(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 949 163**

(51) Int. CI.:

***C07K 16/18*** (2006.01)
***C12N 9/00*** (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **27.01.2017 PCT/US2017/015455**

(87) Fecha y número de publicación internacional: **03.08.2017 WO17132580**

(96) Fecha de presentación y número de la solicitud europea: **27.01.2017 E 17745022 (8)**

(97) Fecha y número de publicación de la concesión europea: **26.04.2023 EP 3408292**

(54) Título: **Inteínas divididas con actividad de corte y empalme excepcional**

(30) Prioridad:

**29.01.2016 US 201662288661 P**

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
**26.09.2023**

(73) Titular/es:

**THE TRUSTEES OF PRINCETON UNIVERSITY (100.0%)**
**87 Prospect Avenue**
**Princeton, NJ 08544, US**

(72) Inventor/es:

**MUIR, TOM, W.;**
**STEVENS, ADAM, J. y**
**SHAH, NEEL, H.**

(74) Agente/Representante:

**ARIAS SANZ, Juan**

**Observaciones:**

**Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes**

## DESCRIPCIÓN

Inteínas divididas con actividad de corte y empalme excepcional

**Antecedentes**

1. Campo técnico

El campo de las realizaciones actualmente reivindicadas de la presente invención se refiere a inteínas, inteínas divididas, composiciones que comprenden inteínas y métodos para el uso de las mismas para la ingeniería de proteínas.

2. Discusión de la técnica relacionada

El corte y empalme de proteínas es un evento de autoprocesamiento postraduccional en el que un dominio proteico intermedio denominado inteína se escinde a sí mismo de una proteína huésped sin dejar rastro, de modo que las secuencias polipeptídicas flanqueantes (exteínas) se ligan entre sí por medio de un enlace peptídico normal.[1] Mientras que el corte y empalme de proteínas normalmente se produce espontáneamente después de la traducción de un polipéptido contiguo, algunas inteínas existen de manera natural en forma dividida.[1] Los dos trozos de la inteína dividida se expresan por separado y permanecen inactivos hasta encontrarse con su compañero complementario, tras lo cual se pliegan cooperativamente y experimentan corte y empalme en trans. Esta actividad se ha aprovechado en una serie de métodos de ingeniería de proteínas que proporcionan control sobre la estructura y la actividad de las proteínas tanto *in vitro* como *in vivo*.[1] Las dos primeras inteínas divididas en caracterizarse, de las especies de cianobacterias *Synechocystis* PCC6803 (Ssp) y *Nostoc punctiforme* PCC73102 (Npu), son ortólogos que se encuentran de manera natural insertados en la subunidad alfa de la ADN polimerasa III (DnaE).[2-4] Npu es especialmente notable debido a su velocidad notablemente rápida de corte y empalme en trans de proteínas (PTS) ($t_{1/2}$=50 s a 30 °C).[5] Esta semivida es significativamente más corta que la de Ssp ($t_{1/2}$=80 min a 30 °C),[5] un atributo que ha ampliado la gama de aplicaciones abiertas a PTS.[1]

A pesar del descubrimiento continuo de nuevas inteínas rápidas,[6,7] se sabe poco sobre qué las separa de sus homólogas más lentas. Tal comprensión debe ayudar a identificar nuevas inteínas que probablemente experimenten corte y empalme rápidamente y permitan potencialmente la ingeniería de inteínas divididas con propiedades de PTS superiores.

**Sumario**

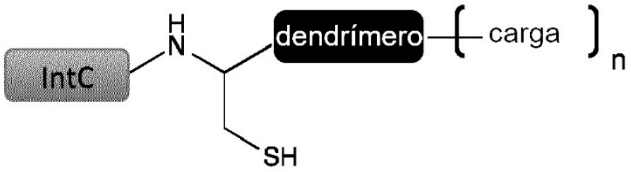
En un aspecto, la invención se refiere a un fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1) o con CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 2).

En otro aspecto, la invención se refiere a un complejo que comprende el fragmento N de inteína dividida de la invención y un compuesto.

En otro aspecto, la invención se refiere a un fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3), con MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 4) o con VKIISRKSLGTQNVYDIGVGEPHNFLLKNGLVASN (SEQ ID NO: 389).

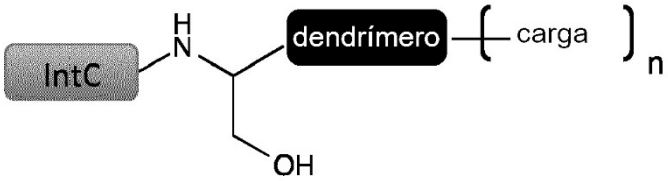
En otro aspecto, la invención se refiere a un complejo que comprende el fragmento C de inteína dividida de la invención y un compuesto.

Un complejo de la estructura




,

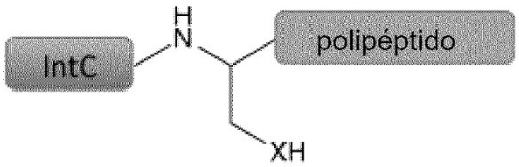
en donde IntC es el fragmento C de inteína dividida de la invención y en donde n es de 0 a 8, la estructura




,

en donde IntC es el fragmento C de inteína dividida de la invención y en donde n es de 0 a 8,

o la estructura




,

en donde IntC es el fragmento C de inteína dividida de la invención y en donde X es azufre (S) u oxígeno (O).

En otro aspecto, la invención se refiere a una composición que comprende:

el fragmento N de inteína dividida de la invención; y el fragmento C de inteína dividida de la invención.

En otro aspecto, la invención se refiere a un plásmido de nucleótidos que comprende una secuencia de nucleótidos que codifica el fragmento N de inteína dividida de la invención o el fragmento C de inteína dividida de la invención.

En otro aspecto, la invención se refiere a un método para cortar y empalmar dos complejos que comprende:

poner en contacto un primer complejo que comprende un primer compuesto y el fragmento N de inteína dividida de la invención y un segundo complejo que comprende un segundo compuesto y el fragmento C de inteína dividida de la invención,

en donde la puesta en contacto se realiza en condiciones que permiten la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

hacer reaccionar el producto intermedio de inteína para formar un conjugado del primer compuesto con el segundo compuesto.

En otro aspecto, la invención se refiere a un método seleccionado del grupo que comprende:

(i) un método que comprende

poner en contacto un primer complejo que comprende un primer compuesto y el fragmento N de inteína dividida de la invención y un segundo complejo que comprende un segundo compuesto y el fragmento C de inteína dividida de la invención,

en donde la puesta en contacto se realiza en condiciones que permiten la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

hacer reaccionar el producto intermedio de inteína con un nucleófilo para formar un conjugado del primer compuesto con el nucleófilo y

(ii) un método que comprende

fusionar una primera secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento N de inteína dividida de la invención, con una segunda secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento C de inteína dividida de la invención, de modo que la fusión de la primera secuencia de nucleótidos y la segunda secuencia de nucleótidos codifica una inteína contigua.

En otro aspecto, la invención se refiere a una inteína que comprende una secuencia de aminoácidos de al menos el 90 %, 95 %, 98 %, 99 % o 100 % de identidad de secuencia con
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLPVKIISRKSLGTQNVYDIGVEKDH

NFLLKNGLVASN (SEQ ID NO: 390).

En otro aspecto, la invención se refiere a un kit para cortar y empalmar dos complejos juntos que comprende:

el fragmento N de inteína dividida de la invención;

el fragmento C de inteína dividida de la invención;

un reactivo para unir el fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

un agente nucleófilo.

**Breve descripción de los dibujos**

La figura 1 muestra una alineación y un modelo generado por ordenador del diseño de la inteína dividida Cfa de acuerdo con una realización de la invención;

la figura 2 muestra gráficos que muestran la caracterización de la inteína Cfa de acuerdo con una realización de la invención;

la figura 3 muestra la expresión y modificación de un anticuerpo monoclonal de ratón usando la inteína Cfa de acuerdo con una realización de la invención;

la figura 4 muestra la identificación de residuos “aceleradores” de la segunda cubierta importantes para el corte y empalme en trans rápido de proteínas de acuerdo con una realización de la invención;

la figura 5 muestra el análisis cinético de mutaciones del lote 2 y modelos generados por ordenador de acuerdo con una realización de la invención;

la figura 6 muestra un análisis de mutaciones del lote 1 y modelos generados por ordenador de acuerdo con una realización de la invención;

la figura 7A y la figura 7B muestran una alineación y el refinamiento de la familia de inteínas DnaE de acuerdo con el aspecto divulgado de la invención (aspecto no cubierto por la invención reivindicada);

la figura 8 es una imagen de un análisis de SDS-PAGE de la expresión de prueba de $His_6$-SUMO-$Npu^N$ e $His_6$-SUMO-$Cfa^N$ de acuerdo con una realización de la invención;

la figura 9 muestra un esquema y gráfico que muestran una mayor promiscuidad de $Cfa_{GEP}$ de acuerdo con una realización de la invención;

la figura 10 muestra gráficos y esquemas que muestran la ciclación de eGFP en *E. coli* con residuos variables de acuerdo con una realización de la invención; y

la figura 11 muestra una tabla que ilustra varios complejos y compuestos de acuerdo con una realización de la invención.

**Descripción detallada**

A continuación se comentan realizaciones de la invención en detalle.

Las realizaciones de la invención incluyen un fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 80 %, 85 %, 90 %, 95 %, 98 %, 99 % o 100 % de identidad de secuencia con

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1).

Las realizaciones de la invención incluyen un fragmento N de inteína dividida que comprende una secuencia de aminoácidos, en donde dicha secuencia de aminoácidos comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 2).

Las realizaciones de la invención incluyen un fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con

VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3).

Las realizaciones de la invención incluyen un fragmento C de inteína dividida que comprende una secuencia de aminoácidos, en donde dicha secuencia de aminoácidos de dicho fragmento C comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con
MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 4).

Las realizaciones de la invención incluyen una composición que comprende lo siguiente: un fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1); y un fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3).

Las realizaciones de la invención incluyen un plásmido de nucleótidos que comprende una secuencia de nucleótidos que codifica un fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1).

Las realizaciones de la invención incluyen un plásmido de nucleótidos que comprende una secuencia de nucleótidos que codifica un fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con
VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3).

Las realizaciones de la invención incluyen un método para cortar y empalmar dos complejos que comprende: poner en contacto un primer complejo que comprende un primer compuesto y un fragmento N de inteína dividida y un segundo complejo que comprende un segundo compuesto y un fragmento C de inteína dividida, en donde la puesta en contacto se realiza en condiciones que permiten la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y hacer reaccionar el producto intermedio de inteína para formar un conjugado del primer compuesto con el segundo compuesto, en donde dicho fragmento N de inteína dividida comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1), y en donde dicho fragmento C de inteína dividida comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3). En algunas realizaciones, hacer reaccionar el producto intermedio de inteína comprende poner en contacto el producto intermedio de inteína con un nucleófilo. En algunas realizaciones, dicho primer compuesto es un polipéptido. En algunas realizaciones, dicho primer compuesto es un anticuerpo.

Las realizaciones de la invención incluyen una inteína que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED

GSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLPVKIISRKSLGTQNVYDIGVEKDH

NFLLKNGLVASN (SEQ ID NO: 390).

Las realizaciones de la invención incluyen un kit para cortar y empalmar dos complejos juntos que comprende lo siguiente: un fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1); un fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, 99 % o 100 % de identidad de secuencia con VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3); reactivos para permitir la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y un agente nucleófilo.

Se da a conocer un método para generar una secuencia peptídica de inteína consenso sintética (aspecto no cubierto por la invención reivindicada) que comprende: generar una población de una pluralidad de secuencias peptídicas de inteína homólogas; identificar aminoácidos asociados con el corte y empalme rápido dentro de dicha población de una pluralidad de secuencias peptídicas de inteína homólogas; generar una subpoblación de una segunda pluralidad de secuencias peptídicas de inteína homólogas, en donde dicha segunda pluralidad de secuencias peptídicas de inteína homólogas comprende aminoácidos asociados con el corte y empalme rápido; crear un alineamiento de al menos tres secuencias peptídicas de dicha subpoblación; determinar un residuo de aminoácido que aparece con mayor frecuencia en cada posición de dichas al menos tres secuencias peptídicas; y generar una secuencia peptídica de inteína consenso sintética basándose en dicho residuo de aminoácido que aparece con mayor frecuencia en cada posición de dichas al menos tres secuencias peptídicas.

Las realizaciones de la invención incluyen un método que comprende: fusionar una primera secuencia de nucleótidos que codifica una secuencia de aminoácidos de un primer fragmento de inteína que comprende CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1) con una segunda secuencia de nucleótidos que codifica una secuencia de aminoácidos de un segundo fragmento de inteína que comprende VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3), de modo que la fusión de la primera secuencia de nucleótidos y la segunda secuencia de nucleótidos codifica una inteína contigua.

Las realizaciones de la invención incluyen una fusión génica que comprende una primera secuencia de nucleótidos que codifica una secuencia de aminoácidos de un primer fragmento de inteína que comprende CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLED GSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1) fusionado con una segunda secuencia de nucleótidos que codifica una secuencia de aminoácidos de un segundo fragmento de inteína que comprende VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3).

Las realizaciones de la invención incluyen una inteína contigua que puede usarse, por ejemplo, en aplicaciones de semisíntesis tradicionales tales como ligamiento de proteínas expresadas.

En algunas realizaciones, los diversos fragmentos de inteína descritos se unen, se fusionan, se enlazan químicamente, se complejan o se acoplan por métodos convencionales conocidos en la técnica a polímeros, péptidos, polipéptidos, oligopéptidos, moléculas pequeñas, nucleótidos, polinucleótidos, oligonucleótidos, fármacos, moléculas citotóxicas o combinaciones de los mismos.

**Ejemplo 1**

En algunos aspectos, se investigó la base del corte y empalme rápido de proteínas a través de un estudio comparativo de las dos primeras inteínas divididas caracterizadas, Npu y Ssp. La diferencia sustancial en la velocidad de corte y empalme entre estas dos proteínas es especialmente desconcertante dada sus secuencias altamente similares (63 % de identidad) y estructuras de sitio activo casi superponibles. Estudios previos de mutagénesis en Npu y Ssp sugieren que la diferencia en la actividad entre los dos probablemente se deba a los efectos combinados de varios residuos, en lugar de un solo sitio.[6,8] Sin embargo, sigue sin estar claro cuántos residuos son responsables de las velocidades de reacción rápidas frente a lentas y, por extensión, si estos residuos “aceleradores” de proteínas contribuyen por igual a las etapas químicas individuales en el proceso global de corte y empalme de proteínas. En consecuencia, los inventores comenzaron su estudio explorando estas preguntas, con la esperanza de que esto proporcionara un punto de partida para desarrollar un sistema de PTS mejorado.

El alto nivel de conservación dentro de los sitios activos de Npu y Ssp sugiere que las diferencias de aminoácidos más distales explican la disparidad en la velocidad de corte empalme entre las dos. Por lo tanto, la atención se centró en los residuos de la “segunda cubierta”, aquellos directamente adyacentes al sitio activo. Para simplificar este análisis, se empleó una estrategia de mutagénesis por lotes junto con un ensayo de PTS *in vitro* previamente notificado.[5] Este ensayo usa constructos de inteína dividida con secuencias de exteína nativa cortas y permite que las velocidades de formación de productos intermedios ramificados ($k_1$, $k_2$) y su resolución a los productos finales de corte y empalme ($k_3$)

se determinen usando un modelo cinético de tres estados.

La reactividad cruzada conocida de los fragmentos de inteína Npu y Ssp sirvió como plataforma conveniente para evaluar qué mitad de la inteína dividida contribuye más significativamente a la diferencia de actividad.[3] Ambas quimeras $Ssp^N$-$Npu^C$ (quimera 1) y $Npu^N$-$Ssp^C$ (quimera 2) muestran una disminución en las velocidades de formación y resolución de la ramificación en comparación con la de Npu nativa (figuras 4C, 4D). Esto indica que los residuos en ambos fragmentos de inteína N y C de Npu y Ssp contribuyen a la diferencia en su velocidad de corte y empalme. A continuación, se eligieron cuatro grupos de posiciones de la segunda cubierta en cada una de estas quimeras basándose en su proximidad al sitio activo, y los residuos de Ssp correspondientes se mutaron a los de Npu (figuras 4A y 4B). A partir de los mutantes de la quimera 1, el lote 2 (L56F, S70K, A83P, E85D) restableció completamente la actividad de formación de ramificaciones a la de Npu nativa (figura 4C), mientras que el lote 1 (R73K, L75M, Y79G, L81M) restableció la mayor parte de la actividad de resolución de ramificaciones (figura 4D). Los efectos de las mutaciones sobre los antecedentes de la quimera 2 fueron más prosaicos, sin un solo lote capaz de restablecer la actividad de corte y empalme a la de Npu nativa (figura 4C y 4D). Por último, previamente se ha demostrado que la mutación A136S en $Ssp^C$ acelera el corte y empalme de proteínas y se examinó por separado.[8] Esta mutación A136S aumenta la velocidad de resolución de ramificaciones dos veces, pero no tiene impacto sobre la formación de ramificaciones (figuras 4C y 4D).

La figura 4 muestra la identificación de residuos “aceleradores” de la segunda cubierta importantes para el corte y empalme en trans rápido de proteínas de acuerdo con una realización de la invención. En los paneles A y B, se muestra el diseño de mutantes de lote de la segunda cubierta en la quimera 1 ($Ssp^N$-$Npu^C$) y la quimera 2 ($Npu^N$-$Ssp^C$). En cada caso, la ubicación de los mutantes (representados como barras) se muestra usando la estructura cristalina de Npu (pdb = 4kl5). Los residuos catalíticos se muestran en negro (representados como barras). El panel C muestra las velocidades hacia adelante ($k_1$, azul) y hacia atrás ($k_2$, rojo) de formación de productos intermedios ramificados a partir de materiales de partida para los diversos constructos descritos en este estudio (error = DE (n = 3)). El panel D muestra la velocidad de resolución de ramificaciones ($k_3$) de los diversos constructos (error = DE (n = 3)).

A continuación, se investigaron las contribuciones individuales de los residuos dentro de los mutantes de lote 1 y 2, ya que estos tuvieron el efecto más profundo sobre la actividad de corte y empalme. Para el lote 2, la mutagénesis adicional muestra que la interacción entre F56, K70 y D85 es probablemente responsable de la mayor velocidad de formación de ramificaciones en $Npu^N$ (figura 5A). Pruebas estructurales respaldan estos datos, ya que K70 es una parte del bucle B del bloque TXXH altamente conservado en $Npu^N$ (residuos 69-72) que cataliza el desplazamiento inicial de acilo de N a S en el corte y empalme de proteínas.[9] Por lo tanto, la posición y la dinámica de K70 (empaquetado contra F56 y D85) deben tener un impacto directo en los residuos catalíticos T69 y H72 (figura 5B).[10-12] Del lote 1, K73, M75 y M81 son responsables de la velocidad más rápida de resolución de ramificaciones en $Npu^N$ (figura 6A). Estos residuos se empaquetan alrededor de la asparagina terminal de la C-inteína, que debe experimentar la formación de succinimida en la etapa final del corte y empalme de proteínas (figura 6B). Tomados en conjunto, los datos de mutagénesis apuntan al papel clave que desempeñan los residuos “aceleradores” de la segunda cubierta en el ajuste de la actividad de las inteínas divididas.

La figura 5 muestra el análisis cinético de mutaciones del lote 2 y modelos generados por ordenador de acuerdo con una realización de la invención. El panel A muestra las velocidades de equilibrio de la formación de ramificaciones ($k_1$, $k_2$) y las velocidades de resolución de ramificaciones ($k_3$) para los mutantes puntuales individual (A83P), doble (A83P, S70K) y triple (L56F, S70K, A83P) de $Ssp^N$ que comprenden el lote 2 (L56F, S70K, A83P, E85D) (error = DE (n=3)). El panel B muestra una vista ampliada del lote 2 (barras verdes junto a las etiquetas F56, K70, P83 y D85) en el sitio activo de Npu (pdb = 4kl5). Los residuos catalíticos se presentan como barras negras.

La figura 6 muestra un análisis de mutaciones del lote 1 y modelos generados por ordenador de acuerdo con una realización de la invención. El panel A muestra las velocidades de equilibrio de la formación de ramificaciones ($k_1$, $k_2$) y las velocidades de resolución de ramificaciones ($k_3$) para los mutantes puntuales individual (R73K), doble (R73K, Y79G) y triple (R73K, Y79G, L81M) que comprenden el lote 1 (error = DE (n=3)). El panel B muestra una vista ampliada del lote 1 (barras rojas junto a las etiquetas K73, M75, G79 y M81) en la estructura de Npu (pdb = 4kl5). Los residuos catalíticos se presentan como barras negras.

Los residuos “aceleradores” que se encuentra que afectan a la velocidad de corte y empalme permiten un enfoque guiado por la actividad para diseñar una inteína DnaE consenso. La ingeniería de proteínas de consenso es una herramienta aplicada a un conjunto homólogo de proteínas para crear una variante termoestable derivada de la familia original.[13,14] En primer lugar, se genera una alineación de secuencias múltiples (MSA) a partir de homólogos de una proteína particular, de la cual se elige el residuo más frecuente estadísticamente en cada posición como representante en la secuencia consenso. Para las inteínas DnaE, se identificaron 105 secuencias a través de una búsqueda BLAST[15] de las bases de datos JGI[16] y NCBI[17] (figura 7A). A continuación, la alineación se filtró para que contuviera solo secuencias que portaban los indicadores de segunda cubierta de corte y empalme rápido: K70, M75, M81 y S136. Las 73 inteínas teóricamente rápidas que quedaban en la MSA (figura 7B) se usaron entonces para generar una secuencia de inteína DnaE rápida consenso (Cfa) (figura 1). Las diversas secuencias divulgadas en las figuras 7A y 7B se presentan a continuación:

>NpuPCC73102/1-137

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQE
VFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 5)

IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 6)

>CthPCC7203:/1-137 *Chroococcidiopsis thermalis* PCC 7203

CLSYDTEILTVEYGAIPIGKIVEERIECTVYSVDNNGFIYTQPIAQWHNRGQQEV
FEYCLEDGSIIRATKDHKFMTFEGKMLPIDEIFEQELDLKQVKSIQN (SEQ ID NO: 7)

VKIISRKSLGIQPVYDIGVERDHKFVLKNGLVASN (SEQ ID NO:8)

>NspCCY9414:/1-137 genoma de *Nodularia spumigena* CCY9414

CLSYDTEILTVEYGYIPIGEIVEKAIECSVYSVDNNGNVYTQPIAQWHNRGEQE
VFEYSLEDGSTIRATKDHKFMTTDGQMLPIDEIFAQELDLLQVHGLPK (SEQ ID NO: 9)

VKITARKFVGRENVYDIGVERYHNFAIKNGLIASN (SEQ ID NO: 10)

>AcyPCC7122:/1-137 *Anabaena cylindrica* PCC 7122

CLSYDTEVLTVEYGFIPIGEIVEKRIECSIFSVDKNGNVYTQPIAQWHNRGRQEI
YEYCLDDGSKIRATKDHKFMTTAGEMLPIDEIFERDLDLLKVEGLPE (SEQ ID NO: 11)

VKIISRQYLGQADVYDIGVEEDHNFAIKNGFIASN (SEQ ID NO: 12)

>CspPCC7507:/1-137 *Calothrix sp.* PCC 7507, genoma completo

CLSYDTEVLTVEYGLLPIGEIVEKGIECRVFSVDNHGNVYTQPIAQWHNRGQQE
VFEYGLDDGSVIRATKDHKFMTTDGKMLPIDEIFERGLDLLQVQGLPE (SEQ ID NO: 13)

VKVITRKYIGKENVYDIGVELDHNFAIRNGLVASN (SEQ ID NO: 14)

>NspPCC7524:/1-137 *Nostoc sp.* PCC 7524

CLSYDTEILTVEYGFLPIGEIVEKGIECTVFSVASNGIVYTQPIAQWHNRGQQEIF
EYCLEDGSIIRATKDHKFMTQDGQMLPIDEIFACELDLLQVQGLPE (SEQ ID NO: 15)

VKVVTRKYIGKENVYDIGVERDHNFVIRNGLVASN (SEQ ID NO: 16)

>Naz0708:/1-137 '*Nostoc azollae*' 0708

CLSYKTEVLTVEYGLIPIGEIVEKRIECSLFSVDENGNIYTQPIAQWHHRGVQEV
YEYCLDDGTIIRATKDHKFMTTIGEMLPIDEIFERDLNLLQVNGLPT (SEQ ID NO: 17)

VKIISRQFLGPANVYDIGVAQDHNFAIKNGLIASN (SEQ ID NO: 18)

>NspPCC7120:/1-137 ADN de *Nostoc sp.* PCC 7120

CLSYDTEVLTVEYGFVPIGEIVEKGIECSVFSINNNGIVYTQPIAQWHHRGKQEV
FEYCLEDGSIIKATKDHKFMTQDGKMLPIDEIFEQELDLLQVKGLPE (SEQ ID NO: 19)

IKIASRKFLGVENVYDIGVRRDHNFFIKNGLIASN (SEQ ID NO: 20)

>AvaATCC29413/1-137 *Anabaena variabilis* ATCC 29413

CLSYDTEVLTVEYGFVPIGEIVDKGIECSVFSIDSNGIVYTQPIAQWHHRGKQEV
FEYCLEDGSIIKATKDHKFMTQDGKMLPIDEIFEQELDLLQVKGLPE (SEQ ID NO: 21)

IKIASRKFLGVENVYDIGVGRDHNFFVKNGLIASN (SEQ ID NO: 22)

>PspPCC7327:/1-135 *Pleurocapsa sp.* PCC 7327.

CLSYDTKILTVEYGAMPIGKIVEEQIDCTVYTVNQNGFVYTQPIAQWHDRGKQ
EIFEYCLEDGSIIRATKDHKFMTTDGQMLPIDKIFEKGLDLKTINCD (SEQ ID NO: 23)

VKILSRKSLGIQSVYDIGVEKDHNFLLANGLVASN (SEQ ID NO: 24)

>CspPCC7424:/1-135 *Cyanothece sp.* PCC 7424

CLSYETQIMTVEYGLMPIGKIVEEQIDCTVYTVNKNGFVYTQPIAQWHYRGEQ
EVFEYCLEDGSTIRATKDHKFMTTDGQMLPIDEIFEQGLELKQIHLS (SEQ ID NO: 25)

VKIISRQSLGIQPVYDIGVEKDHNFLISDGLIASN (SEQ ID NO: 26)

>CspPCC7822:/1-134 *Cyanothece sp.* PCC 7822

CLSYDTEILTVEYGPMPIGKIVEEQIECTVYTVDKNGLVYTQPIAQWHHRGQQE
VFEYCLEDGSIIRATKDHKFMTDDGQMLPIEEIFEKGLELKQIIL (SEQ ID NO: 27)

VKIISRQLAGNQTVYDLGVEKDHNFLLANGLIASN (SEQ ID NO: 28)

>NspPCC7107:/1-137 *Nostoc sp.* PCC 7107

CLSYDTQVLTVEYGLVPIGEIVEKQLECSVFTIDGHGYVYTQAIAQWHNRGQQ
EVFEYGLEDGSVIRATKDHKFMTTDGQMLPIDEIFERELDLLQVQGLRW (SEQ ID NO: 29)

VKIITRKYIGQANVYDIGVAQDHNFVIENRLIASN (SEQ ID NO: 30)

>Tbolicb1/1-136 *Tolypothrix bouteillei* licb1

CLSYDTEILTVEYGFLPIGKIVEKGIECNVYSVDKNGNIYTQPIAQWHDRGEQE
VFEYCLENGSVIRATKDHKFMTTSGEMLPIDEIFERGLDLIRVEDLP (SEQ ID NO: 31)

VKILTRKSIGKQTVYDIGVERDHNFVIKNGSVASN (SEQ ID NO: 32)

>Aov:/1-136 gen precursor de DnaE (dnaE) de *Aphanizomenon ovalisporum*

CLSADTEILTVEYGFLPIGEIVGKAIECRVYSVDGNGNIYTQSIAQWHNRGEQEV
FEYTLEDGSIIRATKDHKFMTTDGEMLPIDEXFARQLDLMQVQGLH (SEQ ID NO: 33)

VKITARKFVGRENVYDIGVEHHHNFAIKNGLIASN (SEQ ID NO: 34)

>OnvPCC7112:/1-137 *Oscillatoria nigro-viridis* PCC 7112

CLSYDTKILTVEYGPMAIGKIVEEKIECTVYSVDSNGYIYTQSIAQWHRRGQQE
VFEYCLEDGSIIRATKDHKFMTVGGQMLPIDEIFEQGLDLKQINSSSD (SEQ ID NO: 35)

VKIISRKSLGTQEVYDIGVEREHNFILENSLVASN (SEQ ID NO: 36)

>RspPCC7116:/1-135 *Rivularia sp.* PCC 7116, genoma completo

CLSYDTEVLTEEFGLIPIGKIVEEKIDCTVYSVDVNGNVYSQPIAQWHNRGMQE
VFEYELEDGSTIRATKDHKFMTVDGEMLAIDEIFEKGLELKRVGIY (SEQ ID NO: 37)

VKIISRKVLKTENVYDIGLEGDHNFIIKDGLIASN (SEQ ID NO: 38)

>TerIMS101:/1-137 *Trichodesmium erythraeum* IMS101

CLTYETEIMTVEYGPLPIGKIVEYRIECTVYTVDKNGYIYTQPIAQWHNRGMQE
VYEYSLEDGTVIRATPEHKFMTEDGQMLPIDEIFERNLDLKCLGTLEL (SEQ ID NO: 39)

VKIVSRKLAKTENVYDIGVTKDHNFVLANGLIASN (SEQ ID NO: 40)

>MspPCC7113:/1-137 *Microcoleus sp.* PCC 7113,

CLSYDSEILTVEYGLMPIGKIVEEGIECTVYSVDSHGYLYTQPIAQWHHRGQQE
VFEYDLEDGSVIRATKDHKFMTSEGQMLAIDEIFERGLELKQVKRSQP (SEQ ID NO: 41)

VKIVRRKSLGIQTVYDIGVERDHNFLLANGLVASN (SEQ ID NO: 42)

>ScyPCC7437:/1-137 *Stanieria cyanosphaera* PCC 7437

CLSYDTEILTVEYGAMPIGKIVKEQIECNVYTVNQNGFIYPQAIAQWHERGKQE
IFEYTLDNGLVIRATKDHKFMTIDGQMLPIDEIFERGLELQRINDYSN (SEQ ID NO: 43)

VKIVSRKSLGKQPVYDIGVTKDHNFLLSNGVVASN (SEQ ID NO: 44)

>CspPCC6303:/1-137 *Calothrix sp.* PCC 6303

CLSYDTEILTWEYGFLKIGEIVEKQILCSVFSVDEQGNVYTQPIAQWHNRGLQE
LFAYQLEDGGVIRATKDHKFMTTDGQMLAIDEIFERQLDLFQVKGLPE (SEQ ID NO: 45)

VKIISRKVLKTENVYDIGLEGDHNFIIKDGLIASN (SEQ ID NO: 46)

>Cst/1-134 PCC7202: *Cyanobacterium stanieri* PCC 7202

CLSYDTEVLTVEYGVLPIGKIVEEQIQCTVYSVDQYGFVYTQAIAQWHDRGEQ
EVFEYELENGATIKATKDHKMMTSDGQMLPIDQIFEQGLDLFMVSF (SEQ ID NO: 47)

VKIVKRRSHGIQKVYDIGVAKDHNFLLHNGLVASN (SEQ ID NO: 48)

>CspATCC51142:/1-134 *Cyanothece sp.* ATCC 51142

CLSYDTEILTVEYGPMPIGKIVEENINCTVYTVDPNGFVYTQAIAQWHYRGEQE
IFEYYLEDGATIRATKDHKFMTMEGKMLPIDEIFENNLDLKQLTL (SEQ ID NO: 49)

VKIIGRQSLGVQKVYDIGVEKEHNFLLHNGLIASN (SEQ ID NO: 50)

>CspPCC8801:/1-134 *Cyanothece sp.* PCC 8801

CLSYDTEILTVEYGAIPIGKVVEENIDCTVYTVDKNGFVYTQNIAQWHLRGQQE
VFEYYLDDGSILRATKDHQFMTLEGEMLPIHEIFERGLELKKIKI (SEQ ID NO: 51)

VKIVSYRSLGKQFVYDIGVAQDHNFLLANGSIASN (SEQ ID NO: 52)

>Asp:/1-136 cromosoma 90 de *Anabaena sp.*

CLSYDTEILTVEYGFLEIGEIVEKQIECKVYTIDSNGMLYTQSIAQWHNRGQQE VYEYLLENGAIIRATKDHKFMTEAGQMLPIDEIFAQGLDLLQVGVAE (SEQ ID NO: 53)

VKIVSRTYVGQANVYDIGVESDHNFVIKNGFIASN (SEQ ID NO:54)

>Aha:/1-137 *Aphanothece halophytica*

CLSYDTEIWTVEYGAMPIGKIVEEKIECSVYTVDENGFVYTQPIAQWHPRGQQE IIEYTLEDGRKIRATKDHKMMTESGEMLPIEEIFQRELDLKVETFHEM (SEQ ID NO: 55)

VKIIKRQSLGRQNVYDVCVETDHNFVLANGCVASN (SEQ ID NO: 56)

>HspPCC7418:/1-137 *Halothece sp.* PCC 7418

CLSYDTEIWTVEYGAMPIGKIVEEKIECSVYTVDENGFVYTQPIAQWHPRGQQE IIEYTLEDGRKIRATKDHKMMTESGEMLPIEEIFQRELDLKVETFHEM (SEQ ID NO: 57)

VKIIKRQSLGRQNVYDIGVETDHNFVLANGCVASN (SEQ ID NO:58)

>CapPCC10605:/1-137 *Cyanobacterium aponinum* PCC 10605

CLSYDTEILTVEYGAISIGKIVEEKINCQVYSVDKNGFIYTQNIAQWHDRGSQEL FEYELEDGRIIKATKDHKMMTKDGQMLAINDIFEQELELYSVDDMGV (SEQ ID NO:59)

VKIVKRRSLGVQPVYDIGVEKDHNFILANGLVASN (SEQ ID NO:60)

>Cat:/1-133 aislado de *Candidatus Atelocyanobacterium thalassa*

CLSYDTKVLTVEYGPLPIGKVVQENIRCRVYTTNDQGLIYTQPIAQWHNRGKQ EIFEYHLDDKTIIRATKEHQFMTVDHVMMPIDEIFEQGLELKKIK (SEQ ID NO:61)

LKIIRRKSLGMHEVFDIGLEKDHNFVLSNGLIASN (SEQ ID NO:62)

>Oli:/1-137 precursor de DnaE 'Solar Lake' de *Oscillatoria limnetica*

CLSYNTEVLTVEYGPLPIGKIVDEQIHCRVYSVDENGFVYTQAIAQWHDRGYQ EIFAYELADGSVIRATKDHQFMTEDGQMFPIDEIWEKGLDLKKLPTVQD (SEQ ID NO:63)

VKIVRRQSLGVQNVYDIGVEKDHNFLLASGEIASN (SEQ ID NO:64)

>Cen:/1-137 Cianobacteria endosimbionte de *Epithemia turgida*

CLSYDTEVLTVEYGAIPIGRMVEESLDCTVYTVDKNGFVYTQSIQQWHSRGQQ EIFEYCFEDGSIIRATKDHKFMTAEGKMSSIHDIFEQGLELKKIIPWSG (SEQ ID NO:65)

AKIISCKSLGKQSVYDIGVVQDHNFLLANGVVASN (SEQ ID NO:66)

>SspPCC7502:/1-133 *Synechococcus sp.* PCC 7502

CLGYDTPVLTVEYGFMPIGKIVEEKIQCHVYSVDQNGLVFTQAIAQWHNRGQQ EVWEYNLDNGDIVRATKDHKFMTIDGQMLPINQIFEQGLELKVIA (SEQ ID NO:67)

VKIVSCKPLRVQTVYDIGVEKDHNFILDNGLVASN (SEQ ID NO:68)

>DsaPCC8305:/1-134 *Dactylococcopsis salina* PCC 8305

CLSYDTEVLTEEYGAIPIGKIVEERMNCHVYSVDENGFIYSQPIAQWHPRGEQE VVEYTLEDGKIIRATADHKMMTETGEMLPIEQIFQQQLDLKISNQ (SEQ ID NO:69)

VKIINRQSLGKQTVYDIGVEKDHNFILGNGLVASN (SEQ ID NO:70)

>CstPCC7417:/1-137 *Cylindrospermum stagnate* PCC 7417

CLSYDTEILTVEYGFIPIGEIVEKRIECSVYSVDNHGNVYTQPIAQWHNRGLQEV FEYCLEDGSTIRATKDHKFMTTDKEMLPIDEIFERGLDLLRVEGLPI (SEQ ID NO:71)

VKIIMRSYVGRENVYDIGVERDHNFVAKNGLIAAN (SEQ ID NO:72)

>SsPCC6803:/1-137 *Synechocystis sp.* PCC 6803

CLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEV LEYELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQ (SEQ ID NO:73)

VKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN (SEQ ID NO:74)

>GspPCC7407:/1-137 *Geitlerinema sp.* PCC 7407

CLSYETPVMTVEYGPLPIGRIVEEQLDCTVYSVDEQGHVYTQPVAQWHHRGL QEVVEYELEDGRRLRATADHRFMTETGEMLPLAEIFERGLELRQVALRVP (SEQ ID NO:75)

VKIVSRRSLGMQLVYDIGVAADHNFVLADGLIAAN (SEQ ID NO:76)

>SspPCC6714:/1-137 *Synechocystis sp.* PCC 6714

CLSFDAEILTVEYGPLSIGKIVGEEINCSVYSVDPQGRIYTQAIAQWHDRGVQEV FEYELEDGSVIRATPDHRFLTTDYELLAIEEIFARQMDLLTLTNLKL (SEQ ID NO:77)

VKVVRRRSLGMHRVFDIGLAQDHNFLLANGAIAAN (SEQ ID NO:78)

>MaePCC7806:/1-135 *Microcystis aeruginosa* PCC 7806

CLGGETLILTEEYGLLPIAKIVSEEVNCTVYSVDKNGFVYSQPISQWHERGLQE VFEYTLENGQTIQATKDHKFMTNDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO: 79)

VKIISRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:80)

>MaeNIES843:/1-135 ADN de *Microcystis aeruginosa* NIES-843 DNA

CLGGETLILTEEYGLLPIAKIVSEEINCTVYTVDQNGFVYSQPISQWHERGLQEV FEYTLENGQTIQATKDHKFMTSDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO:81)

VKIIGRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:82)

>AmaMBIC11017:/1-137 *Acaryochloris marina* MBIC11017,

CLSYDTPVLTLEYGWLPIGQVVQEQIECQVFSINERGHLYTQPIAQWHHRGQQ
EVFEYTLADGSTIQATAEHQFMTTDGQMYPVQQIFEEGLSLKQLPLPWQ (SEQ ID NO:83)

VKIIQRRSLGLQSVYDIGLAQDHNFVMANGWVAAN (SEQ ID NO:84)

>LspPCC7376:/1-137 *Leptolyngbya sp.* PCC 7376

CLDGETPIVTVEYGVLPIREIVEKELLCSVYSIDENGFVYTQPVEQWHQRGDRQ
MFEYQLDNGGVIRATPDHKFLTTEGEMVAIDEIFEKGLNLAEFAPADL (SEQ ID NO:85)

VKILRRHSIGKAKTYDIGVSKNHNFLLANGLFASN (SEQ ID NO:86)

>SelPCC6301:/1-137 *Synechococcus elongatus* PCC 6301

CLAADTEVLTVEYGPIAIGKLVEENIRCQVYCCNPDGYIYSQPIGQWHQRGEQE
VIEYELSDGRIIRATADHRFMTEEGEMLSLDEIFERSLELKQIPTPLL (SEQ ID NO:87)

VKIVRRRSLGVQPVYDLGVATVHNFVLANGLVASN (SEQ ID NO:88)

>SspPCC6312:/1-137 *Synechococcus sp.* PCC 6312

CLSADTELYTVEYGWLPIGRLVEEQIECQVLSVNAHGHVYSQPIAQWHRRAW
QEVFEYQLETGGTIKATTDHQFLTTDGQMYRIEDIFQRGLDLWQLPPDRF (SEQ ID NO:89)

VKIISRCSLGIQPVYDIGVAQDHNFVIRGGLVASN (SEQ ID NO:90)

>Tel:/1-137 ADN de *Thermosynechococcus elongatus* BP-1

CLSGETAVMTVEYGAVPIRRLVQERLSCHVYSLDGQGHLYTQPIAQWHFQGFR
PVYEYQLEDGSTICATPDHRFMTTRGQMLPIEQIFQEGLELWQVAIAPR (SEQ ID NO:91)

GKIVGRRLMGWQAVYDIGLAADHNFVLANGAIAAN (SEQ ID NO:92)

>Tsp:/1-137 genoma de *Thermosynechococcus sp.* NK55

CLSGETAVMTVEYGAVPIRRLVQERLTCHVYSLDAQGHLYTQPIAQWHFQGF
RPVYEYQLEDGSTIWATPDHRFMTTRGQMLPIEQIFQEGLELWQGPIAPS (SEQ ID NO:93)

CKIVGRQLVGWQAVYDIGVARDHNFLLANGAIAAN (SEQ ID NO:94)

>Tvu:/1-137 precursor de DnaE de *Thermosynechococcus vulcanus*

CLSGETAVMTVEYGAIPIRRLVQERLICQVYSLDPQGHLYTQPIAQWHFQGFRP
VYAYQLEDGSTICATPDHRFMTTSGQMLPIEQIFREGLELWQVAIAPP (SEQ ID NO:95)

CKIVGRRLVGWQAVYDIGLAGDHNFLLANGAIAAN (SEQ ID NO:96)

>SspPCC7002:/1-137 *Synechococcus sp.* PCC 7002

CLAGGTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYAQLIEQWHQRGDR
LLYEYELENGQMIRATPDHRFLTTTGELLPIDEIFTQNLDLAAWAVPDS (SEQ ID NO:97)

VKIIRRKFIGHAPTYDIGLSQDHNFLLGQGLIAAN (SEQ ID NO:98)

>ShoPCC7110:/1-136 *Scytonema hofmanni* PCC 7110 contig00136

CLSYDTEVLTAEYGFLPIGKIVEKAIECTVYSVDNDGNIYTQPIAQWHDRGQQE
VFEYSLDDGSVIRATKDHKFMTTGGQMLPIDEIFERGLDLMRIDSLP (SEQ ID NO:99)

VKILTRKSIGKQTVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:100)

>WinUHHT291/1-136 *Westiella intricata* UH HT-29-1

CLSYDTEILTVEYGFLPIGEIVEKRIECTVYTVDTNGYVYTQAIAQWHNRGEQE
VFEYALEDGSIIRATKDHKFMTSEGQMLPIDEIFVKGLDLLQVQGLP (SEQ ID NO: 101)

VKIITRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:102)

>FspPCC9605:/1-136 *Fischerella sp.* PCC 9605 FIS9605DRAFT

CLSYDTEILTVEYGFLPIGEIVEKGIECTVYTVDNNGNVYTQTIAQWHNRGQQE
VFEYCLEDGSVIRATKDHKFMTTDGQMLPIDEIFARGLDLLQVKNLP (SEQ ID NO:103)

VKIVTRRPLGTQNVYDIGVESDHNFVIKNGLVASN (SEQ ID NO:104)

>MrePCC10914:/1-137 *Mastigocladopsis repens* PCC 10914

CLSYDTEVLTVEYGFLPIGEIVEKSIECSVYTVDSNGNVYTQPIAQWHNRGQQE
VFEYCLEDGSIIRATKDHKFMTIHGQMLPIDEIFERGLELMKIQGLPE (SEQ ID NO:105)

AKIITRKSLGTQNVYDIGVERDHNFVTRDGFIASN (SEQ ID NO:106)

>ShoUTEX2349:/1-137 [*Scytonema hofmanni*] UTEX 2349

CLSYNSEVLTVEYGFLPIGKIVEKGIECSVYSVDSYGKIYTQVIAQWHNRGQQE
VFEYCLEDGTIIQATKDHKFMTVDGQMLPIDEIFERGLDLMQVQGLPD (SEQ ID NO:107)

VKIITRKSLGTQNVYDIGVSSDHNFVMKNGLIASN (SEQ ID NO:108)

>AspPCC7108:/1-137 *Anabaena sp.* PCC 7108 Ana7108scaffold_2_Cont3

CLSSDTEVLTVEYGLIPIGEIIEKRIDCSVFSVDKNGNIYTQPIAQWHDRGIQELY
EYCLDDGSTIRATKDHKFMTTAGEMLPIDEIFERGLDLLKVHNLPQ (SEQ ID NO:109)

VKIITRNYVGKENVYDIGVERDHNFAIKNGLIASN (SEQ ID NO:110)

>FspPCC9339:/1-137 *Fischerella sp.* PCC 9339 PCC9339DRAFT

CLSYDTEVLTVEYGFLPIGEIVEKRIECTVYTVDHNGYVYTQPIAQWHNRGYQ
EVFEYGLEDGSVIRATKDHKFMTSEGQMLPIDEIFARELDLLQVTGLVN (SEQ ID NO:111)

VKIVTRRLLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:112)

>Csp336:/1-137 *Calothrix sp.* 336/3

CLSYDTEIFTVEYGFLPIGEIVEKRLECTVLTVDNHGNIYSQPIAQWHHRGQQQI
YEYGLEDGSVIRATKDHKFMTTDGQMLPIDEIFERGLDLLQVTNLDN (SEQ ID NO:113)

VKVITRKLADTENVYDIGVENHHNFLIKNGLVASN (SEQ ID NO:114)

>FthPCC7521:/1-136 *Fischerella thermalis* PCC 7521

CLSYETEILTVEYGFLPIGEIVEKRIECSVYTVDNNGYVCTQPIAQWHNRGYQE
VFEYGLEDGSVIRATKDHKFMTIDRQMLPIDEIFARGLDLLQVTGLP (SEQ ID NO:115)

VKIITRKSLGTQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:116)

>CyaPCC7702/1-137 *Cyanobacterium* PCC 7702 Chl7702

CLSYDTEILTVEYGFLSIGEIVEKEIECTVYTVDSNGYIYTQPIAQWHEQGEQEIF
EYSLEDGSTIRATKDHKFMTIEGEMLPIDQIFARQLDLMQITGLPQ (SEQ ID NO:117)

VKISTKKSLGKQKVYDIGVVRDHNFIIKNGFVASN (SEQ ID NO:118)

>FspPCC943171-136 *Fischerella sp.* PCC 9431

CLSYDTEVLTVEYGFLPIGEIVEKRIECTVYTVDTNGYVYTQAIAQWHNRDEQE
VFEYALEDGSIIRATKDHKFMTSEGQMLPIDEIFAKGLDLLQVQGLP (SEQ ID NO:119)

VKIVTRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:120)

>FmuPCC7414:/1-137 *Fischerella muscicola* PCC 7414

CLSYETEILTVEYGFLPIGEIVEKRIECSVYTVDNNGYVCTQTIAQWHNRGYQE
VFEYGLEDGSVIRATKDHKFMTIDRQMLPIDEIFARGLDLLQVKGLPE (SEQ ID NO:121)

VKIITRQSLGTQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:122)

>FmuPCC73103:/1-137 *Fischerella muscicola* SAG 1427-1 = PCC 73103

CLSYDTEVLTVEYGFLPIGEIVEKTIECNVFTVDSNGYVYTQPIAQWHNRGYQE
VFEYGLEDGSVIRATKDHKFMTSEGKMLPIDEIFARELDLLQVTGLIN (SEQ ID NO:123)

VKIVTRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:124)

>Lae:/1-137 *Lyngbya aestuarii* BL J laest3.contig.3

CLSYDTEILTVEYGAIPIGKVVDEKIECTVYSVDKNGLIYTQPIAQWHNRGKQE
VFEYSLEDGSTIRATKDHKFMTMDNQMLPIDEILEKGLELKQVNADSV (SEQ ID NO:125)

VKIVSRKSLDSQTVYDIGVETDHNFLLANGSVASN (SEQ ID NO:126)

>MspPCC7126:/1-135 *Microchaete sp.* PCC 7126

CLSYKTQVLTVEYGLLAIGEIVEKNIECSVFSVDIHGNVYTQPIAQWHHRGQQE
VFEYGLEDGSIIRATKDHKFMTTQGEMLPIDEIFARGLDLLQVKGV (SEQ ID NO:127)

VKIITRKYIGKENVYDIGVEQDHNFAIKNGLIAAN (SEQ ID NO:128)

>Lsp:/1-137 *Leptolyngbya sp.* JSC-1

CLSYDTEILTVEYGALPIGKIVENQMICSVYSIDNNGYIYIQPIAQWHNRGQQEV
FEYILEDGSIIRSTKDHKFMTKGGEMLPIDEIFERGLELAQVTRLEQ (SEQ ID NO:129)

VKIISRRSVGVQSVYDIGVKQDHNFFLRNGLIASN (SEQ ID NO:130)

>CwaWH8501:/1-137 *Crocosphaera watsonii* WH8501

CLSYDTEILTVEYGAMYIGKIVEENINCTVYTVDKNGFVYTQTIAQWHNRGEQ EIFEYDLEDGSKIKATKDHKFMTIDGEMLPIDEIFEKNLDLKQVVSHPD (SEQ ID NO:131)

VKIIGCRSLGTQKVYDIGVEKDHNFLLANGSIASN (SEQ ID NO:132)

>CchPCC7420:/1-135 *Coleofasciculus chthonoplastes* PCC 7420 (Mcht)

CLSYDTQILTVEYGAVAIGEIVEKQIECTVYSVDENGYVYTQPIAQWHNRGEQE VFEYLLEDGATIRATKDHKFMTDEDQMLPIDQIFEQGLELKQVEVL (SEQ ID NO:133)

VKIIGRKPLGTQPVYDIGVERDHNFLLFNGSVASN (SEQ ID NO:134)

>CspPCC6712/1-133

CLSYDTEVLTVEYGAIPIGKIVEEKIACNVYSVDKNGFVYTQPIAQYHDRGIQE VFEYRLENGSVIRATKDHKMMTADGQMLPIDEIFKQNLDLKQLN (SEQ ID NO:135)

VKIISRQSLGKQSVFDIGVAKDHNFLLANGLVASN (SEQ ID NO:136)

>AflNIES81:/1-132 *Aphanizomenon flos-aquae* NIES-81

CLSYDTEILTVEYGFLQIGEIVEKQIECKVYTVDSNGILYTQSIAQWHNRGQQEV YEYLLENGAIIRATKDHKFMTEEGQMLPIDEIFSQGLDLLQV (SEQ ID NO:137)

VKIISRTYVGQANVYDIGVENDHNFVIKNGFIAAN (SEQ ID NO:138)

>Rbr:/1-137 *Raphidiopsis brookii* D9 D9_5,

CLSYETEVLTLEYGFLPIGEIVDKQMVCTVFSVNDSGNVYTQPIGQWHDRGVQ ELYEYCLDDGSTIRATKDHKFMTTQGEMVPIDEIFHQGWELVQVSGTMN (SEQ ID NO:139)

VKIVSRRYLGKADVYDIGVAKDHNFIIKNGLVASN (SEQ ID NO:140)

>CspCCy0110:/1-134 *Cyanothece sp.* CCY0110 1101676644604

CLSYDTEILTVEYGPMPIGKIVEENINCSVYTVNKNGFVYTQSIAQWHHRGEQE VFEYYLEDGETIRATKDHKFMTTEGKMLPIDEIFENNLDLKKLTV (SEQ ID NO:141)

VKIIERRSLGKQNVYDIGVEKDHNFLLSNNLIASN (SEQ ID NO:142)

>XspPCC7305:/1-135 *Xenococcus sp.* PCC 7305

CLSADTEVLTVEYGAISIGKIVEERIECTVYSVDANGFVYTQEIAQWHNRGEQE VFEYMLDDGSVIRATKDHKLMTIDGQMVAIDEIFSQGLELKQVLGL (SEQ ID NO:143)

VKIVSRKSLGTQTVYDLGVARDHNFLLANGTVASN (SEQ ID NO:144)

>PspPCC7319:/1-135 Pleurocapsa sp. PCC 7319

CLSYDTEIYTVEYGALPIGKIVESRIKCTVLTVDKNGLVYSQPIVQWHDRGIQEV FEYTLDNGATIRATKDHKFMTVEGQMLPIDEIFELGLELKEIQQF (SEQ ID NO:145)

VKIISRQSLGKQSVYDIGVAKDHNFLLANGMVASN (SEQ ID NO:146)

>CraCS505:/1-137 *Cylindrospermopsis raciborskii* CS-505

CLSYETEVLTLEYGFVPIGEIVNKQMVCTVFSLNDSGNVYTQPIGQWHDRGVQ DLYEYCLDDGSTIRATKDHKFMTTQGEMVPIDEIFHQGWELVQVSGISK (SEQ ID NO:147)

VKIVSRRYLGKADVYDIGVAKDHNFIIKNGLVASN (SEQ ID NO:148)

>SmaPCC6313/1-129 *Spirulina major* PCC 6313

CLTYDTLVLTVEYGPVPIGKLVEAQINCQVYSVDANGFIYTQAIAQWHDRGQR QVYEYTLEDGSTIRATPDHKFMTATGEMLPIDQIFEQGLDL (SEQ ID NO:149)

VKIIHRRALPPQSVYDIGVERDHNFLLPSGWVASN (SEQ ID NO:150)

>SsuPCC9445:/1-131 *Spirulina subsalsa* PCC 9445

CLSYDTKIITVEYGAIAIGTIVEQGLHCHVYSVDPNGFIYTQPIAQWHQRGEQEV FAYTLENGSIIQATKDHKFMTQQGKMLPIDTIFEQGLDLLQ (SEQ ID NO:151)

VKIIKRTSLGVRPVYDIGVIQDHNFLLENGLVASN (SEQ ID NO:152)

>MaePCC9807:/1-135 *Microcystis aeruginosa* 9807

CLGGETLILTEEYGLLPIAKIVSEEINCTVYSVDKNGFIYSQPISQWHERGLQEVF EYTLENGQTIQATKDHKFMTSDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO:153)

VKIISRQFLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:154)

>MspGI1:/1-130 *Myxosarcina sp.* GI1 contig_13

CLSYDTEVLTLKYGALPIGEIVEKRINCHVYTRAESGFFYIQSIEQWHDRGEQEV FEYTLENGATIKATKDHKFMTSGGQMLPIDEIFERGLDLL (SEQ ID NO:155)

VKIVSRKSLGKQPVYDLGVAKDHNFLLANGTVASN (SEQ ID NO:156)

>LspPCC6406:/1-136 *Leptolyngbya sp.* PCC 6406

CLSADTQLLTVEYGPLEIGRIVEEQIACHVYSVDANGFVYTQPIAQWHSRGEQE IFEYQLEDGRTLRATADHKFMTTTGEMGRINDIFEQGLDLKQIDLPQ (SEQ ID NO:157)

VKVVSRQSLGVQPVYDIGVATDHNFLLADGLVASN (SEQ ID NO:158)

>AspCCMEE5410:/1-132 *Acaryochloris sp.* CCMEE 5410

CLSYDTPVLTLEYGWLPIGQVVQEQIECQVFSINERGHLYTQPIAQWHHRGQQ EVFEYTLTDGSTIQATAEHQFMTTDGQMYPIQQIFEEGLSLKQL (SEQ ID NO:159)

VKITQRRSLGLQSVYDIGLAQDHNFVIANGWVAAN (SEQ ID NO:160)

>GhePCC6308:/1-133 *Geminocystis herdmanii* PCC 6308

CLSYDTEVLTVEFGAIPMGKIVEERLNCQVYSVDKNGFIYTQNIAQWHDRGVQ EVFEYELEDGRIIKATKDHKMMIENCEMVEIDRIFEEGLELFEVN (SEQ ID NO:161)

VKILKRRSISSQQVYDIGVEKDHNFLLANGLVASN (SEQ ID NO:162)

>NnoPCC7104:/1-133 *Nodosilinea nodulosa* PCC 7104

CLSADTELLTLEYGPLTIGEIVAKRIPCHVFSVDESGYVYTQPVAQWHQRGHQE VFEYQLDDGTTIRATADHQFMTELGEMMAIDEIFQRGLELKQVE (SEQ ID NO: 163)

VKIISRQSLGVQPVYDIGVARDHNFLLADGQVASN (SEQ ID NO:164)

>RlaKORDI51-271-137 *Rubidibacter lacunae* KORDI 51-2

CLSYDTEVLTVEYGPLAIGTIVSERLACTVYTVDRSGFLYAQAISQWHERGRQD VFEYALDNGMTIRATKDHKLMTADGQMVAIDDIFTQGLTLKAIDTAAF (SEQ ID NO:165)

MKIVSRKSLGVQHVYDIGVARDHNFLLANGAIASN (SEQ ID NO:166)

>CfrPCC9212/1-136 *Chlorogloeopsis fritschii* PCC 9212

CLSYDTAILTVEYGFLPIGEIVEKGIECTVYTVDSNGYIYTQPIAQWHNRGEQEL FEYSLEDGSIIRATKDHKFMTIDGQMLPIDEIFARKLELMQVKGLP (SEQ ID NO:167)

VKIIAKKSLGTQNVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:168)

>RinHH01:/1-137 *Richelia intracellularis* HH01 WGS project

CLSYDTQILTVEHGPMSIGEIVEKCLECHVYTVNKNGNICIQTITQWHFRGEQEI FEYELEDGSFIQATKDHKFMTTTGEMLPIHEIFTNGLEILQLSKSLL (SEQ ID NO:169)

VKILARKSLGTQKVYDIGVNDDHNFALSNSFIASN (SEQ ID NO:170)

>SspPCC7117/1-137

CLAGDTPVVTVEYGVLPIQTIVEQELLCQVYSVDAQGLIYTQPIEQWHNRGDR LLYEYELENGQMIRATPDHKFLTTTGELLPIDEIFTQNLDLAAWAVPDS (SEQ ID NO:171)

VKIIRRKFIGHAPTYDIGLSQDHNFLLGQGLIAAN (SEQ ID NO:172)

>SspPCC8807/1-137

CLAGDTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYTQPIEQWHQRGDRF LYEYELENGQMIRATPDHKFLTTTGKLLPIDEIFTQNLDLAAWAVPDS (SEQ ID NO:173)

VKIIRRKFIGHAPTYDIGLSQDHNFLLGQGFIAAN (SEQ ID NO:174)

>SspNKBG042902:/1-137 *Synechococcus sp.* NKBG 042902

CLAGDTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYTQPIEQWHQRGDR LLYEYELENGQMIRATPDHKFLTTTGELLPIDEIFTQNLDLAAWAVPDS (SEQ ID NO:175)

VKILRRKFIGRAPTYDIGLSQDHNFLLGQGLVAAN (SEQ ID NO:176)

>SspNKBG15041:/1-129 *Synechococcus sp.* NKBG15041

CLAGDTPVVTVEYGVLPIRTIVDQELLCHVYSLDPQGFIYAQPVEQWHRRGDR
LLYEYELETGAVIRATPDHKFLTATGEMLPIDEIFVRNLDL (SEQ ID NO:177)

VKIIRRNLIGEAATYDIGLGKDHNFLLGQGLIASN (SEQ ID NO:178)

>SspPCC73109/1-130

CLAGGTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYTQPIEQWHQRGDR
LLYEYELENGQMIRATPDHKFLTTTGELLPIDEIFTQNLDLL (SEQ ID NO:179)

VKIIRRKFIGHAPTYDIGLSQDHNFLLGQGLIAAN (SEQ ID NO:180)

>SspPCC7003/1-130

CLAGDTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYTQPIEQWHKRGDR
LLYEYELENGQIIRATPDHKFLTTTGEMRPIDEIFAKNLSLL (SEQ ID NO: 181)

VKIIRRKFVGHAPTYDIGLSQDHNFLLGQGLIAAN (SEQ ID NO:182)

>CspPCC8802/1-134: *Cyanothece sp.* PCC 8802

CLSYDTEILTVEYGAIPIGKVVEENIDCTVYTVDKNGFVYTQNIAQWHLRGQQE
VFEYYLDDGSILRATKDHQFMTLEGEMLPIHEIFERGLELKKIKI (SEQ ID NO:183)

VKIVSYRSLGKQFVYDIGVAQDHNFLLANGSIASN (SEQ ID NO:184)

>SelPCC7942:/1-137 *Synechococcus elongatus* PCC 7942

CLAADTEVLTVEYGPIAIGKLVEENIRCQVYCCNPDGYIYSQPIGQWHQRGEQE
VIEYELSDGRIIRATADHRFMTEEGEMLSLDEIFERSLELKQIPTPLL (SEQ ID NO:185)

VKIVRRRSLGVQPVYDLGVATVHNFVLANGLVASN (SEQ ID NO:186)

>CfrPCC6912:/1-137 *Chlorogloeposis fritschii* PCC 6912

CLSYDTAILTVEYGFLPIGEIVEKGIECTVYTVDSNGYIYTQPIAQWHNRGEQEL
FEYSLEDGSIIRATKDHKFMTIDGQMLPIDEIFARKLELMQVKGLPE (SEQ ID NO:187)

VKIIAKKSLGTQNVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:188)

>CspATC51472:/1-132 *Cyanothece sp.* ATCC 51472

CLSYDTEILTVEYGPMPIGKIVEENINCTVYTVDPNGFVYTQAIAQWHYRGEQE
IFEYYLEDGATIRATKDHKFMTMEGKMLPIDEIFENNLDLKQL (SEQ ID NO:189)

VKIIGRQSLGVQKVYDIGVEKEHNFLLHNGLIASN (SEQ ID NO:190)

>Lma:/1-132 *Lyngbya majuscula*

CLSYDTEIITVEYGPIAIGEIVEKGIPCTVYSVDSNGYVYTQPIAQWHNRGEQEV
FEYTLDDGSVIRATKDHKFMTIDGQMLPIDEIFEGGLELKQL (SEQ ID NO:191)

VKIISRKSLGTQPVYDIGVKDDHNFILANGMVASN (SEQ ID NO:192)

>CspESFC/1-137

CLSYDTEVLTVEYGAVPIGKLVEEKLNCSVYTVDPNGYIYTQAIAQWHDRGIQ EVFEYQLEDNTIIRATKDHKFMTEDHQMLPIDEIFERGLELKKCPQPQQ (SEQ ID NO:193)

VKIIRRRSLGFQPVYDIGLEQDHNFLLNQGAIASN (SEQ ID NO:194)

>SspPCC7002:/1-129 *Synechococcus sp.* PCC 7002

CLAGGTPVVTVEYGVLPIQTIVEQELLCHVYSVDAQGLIYAQLIEQWHQRGDR LLYEYELENGQMIRATPDHRFLTTTGELLPIDEIFTQNLDL (SEQ ID NO:195)

VKIIRRKFIGHAPTYDIGLSQDHNFLLGQGLIAAN (SEQ ID NO:196)

>AmaMBIC11017:/1-132 *Acaryochloris marina* MBIC11017

CLSYDTPVLTLEYGWLPIGQVVQEQIECQVFSINERGHLYTQPIAQWHHRGQQ EVFEYTLADGSTIQATAEHQFMTTDGQMYPVQQIFEEGLSLKQL (SEQ ID NO:197)

VKIIQRRSLGLQSVYDIGLAQDHNFVMANGWVAAN (SEQ ID NO:198)

>Mae905:/1-129 *Microcystis aeruginosa* DIANCHI905

CLGGETLILTEEYGLLPIAKIVSEEVNCTVYSVDKNGFVYSQPISQWHERGLQE VFEYTLENGQTIQATKDHKFMTNDGEMLAIDTIFERGLDL (SEQ ID NO:199)

VKIISRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:200)

>AciAWQC310F:/1-125 AWQC: *Anabaena circinalis* AWQC310F

CLSYDTEILTVEYGFLEIGEIVEKQIECKVYTVDSNGILYTQPIAQWHHRGQQEV YEYLLENGAIIRATKDHKFMTEAGEMLPIDDIFTQ (SEQ ID NO:201)

VKIISRTYVGQANVYDIGVENDHNFVIKNGFVAAN (SEQ ID NO:202)

>AciAWQC131C:/1-125 *Anabaena circinalis* AWQC131C

CLSYDTEILTVEYGFLEIGEIVEKQIECRVYTVDSNGILYTQPIAQWHYRGQQEV YEYLLENGAIIRATKDHNFMTEAGEMLPIDDIFTQ (SEQ ID NO: 203)

IKIISRKYVGQANVYDIGVENDHNFVIKNGFVAAN (SEQ ID NO: 204)

>CspUCYN:/1-124 *Cyanobacterium sp.* UCYN-A2

CLSYDTKVLTVEYGPLPIGKVVQENIRCRVYTTNDQGLIYTQPIAQWHNRGKQ EIFEYHLDDKTIIRATKEHQFMTVDHVMMPIDEIFEQ (SEQ ID NO:205)

KIIRRKSLGMHEVFDIGLEKDHNFVLSNGLIASN (SEQ ID NO:206)

>Pst:/1-129 *Planktothrix* st147: st147_cleanDRAFT_c6

CLSYDTEVLTVEYGLIPISKIVEEKIECTVYTVNNQGYVYTQPIAQWHNRGEQE VFEYYLEDGSVIRATKDHKFMTVEGQMLPIDEIFEKELDL (SEQ ID NO:207)

VKIISRKSLGTQPVYDIGVQEDHNFVLNNGLVASN (SEQ ID NO:208)

>PlaCYA98/1-129: *Planktothrix* NIVA-CYA 98

CLSYDTEILTVEYGLMPIGKIVKEKIECTVYTVNNQGYVYTQPIAQWHHRGEQ EVFEYCLEDGSVIRATKDHKFMTVQGQMLPIDEIFEKELDL (SEQ ID NO:209)

VKIISRKSLGTQPVYDIGVQEDHNFLLNNGLVASN (SEQ ID NO:210)

>FdiUTEX481:/1-137 *Fremyella diplosiphon* UTEX 481

CLSYDTEVLTVEYGLIPIGEIVEKRLECSVYSVDINGNVYTQPIAQWHHRGQQE VFEYALEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLLQVPHLPE (SEQ ID NO:211)

VKIVTRRAIGAANVYDIGVEQDHNFAIKNGLIAAN (SEQ ID NO:212)

> Pst585:/1-129 *Planktothrix sp.* 585: Longitud=1586997

CLSYDTEILTVEYGLIPISKIVEEKIECTVYTVNNQGYVYTQPIAQWHNRGEQEV FEYYLEDGSVIRATKDHKFMTVDGQMLPIDEIFEKELDL (SEQ ID NO:213)

VKIISRKSLGTQPVYDIGVQEDHNFVLNNGLVASN (SEQ ID NO:214)

>NpuPCC73102/1-137

CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQE VFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO:215)

IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO:216)

>CthPCC7203:/1-137 *Chroococcidiopsis thermalis* PCC 7203

CLSYDTEILTVEYGAIPIGKIVEERIECTVYSVDNNGFIYTQPIAQWHNRGQQEV FEYCLEDGSIIRATKDHKFMTFEGKMLPIDEIFEQELDLKQVKSIQN (SEQ ID NO:217)

VKIISRKSLGIQPVYDIGVERDHKFVLKNGLVASN (SEQ ID NO:218)

>NspCCY9414:/1-137 genoma de *Nodularia spumigena* CCY9414

CLSYDTEILTVEYGYIPIGEIVEKAIECSVYSVDNNGNVYTQPIAQWHNRGEQE VFEYSLEDGSTIRATKDHKFMTTDGQMLPIDEIFAQELDLLQVHGLPK (SEQ ID NO:219)

VKITARKFVGRENVYDIGVERYHNFAIKNGLIASN (SEQ ID NO:220)

>AcyPCC7122:/1-137 *Anabaena cylindrica* PCC 7122

CLSYDTEVLTVEYGFIPIGEIVEKRIECSIFSVDKNGNVYTQPIAQWHNRGRQEI YEYCLDDGSKIRATKDHKFMTTAGEMLPIDEIFERDLDLLKVEGLPE (SEQ ID NO:221)

VKIISRQYLGQADVYDIGVEEDHNFAIKNGFIASN (SEQ ID NO:222)

>CspPCC7507:/1-137 *Calothrix sp.* PCC 7507, genoma completo

CLSYDTEVLTVEYGLLPIGEIVEKGIECRVFSVDNHGNVYTQPIAQWHNRGQQE VFEYGLDDGSVIRATKDHKFMTTDGKMLPIDEIFERGLDLLQVQGLPE (SEQ ID NO:223)

VKVITRKYIGKENVYDIGVELDHNFAIRNGLVASN (SEQ ID NO:224)

>NspPCC7524:/1-137 *Nostoc sp.* PCC 7524

CLSYDTEILTVEYGFLPIGEIVEKGIECTVFSVASNGIVYTQPIAQWHNRGQQEIF EYCLEDGSIIRATKDHKFMTQDGQMLPIDEIFACELDLLQVQGLPE (SEQ ID NO:225)

VKVVTRKYIGKENVYDIGVERDHNFVIRNGLVASN (SEQ ID NO:226)

>Naz0708:/1-137 '*Nostoc azollae*' 0708

CLSYKTEVLTVEYGLIPIGEIVEKRIECSLFSVDENGNIYTQPIAQWHHRGVQEV YEYCLDDGTIIRATKDHKFMTTIGEMLPIDEIFERDLNLLQVNGLPT (SEQ ID NO:227)

VKIISRQFLGPANVYDIGVAQDHNFAIKNGLIASN (SEQ ID NO:228)

>NspPCC7120:/1-137 ADN de *Nostoc sp.* PCC 7120

CLSYDTEVLTVEYGFVPIGEIVEKGIECSVFSINNNGIVYTQPIAQWHHRGKQEV FEYCLEDGSIIKATKDHKFMTQDGKMLPIDEIFEQELDLLQVKGLPE (SEQ ID NO:229)

IKIASRKFLGVENVYDIGVRRDHNFFIKNGLIASN (SEQ ID NO:230)

>AvaATCC29413/1-137 *Anabaena variabilis* ATCC 29413

CLSYDTEVLTVEYGFVPIGEIVDKGIECSVFSIDSNGIVYTQPIAQWHHRGKQEV FEYCLEDGSIIKATKDHKFMTQDGKMLPIDEIFEQELDLLQVKGLPE (SEQ ID NO:231)

IKIASRKFLGVENVYDIGVGRDHNFFVKNGLIASN (SEQ ID NO:232)

>PspPCC7327:/1-135 *Pleurocapsa sp.* PCC 7327.

CLSYDTKILTVEYGAMPIGKIVEEQIDCTVYTVNQNGFVYTQPIAQWHDRGKQ EIFEYCLEDGSIIRATKDHKFMTTDGQMLPIDKIFEKGLDLKTINCD (SEQ ID NO: 233)

VKILSRKSLGIQSVYDIGVEKDHNFLLANGLVASN (SEQ ID NO:234)

>CspPCC7424:/1-135 *Cyanothece sp.* PCC 7424

CLSYETQIMTVEYGLMPIGKIVEEQIDCTVYTVNKNGFVYTQPIAQWHYRGEQ EVFEYCLEDGSTIRATKDHKFMTTDGQMLPIDEIFEQGLELKQIHLS (SEQ ID NO:235)

VKIISRQSLGIQPVYDIGVEKDHNFLISDGLIASN (SEQ ID NO:236)

>CspPCC7822:/1-134 *Cyanothece sp.* PCC 7822

CLSYDTEILTVEYGPMPIGKIVEEQIECTVYTVDKNGLVYTQPIAQWHHRGQQE VFEYCLEDGSIIRATKDHKFMTDDGQMLPIEEIFEKGLELKQIIL (SEQ ID NO:237)

VKIISRQLAGNQTVYDLGVEKDHNFLLANGLIASN (SEQ ID NO:238)

>NspPCC7107:/1-137 *Nostoc sp.* PCC 7107

CLSYDTQVLTVEYGLVPIGEIVEKQLECSVFTIDGHGYVYTQAIAQWHNRGQQ EVFEYGLEDGSVIRATKDHKFMTTDGQMLPIDEIFERELDLLQVQGLRW (SEQ ID NO:239)

VKIITRKYIGQANVYDIGVAQDHNFVIENRLIASN (SEQ ID NO:240)

>Tbolicbl/1-136 *Tolypothrix bouteillei* licb1

CLSYDTEILTVEYGFLPIGKIVEKGIECNVYSVDKNGNIYTQPIAQWHDRGEQE
VFEYCLENGSVIRATKDHKFMTTSGEMLPIDEIFERGLDLIRVEDLP (SEQ ID NO:241)

VKILTRKSIGKQTVYDIGVERDHNFVIKNGSVASN (SEQ ID NO:242)

>Aov:/1-136 gen de precursor de DnaE (dnaE) de *Aphanizomenon ovalisporum*

CLSADTEILTVEYGFLPIGEIVGKAIECRVYSVDGNGNIYTQSIAQWHNRGEQEV
FEYTLEDGSIIRATKDHKFMTTDGEMLPIDEXFARQLDLMQVQGLH (SEQ ID NO:243)

VKITARKFVGRENVYDIGVEHHHNFAIKNGLIASN (SEQ ID NO:244)

>OnvPCC7112:/1-137 *Oscillatoria nigro-viridis* PCC 7112

CLSYDTKILTVEYGPMAIGKIVEEKIECTVYSVDSNGYIYTQSIAQWHRRGQQE
VFEYCLEDGSIIRATKDHKFMTVGGQMLPIDEIFEQGLDLKQINSSSD (SEQ ID NO:245)

VKIISRKSLGTQEVYDIGVEREHNFILENSLVASN (SEQ ID NO:246)

>RspPCC7116:/1-135 *Rivularia sp.* PCC 7116, genoma completo

CLSYDTEVLTEEFGLIPIGKIVEEKIDCTVYSVDVNGNVYSQPIAQWHNRGMQE
VFEYELEDGSTIRATKDHKFMTVDGEMLAIDEIFEKGLELKRVGIY (SEQ ID NO:247)

VKIISRKVLKTENVYDIGLEGDHNFIIKDGLIASN (SEQ ID NO:248)

>MspPCC7113:/1-137 *Microcoleus sp.* PCC 7113,

CLSYDSEILTVEYGLMPIGKIVEEGIECTVYSVDSHGYLYTQPIAQWHHRGQQE
VFEYDLEDGSVIRATKDHKFMTSEGQMLAIDEIFERGLELKQVKRSQP (SEQ ID NO:249)

VKIVRRKSLGIQTVYDIGVERDHNFLLANGLVASN (SEQ ID NO:250)

>ScyPCC7437:/1-137 *Stanieria cyanosphaera* PCC 7437

CLSYDTEILTVEYGAMPIGKIVKEQIECNVYTVNQNGFIYPQAIAQWHERGKQE
IFEYTLDNGLVIRATKDHKFMTIDGQMLPIDEIFERGLELQRINDYSN (SEQ ID NO:251)

VKIVSRKSLGKQPVYDIGVTKDHNFLLSNGVVASN (SEQ ID NO:252)

>CspPCC6303:/1-137 *Calothrix sp.* PCC 6303

CLSYDTEILTWEYGFLKIGEIVEKQILCSVFSVDEQGNVYTQPIAQWHNRGLQE
LFAYQLEDGGVIRATKDHKFMTTDGQMLAIDEIFERQLDLFQVKGLPE (SEQ ID NO:253)

VKIISRKVLKTENVYDIGLEGDHNFIIKDGLIASN (SEQ ID NO:254)

>Cst:/1-134 PCC7202: *Cyanobacterium stanieri* PCC 7202

CLSYDTEVLTVEYGVLPIGKIVEEQIQCTVYSVDQYGFVYTQAIAQWHDRGEQ
EVFEYELENGATIKATKDHKMMTSDGQMLPIDQIFEQGLDLFMVSF (SEQ ID NO:255)

VKIVKRRSHGIQKVYDIGVAKDHNFLLHNGLVASN (SEQ ID NO:256)

>CspATCC51142:/1-134 *Cyanothece sp.* ATCC 51142

CLSYDTEILTVEYGPMPIGKIVEENINCTVYTVDPNGFVYTQAIAQWHYRGEQE IFEYYLEDGATIRATKDHKFMTMEGKMLPIDEIFENNLDLKQLTL (SEQ ID NO:257)

VKIIGRQSLGVQKVYDIGVEKEHNFLLHNGLIASN (SEQ ID NO:258)

>CspPCC8801:/1-134 *Cyanothece sp.* PCC 8801

CLSYDTEILTVEYGAIPIGKVVEENIDCTVYTVDKNGFVYTQNIAQWHLRGQQE VFEYYLDDGSILRATKDHQFMTLEGEMLPIHEIFERGLELKKIKI (SEQ ID NO:259)

VKIVSYRSLGKQFVYDIGVAQDHNFLLANGSIASN (SEQ ID NO:260)

>Asp:/1-136 cromosoma 90 de *Anabaena sp.*

CLSYDTEILTVEYGFLEIGEIVEKQIECKVYTIDSNGMLYTQSIAQWHNRGQQE VYEYLLENGAIIRATKDHKFMTEAGQMLPIDEIFAQGLDLLQVGVAE (SEQ ID NO:261)

VKIVSRTYVGQANVYDIGVESDHNFVIKNGFIASN (SEQ ID NO:262)

>Aha:/1-137 *Aphanothece halophytica*

CLSYDTEIWTVEYGAMPIGKIVEEKIECSVYTVDENGFVYTQPIAQWHPRGQQE IIEYTLEDGRKIRATKDHKMMTESGEMLPIEEIFQRELDLKVETFHEM (SEQ ID NO:263)

VKIIKRQSLGRQNVYDVCVETDHNFVLANGCVASN (SEQ ID NO:264)

>HspPCC7418:/1-137 *Halothece sp.* PCC 7418

CLSYDTEIWTVEYGAMPIGKIVEEKIECSVYTVDENGFVYTQPIAQWHPRGQQE IIEYTLEDGRKIRATKDHKMMTESGEMLPIEEIFQRELDLKVETFHEM (SEQ ID NO:265)

VKIIKRQSLGRQNVYDIGVETDHNFVLANGCVASN (SEQ ID NO:266)

>CapPCC10605:/1-137 *Cyanobacterium aponinum* PCC 10605

CLSYDTEILTVEYGAISIGKIVEEKINCQVYSVDKNGFIYTQNIAQWHDRGSQEL FEYELEDGRIIKATKDHKMMTKDGQMLAINDIFEQELELYSVDDMGV (SEQ ID NO:267)

VKIVKRRSLGVQPVYDIGVEKDHNFILANGLVASN (SEQ ID NO:268)

>Cat:/1-133 aislado de *Candidatus Atelocyanobacterium thalassa*

CLSYDTKVLTVEYGPLPIGKVVQENIRCRVYTTNDQGLIYTQPIAQWHNRGKQ EIFEYHLDDKTIIRATKEHQFMTVDHVMMPIDEIFEQGLELKKIK (SEQ ID NO:269)

LKIIRRKSLGMHEVFDIGLEKDHNFVLSNGLIASN (SEQ ID NO:270)

>Oli:/1-137 precurosr de DnaE 'Solar Lake' de *Oscillatoria limnetica*

CLSYNTEVLTVEYGPLPIGKIVDEQIHCRVYSVDENGFVYTQAIAQWHDRGYQ EIFAYELADGSVIRATKDHQFMTEDGQMFPIDEIWEKGLDLKKLPTVQD (SEQ ID NO:271)

VKIVRRQSLGVQNVYDIGVEKDHNFLLASGEIASN (SEQ ID NO:272)

>Cen:/1-137 Cianobacteria endosimbionte de *Epithemia turgida*

CLSYDTEVLTVEYGAIPIGRMVEESLDCTVYTVDKNGFVYTQSIQQWHSRGQQ EIFEYCFEDGSIIRATKDHKFMTAEGKMSSIHDIFEQGLELKKIIPWSG (SEQ ID NO:273)

AKIISCKSLGKQSVYDIGVVQDHNFLLANGVVASN (SEQ ID NO:274)

>SspPCC7502:/1-133 *Synechococcus sp.* PCC 7502

CLGYDTPVLTVEYGFMPIGKIVEEKIQCHVYSVDQNGLVFTQAIAQWHNRGQQ EVWEYNLDNGDIVRATKDHKFMTIDGQMLPINQIFEQGLELKVIA (SEQ ID NO:275)

VKIVSCKPLRVQTVYDIGVEKDHNFILDNGLVASN (SEQ ID NO:276)

>CspUCYN:/1-124 *Cyanobacterium sp.* UCYN-A2

CLSYDTKVLTVEYGPLPIGKVVQENIRCRVYTTNDQGLIYTQPIAQWHNRGKQ EIFEYHLDDKTIIRATKEHQFMTVDHVMMPIDEIFEQ (SEQ ID NO:277)

KIIRRKSLGMHEVFDIGLEKDHNFVLSNGLIASN (SEQ ID NO:278)

>Pst:/1-129 *Planktothrix* st147: st147_cleanDRAFT_c6

CLSYDTEVLTVEYGLIPISKIVEEKIECTVYTVNNQGYVYTQPIAQWHNRGEQE VFEYYLEDGSVIRATKDHKFMTVEGQMLPIDEIFEKELDL (SEQ ID NO:279)

VKIISRKSLGTQPVYDIGVQEDHNFVLNNGLVASN (SEQ ID NO:280)

>PlaCYA98/1-129: *Planktothrix* NIVA-CYA 98

CLSYDTEILTVEYGLMPIGKIVKEKIECTVYTVNNQGYVYTQPIAQWHHRGEQ EVFEYCLEDGSVIRATKDHKFMTVQGQMLPIDEIFEKELDL (SEQ ID NO:281)

VKIISRKSLGTQPVYDIGVQEDHNFLLNNGLVASN (SEQ ID NO:282)

>Pst585:/1-129 *Planktothrix sp.* 585: longitud=1586997

CLSYDTEILTVEYGLIPISKIVEEKIECTVYTVNNQGYVYTQPIAQWHNRGEQEV FEYYLEDGSVIRATKDHKFMTVDGQMLPIDEIFEKELDL (SEQ ID NO:283)

VKIISRKSLGTQPVYDIGVQEDHNFVLNNGLVASN (SEQ ID NO:284)

>CspPCC8802/1-134: *Cyanothece sp.* PCC 8802

CLSYDTEILTVEYGAIPIGKVVEENIDCTVYTVDKNGFVYTQNIAQWHLRGQQE VFEYYLDDGSILRATKDHQFMTLEGEMLPIHEIFERGLELKKIKI (SEQ ID NO:285)

VKIVSYRSLGKQFVYDIGVAQDHNFLLANGSIASN (SEQ ID NO:286)

>CfrPCC6912:/1-137 *Chlorogloeposis fritschii* PCC 6912

CLSYDTAILTVEYGFLPIGEIVEKGIECTVYTVDSNGYIYTQPIAQWHNRGEQEL FEYSLEDGSIIRATKDHKFMTIDGQMLPIDEIFARKLELMQVKGLPE (SEQ ID NO:287)

VKIIAKKSLGTQNVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:288)

>CspATC51472:/1-132 *Cyanothece sp.* ATCC 51472

CLSYDTEILTVEYGPMPIGKIVEENINCTVYTVDPNGFVYTQAIAQWHYRGEQE IFEYYLEDGATIRATKDHKFMTMEGKMLPIDEIFENNLDLKQL (SEQ ID NO:289)

VKIIGRQSLGVQKVYDIGVEKEHNFLLHNGLIASN (SEQ ID NO:290)

>Lma:/1-132 *Lyngbya majuscula*

CLSYDTEIITVEYGPIAIGEIVEKGIPCTVYSVDSNGYVYTQPIAQWHNRGEQEV FEYTLDDGSVIRATKDHKFMTIDGQMLPIDEIFEGGLELKQL (SEQ ID NO:291)

VKIISRKSLGTQPVYDIGVKDDHNFILANGMVASN (SEQ ID NO:292)

>CspESFC/1-137

CLSYDTEVLTVEYGAVPIGKLVEEKLNCSVYTVDPNGYIYTQAIAQWHDRGIQ EVFEYQLEDNTIIRATKDHKFMTEDHQMLPIDEIFERGLELKKCPQPQQ (SEQ ID NO:293)

VKIIRRRSLGFQPVYDIGLEQDHNFLLNQGAIASN (SEQ ID NO:294)

>Mae905:/1-129 *Microcystis aeruginosa* DIANCHI905

CLGGETLILTEEYGLLPIAKIVSEEVNCTVYSVDKNGFVYSQPISQWHERGLQE VFEYTLENGQTIQATKDHKFMTNDGEMLAIDTIFERGLDL (SEQ ID NO:295)

VKIISRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:296)

>RlaKORDI51-2:/1-137 *Rubidibacter lacunae* KORDI 51-2

CLSYDTEVLTVEYGPLAIGTIVSERLACTVYTVDRSGFLYAQAISQWHERGRQD VFEYALDNGMTIRATKDHKLMTADGQMVAIDDIFTQGLTLKAIDTAAF (SEQ ID NO:297)

MKIVSRKSLGVQHVYDIGVARDHNFLLANGAIASN (SEQ ID NO:298)

>CfrPCC9212/1-136 *Chlorogloeopsis fritschii* PCC 9212

CLSYDTAILTVEYGFLPIGEIVEKGIECTVYTVDSNGYIYTQPIAQWHNRGEQEL FEYSLEDGSIIRATKDHKFMTIDGQMLPIDEIFARKLELMQVKGLP (SEQ ID NO:299)

VKIIAKKSLGTQNVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:300)

>RinHH01:/1-137 *Richelia intracellularis* HH01 WGS project

CLSYDTQILTVEHGPMSIGEIVEKCLECHVYTVNKNGNICIQTITQWHFRGEQEI FEYELEDGSFIQATKDHKFMTTTGEMLPIHEIFTNGLEILQLSKSLL (SEQ ID NO:301)

VKILARKSLGTQKVYDIGVNDDHNFALSNSFIASN (SEQ ID NO:302)

>GhePCC6308:/1-133 *Geminocystis herdmanii* PCC 6308

CLSYDTEVLTVEFGAIPMGKIVEERLNCQVYSVDKNGFIYTQNIAQWHDRGVQ EVFEYELEDGRIIKATKDHKMMIENCEMVEIDRIFEEGLELFEVN (SEQ ID NO:303)

VKILKRRSISSQQVYDIGVEKDHNFLLANGLVASN (SEQ ID NO:304)

>SsuPCC9445:/1-131 *Spirulina subsalsa* PCC 9445

CLSYDTKIITVEYGAIAIGTIVEQGLHCHVYSVDPNGFIYTQPIAQWHQRGEQEV FAYTLENGSIIQATKDHKFMTQQGKMLPIDTIFEQGLDLLQV (SEQ ID NO:305)

KIIKRTSLGVRPVYDIGVIQDHNFLLENGLVASN (SEQ ID NO:306)

>MaePCC9807:/1-135 *Microcystis aeruginosa* 9807

CLGGETLILTEEYGLLPIAKIVSEEINCTVYSVDKNGFIYSQPISQWHERGLQEVF EYTLENGQTIQATKDHKFMTSDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO:307)

VKIISRQFLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:308)

>MspGI1:/1-130 *Myxosarcina sp.* GI1 contig_13

CLSYDTEVLTLKYGALPIGEIVEKRINCHVYTRAESGFFYIQSIEQWHDRGEQEV FEYTLENGATIKATKDHKFMTSGGQMLPIDEIFERGLDLL (SEQ ID NO:309)

VKIVSRKSLGKQPVYDLGVAKDHNFLLANGTVASN (SEQ ID NO:310)

>ShoPCC7110:/1-136 *Scytonema hofmanni* PCC 7110 contig00136

CLSYDTEVLTAEYGFLPIGKIVEKAIECTVYSVDNDGNIYTQPIAQWHDRGQQE VFEYSLDDGSVIRATKDHKFMTTGGQMLPIDEIFERGLDLMRIDSLP (SEQ ID NO:311)

VKILTRKSIGKQTVYDIGVERDHNFVIKNGLVASN (SEQ ID NO:312)

>WinUHHT291/1-136 *Westiella intricata* UH HT-29-1

CLSYDTEILTVEYGFLPIGEIVEKRIECTVYTVDTNGYVYTQAIAQWHNRGEQE VFEYALEDGSIIRATKDHKFMTSEGQMLPIDEIFVKGLDLLQVQGLP (SEQ ID NO:313)

VKIITRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:314)

>FspPCC9605:/1-136 *Fischerella sp.* PCC 9605 FIS9605DRAFT

CLSYDTEILTVEYGFLPIGEIVEKGIECTVYTVDNNGNVYTQTIAQWHNRGQQE VFEYCLEDGSVIRATKDHKFMTTDGQMLPIDEIFARGLDLLQVKNLP (SEQ ID NO:315)

VKIVTRRPLGTQNVYDIGVESDHNFVIKNGLVASN (SEQ ID NO:316)

>MrePCC10914:/1-137 *Mastigocladopsis repens* PCC 10914

CLSYDTEVLTVEYGFLPIGEIVEKSIECSVYTVDSNGNVYTQPIAQWHNRGQQE VFEYCLEDGSIIRATKDHKFMTIHGQMLPIDEIFERGLELMKIQGLPE (SEQ ID NO:317)

AKIITRKSLGTQNVYDIGVERDHNFVTRDGFIASN (SEQ ID NO:318)

>ShoUTEX2349:/1-137 [*Scytonema hofmanni*] UTEX 2349

CLSYNSEVLTVEYGFLPIGKIVEKGIECSVYSVDSYGKIYTQVIAQWHNRGQQE
VFEYCLEDGTIIQATKDHKFMTVDGQMLPIDEIFERGLDLMQVQGLPD (SEQ ID NO:319)

VKIITRKSLGTQNVYDIGVSSDHNFVMKNGLIASN (SEQ ID NO:320)

>AspPCC7108:/1-137 *Anabaena sp.* PCC 7108 Ana7108scaffold_2_Cont3

CLSSDTEVLTVEYGLIPIGEIIEKRIDCSVFSVDKNGNIYTQPIAQWHDRGIQELY
EYCLDDGSTIRATKDHKFMTTAGEMLPIDEIFERGLDLLKVHNLPQ (SEQ ID NO:321)

VKIITRNYVGKENVYDIGVERDHNFAIKNGLIASN (SEQ ID NO:322)

>FspPCC9339:/1-137 *Fischerella sp.* PCC 9339 PCC9339DRAFT

CLSYDTEVLTVEYGFLPIGEIVEKRIECTVYTVDHNGYVYTQPIAQWHNRGYQ
EVFEYGLEDGSVIRATKDHKFMTSEGQMLPIDEIFARELDLLQVTGLVN (SEQ ID NO:323)

VKIVTRRLLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:324)

>Csp336:/1-137 *Calothrix sp.* 336/3

CLSYDTEIFTVEYGFLPIGEIVEKRLECTVLTVDNHGNIYSQPIAQWHHRGQQQI
YEYGLEDGSVIRATKDHKFMTTDGQMLPIDEIFERGLDLLQVTNLDN (SEQ ID NO:325)

VKVITRKLADTENVYDIGVENHHNFLIKNGLVASN (SEQ ID NO:326)

>FthPCC7521:/1-136 *Fischerella thermalis* PCC 7521

CLSYETEILTVEYGFLPIGEIVEKRIECSVYTVDNNGYVCTQPIAQWHNRGYQE
VFEYGLEDGSVIRATKDHKFMTIDRQMLPIDEIFARGLDLLQVTGLP (SEQ ID NO:327)

VKIITRKSLGTQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:328)

>CyaPCC7702/1-137 *Cyanobacterium* PCC 7702 Chl7702

CLSYDTEILTVEYGFLSIGEIVEKEIECTVYTVDSNGYIYTQPIAQWHEQGEQEIF
EYSLEDGSTIRATKDHKFMTIEGEMLPIDQIFARQLDLMQITGLPQ (SEQ ID NO:329)

VKISTKKSLGKQKVYDIGVVRDHNFIIKNGFVASN (SEQ ID NO:330)

>FspPCC9431:/1-136 *Fischerella sp.* PCC 9431

CLSYDTEVLTVEYGFLPIGEIVEKRIECTVYTVDTNGYVYTQAIAQWHNRDEQE
VFEYALEDGSIIRATKDHKFMTSEGQMLPIDEIFAKGLDLLQVQGLP (SEQ ID NO:331)

VKIVTRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:332)

>FmuPCC7414:/1-137 *Fischerella muscicola* PCC 7414

CLSYETEILTVEYGFLPIGEIVEKRIECSVYTVDNNGYVCTQTIAQWHNRGYQE
VFEYGLEDGSVIRATKDHKFMTIDRQMLPIDEIFARGLDLLQVKGLPE (SEQ ID NO:333)

VKIITRQSLGTQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:334)

>FmuPCC73103:/1-137 *Fischerella muscicola* SAG 1427-1 = PCC 73103

CLSYDTEVLTVEYGFLPIGEIVEKTIECNVFTVDSNGYVYTQPIAQWHNRGYQE VFEYGLEDGSVIRATKDHKFMTSEGKMLPIDEIFARELDLLQVTGLIN (SEQ ID NO:335)

VKIVTRKFLGIQNVYDIGVEQNHNFVIKNGLVASN (SEQ ID NO:336)

>Lae:/1-137 *Lyngbya aestuarii* BL J laest3.contig.3

CLSYDTEILTVEYGAIPIGKVVDEKIECTVYSVDKNGLIYTQPIAQWHNRGKQE VFEYSLEDGSTIRATKDHKFMTMDNQMLPIDEILEKGLELKQVNADSV (SEQ ID NO:337)

VKIVSRKSLDSQTVYDIGVETDHNFLLANGSVASN (SEQ ID NO:338)

>Lsp:/1-137 *Leptolyngbya sp.* JSC-1

CLSYDTEILTVEYGALPIGKIVENQMICSVYSIDNNGYIYIQPIAQWHNRGQQEV FEYILEDGSIIRSTKDHKFMTKGGEMLPIDEIFERGLELAQVTRLEQ (SEQ ID NO:339)

VKIISRRSVGVQSVYDIGVKQDHNFFLRNGLIASN (SEQ ID NO:340)

>CwaWH8501:/1-137 *Crocosphaera watsonii* WH8501

CLSYDTEILTVEYGAMYIGKIVEENINCTVYTVDKNGFVYTQTIAQWHNRGEQ EIFEYDLEDGSKIKATKDHKFMTIDGEMLPIDEIFEKNLDLKQVVSHPD (SEQ ID NO:341)

VKIIGCRSLGTQKVYDIGVEKDHNFLLANGSIASN (SEQ ID NO:342)

>CchPCC7420:/1-135 *Coleofasciculus chthonoplastes* PCC 7420

CLSYDTQILTVEYGAVAIGEIVEKQIECTVYSVDENGYVYTQPIAQWHNRGEQE VFEYLLEDGATIRATKDHKFMTDEDQMLPIDQIFEQGLELKQVEVL (SEQ ID NO:343)

VKIIGRKPLGTQPVYDIGVERDHNFLLFNGSVASN (SEQ ID NO:344)

>CspPCC6712/1-133

CLSYDTEVLTVEYGAIPIGKIVEEKIACNVYSVDKNGFVYTQPIAQYHDRGIQE VFEYRLENGSVIRATKDHKMMTADGQMLPIDEIFKQNLDLKQLN (SEQ ID NO:345)

VKIISRQSLGKQSVFDIGVAKDHNFLLANGLVASN (SEQ ID NO:346)

>Rbr:/1-137 *Raphidiopsis brookii* D9 D9_5,

CLSYETEVLTLEYGFLPIGEIVDKQMVCTVFSVNDSGNVYTQPIGQWHDRGVQ ELYEYCLDDGSTIRATKDHKFMTTQGEMVPIDEIFHQGWELVQVSGTMN (SEQ ID NO:347)

VKIVSRRYLGKADVYDIGVAKDHNFIIKNGLVASN (SEQ ID NO:348)

>CspCCy0110:/1-134 *Cyanothece sp.* CCY0110 1101676644604

CLSYDTEILTVEYGPMPIGKIVEENINCSVYTVNKNGFVYTQSIAQWHHRGEQE VFEYYLEDGETIRATKDHKFMTTEGKMLPIDEIFENNLDLKKLTV (SEQ ID NO:349)

VKIIERRSLGKQNVYDIGVEKDHNFLLSNNLIASN (SEQ ID NO:350)

>XspPCC7305:/1-135 *Xenococcus sp.* PCC 7305

CLSADTEVLTVEYGAISIGKIVEERIECTVYSVDANGFVYTQEIAQWHNRGEQE
VFEYMLDDGSVIRATKDHKLMTIDGQMVAIDEIFSQGLELKQVLGL (SEQ ID NO:351)

VKIVSRKSLGTQTVYDLGVARDHNFLLANGTVASN (SEQ ID NO:352)

>PspPCC7319:/1-135 *Pleurocapsa sp.* PCC 7319

CLSYDTEIYTVEYGALPIGKIVESRIKCTVLTVDKNGLVYSQPIVQWHDRGIQEV
FEYTLDNGATIRATKDHKFMTVEGQMLPIDEIFELGLELKEIQQF (SEQ ID NO:353)

VKIISRQSLGKQSVYDIGVAKDHNFLLANGMVASN (SEQ ID NO:354)

>CraCS505:/1-137 *Cylindrospermopsis raciborskii* CS-505

CLSYETEVLTLEYGFVPIGEIVNKQMVCTVFSLNDSGNVYTQPIGQWHDRGVQ
DLYEYCLDDGSTIRATKDHKFMTTQGEMVPIDEIFHQGWELVQVSGISK (SEQ ID NO:355)

VKIVSRRYLGKADVYDIGVAKDHNFIIKNGLVASN (SEQ ID NO:356)

>MaePCC7806:/1-135 *Microcystis aeruginosa* PCC 7806

CLGGETLILTEEYGLLPIAKIVSEEVNCTVYSVDKNGFVYSQPISQWHERGLQE
VFEYTLENGQTIQATKDHKFMTNDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO:357)

VKIISRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:358)

>MaeNIES843 :/1-135 ADN de *Microcystis aeruginosa* NIES-843

CLGGETLILTEEYGLLPIAKIVSEEINCTVYTVDQNGFVYSQPISQWHERGLQEV
FEYTLENGQTIQATKDHKFMTSDGEMLAIDTIFERGLDLKSSDFS (SEQ ID NO:359)

VKIIGRQSLGRKPVYDIGVEKDHNFLLGNGLIASN (SEQ ID NO:360)

La figura 1 muestra una alineación y un modelo generado por ordenador del diseño de la inteína dividida Cfa según una realización de la invención. El panel A muestra una alineación de secuencias de Npu DnaE y Cfa DnaE. Las secuencias comparten un 82 % de identidad con las diferencias (subrayadas, cian) distribuidas uniformemente a través de la secuencia primaria. Los residuos catalíticos y los residuos “aceleradores” de la segunda cubierta se muestran en signo de intercalación, naranja y asterisco, verde, respectivamente. El panel B muestra los mismos residuos resaltados en el panel que se mapearon en la estructura de Npu (pdb = 4kl5).

La inteína Cfa tiene una alta similitud de secuencia con Npu (82 %), y los residuos no idénticos se extienden por toda la estructura 3D de la proteína.

Se generaron fragmentos de inteína Cfa fusionados con exteínas modelo y se midió su actividad de PTS usando el ensayo *in vitro* mencionado anteriormente (figura 2). Esto reveló que la inteína Cfa experimenta corte y empalme 2,5 veces más rápido a 30 °C que Npu ($t_{1/2}$ 20 s frente a 50 s), una mejora notable en la actividad ya que esta última es la inteína dividida DnaE más rápida caracterizada (figura 2A). Esta velocidad acelerada se manifiesta tanto en la formación de ramificaciones (aumento de 3 veces) como en la resolución de ramificaciones (aumento de 2 veces). En línea con las inteínas DnaE originales, Cfa retiene la preferencia por un residuo hidrófobo voluminoso en la posición +2 de la C-exteína. Sorprendentemente, Cfa muestra una mayor velocidad de corte y empalme en función de la temperatura y es consistentemente más rápida que Npu (figura 2A). La inteína Cfa incluso mantiene la actividad a 80 °C, aunque con un rendimiento reducido de productos de corte y empalme, mientras que Npu es inactiva a esta temperatura. Estos resultados demuestran que la ingeniería de consenso es eficaz en la producción de una inteína que es altamente activa en un amplio intervalo de temperaturas.

Las aplicaciones de PTS normalmente requieren la fisión de una proteína diana y la fusión de los fragmentos

resultantes con los segmentos de inteína dividida apropiados.[1] Como consecuencia, la solubilidad de estas proteínas de fusión a veces puede ser escasa. Debido a que se usan con frecuencia desnaturalizantes de proteínas tales como clorhidrato de guanidina (GuHCl) y urea para mantener estos fragmentos menos solubles en disolución, se sometió a prueba la capacidad de Cfa para experimentar corte y empalme en presencia de estos agentes caotrópicos. Se encontró que la inteína Cfa experimenta corte y empalme en presencia de GuHCl hasta 4 M (con poca disminución en la actividad observada hasta 3 M), mientras que no se observó actividad para Npu en GuHCl ≥3 M (figura 2B). Sorprendentemente, el corte y empalme de Cfa no se ve afectado en gran medida hasta urea 8 M, mientras que el corte y empalme de Npu disminuye drásticamente por encima de urea 4 M (figura 2C).

La figura 2 muestra gráficos que muestran la caracterización de la inteína Cfa de acuerdo con una realización de la invención. En el panel A, se muestran las velocidades de corte y empalme para Cfa y Npu en función de la temperatura. Npu es inactiva a 80 °C (error = DE (n=3)). En los paneles B y C, se muestran las velocidades de corte y empalme para Cfa y Npu en función del caotropo añadido. Npu es inactiva en GuHCl 3 M o urea 8 M. Obsérvese que Cfa tiene actividad residual en GuHCl 4 M ($k = 7 \times 10^{-5}$) (error = DE (n = 3)).

La tolerancia sin precedentes e inesperada de Cfa a altas concentraciones de GuHCl y urea sugiere que la inteína podría retener la actividad directamente después de la extracción caotrópica de proteínas insolubles de cuerpos de inclusión bacterianos, acelerando así los estudios basados en PTS. En consecuencia, la proteína de fusión modelo, $His_6$-Sumo-$Cfa^N$, se sobreexpresó en células de *E. coli* y se extrajo la proteína de los cuerpos de inclusión con urea 6 M. La proteína se purificó a partir de este extracto mediante cromatografía de afinidad de níquel y luego se modificó directa y eficientemente por PTS en condiciones desnaturalizantes, es decir, sin la necesidad de etapas de replegamiento intermedias. En general, se espera que la actividad robusta de Cfa en presencia de agentes caotrópicos resulte útil cuando se trabaja con fragmentos de proteínas que demuestran poca solubilidad en condiciones nativas.

La fusión de una proteína de interés a una inteína dividida puede dar como resultado una reducción marcada en los niveles de expresión celular en comparación con la proteína sola.[6] Esta situación se encuentra con mayor frecuencia para fusiones a N-inteínas que a C-inteínas, lo que probablemente se deba al mayor tamaño de las primeras y a su estado parcialmente plegado.[18] Por lo tanto, se investigó si la estabilidad térmica y caotrópica mejorada de Cfa se traduciría en niveles de expresión aumentados de fusiones de $Cfa^N$. De hecho, estudios modelo en *E. coli* revelaron un aumento significativo (30 veces) en la expresión de proteína soluble para una fusión de $Cfa^N$ en comparación con la fusión de $Npu^N$ correspondiente (figura 8). Dado este resultado, se investigó si las fusiones de $Cfa^N$ también presentarían niveles de expresión de proteína aumentados en células de mamífero. En particular, las fusiones de inteínas a la cadena pesada (HC) de anticuerpos monoclonales (mAb) se han convertido en una herramienta poderosa para la conjugación específica de sitio de cargas sintéticas.[19-21] Se exploraron los niveles de expresión en células HEK293 de un mAb (αDec205) en función de la N-inteína fusionada a su HC. De acuerdo con los resultados de la expresión bacteriana, la producción de la fusión HC-$Cfa^N$ fue significativamente mayor que para las otras inteínas examinadas; por ejemplo, los niveles secretados del constructo mAb-Cfa fueron ~10 veces más altos que para la fusión de Npu correspondiente (figuras 3A y 3B). De manera importante, mAb-Cfa retuvo la actividad de PTS y pudo modificarse de manera específica de sitio con un péptido sintético mediante corte y empalme directamente en el medio de crecimiento después de la expresión de cuatro días a 37 °C.

La figura 8 es un análisis de SDS-PAGE de la expresión de prueba de $His_6$-SUMO-$Npu^N$ e $His_6$-SUMO-$Cfa^N$. Gel teñido con azul brillante de Coomassie a partir de una purificación con Ni-NTA en un volumen de columna (CV) de 4 ml de la fracción soluble de 1 l de cultivo de *E. coli*. Los carriles corresponden a (P) el sedimento de cuerpo de inclusión, (FT) fracción no retenida de la disolución de Ni-NTA unida al lote, (W1) un lavado de 5 CV con imidazol 5 mM, (W2) un lavado de 5 CV con imidazol 25 mM, (E1-E4) y cuatro eluciones de 1,5 CV de imidazol 250 mM.

Finalmente, para explorar adicionalmente la utilidad de la inteína Cfa en el contexto de la conjugación de anticuerpos, se investigó si el sistema de PTS podría usarse para unir múltiples copias de una carga sintética a la cadena pesada del mAb. En consecuencia, se usó semisíntesis para preparar un constructo en la que la mitad C-terminal de Cfa ($Cfa^C$) se fusionó con una C-exteína que contenía un andamio dendrimérico que permite la unión multimérica de la carga, en este caso, fluoresceína (figura 3C). Esta carga dendrítica se unió con éxito al anticuerpo αDec205 por medio de PTS mediada por Cfa, nuevamente realizada directamente *in situ* dentro del medio de crecimiento celular (figuras 3D y 3E). Esto representa la primera vez que se ha usado PTS para unir un constructo de exteína ramificada a una proteína diana, resaltando el potencial del sistema para manipular la cantidad de carga útil de conjugados de anticuerpo-fármaco.[22]

La figura 3 muestra la expresión y modificación de un anticuerpo monoclonal de ratón usando la inteína Cfa de acuerdo con una realización de la invención. El panel A muestra la expresión de prueba en células HEK293T de diversos homólogos de $Int^N$ (Npu, Mcht, Ava y Cfa) fusionados al extremo C terminal de la cadena pesada de un anticuerpo monoclonal de ratón αDec205. Parte superior: Análisis de inmunotransferencia de tipo Western (IgG de ratón α) de los niveles de anticuerpos presentes en el medio después de la expresión de 96 horas. Parte inferior: inmunotransferencia de tipo Western de α-actina del lisado celular como control de carga. El panel B muestra la cuantificación del rendimiento de expresión normalizado por densitometría de la señal de αDEC205 HC-$Int^N$ en el panel A (error = DE (n=4)). El panel C muestra la estructura del constructo de $Cfa^C$-dendrímero usado en las reacciones

de PTS con la fusión de αDEC205 HC-Int$^N$. Por simplicidad, la secuencia peptídica de Cfa$^C$ se representa simbólicamente en verde (como un rectángulo con un corte triangular a la izquierda). El panel D es un esquema del enfoque de PTS *in situ* usado para modificar la HC de un mAb con una carga multivalente. El panel E es un análisis de SDS-PAGE de la reacción de PTS. Carril 1: mAB αDEC205 de ratón de tipo silvestre. Carril 2: Fusión de mAB de αDEC205 de ratón-Cfa$^N$. Carril 3: adición del Cfa$^C$-dendrímero a los medios que contienen el mAB αDEC205-Cfa$^N$. La reacción de corte y empalme se analizó por fluorescencia (parte inferior) e inmunotransferencia de tipo Western (parte superior, IgG de ratón α).

El descubrimiento de inteínas divididas rápidas ha revolucionado las aplicaciones del corte y empalme en trans de proteínas. La notable robustez de la inteína Cfa descrita en este estudio debe ampliar la utilidad de muchas de estas tecnologías al permitir que se realice PTS en un intervalo más amplio de condiciones de reacción. Además, la capacidad de Cfa para aumentar los rendimientos de expresión de fusiones de N-inteína debe fomentar el uso adicional de inteínas divididas para la semisíntesis de proteínas. El enfoque guiado por la actividad que se usa para diseñar por ingeniería genética esta inteína puede aplicarse a otras familias de inteínas o actuar como estrategia general para el refinamiento de múltiples alineamientos de secuencias usadas para la ingeniería de consenso.

**Materiales y métodos**

Materiales

Se adquirieron oligonucleótidos y genes sintéticos de Integrated DNA Technologies (Coralville, IA). El kit de mutagénesis dirigida al sitio QuickChange XL II y la polimerasa de fusión Pfu Ultra II Hotsart se adquirieron de Agilent (La Jolla, CA). Todas las enzimas de restricción y 2x Gibson Assembly Master Mix se adquirieron de New England Biolabs (Ipswich, MA). Las células "internas" de alta competencia usadas para la clonación y expresión de proteínas se generaron a partir de *E. coli* One Shot Bl21 (DE3) químicamente competente y células competentes DH5α con eficiencia de subclonación adquiridas de Invitrogen (Carlsbad, CA). Medio de Eagle modificado por Dulbecco (DMEM), Lipofectamine 2000 y suero bovino fetal con bajo contenido en IgG se adquirieron también de Invitrogen. Se adquirieron kits de purificación de ADN de Qiagen (Valencia, CA). Todos los plásmidos se secuenciaron por GENEWIZ (South Plainfield, NJ). Se adquirieron N,N-diisopropiletilamina (DIPEA), medio Luria Bertani (LB) y todas las sales tamponantes de Fisher Scientific (Pittsburgh, PA). Se adquirieron dimetilformamida (DMF), diclorometano (DCM), azul brillante de Coomassie, triisopropilsilano (TIS), β-mercaptoetanol (BME), DL-ditiotreitol (DTT), 2-mercaptoetanosulfonato de sodio (MESNa), tetraquis(trifenilfosfina)paladio (0) (Pd(PPh$_3$)$_4$) y 5(6)-carboxifluoresceína de Sigma-Aldrich (Milwaukee, WI) y se usaron sin purificación adicional. Se adquirieron clorhidrato de tris(2-carboxietil)fosfina (TCEP) e isopropil-β-D-tiogalactopiranósido (IPTG) de Gold Biotechnology (St. Louis, MO). El inhibidor de proteasa usado fue el inhibidor de proteasa Roche Complete (Roche, Branchburg, NJ). Se adquirió resina de ácido níquel-nitrilotriacético (Ni-NTA) de Thermo Scientific (Rockford, IL). Se adquirieron aminoácidos Fmoc de Novabiochem (Darmstadt, Alemania) o Bachem (Torrance, CA). Se adquirieron hexafluorofosfato de (7-azabenzotriazol-1-iloxi)tripirrolidinofosfonio (PyAOP) y hexafluorofosfato de *O*-(benzotriazol-1-il)-N,N,N',N'-tetrametiluronio (HBTU) de Genscript (Piscataway, NJ). Se adquirió la resina Rink Amide-ChemMatrix de Biotage (Charlotte, NC). Se adquirió ácido trifluoroacético (TFA) de Halocarbon (North Augusta, SC). Se adquirieron la membrana de PVDF de inmunotransferencia (0,2 μm) y los geles Criterion XT Bis-Tris (poliacrilamida al 12 %) de Bio-Rad (Hercules, CA). El tampón de ejecución de MES-SDS se adquirió de Boston Bioproducts (Ashland, MA). El anticuerpo secundario anti-IgG de ratón (Licor mouse 800) y el anticuerpo primario de αActina de ratón se adquirieron de Li-COR biotechnology (Lincoln, NE).

**Equipo**

La RP-HPLC analítica se realizó en instrumentos Hewlett-Packard series 1100 y 1200 equipados con una columna $C_{18}$ Vydac (5 μm, 4,6 x 150 mm) a un caudal de 1 ml/min. La RP-HPLC preparativa se realizó en un sistema de LC preparativa Waters compuesto por un módulo de gradiente binario Waters 2545 y un detector UV Waters 2489. Las purificaciones se llevaron a cabo en una columna $C_{18}$ Vydac 218TP1022 (10 μM; 22 x 250 mm) a un caudal de 18 ml/min. Todas las ejecuciones usaron TFA al 0,1 % (ácido trifluoroacético) en agua (disolvente A) y acetonitrilo al 90 % en agua con TFA al 0,1 % (disolvente B). A menos que se indique lo contrario, los péptidos y las proteínas se analizaron usando el siguiente gradiente: 0 % de B durante 2 minutos (isocrático) seguido del 0-73 % de B durante 30 minutos. El análisis espectrométrico de masas de ionización por electropulverización (ESI-MS) se realizó en un espectrómetro de masas Bruker Daltonics MicroTOF-Q II. La cromatografía de exclusión molecular se llevó a cabo en un sistema AKTA FPLC (GE Healthcare) usando una columna Superdex S75 16/60 (CV = 125 ml). Se obtuvieron imágenes de geles teñidos con Coomassie e inmunotransferencias de tipo Western usando un generador de imágenes de infrarrojo LI-COR Odyssey. Se obtuvieron imágenes de geles fluorescentes usando un generador de imágenes GE ImageQuant LAS 4000. El ensayo de crecimiento de *E. coli* dependiente de corte y empalme se realizó en un lector de microplacas sintonizable VersaMax de Molecular Devices. La lisis celular se llevó a cabo usando un sonificador digital S-450D Branson.

**Clonación de plásmidos de ADN**

Todos los constructos de N-inteína para la expresión en *E. coli.* se clonaron en los vectores pET y pTXB1 usados previamente.[1] Los plásmidos que codifican para WT pet30-$His_6$-SUMO-AEY-$Ssp^N$, pet30-$His_6$-SUMO-AEY-$Npu^N$, pTXB1-$Ssp^C$-MxeGyrA-$His_6$ y pTXB1-$Npu^C$-MxeGyrA-$His_6$ se clonaron como se describió anteriormente[1] y codifican las siguientes secuencias de proteínas. Los productos proteicos después de la escisión de SUMO (N-inteínas) o tiólisis (C-inteínas) se muestran en negrita para todos los plásmidos.

**Plásmido 1:**

WT $Ssp^N$: pet30-$His_6$-SUMO-AEY-$Ssp^N$

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA HREQIGG**AEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHD RGEQEVLEYELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALD NHRLPFPLLDAGTIK** (SEQ ID NO:361)

**Plásmido 2:**

WT $Npu^N$: pet30-$His_6$-SUMO-AEY-$Npu^N$

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA HREQIGG**AEYALSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHD RGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN** (SEQ ID NO:362)

**Plásmido 3:**

WT $Ssp^C$: pTXB1-$Ssp^C$-MxeGyrA-$His_6$

**MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTAN HPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPG LVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHAHHHHHH (SEQ ID NO:363)

**Plásmido 4:**

WT $Npu^C$: pTXB1-$Npu^C$-MxeGyrA-$His_6$

**MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CITGDALVALPEG ESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTA NHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVP GLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNG FVSHAHHHHHH (SEQ ID NO:364)

Todos los mutantes del lote $Ssp^N$ se clonaron usando el kit de mutagénesis dirigida al sitio QuikChange usando el plásmido 1 como molde y codifican las secuencias de proteínas que se muestran a continuación. La secuencia de N-inteína se muestra en negrita con los residuos correspondientes a la mutación del lote subrayada.

**Plásmido 5:**

Lote 1: Pet30-$His_6$-SUMO-AEY-$Ssp^N$ (R73K, L75M, Y79G, L81M)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHKFMTTDGQMLAIEEIFARQLDLLTLENIKQTEEALDNHRLPF
PLLDAGTIK (SEQ ID NO:365)

**Plásmido 6:**

Ssp$^N$ R73K: Pet30-His$_6$-SUMO-AEY-Ssp$^N$ (R73K)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHKFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:366)

**Plásmido 7:**

Ssp$^N$ R73K Y79G: Pet30-His$_6$-SUMO-AEY-Ssp$^N$ (R73K, Y79G)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHKFLTTDGQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:367)

**Plásmido 8:**

Ssp$^N$ R73K Y79G L81M: Pet30-His$_6$-SUMO-AEY-Ssp$^N$ (R73K, Y79G, L81M)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHKFLTTDGQMLAIEEIFARQLDLLTLENIKQTEEALDNHRLPF
PLLDAGTIK (SEQ ID NO:368)

**Plásmido 9:**

Lote 2: Pet30-His$_6$-SUMO-AEY-Ssp$^N$ (L56F, S70K, A83P, E85D)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVFEYELEDGSVIRATKDHRFLTTDYQLLPIDEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:369)

**Plásmido 10:**

Ssp$^N$ A83P: Pet30-His$_6$-SUMO-AEY-Ssp$^N$ (A83P)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHRFLTTDYQLLPIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:370)

**Plásmido 11:**

$Ssp^N$ S70K A83P: Pet30-$His_6$-SUMO-AEY-$Ssp^N$ (S70K, A83P)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATKDHRFLTTDYQLLPIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:371)

**Plásmido 12:**

$Ssp^N$ L56, S70K, A83P: Pet30-$His_6$-SUMO-AEY-$Ssp^N$ (L56F, S70K, A83P)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGE
QEVFEYELEDGSVIRATKDHRFLTTDYQLLPIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:372)

**Plásmido 13:**

Lote 3: Pet30-$His_6$-SUMO-AEY-$Ssp^N$ (S23E, E24K, E25R, N27E)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVEKRIECSVYSVDPEGRVYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:373)

**Plásmido 14:**

Lote 4: Pet30-$His_6$-SUMO-AEY-$Ssp^N$ (P35N, E36N, R38N, V39I)

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGGAEYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDNNGNIYTQAIAQWHDRGE
QEVLEYELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFP
LLDAGTIK (SEQ ID NO:374)

Los cuatro mutantes de lote (lotes 5-8) y mutante puntual A136S en la inteína $Ssp^C$ se clonaron por PCR inversa usando polimerasa Pfu Ultra II HS (Agilent) usando el plásmido 3 como molde y codifican las secuencias de proteínas que se muestran a continuación:

**Plásmido 15:**

Lote 5: pTXB1-Ssp$^{C}$-MxeGyrA-His$_6$ (V103I, V105I, I106A, G107T)

**M**I**K**I**A**T**RRSLGVQRIFDIGLPQDHNFLLANGAIAAN**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTAN HPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPG LVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHAHHHHHH (SEQ ID NO:375)

**Plásmido 16:**

Lote 6: pTXB1-Ssp$^{C}$-MxeGyrA-His$_6$ (R115N, 1116V, F117Y)

**MVKVIGRRSLGVQ**N**V**Y**DIGLPQDHNFLLANGAIAAN**CITGDALVALPEG ESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTA NHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVP GLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNG FVSHAHHHHHH (SEQ ID NO:376)

**Plásmido 17:**

Lote 7 pTXB1-Ssp$^{C}$-MxeGyrA-His$_6$ (L121V, P122E, Q123R)

**MVKVIGRRSLGVQRIFDIG**V**E**R**DHNFLLANGAIAAN**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTAN HPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPG LVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHAHHHHHH (SEQ ID NO:377)

**Plásmido 18:**

Lote 8: pTXB1-Ssp$^{C}$-MxeGyrA-His$_6$ (L128A, A130K, A133F)

**MVKVIGRRSLGVQRIFDIGLPQDHNF**A**L**K**NG**F**IAAN**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTAN HPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPG LVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHAHHHHHH (SEQ ID NO:378)

**Plásmido 19:**

Ssp$^{C}$ A136S: pTXB1-Ssp$^{C}$-MxeGyrA-His$_6$ (A136S)

**MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIA**S**N**CITGDALVALPEGE SVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTAN HPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPG LVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGF VSHAHHHHHH (SEQ ID NO:379)

Se optimizaron los codones del gen para la secuencia de $Dna^E$ consenso fusionada para la expresión en *E. coli* a través de ADN IDT y se adquirió como gBlock. La secuencia de ADN gBlock se muestra a continuación:

TGCCTGTCTTACGACACAGAGATTCTGACCGTTGAATATGGATTCCTTCC
TATCGGTAAGATCGTGGAGGAACGGATTGAATGCACAGTCTATACGGTAGATAAAAA
TGGCTTTGTGTATACACAACCTATTGCTCAGTGGCATAACCGGGGAGAACAGGAAGTT
TTCGAATACTGCTTAGAAGACGGTTCGATTATCCGTGCAACGAAAGATCACAAATTTA
TGACGACCGACGGTCAGATGTTACCGATTGATGAGATTTTCGAACGGGGGTTAGACCT
GAAACAAGTTGATGGTTTGCCGATGGTCAAGATCATTAGTCGTAAGAGTCTGGGCACT
CAAAACGTCTACGATATTGGAGTAGAAAAAGATCATAATTTTTTGCTGAAGAATGGGC
TGGTGGCCTCTAAC (SEQ ID NO:380)

El plásmido de expresión para $Cfa^N$ se clonó usando ensamblaje de Gibson en el **plásmido 1**, produciendo un vector que codifica para la siguiente proteína que se muestra a continuación:

**Plásmido 20:**

CfaN: pET30-$His_6$-SUMO-AEY-$Cfa^N$

MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVS
DGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEA
HREQIGG**AEYCLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHN**
**RGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP** (SEQ ID
NO:381)

El plásmido de expresión para la C-inteína consenso se clonó usando ensamblaje de Gibson en el **plásmido 3**, produciendo un vector que codifica para el siguiente gen:

**Plásmido 21:**

$Cfa^C$: pTXB1-$Cfa^C$-MxeGyrA-H6

**MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN**CITGDALVALPEG
ESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTA
NHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVP
GLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNG
FVSHAHHHHHH (SEQ ID NO:382)

Constructos de Cfa usados para el examen de crecimiento de *E. coli*.

Los plásmidos Cfa usados para examinar la dependencia del corte y empalme en la posición +2 de la C-exteína se generaron usando clonación por restricción en un plásmido generado previamente[2] que contiene un sistema de expresión doble del gen de aminoglucósido fosfotransferasa dividida ($Kan^R$). El constructo de expresión doble de Cfa se muestra a continuación:

**Plásmidos 22-25**

[Promotor de KanR]-[RBS]-[$KanR^N$]-[$Cfa^N$]-[iRBS]-[$Cfa^C$-[CXN-$KanR^C$]

Después de la secuencia promotora, hay dos sitios de unión al ribosoma de *E. coli* separados en este vector (RBS e iRBS). Cada RBS va seguido por una mitad del constructo de KanR-inteína dividida, cuyas secuencias de proteínas se muestran a continuación (la inteína Cfa se resalta en negrita).

KanR$^N$-Cfa$^N$:

MEQKLISEEDLSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQSGATI
YRLYGKPDAPELFLKHGKGSVANDVTDEMVRLNWLTEFMPLPTIKHFIRTPDDAWLLTTA
IPGKTAFQVLEEYPDSGENIVDALAVFLRRLHSIPVCNCPFNSDRVFRLAQAQSRMNNGLV
DASDFDDERNGWPVEQVWKEMHKLLPF**CLSYDTEILTVEYGFLPIGKIVEERIECTVYT**
**VDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFE**
**RGLDLKQVDGLP** (SEQ ID NO:384)

Cfa$^C$-KanR$^C$

**MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASNC<u>X</u>N**SVVTHGDFSL
DNLIFDEGKLIGCIDVGRVGIADRYQDLAILWNCLGEFSPSLQKRLFQKYGIDNPDMNKLQ
FHLMLDEFF (SEQ ID NO:385)

La posición +2 de la C-exteína está subrayada, y es o bien fenilalanina, DGGMR, glicina, arginina o bien glutamato.

αDEC205-HC-Cfa$^N$

Los plásmidos de pCMV que contienen la cadena ligera (LC), la cadena pesada (HC) y fusiones de HC-inteína (HC-Npu$^N$, HC-Mcht$^N$, HC-Ava$^N$) del anticuerpo αDEC205 se obtuvieron como se describió anteriormente.[3] Se generó una secuencia de Cfa DnaE con codones optimizados para la expresión en células de mamífero usando JCAT[4] y se adquirió como gBlock a través de ADN IDT. La secuencia se muestra a continuación:

TGCCTGAGCTACGACACCGAGATCCTGACCGTGGAGTACGGCTTCCTGC
CCATCGGCAAGATCGTGGAGGAGCGCATCGAGTGCACCGTGTACACCGTGGACAAGA
ACGGCTTCGTGTACACCCAGCCCATCGCCCAGTGGCACAACCGCGGCGAGCAGGAGG
TGTTCGAGTACTGCCTGGAGGACGGCAGCATCATCCGCGCCACCAAGGACCACAAGTT
CATGACCACCGACGGCCAGATGCTGCCCATCGACGAGATCTTCGAGCGCGGCCTGGA
CCTGAAGCAGGTGGACGGCCTGCCCGTGAAGATCATCAGCCGCAAGAGCCTGGGCAC
CCAGAACGTGTACGACATCGGCGTGGAGAAGGACCACAACTTCCTGCTGAAGAACGG
CCTGGTGGCCAGCAAC (SEQ ID NO:386)

La secuencia de Cfa$^N$ con codones optimizados para mamíferos se clonó luego en el plásmido pCMV HC-Npu$^N$ usando clonación por restricción para dar una secuencia que codifica la siguiente proteína:

**Plásmido 26:**

HC-Cfa$^N$: pCMV-HC-Cfa$^N$

MGWSCIILFLVATATGVHSEVKLLESGGGLVQPGGSLRLSCAASGFTFNDF
YMNWIRQPPGQAPEWLGVIRNKGNGYTTEVNTSVKGRFTISRDNTQNILYLQMNSLRAED
TAIYYCARGGPYYYSGDDAPYWGQGVMVTVSSATTKGPSVYPLAPGSAAQTNSMVTLG
CLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLYTLSSSVTVPSSTWPSETVTCNVAHP
ASSTKVDKKIVPRDCGCKPCICTVPEVSSVFIFPPKPKDVLTITLTPKVTCVVVAISKDDPEV
QFSWFVDDVEVHTAQTQPREEQFNSTFRSVSELPIMHQDWLNGKEFKCRVNSAAFPAPIE
KTISKTKGRPKAPQVYTIPPPKEQMAKDKVSLTCMITDFFPEDITVEWQWNGQPAENYKN
TQPIMDTDGSYFVYSKLNVQKSNWEAGNTFTCSVLHEGLHNHHTEKSLSHSPGKASGG**C**
**LSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYC**
**LEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP**GHHHHHHG (SEQ ID NO:387)

Inteína Cfa$^{C}$ para el ligamiento del dendrímero:

Un plásmido que contenía la C-inteína Cfa con un enlazador de C-exteína se clonó por PCR inversa en el **plásmido 21** y codifica la secuencia de proteína que se muestra a continuación:

**Plásmido 27:**

**Cfa$^{C}$-link: pTXB1-H6-Cfa$^{C}$-CFNSGG-MxeGyrA-H6**

**MGHHHHHHSGVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASNCFN**
**SGG**CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPV
YTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFA
RGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPV
YSLRVDTADHAFITNGFVSHAHHHHHH (SEQ ID NO:388)

Los protocolos de expresión y purificación de todos los constructos His$_6$-SUMO-AEY-Int$^{N}$ (plásmidos 1, 2, 5-14, 20) e Int$^{C}$-GyrAHis$_6$ (plásmidos 3, 4, 15-19, 21, 27) se adaptaron a partir de métodos previamente descritos.[1]

Expresión de todos los constructos His$_6$-SUMO-AEY-Int$^{N}$

Se transformaron células de *E. coli* BL21 (DE3) con un plásmido de N-inteína y se hicieron crecer a 37 °C en 1 l de LB que contenía 50 μg/ml de kanamicina. Una vez que el cultivo había alcanzado una $DO_{600}$=0,6, se añadió IPTG 0,5 mM para inducir la expresión (concentración final 0,5 mM, 3 h a 37 °C). Las células se sedimentaron mediante centrifugación (10.500 rcf, 30 min) y se almacenaron a -80 °C.

Purificación de todos los constructos His$_6$-SUMO-AEY-Int$^{N}$

Purificación de constructos de N-inteína para mutagénesis por lotes

Los sedimentos celulares (de la expresión de los plásmidos 1, 2, 5-14) se resuspendieron en 30 ml de tampón de lisis (fosfato 50 mM, NaCl 300 mM, imidazol 5 mM, pH 8,0) que contenía cóctel inhibidor de proteasa Roche Complete. Las células resuspendidas se lisaron luego por sonicación en hielo (amplitud del 35 %, 8 pulsos de 20 segundos de encendido / 30 segundos de apagado). El cuerpo de inclusión insoluble que contenía la N-inteína se recuperó por centrifugación (35.000 rcf, 30 min). El sobrenadante se desechó y el sedimento se resuspendió en 30 ml de tampón de lavado Triton (tampón de lisis con tritón X-100 al 0,1 %) y se incubó a temperatura ambiente durante 30 minutos. El lavado con Triton se centrifugó a continuación a 35.000 rcf durante 30 minutos. El sobrenadante se desechó, el sedimento de cuerpos de inclusión se resuspendió en 30 ml de tampón de lisis que contenía urea 6 M, y la suspensión se incubó durante la noche a 4 °C para extraer y resolubilizar la proteína. Esta mezcla se centrifugó luego a 35.000 rcf durante 30 minutos.

El sobrenadante se mezcló luego con 4 ml de resina Ni-NTA (para purificación por afinidad usando la etiqueta His$_6$) y se incubó a 4 °C durante 30 minutos para unir por lotes la proteína. Esta mezcla se cargó en una columna con frita, se recogió la fracción no retenida y se lavó la columna con 5 volúmenes de columna (CV) de tampón de lisis con urea

6 M y 5 CV de tampón de lisis con imidazol 25 mM y urea 6 M. La proteína se eluyó entonces en cuatro fracciones de 1,5 CV de tampón de lisis con imidazol 250 mM y urea 6 M. Se encontró generalmente por SDS-PAGE (gel Bis-Tris al 12 %, ejecutado durante 50 minutos a 170 V) que las dos primeras fracciones de elución contenían la proteína expresada y se combinaron para el replegamiento.

Las N-inteínas se replegaron por diálisis gradual en tampón de lisis con DTT 0,5 mM a 4 °C. Esta proteína replegada se trató luego con TCEP 10 mM y proteasa Ulp1 (durante la noche, TA) para escindir la etiqueta de expresión $His_6$-SUMO. Después, la disolución se mezcló con 4 ml de resina de Ni-NTA y se incubó durante 30 minutos a 4 °C. La suspensión se aplicó a una columna con frita y se recogió la fracción no retenida junto con un lavado de 3 CV con tampón de lisis. La proteína se trató luego con TCEP 10 mM, se concentró hasta 10 ml y se purificó adicionalmente por cromatografía de exclusión molecular usando una columna de filtración en gel S75 16/60 empleando tampón de corte y empalme desgasificado (fosfato de sodio 100 mM, NaCl 150 mM, EDTA 1 mM, pH 7,2) como fase móvil. Las fracciones se analizaron por SDS-PAGE, RP-HPLC analítica y ESI-MS. La proteína pura se almacenó mediante congelación instantánea en $N_2$ líquido después de la adición de glicerol (20 % v/v). Nota: durante la etapa de replegamiento, se observó una precipitación significativa de proteínas para el lote 3, lo que sugiere que es propenso a la agregación.

Purificación de $Cfa^N$:

El sedimento celular (de la expresión del plásmido 20) se resuspendió en primer lugar en 30 ml de tampón de lisis (fosfato 50 mM, NaCl 300 mM, imidazol 5 mM, pH 8,0) que contenía el cóctel inhibidor de proteasa Roche Complete. Las células se lisaron luego por sonicación (amplitud del 35 %, 8 pulsos de 20 segundos de encendido / 30 segundos de apagado), y el lisado se sedimentó por centrifugación (35.000 rcf, 30 min). El sobrenadante se incubó con 4 ml de resina Ni-NTA durante 30 minutos a 4 °C para enriquecer la proteína $Cfa^N$ soluble. A continuación, la suspensión se cargó en una columna con frita, y la columna se lavó con 20 ml de tampón de lavado 1 (tampón de lisis) seguido de 20 ml de tampón de lavado 2 (tampón de lisis con imidazol 25 mM). Finalmente, la proteína se eluyó de la columna con 4 x 1,5 CV de tampón de elución (tampón de lisis + imidazol 250 mM).

La proteína deseada, que estaba presente en las fracciones de elución 1 y 2 tal como se determinó por SDS-PAGE (gel de bis-tris al 12 % ejecutado en tampón de ejecución MES-SDS a 170 V durante 50 minutos), se dializó entonces en tampón de lisis durante 4 horas a 4 °C. Después de la diálisis, la proteína se trató con TCEP 10 mM y proteasa Ulp1 durante la noche a temperatura ambiente para escindir la etiqueta de expresión $His_6$-SUMO. A continuación, la disolución se incubó con 4 ml de resina de Ni-NTA durante 30 minutos a 4 °C. La suspensión se aplicó a una columna con frita y se recogió la fracción no retenida junto con un lavado de 3 CV con tampón de lisis. La proteína se trató luego con TCEP 10 mM, se concentró hasta 10 ml y se purificó sobre una columna de filtración en gel S75 16/60 empleando tampón de corte y empalme desgasificado (fosfato de sodio 100 mM, NaCl 150 mM, EDTA 1 mM, pH 7,2) como fase móvil. Las fracciones se analizaron mediante SDS-PAGE (gel de bis-tris al 12 % ejecutado en tampón de ejecución MES-SDS a 170 V durante 60 minutos), RP-HPLC analítica y ESI-MS. La proteína pura se almacenó en glicerol (20 % v/v) y se congeló instantáneamente en $N_2$ líquido.

Semisíntesis de constructos de $Int^C$-CFN

Se transformaron células *E. coli* BL21 (DE3) con el plásmido pTXB1-$Int^C$-GyrA-$H_6$ apropiado (plásmidos 3, 4, 15-19, 21) y se hicieron crecer en 2 l de medio LB que contenía ampicilina (100 μg/ml) a 37 °C. Una vez que el cultivo había alcanzado una $DO_{600}$ = 0,6, se indujo la expresión mediante la adición de IPTG (0,5 mM, 3 horas, 37 °C). Los sedimentos celulares se recogieron por centrifugación (10.500 rcf, 30 min), se resuspendieron en tampón de lisis y se lisaron por sonicación en hielo (amplitud del 35 %, 10 pulsos de 20 segundos de encendido / 30 segundos de apagado). La proteína en la fracción soluble se aisló por centrifugación (35.000 rcf, 30 min) y después se enriqueció mediante purificación con Ni-NTA (4 ml de perlas, llevado a cabo como se describe para constructos de N-inteína). Después de la elución en tampón de lisis con imidazol 250 mM, se retiró el imidazol por diálisis en tampón de lisis nuevo. Entonces se llevó a cabo el ligamiento durante la noche a temperatura ambiente con la adición de TCEP 10 mM, el cóctel inhibidor de proteasa Roche Complete, MESNa 100 mM, EDTA 5 mM y CFN-$NH_2$ 5 mM (pH 7,0). El péptido de $Int^C$-CFN ligado se acidificó con TFA al 0,5 % y se purificó mediante RP-HPLC en una columna preparativa $C_{18}$: Gradiente = 10 % de B durante 10 minutos (isocrático) seguido del 20-60 % de B durante 60 minutos. La pureza de cada proteína se determinó mediante RP-HPLC analítica y su identidad se confirmó mediante ESI-MS.

Aislamiento de $Cfa^C$-link-MESNa

El péptido $Cfa^C$-link-MESNa usado para la semisíntesis de la fusión inteína-dendrímero se expresó y purificó exactamente como se describió anteriormente para los constructos $Int^C$-CFN (expresión del plásmido 27). Sin embargo, no se añadió tripéptido durante la etapa de ligamiento final, dando como resultado en su lugar tiólisis de la inteína y formación de un α-tioéster. Este α-tioéster de $Cfa^C$-MESNa se purificó luego por RP-HPLC preparativa. Las fracciones se analizaron por ESI-MS, se combinaron y se liofilizaron.

Análisis del corte y empalme en trans de proteínas mediante RP-HPLC y ESI-MS para mutantes de lote.

Se llevaron a cabo reacciones de corte y empalme según una adaptación de un protocolo descrito anteriormente.[1] Brevemente, se preincubaron N- y C-inteínas ($Int^N$ 15 μM, $Int^C$ 10 μM) individualmente en tampón de corte y empalme (fosfatos de sodio 100 mM, NaCl 150 mM, EDTA 1 mM, pH 7,2) con TCEP 2 mM durante 15 minutos. Todas las reacciones de corte y empalme se llevaron a cabo a 30 °C a menos que se indicara lo contrario. Las reacciones de corte y empalme que comparaban la tolerancia de Npu y Cfa a agentes caotrópicos se llevaron a cabo con la concentración indicada de o bien urea o bien clorhidrato de guanidina. El corte y empalme se inició mezclando volúmenes iguales de N- y C-inteínas con alícuotas retiradas en los momentos indicados y se extinguió mediante la adición de clorhidrato de guanidina 8 M, TFA al 4 % (3:1 v/v). Para todas las reacciones de corte y empalme que contenían $Npu^C$-CFN o $Cfa^C$-CFN, el progreso de la reacción se monitorizó mediante RP-HPLC. Para todas las reacciones de corte y empalme que contenían $Ssp^C$-CFN, el progreso de la reacción se monitorizó mediante ESI-MS (muestras desalinizadas con ZipTip antes de la inyección) debido a la mala resolución cromatográfica de cada estado como se observó anteriormente.[1] Se observó que el corte y empalme tanto para el lote 3 como para Cfa a 80 °C (preincubación de 15 minutos) era ineficaz, alcanzando ~50 % de finalización. Esto probablemente se deba a agregación (e inactivación) de la N-inteína. Obsérvese que preincubaciones más cortas de Cfa a 80 °C condujeron a un corte y empalme más eficaz.

Análisis cinético de reacciones de corte y empalme en trans de mutantes de lote:

El análisis cinético se llevó a cabo como se describió anteriormente.[1] Brevemente, se separan cinco especies (1-5) por RP-HPLC, y se determinan las áreas de los picos. Para ESI-MS, se calculan las áreas de los picos para las especies 1-4. Cada pico individual se normalizó frente al área total de todos los picos combinados y se representaron gráficamente las curvas de progreso de la reacción (n=3). Los datos se ajustaron entonces en ProFit a la solución analítica de la ecuación de velocidad diferencial acoplada para el modelo de corte y empalme cinético de tres estados. Debido a que el material de partida no puede separarse del tioéster lineal usando este ensayo, el modelo cinético de tres estados colapsa la etapa de unión y las dos primeras etapas de la reacción de corte y empalme en un equilibrio. Cada reacción de corte y empalme se llevó a cabo por triplicado con cada réplica analizada por separado. Se notifican la media y la desviación estándar para todos los valores (n = 3).

Análisis cinético de reacciones de corte y empalme en trans globales para Npu y Cfa

Todas las reacciones de corte y empalme que comparaban Npu y Cfa se separaron por RP-HPLC con áreas de los picos una vez más calculadas usando el software del fabricante. Para estas reacciones, se calcularon las áreas de los picos para el material de partida y producto intermedio ramificado (especie 1 y 2) y producto (especie 3, 4, 5). Los datos se ajustaron a continuación a la ecuación de velocidad de primer orden usando el software GraphPad Prism.

$$[P](t) = [P]_{máx} \cdot (1 - e^{-kt})$$

Donde [P] es la intensidad normalizada del producto, $[P]_{máx}$ es este valor a $t=\infty$ (la meseta de reacción) y k es la constante de velocidad ($s^{-1}$). Se notifican la media y la desviación estándar (n = 3).

Generación y refinamiento de la alineación de secuencias múltiples de inteína DnaE.

Se identificaron homólogos de Npu DnaE a través de una búsqueda BLAST[5] de las bases de datos NCBI[6] (colección de nucleótidos) y JGI[7] usando las secuencias de proteínas Npu DnaE. Esto condujo a la identificación de 105 proteínas con >60 % de identidad de secuencia. Para N-inteínas con colas C-terminales largas, las proteínas se truncaron a 102 residuos, la longitud de Npu. Para las N-inteínas de la base de datos JGI, el punto de truncamiento se determinó mediante los resultados del programa BLAST (el último residuo identificado en la búsqueda Blast se seleccionó como punto de truncamiento). A continuación, se generó una alineación de secuencias múltiples (MSA) de la secuencia fusionada (es decir, la N-inteína conectada a la C-inteína) de las 105 inteínas en Jalview (figura 7A).[8] Para refinar el MSA para inteínas que se predice que experimentan corte y empalme rápido, se eliminaron todas las secuencias que no contenían K70, M75, M81 y S136 (los residuos “aceleradores”) de la alineación, dejando 73 inteínas que se predice que tienen una cinética de corte y empalme rápida (figura 7B). La secuencia consenso de esta alineación refinada de inteínas rápidas (Cfa) se calculó en Jalview determinando el aminoácido que aparecía con mayor frecuencia en cada posición. No se identificó un residuo consenso en las posiciones 98 y 102 debido a la falta de homología en la alineación y, por lo tanto, la secuencia consenso se truncó a 101 aminoácidos y la posición 98 se fijó al residuo encontrado en Npu DnaE. Esta secuencia consenso se alineó luego con Npu DnaE en Jalview para calcular su porcentaje de identidad. Se mapearon residuos no idénticos sobre la estructura cristalina de Npu DnaE (pdb = 4K15) (figura 1).

Las figuras 7A y 7B muestran una alineación y refinamiento de la familia de inteínas DnaE. La figura 7A muestra la alineación de secuencias múltiples (MSA) de los 105 miembros de la familia de inteínas DnaE encontradas a partir de una búsqueda BLAST de las bases de datos de secuencias JGI y NCBI. Las ubicaciones de los residuos “aceleradores” usados para filtrar la alineación se indican con flechas negras. La figura 7B muestra la MSA de las 73 inteínas DnaE que se predice que demuestran una cinética de corte y empalme rápida debido a la presencia de los cuatro residuos “aceleradores”.

Examen de *E. coli* $Kan^R$ para determinar la dependencia de Cfa exteína.

El ensayo de resistencia a la kanamicina acoplado al corte y empalme de proteínas ($Kan^R$) se llevó a cabo como se describió anteriormente.[2,9] Brevemente, un plásmido que codifica una aminoglucósido fosfotransferasa fragmentada fusionada a una inteína dividida (Cfa) con cualquiera de F, G, R o E presente en la posición +2 de la C-exteína (plásmidos 22-25) se transformó en células competentes DH5α y se cultivó en cultivos iniciadores durante la noche (caldo LB, 100 μg/ml de ampicilina, 18 h). Estos cultivos se diluyeron luego veinte veces en una placa de 96 pocillos, y se midió el crecimiento de *E. coli* a diversas concentraciones de kanamicina (2,5, 10, 25, 50, 100, 250, 1000 μg/ml de kanamicina con 100 μg/ml de ampicilina). La densidad óptica celular a 650 nm ($DO_{650}$) en el punto final de 24 horas se ajustó a una curva de respuesta a la dosis con pendiente variable.

$$DO_{obs} = DO_{mín} + \frac{(DO_{máx} - DO_{mín})}{1 + 10^{[logCI_{50} - \log[Kan])\cdot pendiente}}$$

Donde la $DO_{mín}$ se fijó a la absorbancia del fondo a 650 nm. Cada ensayo se llevó a cabo por triplicado, se ajustó por separado y los valores de $CI_{50}$ se notifican como la media y la desviación estándar de $CI_{50}$ para estas tres mediciones separadas.

Corte y empalme en trans de proteínas del cuerpo de inclusión extraído

Los cuerpos de inclusión de *E. coli* que contenían expresión de $His_6$-Sumo-$Cfa^N$ (plásmido 20) se resuspendieron y se extrajeron durante la noche a 4 °C en tampón de lisis que contenía urea 6 M. Después de la centrifugación (35.000 rcf, 30 min), se retiró el sobrenadante y se enriqueció la proteína enriquecida con Ni-NTA en condiciones desnaturalizantes (como se describió anteriormente). Sin embargo, en lugar de replegar la proteína, se inició directamente el corte y empalme en trans mediante la adición de $Cfa^C$-CFN ($Cfa^C$ 10 μM, TCEP 2 mM, EDTA 2 mM, 2 horas, TA). El progreso de la reacción se monitorizó mediante SDS-PAGE.

Expresión de la prueba de αDec205-HC-$Int^N$ y corte y empalme

Prueba de expresión de HC-$Npu^N$, HC-$Mcht^N$, HC-$Ava^N$, HC-$Cfa^N$

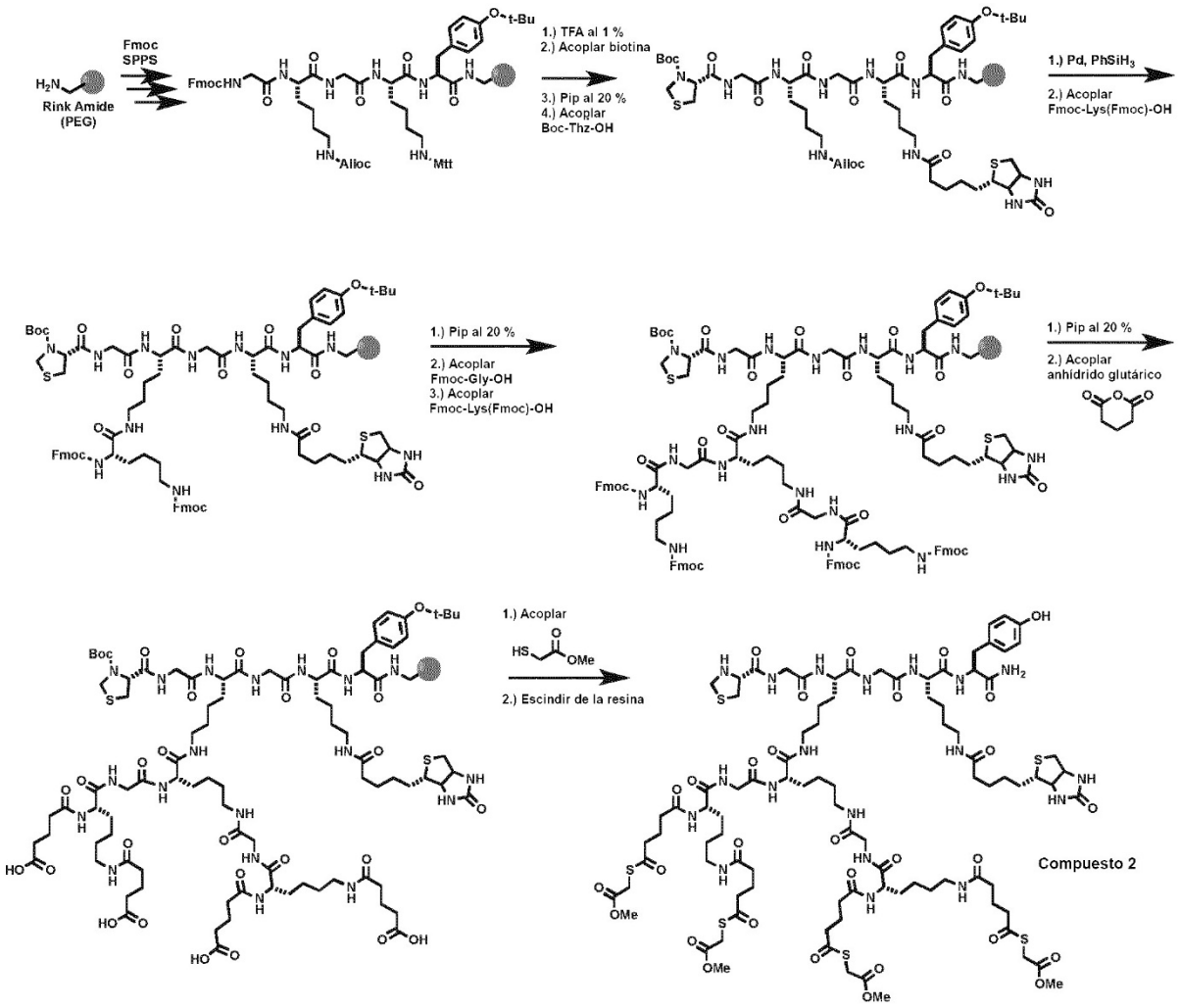
La expresión de todos los constructos de mAb se llevó a cabo como se describió anteriormente.[3] Brevemente, se cotransfectaron plásmidos que codificaban el αDec205-LC y el αDec205-HC-$Int^N$ en células HEK293T y se incubaron durante 96 h (5 % de $CO_2$). Las células se centrifugaron (5 minutos, 1.000 rcf), se mezclaron 15 μl de medio para cada fusión de inteína con 5 μl de colorante de carga 4x y se ejecutaron en un gel Bis-Tris al 12 % en tampón de ejecución MES-SDS (170 V durante 50 minutos). Entonces, se analizó la proteína mediante inmunotransferencia de tipo Western (transferida a una membrana de PVDF, transferencia frente a IgG de ratón α). El rendimiento de expresión se midió como la cantidad de HC-$Int^N$ en los medios tal como se determinó por densitometría. Para tener en cuenta el crecimiento y la supervivencia variables de las células, el rendimiento se normalizó usando una transferencia de α-actina del lisado de células HEK293T (sonicación de 5 s, amplitud del 35 %, en colorante de carga 1x) y después se representó en relación con la expresión de HC-$Cfa^N$. Se llevaron a cabo cuatro réplicas de esta expresión de prueba, y se calculó la media con el error representado como la desviación estándar.

Corte y empalme en trans de proteínas en medios de crecimiento

Después de la expresión de 96 h a 37 °C de los constructos mAB-$Ava^N$ y mAB-$Cfa^N$ descritos anteriormente, el medio se centrifugó (1.000 rcf, 5 minutos). El sobrenadante se mezcló luego con el péptido $Cfa^C$-CFN (semisíntesis del plásmido expresado 21) y se incubó durante 2 horas a temperatura ambiente ($Cfa^C$-CFN 1 μM, TCEP 2 mM, EDTA 2 mM). Las reacciones de corte y empalme se analizaron mediante SDS-PAGE (Bis-Tris al 12 % en tampón de ejecución MES-SDS a 170 V durante 50 minutos) seguido de inmunotransferencia de tio Western (IgG de ratón α).

Síntesis de péptidos y dendrímeros

**Cys-Gly-Lys (fluoresceína)**. Este péptido se sintetizó mediante la adición manual de reactivos en la resina Rink Amide de acuerdo con un procedimiento publicado anteriormente.[2]




Esquema complementario 1

**Compuesto 2 (tioéster de dendrímero).** Este compuesto se sintetizó en la fase sólida usando la ruta descrita en el esquema complementario 1 en una escala de 400 mg de resina Rink Amide (sustitución: 0,47 mmol/g, 188 μmol). En primer lugar se proporcionan los procedimientos generales, seguido de cualquier método específico para este péptido. El grupo Fmoc se retiró con 3 ml de piperidina al 20 % en DMF y se realizó dos veces (una desprotección durante 30 segundos seguido de una desprotección adicional durante 15 minutos). Después de cada etapa de desprotección, así como todas las etapas de síntesis posteriores, se usaron lavados de flujo (3 x 5 s con ~5 ml de DMF cada uno). El acoplamiento se realizó usando 4 eq. de monómero, 4 eq. de HBTU y 8 eq. de DIPEA sin preactivación a menos que se indique lo contrario. Se usaron acoplamientos dobles para todos los residuos para garantizar la acilación completa.

El grupo protector de tritilo se retiró selectivamente usando TFA al 1 %, TIS al 5 % en DCM usando un total de 30 ml (10x 3 ml) de cóctel de desprotección. Un lavado exhaustivo de la resina con DCM tanto durante como después de estos ciclos garantizó la eliminación de cualquier especie de tritilo liberada. La resina también se neutralizó con DIPEA al 5 % en DMF antes de que se realizara el siguiente acoplamiento. El grupo Alloc se desprotegió usando 0,1 eq. de tetraquis(trifenilfosfina)paladio (0), 20 eq. de fenilsilano en DCM durante 3x 45 min cada uno. Se usó un lavado exhaustivo de la resina con DCM durante y después de estos ciclos, así como un lavado con DIPEA al 5 % en DMF antes del siguiente acoplamiento. El monómero de anhídrido glutárico se usó como ácido dicarboxílico preactivado para permitir la formación de los tioésteres (es decir, para funcionalizar un ácido carboxílico unido a resina libre). Se añadieron 20 eq. de anhídrido glutárico y 10 eq. de DIPEA (en relación con el número de aminas que van a acilarse) a la resina y se dejó reaccionar durante una hora. Después, se lavó la resina y se repitió el acoplamiento para garantizar la reacción completa de las aminas primarias unidas a la resina. Para formar los tioésteres unidos a resina, se añadieron 30 eq. de tioglicolato de metilo, 5 eq. de PyAOP y 10 eq. de DIPEA (en relación con el número de carboxilatos) en DMF a la resina y se dejó reaccionar durante una hora. La resina se lavó con DMF en exceso y el procedimiento de acoplamiento se repitió dos veces más.

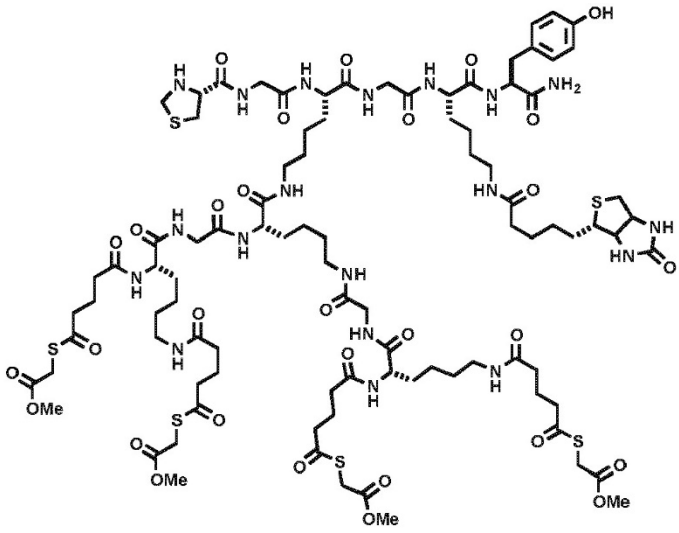
La escisión se realizó con el 95 % de TFA, el 2,5 % de TIS y el 2,5 % de $H_2O$ durante dos horas a temperatura ambiente. Después, el péptido se precipitó con dietil éter, se disolvió en agua con TFA al 0,1 % y se analizó mediante

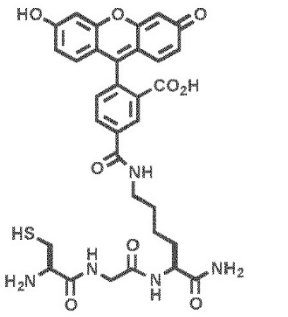
RP-HPLC. El material bruto se purificó mediante RP-HPLC a escala semipreparativa, y se analizaron las fracciones deseadas, se agruparon y se liofilizaron. Caracterización por RP-HPLC: gradiente 0-73 % de B, $t_r$ = 18,4 min. Masa esperada: 2198,86 Da. Hallada: 2198,82 Da.

Compuesto 3 (dendrímero fluoresceína).

El compuesto 3 se sintetizó mediante ligamiento químico nativo (esquema 2). El compuesto 2 se disolvió en tampón de ligamiento y se mezcló con cinco eq. de Cys-Gly-Lys(fluoresceína) (**2** 1 mM, péptido 5 mM, guanidina 4 M, fosfato 100 mM, NaCl 150 mM, MPAA 100 mM, TCEP 20 mM, pH 7,0) y se dejó reaccionar durante la noche a temperatura ambiente. La desprotección de la tiazolidina se logró mediante la adición de metoxiamina 0,1 M (concentración final) y disminuyendo el pH del tampón de ligamiento hasta 4,0 (durante la noche, TA).

Al intentar purificar el compuesto 3 por RP-HPLC, los inventores notaron que presentaba poca solubilidad cuando se acidificó y se diluyó en agua. Sin embargo, Cys-Gly-Lys(fluoresceína), MPAA y metoxiamina permanecieron en disolución. A partir de esta observación, se purificó 3 por precipitación selectiva después de una dilución de 10 veces en agua con TFA al 0,1 %. El polvo precipitado se aisló por centrifugación (17.000 rcf, 5 min) y luego se redisolvió (fosfato 100 mM, NaCl 150 mM, pH 7,2) para eliminar por lavado cualquier contaminante restante. Una vez más, la disolución se precipitó por acidificación y se aisló por centrifugación (17.000 rcf, 5 min). Este polvo aislado se liofilizó después. Masa esperada: 4417,8 Da. Hallada: 4417,5 Da.



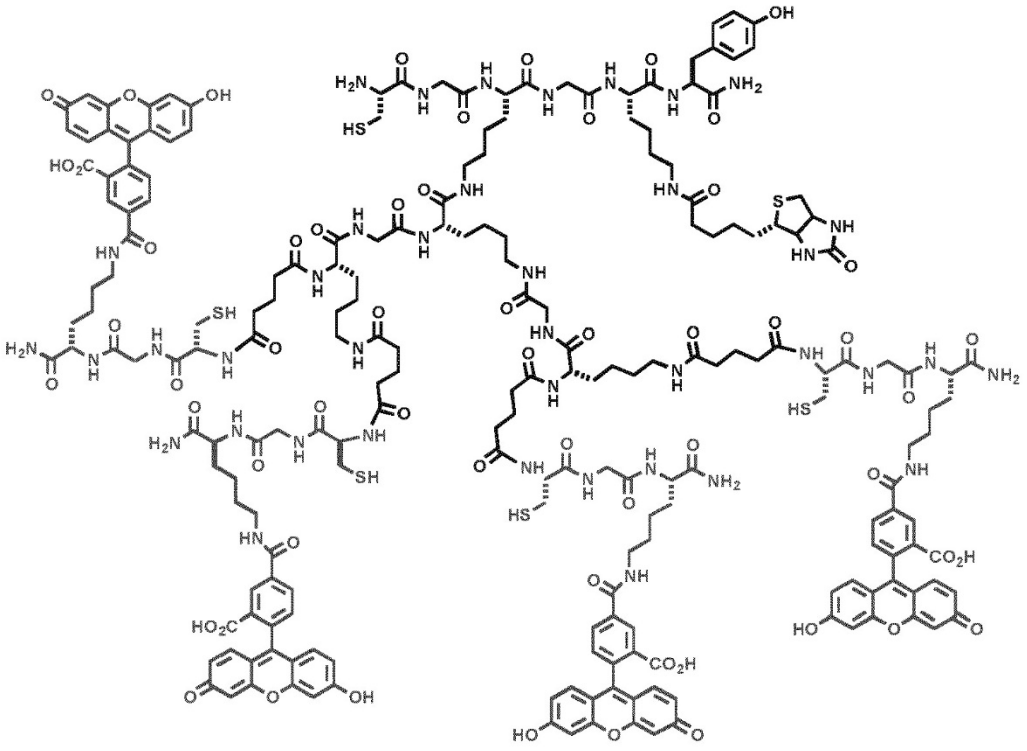


CGK (Fluoresceína)

1.) Ligamiento químico nativo

2.) Desprotección de tiazolidina
$H_2N$-OMe

Compuesto 2



Compuesto 3

**Esquema complementario 2.** El ligamiento químico nativo se usó para elaborar el compuesto 2 que contiene tetratioéster con un tripéptido fluorescente. La desprotección posterior usando metoxilamina se usó para exponer la cisteína N-terminal para ligamientos adicionales que produjeron dendrímero tetrafuncionalizado, compuesto 3.

**Compuesto 1**: ($Cfa^C$-dendrímero)

El compuesto 1 se sintetizó mediante ligamiento de proteínas expresadas. El compuesto 3 se disolvió en tampón de ligamiento y se mezcló con 1,5 eq. de tioéster de $Cfa^C$-MESNa (3 100 μM, $Cfa^C$-MESNa 150 μM, guanidina 4 M, fosfato 100 mM, NaCl 150 mM, TCEP 20 mM, MPAA 100 mM). Se dejó que la reacción prosiguiera durante la noche a temperatura ambiente. El producto ligado se purificó luego por RP-HPLC semipreparativa. Las fracciones deseadas se agruparon y se liofilizaron. Masa esperada: 9860,8 Da. Hallada: 9860,3 Da.

Corte y empalme en trans de proteínas de dendrímero con mAb αDec205

El mAb αDec205 con $Cfa^N$ fusionado a su extremo C-terminal se expresó como se describió anteriormente. Después de la expresión de 96 h, el medio se concentró 10 veces en un concentrador Amicon 30K (0,5 ml). El compuesto 1 se disolvió en tampón de corte y empalme (fosfato 100 mM, NaCl 150 mM, EDTA 1 mM, pH 7,2) y después se mezcló con los medios concentrados (compuesto 1 2 μM, TCEP 2 mM, EDTA 1 mM) y dejó que la reacción prosiguiera durante 2 horas a temperatura ambiente. Entonces se analizó la mezcla de corte y empalme por SDS-PAGE (Bis-Tris al 12 % ejecutado en tampón de ejecución MES-SDS a 170 V durante 50 minutos) y se obtuvieron imágenes en un generador de imágenes de fluorescencia. A esto le siguió la transferencia a una membrana de PVDF y el análisis de inmunotransferencia de tipo Western (IgG de ratón α).

La invención permite la formación de diversos complejos entre un fragmento de inteína dividida y un compuesto. Varios de tales complejos y compuestos se ilustran en la tabla de la figura 11. IntC es un fragmento de inteína dividida, por ejemplo, un fragmento C de inteína dividida. Por ejemplo, el dendrímero puede tener la forma del compuesto 2, compuesto 3, o porciones de estos. Por ejemplo, la carga puede ser un colorante (por ejemplo, fluoresceína), otra molécula marcadora, un fármaco (por ejemplo, una molécula citotóxica, tal como se usa en el tratamiento del cáncer) o un nucleótido. Por ejemplo, el polipéptido puede ser un polipéptido total o parcialmente sintético o de origen natural o una porción del mismo. Un dendrímero puede ser una molécula que tiene una estructura química ramificada sobre la que pueden “cargarse” una o más moléculas de “carga”. Una molécula de “carga” puede ser una molécula sintética de origen natural. La molécula de carga puede estructurarse para que no tenga 1,2-aminotioles o 1,2-aminoalcoholes libres. Cuando la inteína se une a través de un aminotiol o aminoalcohol a un polipéptido, como se muestra en la fila 3 de la tabla de la figura 11, el complejo formado puede considerarse como una proteína de fusión recombinante.

*Ejemplo 2*

Una advertencia importante para los métodos basados en corte y empalme es que todas las inteínas caracterizadas presentan una preferencia de secuencia en los residuos de exteína adyacentes al sitio de corte y empalme. Además de un residuo de Cys, Ser o Thr catalítico obligatorio en la posición +1 (es decir, el primer residuo dentro de la C-exteína), existe un sesgo para residuos que se asemejan a la secuencia de N- y C-exteína proximal encontrada en el sitio de inserción nativo. La desviación de este contexto de secuencia preferido conduce a una marcada reducción en la actividad de corte y empalme, lo que limita la aplicabilidad de los métodos basados en PTS.[23, 24] En consecuencia, existe la necesidad de inteínas divididas cuyas actividades se vean mínimamente afectadas por el entorno de secuencia local. Para las inteínas DnaE, las preferencias de secuencia de exteína se limitan en gran medida a la cisteína catalítica en la posición +1 y los residuos hidrófobos grandes que se prefieren en la posición +2.[25]

En este ejemplo, se diseñó una mutación de bucle “EKD” a “GEP” en los residuos 122-124 de Cfa ($Cfa_{GEP}$) y dio como resultado una mayor promiscuidad en la posición +2 de la C-exteína en un ensayo de resistencia a kanamicina (figura 9). La mutación EKD → GEP aumenta la actividad de Cfa en una amplia gama de contextos de exteína. Además, puede esperarse razonablemente que estas mismas mutaciones (o similares) aumenten la promiscuidad entre otros miembros de la familia de inteínas DnaE (incluidas Npu y las enumeradas en las figuras 7A y 7B).

Las siguientes secuencias representan las inteínas modificadas por ingeniería genética:

La C-inteína Cfa con la mutación “GEP” que confiere más actividad “promiscua” de acuerdo con una realización de la invención es:

VKIISRKSLGTQNVYDIGVGEPHNFLLKNGLVASN (SEQ ID NO: 389).

Un ejemplo de una inteína de fusión de la N-inteína Cfa y la C-inteína Cfa con la mutación “GEP” de SEQ ID: 389) es:

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGE

QEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLPVKIISRKSLGTQN

VYDIGVGEPHNFLLKNGLVASN (SEQ ID NO: 390).

La figura 9 muestra un esquema y una tabla que muestran el aumento de la promiscuidad de $Cfa_{GEP}$. El panel A muestra un esquema que representa el sistema de selección de *E. coli* dependiente de PTS con la inteína dividida Cfa. La proteína de resistencia a kanamicina, KanR, se divide y se fusiona con fragmentos de N- y C-inteína ($Cfa^N$ y $Cfa^C$). El residuo de C-exteína +2 (X roja) varía en el sistema. En el panel B, se muestran los valores de $CI_{50}$ para la resistencia a kanamicina de las inteínas $Cfa_{EKD}$ (WT) y $Cfa_{GEP}$ (GEP) con el residuo de C-exteína +2 indicado (error = error estándar (n = 3)).

Además, esta misma tolerancia para secuencias de exteína variables también se observó en la ciclación de eGFP en *E. coli* (figura 10). La inteína $Cfa_{GEP}$ demostró rendimientos mejorados del producto ciclado en todos los contextos desfavorables de C-exteína +2 sometidos a prueba (figura 10 panel A, figura 10 panel B). Además, $Cfa_{GEP}$ mantiene esta actividad de ciclación mejorada incluso cuando las posiciones de exteína -1 y +3 varían (figura 10 panel C, figura

10 panel D). Esta secuencia de bucle “GEP” modificada por ingeniería genética, que no se ha identificado en una inteína DnaE dividida de manera natural de tipo silvestre, por lo tanto, debe expandir la amplitud de las proteínas y péptidos accesibles a las tecnologías basadas en PTS.

La figura 10 muestra esquemas y gráficos que muestran la ciclación de eGFP con la inteína dividida $Cfa_{GEP}$. El panel A es un esquema que representa la ciclación de eGFP en *E. coli* con residuos variables en la posición de C-exteína +2 (X roja). En el panel B, se muestra la fracción de eGFP ciclada formada después de la expresión durante la noche en *E. coli* para $Cfa_{EKD}$ (WT) y $Cfa_{GEP}$ (GEP) con el residuo de C-exteína +2 indicado (media ± desviación estándar, n = 3). El panel C es un esquema que representa la ciclación de eGFP en *E. coli* con residuos variables en la posición de C-exteína +3 (X azul) y la posición de N-exteína -1 (X roja). El panel D muestra una fracción de eGFP ciclada formada después de la expresión durante la noche en *E. coli* para $Cfa_{EKD}$ (WT) y $Cfa_{GEP}$ (GEP) con los residuos de C-exteína +3 y N-exteína -1 indicados (media ± desviación estándar, n = 3).

Los expertos en la técnica apreciarán que pueden configurarse diversas adaptaciones y modificaciones de la realización preferida recién descrita sin apartarse del alcance de la invención. La realización ilustrada se ha expuesto solo con fines de ejemplo y no debe tomarse como limitativa de la invención. Por lo tanto, debe entenderse que, dentro del alcance de las reivindicaciones adjuntas, la invención puede ponerse en práctica de forma distinta a la descrita específicamente en el presente documento.

**Bibliografía**

(1) Shah, N. H.; Muir, T. W. *Chem. Sci.* 2014, 5, 15.

(2) Wu, H.; Hu, Z.; Liu, X. Q. *Proc. Natl. Acad. Sci. U. S. A.* 1998, 95, 9226.

(3) Iwai, H.; Zuger, S.; Jin, J.; Tam, P. H. *FEBS Lett*. 2006, 580, 1853.

(4) Zettler, J.; Schutz, V.; Mootz, H. D. FEES Lett. 2009, 583, 909.

(5) Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc*. 2013, 135, 5839.

(6) Shah, N. H.; Dann, G. P.; Vila-Perello, M.; Liu, Z.; Muir, T. W. *J. Am. Chem. Soc*. 2012, 134, 11338.

(7) Carvajal-Vallejos, P.; Pallisse, R.; Mootz, H. D.; Schmidt, S. R. *J. Biol. Chem* 2012, 287, 28686.

(8) Wu, Q.; Gao, Z.; Wei, Y.; Ma, G.; Zheng, Y.; Dong, Y.; Liu, Y. *Biochem. J.* 2014, 461, 247.

(9) Aranko, A. S.; Oeemig, J. S.; Kajander, T.; Iwai, H. *Nat. Chem. Biol*. 2013, 9, 616.

(10) Pietrokovski, S. *Protein Sci.* 1994, 3, 2340.

(11) Dearden, A. K.; Callahan, B.; Roey, P. V.; Li, Z.; Kumar, U.; Belfort, M.; Nayak, S. K. *Protein Sci*. 2013, 22, 557.

(12) Du, Z.; Shemella, P. T.; Liu, Y.; McCallum, S. A.; Pereira, B.; Nayak, S. K.; Belfort, G.; Belfort, M.; Wang, C. *J. Am. Chem. Soc*. 2009, 131, 11581.

(13) Lehmann, M.; Kostrewa, D.; Wyss, M.; Brugger, R.; D'Arcy, A.; Pasamontes, L.; van Loon, A. P. *Protein Eng.* 2000, 13, 49.

(14) Steipe, B. *Methods Enzymol*. 2004, 388, 176.

(15) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol*. 1990, 215, 403.

(16) Grigoriev, I. V.; Nordberg, H.; Shabalov, I.; Aerts, A.; Cantor, M.; Goodstein, D.; Kuo, A.; Minovitsky, S.; Nikitin, R.; Ohm, R. A.; Otillar, R.; Poliakov, A.; Ratnere, I.; Riley, R.; Smirnova, T.; Rokhsar, D.; Dubchak, I. *Nucleic Acids Res*. 2012, 40, D26.

(17) Tatusova, T.; Ciufo, S.; Fedorov, B.; O'Neill, K.; Tolstoy, I. *Nucleic Acids Res*. 2014, 42, D553.

(18) Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc*. 2013, 135, 18673.

(19) Mohlmann, S.; Bringmann, P.; Greven, S.; Harrenga, A. *BMC Biotechnol*. 2011, 11, 76.

(20) Barbuto, S.; Idoyaga, J.; Vila-Perello, M.; Longhi, M. P.; Breton, G.; Steinman, R. M.; Muir, T. W. *Nat. Chem. Biol*. 2013, 9, 250.

(21) Vila-Perello, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W. *J. Am. Chem. Soc*. 2013, 135, 286.

(22) Shah, N. D.; Parekh, H. S.; Steptoe, R. J. *Pharm. Res*. 2014, 31, 3150.

(23) Iwai, H.; Zuger, S.; Jin, J.; Tam, P. H. *FEBS Lett*. 2006, 580, 1853.

(24) Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M. *Proc Natl Acad Sci U S A* 2009, 106, 11005.

(25) Cheriyan, M.; Pedamallu, C. S.; Tori, K.; Perler, F. *J Biol Chem* 2013, 288, 6202.

**REIVINDICACIONES**

1. Fragmento N de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, el 99 % o el 100 % de identidad de secuencia con

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEV FEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGL (SEQ ID NO: 1) o con

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEV FEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 2).

2. Complejo que comprende el fragmento N de inteína dividida según la reivindicación 1 y un compuesto.

3. Complejo según la reivindicación 2, en el que el compuesto se selecciona del grupo que consiste en un péptido o un polipéptido, una cadena de anticuerpo, una cadena pesada de anticuerpo, un péptido, un oligonucleótido, un fármaco o una molécula citotóxica.

4. Fragmento C de inteína dividida que comprende una secuencia de aminoácidos de al menos el 98 %, el 99 % o el 100 % de identidad de secuencia con

VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 3), con

MVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 4) o con

VKIISRKSLGTQNVYDIGVGEPHNFLLKNGLVASN (SEQ ID NO: 389).

5. Complejo que comprende el fragmento C de inteína dividida según la reivindicación 4 y un compuesto.

6. Complejo según la reivindicación 5, en el que el compuesto se selecciona del grupo que consiste en:

(i) un péptido o un polipéptido,

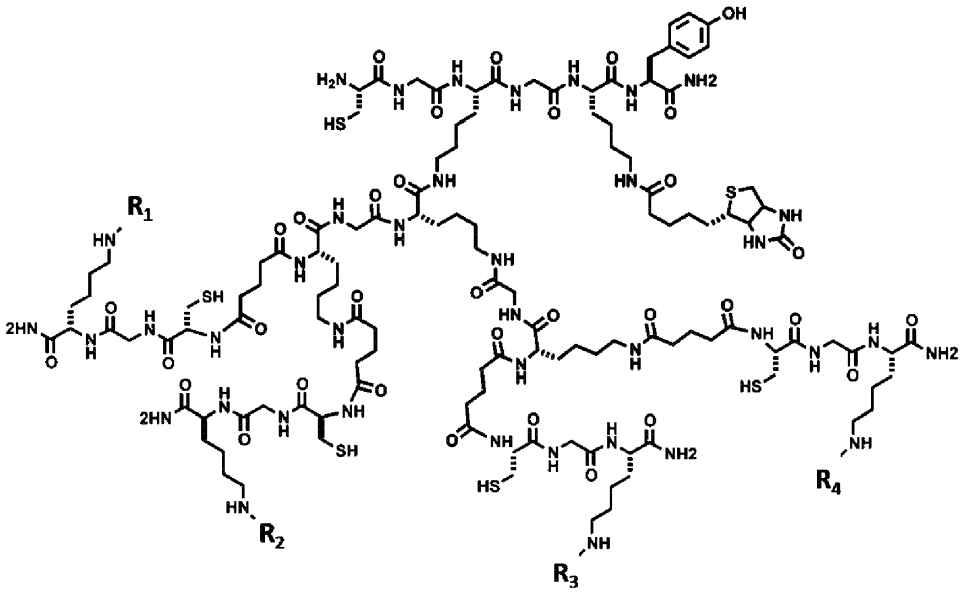
(ii) un compuesto que comprende un péptido, un oligonucleótido, un fármaco o una molécula citotóxica,

(iii) un 1,2-aminotiol unido a un péptido, un oligonucleótido, un fármaco o una molécula citotóxica,

(iv) un 1,2-aminoalcohol unido a un péptido, un oligonucleótido, un fármaco o una molécula citotóxica y

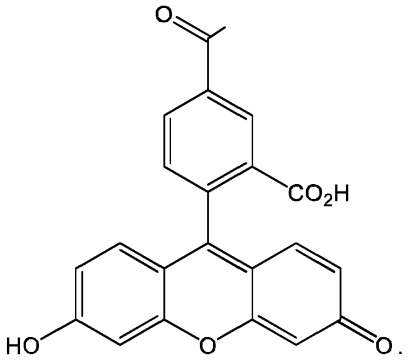
(v) un dendrímero,

(vi) un dendrímero que tiene la estructura

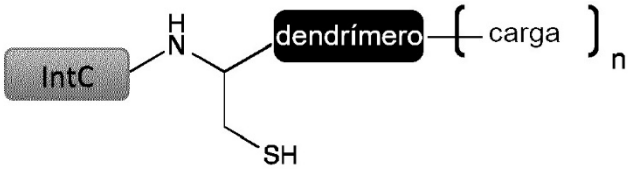


en donde R1, R2, R3 y R4 se seleccionan independientemente del grupo que consiste en hidrógeno (H) y moléculas de carga.

7. Complejo según la reivindicación 6, en donde R1, R2, R3 y R4 son cada uno una molécula de colorante o en donde R1, R2, R3 y R4 son cada uno un derivado de fluoresceína que tiene la estructura



8. Complejo de la estructura

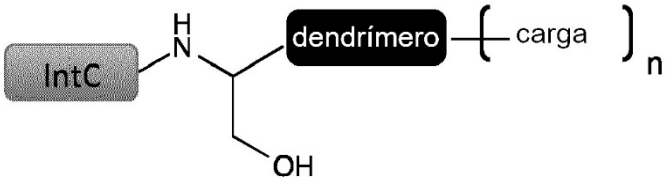



,

en donde IntC es el fragmento C de inteína dividida según la reivindicación 4 y

en donde n es de 0 a 8,

la estructura

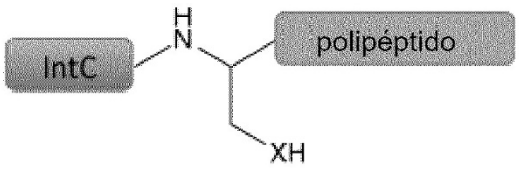



,

en donde IntC es el fragmento C de inteína dividida según la reivindicación 4 y

en donde n es de 0 a 8,

o la estructura




,

en donde IntC es el fragmento C de inteína dividida según la reivindicación 4 y

en donde X es azufre (S) u oxígeno (O).

9. Composición que comprende:

el fragmento N de inteína dividida según la reivindicación 1; y

el fragmento C de inteína dividida según la reivindicación 4.

10. Plásmido de nucleótidos que comprende una secuencia de nucleótidos que codifica el fragmento N de inteína dividida según la reivindicación 1 o el fragmento C de inteína dividida según la reivindicación 4.

11. Método para cortar y empalmar dos complejos que comprende:

poner en contacto un primer complejo que comprende un primer compuesto y el fragmento N de inteína dividida según la reivindicación 1 y un segundo complejo que comprende un segundo compuesto y el fragmento C de inteína dividida según la reivindicación 4,

en donde la puesta en contacto se realiza en condiciones que permiten la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

hacer reaccionar el producto intermedio de inteína para formar un conjugado del primer compuesto con el segundo compuesto.

12. Método seleccionado del grupo que comprende:

(i) un método que comprende

poner en contacto un primer complejo que comprende un primer compuesto y el fragmento N de inteína dividida según la reivindicación 1 y un segundo complejo que comprende un segundo compuesto y el fragmento C de inteína dividida según la reivindicación 4,

en donde la puesta en contacto se realiza en condiciones que permiten la unión del fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

hacer reaccionar el producto intermedio de inteína con un nucleófilo para formar un conjugado del primer compuesto con el nucleófilo

y

(ii) un método que comprende

fusionar una primera secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento N de inteína dividida según la reivindicación 1,

con una segunda secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento C de inteína dividida según la reivindicación 4,

de modo que la fusión de la primera secuencia de nucleótidos y la segunda secuencia de nucleótidos codifica una inteína contigua.

13. Método según la reivindicación 12, en el que el primer compuesto es un polipéptido o un anticuerpo y/o en el que el segundo compuesto es un dendrímero o un polipéptido.

14. Inteína que comprende una secuencia de aminoácidos de al menos el 90 %, el 95 %, el 98 %, el 99 % o el 100 % de identidad de secuencia con

CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEV

FEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLPVKIISRKSL

GTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 390).

15. Kit para cortar y empalmar dos complejos entre sí que comprende:

el fragmento N de inteína dividida según la reivindicación 1;

el fragmento C de inteína dividida según la reivindicación 4;

un reactivo para unir el fragmento N de inteína dividida al fragmento C de inteína dividida para formar un producto intermedio de inteína; y

un agente nucleófilo.

16. Fusión génica que comprende:

una primera secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento N de inteína dividida según la reivindicación 1

fusionada con una segunda secuencia de nucleótidos que codifica una secuencia de aminoácidos del fragmento C de inteína dividida según la reivindicación 4.

17. Polinucleótido que codifica el fragmento N de inteína dividida según la reivindicación 1 o que codifica el fragmento C de inteína dividida según la reivindicación 4.

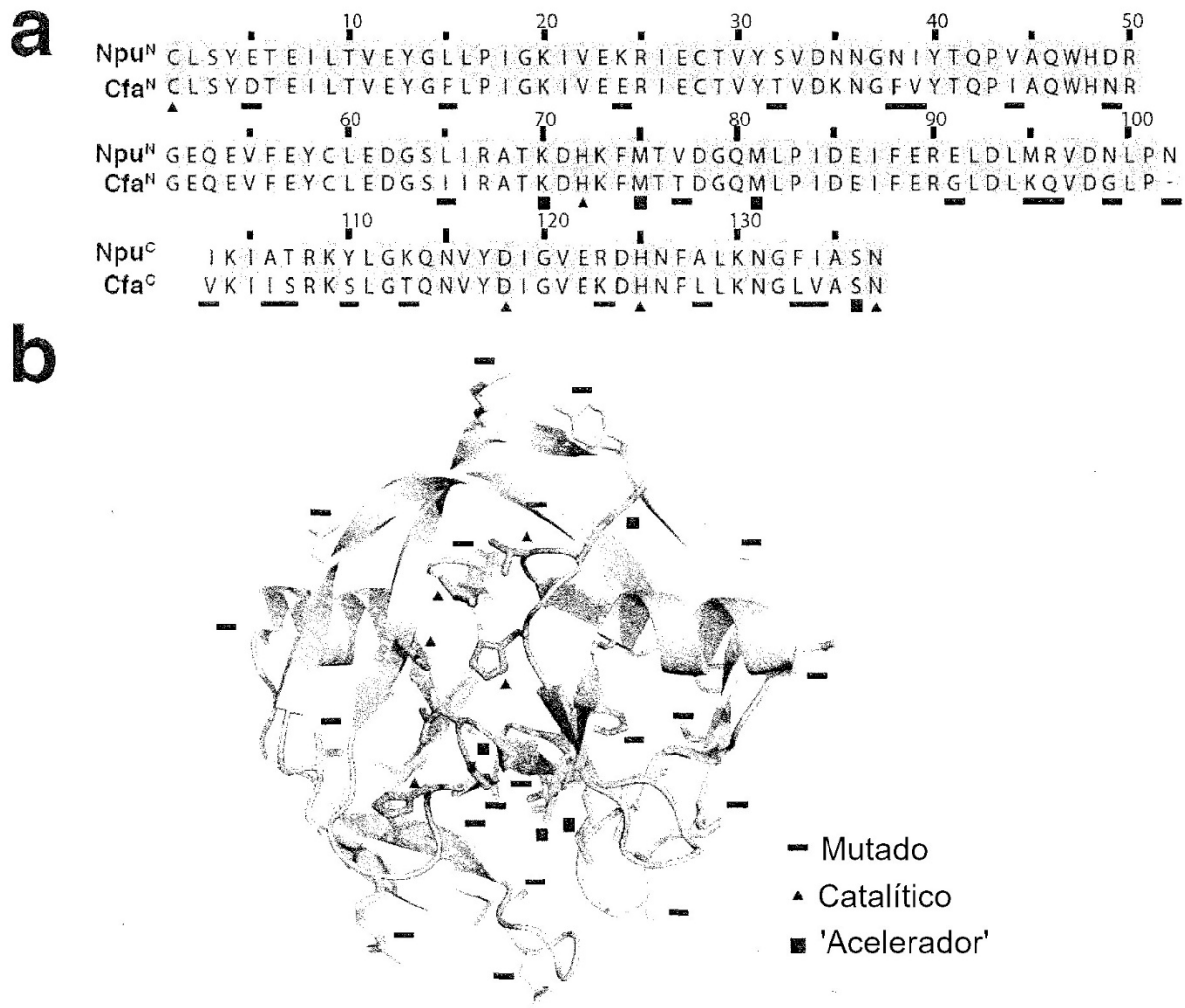

a
10
20
30
40
50
Npu^N CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDR
Cfa^N CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNR
60
70
80
90
100
Npu^N GEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN
Cfa^N GEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP-
110
120
130
Npu^C IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN
Cfa^C VKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN
b
Mutado
Catalítico
'Acelerador'

FIG. 1

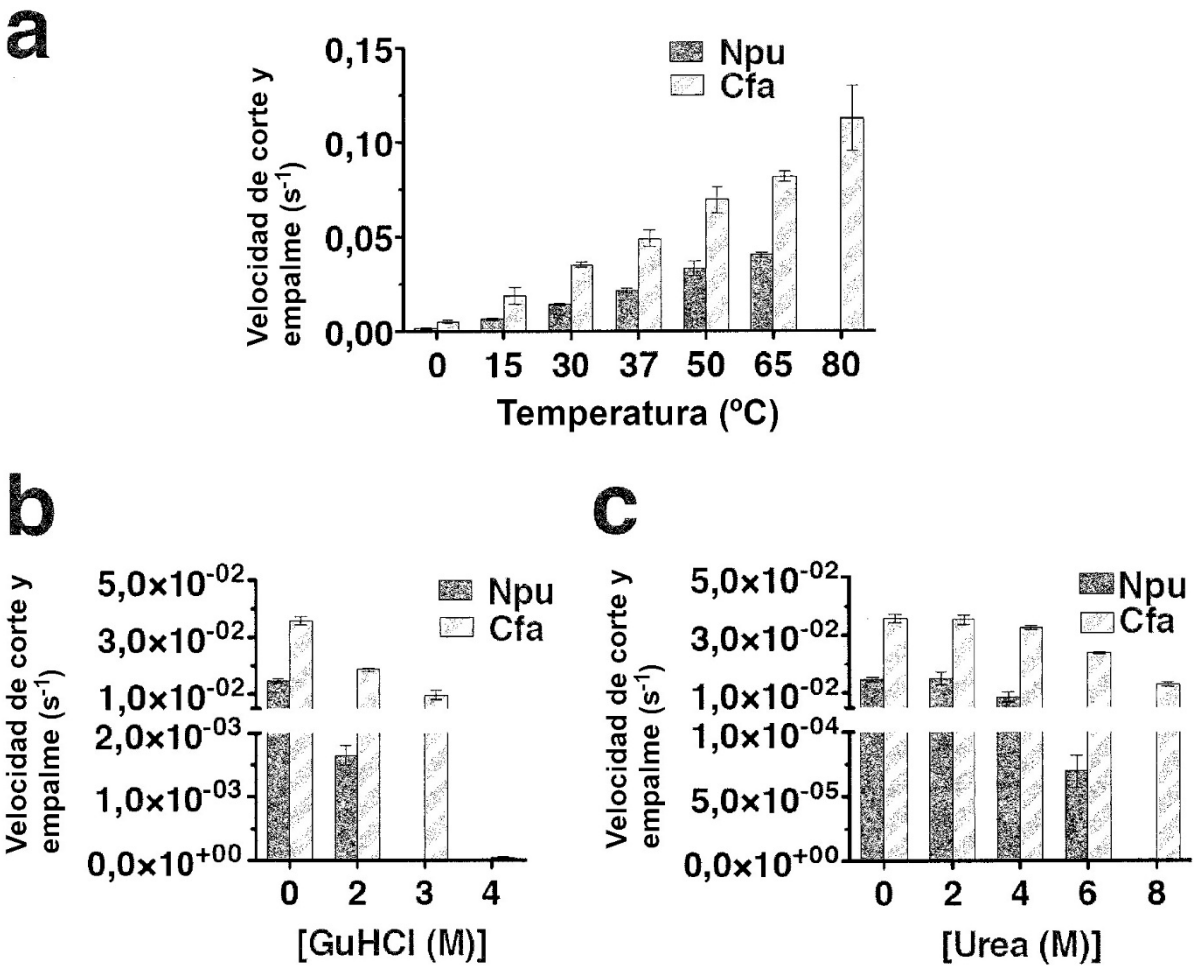
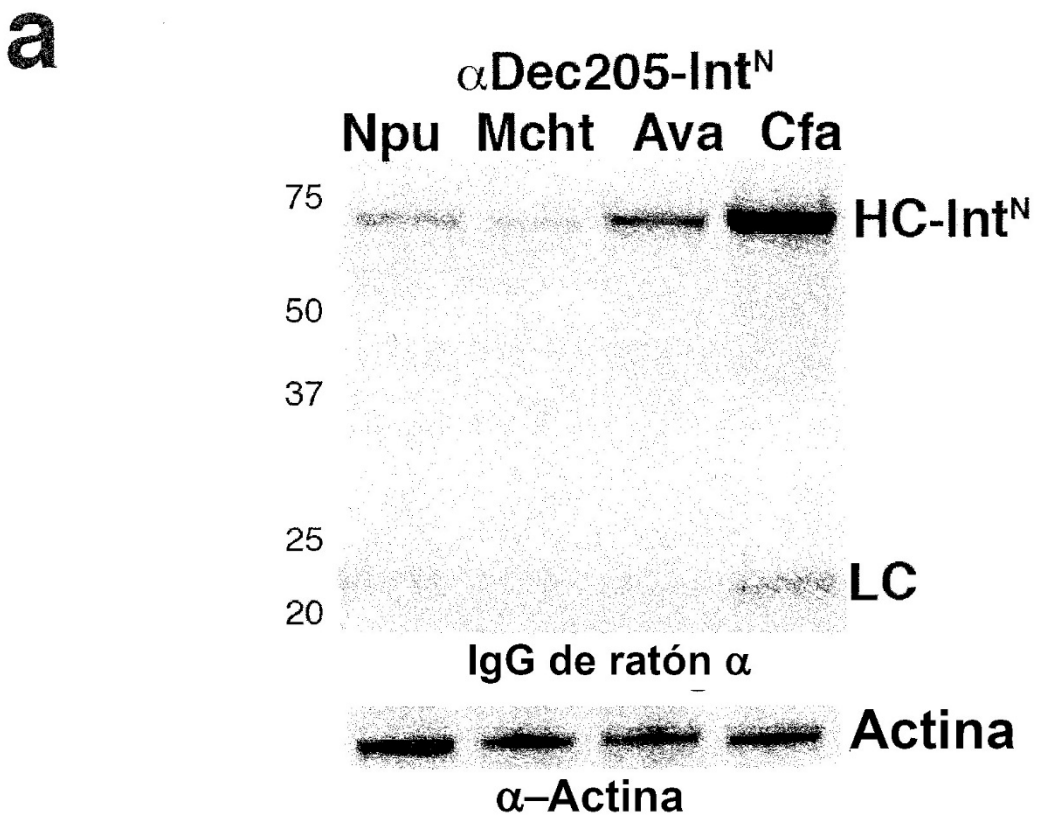

a
Velocidad de corte y empalme ($s^{-1}$)
0,15
0,10
0,05
0,00
Npu
Cfa
0 15 30 37 50 65 80
Temperatura (ºC)
b
Velocidad de corte y empalme ($s^{-1}$)
$5,0\times10^{-02}$
$3,0\times10^{-02}$
$1,0\times10^{-02}$
$2,0\times10^{-03}$
$1,0\times10^{-03}$
$0,0\times10^{+00}$
Npu
Cfa
0 2 3 4
[GuHCl (M)]
c
Velocidad de corte y empalme ($s^{-1}$)
$5,0\times10^{-02}$
$3,0\times10^{-02}$
$1,0\times10^{-02}$
$1,0\times10^{-04}$
$5,0\times10^{-05}$
$0,0\times10^{+00}$
Npu
Cfa
0 2 4 6 8
[Urea (M)]

FIG. 2


a
αDec205-Int^N
Npu Mcht Ava Cfa
75
50
37
25
20
HC-Int^N
LC
IgG de ratón α
Actina
α–Actina

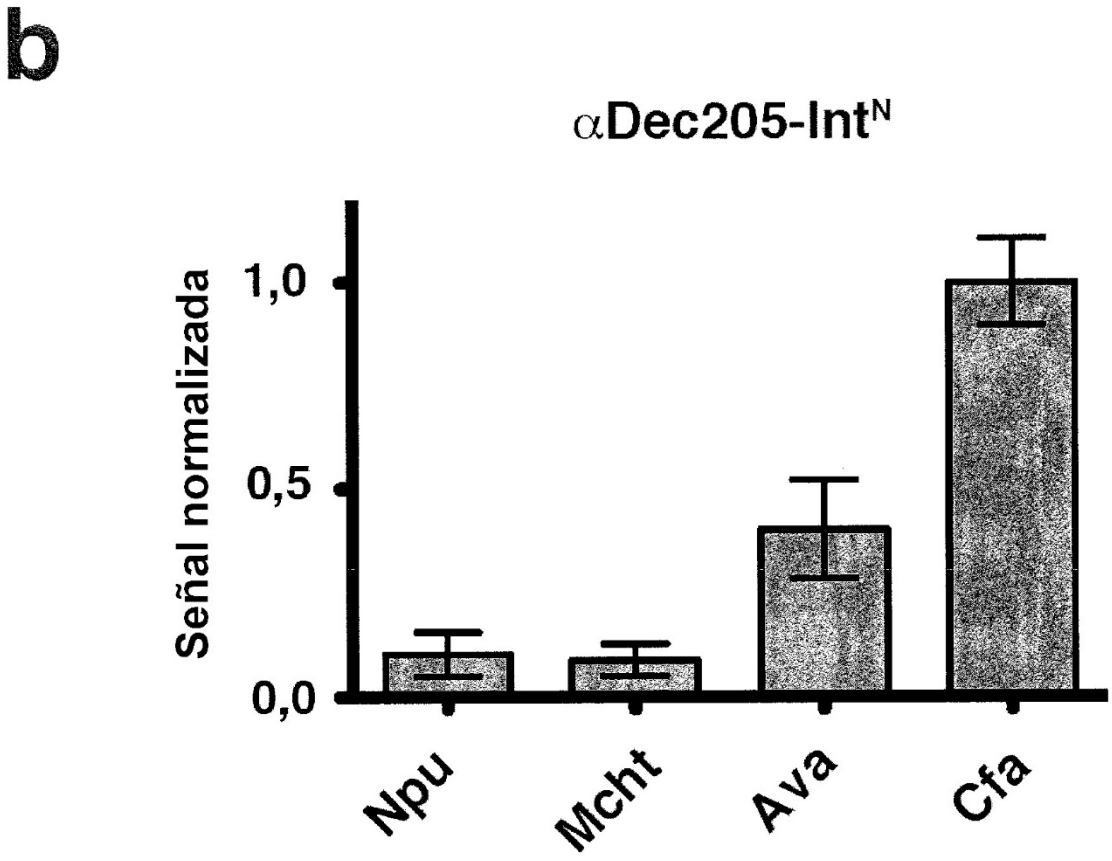

b
αDec205-Int^N
Señal normalizada
1,0
0,5
0,0
Npu
Mcht
Ava
Cfa

FIG. 3

C

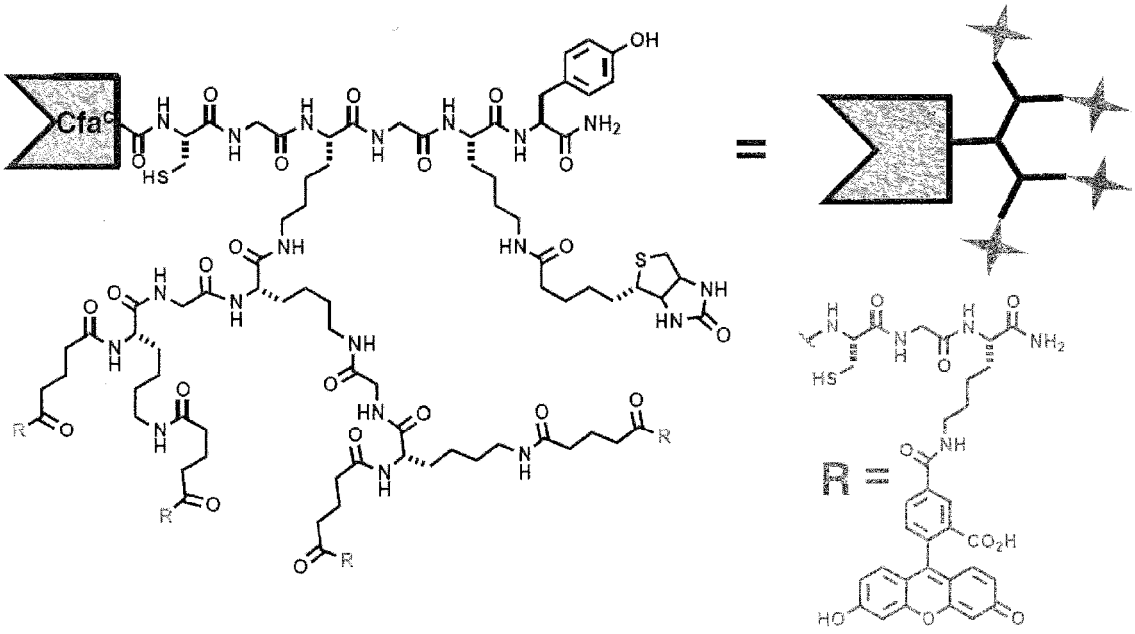
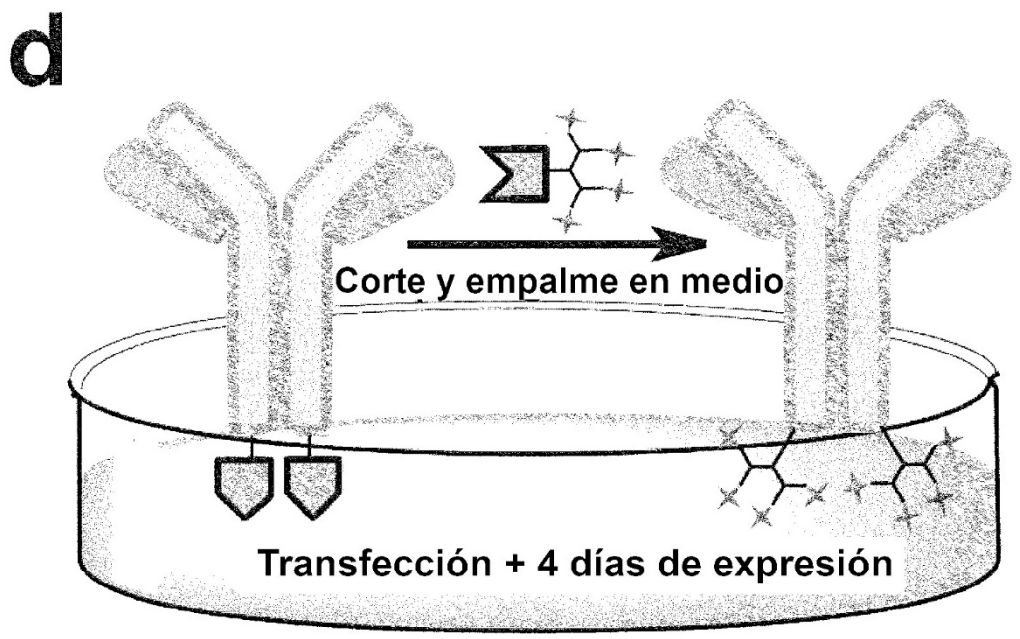



FIG. 3


d
Corte y empalme en medio
Transfección + 4 días de expresión

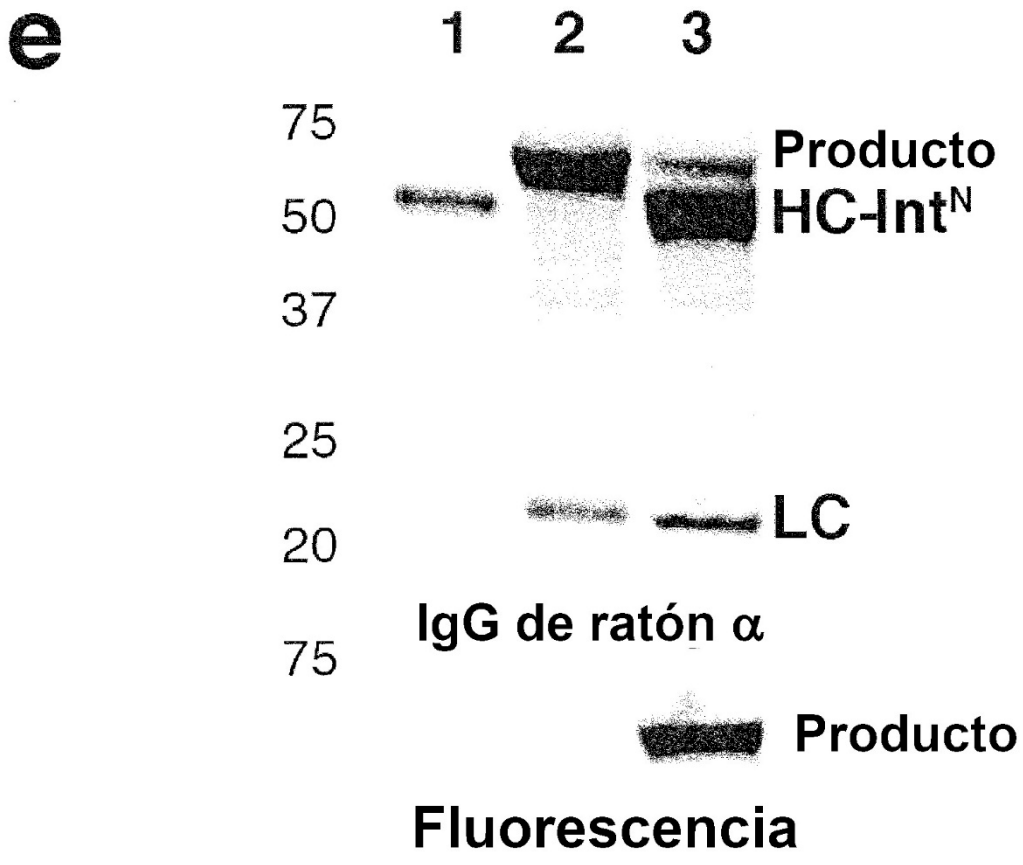

e
1
2
3
75
50
37
25
20
75
Producto
HC-Int^N
LC
IgG de ratón α
Producto
Fluorescencia

FIG. 3

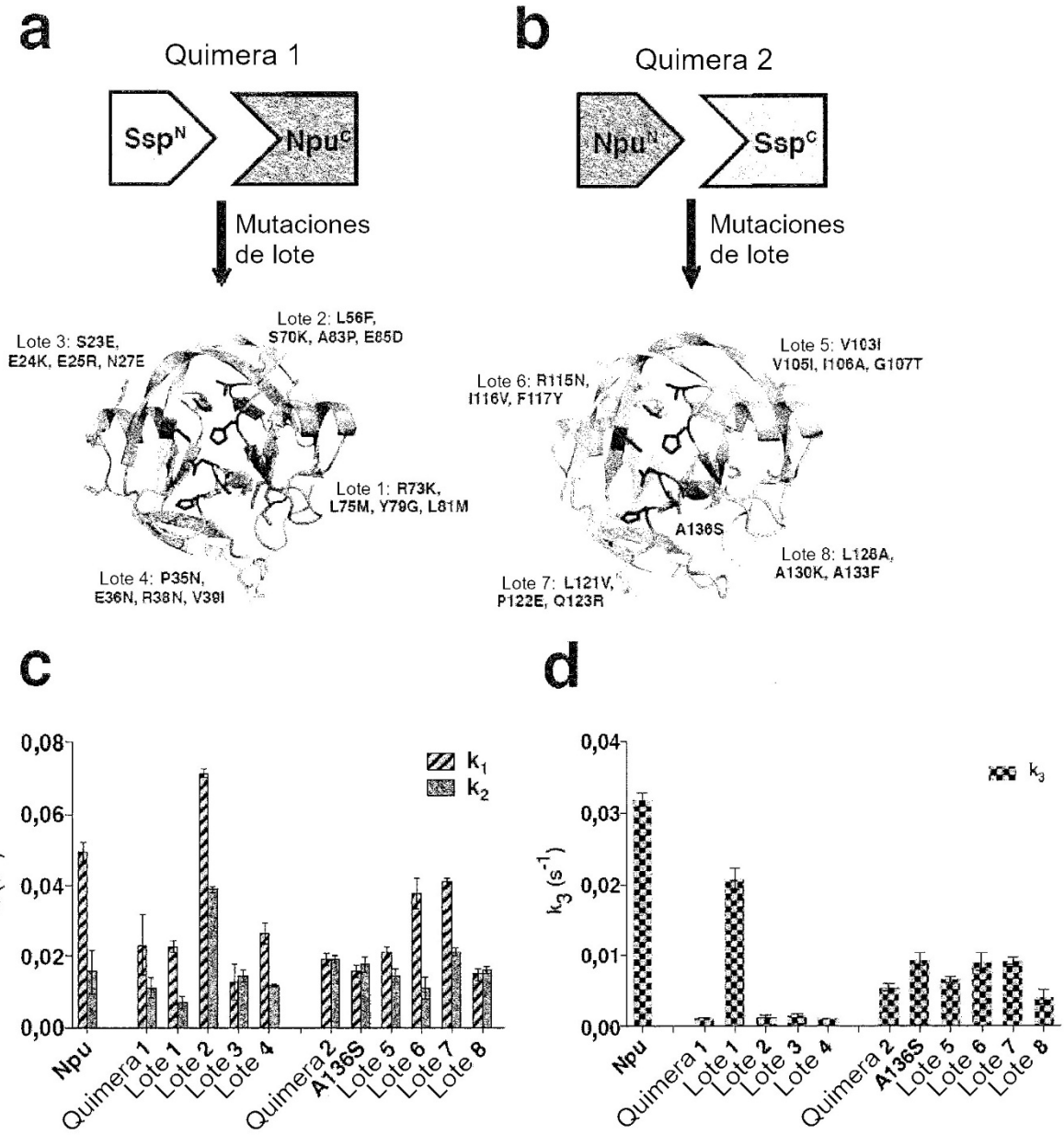

a
Quimera 1
Ssp^N
Npu^C
Mutaciones de lote
Lote 2: L56F, S70K, A83P, E85D
Lote 3: S23E, E24K, E25R, N27E
Lote 1: R73K, L75M, Y79G, L81M
Lote 4: P35N, E36N, R38N, V39I
b
Quimera 2
Npu^N
Ssp^C
Mutaciones de lote
Lote 5: V103I V105I, I106A, G107T
Lote 6: R115N, I116V, F117Y
A136S
Lote 8: L128A, A130K, A133F
Lote 7: L121V, P122E, Q123R
c
k (s^-1)
0,08
0,06
0,04
0,02
0,00
k1
k2
Npu
Quimera 1
Lote 1
Lote 2
Lote 3
Lote 4
Quimera 2
A136S
Lote 5
Lote 6
Lote 7
Lote 8
d
k3 (s^-1)
0,04
0,03
0,02
0,01
0,00
k3
Npu
Quimera 1
Lote 1
Lote 2
Lote 3
Lote 4
Quimera 2
A136S
Lote 5
Lote 6
Lote 7
Lote 8

FIG. 4

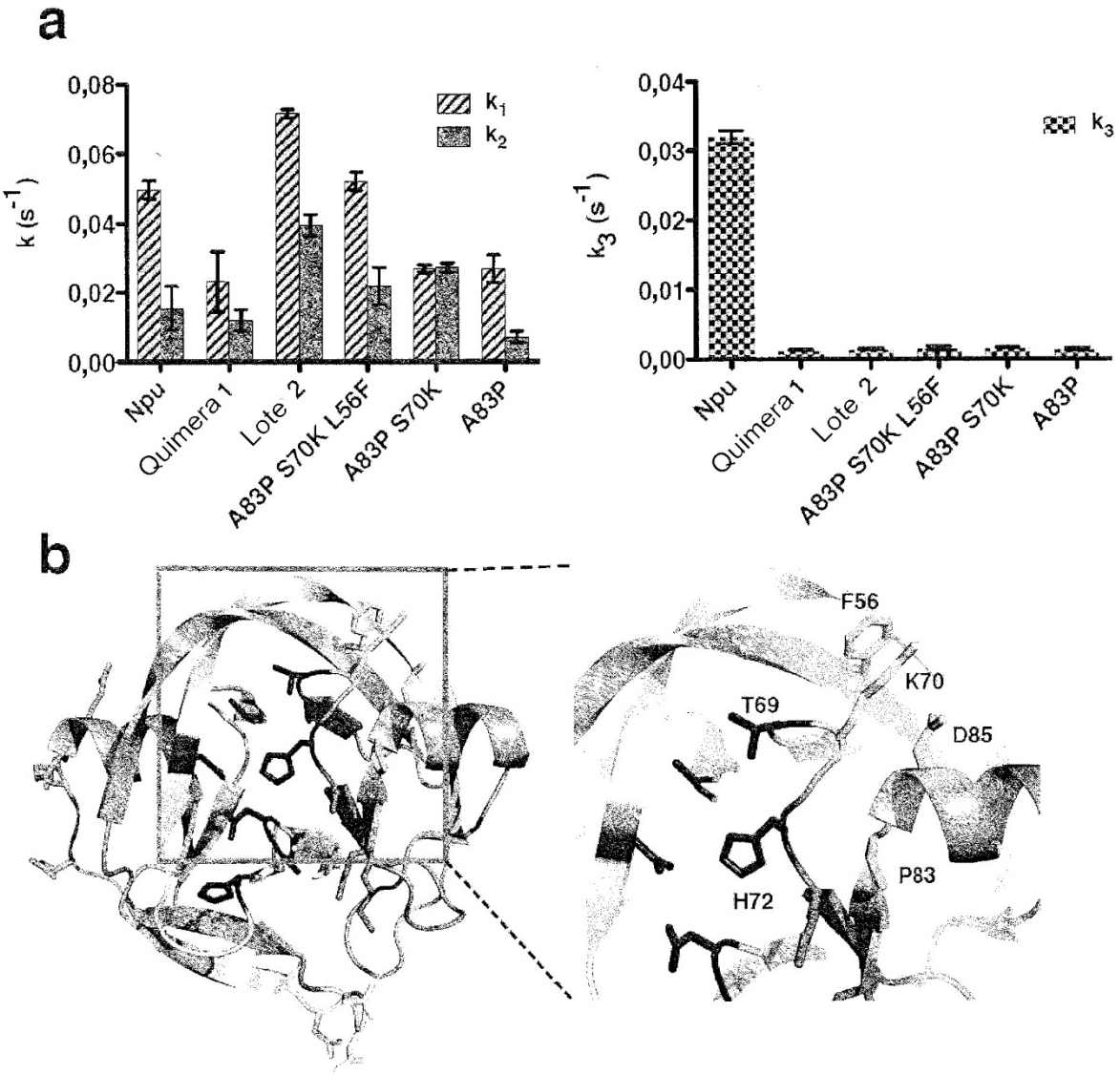

a
0,08
0,06
0,04
0,02
0,00
k (s-1)
k1
k2
Npu
Quimera 1
Lote 2
A83P S70K L56F
A83P S70K
A83P
0,04
0,03
0,02
0,01
0,00
k3 (s-1)
k3
Npu
Quimera 1
Lote 2
A83P S70K L56F
A83P S70K
A83P
b
F56
K70
T69
D85
P83
H72

FIG. 5

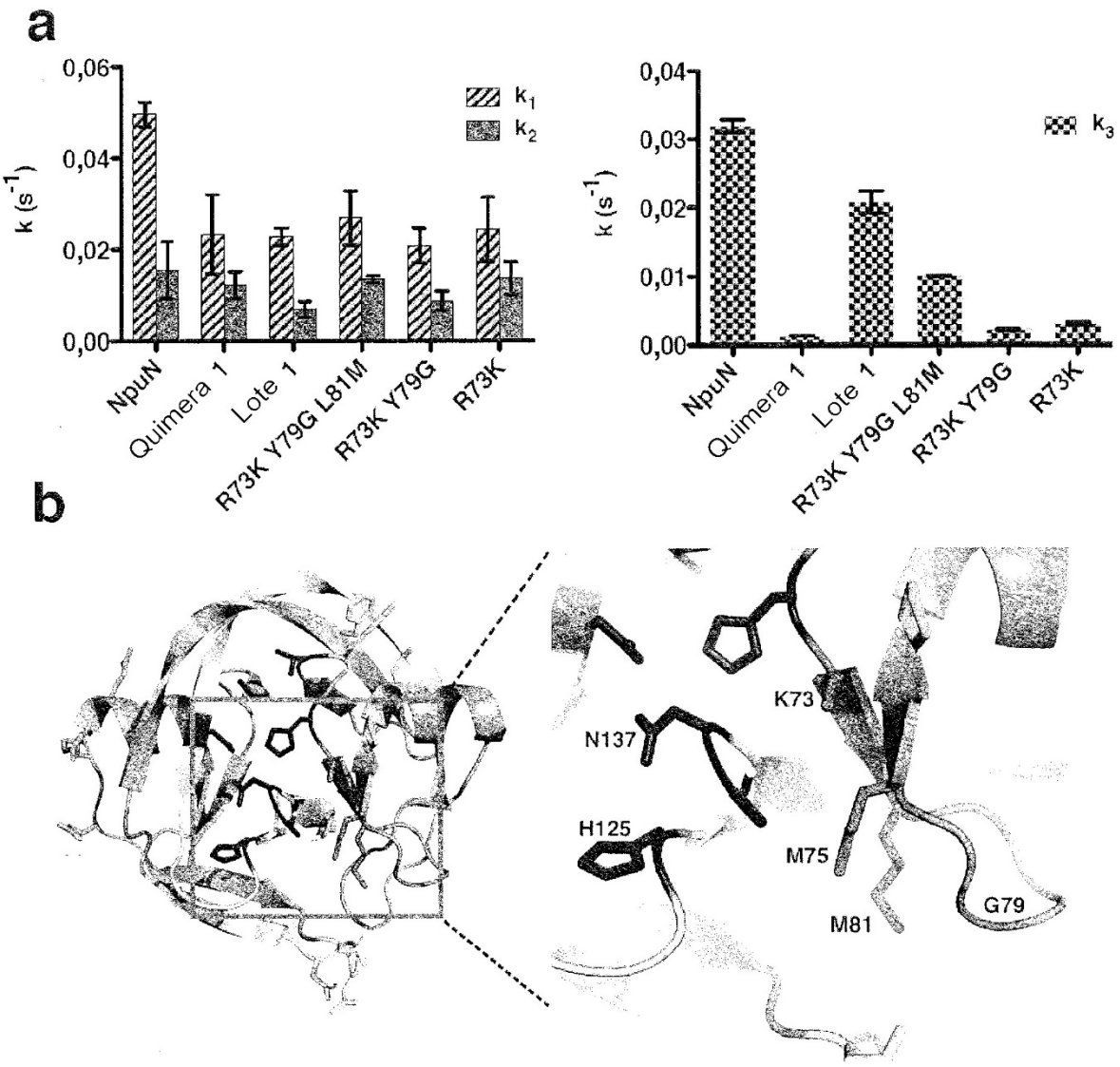

a
0,06
0,04
0,02
0,00
k (s-1)
k1
k2
NpuN
Quimera 1
Lote 1
R73K Y79G L81M
R73K Y79G
R73K
0,04
0,03
0,02
0,01
0,00
k (s-1)
k3
NpuN
Quimera 1
Lote 1
R73K Y79G L81M
R73K Y79G
R73K
b
K73
N137
H125
M75
G79
M81

FIG. 6

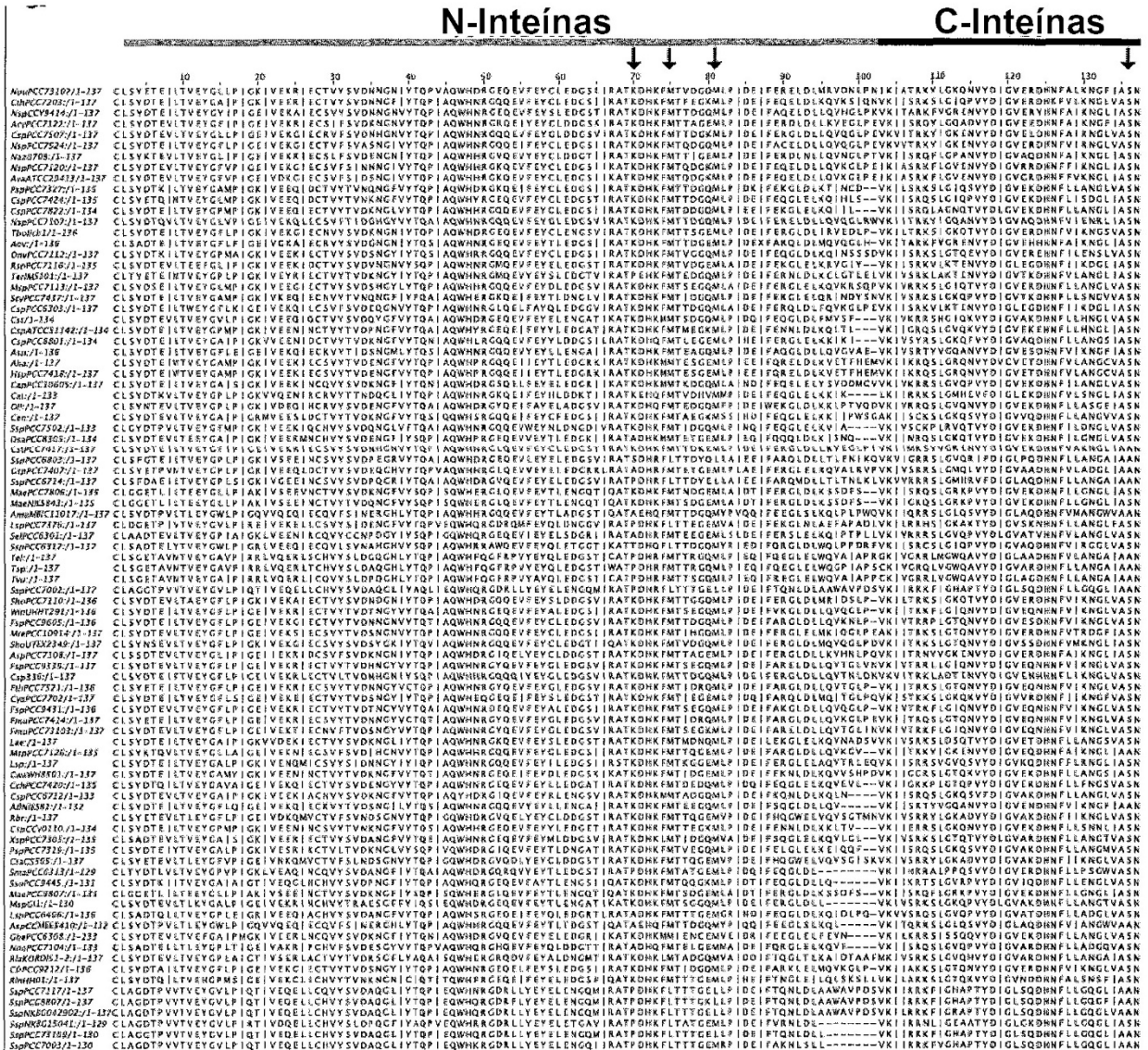



FIG. 7A

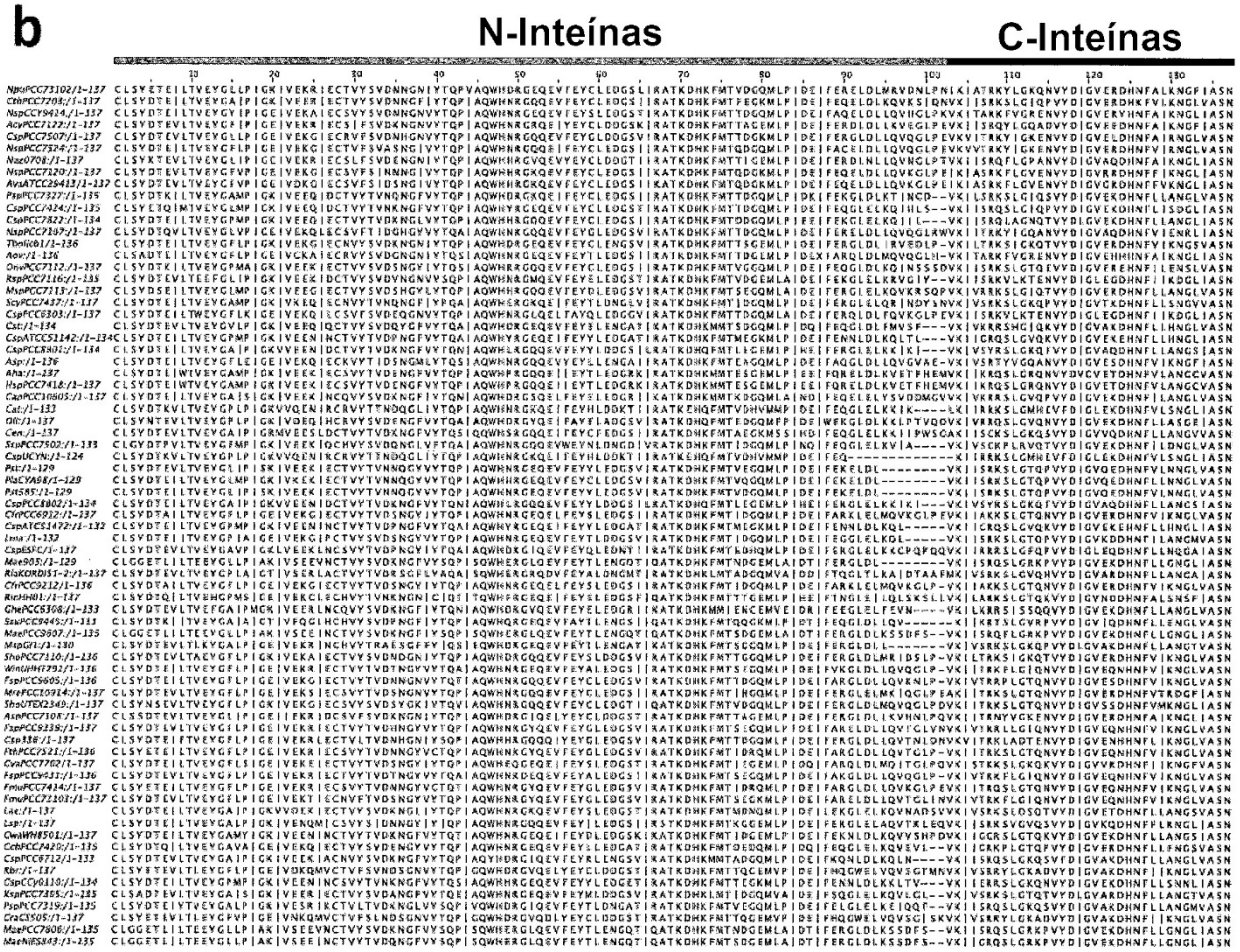



FIG. 7B


SUMO-Npu$^N$
SUMO-Cfa$^N$
P FT W1 W2 E1 E2 E3 E4
P FT W1 W2 E1 E2 E3 E4
75
50
37
25
20
15
Sumo-Int$^N$

FIG. 8

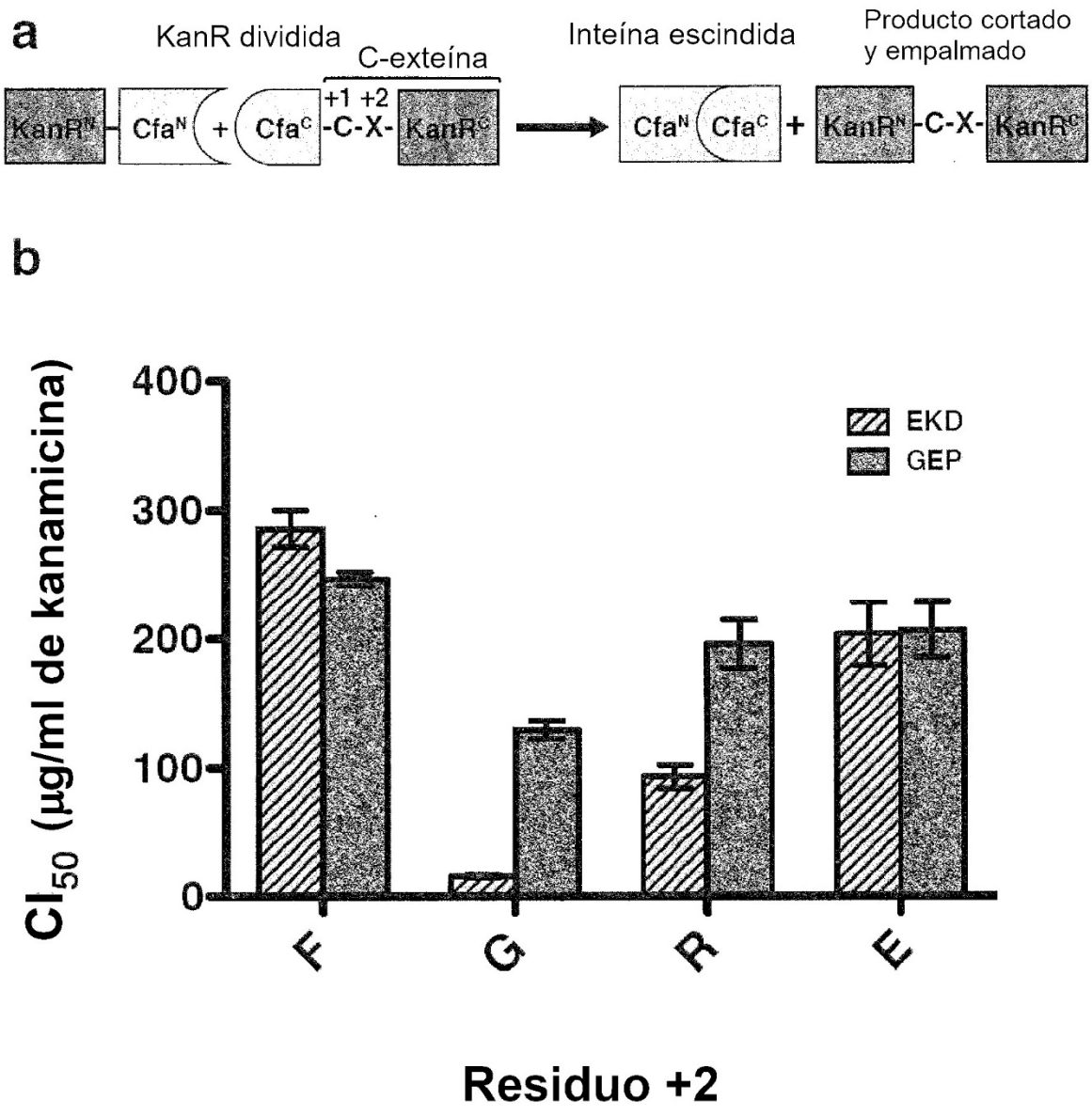

a
KanR dividida
C-exteína
Inteína escindida
Producto cortado y empalmado
KanR^N
Cfa^N
+
Cfa^C
+1 +2
-C-X-
KanR^C
Cfa^N Cfa^C
+
KanR^N
-C-X-
KanR^C
b
CI50 (µg/ml de kanamicina)
400
300
200
100
0
EKD
GEP
F
G
R
E
Residuo +2

FIG. 9

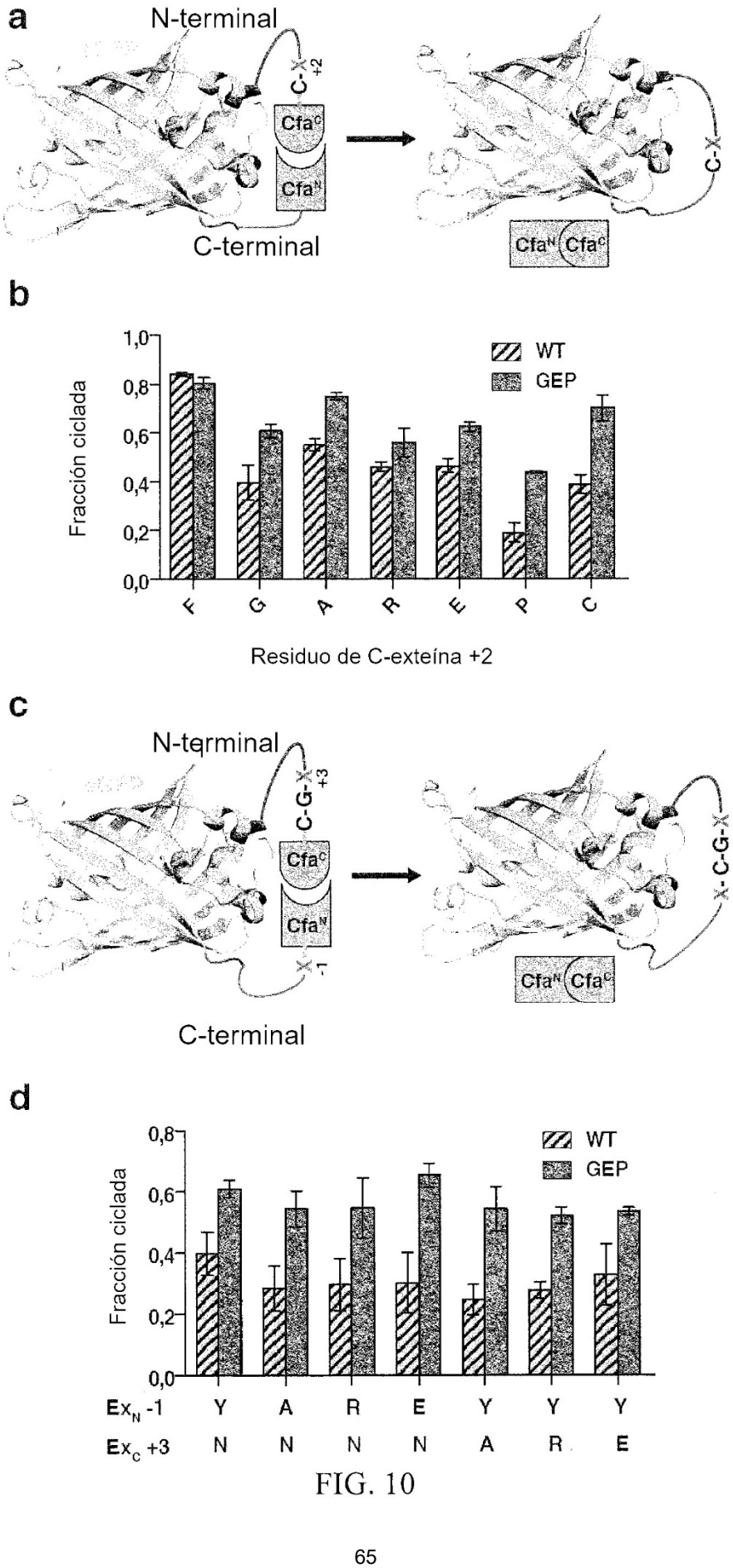

a
N-terminal
C-terminal
Cfa^C
Cfa^N
C-X +2
C-X
Cfa^N Cfa^C
b
Fracción ciclada
1,0
0,8
0,6
0,4
0,2
0,0
WT
GEP
F
G
A
R
E
P
C
Residuo de C-exteína +2
c
N-terminal
C-terminal
C-G-X +3
X -1
Cfa^C
Cfa^N
X-C-G-X
Cfa^N Cfa^C
d
Fracción ciclada
0,8
0,6
0,4
0,2
0,0
WT
GEP
Ex_N -1 Y A R E Y Y Y
Ex_C +3 N N N N A R E

FIG. 10

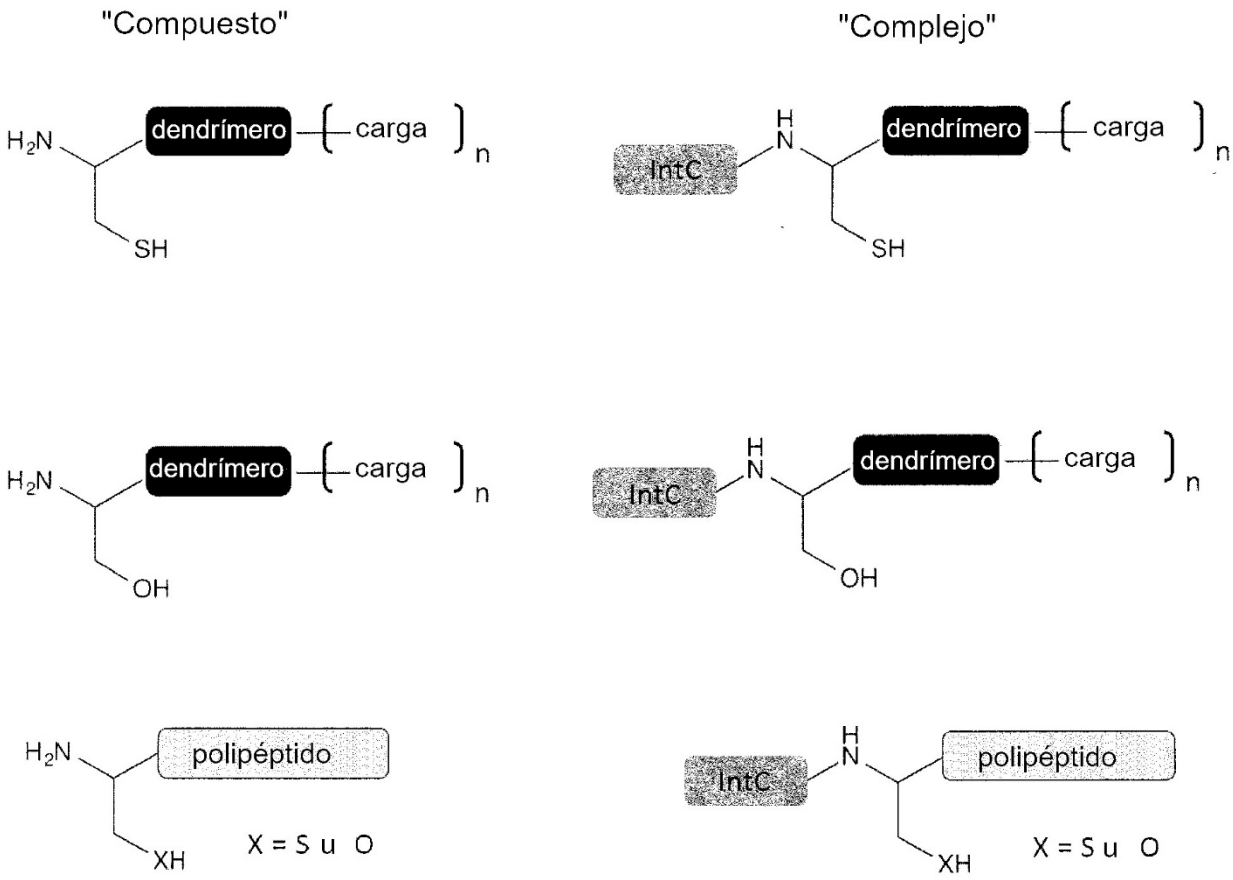

"Compuesto"
"Complejo"
H2N
dendrímero
carga
n
SH
IntC
H
N
OH
polipéptido
XH
X = S u O

FIG. 11