 66)	0,36	0,72	0,25	0,77	2,72	3,11	0,06	0,06	0,05
TKHGTPKHREDK P (SEQ ID NO: 13)	0,01	0,01	0,01	0,01	0,01	0,79	0,00	0,00	0,00

5 Para respaldar y ampliar los resultados obtenidos en los estudios de ELISA, se evaluaron ZFP con enlazadores seleccionados para la determinar la actividad de escisión *in vivo* en diversos sitios diana usando el ensayo de detección de levadura descrito en la Publicación de Patente Internacional WO 2009/042163. Como estos experimentos requieren más mano de obra, se realizaron con menos ZFP. Para estos estudios, se ensamblaron seis ZFP de la Tabla 5 y cinco ZFP de la Tabla 6 en construcciones que permitieron la expresión como nucleasas de dedos de cinc (ZFN) como se describe en el documento WO 2007/139982. La actividad *in vivo* se midió después mediante la evaluación de la secreción de MEL-1 de cepas de levadura que tenían diversos sitios diana. Las secuencias diana utilizadas para estos estudios se proporcionan en las Tablas 7 y 8, e incluyen variaciones de los sitios de unión 7263 y 7264 con inserciones centrales de 0, 1- o 2 pb.

10

Tabla 7: Dianas utilizadas para el ensayo de detección de levaduras de ZFP seleccionadas para omitir un hueco de 1 pb en ZFP7263

ZFP con enlazador aleatorizado	Secuencia de hueco	Sitios diana
7263	-	ACTCTGTGGAAG (SEQ ID NO:95)
	A	ACTCTGaTGGAAG (SEQ ID NO:96)
	C	ACTCTGcTGGAAG (SEQ ID NO:97)
	G	ACTCTGgTGGAAG (SEQ ID NO:98)
	T	ACTCTGtTGGAAG (SEQ ID NO:99)
	AC	ACTCTGacTGGAAG (SEQ ID NO:100)
	AT	ACTCTGatTGGAAG (SEQ ID NO:101)
	CT	ACTCTGctTGGAAG (SEQ ID NO:102)

Se construyeron plásmidos indicadores que llevaban sitios diana de nucleasa esencialmente como se describe en la Publicación de Patente Internacional WO 2009/042163, excepto por que los casetes de diana de nucleasa tenían la forma general de GATCTGTT*CGGAGCCGCTTTA*ACCC(X)₁₇₋₁₄TGCTCGCG (SEQ ID NO: 103) donde (1) las cuatro bases subrayadas en cada extremo representan los salientes utilizados para la clonación en el plásmido indicador digerido con BamHI/BssHII, (2) la secuencia en cursiva representa el sitio de unión para la ZFN 7264 que se une a la cadena antisentido y que era invariante para estas detecciones, y (3) se reemplazó (X)₁₂₋₁₄ con las secuencias enumeradas en la tabla. Las bases en mayúsculas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las letras en minúsculas indican nucleótidos insertados (o bases "hueco").

Tabla 8: Dianas utilizadas para el ensayo de detección de levaduras de ZFP seleccionadas para omitir un hueco de 1 pb en ZFP7264

ZFP con enlazador aleatorizado	Secuencia de hueco	Sitios diana
7264	-	AAAGCGGCTCCG (SEQ ID NO:104)
	A	AAAGCGaGCTCCG (SEQ ID NO:105)
	C	AAAGCGcGCTCCG (SEQ ID NO:106)
	G	AAAGCGgGCTCCG (SEQ ID NO:107)
	T	AAAGCGtGCTCCG (SEQ ID NO:108)
	TT	AAAGCGttGCTCCG (SEQ ID NO:109)
	TA	AAAGCGtaGCTCCG (SEQ ID NO:110)
	CT	AAAGCGctGCTCCG (SEQ ID NO:111)

Se construyeron plásmidos indicadores que llevaban sitios diana de nucleasa esencialmente como se describe en la Publicación de Patente Internacional WO 2009/042163, excepto por que los casetes diana de nucleasa tenían la forma general de GATCTGTT(X)₁₂₋₁₄AACCCACTCTGTGGAAGTGCTCGCG (SEQ ID NO: 112) donde (1) las cuatro bases subrayadas en cada extremo representan los salientes utilizados para la clonación en el plásmido indicador digerido con BamHI/BssHII, (2) la secuencia en cursiva representa el sitio de unión para la ZFN 7263 que era invariable para estas detecciones, y (3) se reemplazó (X)₁₂₋₁₄ con las secuencias enumeradas en la tabla. Las bases en mayúsculas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las letras en minúsculas indican nucleótidos insertados (o bases "hueco"). Téngase en cuenta que los sitios diana enumerados en la tabla son el complemento inverso de lo que está presente en el casete diana cuando la ZFN 7264 se une a la cadena antisentido.

Los datos para estos experimentos se muestran en las Tablas 9 y 10, enumerando cada tabla datos para proteínas derivadas de una ZFN hospedadora diferente. La Tabla 9 proporciona datos para ZFN derivadas de 7263 y la Tabla 10 proporciona datos para ZFN derivadas de 7264. En cada tabla, se enumeran datos de actividad de nucleasa para la ZFN hospedadora en la fila superior, seguidos de datos de actividad de nucleasa para una proteína de control en la fila 2, seguidos de datos para las ZFP seleccionadas entre las bibliotecas de presentación de fagos. Puesto que ZFP7263 y ZFP7264 son dos mitades del mismo dímero de nucleasa de dedos de cinc, los datos para la ZFN hospedadora son los mismos en cada tabla. Los resultados de estos estudios coincidieron ampliamente con los patrones observados en los estudios de ELISA, en que las ZFN que llevaban enlazadores seleccionados por fagos mostraron tanto una mayor actividad como una mejor preferencia por dianas que llevaban una inserción de 1 pb que las ZFP que llevaban enlazadores de control.

Tabla 9: Resultados de detección de levadura para variantes de ZFP7263 con diferentes enlazadores centrales

Secuencia del enlazador central	Puntuación de MEL-1 para la actividad nucleasa en dianas que tienen el hueco indicado [la puntuación se normaliza a ZFP7263 unida a su diana sin hueco (entrada subrayada)]								
	hueco de 0 pb	hueco de 1 pb					hueco de 2 pb		
	-	A	C	G	T	relación promedio de puntuación de 1 pb:0 pb	AC	AT	CT
TGSQKP (SEQ ID NO: 72)	<u>1,00</u>	0,05	0,05	0,09	0,06	0,06	0,09	0,02	0,08
TGGGGSQKP (SEQ ID NO: 2)	0,11	0,14	0,04	0,60	0,31	2,57	0,04	0,02	0,04
TPRPPIPKP (SEQ ID NO: 4)	0,04	1,30	1,41	2,18	1,40	35,90	0,05	0,03	0,04
TQRPQIPPKP (SEQ ID NO: 62)	0,04	1,05	0,43	2,18	1,19	34,18	0,03	0,02	0,03
TPNRCPTKP (SEQ ID NO: 63)	0,05	1,30	0,34	2,85	1,59	33,22	0,04	0,03	0,06
TYPRLLP (SEQ ID NO: 7)	0,05	0,69	0,37	2,21	1,06	20,15	0,10	0,05	0,05
TPLCQRPMKQK P (SEQ ID NO: 8)	0,04	0,97	0,30	1,52	1,19	27,52	1,05	0,01	0,03

Tabla 10: Resultados de detección de levadura para variantes de ZFP7264 con diferentes enlazadores centrales

Secuencia del enlazador central	Puntuación de MEL-1 para la actividad nucleasa en dianas que tienen el hueco indicado [la puntuación se normaliza a ZFP7264 unida a su diana sin hueco (entrada subrayada)]								
	hueco de 0 pb	hueco de 1 pb					hueco de 2 pb		
	-	A	C	G	T	relación promedio de puntuación de 1 pb:0 pb	TT	TA	CT
TGSQKP (SEQ ID NO: 72)	<u>1,00</u>	0,05	0,05	0,09	0,06	0,06	0,09	0,02	0,08
TGGGGSQKP (SEQ ID NO: 2)	0,28	0,08	0,07	0,14	0,34	0,57	0,12	0,09	0,15
TGLPKPKP (SEQ ID NO: 64)	0,04	0,08	0,07	0,42	0,79	8,03	0,11	0,08	0,09
TSRPRPKP (SEQ ID NO: 11)	0,08	0,18	0,08	0,61	3,38	13,33	0,03	0,07	0,10
TVPRPTPPKP (SEQ ID NO: 12) (designado 1e)	0,08	1,24	0,14	1,40	2,34	20,28	0,12	0,09	0,11

Puesto que el ELISA estuvo en estrecha concordancia con los datos de detección de levadura, los inventores eligieron un conjunto de ejemplos de enlazadores de omisión de 1 pb que se realizaron entre los mejores en el ensayo de ELISA. Éstos se enumeran en la Tabla 11 y se designan "1c", "1d", "1e" y "1f". Estas designaciones también se incluyen en los datos presentados en las Tablas 4, 6 y 10. Los datos de ELISA para estos enlazadores de ejemplo también se muestran en las Figuras 4 y 5.

Tabla 11: Diseños de enlazadores de ejemplo

Secuencia de enlazador	Designación de enlazador	Número de bases omitidas
THPRAPIPKP (SEQ ID NO: 55)	1c	1
TPNRRPAPKP (SEQ ID NO: 56)	1d	1
TVPRPTPPKP (SEQ ID NO: 12)	1e	1
TYPRPIAAKP (SEQ ID NO: 54)	1f	1

Como se ha indicado anteriormente en el Ejemplo 1, los sitios diana utilizados para la selección (Tabla 2a) contenían bases degeneradas en el hueco con el fin de favorecer la selección de enlazadores que no presentaban ninguna preferencia inherente por secuencias de hueco particulares. Los datos que se muestran en las figuras 4 y 5 sugieren que esta estrategia de selección fue satisfactoria: las proteínas que llevan los enlazadores de ejemplo presentan poca variación en la unión entre dianas con bases de hueco de A, G, C o T. Además, la cantidad de variación menor que se observa se refleja en los resultados obtenidos con los enlazadores flexibles de control (véase, por ejemplo, la Figura 5D para el enlazador flexible (TGGGGSQKP) (SEQ ID NO: 2)) lo que indica que la variación es una propiedad de los dedos flanqueantes.

Se realizó un experimento de ELISA análogo para los enlazadores seleccionados para omitir un hueco de 2 pares de

bases en el contexto de ZFP8196. Se generaron dos proteínas de control adicionales reemplazando el enlazador central de cada ZFP hospedadora con dos secuencias de enlazador alternativas, caracterizadas anteriormente, que representaban colectivamente el estado de la técnica para abarcar 2 pb. Las secuencias de estos enlazadores de control eran TGGGGSGGSQKP (SEQ ID NO: 14) y LRQKDGGGSERP (SEQ ID NO: 68). Estas proteínas de control, así como las ZFP hospedadoras, también se incluyeron en los estudios de ELISA. Se sometieron a ensayo ZFP para determinar la unión a sitios diana que no contenían hueco, cada uno de los 4 huecos posibles de 1 par de bases y cada uno de los 16 huecos posibles de 2 pares de bases. Los sitios diana se enumeran en la Tabla 12. Las puntuaciones de ELISA se normalizaron a la puntuación de la ZFP 8196 unida a su sitio diana sin huecos. Se eligieron enlazadores de ejemplo basándose en su comportamiento de ELISA de manera similar a los enlazadores de omisión de 1 pb. Estos enlazadores de ejemplo se designaron "2d", "2e" y "2f". Los resultados de ELISA para los enlazadores de ejemplo en la ZFP 8196 y las ZFP de control se muestran en la Figura 6. Como se observa en la Figura 6, ninguna de las ZFP que llevan los enlazadores de control muestra una preferencia por un sitio diana con un hueco de 2 pares de bases. Sin embargo, las ZFP con los enlazadores seleccionados mostraron preferencias claras por un hueco de 2 pares de bases sobre los sitios diana de 1 par de bases y sin hueco.

Tabla 12: Sitios diana utilizados para la caracterización de ELISA de enlazadores de omisión de 2 pb

ZFP con enlazador aleatorizado	Secuencia de hueco	Sitios diana
8196	-	ATAAACTGCAAAAAGGC (SEQ ID NO:32)
	A	ATAAACTG <u>a</u> CAAAAAGGC (SEQ ID NO:73)
	C	ATAAACTG <u>c</u> CAAAAAGGC (SEQ ID NO:74)
	G	ATAAACTG <u>g</u> CAAAAAGGC (SEQ ID NO:75)
	T	ATAAACTG <u>t</u> CAAAAAGGC (SEQ ID NO:76)
	AA	ATAAACTG <u>aa</u> CAAAAAGGC (SEQ ID NO:113)
	AC	ATAAACTG <u>ac</u> CAAAAAGGC (SEQ ID NO:114)
	AG	ATAAACTG <u>ag</u> CAAAAAGGC (SEQ ID NO:115)
	AT	ATAAACTG <u>at</u> CAAAAAGGC (SEQ ID NO:116)
	CA	ATAAACTG <u>ca</u> CAAAAAGGC (SEQ ID NO:117)
	CC	ATAAACTG <u>cc</u> CAAAAAGGC (SEQ ID NO:118)
	CG	ATAAACTG <u>cg</u> CAAAAAGGC (SEQ ID NO:119)
	CT	ATAAACTG <u>ct</u> CAAAAAGGC (SEQ ID NO:120)
	GA	ATAAACTG <u>ga</u> CAAAAAGGC (SEQ ID NO:121)
	GC	ATAAACTG <u>gc</u> CAAAAAGGC (SEQ ID NO:122)
	GG	ATAAACTG <u>gg</u> CAAAAAGGC (SEQ ID NO:123)
	GT	ATAAACTG <u>gt</u> CAAAAAGGC (SEQ ID NO:124)
	TA	ATAAACTG <u>ta</u> CAAAAAGGC (SEQ ID NO:125)
	TC	ATAAACTG <u>tc</u> CAAAAAGGC (SEQ ID NO:126)
	TG	ATAAACTG <u>tg</u> CAAAAAGGC (SEQ ID NO:127)
	TT	ATAAACTG <u>tt</u> CAAAAAGGC (SEQ ID NO:128)

Tabla 12: Los sitios diana de ADN bicatenario utilizados en los estudios de caracterización de ELISA tenían la forma general: TTAG(X)₁₆₋₁₈TATC, (SEQ ID NO: 94) donde (X)₁₆₋₁₈ se reemplazó con las secuencias enumeradas en la tabla. Los dúplex de ADN se prepararon hibridando oligonucleótidos complementarios. Los oligonucleótidos complementarios a las secuencias enumeradas en la tabla contenían una biotina 5'. Las bases subrayadas indican las secuencias de unión para los cuatro dedos de cada ZFP hospedadora, mientras que las bases en minúsculas indican nucleótidos insertados (o bases "hueco").

Como se ha indicado anteriormente en el Ejemplo 1, los sitios diana utilizados para la selección (Tabla 2a) contenían bases degeneradas en el hueco con el fin de favorecer la selección de enlazadores que no presentaban ninguna preferencia inherente por secuencias de hueco particulares. En la Figura 6 se muestra una ampliación de la escala para uno de los enlazadores flexibles (TGGGGSGGSQKP (SEQ ID NO: 14)). Este enlazador flexible no debería tener ninguna interacción con el sitio diana y, por tanto, el patrón observado probablemente se deba a la unión de las proteínas de dedos de cinc. El hecho de que los enlazadores de ejemplo muestren un patrón similar de unión a sitios diana de hueco de 2 pares de bases sugiere que los enlazadores seleccionados tampoco deberían imponer ningún sesgo de composición de hueco en la unión de ZFP.

Un resumen más conciso de estos datos se presenta en la Tabla 13, donde las puntuaciones de ELISA se promediaron en todas las composiciones de hueco de 1 o 2 pares de bases. También se informa la preferencia en número de veces por un hueco de 2 pares de bases sobre el hueco de 1 par de bases y la diana sin hueco. El enlazador más selectivo (TPNPHRRTPSHKP (SEQ ID NO: 69), "2f") representa una mejora en la selectividad de hueco de 2 pares de bases de >100 veces sobre un hueco de cero de pares de bases y >20 veces sobre un hueco de 1 par de bases en comparación con los enlazadores de control.

Tabla 13: Resumen de los datos de ELISA para enlazadores de omisión de 2 pb

Secuencia de enlazador	Designación	Puntuación normalizada de ELISA promedio (hueco)			Selectividad de hueco de 2 pb frente a:	
		0bp	1bp	2bp	0bp	1bp
TGGGGSGGSQKP (SEQ ID NO: 14)	flexible	0,019	0,041	0,010	0,6	0,2
LRQKDGGGSERP (SEQ ID NO: 68)	Kim, Pabo	0,010	0,047	0,003	0,4	0,1
TPNPHRRTPSHKP (SEQ ID NO: 69)	2f	0,003	0,046	0,219	64,6	4,7
TLAPRPYRPPKP (SEQ ID NO: 70)	2d	0,005	0,035	0,127	24,4	3,6
TPGGKSSRTDRNKP (SEQ ID NO: 71)	2e	0,005	0,099	0,100	22,0	1,0

Ejemplo 3: Caracterización de ELISA de enlazadores en diversas ZFP hospedadoras

5 Para demostrar la generalidad de los enlazadores de ejemplo, los cuatro enlazadores de omisión de 1 pb enumerados en la Tabla 11 (1c-1f) se clonaron en doce ZFP hospedadoras diferentes. Las ZFP hospedadoras se designaron ZFP1, ZFP2 etc. Las proteínas resultantes se expresaron a través de transcripción y traducción *in vitro* y se sometieron a ensayo a través de ELISA, como se ha descrito anteriormente. A efectos comparativos, los inventores también sometieron a ensayo las ZFP hospedadoras con un enlazador flexible (TGGGGSQKP (SEQ ID NO: 2)) y los resultados se presentan en la Figura 7. Estos datos demuestran que, con respecto a un enlazador flexible convencional, los nuevos enlazadores aumentaron significativamente la puntuación de ELISA de la mayoría de las ZFP hospedadoras, con la única excepción de las ZFP que saturan el ensayo (ZFP1 y ZFP2) o cuya unión es indetectablemente baja (ZFP 11 y ZFP 12). Los aumentos promedio en número de veces de la puntuación de ELISA en todas las ZFP hospedadoras fueron de 3-5.

15 En un estudio similar, los tres enlazadores de ejemplo seleccionados para omitir un hueco de 2 pb enumerados en la Tabla 13 (2d-2f) se sometieron a ensayo en seis ZFP hospedadoras diferentes como se han descrito anteriormente (ZFP13, ZFP14, etc.) y estos resultados se presentan en la Figura 8. En estos experimentos, las mejoras promedio en número de veces en la puntuación de ELISA en todas las ZFP hospedadoras variaron de 1,9 a 2,4.

Ejemplo 4: Caracterización de ZFN con enlazadores de ejemplo en loci endógenos en células de mamífero

25 Después se sometieron a ensayo ZFN para determinar su capacidad para inducir roturas bicatenarias en loci endógenos. Brevemente, se emparejaron plásmidos que codificaban las 18 ZFN (fusiones ZFP-*FokI*) descritas anteriormente (Ejemplo 3) con sus ZFN compañeras apropiadas y se introdujeron en células K562 mediante transfección usando el kit Amaxa™ Nucleofection como especifica el fabricante. Para determinar la actividad de ZFN en el locus diana medida mediante el nivel de unión de extremos no homólogos (NHEJ), se realizaron ensayos de falta de coincidencia de CEL-I esencialmente según las instrucciones del fabricante (Transgenomic SURVEYOR™). Se recogieron células y se preparó ADN cromosómico usando un kit Quickextract™ de acuerdo con las instrucciones del fabricante (Epicentre®). La región apropiada del locus diana se amplificó por PCR usando ADN polimerasa de alta fidelidad Taq Accuprime™ (Invitrogen) seguida de tratamiento con la enzima CEL-I.

35 Los geles de ejemplo generados para el ensayo de CEL-1 se muestran en la Figura 9. La Figura 9A muestra datos de detección de ZFN3 y ZFN4 (ZFN que omiten 1 par de bases) como ZFN hospedadora, mientras que la Figura 9B muestra los datos de detección de ZFN14 (ZFN que omite 2 pares de bases) como ZFN hospedadora. Los datos para todos los enlazadores de omisión de 1 pb de ejemplo se resumen en la Figura 10 (ZFN1-ZFN12). Algunas de las ZFN se expresaron usando una condición de expresión alta. La expresión alta se obtiene después de la transfección mediante incubación de células a 37 °C durante 24 horas y después incubando a 30 °C durante 48 horas antes de aislar el ADN genómico. Las ZFN que utilizan estas condiciones se resaltan en la Figura 10. En concreto, tres ZFN que eran inactivas con el enlazador TGGGGSQKP (SEQ ID NO: 2) ("enlazador flexible") (ZFN 4, 9 y 10) se vuelven activas cuando se usa un enlazador como se describe en el presente documento. Para estos casos, se asignó un valor de modificación del 1,0 % al enlazador flexible con fines de normalización (el límite de detección del ensayo). En el 85 % de las ZFN sometidas a ensayo con los nuevos enlazadores, se observó un aumento en el nivel de modificación génica, con un aumento promedio de aproximadamente 1,8-2,8 veces en los nueve pares de ZFN activas.

50 De manera similar, las ZFN descritas anteriormente (Ejemplo 3) que contenían los enlazadores de ejemplo de omisión de 2 pb (ZFN13-ZFN18) se sometieron a ensayo en loci endógenos y los resultados se resumen en la Figura 11. En este estudio, la sustitución de los enlazadores que se describen en el presente documento mejoró la actividad en comparación con el enlazador flexible para 3 de las 4 ZFN activas y la mejora promedio fue de 1,5-2 veces en todos los pares de ZFN activas.

Ejemplo 5: Selecciones secundarias para un enlazador de omisión de 2 pb

5 Se construyó un conjunto secundario de bibliotecas basándose en la información obtenida de las selecciones iniciales para un enlazador de omisión de 2 pb (Ejemplo 3 y Figura 3). Estas bibliotecas fijaron los tres restos carboxi-terminales del enlazador como RPP (lisina, prolina, prolina) y aleatorizaron los restos amino-terminales restantes. El diseño de la biblioteca se muestra en la Figura 12.

10 Las selecciones se realizaron de la misma manera que en el Ejemplo 1 usando ZFP8196 como proteína hospedadora. La selectividad de hueco del grupo de fagos seleccionado se muestra en la Figura 13A y las secuencias de los enlazadores de clones individuales se muestran en la Figura 13B.

15 Se realizó un experimento de ELISA en cada uno de los clones individuales de la selección secundaria (Figura 13B), similar al del Ejemplo 2. Se sometieron a ensayo ZFP para determinar la unión a sitios diana que no contenían hueco, un grupo de los 4 huecos posibles de 1 par de bases y un grupo de los 16 huecos posibles de 2 pares de bases. Los sitios diana se enumeran en la Tabla 12. Las puntuaciones de ELISA se normalizaron a la puntuación de la ZFP8196 hospedadora unida a su sitio diana sin hueco. Los resultados de ELISA para las ZFP que llevaban enlazadores que mostraron una buena puntuación de ELISA normalizada en el grupo de sitios diana de hueco de 2 pb y una buena selectividad de hueco se muestran en la Tabla 14.

Tabla 14: Resumen de los datos de ELISA para enlazadores de omisión de 2 pb

Secuencia de enlazador	Puntuación normalizada de ELISA promedio (hueco)			Selectividad de hueco de 2 pb frente a:	
	Obp	1bp	2bp	0bp	1bp
TETTRPFRPPKP (SEQ ID NO: 183)	0,001	0,001	0,570	570,0	570,0
TGSLRPYRRPKP (SEQ ID NO: 177)	0,001	0,010	0,310	310,0	31,0
TSINRPFRRPKP (SEQ ID NO: 184)	0,010	0,020	0,570	57,0	28,5
TNTTRPYRPPKP (SEQ ID NO: 175)	0,001	0,010	0,410	410,0	41,0
TASCPRPFRPPKP (SEQ ID NO: 194)	0,010	0,020	0,370	37,0	18,5
TGEARPYRPPKP (SEQ ID NO: 178)	0,001	0,010	0,610	610,0	61,0

25 Como se muestra, las ZFP con los enlazadores seleccionados mostraron preferencias claras por un hueco de 2 pares de bases sobre los sitios diana de 1 par de bases y los que no tienen hueco.

REIVINDICACIONES

1. Una proteína de dedos de cinc de múltiples dedos que se une específicamente a un sitio diana, comprendiendo la proteína de dedos de cinc de múltiples dedos módulos de dedos de cinc de origen no natural, en donde cada módulo de dedos de cinc se une a un subsitio diana y al menos dos de los módulos de unión a ADN de dedos de cinc de origen no natural que se unen a subsitios diana separados por 1 o 2 pares de bases se unen mediante una secuencia de enlazador de aminoácidos de 5 a 20 restos de aminoácidos entre el último resto del módulo de dedos de cinc N-terminal y el primer resto del módulo de dedos de cinc C-terminal, comprendiendo el enlazador de aminoácidos un resto de aminoácido N-terminal adyacente al módulo de dedos de cinc N-terminal, un resto de aminoácido C-terminal adyacente al módulo de dedos de cinc C-terminal y restos de aminoácidos internos a los restos de enlazador de aminoácidos N y C-terminales, en donde dicha secuencia de enlazador se selecciona entre el grupo que consiste en: TPDAPKPKP, TPGLHRPKP, TEPRAKPPKP, TPSHTPRPKP, TGYSIPRPKP, TYPRPIAAKP, THPRAPIPKP, TPNRRPAPKP, TSPRLPAPKP, TCPRPPTTRKP, TSSPRSNAKP, TVSPAPCRSKP, TPDRPISTCKP, TPRPPIPKP, TQRQIIPPKP, TPNRCPPTKP, TYPRLLAKP, TPLCQRPMKQKP, TGLPKPKP, TSRPRPKP, TLPLPRPKP, TVPRPTPPKP, TLPPCFRPKP, cuando los subsitios diana están separados por 1 par de bases, y TLAPRPYRPPKP, TPNPHRRTDPSHKP, TPGGKSSRTDRNKP, TNTRPYRPPKP, TGSLRPYRPPKP, TGEARPYRPPKP, TETTRPFRPPKP, TSINRPFRRPKP, TASCPRPFRPPKP cuando los subsitios diana están separados por 2 pares de bases.
2. La proteína de dedos de cinc de múltiples dedos de la reivindicación 1, en donde el enlazador comprende adicionalmente al menos dos restos de aminoácidos básicos.
3. La proteína de dedos de cinc de múltiples dedos de la reivindicación 2, en donde los restos de aminoácidos básicos se seleccionan entre el grupo que consiste en restos de arginina, restos de histidina, restos de lisina o combinaciones de los mismos.
4. Una proteína de fusión que comprende la proteína de dedos de cinc de múltiples dedos de cualquiera de las reivindicaciones 1 a 3 y un dominio regulador.
5. La proteína de fusión de la reivindicación 4, en donde el dominio regulador es un dominio modulador de la transcripción.
6. La proteína de fusión de la reivindicación 5, en donde el dominio regulador es un dominio de activación o un dominio de represión.
7. La proteína de fusión de la reivindicación 4, en donde el dominio regulador es un dominio de escisión o semidominio de escisión.
8. Un polinucleótido que codifica cualquiera de las proteínas de las reivindicaciones 1 a 7.
9. Una célula que comprende cualquiera de las proteínas de las reivindicaciones 1 a 7 y/o un polinucleótido de acuerdo con la reivindicación 8.
10. Un método para modular la expresión de un gen en una célula *in vitro*, comprendiendo el método administrar una proteína de acuerdo con cualquiera de las reivindicaciones 1 a 7 o un polinucleótido de acuerdo con la reivindicación 8 a la célula.
11. El método de la reivindicación 10, en donde la modulación comprende la regulación positiva o negativa del gen.
12. El método de la reivindicación 10, en donde la modulación comprende la escisión del gen.
13. El método de cualquiera de las reivindicaciones 10 a 12, en donde el gen es un gen celular endógeno.
14. La proteína de acuerdo con una cualquiera de las reivindicaciones 1-7 o el polinucleótido de acuerdo con la reivindicación 8 para su uso en terapia.

Figura 1A

ZFP hospedadoras utilizadas para la selección

ZFP 8196	F1	MAERPFQCRICMRNFSRSDNLSVHIRTHTGE	-1	+6
	F2	<u>KPFACD</u> ICGRKFAQKINLQVHTKIHTGE		
	F3	<u>KPFQCR</u> ICMRNFSRSDVLSEHIRTHTGE		
	F4	<u>KPFACD</u> ICGRKFAQRNHRTHTKIHLRGS		
ZFP 7263	F1	MAERPFQCRICMRNFSRSDNLSVHIRTHTGE	-1	+6
	F2	<u>KPFACD</u> ICGRKFAARNNAHRINHHTKIHTGSQ		
	F3	<u>KPFQCR</u> ICMRNFSRSDDTSEHIRTHTGE		
	F4	<u>KPFACD</u> ICGRKFAARSTRTNHTKIHLRGS		
ZFP 7264	F1	MAERPFQCRICMRNFSRSDTLSEHIRTHTGE	-1	+6
	F2	<u>KPFACD</u> ICGRKFAARSTRTHTKIHTGSQ		
	F3	<u>KPFQCR</u> ICMRNFSRSDSLSKHIRTHTGE		
	F4	<u>KPFACD</u> ICGRKFAQRSNLKVHTKIHLRGS		

↑
Hélice de reconocimiento

Figura 1B Diseños de bibliotecas de enlazadores utilizados para la selección

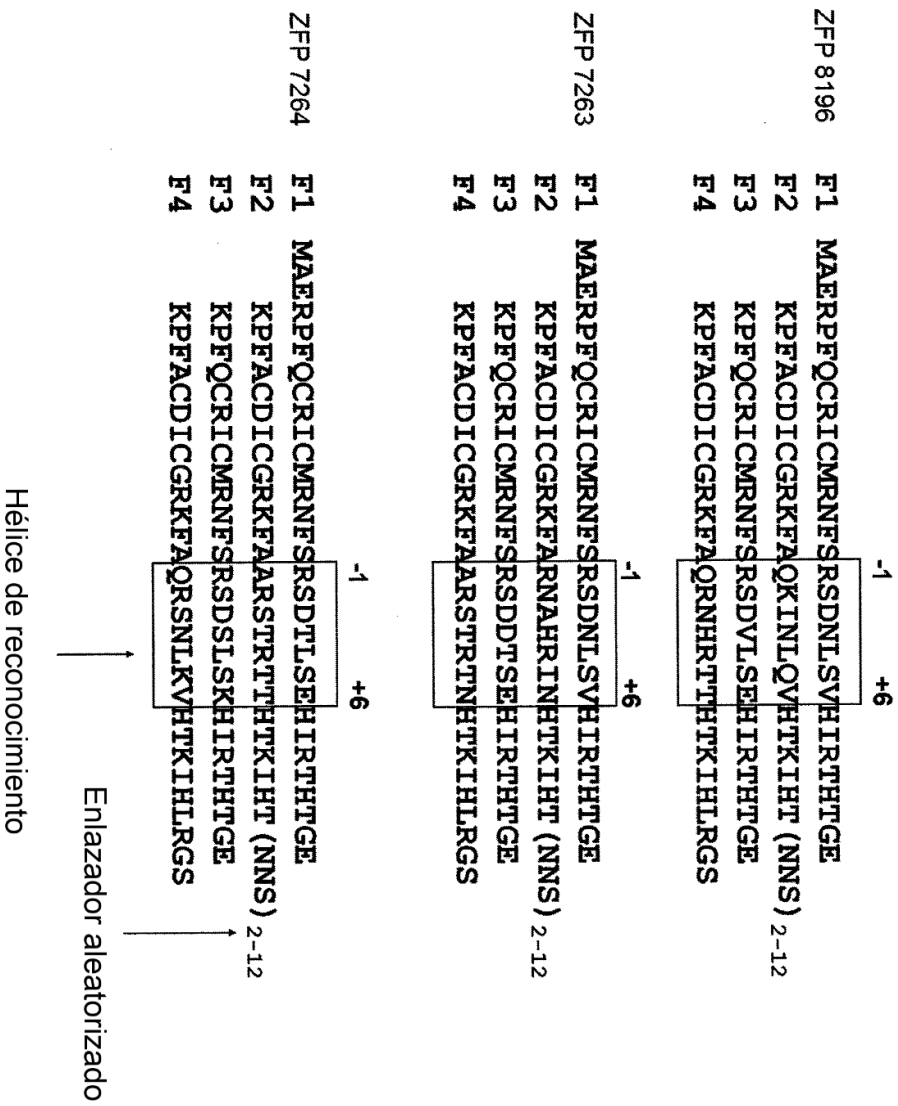


Figura 2: Selectividad de hueco de grupos de fagos seleccionados

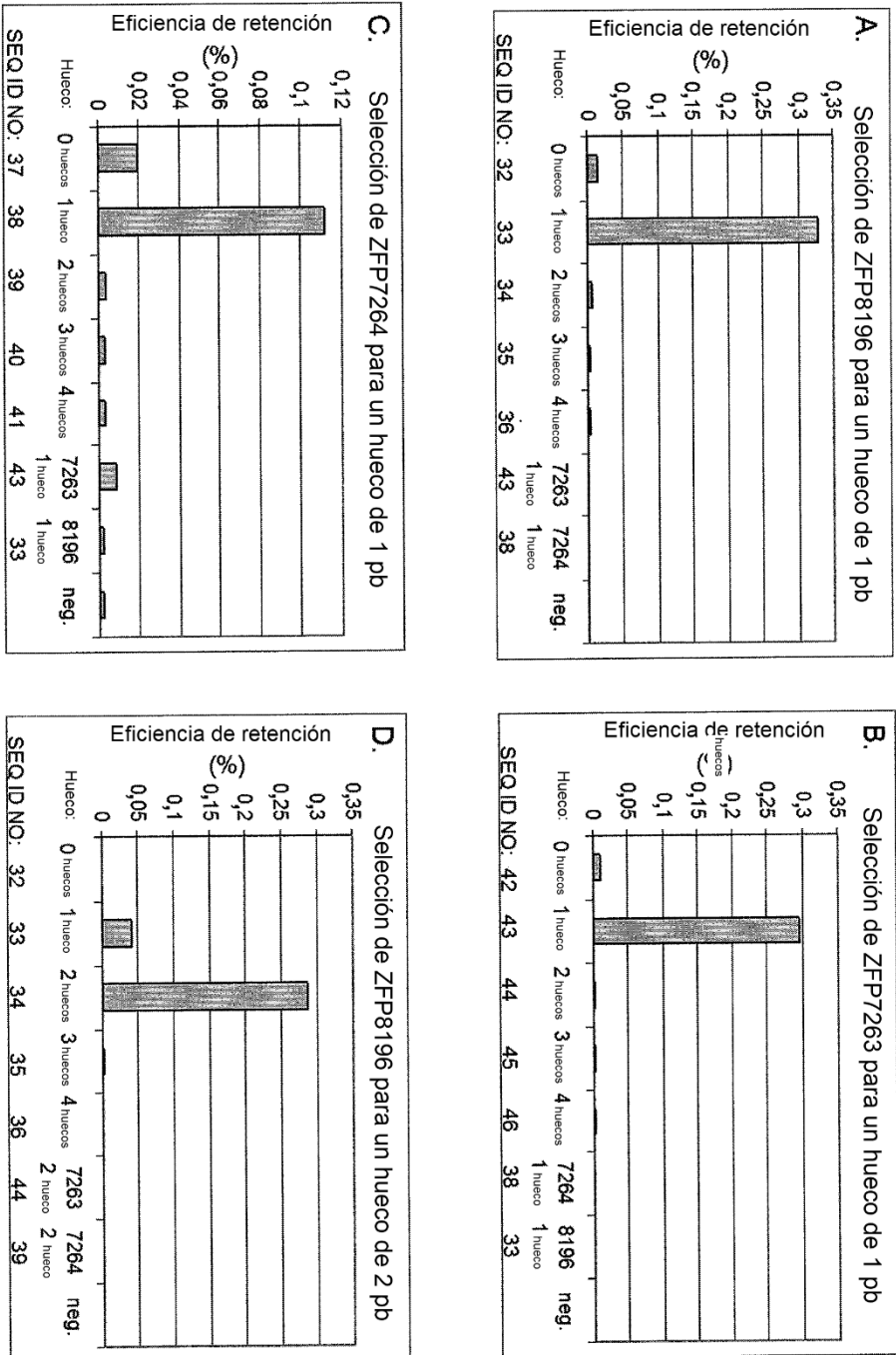


Figura 3

A.

ZFP8196			ZFP7263			ZFP7264					
n.º de clones			n.º de clones			n.º de clones					
1	HTK IHTL	PDAPIKRP	KPFQCRIC	1	HTK IHTL	RPRPIP	KPFQCRIC	1	HTK IHTG	LPRKR	KPFQCRIC
1	HTK IHTL	GLHKKP	KPFQCRIC	3	HTK IHTQ	RQIIP	KPFQCRIC	3	HTK IHTS	RRRP	KPFQCRIC
2	HTK IHM	ERAKRP	KPFQCRIC	3	HTK IHTN	NRCP	KPFQCRIC	1	HTK IHTL	PLFRP	KPFQCRIC
1	HTK IHTL	SHTPRP	KPFQCRIC	2	HTK IHTY	BRBLLA	KPFQCRIC	4	HTK IHTV	RRPTRP	KPFQCRIC
1	HTK IHTG	YSLPRP	KPFQCRIC	10	HTK IHTF	LCQFRPMKKO	KPFQCRIC	1	HTK IHTL	PPCFRP	KPFQCRIC
1	HTK IHTV	PRPIA	KPFQCRIC	1	HTK IHTP	LCQFRPMKKO	KPFQCRIC	1	HTK IHTL	PPBCFRP	KPFQCRIC
1	HTK IHTH	RRAPI	KPFQCRIC								
1	HTK IHTN	RRAIP	KPFQCRIC								
1	HTK IHTS	PLPAP	KPFQCRIC								
1	HTK IHTC	RRPPT	KPFQCRIC								
1	HTK IHTV	SRBSNA	KPFQCRIC								
1	HTK IHTS	RAFCR	SKPFQCRIC								
1	HTK IHM	DRPIST	CKPFQCRIC								

B.

ZFP8196			
n.º de clones			
13	HTK IHTL	NACKPYRT	KPFQCRIC
2	HTK IHTL	LAPRPYRPP	KPFQCRIC
1	HTK IHTG	SEHVRRANS	KPFQCRIC
1	HTK IHTD	DAAPRRBRDDT	KPFQCRIC
1	HTK IHTE	YCTRRFRP	KPFQCRIC
1	HTK IHTN	PNPRRYDPSH	KPFQCRIC
1	HTK IHTN	TPRRYRLRPP	KPFQCRIC
1	HTK IHTL	GGKSSRTDR	NKPFQCRIC

Figura 4

Selectividad de hueco de enlazadores seleccionados para omitir un par de bases en ZFP8196

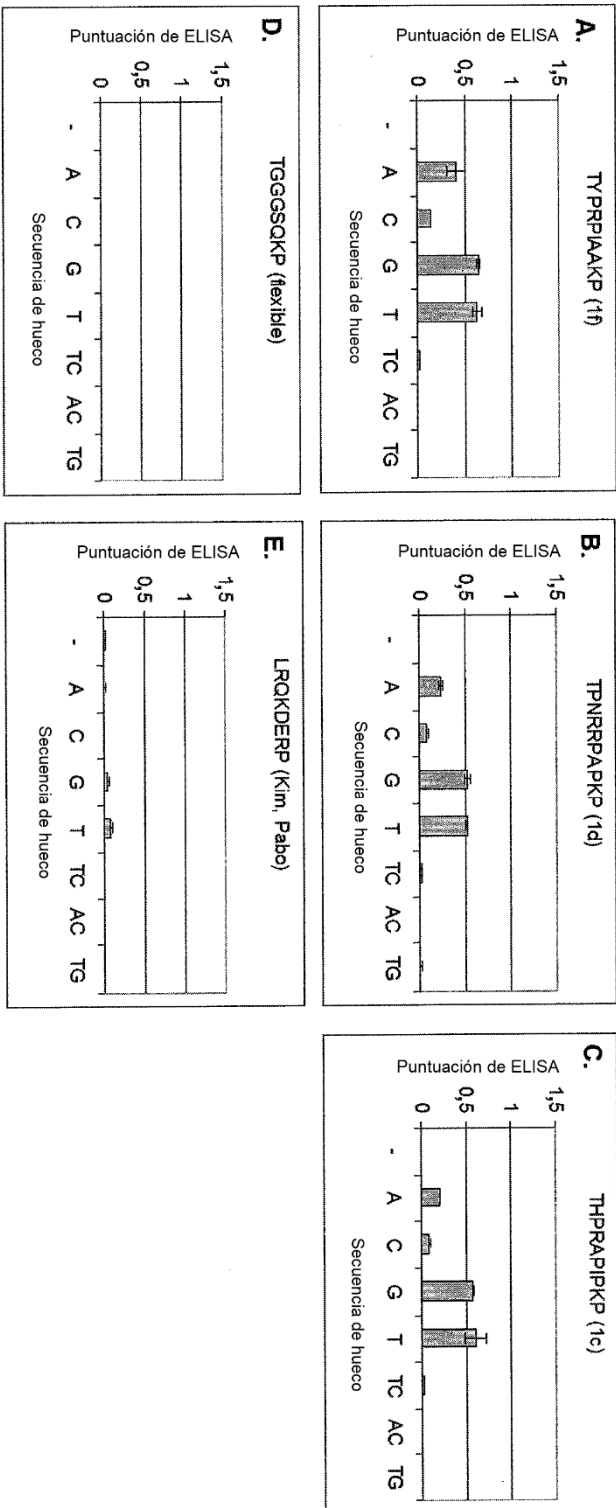


Figura 5

Selectividad de hueco de enlazadores seleccionados para omitir un par de bases en ZFP7264

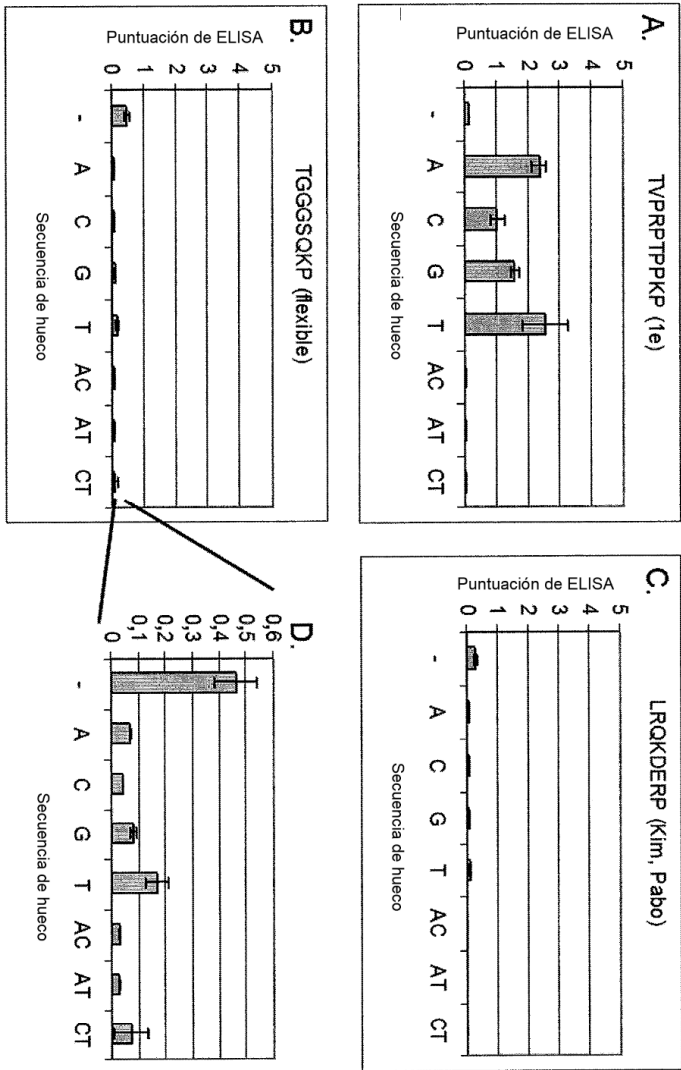


Figura 6

Selectividad de hueco de enlazadores seleccionados para omitir dos pares de bases en ZFP8196

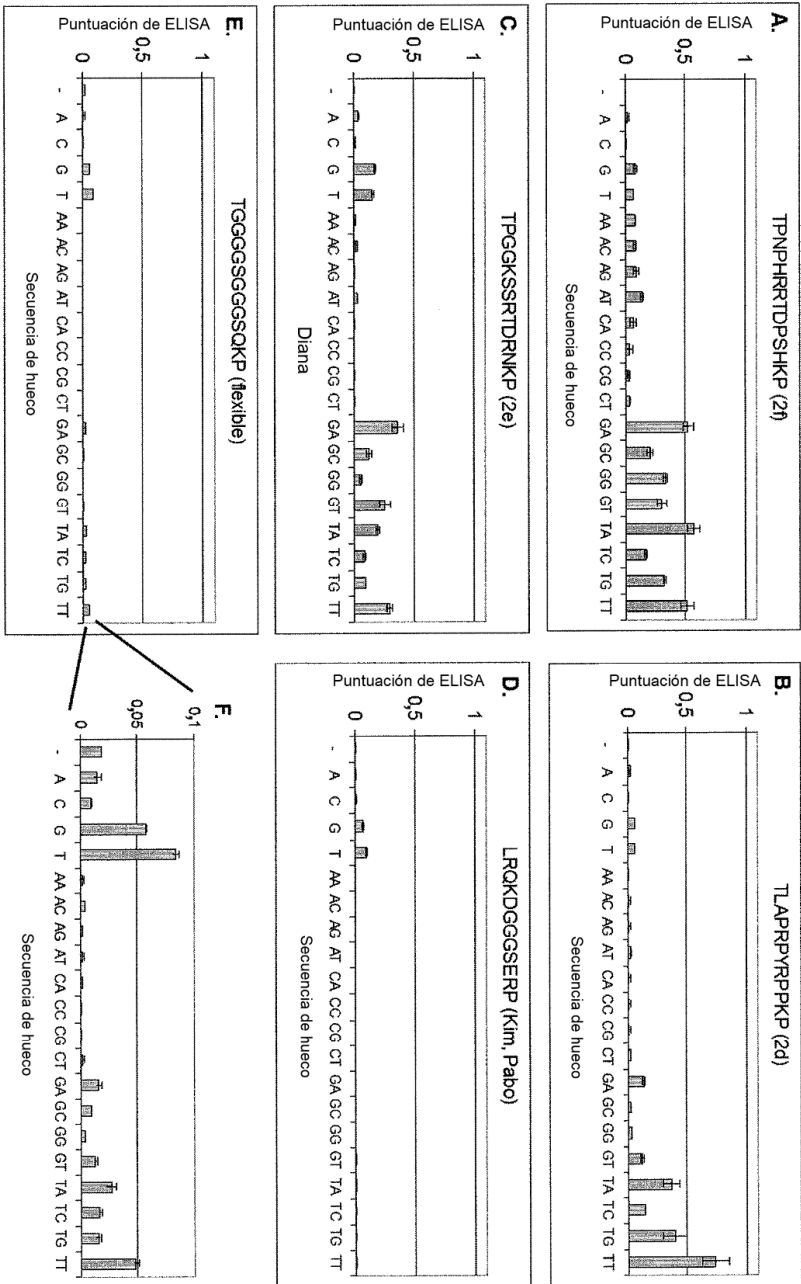


Figura 7

Estudios de portabilidad con enlazadores seleccionados para omitir un par de bases

A

Enlazador Secuencia	Designación	Puntuación de ELISA											
		ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12
TGGGSSQKP	flexible	0,75	1,08	0,21	0,26	0,08	0,05	0,05	0,05	0,05	0,05	0,05	0,05
TVPRPTPKP	1e	1,19	1,30	1,05	1,10	0,75	0,29	0,48	0,38	0,50	0,18	0,05	0,05
TYRRPIAAKP	1f	1,25	1,35	1,03	0,94	0,96	0,30	0,33	0,12	0,30	0,11	0,05	0,05
TPNRRPAPKP	1d	1,20	1,08	1,04	0,83	0,85	0,25	0,39	0,28	0,24	0,12	0,05	0,05
THPRAPIPKP	1c	0,92	1,18	0,66	0,56	0,91	0,23	0,21	0,26	0,18	0,05	0,05	0,05

B

Enlazador Secuencia	Designación	Puntuación de ELISA normalizado a enlazador flexible												Promedio	
		ZFP1	ZFP2	ZFP3	ZFP4	ZFP5	ZFP6	ZFP7	ZFP8	ZFP9	ZFP10	ZFP11	ZFP12		
TGGGSSQKP	flexible	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	5,0
TVPRPTPKP	1e	1,6	1,2	<u>4,9</u>	<u>4,1</u>	<u>9,2</u>	<u>5,9</u>	<u>9,6</u>	<u>7,6</u>	<u>10,1</u>	3,6	1,0	1,0	1,0	<u>4,0</u>
TYRRPIAAKP	1f	1,7	1,2	<u>4,9</u>	3,5	<u>11,7</u>	<u>5,9</u>	<u>6,5</u>	2,3	<u>6,0</u>	2,2	1,0	1,0	1,0	<u>4,0</u>
TPNRRPAPKP	1d	1,6	1,0	<u>4,9</u>	3,1	<u>10,4</u>	<u>4,9</u>	<u>7,8</u>	5,7	<u>4,8</u>	2,4	1,0	1,0	1,0	<u>4,0</u>
THPRAPIPKP	1c	1,2	1,1	3,1	2,1	<u>11,2</u>	<u>4,7</u>	<u>4,3</u>	5,3	3,6	1,0	1,0	1,0	1,0	3,3

los valores subrayados muestran una mejora >4 veces

Figura 8

Estudios de portabilidad con enlazadores seleccionados para omitir dos pares de bases

Enlazador		Puntuación de ELISA									
Secuencia	Designación	ZFP13	ZFP14	ZFP15	ZFP16	ZFP17	ZFP18				Promedio
TGGGSSGGSQKP	flexible	0,25	0,61	0,05	0,13	0,35	0,51				
TPNPHRRTDPSHKP	2f	0,48	1,13	0,27	0,31	0,79	0,47				
TLAPRPYRPPKP	2d	0,44	1,13	0,11	0,34	0,82	0,43				
TPGGKSSRTDRNKP	2e	0,42	0,75	0,17	0,27	0,68	0,52				

Enlazador		Puntuación de ELISA normalizado a enlazador flexible									
Secuencia	Designación	ZFP13	ZFP14	ZFP15	ZFP16	ZFP17	ZFP18				Promedio
TGGGSSGGSQKP	flexible	1,0	1,0	1,0	1,0	1,0	1,0				1,0
TPNPHRRTDPSHKP	2f	1,9	1,9	<u>5,1</u>	<u>2,3</u>	<u>2,3</u>	0,9				<u>2,4</u>
TLAPRPYRPPKP	2d	1,7	1,9	2,1	<u>2,5</u>	2,4	0,8				1,9
TPGGKSSRTDRNKP	2e	1,7	1,2	<u>3,3</u>	<u>2,0</u>	<u>2,0</u>	1,0				1,9

los valores subrayados muestran una mejora >2 veces

Figura 9

Geles de ensayo de ejemplo

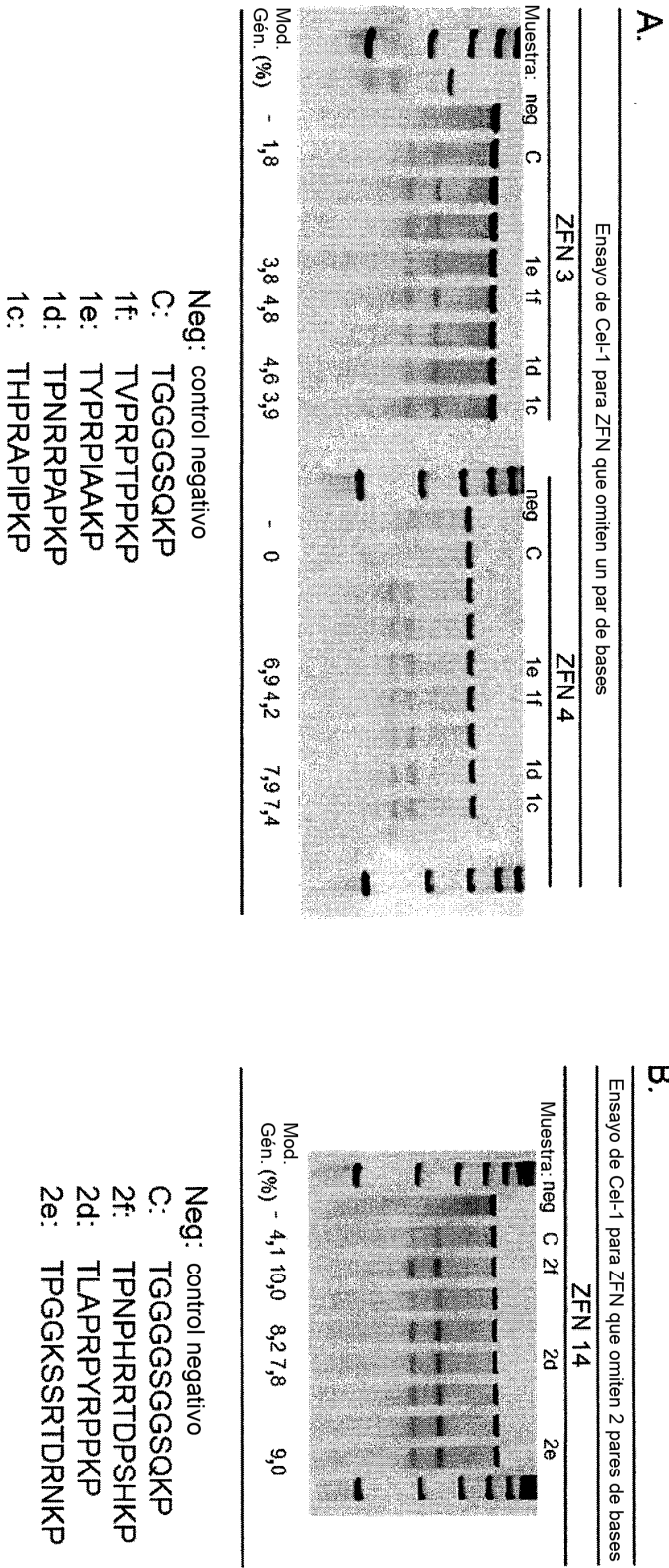


Figura 10

Modificación génica para ZFN que contiene nuevos enlazadores que omiten un par de bases

A

Enlazador Secuencia	Designación	Modificación génica (%)											
		ZFN1	ZFN2	ZFN3	ZFN4	ZFN5	ZFN6	ZFN7	ZFN8	ZFN9	ZFN10	ZFN11	ZFN12
TGGGGSQKP	flexible	6,1	1,8	5,9	0,0	3,5	-	1,1	-	0,0	0,0	-	1,7
TVRRTPPKP	1e	9,5	3,8	9,3	1,0	4,9	-	2,6	-	3,7	6,9	-	1,9
TYRPIAAKP	1f	9,6	4,8	12,0	1,0	2,5	-	1,5	-	1,5	4,2	-	2,8
TPNRRPAPKP	1d	6,4	4,6	12,5	1,5	6,2	-	2,6	-	4,5	7,9	-	1,9
THPRAPPKP	1c	9,9	3,9	12,8	1,4	5,1	-	2,7	-	NA	7,4	-	NA

B

Enlazador Secuencia	Designación	Modificación génica normalizada a enlazador flexible												Promedio
		ZFN1	ZFN2	ZFN3	ZFN4	ZFN5	ZFN6	ZFN7	ZFN8	ZFN9	ZFN10	ZFN11	ZFN12	
TGGGGSQKP	flexible	1,0	1,0	1,0	-	1,0	-	1,0	-	-	-	-	1,0	
TVRRTPPKP	1e	1,6	2,1	1,6	>1	1,4	-	2,4	-	>3,5	>7,0	-	1,1	
TYRPIAAKP	1f	1,6	2,7	2,0	>1	0,7	-	1,4	-	>1,5	>4,0	-	1,6	
TPNRRPAPKP	1d	1,0	2,6	2,1	>1,5	1,8	-	2,4	-	>4,5	>8,0	-	1,1	
THPRAPPKP	1c	1,6	2,2	2,2	>1,5	1,5	-	2,5	-	NA	>7,0	-	NA	

condiciones de expresión alta

los valores subrayados muestran una mejora >2 veces

- | ninguna señal cuantificable
NA | no sometido a ensayo

Figura 11

Modificación génica para ZFN que contienen nuevos enlazadores que omiten dos pares de bases

A

Enlazador		Modificación génica (%)							
Secuencia	Designación	ZFN13	ZFN14	ZFN15	ZFN16	ZFN17	ZFN18		
TGGGGSGSQKP	flexible	-	7,5	2,5	2,9	5,3	-		
TPNPHRRTDPSHKP	2f	-	17,0	5,6	7,3	4,7	-		
TLAPRPYRPPKP	2d	-	11,3	5,0	4,8	4,9	-		
TPGGKSSRTDRNKP	2e	-	11,7	5,0	4,3	6,3	-		

B

Enlazador		Modificación génica normalizada a enlazador flexible							
Secuencia	Designación	ZFN13	ZFN14	ZFN15	ZFN16	ZFN17	ZFN18	Promedio	
TGGGGSGSQKP	flexible	-	1,0	1,0	1,0	1,0	-	2,0	
TPNPHRRTDPSHKP	2f	-	2,3	2,2	2,5	0,9	-	1,5	
TLAPRPYRPPKP	2d	-	1,5	2,0	1,7	0,9	-	1,5	
TPGGKSSRTDRNKP	2e	-	1,6	2,0	1,5	1,2	-	1,6	

condiciones de expresión alta

los valores subrayados muestran una mejora >2 veces

ninguna señal cuantificable

