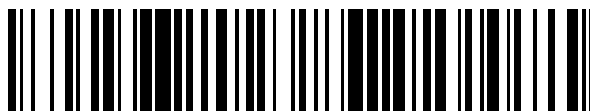


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 661 095**

51 Int. Cl.:

**C07K 14/00** (2006.01)

**C07C 11/00** (2006.01)

**C07C 11/22** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.02.2012 PCT/EP2012/051885**

87 Fecha y número de publicación internacional: **09.08.2012 WO12104422**

96 Fecha de presentación y número de la solicitud europea: **03.02.2012 E 12702538 (5)**

97 Fecha y número de publicación de la concesión europea: **20.12.2017 EP 2670767**

54 Título: **Aminoácidos no naturales que comprenden un grupo análogo de ciclooctinilo o trans-ciclooctenilo y usos de los mismos**

30 Prioridad:

**03.02.2011 US 201161462477 P**

**16.03.2011 US 201161453358 P**

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

**27.03.2018**

73 Titular/es:

**EUROPEAN MOLECULAR BIOLOGY  
LABORATORY (100.0%)  
Meyerhofstrasse 1  
69117 Heidelberg, DE**

72 Inventor/es:

**LEMKE, EDWARD;  
SCHULTZ, CARSTEN;  
PLASS, TILMAN;  
MILLES, SIGRID y  
KOEHLER, CHRISTINE**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**Observaciones:**

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

**ES 2 661 095 T3**

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

## DESCRIPCIÓN

Aminoácidos no naturales que comprenden un grupo análogo de ciclooctinilo o trans-ciclooctenilo y usos de los mismos

5

**Campo de la invención**

La invención se refiere a aminoácidos no naturales (AANN) que comprenden un grupo análogo de ciclooctinilo o *trans*-ciclooctenilo, y su uso y procesos para la preparación de polipéptidos que comprende un o más de un grupo análogo de ciclooctinilo o *trans*-ciclooctenilo. Estos polipéptidos pueden modificarse covalentemente mediante reacción *in vitro* o *in vivo* con compuestos que comprenden un grupo azida, óxido de nitrilo, nitrona, diazocarbonilo o 1,2,4,5-tetrazina.

10

**Antecedentes de la invención**

15

La capacidad de visualizar biomoléculas dentro de muestras de ensayo vivas mediante marcadores de fluorescencia modificados técnicamente se ha convertido en una herramienta importante en la biotecnología, la biología celular y la ciencia biosanitaria modernas. La codificación de proteínas de fusión con proteínas autofluorescentes comparativamente grandes es en la actualidad la técnica más ampliamente aplicada. Como los colorantes sintéticos ofrecen normalmente mejores propiedades fotofísicas que las proteínas autofluorescentes, se han desarrollado estrategias alternativas basándose en la codificación genética de marcadores únicos tales como marcadores Halo y SNAP, que ofrecen alta especificidad pero aún tienen un tamaño bastante grande. Pueden usarse marcadores pequeños como motivos de multihistidina o multicisteína para reconocer fluoróforos más pequeños, pero dentro del ambiente celular adolecen frecuentemente de problemas de especificidad ya que su elemento de reconocimiento básico se construye a partir de cadenas laterales de aminoácidos nativos. Dichas desventajas pueden superarse utilizando química bioortogonal que se basa en la unión de restos no naturales en condiciones fisiológicas suaves.

20

25

Una química potente que actúa eficazmente a temperaturas fisiológicas y en ambientes biológicos abundantemente funcionalizados es la cicloadición (3+2) de tipo Huisgen catalizada por cobre (I) entre azidas y alquinos lineales o la reacción de cicloadición (4+2) de Diels-Alder de demanda de electrones inversa entre un dienófilo restringido tal como *trans*-cicloocteno o norborneno y una 1,2,4,5-tetrazina, ambas formas de química clic (Kolb et al., *Angew Chem Int Ed Engl* 2001, 40:2004; Devaraj et al., *Angew Chem Int Ed Engl* 2009, 48:7013). Sin embargo, la cicloadición (3+2) más establecida requiere un catalizador de cobre que es tóxico para bacterias y células de mamífero, que reduce en gran medida la biocompatibilidad de este tipo de química clic. Esta limitación ha sido superada por Bertozzi y colaboradores, que han mostrado que la reacción clic actúa fácilmente sin necesidad de un catalizador citotóxico cuando se utilizan alquinos con tensión anular como sustrato (Agard et al., *J Am Chem Soc* 2004, 126:15046). La química clic sin cobre ha encontrado cada vez más aplicaciones en el marcaje de biomoléculas. Se usaron colorantes fluorescentes que comprenden grupos ciclooctinilo para marcar carbohidratos y proteínas que comprenden restos de azida unidos enzimáticamente *in vivo* (Chang et al., *Proc Natl Acad Sci USA* 2010, 107:1821) y el marcaje de ácido fosfatídico modificado con alquino/cicloalquino con fluoróforos de azido se describe en Neef y Schultz, *Angew Chem Int Ed Engl* 2009, 48:1498. La cicloadición (4+2) de Diels-Alder para marcar moléculas *in vivo* requiere la reacción de un grupo dienófilo con tensión tal como un grupo *trans*-ciclooctenilo o grupo norbornenilo con una 1,2,4,5-tetrazina fusionado con una sonda de molécula pequeña, por ejemplo un fluoróforo (Devaraj et al., *Bioconjugate Chem* 2008, 19:2297; Devaraj et al., *Angew Chem Int Ed Engl* 2009, 48:7013; Devaraj et al., *Angew Chem Int Ed Engl* 2010, 49:2869). No se requiere catalizador. Se describen compuestos que comprenden grupos ciclooctinilo o grupos *trans*-ciclooctenilo y su uso en química clic en Tiefenbrunn et al. (*Biopolymers* 2009, 94(1):95-106), DeForest et al. (*Nature Materials* 2009, 8:659-664), documento WO 2007/039858, Palomo (*Eur J Org Chem* 2010, 33:6303-6314) y documento WO 2010/119389.

30

35

40

45

La modificación traduccional de proteínas por codificación genética directa de aminoácidos no naturales fluorescentes usando un par de aminoacil ARNt/sintetasa ortogonal ofrece especificidad exquisita, libertad de colocación dentro de la proteína diana y, si lo hay, un cambio de estructura mínimo. Este enfoque fue aplicado por primera vez con éxito por Summerer et al. (*Proc Natl Acad Sci USA* 2006, 103:9785), que desarrolló un par de leucil ARNt/sintetasa de *Escherichia coli* para codificar genéticamente el AANN dansilalanina en *Saccharomyces cerevisiae*. En respuesta al codón de terminación ámbar TAG, la maquinaria de traducción del hospedador incorporó inmediatamente dansilalanina. Este enfoque se ha usado mientras tanto para codificar genéticamente varios colorantes pequeños y otros restos de interés. Por ejemplo, se han usado pares de tirosil ARNt<sup>tyr</sup>/sintetasa de *Methanococcus jannaschii* modificada técnicamente, leucil ARNt<sup>leu</sup>/sintetasa de *E. coli* así como pirrolisina ARNt<sup>pyl</sup>/sintetasa de *Methanosarcina maize* y *M. barkeri* para codificar genéticamente restos de azida en polipéptidos (Chin et al., *J Am Chem Soc* 2002, 124:9026; Chin et al., *Science* 2003, 301:964; Nguyen et al., *J Am Chem Soc* 2009, 131:8720, Yanagisawa et al., *Chem Biol* 2008, 15:1187). Sin embargo, debido a la necesidad de desarrollar nuevos pares de aminoacil ARNt/sintetasa y limitaciones potenciales del tamaño impuestas por la maquinaria de traducción, aún no se han codificado colorantes mayores con propiedades fotofísicas potenciadas y otros restos voluminosos. Se encuentra una revisión sobre pares de aminoacil-ARNt sintetasa/ARNt ortogonales y aminoácidos no naturales codificados de este modo en Liu et al. (*Annual Review of Biochemistry* 2010, 79:413-444).

50

55

60

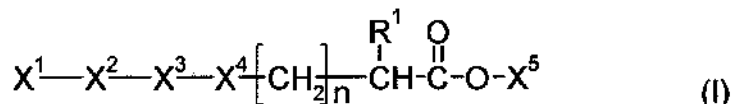
65

A pesar de grandes esfuerzos, sigue existiendo una alta demanda de estrategias que faciliten el marcaje específico de sitio de proteínas *in vitro* e *in vivo*. Por lo tanto, fue un objetivo de la presente invención proporcionar aminoácidos o análogos de los mismos que puedan incorporarse en la traducción en cadenas polipeptídicas y permitir el marcaje del polipéptido resultante *in vitro* así como *in vivo*.

5

**Sumario de la invención**

La presente invención se refiere a compuestos de fórmula I

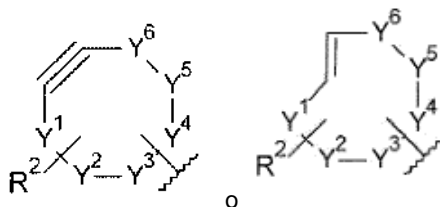


10

en la que:

$X^1$  tiene la fórmula

15



en la que:

20

$Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$  con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ ;

25

$R^2$  es hidrógeno, halógeno, alquilo  $C_1-C_4$ ,  $CF_3$ , CN, alcoxi  $C_1-C_4$ ,  $-O-CF_3$ , alcanoiloxi  $C_2-C_5$ , alquilaminocarboniloxi  $C_1-C_4$  o alquiltio  $C_1-C_4$ ;

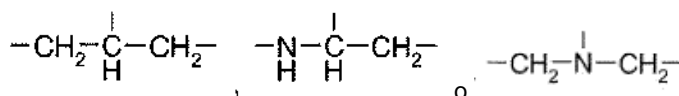
$X^2$  es  $-CH_2-$ ,  $-O-$ ,  $-S-$ ,  $-NH-$ ,  $-C(O)-$ ,  $-OC(O)-$ ,  $-C(O)O-$ ,  $-NH-C(O)-$  o  $-C(O)-NH-$ , o

$X^2$  es  $>CH-$  o  $>N-$  en el que el carbono o el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forma un anillo de 3 miembros, o

30

$X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 4 miembros, o

$X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



35

en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 5 miembros;

$X^3$  es alquileo  $C_1-C_6$ ,  $-(CH_2-CH_2-O)_m-$ ,  $-(CH_2-O)_p-$  o un enlace sencillo;

40

$X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-C(O)-$ ,  $-NH-CH(NH_2)-$ ,  $-CH(NH_2)-NH-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$ ,  $-C(O)-NH-C(NH)-NH-$ ,  $NH-CH(NH_2)-C(O)-$  o  $-NH-C(NH)-NH-C(O)-$ ;

$X^5$  es hidrógeno, alquilo  $C_1-C_6$ , alcoxi  $C_1-C_6$ -alquilo  $C_1-C_2$ , alcanoiloxi  $C_2-C_7$ -alquilo  $C_1-C_2$  o alcanoilulfanil  $C_2-C_7$ -alquilo  $C_1-C_2$ ;

$R^1$  es  $-OH$  o  $-NH_2$ ;

45

$n$  es un número entero de 1 a 4;

$m$  es un número entero de 1 a 6; y

$p$  es un número entero de 1 a 6,

o una sal de adición de ácidos o bases del mismo.

50

Los compuestos o sales de la invención pueden incorporarse en la traducción en un polipéptido que está codificado por un polinucleótido que comprende uno o varios codones selectores.

La presente invención también se refiere por lo tanto a un proceso para preparar un polipéptido diana que tiene uno o varios grupos análogos de ciclooctinilo o *trans*-ciclooctenilo, comprendiendo el proceso:

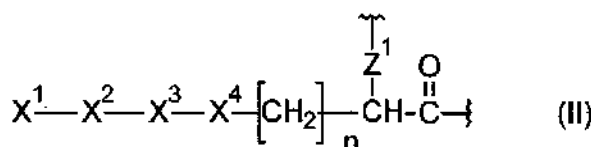
a) proporcionar un sistema de traducción que comprende:

- (i) una aminoacil ARNt sintetasa o un polinucleótido que la codifica;  
 (ii) un compuesto o una sal de la invención;  
 (iii) un ARNt que tiene un anticodón para un codón selector o un polinucleótido que codifica dicho ARNt; y  
 (iv) un polinucleótido que codifica el polipéptido diana y que comprende uno o varios codones selectores,

en el que la aminoacil ARNt sintetasa (i) es capaz de acilar específicamente el ARNt (iii) con el compuesto o sal (ii);

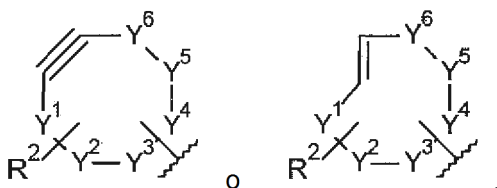
- b) permitir la traducción del polinucleótido (iv) incorporando de este modo el compuesto (ii) al polipéptido diana en la posición o las posiciones codificadas por el codón o los codones selectores; y  
 c) opcionalmente recuperar el polipéptido resultante.

Por lo tanto, también se describe en el presente documento un polipéptido que comprende uno o más de un residuo de fórmula II



en la que:

$X^1$  tiene la fórmula



en la que:

$Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$

son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$  con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ ;

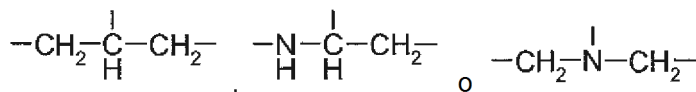
$R^2$  es hidrógeno, halógeno, alquilo  $C_1-C_4$ ,  $CF_3$ , CN, alcoxi  $C_1-C_4$ ,  $-O-CF_3$ , alcanoiloxi  $C_2-C_5$ , alquilaminocarboniloxi  $C_1-C_4$  o alquiltio  $C_1-C_4$ ;

$X^2$  es  $-CH_2-$ ,  $-O-$ ,  $-S-$ ,  $-NH-$ ,  $-C(O)-$ ,  $-OC(O)-$ ,  $-C(O)O-$ ,  $-NH-C(O)-$  o  $-C(O)-NH-$ , o

$X^2$  es  $>CH-$  o  $>N-$  en el que el carbono o el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forma un anillo de 3 miembros, o

$X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 4 miembros, o

$X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 5 miembros;

$X^3$  es alquileo  $C_1-C_6$ ,  $-(CH_2-CH_2-O)_m-$ ,  $-(CH_2-O)_p-$  o un enlace sencillo;

$X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-C(O)-$ ,  $-NH-CH(NH_2)-$ ,  $-CH(NH_2)-NH-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$ ,  $-C(O)-NH-C(NH)-NH-$ ,  $NH-CH(NH_2)-C(O)-$  o  $-NH-C(NH)-NH-C(O)-$ ;

$Z^1$  es  $-O-$  o  $-NH-$ ;

$n$  es un número entero de 1 a 4;

$m$  es un número entero de 1 a 6; y

$p$  es un número entero de 1 a 6.

Se describen además kits para preparar un polipéptido que tiene uno o varios grupos análogos de ciclooctinilo o *trans*-ciclooctenilo (polipéptido diana). Los kits comprenden un compuesto o una sal de la invención y opcionalmente uno o más medios para preparar el polipéptido, tales como uno o varios polinucleótidos que codifican una aminoacil ARNt sintetasa, un ARNt como se describe en el presente documento; un polinucleótido que codifica una proteína indicadora; y/o medios adicionales para traducción de un polinucleótido que codifica dicho polipéptido diana.

Como los grupos *trans*-ciclooctenilo se sabe que los grupos norbornenilo reaccionan con 1,2,4,5-tetrazinas en cicloadiciones de Diels-Alder de demanda de electrones inversa (Devaraj et al., Bioconjugate Chem 2008, 19:2297). Por lo tanto, también se describen en el presente documento aminoácidos no naturales que comprenden un grupo norbornenilo, su uso, kits y procesos para la preparación de polipéptidos que comprenden uno o varios grupos norbornenilo, en los que lo que se desvela en el presente documento con respecto a *trans*-ciclooctenilo se aplica de manera análoga a norbornenilo.

### Breve descripción de los dibujos

La Figura 1 muestra imágenes fluorescentes de cultivos de *E. coli* que expresan GFP<sup>TAG</sup> (a) y bandas de GFP<sup>tag</sup> en un gel con tinción de Coomassie de un SDS-PAGE de una proteína marcada con His purificada de cultivos de expresión de proteínas de *E. coli* (b). Se expresó GFP<sup>TAG</sup> en presencia de pares de pirrolisil ARNt/pirrolisil ARNt sintetasa de tipo silvestre (tRNA<sup>pyl</sup>/RS<sup>WT</sup>) o mutante (tRNA<sup>pyl</sup>/RS<sup>AF</sup>) en medio complementado con NaOH ("-" ), compuesto **1** ("+1") o compuesto **2** ("+2"). Los tamaños de proteínas marcadoras se proporcionan en kDa.

La Figura 2 muestra el gel con tinción de Coomassie de la Figura 1b a tamaño completo.

La Figura 3 muestra el análisis de lisado de células completas de cultivos de *E. coli* descrito en el ejemplo B. Después de añadir azido cumarina fluorogénica (compuesto **3**) a cultivos que expresaban GFP<sup>tag->1</sup> y GFP<sup>WT</sup>, se tomaron muestras pequeñas en los puntos temporales indicados y se sometieron a SDS-PAGE. Después de electroforesis, se tomó una imagen fluorescente del gel (a) a una longitud de onda de excitación de 365 nm detectando la emisión con un ajuste de filtro de bromuro de etidio. Las proteínas separadas en el gel de poliácridamida SDS se visualizaron por tinción de Coomassie (b). La altura de procesamiento de GFP se indica con una flecha. Los tamaños de proteínas marcadoras se proporcionan en kDa.

La Figura 4 muestra un histograma bidimensional (S frente a E<sub>fret</sub>) de datos de smFRET de moléculas de GFP<sup>TAG->1,Atto647N</sup> individuales de difusión libre basados en los datos sin procesar mostrados en la Figura 5.

La Figura 5 muestra la traza de datos sin procesar (agrupados a un tiempo de resolución de 1 ms) de GFP<sup>TAG->1,Atto647N</sup> individuales de difusión libre correspondiente a los datos mostrados en la Figura 4. Las ráfagas fluorescentes detectadas en el canal verde (donante, D) surgen del cromóforo GFP directamente excitado (y poca o ninguna transferencia de energía a A) y las ráfagas en el canal rojo (aceptor, A) se originan de transferencia de energía por resonancia al colorante Atto647N.

La Figura 6 muestra imágenes fluorescentes de suspensiones de *E. coli* en tubos de microcentrífuga que expresan QPP<sup>TAG</sup> en ausencia (" - ") y presencia (" + ") de UAA **1**, **13**, **16** o **17**, respectivamente, con el gel de poliácridamida SDS teñido con Coomassie correspondiente después de purificación del GFP<sup>TAG</sup> obtenido en *E. coli* cotransfectada con plásmidos que codifican RS<sup>WT</sup> (a) o RS<sup>AF</sup> (b).

La Figura 7 muestra análisis de proteínas producidas por *E. coli* como se describe en el ejemplo D. Los cultivos que expresaban GFP<sup>tag->1</sup> (+1), GFP<sup>tag->13</sup> (+13), GFP<sup>tag->16</sup> (+16), GFP<sup>tag->17</sup> (+17) o GFP<sup>TAG</sup> con propargilisina (control negativo) se trataron con TAMRA-azida (Az) o TAMRA-tetrazina (Tet) y se sometieron a SDS-PAGE. Después de electroforesis, se tomó una imagen fluorescente del gel (a). Las proteínas separadas en el gel de poliácridamida SDS se visualizaron por tinción de Coomassie (b).

La Figura 8 ilustra FRET creciente de GFP a TAMRA observado con GFP<sup>tag->13</sup> expresada en *E. coli* durante el marcaje con TAMRA-tetrazina. Los espectros de fluorescencia (excitación a 450 nm, emisión a 470-650 nm) se registraron a lo largo del tiempo (gráfico coloreado de oscuro a claro).

### Descripción detallada de la invención

Los compuestos y sales de la invención son aminoácidos no naturales que pueden incorporarse translacionalmente en cadenas de polipéptidos.

La expresión "aminoácido no natural" se refiere a un aminoácido que no es uno de los 20 aminoácidos canónicos o selenocisteína. La expresión también se refiere a análogos de aminoácido, por ejemplo, en los que el grupo α-amino está reemplazado por un grupo hidroxilo.

Los compuestos o sales de la invención poseen centro de asimetría y pueden existir en disposiciones espaciales diferentes o como diferentes tautómeros. Para la preparación de polipéptidos con grupos análogos de ciclooctinilo o

*trans*-ciclooctenilo, pueden usarse mezclas enantioméricas, en particular racematos, mezclas diastereoméricas y mezclas tautoméricas. Como alternativa, los enantiómeros esencialmente puros respectivos, diastereómeros y tautómeros de los compuestos o sales de la invención pueden usarse para tal propósito.

- 5 Los restos orgánicos mencionados en las definiciones anteriores de las variables son - como el término alquilo - términos colectivos para listados individuales de los miembros individuales del grupo. El sufijo  $C_{n-m}$  indica en cada caso el número posible de átomos de carbono en el grupo.

10 El término halógeno representa en cada caso un radical de flúor, bromo, cloro o yodo, en particular un radical de flúor.

15 Alquilo  $C_1-C_6$  es un grupo alquilo de cadena lineal o ramificado que tiene de 1 a 6, en particular de 1 a 4 o de 1 a 3 átomos de carbono. Los ejemplos incluyen metilo, alquilo  $C_2-C_4$  tal como etilo, n-propilo, *iso*-propilo, n-butilo, 2-butilo, *iso*-butilo o *terc*-butilo y también pentilo, 1-metilbutilo, 2-metilbutilo, 3-metilbutilo, 2,2-dimetilpropilo, 1-etilpropilo, hexilo, 1,1-dimetilpropilo, 1,2-dimetilpropilo, 1-metilpentilo, 2-metilpentilo, 3-metilpentilo, 4-metilpentilo, 1,1-dimetilbutilo, 1,2-dimetil-butilo, 1,3-dimetilbutilo, 2,2-dimetilbutilo, 2,3-dimetilbutilo, 3,3-dimetilbutilo, 1-etilbutilo, 2-etilbutilo, 1,1,2-trimetilpropilo, 1,2,2-trimetilpropilo, 1-etil-1-metilpropilo y 1-etil-2-metilpropilo.

20 Alquileno  $C_1-C_4$  es un grupo alquileno de cadena lineal o ramificado que tiene de 1 a 4 átomos de carbono. Los ejemplos incluyen metileno y 1,2-etileno. Un ejemplo adicional es 1,3-propileno.

Alquileno  $C_1-C_6$  es un grupo alquileno de cadena lineal o ramificado que tiene de 1 a 6 átomos de carbono. Los ejemplos incluyen metileno, etileno, 1,2-etileno, 1,3-propileno, isopropileno, 1-4-butileno y 1-5-pentileno.

25 Alcoxi  $C_1-C_6$  es un radical de la fórmula R-O-, en la que R es un grupo alquilo de cadena lineal o ramificado que tiene de 1 a 6, en particular de 1 a 4 o de 1 a 3 átomos de carbono como se define en el presente documento.

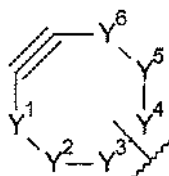
30 Alcanoiloxi  $C_2-C_7$  es un radical de la fórmula R-(CO)-O-, en la que R es un grupo alquilo de cadena lineal o ramificado que tiene de 1 a 6, en particular de 1 a 4 o de 1 a 3 átomos de carbono como se define en el presente documento.

35 Alquilaminocarboniloxi  $C_1-C_6$  es un radical de la fórmula R-NH-(CO)-O-, en la que R es un grupo alquilo de cadena lineal o ramificado que tiene de 1 a 6, en particular de 1 a 4 o de 1 a 3 átomos de carbono como se define en el presente documento.

Alquiltio  $C_1-C_4$  es un radical de la fórmula R-S-, en la que R es un radical alquilo que tiene de 1 a 4, preferiblemente de 1 a 3 átomos de carbono como se define en el presente documento.

40 Alcanoilsulfanilo  $C_2-C_7$  es un radical de la fórmula R-(CO)-S-, en la que R es un grupo alquilo de cadena lineal o ramificado que tiene de 1 a 6, en particular de 1 a 4 o de 1 a 3 átomos de carbono como se define en el presente documento.

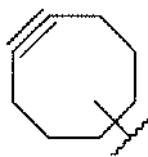
45 El término grupo análogo de ciclooctinilo representa un radical cicloalifático insaturado que tiene 8 átomos de carbono y un triple enlace en la estructura de anillo, en el que 1 o 2 átomos de carbono pueden estar reemplazados por un átomo de oxígeno, azufre y/o nitrógeno. En particular, el término grupo análogo de ciclooctinilo representa un resto de la fórmula:



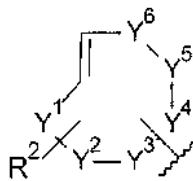
50 en la que  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$ , con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ .

El término grupo ciclooctinilo representa un grupo análogo de ciclooctinilo como se ha definido anteriormente, en el que todos los  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son  $-CH_2-$ . En particular, el término ciclooctinilo representa un resto de la fórmula:

55

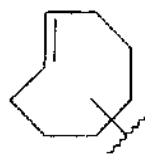


5 El término grupo análogo de *trans*-ciclooctenilo representa un radical cicloalifático insaturado que tiene 8 átomos de carbono y un doble enlace que está en configuración *trans* en la estructura de anillo, en el que 1 o 2 átomos de carbono pueden estar reemplazados por un átomo de oxígeno, azufre y/o nitrógeno. En particular, el término grupo análogo de *trans*-ciclooctenilo representa un resto de la fórmula:



10 en la que  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$ , con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ .

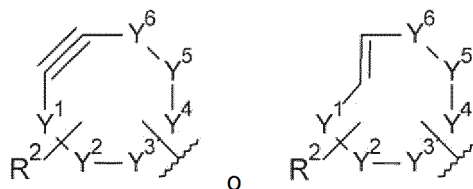
15 El término *trans*-ciclooctenilo representa un grupo análogo de *trans*-ciclooctenilo como se ha definido anteriormente, en el que todos los  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ . En particular, el término *trans*-ciclooctenilo representa un resto de la fórmula:



20 A menos que se indique lo contrario, el término "sustituido" significa que un radical está sustituido con 1, 2 o 3, especialmente 1 o 2, sustituyentes que se seleccionan en particular entre el grupo que consiste en halógeno, alquilo  $C_1-C_4$ ,  $CN$ ,  $CF_3$ ,  $-O-CF_3$ , alcoxi  $C_1-C_4$ , alcanoiloxi  $C_2-C_4$ , alquilaminocarboniloxi  $C_1-C_4$  y alquiltio  $C_1-C_4$ .

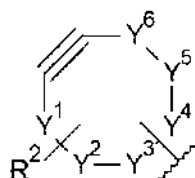
25 Con respecto a la capacidad de los compuestos de incorporarse translacionalmente en una cadena de polipéptido, las variables  $X^1, X^2, X^3, X^4, X^5, Y^1, Y^2, Y^3, Y^4, Y^5, Y^6, n, m, p, R^1$  y  $R^2$  tienen preferiblemente los siguientes significados que, cuando se toman solos o en combinación, representan realizaciones particulares de los aminoácidos no naturales de la fórmula I o cualquiera de las otras fórmulas desveladas en el presente documento.

$X^1$  tiene la fórmula:



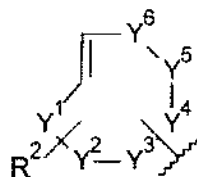
30 en la que  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$ , con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ ; y en la que  $R^2$  es como se define en el presente documento.

35 De acuerdo con una realización,  $X^1$  es un grupo análogo de ciclooctinilo de fórmula



en la que  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  y  $R^2$  son como se definen en el presente documento.

De acuerdo con otra realización,  $X^1$  es un grupo análogo de *trans*-ciclooctenilo de fórmula



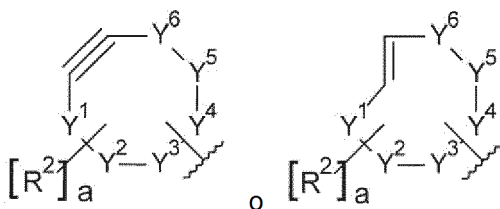
5

en la que  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  y  $R^2$  son como se definen en el presente documento.

10 El grupo  $X^1$  análogo de ciclooctinilo o *trans*-ciclooctenilo puede estar unido a  $X^2$  mediante un átomo del anillo en una posición  $\alpha, \beta$  o  $\gamma$  relativa al triple o doble enlace. En caso de que  $X^2$  junto con dos átomos en el anillo adyacentes de  $X^1$  forme un anillo de 3, 4 o 5 miembros, el grupo  $X^1$  puede estar unido a  $X^2$  mediante  $Y^1$  e  $Y^2, Y^2$  e  $Y^3, Y^3$  e  $Y^4, Y^4$  e  $Y^5, Y^5$  e  $Y^6$ . Se apreciará fácilmente que  $X^2$  es está enlazado mediante el núcleo de N a uno o dos de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  en  $X^1$  (por lo cual  $-CH_2-$  o  $-NH-$  se vuelven  $>CH-$  o  $>N-$ , respectivamente).

15 De acuerdo con una realización particular, el grupo análogo de ciclooctinilo o *trans*-ciclooctenilo está unido a  $X^2$  mediante el átomo del anillo en posición  $\alpha$  en relación al triple o doble enlace, es decir mediante  $Y^1$  o  $Y^6$ . En caso de que  $X^2$  junto con dos átomos adyacentes en el anillo de  $X^1$  forme un anillo de 3, 4 o 5 miembros, esta es una realización particular si el grupo análogo de ciclooctinilo o *trans*-ciclooctenilo está unido a  $X^2$  mediante  $Y^3$  e  $Y^4$ .

20 El grupo análogo de ciclooctinilo o *trans*-ciclooctenilo puede estar sin sustituir (es decir  $R^2$  es hidrógeno) o sustituido con uno o más de un radical  $R^2$ . Por lo tanto, puede haber uno o más de un sustituyente  $R^2$ . Más particularmente, puede haber hasta 5 sustituyentes  $R^2$ . Preferiblemente hay 1, 2 o 3 sustituyentes  $R^2$ . La fórmula (I) puede representarse por tanto de la siguiente manera:



25

en la que a es cero, 1, 2, 3, 4 o 5.

30 Si hay más de un radical  $R^2$ , estos pueden ser radicales iguales o diferentes y dos radicales  $R^2$  pueden estar enlazados al mismo átomo o a átomos diferentes. Por ejemplo,  $R^2$  puede ser dos átomos de flúor enlazados a un átomo del anillo de carbono.

35  $R^2$  es hidrógeno, halógeno, alquilo  $C_1-C_4$ ,  $CF_3$ , CN, alcoxi  $C_1-C_4$ ,  $-O-CF_3$ , alcanoiloxi  $C_2-C_5$ , alquilamino  $C_1-C_4$ -carboniloxi o alquiltio  $C_1-C_4$ .

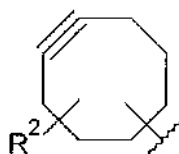
De acuerdo con una realización particular,  $R^2$  es halógeno, preferiblemente flúor.

40 De acuerdo con una realización particular adicional,  $X^1$  es un grupo análogo de ciclooctinilo y  $R^2$  es halógeno, preferiblemente flúor.

Cuando  $X^1$  es un grupo análogo de ciclooctinilo y  $R^2$  es halógeno, por ejemplo flúor, se prefiere particularmente que  $R^2$  esté unido al átomo del anillo que es adyacente al triple enlace.

45 De acuerdo con una realización particular adicional,  $R^2$  es alcanoiloxi  $C_2-C_5$  o alquilaminocarboniloxi  $C_1-C_4$ .

De acuerdo con una realización particular,  $X^1$  es ciclooctinilo (todos de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son  $-CH_2-$ ), es decir  $X^1$  tiene la fórmula

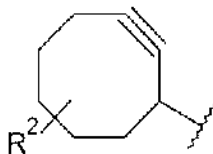




en la que  $R^2$  es como se define en el presente documento.

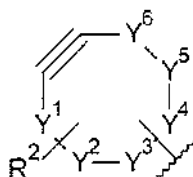
Estos grupos ciclooctilo sustituido o sin sustituir pueden estar unidos a  $X^2$  mediante un átomo del anillo en posición  $\alpha$ ,  $\beta$  o  $\gamma$  relativa al triple enlace. De acuerdo con una realización particular,  $X^1$  tiene una fórmula

5



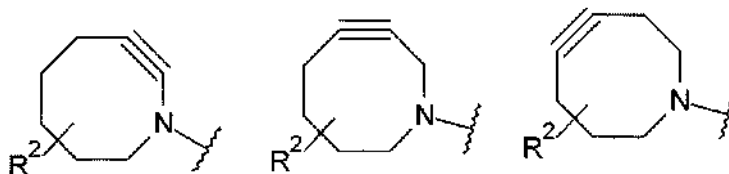
en la que  $R^2$  es como se define en el presente documento.

10 De acuerdo con otra realización particular,  $X^1$  es azaciclooctinilo, es decir  $X^1$  tiene la fórmula



15

en la que uno de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  es  $-NH-$  mientras que los cinco restantes de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son  $-CH_2-$  y  $R^2$  es como se define en el presente documento. Los residuos de azaciclooctinilo particulares incluyen radicales de 1-azaciclooctin-1-ilo que están enlazados a  $X^2$  mediante el átomo de nitrógeno, por ejemplo en los que  $X^1$  tiene una fórmula seleccionada entre



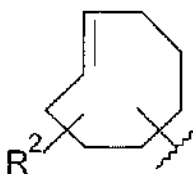
20

en la que  $R^2$  es como se define en el presente documento.

De acuerdo con una realización particular,  $X^1$  es ciclooctinilo no sustituido.

25 De acuerdo con otra realización particular,  $X^1$  es ciclooctinilo sustituido con uno o dos átomos de halógeno, por ejemplo átomos de flúor, unidos al átomo del anillo que es adyacente al triple enlace.

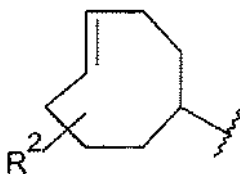
De acuerdo con otra realización particular,  $X^1$  es un *trans*-ciclooctenilo, es decir  $X^1$  tiene la fórmula



30

en la que  $R^2$  es como se define en el presente documento.

35 Estos grupos *trans*-ciclooctenilo sustituidos o sin sustituir pueden estar unidos a  $X^2$  mediante un átomo del anillo en posición  $\alpha$ ,  $\beta$  o  $\gamma$  relativa al doble enlace. De acuerdo con una realización particular,  $X^1$  tiene la fórmula



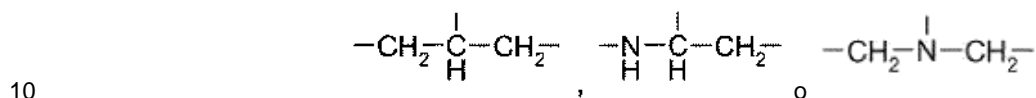
40

en la que  $R^2$  es como se define en el presente documento.

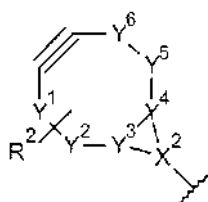
$X^2$  es  $-\text{CH}_2-$ ,  $-\text{O}-$ ,  $-\text{S}-$ ,  $-\text{NH}-$ ,  $-\text{C}(\text{O})-$ ,  $-\text{OC}(\text{O})-$ ,  $-\text{C}(\text{O})\text{O}-$ ,  $-\text{NHC}(\text{O})-$  o  $-\text{C}(\text{O})\text{NH}-$ .

Preferentemente,  $X^2$  es  $-\text{O}-$ .

- 5 Como alternativa,  $X^2$  es  $>\text{CH}-$  o  $>\text{N}-$  en el que el átomo de carbono o de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros; o  $X^2$  es  $-\text{CH}_2-\text{CH}<$ ,  $-\text{NH}-\text{CH}<$  o  $-\text{CH}_2-\text{N}<$  en el que los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 4 miembros; o  $X^2$  es  $-\text{CH}_2-\text{CH}_2-\text{CH}<$ ,  $-\text{NH}-\text{CH}_2-\text{CH}<$ ,  $-\text{CH}_2-\text{NH}-\text{CH}<$ ,  $-\text{CH}_2-\text{CH}_2-\text{N}<$ ,

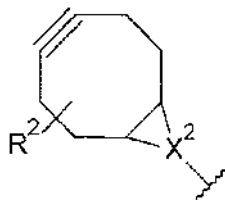


en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 5 miembros. Por ejemplo,  $X^1-X^2-$  tiene una fórmula



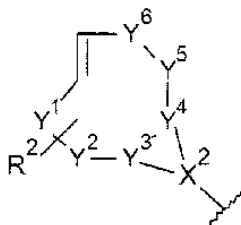
- 20 en la que  $Y^1$ ,  $Y^2$ ,  $Y^5$ ,  $Y^6$ ,  $X^2$  y  $R^2$  son como se definen en el presente documento, e  $Y^3$  e  $Y^4$  se seleccionan independientemente entre  $>\text{CH}-$  o  $>\text{N}-$ , en el que se prefiere si ambos  $Y^3$  e  $Y^4$  son  $>\text{CH}-$ . Se entiende que tales grupos análogos de biciclo[6,1,0]noninilo, biciclo[6,2,0]decinilo y biciclo[6,3,0]undecinilo comprenden el grupo análogo de ciclooctinilo sustituido o sin sustituir de la invención.

En caso de que  $X^2$  junto con dos átomos adyacentes en el anillo de  $X^1$  forme un anillo de 3, 4 o 5 miembros, esta es una realización particular si  $X^1-X^2-$  tiene una fórmula



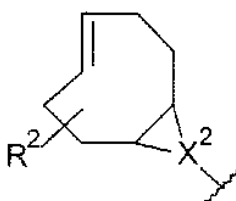
- en la que  $R^2$  y  $X^2$  son como se definen en el presente documento. Los grupos biciclo[6,1,0]noninilo (es decir  $X^2$  es  $>\text{CH}-$ ) representan una realización particular de la invención.

- 30 De acuerdo con otra realización,  $X^1-X^2-$  tiene una fórmula



- en la que  $Y^1$ ,  $Y^2$ ,  $Y^5$ ,  $Y^6$ ,  $X^2$  y  $R^2$  son como se definen en el presente documento, e  $Y^3$  e  $Y^4$  se seleccionan independientemente entre  $>\text{CH}-$  o  $>\text{N}-$ , en el que se prefiere si ambos  $Y^3$  e  $Y^4$  son  $>\text{CH}-$ . Se entiende que tales grupos análogos de biciclo[6,1,0]-*trans*-nonenilo, biciclo[6,2,0] -*trans*-decenilo y biciclo[6,3,0]-*trans*-undecenilo comprenden el grupo análogo de *trans*-ciclooctinilo sustituido o sin sustituir de la invención.

- 40 En caso de que  $X^2$  junto con dos átomos adyacentes en el anillo de  $X^1$  forme un anillo de 3, 4 o 5 miembros, esta es una realización particular si  $X^1-X^2-$  tiene una fórmula



en la que  $R^2$  y  $X^2$  son como se definen en el presente documento. Los grupos biciclo[6,1,0]-*trans*-nonenilo (es decir  $X^2$  es  $>CH-$ ) representa una realización particular de la invención.

5  $X^3$  es alquileo  $C_1-C_6$ ,  $-(CH_2-CH_2-O)_m-$  o un enlace sencillo; y  $m$  es 1, 2, 3, 4, 5 o 6.

En conexión con  $X^3$ , alquileo  $C_1-C_6$  se refiere preferiblemente a alquileo de cadena lineal.

10 De acuerdo con una realización preferida,  $X^3$  es  $-CH_2-CH_2-O-$  o un enlace sencillo.

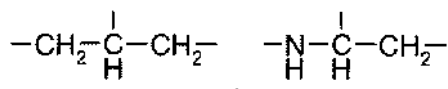
Como alternativa,  $X^3$  es  $-(CH_2-O)_p-$ ; y  $p$  es 1, 2, 3, 4, 5 o 6. De acuerdo con una realización particular,  $X^3$  es  $-CH_2-O-$  (es decir,  $p$  es 1).

15 De acuerdo con una realización particular, el elemento estructural  $-X^2-X^3-$  comprende de 1 a 6 átomos en la cadena principal, tal como 1, 2, 3 o 4 átomos en la cadena principal.

De acuerdo con una realización particular,  $-X^2-X^3-$  es  $-O-$  u  $-O-(CH_2)_2-O-$ . En caso de que  $X^2$  junto con dos átomos adyacentes en el anillo de  $X^1$  forme un anillo de 3, 4 o 5 miembros, siendo  $X^3$   $-CH_2-O-$  representa una realización particular adicional.

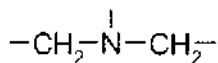
20 De acuerdo con una realización particular adicional,  $X^1-X^2-X^3-$  es  $X^1-O-$  o  $X^1-O-(CH_2)_2-O-$ , en el que  $X^1$  es como se define en el presente documento, preferiblemente ciclooctinilo no sustituido o *trans*-ciclooctenilo no sustituido.

25 De acuerdo con una realización particular adicional,  $X^1-X^2-X^3-$  es  $X^1-X^2-CH_2-O-$ , en el que  $X^1$  es como se define en el presente documento, preferiblemente ciclooctinilo no sustituido, y  $X^2$  es  $>CH-$  o  $>N-$  en el que el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros; o  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 4 miembros; o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



o

35



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 5 miembros.

40  $X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-C(O)-$ ,  $-NH-CH(NH_2)-$ ,  $-CH(NH_2)-NH-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$ ,  $-C(O)-NH-C(NH)-NH-$ ,  $NH-CH(NH_2)-C(O)-$  o  $-NH-C(NH)-NH-C(O)-$ .

En particular,  $X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-CH(NH_2)-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$  o  $-C(O)-NH-C(NH)-NH-$ .

45 De acuerdo con una realización preferida,  $X^4$  es  $-C(O)-NH-$ .

$n$  es un número entero de 1 a 4.

50 De acuerdo con una realización particular,  $n$  es 3 o 4.

De acuerdo con una realización preferida,  $n$  es 4.

55 De acuerdo con una realización particular,  $-X^4-(CH_2)_n-$  es  $-NH-(CH_2)_n-$ ,  $-NH-C(O)-(CH_2)_n-$ ,  $-NH-CH(NH_2)-(CH_2)_n-$ ,  $-NH-C(NH)-NH-(CH_2)_n-$ ,  $-C(O)-NH-CH(NH_2)-(CH_2)_n-$  o  $-C(O)-NH-C(NH)-NH-(CH_2)_n-$ , en los que  $n$  es preferiblemente 3 o 4.

De acuerdo con una realización preferida,  $-X^4-(CH_2)_n-$  es  $-C(O)-NH-(CH_2)_n-$ , en los que n es preferiblemente 3 o 4.

De acuerdo con una realización particular adicional,  $-X^4-(CH_2)_n-$  es  $-NH-(CH_2)_4-$ ,  $-NH-C(O)-CH_2-$ ,  $-NH-C(O)-(CH_2)_2-$ ,  $-NH-CH(NH_2)-(CH_2)_3-$ ,  $-NH-CH(NH_2)-(CH_2)_4-$ ,  $-NH-C(NH)-NH-(CH_2)_3-$ ,  $-C(O)-NH-CH(NH_2)-(CH_2)_3-$ ,  $-C(O)-NH-CH(NH_2)-(CH_2)_4-$  o  $-C(O)-NH-C(NH)-NH-(CH_2)_3-$ .

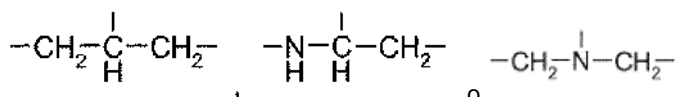
De acuerdo con una realización preferida,  $-X^4-(CH_2)_n$  es  $-C(O)-NH-(CH_2)_4-$ .

De acuerdo con un aspecto particular de la invención,  $-X^2-X^3-X^4-$  comprende una funcionalidad de carbamato  $-O-C(O)-NH-$  (por ejemplo,  $X^2$  es  $-O-$ ,  $X^3$  es un enlace y  $X^4$  es  $-C(O)-NH-$ , o  $X^3$  es  $-(CH_2-CH_2-O)_m-$  o  $-(CH_2-O)_p-$  y  $X^4$  es  $-C(O)-NH-$ ).

De acuerdo con una realización particular, el elemento estructural  $-X^2-X^3-X^4-(CH_2)_n-$  comprende de 5 a 12 átomos en la cadena principal, tal como 6, 7, 8, 9, 10 o 11 átomos en la cadena principal.

De acuerdo con una realización particular,  $-X^2-X^3-X^4-$  es  $-O-C(O)-NH-$ ,  $-O-CH_2-O-C(O)-NH-$  u  $-O-(CH_2)_2-O-C(O)-NH-$ .

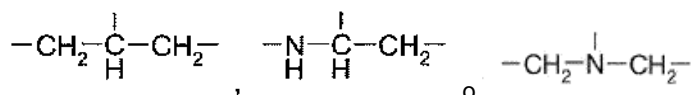
De acuerdo con una realización particular adicional,  $-X^2-X^3-X^4-$  es  $-X^2-CH_2-O-C(O)-NH-$  en el que  $X^2$  es  $>CH-$  o  $>N-$ , en el que el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros; o  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 4 miembros; o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 5 miembros.

De acuerdo con una realización preferida,  $X^1-X^2-X^3-X^4-(CH_2)_n-$  es  $X^1-O-C(O)-NH-(CH_2)_4-$ ,  $X^1-O-CH_2-O-C(O)-NH-(CH_2)_4-$  o  $X^1-O-(CH_2)_2-O-C(O)-NH-(CH_2)_4-$ , en el que  $X^1$  es como se define en el presente documento y es preferiblemente ciclooctinilo sin sustituir o *trans*-ciclooctenilo sin sustituir.

De acuerdo con una realización particular adicional,  $X^1-X^2-X^3-X^4-(CH_2)_n-$  es  $X^1-X^2-CH_2-O-C(O)-NH-(CH_2)_4-$ , en el que  $X^2$  es  $>CH-$  o  $>N-$ , en el que el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros; o  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 4 miembros; o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 5 miembros.

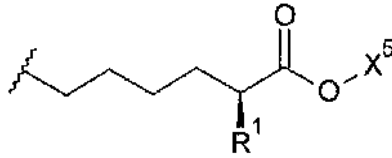
$X^5$  es hidrógeno, alquilo  $C_1-C_6$ , alcoxi  $C_1-C_6$ -alquilo  $C_1-C_2$ , alcanoiloxi  $C_2-C_7$ -alquilo  $C_1-C_2$  o alcanoilulfanil  $C_2-C_7$ -alquilo  $C_1-C_2$ .

De acuerdo con una realización particular,  $X^5$  es hidrógeno, alcóximetilo  $C_1-C_6$ , alcóxiet-1-ilo  $C_1-C_6$  (especialmente 1-(alcoxi  $C_1-C_6$ )et-1-ilo), alcanoiloximetilo  $C_2-C_7$  o alcanoilulfaniletilo  $C_2-C_7$ .

De acuerdo con una realización preferida,  $X^5$  es hidrógeno.

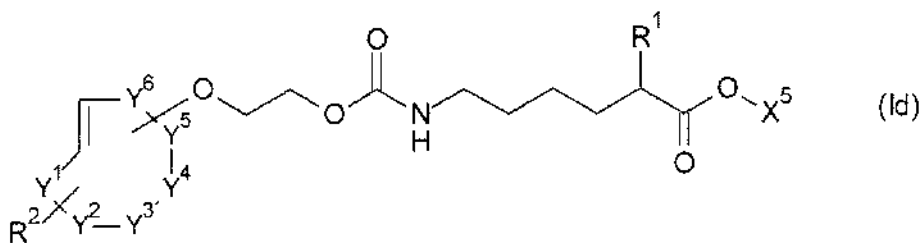
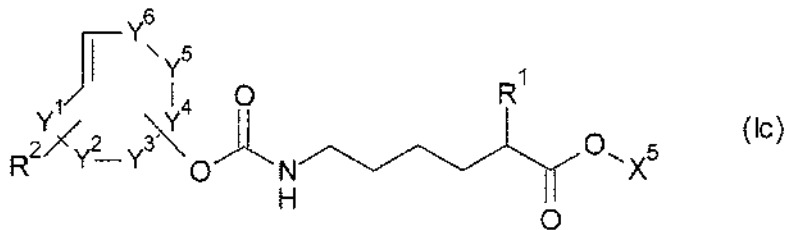
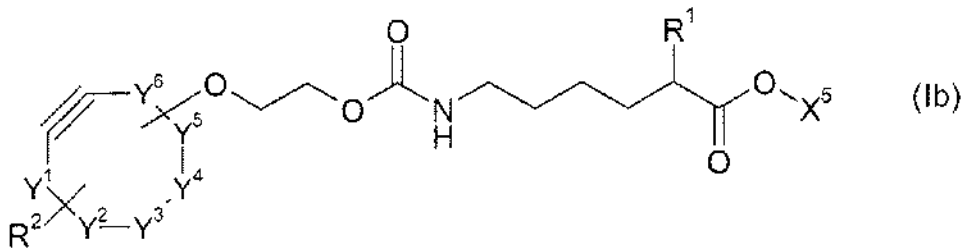
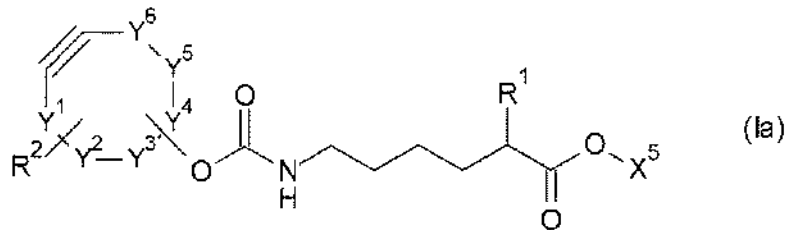
Con respecto al átomo de carbono asimétrico que porta  $R^1$ , el compuesto de la invención puede tener configuración S o R (de acuerdo con las reglas de prioridad de Cahn-Ingold-Prelog), prefiriéndose la configuración S.

De acuerdo con una realización preferida,  $-(CH_2)_n-CHR^1-C(O)O-X^5$  tiene la fórmula



5 en la que  $R^1$  y  $X^5$  son como se definen en el presente documento y  $X^5$  es en particular hidrógeno.

De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de una cualquiera de las fórmulas la, lb, lc e ld

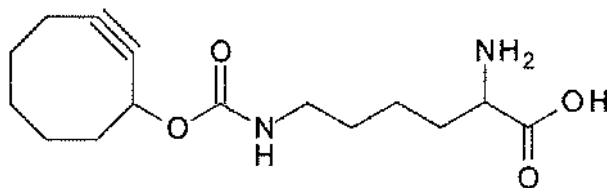


10

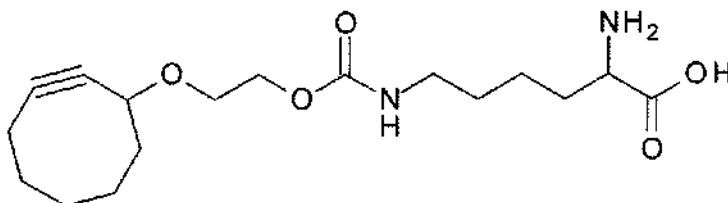
en las que  $R^1$ ,  $R^2$ ,  $X^5$ , e  $Y^1$  a  $Y^6$  son como se definen en el presente documento, o una sal de adición de ácidos o bases del mismo.

15 De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula la o lb, en la que  $R^1$ ,  $X^5$ , e  $Y^1$  a  $Y^6$  son como se definen en el presente documento y  $R^2$  es hidrógeno o halógeno, en particular flúor, o una sal de adición de ácidos o bases del mismo.

De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula



5 o

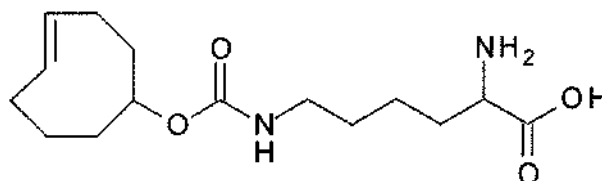


;

o una sal de adición de ácidos o bases del mismo.

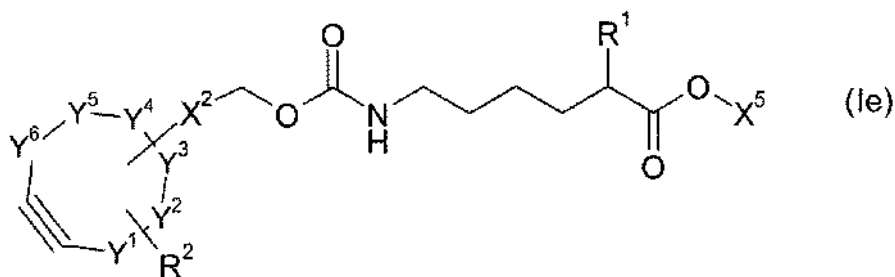
10

De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula



15 o una sal de adición de ácidos o bases del mismo.

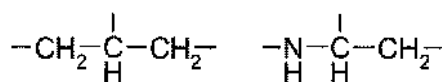
De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula



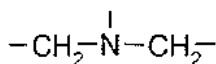
20

en la que  $R^1$ ,  $R^2$ ,  $X^5$ , e  $Y^1$  a  $Y^6$  son como se definen en el presente documento y  $X^2$  es  $>CH-$  o  $>N-$  en el que el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros; o  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 4 miembros; o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,

25



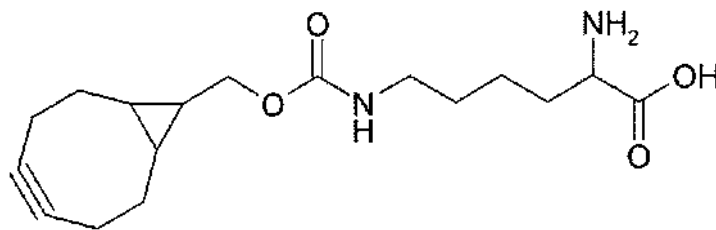
30 o



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos

adyacentes en el anillo de X<sup>1</sup> forma un anillo de 5 miembros, o una sal de adición de ácidos o bases del mismo.

De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula

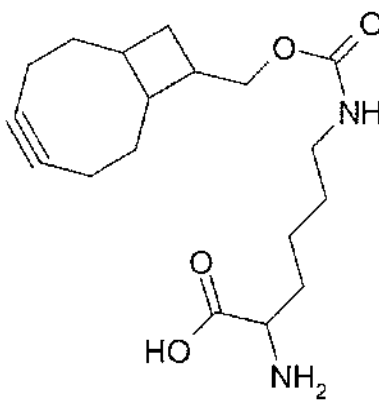


5

o una sal de adición de ácidos o bases del mismo.

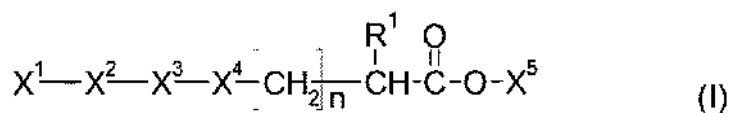
De acuerdo con una realización particular adicional, el compuesto o sal de la invención es un compuesto de fórmula

10



o una sal de adición de ácidos o bases del mismo.

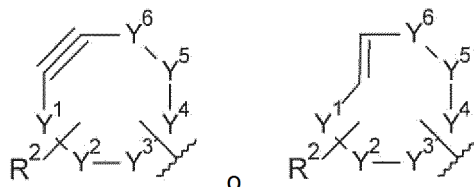
15 De acuerdo con una realización, los compuestos de la invención tienen por tanto la fórmula I



en la que:

20

X<sup>1</sup> tiene la fórmula



25

en la que:

Y<sup>1</sup>, Y<sup>2</sup>, Y<sup>3</sup>, Y<sup>4</sup>, Y<sup>5</sup>, Y<sup>6</sup>

son independientemente -CH<sub>2</sub>-, -NH-, -S- u -O- con la condición de que al menos 4 de Y<sup>1</sup>, Y<sup>2</sup>, Y<sup>3</sup>, Y<sup>4</sup>, Y<sup>5</sup>, Y<sup>6</sup> sean -CH<sub>2</sub>-;

30

R<sup>2</sup> es hidrógeno, halógeno, alquilo C<sub>1</sub>-C<sub>4</sub>, CF<sub>3</sub>, CN, alcoxi C<sub>1</sub>-C<sub>4</sub>, -O-CF<sub>3</sub>, alcanoiloxi C<sub>2</sub>-C<sub>5</sub>, alquilaminocarboniloxi C<sub>1</sub>-C<sub>4</sub> o alquiltio C<sub>1</sub>-C<sub>4</sub>;

X<sup>2</sup> es -CH<sub>2</sub>-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

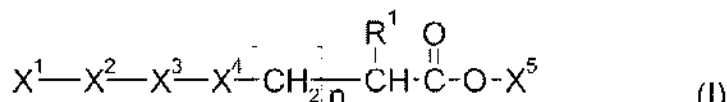
X<sup>3</sup> es alquilenos C<sub>1</sub>-C<sub>6</sub>, -(CH<sub>2</sub>-CH<sub>2</sub>-O)<sub>m</sub>- o un enlace sencillo;

X<sup>4</sup> es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH<sub>2</sub>)-, -CH(NH<sub>2</sub>)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH<sub>2</sub>)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH<sub>2</sub>)-C(O)- o -NH-C(NH)-NH-C(O)-;

35

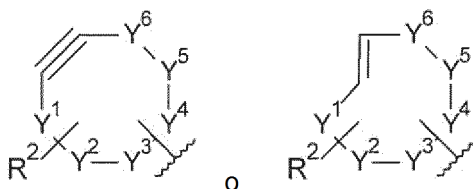
- $X^5$  es hidrógeno, alquilo  $C_1-C_6$ , alcoxi  $C_1-C_6$ -alquilo  $C_1-C_2$ , alcoiloxi  $C_2-C_7$ -alquilo  $C_1-C_2$  o alcoilsulfanil  $C_2-C_7$ -alquilo  $C_1-C_2$ ;  
 $R^1$  es  $-OH$  o  $-NH_2$ ;  
 $n$  es un número entero de 1 a 4; y  
 $m$  es un número entero de 1 a 6.

De acuerdo con otra realización, los compuestos de la invención tienen por tanto la fórmula I



10 en la que:

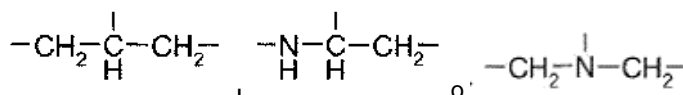
$X^1$  tiene la fórmula



15 en la que:  
 $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$

son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  o  $-O-$  con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ ;

- $R^2$  es hidrógeno, halógeno, alquilo  $C_1-C_4$ ,  $CF_3$ ,  $CN$ , alcoxi  $C_1-C_4$ ,  $-O-CF_3$ , alcoiloxi  $C_2-C_5$ , alquilaminocarbonilo  $C_1-C_4$  o alquiltio  $C_1-C_4$ ;  
 $X^2$  es  $>CH-$  o  $>N-$  en el que el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forma un anillo de 3 miembros, o  
 $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en el que los dos átomos de carbono o el carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 4 miembros, o  
 $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



en el que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes del anillo de  $X^1$  forman un anillo de 5 miembros;

- $X^3$  es alquilenos  $C_1-C_6$ ,  $-(CH_2-CH_2-O)_m-$ ,  $-(CH_2-O)_p-$  o un enlace sencillo;  
 $X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-C(O)-$ ,  $-NH-CH(NH_2)-$ ,  $-CH(NH_2)-NH-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$ ,  $-C(O)-NH-C(NH)-NH-$ ,  $NH-CH(NH_2)-C(O)-$  o  $-NH-C(NH)-NH-C(O)-$ ;  
 $X^5$  es hidrógeno, alquilo  $C_1-C_6$ , alcoxi  $C_1-C_8$ -alquilo  $C_1-C_2$ , alcoiloxi  $C_2-C_7$ -alquilo  $C_1-C_2$  o alcoilsulfanil  $C_2-C_7$ -alquilo  $C_1-C_2$ ;  
 $R^1$  es  $-OH$  o  $-NH_2$ ;  
 $n$  es un número entero de 1 a 4;  
 $m$  es un número entero de 1 a 6; y  
 $p$  es un número entero de 1 a 6.

Las sales de adición de ácidos o bases de los compuestos de la invención son especialmente sales de adición con ácidos o bases fisiológicamente toleradas. Las sales de adición de ácidos fisiológicamente toleradas pueden formarse por tratamiento de la forma de base de un compuesto de la invención con ácidos orgánicos o inorgánicos adecuados. Los compuestos de la invención que contienen un protón ácido pueden convertirse en sus formas de sal de adición no tóxicas de metal o amina mediante tratamiento con bases orgánicas e inorgánicas adecuadas. Los compuestos y sales de la invención también comprenden los hidratos y formas de adición de disolventes de los mismos, por ejemplo hidratos, alcoholatos y similares.

Los ácidos o bases fisiológicamente tolerables son aquellos que son tolerados mediante el sistema de translación usado para la preparación de polipéptidos con grupos análogos de ciclooctinilo o *trans*-ciclooctenilo, por ejemplo son



sustancialmente no tóxicos para células vivas.

5 Cuando se usan compuestos o sales de la invención, en los que X5 es distinto de hidrógeno, para la preparación de polipéptidos en un sistema de traducción, se cree que X5 se retira *in situ*, por ejemplo de forma enzimática dentro del sistema de traducción antes de incorporarse en el polipéptido. En consecuencia, X5 se elige de manera conveniente para que sea compatible con la capacidad del sistema de traducción de convertir el compuesto o sales de la invención en una forma que sea reconocida y procesada por la aminoacil ARNt sintetasa.

10 Los compuestos y sales de la invención pueden prepararse por analogía a métodos que son bien conocidos en la técnica. Se encuentran métodos adecuados para la preparación de compuestos de fórmula (I) en las diversas publicaciones citadas en el presente documento, todas las cuales se incorporan en el presente documento por referencia en su totalidad. Algunos métodos se indican en el presente documento.

15 Los compuestos de la invención que comprenden un grupo biciclo[6,1,0]noninilo pueden prepararse a partir de un precursor, tal como 9-hidroximetilbiciclo[6,1,0]nonino, que pueden sintetizarse de acuerdo con Dommerholt et al. (Angew. Chem. Int. Ed. 2010, 49:9422).

20 Los compuestos y sales de la invención pueden usarse para la preparación de polipéptidos que comprenden uno o varios grupos análogos de ciclooctinilo o *trans*-ciclooctenilo. La invención proporciona procesos para preparar dichos polipéptidos, *in vivo* o *in vitro*. En particular, Los compuestos o sales de la invención pueden incorporarse en la traducción en un polipéptido que está codificado por un polinucleótido que comprende uno o varios codones selectores.

25 La presente invención también se refiere por lo tanto a un proceso para preparar un polipéptido diana que tiene uno o varios grupos análogos de ciclooctinilo o *trans*-ciclooctenilo, comprendiendo el proceso:

a) proporcionar un sistema de traducción que comprende:

- 30 (i) una aminoacil ARNt sintetasa o un polinucleótido que la codifica;  
 (ii) un compuesto o una sal de la invención;  
 (iii) un ARNt que tiene un anticodón para un codón selector o un polinucleótido que codifica dicho ARNt; y  
 (iv) un polinucleótido que codifica el polipéptido diana y que comprende uno o varios codones selectores,

35 en el que la aminoacil ARNt sintetasa (i) es capaz de acilar específicamente el ARNt (iii) con el compuesto o sal (ii);

- b) permitir la traducción del polinucleótido (iv); y  
 c) opcionalmente recuperar el polipéptido resultante.

40 Los grupos norbornenilo reaccionan con 1,2,4,5-tetrazinas de una manera análoga a grupos *trans*-ciclooctenilo. Lo que se desvela en el presente documento con respecto a *trans*-ciclooctenilo se aplica por lo tanto de manera análoga a norbornenilo. Por lo tanto, se describen en el presente documento compuestos que tienen una fórmula correspondiente a fórmula (I), en la que X<sup>1</sup> es norbornen-2-ilo o norbornen-7-ilo, y en particular es norbornen-2-ilo.

45 La expresión «sistema de traducción» se refiere a los componentes necesarios para incorporar un aminoácido de origen natural en una cadena polipeptídica creciente (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetetasas, ARNm y similares.

El sistema de traducción puede ser un sistema de traducción *in vivo* o *in vitro*.

50 Un sistema de traducción *in vitro* puede ser un sistema de traducción sin células. Un sistema de traducción sin células es un sistema para sintetizar una proteína deseada obteniendo factores proteicos requeridos para la traducción de ARNm, por ejemplo, en forma de un extracto celular, seguido de reconstitución de esta reacción *in vitro*. Dichos sistemas sin células y su uso para la síntesis de proteínas se conocen en la técnica. Los ejemplos incluyen extractos de *E. coli*, extracto de germen de trigo o lisado de reticulocitos de conejo (Spirin y Swartz, Cell-free Protein Synthesis, Wiley VCH Verlag, Weinheim, Alemania, 2008).

60 Preferentemente, el sistema de traducción usado en el proceso de la invención es un sistema de traducción *in vivo*. Un sistema de traducción *in vivo* puede ser una célula, por ejemplo una célula procariota o eucariota. La célula puede ser una célula bacteriana, por ejemplo *E. coli*; una célula fúngica tal como una célula de levadura, por ejemplo *S. cerevisiae*; una célula vegetal o una célula animal tal como una célula de insecto o una célula de mamífero, por ejemplo una célula HeLa. Las células eucariotas usadas para expresión de polipéptidos pueden ser células individuales o partes de un organismo multicelular.

De acuerdo con una realización particular, el sistema de traducción es una célula *E. coli*.

65 De acuerdo con una realización particular adicional, el sistema de traducción es una célula de mamífero, por ejemplo

una célula HeLa.

Un sistema de traducción útil para la preparación de polipéptidos de la invención comprende, en particular, una aminoacil ARNt sintetasa o un polinucleótido que la codifica; un compuesto o una sal de la invención; un ARNt que  
5 tiene un anticodón para un codón selector o un polinucleótido que codifica dicho ARNt; un polinucleótido que codifica el polipéptido de la invención y que comprende uno o varios codones selectores.

Por ejemplo, pueden introducirse polinucleótidos que codifican la aminoacil ARNt sintetasa, el ARNt y el polipéptido de la invención en una célula por transfección/transformación conocidas en la técnica.

Una aminoacil ARNt sintetasa (RS) es una enzima capaz de acilar un ARNt con un aminoácido o análogo de aminoácido. De manera conveniente, la RS usada en procesos de la invención es capaz de acilar un ARNt con el aminoácido no natural de la invención.

15 Los procesos de la invención utilizan de manera conveniente un par de aminoacil ARNt ARNt sintetasa (ARNt/RS). Preferentemente, el par ARNt/RS usado en el proceso de la invención es ortogonal al sistema de traducción.

El término «ortogonal» como se usa en el presente documento se refiere a una molécula (por ejemplo, un ARNt ortogonal (ARNt-O) y/o una aminoacil ARNt sintetasa ortogonal (RS-O)) que se usa con eficacia reducida por un sistema de traducción de interés (por ejemplo, una célula). Ortogonal se refiere a la incapacidad o eficacia reducida, por ejemplo, menos del 20 % de eficacia, menos del 10 % de eficacia, menos del 5 % de eficacia o, por ejemplo, menos del 1 % de eficacia, de un ARNt ortogonal o una aminoacil ARNt sintetasa ortogonal para actuar con las aminoacil ARNt sintetatas endógenas o los ARNt endógenos del sistema de traducción de interés.

25 Por ejemplo, un ARNt ortogonal en un sistema de traducción de interés es acilado por cualquier aminoacil ARNt sintetasa endógena de un sistema de traducción de interés con eficacia reducida o incluso cero, en comparación con la acilación de un ARNt endógeno por la aminoacil ARNt sintetasa endógena. En otro ejemplo, una aminoacil ARNt sintetasa ortogonal acila cualquier ARNt endógeno en el sistema de traducción de interés con eficacia reducida o incluso cero, en comparación con la acilación del ARNt endógeno por una aminoacil ARNt sintetasa endógena.

30 Los pares de ARNt/RS ortogonales usados en procesos de la invención tienen preferentemente las siguientes propiedades: el ARNt-O es preferentemente acilado con el aminoácido no natural de la invención por la RS-O. Además, el par ortogonal actúa en el sistema de traducción de interés, por ejemplo, el sistema de traducción usa el ARNt-O acilado con aminoácido no natural para incorporar el aminoácido no natural de la invención en una cadena polipeptídica. La incorporación se produce de una manera específica de sitio, por ejemplo, el ARNt-O reconoce un  
35 codón selector, por ejemplo, un codón de terminación ámbar, en el ARNm que codifica el polipéptido.

La expresión «acila preferentemente» se refiere a una eficacia de, por ejemplo, aproximadamente 50 % de eficacia, aproximadamente 70 % de eficacia, aproximadamente 75 % de eficacia, aproximadamente 85 % de eficacia, aproximadamente 90 % de eficacia, aproximadamente 95 % de eficacia, o aproximadamente 99 % o más de eficacia, a la que una RS-O acila un ARNt-O con un aminoácido no natural en comparación con un ARNt endógeno o aminoácido de un sistema de traducción de interés. El aminoácido no natural se incorpora después en una cadena polipeptídica creciente con alta fidelidad, por ejemplo, a más de aproximadamente 75 % de eficacia para un codón selector dado, a más de aproximadamente 80 % de eficacia para un codón selector dado, a más de aproximadamente 90 % de eficacia para un codón selector dado, a más de aproximadamente 95 % de eficacia para un codón selector dado, o a más de aproximadamente 99 % o más de eficacia para un codón selector dado.

La expresión «codón selector» se refiere a codones reconocidos por el ARNt-O en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle anticodónico de ARNt-O reconoce el codón selector en el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido no natural, en este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones sin sentido, tales como codones de terminación, por ejemplo, codones ámbar, ocre y ópalo; codones de cuatro o más bases; codones derivados de pares de bases naturales o no naturales y similares. Para un sistema dado, un codón selector también puede incluir uno de los codones naturales de tres bases (es decir tripletes naturales), en el que el sistema endógeno no usa dicho triplete natural, por ejemplo, un sistema que carece de un ARNt que reconoce el triplete natural o un sistema en el que el triplete natural es un  
55 codón poco habitual.

Un anticodón tiene la secuencia complementaria inversa del codón correspondiente.

60 Un par ARNt-O/RS-O está compuesto de un ARNt-O, por ejemplo, un ARNt supresor, o similares, y una RS-O.

Un ARNt supresor es un ARNt que altera la lectura de un ARN mensajero (ARNm) en un sistema de traducción dado. Un ARNt supresor puede leer hasta, por ejemplo, un codón de terminación, un codón de cuatro bases o un codón poco habitual.

65 El ARNt-O no es acilado por sintetatas endógenas y es capaz de descodificar un codón selector, como se describe

en el presente documento. La RS-O reconoce el ARNt-O, por ejemplo, con un bucle anticodónico extendido, y preferentemente acila el ARNt-O con un aminoácido no natural.

5 El ARNt y la RS usados en los procesos de la invención pueden ser de origen natural o pueden obtenerse por mutación de un ARNt y/o una RS de origen natural de diversos organismos. En diversas realizaciones, el ARNt y la RS se obtienen de al menos un organismo. En otra realización, el ARNt se obtiene de un ARNt de origen natural o de origen natural mutado de un primer organismo y la RS se obtiene de RS de origen natural o de origen natural mutada de un segundo organismo.

10 Un par de ARNt/RS adecuado puede seleccionarse de bibliotecas de ARNt y RS mutantes, por ejemplo basándose en los resultados de una exploración de biblioteca. Como alternativa, un par de ARNt/RS adecuado puede ser un par de ARNt/sintetasa heterólogo que se importa de una especie fuente al sistema de traducción. Preferentemente, la célula usada como sistema de traducción es diferente de dichas especies fuente.

15 Se describen métodos para desarrollar pares de ARNt/RS, por ejemplo, en los documentos WO 02/085923 y WO 02/06075.

20 Preferentemente, la RS es una pirrolisil ARNt sintetasa (pyIRS) capaz de acilar un ARNt con el aminoácido no natural de la invención.

La pirrolisil ARNt sintetasa usada en procesos de la invención puede ser una pyIRS de tipo silvestre o una modificada por ingeniería genética. Los ejemplos de pyIRS de tipo silvestre incluyen, pero sin limitación pyIRS de arqueobacterias y eubacterias tales como *Methanosarcina maize*, *Methanosarcina barkeri*, *Methanococcoides burtonii*, *Methanosarcina acetivorans*, *Methanosarcina thermophila*, y *Desulfitobacterium hafniense*.

25 Se han descrito pyIRS modificadas por ingeniería genética, por ejemplo, en Neumann et al. (Nat Chem Biol 4:232, 2008), en Yanagisawa et al. (Chem Biol 2008, 15:1187) y en el documento EP2192185A1).

30 De acuerdo con una realización particular, la pirrolisil ARNt sintetasa usada para la preparación de polipéptidos de la invención es pirrolisil ARNt sintetasa de tipo silvestre de *M. maize*.

De acuerdo con una realización particular, la pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos de pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre expuesta en SEQ ID NO: 1, o un fragmento funcional de la misma.

35 SEQ ID NO: 1:

```
MDKKPLNLTLSATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL 60
RHHKYRKTCKRCRVSDDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLE 120
NTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMS 180
APVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRKKDLQQIYAEERE 240
NYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM 300
LAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEF TMLNFCQMGSGCTRENLE 360
SIITDFLNHLGIDFKIVGDS CMVYGD TLDVMHGDLELSSAVVGP I PLDREW GIDK PWIGA 420
GFGLERLLKVKHDFKNIKRAARSESYNGISTNL 454
```

40 De acuerdo con otra realización particular, la pirrolisil ARNt sintetasa es pirrolisil ARNt sintetasa de *M. maize* que comprende una o varias alteraciones de aminoácidos, preferentemente seleccionadas de las sustituciones de aminoácidos Y306A y Y384F.

45 De acuerdo con una realización particular, la pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos de pirrolisil ARNt sintetasa de *M. maize* mutante expuesta en SEQ ID NO: 2, o un fragmento funcional de la misma.

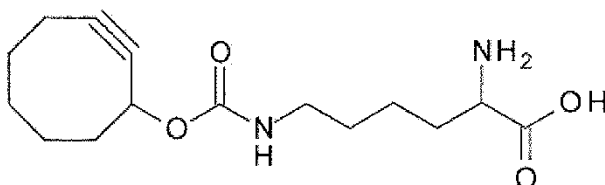
SEQ ID NO: 2:

MDKKPLNNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL 60  
 RHHKYRKTCKRCRVSDLDLNKFLTKANEDQTSVKVKVVSAPTTRTKKAMPKSVARAPKPLE 120  
 NTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMS 180  
 APVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQIYAEERE 240  
 NYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM 300  
 LAPNLANYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFMNLNFCOMGSGCTRENLE 360  
 SIITDFLNHLGIDFKIVGDSMVFGDTLDVMHGDELELSSAVVGP IPLDREWGIDKPWIGA 420  
 GFGLERLLKVKHDFKNIKRAARSESYYNGISTNL 454

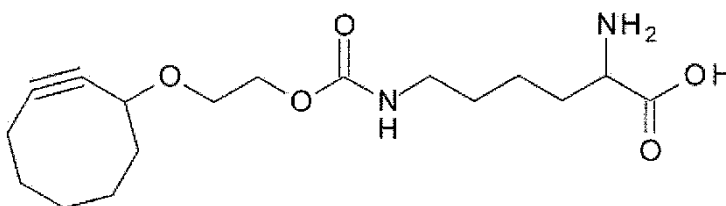
- 5 Cualquier aminoacil ARNt sintetasa descrita en el presente documento puede usarse para acilación de un ARNt con el compuesto de la invención.

De acuerdo con una realización preferida, se usa pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre para acilación de un ARNt con un compuesto de fórmula

10



o

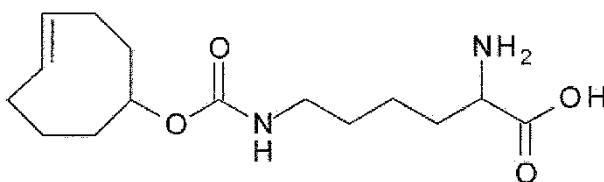


15

o una sal del mismo.

De acuerdo con una realización preferida adicional, se usa pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre para acilación de un ARNt con un compuesto de fórmula

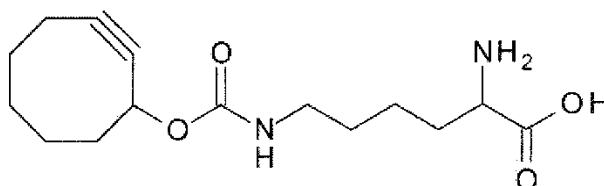
20



o una sal del mismo.

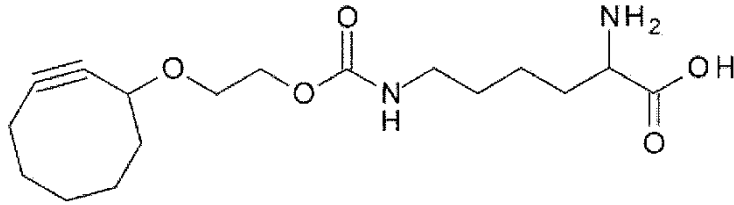
25

De acuerdo con otra realización preferida, se usa pirrolisil ARNt sintetasa de *M. maize* mutante que comprende las sustituciones de aminoácidos Y306A y Y384F para acilación de un ARNt con un compuesto de fórmula



30

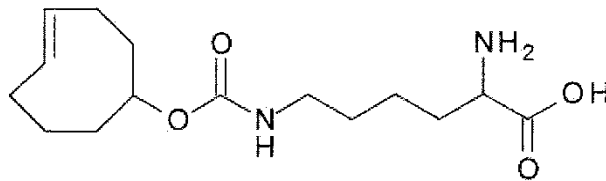
o



5 o una sal del mismo.

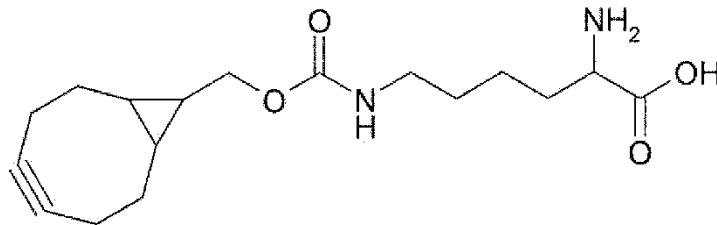
De acuerdo con una realización preferida adicional, se usa pirrolisil ARNt sintetasa de *M. maize* mutante que comprende las sustituciones de aminoácidos Y306A y Y384F para acilación de un ARNt con un compuesto de fórmula

10



o una sal del mismo.

15 De acuerdo con una realización preferida adicional, se usa pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre o una pirrolisil ARNt sintetasa de *M. maize* mutante que comprende las sustituciones de aminoácidos Y306A y Y384F para acilación de un ARNt con un compuesto de fórmula



20

o una sal del mismo.

El ARNt que se usa en combinación con la *pylRS* (ARNt<sup>pyl</sup>) puede ser un ARNt de tipo silvestre o uno modificado por ingeniería genética. Los ejemplos de ARNt<sup>pyl</sup> de tipo silvestre incluyen, pero sin limitación, ARNt de arqueobacterias y eubacterias, tal como se ha mencionado anteriormente, que facilitan la incorporación en la traducción de restos de pirrolisilo.

25

Codones selectores utilizados en procesos de la presente invención expanden el marco de codones genéticos de la maquinaria biosintética de proteínas del sistema de traducción usado. Por ejemplo, un codón selector incluye, por ejemplo, un codón de tres bases único, un codón sin sentido, tal como un codón de terminación, por ejemplo, un codón ámbar o un codón ópalo, un codón no natural, al menos un codón de cuatro bases o similares. Pueden introducirse varios codones selectores en un polinucleótido que codifica un polipéptido deseado (polipéptido diana), por ejemplo, uno o más, dos o más, más de tres, etc.

30

Los 64 codones genéticos codifican 20 aminoácidos y tres codones de terminación. Debido a que solo es necesario un codón de terminación para terminación de la traducción, los otros dos pueden usarse en principio para codificar aminoácidos no proteínogénicos. El codón de terminación ámbar, UAG, se ha usado con éxito en sistema biosintético *in vitro* y en oocitos de *Xenopus* para dirigir la incorporación de aminoácidos no naturales. Entre los tres codones de terminación, UAG es el codón de terminación menos usado en *E. coli*. Algunas cepas de *E. coli* contienen ARNt supresores naturales, que reconocen UAG e insertan un aminoácido natural. Además, estos ARNt supresores de ámbar se han usado en mutagénesis de proteínas convencional.

40

En una realización, los métodos implican el uso de un codón selector que es un codón de terminación para la incorporación de un compuesto de la invención. Por ejemplo, se genera un ARNt-O que reconoce el codón de terminación, preferentemente el codón de terminación ámbar, y es acilado por una RS-O con un compuesto de la

45

invención. Este ARNt-O no es reconocido por las aminoacil ARNt sintetetasas de origen natural. Puede usarse mutagénesis dirigida convencional para introducir el codón de terminación, por ejemplo, el codón de terminación ámbar, en el sitio de interés en la secuencia polinucleotídica que codifica el polipéptido diana. Cuando la RS-O, el ARNt-O y el gen mutante se combinan en un sistema de traducción, el aminoácido no natural se incorpora en respuesta al codón de terminación ámbar para proporcionar un polipéptido que contiene el análogo de aminoácido no natural, es decir el compuesto de la invención, en la posición o las posiciones especificadas.

La incorporación de los compuestos de la invención *in vivo* puede realizarse sin perturbación significativa del hospedador, por ejemplo, una célula de *E. coli* o HeLa. Por ejemplo, debido a que la eficacia de supresión para el codón de terminación ámbar depende de la competición entre el ARNt-O, por ejemplo, el ARNt supresor de ámbar, y el factor de liberación 1 (RF1) (que se une con el codón de terminación ámbar e inicia la liberación del péptido creciente del ribosoma), la eficacia de supresión puede modularse, por ejemplo, aumentando el nivel de expresión de ARNt-O, por ejemplo, el ARNt supresor, o usando una cepa deficiente en RF1.

De acuerdo con realizaciones particulares, el ARNt<sup>pyl</sup> usado en procesos de la invención comprende el anticodón CUA para el codón de terminación ámbar.

Otros codones selectores útiles para codificar compuestos de la invención son codones poco habituales. Por ejemplo, cuando se reduce la concentración de arginina en una reacción de síntesis de proteínas *in vitro*, el codón de arginina poco habitual, AGG, ha demostrado ser eficaz para la inserción de Ala por un ARNt sintético acilado con alanina. En este caso, el ARNt sintético compete con el ARNt<sup>Arg</sup> de origen natural, que existe como una especie menor en *E. coli*. Algunos organismos no usan todos los codones tripletes. Por ejemplo, se ha utilizado un codón no asignado AGA en *Micrococcus luteus* para la inserción de aminoácidos en un extracto de transcripción/traducción *in vitro*. En consecuencia, cualquier codón triplete no usado por el sistema de traducción aplicado en los procesos de la invención puede actuar como codón selector.

El sistema de traducción se mantiene durante un tiempo adecuado en condiciones que permiten la formación del polipéptido diana por un ribosoma. El ribosoma se une a ARNm que codifica el polipéptido diana y comprende uno o varios codones selectores. A continuación, el polipéptido se forma por unión por etapas de aminoácidos en posiciones codificadas por codones que se unen con los aminoacil ARNt respectivos. Por lo tanto, el compuesto de la invención se incorpora en el polipéptido diana en la posición o las posiciones codificadas por el codón o los codones selectores.

Puede efectuarse traducción del polipéptido diana por un sistema de traducción por procedimientos bien conocidos en la técnica. Para facilitar la traducción eficaz, los componentes del sistema de traducción pueden mezclarse. Las células usadas como sistema de traducción se cultivan de manera conveniente y se mantienen en un medio de expresión adecuado en condiciones y durante un tiempo adecuados para producir el polipéptido diana. Puede ser necesario inducir la expresión mediante la adición de un compuesto, tal como arabinosa, isopropil β-D-tiogalactósido (IPTG) o tetraciclina que permite la transcripción del gen del polipéptido diana.

Opcionalmente, después de la traducción el polipéptido diana puede recuperarse del sistema de traducción. A tal fin, los polipéptidos diana pueden recuperarse y purificarse, bien parcialmente o bien sustancialmente hasta su homogeneidad, según procedimientos conocidos y usados por los expertos en la materia. Los procedimientos convencionales bien conocidos en la técnica incluyen, por ejemplo, precipitación con sulfato de amonio o etanol, extracción con ácido o base, cromatografía en columna, cromatografía en columna de afinidad, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de hidroxilapatita, cromatografía de lectina, electroforesis en gel y similares. Pueden usarse etapas de replegamiento de proteínas, según se desee, para preparar proteínas maduras correctamente plegadas. Pueden emplearse cromatografía líquida de alto rendimiento (HPLC), cromatografía de afinidad u otros métodos adecuados en etapas de purificación final cuando se desee alta pureza. Pueden usarse anticuerpos preparados contra el aminoácido no natural de la invención o contra los polipéptidos diana como reactivos de purificación, es decir para purificación basada en afinidad de los polipéptidos.

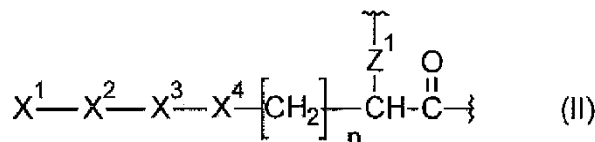
Diversos métodos de purificación/plegamiento de proteínas son bien conocidos en la técnica, incluyendo, por ejemplo, los expuestos en Scopes, Protein Purification, Springer, Berlín (1993); y Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press (1990); y las referencias citadas en esos documentos.

Como se ha indicado, los expertos en la materia reconocerán que, después de la síntesis, expresión y/o purificación, los polipéptidos pueden poseer una conformación diferente de las conformaciones deseadas de los polipéptidos relevantes. Por ejemplo, los polipéptidos producidos por sistemas procariontes con frecuencia se optimizan por exposición a agentes caotrópicos para conseguir plegamiento apropiado. Durante la purificación de, por ejemplo, lisados derivados de *E. coli*, el polipéptido expresado opcionalmente se desnaturaliza y después se renaturaliza. Esto se consigue, por ejemplo, solubilizando las proteínas en un agente caotrópico tal como guanidina HCl. En general, es en ocasiones deseable desnaturalizar y reducir los polipéptidos expresados y después hacer que los polipéptidos se replieguen en su conformación preferida. Por ejemplo, guanidina, urea, DTT, DTE y/o una chaperonina pueden añadirse a un producto de traducción de interés. Los expertos en la materia conocen bien

métodos para reducir, desnaturalizar y renaturalizar proteínas. Los polipéptidos pueden replegarse en un tampón redox que contiene, por ejemplo, glutatión oxidado y L-arginina.

5 También se describen polipéptidos producidos por los procesos de la invención. Dichos polipéptidos pueden prepararse por un proceso que hace uso de un sistema de traducción.

Por lo tanto también se describe un polipéptido que comprende uno o varios restos de fórmula II



10 en donde  $X^1$ ,  $X^2$ ,  $X^3$ ,  $X^4$ , y  $n$  son como se define en el presente documento y  $Z^1$  es -O- o -NH-.

El grupo análogo de ciclooctinilo o trans-ciclooctenilo de los polipéptidos facilita la unión covalente de una molécula de interés por reacciones clic sin metales.

15 Dichas reacciones incluyen cicloadiciones de grupos análogos de ciclooctinilo con azidas, óxidos de nitrilo, nitronas y reactivos de diazocarbonilo (Sanders et al., J Am Chem Soc 2010, 133:949; Agard et al., J Am Chem Soc 2004, 126:15046). Los óxidos de nitrilo pueden prepararse convenientemente por oxidación directa de las oximas correspondientes. De manera conveniente, esta oxidación y la cicloadición de los óxidos de nitrilo resultantes con grupos análogos de ciclooctinilo de polipéptidos se realizan como un procedimiento de un solo reactor. En consecuencia, las moléculas adecuadas de interés para unión con polipéptidos que portan uno o varios grupos análogos de ciclooctinilo pueden tener uno o varios grupos azida, óxido de nitrilo, oxima, nitrona o diazocarbonilo.

20 Se ha informado de que grupos *trans*-Ciclooctenilo reaccionan de forma eficaz con compuestos que comprenden un grupo 1,2,4,5-tetrazina por una cicloadición de Diels-Alder de demanda de electrones inversa (Devaraj et al., Angew Chem Int Ed Engl 2009, 48:7013). En consecuencia, las moléculas adecuadas de interés para unión con polipéptidos que portan uno o varios grupos análogos de trans-ciclooctenilo pueden tener uno o varios grupos 1,2,4,5-tetrazina.

25 Las moléculas de interés que pueden unirse con un polipéptido como se describe en el presente documento incluyen, pero sin limitación, un marcador detectable; un fármaco; una toxina; un enlazador; un péptido; un miembro de un par de unión específico; una etiqueta epitópica; y similares.

30 Los marcadores detectables que pueden unirse con un polipéptido como se describe en el presente documento incluyen, pero sin limitación, moléculas fluorescentes (por ejemplo, moléculas autofluorescentes o moléculas capaces de emitir fluorescencia tras su contacto con un reactivo), marcadores de espín o cromóforos para estudios de TERF (por ejemplo, para estudiar la estructura de polipéptidos *in vivo*), marcadores radiactivos (por ejemplo,  $^{11}In$ ,  $^{125}I$ ,  $^{131}I$ ,  $^{212}B$ ,  $^{90}Y$ ,  $^{186}Rh$ , y similares), biotina (por ejemplo, para detectar mediante reacción de biotina y avidina), etiquetas de purificación (distintos de biotina), y similares. Los marcadores detectables también incluyen péptidos o polipéptidos que pueden detectarse mediante unión a anticuerpos, por ejemplo, mediante unión de un anticuerpo marcado de forma detectable o mediante detección de anticuerpo unido a través de un ensayo de tipo sándwich.

35 Los fármacos que pueden unirse con un polipéptido como se describe en el presente documento incluyen, pero sin limitación, compuestos citotóxicos (por ejemplo, compuestos quimioterapéuticos del cáncer); compuestos antivíricos; modificadores de la respuesta biológica (por ejemplo, hormonas, quimiocinas, citocinas, interleucinas, etc.); agentes que afectan a los microtúbulos; moduladores de hormonas; compuestos esteroideos; y similares.

40 Los compañeros de unión específicos que pueden unirse con un polipéptido como se describe en el presente documento incluyen, pero sin limitación, un miembro de un par de receptor/ligando; un miembro de un par de anticuerpo/antígeno; un miembro de un par de lectina/carbohidrato; un miembro de un par de enzima/sustrato; biotina/avidina; biotina/estreptavidina; digoxina/antidigoxina; y similares.

45 Para facilitar la unión de dicha molécula de interés esta se pone en contacto con un polipéptido como se describe en el presente documento. En muchos casos, esta puesta en contacto se puede llevar a cabo en condiciones fisiológicas, es decir condiciones compatibles con células vivas. La reacción entre grupo análogo de ciclooctinilo y azida o entre grupo análogo de ciclooctinilo y grupo óxido de nitrilo, respectivamente, es selectiva y compatible con ambientes acuosos. Puede llevarse a cabo unión de las moléculas de interés anteriormente descritas grupos análogos de ciclooctinilo o *trans*-ciclooctenilo de los polipéptidos descritos en el presente documento *in vitro*. A tal fin, el polipéptido puede purificarse o proporcionarse como parte del sistema de expresión usado para su preparación. Como alternativa, la reacción puede llevarse a cabo *in vivo* poniendo en contacto dicha molécula de interés con una célula, en la que se ha expresado un polipéptido como se describe en el presente documento. En dicha célula el polipéptido puede localizarse en la superficie celular, dentro de la membrana celular o puede ser intracelular.

Los polipéptidos como se describen en la presente invención, opcionalmente marcados, fosforilados y/o glucosilados, pueden usarse como componentes de ensayo, por ejemplo para la detección de compuestos en bioensayos, para tratamientos terapéuticos, profilácticos o cosméticos o como inmunógenos para la producción de anticuerpos. Para dichos fines, los polipéptidos pueden aplicarse en forma purificada o proporcionarse como parte del sistema de expresión usado para su preparación.

Los compuestos y sales de la invención pueden ser parte de un kit para preparar un polipéptido con uno o más grupos análogos de ciclooctinilo o *trans*-ciclooctenilo.

Por lo tanto, se describen además kits para preparar un polipéptido que tiene uno o varios grupos análogos de ciclooctinilo o *trans*-ciclooctenilo (polipéptido diana). Los kits comprenden un compuesto o una sal de la invención y opcionalmente uno o más medios para preparar el polipéptido. Dichos medios incluyen, pero sin limitación

- i) una aminoacil ARNt sintetasa o un polinucleótido que la codifica;
- ii) un ARNt como se describe en el presente documento o un polinucleótido que lo codifica.

Tanto la aminoacil ARNt sintetasa como el ARNt pueden, por ejemplo, proporcionarse en forma de uno o varios vectores de expresión para dicha aminoacil ARNt sintetasa y ARNt correspondiente.

Dicho kit también puede comprender un polinucleótido que codifica una proteína indicadora, por ejemplo un vector de expresión para, por ejemplo, GFP, en el que la secuencia polinucleotídica que codifica dicha proteína indicadora comprende un codón de terminación ámbar. Dicho polinucleótido que codifica proteína indicadora puede actuar como un control positivo para confirmar la expresión de un polipéptido con grupo o grupos análogos de ciclooctinilo o *trans*-ciclooctenilo.

Además, dicho kit puede comprender medios adicionales para traducción de un polinucleótido que codifica dicho polipéptido, por ejemplo un sistema de traducción, tal como células *E. coli*, células HeLa, extracto de *E. coli*, extracto de germen de trigo o lisado de reticulocitos de conejo, e instrucciones para su uso.

## Ejemplos

### Ejemplos de preparación

#### Materiales y métodos generales

A menos que se indique otra cosa, los materiales para las síntesis químicas se obtuvieron de proveedores comerciales (Sigma-Aldrich, Aldrich, Sigma, Fluka, Acros, Iris) con la pureza más alta disponible y se usaron sin purificación adicional. Se adquirieron disolventes secos de Sigma-Aldrich, Acros, y Fluka; almacenados sobre tamices moleculares; y se usaron según se suministraron. Los disolventes usados para extracción y cromatografía se adquirieron de Fluka, Thermo Fisher Scientific, Merck, y BDH Prolabo (VWR). Se realizó cromatografía ultrarrápida (FC) usando gel de sílice 60 de Merck (malla 63 - 200), y se realizó cromatografía de capa fina (TLC) en placas de gel de sílice prerrecubiertas, con recubrimiento de aluminio (Macherey-Nagel Alugram Sil G/UV<sub>254</sub> y gel de sílice 60 Merck WF<sub>254s</sub>) con mezclas de *n*-Hex/EtOAc o DCM/MeOH/AcOH como fases móviles. Las manchas se detectaron mediante una lámpara de mano UV a 254 nm o 366 nm o tintando con A) solución de tinte de anisaldehído (85 ml de EtOH, 10 ml de AcOH, 5 ml de H<sub>2</sub>SO<sub>4</sub> concentrado, 0,5 ml de anisaldehído), B) solución de tinte de KMnO<sub>4</sub> (3,0 g de KMnO<sub>4</sub>, 20 g de K<sub>2</sub>CO<sub>3</sub> en 300 ml de NaOH acuoso al 5 %), o C) solución de tinte de ninhidrina (250 ml de EtOH, 1,5 ml de AcOH, 0,5 g de ninhidrina) y posterior tratamiento térmico.

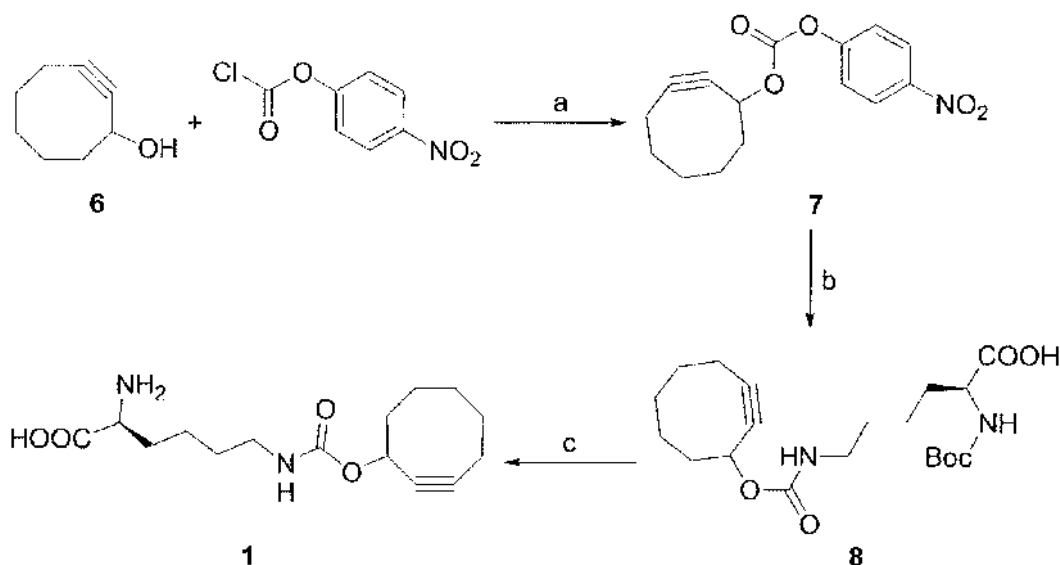
Los espectros de RMN se registraron usando un espectrómetro Bruker UltraShield™ Advance 400 (400 MHz, <sup>1</sup>H; 100 MHz, <sup>13</sup>C) y calibrado usando disolvente no deuterado residual como referencia interna. Los espectros de masas de alta resolución (HR) se registraron en la University of Heidelberg usando ionización por electronebulización (IEN) EM en un espectrómetro de masas Bruker ApexQe híbrido 9,4 T FT-ICR. Los productos se caracterizaron por RMN (<sup>1</sup>H, <sup>13</sup>C) y EM HR.

#### Ejemplo 1: *N*-ε-((Ciclooct-2-in-1-iloxi)carbonil)-L-lisina

La *N*-ε-((ciclooct-2-in-1-iloxi)carbonil)-L-lisina (1) puede prepararse como se indica en el esquema 1.

Esquema 1: Síntesis de derivado de ciclooctina lisina 1. Reactivos y condiciones: a) TEA, THF, de -10 °C a TA, 83 %; b) Boc-L-Lys-OH, TEA, DMF, de 0 °C a TA, 91 %; c) ácido fórmico, CHCl<sub>3</sub>, TA, 96 %.





El Compuesto **6**, ciclooct-2-in-1-ol se sintetizó de acuerdo con Reese y Shaw (Chem Commun 1970, 1142).

5 **a) Carbonato de ciclooct-2-in-1-il 4-nitrofenilo (7)**

El Compuesto **6** (3,12 g, 25,1 mmol) y TEA (4,20 ml, 30,2 mmol, 1,2 equiv.) se disolvieron en THF (0,2 M, 126 ml) y se añadieron gota a gota a una solución agitada de cloroformato de 4-nitrofenilo (15,20 g, 75,4 mmol, 3,0 equiv.) y THF (0,7 M, 36 ml) durante un periodo de 1 h a  $-10\text{ }^{\circ}\text{C}$ . La mezcla de reacción se dejó calentar hasta TA y se agitó durante una noche. Después, se añadió cHex (100 ml) y el THF se retiró a presión reducida. La filtración de la mezcla de reacción seguido de FC (cHex:EtOAc 9:1) dio **7** en forma de un aceite de color amarillo (6,03 g, 20,9 mmol, 83 %).

$F_r$  (cHex/EtOAc 4:1) = 0,70.

RMN  $^1\text{H}$  ( $\text{CDCl}_3$ )  $\delta$  = 8,28 (dt,  $^3J$  = 9,30,  $^3J$  = 2,72, 2H,  $\text{CH}^{\text{aromático}}$ ), 7,40 (dt,  $^3J$  = 9,30,  $^3J$  = 2,72, 2H,  $\text{CH}^{\text{aromático}}$ ), 5,30-5,35 (m, 1H,  $\text{CH}^{\text{propargilo}}$ ), 2,10-2,38 (m, 3H,  $\text{CH}_2^{\text{anillo}}$ ), 1,92-1,99 (m, 2H,  $\text{CH}_2^{\text{anillo}}$ ), 1,74-1,88 (m, 3H,  $\text{CH}_2^{\text{anillo}}$ ), 1,59-1,64 (m, 2H,  $\text{CH}_2^{\text{anillo}}$ ) ppm.

20 **b) *N*- $\alpha$ -*tert*-Butiloxicarbonil-*N*- $\epsilon$ -((ciclooct-2-in-1-iloxi)carbonil)-L-lisina (8)**

El Compuesto **7** (0,93 g, 3,21 mmol) se disolvió en DMF (0,2 M, 16 ml) y se añadió gota a gota a una solución agitada de Boc-L-Lys-OH (1,03 g, 4,18 mmol, 1,3 equiv.) y TEA (1,35 ml, 9,64 mmol, 3,0 equiv.) en DMF (0,5 M, 6 ml) durante un periodo de 1 h a  $0\text{ }^{\circ}\text{C}$ . La mezcla de reacción se agitó a TA durante una noche. Después de la retirada de todos los componentes volátiles por evaporación a presión reducida, el residuo se recogió en  $\text{H}_2\text{O}$  (100 ml) y EtOAc (100 ml). La fase acuosa se acidificó con HCl concentrado y se extrajo con EtOAc (3 x 50 ml). Las capas orgánicas combinadas se lavaron con una solución saturada de NaCl y se secaron sobre  $\text{Na}_2\text{SO}_4$ . El disolvente se evaporó a presión reducida y el producto en bruto se purificó por FC (DCM/MeOH/AcOH 97:2:1) para producir **8** en forma de un aceite muy viscoso de color amarillo (1,17 g, 2,94 mmol, 91 %).

$F_r$  (DCM/MeOH/AcOH 97:2:1) = 0,45.

RMN  $^1\text{H}$  ( $\text{CDCl}_3$ )  $\delta$  = 5,26-5,32 (m, 1H,  $\text{CH}^{\text{propargilo}}$ ), 5,19-5,25 (m, 1H, NH), 4,80-4,86 (m, 1H, NH), 4,24-4,34 (m, 1H,  $\alpha\text{-CH}^{\text{Lys}}$ ), 3,16 (c,  $^3J$  = 6,32, 2H,  $\epsilon\text{-CH}_2^{\text{Lys}}$ ), 2,10-2,31 (m, 3H,  $\text{CH}_2^{\text{anillo}}$ ), 1,63-2,04 (m, 9H,  $\text{CH}_2^{\text{anillo}}$ ,  $\text{CH}_2^{\text{Lys}}$ ), 1,49-1,57 (m, 4H,  $\text{CH}_2^{\text{Lys}}$ ), 1,45 (s, 9H, Boc) ppm.

RMN  $^{13}\text{C}$  ( $\text{CDCl}_3$ )  $\delta$  = 176,5 ( $\text{C}(\text{O})\text{O}_2$ ), 156,1 ( $\text{C}^{\text{Lys}}$ ), 154,7 ( $\text{C}^{\text{Boc}}$ ), 101,6, 91,2 (2x  $\text{C}^{\text{anillo}}$ ), 80,1 ( $\text{C}(\text{CH}_3)_3^{\text{Boc}}$ ), 67,1 ( $\text{CH}^{\text{propargilo}}$ ), 53,3 ( $\alpha\text{-CH}^{\text{Lys}}$ ), 41,9 ( $\text{CH}_2^{\text{anillo}}$ ), 40,5 ( $\epsilon\text{-CH}_2^{\text{Lys}}$ ), 34,2 ( $\text{CH}_2^{\text{anillo}}$ ), 31,9 ( $\text{CH}_2^{\text{Lys}}$ ), 29,7 ( $\text{CH}_2^{\text{anillo}}$ ), 29,3 ( $\text{CH}_2^{\text{Lys}}$ ), 28,5 (3x  $\text{CH}_3^{\text{Boc}}$ ), 26,2 ( $\text{CH}_2^{\text{anillo}}$ ), 22,4 ( $\text{CH}_2^{\text{Lys}}$ ), 20,7 ( $\text{CH}_2^{\text{anillo}}$ ) ppm.

35 **c) *N*- $\epsilon$ -((Ciclooct-2-in-1-iloxi)carbonil)-L-lisina (1)**

El Compuesto **8** (1,24 g, 3,13 mmol) se disolvió en ácido fórmico al 70 % en  $\text{CHCl}_3$  (0,2 M, 16 ml) y se agitó durante 36 h a TA. Se añadió DMF (0,2 M, 16 ml) y todos los componentes volátiles se retiraron a presión reducida. El material residual se recogió en HCl 0,1 M (100 ml) y se liofilizó, proporcionando la sal pura de HCl de **1** en forma de un sólido de color amarillo (1,00 g, 3,01 mmol, 96 %).

RMN  $^1\text{H}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 7,43-7,79 (m, 2H,  $\alpha\text{-NH}_2$ ), 7,18 (t,  $^3J$  = 5,72, 1H,  $\epsilon\text{-NH}$ ), 5,09-5,13 (m, 1H,  $\text{CH}^{\text{propargilo}}$ ), 3,05 (t,  $^3J$  = 5,72, 1H,  $\alpha\text{-CH}^{\text{Lys}}$ ), 2,90 (c,  $^3J$  = 5,87, 2H,  $\epsilon\text{-CH}_2^{\text{Lys}}$ ), 2,00-2,26 (m, 3H,  $\text{CH}_2^{\text{anillo}}$ ), 1,77-1,91 (m, 3H,  $\text{CH}_2^{\text{anillo}}$ ), 1,63-1,71 (m, 2H,  $\text{CH}_2^{\text{anillo}}$ ), 1,41-1,59 (m, 4H,  $\text{CH}_2^{\text{anillo}}$ ,  $\text{CH}_2^{\text{Lys}}$ ), 1,23-1,37 (m, 4H,  $\text{CH}_2^{\text{Lys}}$ ) ppm.

RMN  $^{13}\text{C}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 170,8 ( $\text{C}(\text{O})\text{O}_2$ ), 155,7 ( $\text{C}^{\text{Lys}}$ ), 101,2, 92,4 (2x  $\text{C}^{\text{anillo}}$ ), 66,0 ( $\text{CH}^{\text{propargilo}}$ ), 54,4 ( $\alpha\text{-CH}^{\text{Lys}}$ ), 42,1 ( $\text{CH}_2^{\text{anillo}}$ ), 40,6 ( $\epsilon\text{-CH}_2^{\text{Lys}}$ ), 34,3 ( $\text{CH}_2^{\text{anillo}}$ ), 31,1 ( $\text{CH}_2^{\text{Lys}}$ ), 29,7 ( $\text{CH}_2^{\text{anillo}}$ ), 29,5 ( $\text{CH}_2^{\text{Lys}}$ ), 26,3 ( $\text{CH}_2^{\text{anillo}}$ ), 22,8 ( $\text{CH}_2^{\text{Lys}}$ ), 20,4 ( $\text{CH}_2^{\text{anillo}}$ ) ppm.

IEN EM HR:  $[M+H]^+$  calculado: 297,18088,  $[M+H]^+$  encontrado: 297,18083;  $[M+Na]^+$  calculado: 319,16283,  $[M+Na]^+$  encontrado: 319,16282;  $[M+K]^+$  calculado: 335,13677,  $[M+K]^+$  encontrado: 335,13679.

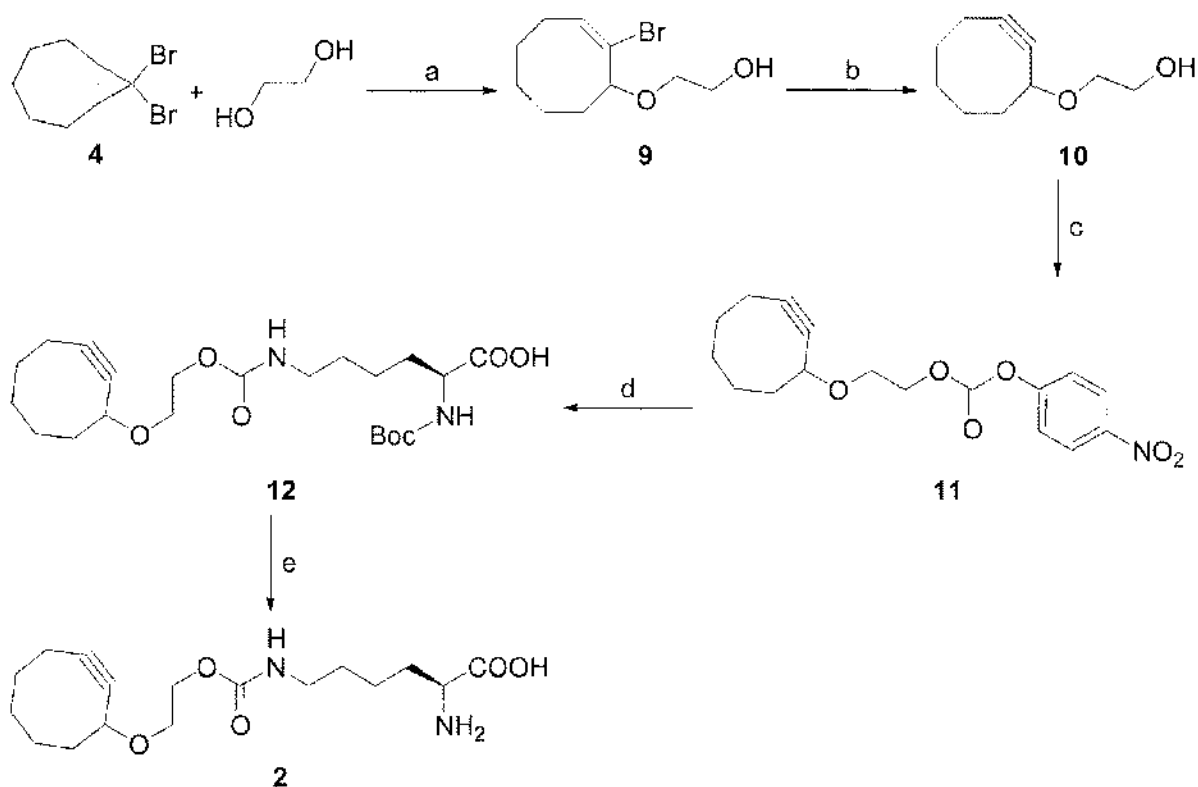
En un procedimiento alternativo para la síntesis del compuesto **1**, el compuesto **8** se disolvió en ácido fórmico al 60 % en  $CHCl_3$ . A parte de eso, la reacción se realizó como se ha descrito anteriormente.

### Ejemplo 2: *N*-ε-((2-(Ciclooct-2-in-1-iloxi)etoxi)carbonil)-L-lisina

La *N*-ε-((2-(ciclooct-2-in-1-iloxi)etoxi)carbonil)-L-lisina (**2**) puede prepararse como se indica en el esquema 2.

10

Esquema 2: Síntesis de derivado de ciclooctina lisina **2**. Reactivos y condiciones: a)  $AgClO_4$ , acetona, TA, negro, 49 %; b) DBU, DMSO, 60 °C, 74%; c) cloroforniato de 4-nitrofenilo, TEA, THF, de -10 °C a TA, 65%; d) Boc-Lys-OH, TEA, DMF, de 0 °C a TA, 79%; e) ácido fórmico,  $CHCl_3$ , TA, 94 %.



15

El Compuesto **4**, 8,8-dibromobicyclo[5,1,0]octano, se sintetizó partiendo de *cis*-ciclohepteno disponible en el mercado según se indica por Neef y Schultz (Angew Chem Int Ed Engl 2009, 48:1498).

#### 20 a) 2-(Bromociclooct-2-en-1-iloxi)etanol (**9**)

El Compuesto **4** (3,12 g, 11,6 mmol) y etano-1,2-diol anhidro (13,0 ml, 23,3 mmol, 20,0 equiv.) se disolvieron en acetona anhidra (0,6 M, 19 ml). Se añadió en pequeñas porciones  $AgClO_4$  anhidro (7,24 g, 34,9 mmol, 3,0 equiv.) con exclusión de la luz y se agitó a TA durante 1 h. Después de la adición de EtOAc (100 ml) y de filtración, se añadió HCl 1 M (100 ml) y la capa acuosa se extrajo con EtOAc (3x 50 ml). Las capas orgánicas combinadas se lavaron con HCl 1 M/ $H_2O$ /solución saturada de NaCl (cada uno 100 ml) y se secaron sobre  $Na_2SO_4$ . El disolvente se evaporó a presión reducida y el compuesto **9** (1,42 g, 5,86 mmol, 49 %) se obtuvo en forma de un aceite de color amarillo y se usó sin purificación adicional. RMN  $^1H$  ( $CDCl_3$ )  $\delta$  = 6,20 (dd,  $^3J$  = 11,73,  $^3J$  = 4,09, 1H,  $CH^{vinilo}$ ), 3,91 (dd,  $^3J$  = 10,22,  $^3J$  = 5,09, 1H,  $CH^{alilo}$ ), 3,78 (t,  $^3J$  = 4,51, 2H,  $CH_2^{etilo}$ ), 3,61-3,66 (m, 1H,  $CHH^{etilo}$ ), 3,44-3,49 (m, 1H,  $CHH^{etilo}$ ), 2,27-2,34 (m, 2H,  $CH_2^{anillo}$ ), 1,84-2,06 (m, 2H,  $CH_2^{anillo}$ ), 1,67-1,76 (m, 2H,  $CH_2^{anillo}$ ), 1,43-1,55 (m, 2H,  $CH_2^{anillo}$ ), 1,23-1,34 (m, 2H,  $CH_2^{anillo}$ ) ppm.

25

30

RMN  $^{13}C$  ( $CDCl_3$ )  $\delta$  = 132,8 (CBr), 131,7 ( $CH^{vinilo}$ ), 85,0 ( $CH^{alilo}$ ), 69,9, 61,9 (2x  $CH_2^{etilo}$ ), 39,6, 36,5, 33,3, 28,1, 26,3 (5x  $CH_2^{anillo}$ ) ppm.

IEN EM HR:  $[M+Na]^+$  calculado: 271,03041,  $[M+Na]^+$  encontrado: 271,03046;  $[M+K]^+$  calculado: 287,00435,  $[M+K]^+$  encontrado: 287,00443.

35

**b) 2-(Ciclooct-2-in-1-iloxi)etanol (10)**

El Compuesto 9 (3,76 g, 15,1 mmol) se disolvió en DMSO (0,5 M, 30 ml) y se calentó a 60 °C. Se añadió DBU (4,51 ml, 30,2 mmol, 2,0 equiv.), la solución resultante se agitó durante 15 min y se añadió más DBU (18,0 ml, 121 mmol, 8,0 equiv.). La mezcla se agitó a 60 °C durante una noche y después se enfrió a TA. Se añadieron EtOAc (100 ml) y agua (100 ml). Después de acidificación a pH 1 con HCl concentrado, la fase acuosa se extrajo con EtOAc (3x 50 ml). Las capas orgánicas combinadas se lavaron con HCl 1 M/solución saturada de NaCl (cada uno 100 ml), se secaron sobre Na<sub>2</sub>SO<sub>4</sub> y se evaporaron a presión reducida. La FC (cHex/EtOAc 9:1) proporcionó el compuesto **10** (1,89 g, 11,2 mmol, 74 %) en forma de un aceite de color amarillo claro.

F<sub>r</sub> (cHex/EtOAc 4:1) = 0,24.

RMN <sup>1</sup>H (CDCl<sub>3</sub>) δ = 4,20-4,24 (m, 1H, CH<sup>propargilo</sup>), 3,72-3,77 (m, 2H, CH<sub>2</sub><sup>etilo</sup>), 3,65-3,71 (m, 1H, CHH<sup>etilo</sup>), 3,44-3,50 (m, 1H, CHH<sup>etilo</sup>), 2,10-2,31 (m, 3H, CH<sub>2</sub><sup>anillo</sup>), 1,91-2,03 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,78-1,89 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,58-1,74 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,42-1,51 (m, 1H, CH<sub>2</sub><sup>anillo</sup>) ppm.

RMN <sup>13</sup>C (CDCl<sub>3</sub>) δ = 100,5, 92,5 (2x C<sup>anillo</sup>), 72,8 (CH<sup>propargilo</sup>), 70,4, 61,9 (2x CH<sub>2</sub><sup>etilo</sup>), 42,3, 34,3, 29,8, 26,3, 20,7 (5x CH<sub>2</sub><sup>anillo</sup>) ppm.

**c) Carbonato de 2-(ciclooct-2-in-1-iloxi)etil 4-nitrofenilo (11)**

El Compuesto **10** (2,64 g, 15,7 mmol) y TEA (2,63 ml, 18,8 mmol, 1,2 equiv.) disueltos en THF (0,2 M, 79 ml) se añadieron gota a gota a una solución agitada de clorofornio de 4-nitrofenilo (9,49 g, 47,1 mmol, 3,0 equiv.) y THF (0,7 M, 22 ml) durante un periodo de 1 h a -10 °C. La mezcla de reacción se agitó durante una noche a TA y se diluyó con cHex (100 ml). Después, se añadió cHex (100 ml) y el THF se retiró a presión reducida. La filtración de la mezcla de reacción seguido de FC (cHex:EtOAc 4:1) dieron el compuesto **11** en forma de un aceite de color amarillo (3,42 g, 10,3 mmol, 65 %).

F<sub>r</sub> (cHex/EtOAc 4:1) = 0,44.

RMN <sup>1</sup>H (CDCl<sub>3</sub>) δ = 8,28 (dt, 3J = 9,18, <sup>3</sup>J = 2,12, 2H, CH<sup>aromático</sup>), 7,39 (dt, <sup>3</sup>J = 9,18, <sup>3</sup>J = 2,12, 2H, CH<sup>aromático</sup>), 4,41-4,47 (m, 2H, CH<sub>2</sub><sup>etilo</sup>), 4,25-4,30 (m, 1H, CH<sup>propargilo</sup>), 3,86-3,92 (m, 1H, CHH<sup>etilo</sup>), 3,63-3,69 (m, 1H, CHH<sup>etilo</sup>), 2,11-2,32 (m, 3H, CH<sub>2</sub><sup>anillo</sup>), 1,91-2,05 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,79-1,89 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,62-1,73 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,42-1,52 (m, 1H, CH<sub>2</sub><sup>anillo</sup>) ppm.

RMN <sup>13</sup>C (CDCl<sub>3</sub>) δ = 156,5 (C(O)O<sub>2</sub>), 155,6, 152,5 (2x C<sup>aromático</sup>), 125,3, 122,3 (2x CH<sup>aromático</sup>), 101,0, 92,1 (2x C<sup>anillo</sup>), 73,0 (CH<sup>propargilo</sup>), 68,5, 66,3 (2x CH<sub>2</sub><sup>etilo</sup>), 42,3, 34,3, 29,7, 26,3, 20,7 (5x CH<sub>2</sub><sup>anillo</sup>) ppm.

**d) N-α-terc-Butiloxycarbonil-N-ε-((2-(ciclooct-2-in-1-iloxi)etoxi)carbonil)-L-lisina (12)**

El Compuesto **11** (0,60 g, 1,81 mmol) se disolvió en DMF (0,2 M, 9 ml) y se añadió gota a gota una solución agitada de Boc-L-Lys-OH (0,58 g, 2,35 mmol, 1,3 equiv.) y TEA (0,76 ml, 5,42 mmol, 3,0 equiv.) en DMF (0,5 M, 4 ml) durante un periodo de 1 h a 0 °C. La mezcla de reacción se agitó a TA durante una noche. Después de la retirada de todos los componentes volátiles por evaporación a presión reducida, el residuo se recogió en H<sub>2</sub>O (100 ml) y EtOAc (100 ml). La fase acuosa se acidificó con HCl concentrado y se extrajo con EtOAc (3 x 50 ml). Las capas orgánicas combinadas se lavaron con una solución saturada de NaCl y se secaron sobre Na<sub>2</sub>SO<sub>4</sub>. El disolvente se evaporó a presión reducida y el producto en bruto se purificó por FC (DCM/MeOH/AcOH 97:2:1) para producir el compuesto **12** en forma de un aceite muy viscoso de color amarillo (0,631 g, 1,43 mmol, 79 %).

F<sub>r</sub> (DCM/MeOH/AcOH 97:2:1) = 0,41.

RMN <sup>1</sup>H (CDCl<sub>3</sub>) δ = 5,18-5,23 (m, 1H, NH), 4,79-4,85 (m, 1H, NH), 4,16-4,31 (m, 4H, CH<sup>propargilo</sup>, α-CH<sup>Lys</sup>, CH<sub>2</sub><sup>etilo</sup>), 3,74-3,81 (m, 1H, CHH<sup>etilo</sup>), 3,52-3,60 (m, 1H, CHH<sup>etilo</sup>), 3,15-3,26 (m, 2H, ε-CH<sub>2</sub><sup>Lys</sup>), 2,11-2,31 (m, 3H, CH<sub>2</sub><sup>anillo</sup>), 1,51-2,04 (m, 13H, CH<sub>2</sub><sup>anillo</sup>, CH<sub>2</sub><sup>Lys</sup>), 1,45 (s, 9H, Boc) ppm.

RMN <sup>13</sup>C (CDCl<sub>3</sub>) δ = 175,8 (C(O)O<sub>2</sub>), 156,7 (C<sup>Lys</sup>), 155,9 (C<sup>Boc</sup>), 100,6, 92,4 (2x C<sup>anillo</sup>), 80,2 (C(CH<sub>3</sub>)<sub>3</sub><sup>Boc</sup>), 72,7 (CH<sup>propargilo</sup>), 67,4, 63,9 (2x CH<sub>2</sub><sup>etilo</sup>), 53,2 (α-CH<sup>Lys</sup>), 42,2 (CH<sub>2</sub><sup>anillo</sup>), 40,5 (ε-CH<sub>2</sub><sup>Lys</sup>), 34,3 (CH<sub>2</sub><sup>anillo</sup>), 31,8 (CH<sub>2</sub><sup>Lys</sup>), 29,8 (CH<sub>2</sub><sup>anillo</sup>), 29,3 (CH<sub>2</sub><sup>Lys</sup>), 28,3 (3x CH<sub>3</sub><sup>Boc</sup>), 26,4 (CH<sub>2</sub><sup>anillo</sup>), 22,3 (CH<sub>2</sub><sup>Lys</sup>), 20,7 (CH<sub>2</sub><sup>anillo</sup>) ppm.

IEN EM HR: [M+H]<sup>+</sup> calculado: 441,25953, [M+H]<sup>+</sup> encontrado: 441,25982; [M+Na]<sup>+</sup> calculado: 463,24147, [M+Na]<sup>+</sup> encontrado: 463,24172; [M+K]<sup>+</sup> calculado: 479,21541, [M+K]<sup>+</sup> encontrado: 479,21570.

**e) N-ε-((2-(Ciclooct-2-in-1-iloxi)etoxi)carbonil)-L-lisina (2)**

El Compuesto **12** (2,08 g, 4,71 mmol) se disolvió en ácido fórmico al 70 % en CHCl<sub>3</sub> (0,2 M, 24 ml) y se agitó durante 36 h a TA. Se añadió DMF (0,2 M, 24 ml) y todos los componentes volátiles se retiraron a presión reducida. El material residual se recogió en HCl 0,1 M (100 ml) y se liofilizó, proporcionando la sal pura de HCl de **2** en forma de un sólido de color amarillo (1,50 g, 4,42 mmol, 94 %).

RMN <sup>1</sup>H (DMSO-d<sub>6</sub>) δ = 7,34-7,65 (m, 2H, α-NH<sub>2</sub>), 7,14-7,23 (m, 1H, ε-NH), 4,17-4,25 (m, 1H, CH<sup>propargilo</sup>), 3,95-4,07 (m, 2H, CH<sub>2</sub><sup>etilo</sup>), 3,49-3,60 (m, 1H, CHH<sup>etilo</sup>), 3,38-3,43 (m, 1H, CHH<sup>etilo</sup>), 3,05 (t, <sup>3</sup>J = 6,03, 1H, α-CH<sup>Lys</sup>), 2,92 (c, <sup>3</sup>J = 6,23, 2H, ε-CH<sub>2</sub><sup>Lys</sup>), 1,98-2,24 (m, 2H, CH<sub>2</sub><sup>anillo</sup>), 1,59-1,86 (m, 5H, CH<sub>2</sub><sup>anillo</sup>), 1,44-1,57 (m, 3H, CH<sub>2</sub><sup>anillo</sup>, CH<sub>2</sub><sup>Lys</sup>), 1,22-1,39 (m, 6H, CH<sub>2</sub><sup>anillo</sup>, CH<sub>2</sub><sup>Lys</sup>) ppm.

RMN <sup>13</sup>C (DMSO-d<sub>6</sub>) δ = 171,2 (C(O)O<sub>2</sub>), 156,8 (C<sup>Lys</sup>), 100,4, 93,4 (2x C<sup>anillo</sup>), 72,3 (CH<sup>propargilo</sup>), 67,5, 65,9 (2x CH<sub>2</sub><sup>etilo</sup>), 53,6 (α-CH<sup>Lys</sup>), 42,3 (CH<sub>2</sub><sup>anillo</sup>), 40,5 (ε-CH<sub>2</sub><sup>Lys</sup>), 34,4 (CH<sub>2</sub><sup>anillo</sup>), 30,7 (CH<sub>2</sub><sup>Lys</sup>), 29,8 (CH<sub>2</sub><sup>Lys</sup>), 29,6 (CH<sub>2</sub><sup>anillo</sup>), 26,4 (CH<sub>2</sub><sup>anillo</sup>), 22,5 (CH<sub>2</sub><sup>Lys</sup>), 20,5 (CH<sub>2</sub><sup>anillo</sup>) ppm.

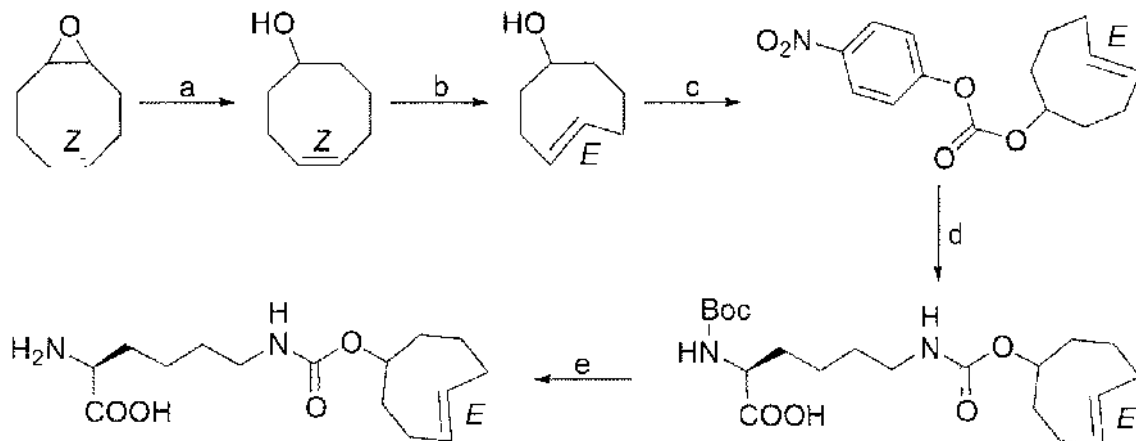
IEN EM HR: [M+H]<sup>+</sup> calculado: 341,20710, [M+H]<sup>+</sup> encontrado: 341,20716; [M+Na]<sup>+</sup> calculado: 363,18917, [M+Na]<sup>+</sup>

encontrado: 363,18904.

En un procedimiento alternativo para la síntesis del compuesto **2**, el compuesto **12** se disolvió en ácido fórmico al 60 % en  $\text{CHCl}_3$ . A parte de eso, la reacción se realizó como se ha descrito anteriormente.

### 5 Ejemplo 3: Síntesis de compuestos de la invención que comprenden grupos análogos de *trans*-ciclooctenilo

Esquema de síntesis 3:



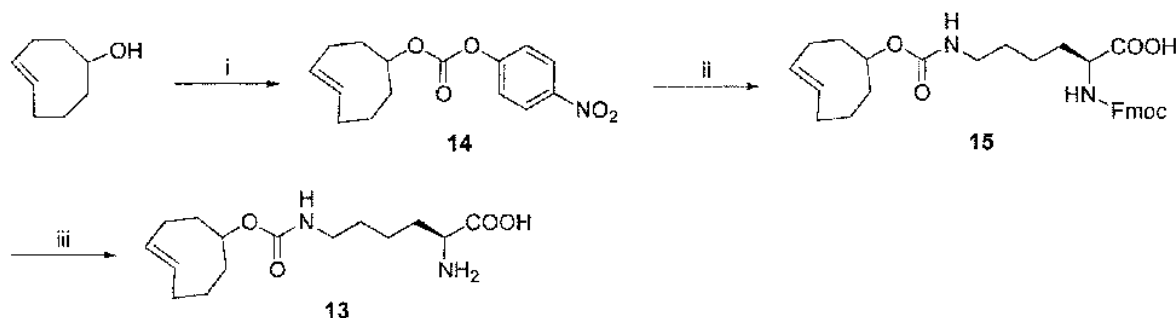
10 Condiciones de reacción:

- 15 a  $\text{LiAlH}_4$ , THF, de 0 °C a TA, durante una noche;  
 b benzoato de metilo, éter dietílico, ciclohexano, irradiación a 254 nm, TA, 6 h  
 c cloroformiato de 4-nitrofenilo,  $\text{NEt}_3$  (TEA), THF, de -10 °C a TA, durante una noche;  
 d Boc-1-Lys-OH,  $\text{NEt}_3$  (TEA), DMF, de -10 °C a TA, durante una noche;  
 e ácido fórmico al 70 % en  $\text{CHCl}_3$ , TA, una noche.

20 En la etapa (e), puede usarse ácido fórmico al 60 % en lugar de ácido fórmico al 70 %.

Pueden encontrarse información específica sobre la síntesis de derivados de cicloocteno en Royzen et al. (J Am Soc 2008, 130:3760) y Hillmyer et al. (Macromol 1995, 28:6311).

25 Esquema de síntesis 4: Síntesis de derivado de *trans*-cicloocteno lisina **13**. Reactivos y condiciones: i) cloroformiato de 4-nitrofenilo, DIEA, THF, de 0 °C a TA, 73%; ii) Fmoc-L-Lys-OH, DIEA, DMSO, TA, 85%; iii) piperidina al 20 % en DMF, TA, 80 %. El rendimiento total después de cinco etapas de síntesis, partiendo de (Z)-9-oxabicyclo[6,1,0]non-4-eno fue 37 % y el rendimiento promedio por etapa fue 83 %.



30 i) Síntesis del compuesto **14**:

35 Se disolvieron *trans*-ciclooct-4-enol (1,00 g, 7,92 mmol, 1,0 equiv.) y DIEA (3,07 g, 4,14 ml, 23,8 mmol, 3,0 equiv.) en THF seco (0,3 M, 26 ml). La solución transparente resultante se añadió gota a gota a 0 °C y en una atmósfera de argón a una solución transparente de cloroformiato de 4-nitrofenilo (4,79 g, 23,8 mmol, 3,0 equiv.) en THF seco (0,3 M, 26 ml) durante un periodo de 2 h. La mezcla de reacción se dejó calentar hasta TA y se agitó durante una noche. Se añadió EtOAc (100 ml) y los componentes insolubles se retiraron por filtración sobre diatomita (Celite). Después, el filtrado se lavó con  $\text{H}_2\text{O}$  (50 ml), HCl 1,0 M (50 ml) y una solución saturada de NaCl (50 ml). Posteriormente, la fase orgánica restante se secó sobre  $\text{Na}_2\text{SO}_4$  y se concentró a presión reducida. El producto en bruto se purificó  
 40 sobre gel de sílice por cromatografía ultrarrápida (gel de sílice 60 de Macherey-Nagel, 0,04-0,063 mm, malla 230-

400; cHex: EtOAc 19:1 v/v). El Compuesto **14** (1,678 g, 5,78 mmol, 73 %) se obtuvo en forma de un sólido de color blanco.

Isómero principal:

5  $F_r$  (cHex: EtOAc 4:1 v/v) = 0,80.  
 RMN  $^1\text{H}$  ( $\text{CDCl}_3$ )  $\delta$  = 8,29-8,24 (m, 2H), 7,39-7,34 (m, 2H), 5,67-5,57 (m, 1H), 5,55-5,45 (m, 1H), 4,48-4,42 (m, 1H), 2,47-2,33 (m, 3H), 2,22-2,07 (m, 2H), 2,05-1,85 (m, 3H), 1,81-1,67 (m, 2H) ppm.  
 10 RMN  $^1\text{H}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 8,31-8,26 (m, 2H), 7,55-7,50 (m, 2H), 5,66-5,57 (m, 1H), 5,51-5,41 (m, 1H), 4,38-4,32 (m, 1H), 2,36-2,24 (m, 3H), 2,13-2,02 (m, 2H), 1,96-1,86 (m, 2H), 1,85-1,76 (m, 1H), 1,73-1,58 (m, 2H) ppm.  
 RMN  $^{13}\text{C}$  ( $\text{CDCl}_3$ )  $\delta$  = 155,7, 152,0, 145,3, 134,9, 133,0, 125,3, 121,8, 86,4, 40,7, 38,3, 34,1, 32,4, 31,1 ppm.  
 RMN  $^{13}\text{C}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 155,9, 151,9, 145,5, 135,4, 133,1, 125,8, 123,1, 86,2, 40,5, 38,0, 34,0, 32,4, 31,1 ppm.

Isómero menor:

15 RMN  $^1\text{H}$  ( $\text{CDCl}_3$ )  $\delta$  = 8,33-8,27 (m, 2H), 7,42-7,38 (m, 2H), 5,69-5,54 (m, 2H), 5,03-4,97 (m, 1H), 2,50-2,27 (m, 4H), 2,23-2,16 (m, 1H). 1,96-1,85 (m, 2H), 1,82-1,71 (m, 1H), 1,67-1,58 (m, 1H), 1,39-1,30 (m, 1H) ppm.  
 RMN  $^{13}\text{C}$  ( $\text{CDCl}_3$ )  $\delta$  = 155,7, 152,0, 145,3, 135,5, 131,5, 125,3, 121,9, 76,0, 40,6, 34,1, 32,1, 29,8, 28,0 ppm.  
 20 EM HR (FAB+)  $m/z$ : calculado para  $\text{C}_{15}\text{H}_{18}\text{NO}_5$   $[\text{M}+\text{H}]^+$ : 292,1185, medido: 292,1151.

ii) Síntesis del compuesto **15**:

Se suspendió Fmoc-L-Lys-OH (0,69 g, 1,87 mmol, 1,2 equiv.) en DIEA (0,24 g, 0,33 ml, 1,87 mmol, 1,2 equiv.) y DMSO anhidro (0,2 M, 8 ml) en una atmósfera de argón. A esta suspensión de color blanco, se añadió gota a gota una solución transparente del compuesto **14** (0,45 g, 1,56 mmol, 1,0 equiv.) en DMSO anhidro (0,2 M, 8 ml) a TA en una atmósfera de argón durante un periodo de 2 h. La mezcla de reacción se agitó durante 4 h más a TA. Se añadieron  $\text{H}_2\text{O}$  (50 ml) y EtOAc (150 ml) y el pH de la capa acuosa se ajustó a 1-3 con HCl concentrado. Las fases se separaron y la capa acuosa se extrajo con EtOAc (2x 50 ml). Las capas orgánicas combinadas se lavaron con una solución saturada de NaCl (2x 50 ml) y se secaron sobre  $\text{Na}_2\text{SO}_4$ . Todos los componentes volátiles se evaporaron a presión reducida y el producto en bruto se purificó por cromatografía ultrarrápida (gel de sílice 60 de Macherey-Nagel, 0,04-0,063 mm, malla 230-400; DCM: MeOH 95:5 v/v) para producir el compuesto **15** (0,69 g, 1,32 mmol, 85 %) en forma de un sólido de color blanco. Para evitar que las condiciones ácidas puedan conducir a una isomerización del doble enlace a partir de una conformación *trans* a *cis*, no se usó AcOH para purificación.

Isómero principal:

35  $F_r$  (DCM: MeOH: AcOH 96:2:2 v/v/v) = 0,16.  
 RMN  $^1\text{H}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 7,87 (d,  $^3J(\text{H,H}) = 7,4$  Hz, 2H), 7,70 (d,  $^3J(\text{H,H}) = 7,3$  Hz, 2H), 7,40 (t,  $^3J(\text{H,H}) = 7,4$  Hz, 2H), 7,31 (t,  $^3J(\text{H,H}) = 7,4$  Hz, 2H), 6,92 (t,  $^3J(\text{H,H}) = 4,4$  Hz, 1H), 5,58-5,47 (m, 1H), 5,44-5,34 (m, 1H), 4,33-4,12 (m, 4H), 3,86-3,73 (m, 1H), 2,96-2,82 (m, 2H), 2,28-2,15 (m, 3H), 1,90-1,76 (m, 4H), 1,69-1,42 (m, 5H), 1,37-1,18 ppm (m, 4H).  
 40 RMN  $^{13}\text{C}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 156,4, 156,2, 144,4, 144,3, 141,2, 135,4, 133,0, 128,1, 127,5, 125,8, 120,6, 79,3, 65,9, 47,2, 41,2, 40,9, 40,5, 38,7, 34,2, 32,6, 31,4, 31,0, 29,7, 23,3 ppm.

iii) Síntesis del compuesto **13**:

El Compuesto **15** (0,30 g, 0,58 mmol, 1,0 equiv.) se disolvió en piperidina al 20 % en DMF (50 mM, 12 ml) y se agitó durante 1 h a TA. Todos los componentes volátiles se retiraron a presión reducida. El producto en bruto se purificó por cromatografía ultrarrápida (gel de sílice 60 Macherey-Nagel, 0,04-0,063 mm, malla 230-400; acetona: MeOH:  $\text{H}_2\text{O}$  65:25:10 v/v/v) sobre gel de sílice para producir el compuesto **13** (0,14 g, 0,46 mmol, 80 %) en forma de un sólido de color blanco.

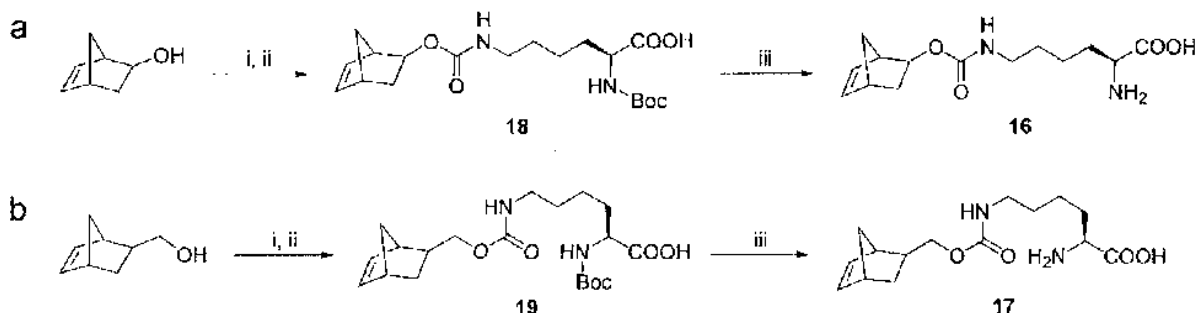
Isómero principal:

55  $F_r$  (acetona:MeOH:  $\text{H}_2\text{O}$  65:25:10 v/v/v) = 0,53.  
 RMN  $^1\text{H}$  ( $\text{CD}_3\text{OD}$ )  $\delta$  = 5,65-5,54 (m, 1H), 5,51-5,41 (m, 1H), 4,38-4,21 (m, 1H), 3,52 (dd,  $^3J(\text{H,H}) = 7,0$ , 5,2 Hz, 1H), 3,11-3,02 (m, 2H), 2,37-2,25 (m, 3H), 2,01-1,65 (m, 8H), 1,63-1,37 (m, 5H) ppm.  
 60 RMN  $^{13}\text{C}$  ( $\text{CD}_3\text{OD}$ )  $\delta$  = 173,1, 157,4, 134,7, 132,3, 80,2, 54,7, 40,8, 39,9, 38,2, 33,8, 32,1, 30,7, 30,6, 29,2, 22,1 ppm. EM HR (IEN)  $m/z$ : calculado para  $\text{C}_{15}\text{H}_{27}\text{N}_2\text{O}_4$   $[\text{M}+\text{H}]^+$ : 299,19653, medido: 299,19656.

#### Ejemplo de referencia 4: Síntesis de compuestos que comprenden grupos norbornenilo

Esquema de síntesis 5: Síntesis de derivados de norborneno lisina **16** (a) y **17** (b). Reactivos y condiciones: (a) i) fosgeno, THF, tolueno, de 0 °C a TA; ii) Boc-L-Lys-OH, THF, NaOH, 97 % (después de dos etapas); iii) ácido fórmico,  $\text{CHCl}_3$ , TA, 94%. (b) i) trifosgeno, THF, de 0 °C a TA; ii) Boc-L-Lys-OH, THF, NaOH, de 0 °C a TA, 96% (después de dos etapas); iii) ácido fórmico,  $\text{CHCl}_3$ , TA, 95%. Todo el material de partida de norborneno se usó según

se suministró en forma de una mezcla de isómeros *endo* y *exo*. No se hicieron intentos para separar los isómeros *endo* y *exo* en ningún punto de la síntesis. La proporción *exo/endo* se determinó por RMN  $^1\text{H}$  y se indica en los protocolos.



5

(a) i), ii) Síntesis del compuesto **18**:

Se añadió gota a gota 5-norbornen-2-ol (5,00 g, 45,4 mmol, 1,0 equiv.) en tolueno seco/THF (1:1 v/v, 1,0 M, 45 ml) a 0 °C y en una atmósfera de argón a una solución al 20 % de fosgeno en tolueno (8,98 g, 90,8 mmol, 2,0 equiv.; 47,8 ml de una solución al 20 % de fosgeno en tolueno) durante un periodo de 1 h. La mezcla de reacción se dejó calentar hasta TA y se agitó durante 3 h más. Posteriormente, todos los componentes volátiles se retiraron a presión reducida y el residuo se secó durante 30 min a alto vacío y se usó directamente en la siguiente etapa. El residuo de color pardo se recogió en THF seco (3,5 M, 13 ml) y se añadió gota a gota a 0 °C a una solución de Boc-L-Lys-OH (14,5 g, 59,0 mmol, 1,3 equiv.) en NaOH 1,0 M/THF (2:1 v/v, 0,3 M, 151 ml). Después de la adición, la mezcla de reacción se dejó calentar hasta TA y se agitó durante 12 h más. Se añadió EtOAc (100 ml) y la capa acuosa se acidificó a pH < 4 con HCl concentrado. Las fases se separaron y la capa acuosa se extrajo con EtOAc (3x 70 ml). Las capas orgánicas combinadas se lavaron con una solución saturada de NaCl (80 ml) y se secó sobre  $\text{Na}_2\text{SO}_4$ . Todos los componentes volátiles se evaporaron a presión reducida. El producto en bruto se purificó por cromatografía ultrarrápida (gel de sílice 60 de Macherey-Nagel, 0,04-0,063 mm, malla 230-400; DCM: MeOH: AcOH 96:3:1 v/v/v; coevaporación con tolueno) para producir el compuesto **18** (16,9 g, 44,3 mmol, 97 %) en forma de un sólido de color pardo en una proporción *exo/endo* de 3:7 (determinada por RMN  $^1\text{H}$ ).

$F_r$  (DCM: MeOH: AcOH 90:8:2 v/v/v) = 0,45.  
 RMN  $^1\text{H}$  ( $\text{CDCl}_3$ )  $\delta$  = 6,34-6,31 (m, 0,7H), 6,25-6,20 (m, 0,3H), 5,99-5,94 (m, 1H), 5,29-5,21 (m, 2H), 4,37-4,24 (m, 1H), 3,22-2,99 (m, 3H), 2,89-2,80 (m, 1H), 2,15-2,08 (m, 0,7H), 1,91-1,66 (m, 3,3H), 1,61-1,37 (m, 14,3H), 1,33-1,23 (m, 1H), 0,99-0,88 ppm (m, 0,7H).

EM HR (IEN)  $m/z$ : calculado para  $\text{C}_{23}\text{H}_{42}\text{N}_3\text{O}_6$  [ $\text{M}+\text{C}_4\text{H}_{11}\text{N}+\text{H}$ ] $^+$ : 456,30681, medido: 456,30678.

(a) iii) Síntesis del compuesto **16**:

30

El Compuesto **18** (11,6 g, 30,4 mmol, 1,0 equiv.) se disolvió en ácido fórmico al 60 % en  $\text{CHCl}_3$  (6:4 v/v, 0,2 M, 152 ml) y se agitó durante 24 h a TA. Se añadió DMF (0,2 M, 152 ml) y todos los componentes volátiles se retiraron a presión reducida. El residuo se recogió en HCl 50 mM y se liofilizó, proporcionando la sal pura de HCl del compuesto **1** (9,14 g, 28,7 mmol, 94 %) en forma de un sólido de color amarillo en una proporción *exo/endo* de 3:7 (determinada por RMN  $^1\text{H}$ ).

35

$F_r$  (DCM: MeOH: AcOH 87:10:3 v/v/v) = 0,04.

RMN  $^1\text{H}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 8,47-8,37 (m, 2H), 7,12-7,04 (m, 0,3H), 6,96-6,89 (m, 0,7H), 6,29 (dd,  $^3J(\text{H,H})$  = 5,4, 2,9 Hz, 0,7H), 6,23 (dd,  $^3J(\text{H,H})$  = 5,5, 2,7 Hz, 0,3H), 5,97 (dd,  $^3J(\text{H,H})$  = 5,5, 3,2 Hz, 0,3H), 5,91 (dd,  $^3J(\text{H,H})$  = 5,4, 2,6 Hz, 0,7H), 5,12-5,01 (m, 0,7H), 4,72-4,63 (m, 0,3H), 3,84-3,76 (m, 1H), 3,04-2,98 (m, 0,7H), 2,94-2,86 (m, 2H), 2,81-2,71 (m, 1,3H), 2,29-2,25 (m, 0,3H), 2,05-1,98 (m, 0,7H), 1,79-1,24 (m, 8H), 0,91-0,71 ppm (m, 1H).

40

RMN  $^{13}\text{C}$  ( $\text{DMSO-d}_6$ )  $\delta$  = 171,4, 156,7, 156,6\*, 141,3\*, 138,6, 133,2\*, 132,3, 74,4, 72,7\*, 52,2, 47,6, 47,5\*, 46,3\*, 46,0, 42,2, 40,5\*, 34,7, 34,5, 33,7\*, 30,0, 29,3, 22,0 ppm (\*=señales que pertenecen a isómero *exo*).

EM HR (IEN)  $m/z$ : calculado para  $\text{C}_{14}\text{H}_{23}\text{N}_2\text{O}_4$  [ $\text{M}+\text{H}$ ] $^+$ : 283,16523, medido: 283,16517.

(b) i), ii) Síntesis del compuesto **19**:

45

Se añadió gota a gota 5-norborneno-2-metanol (4,17 g, 33,6 mmol, 1,0 equiv.) a 0 °C y en una atmósfera de argón a una solución de trifosgeno (9,96 g, 33,6 mmol, 1,0 equiv.) en THF seco (0,5 M, 67 ml) durante un periodo de 2 h y se agitó durante 6 h más a 0 °C. La mezcla de reacción se dejó calentar hasta TA y se filtró. Posteriormente, todos los componentes volátiles se retiraron a presión reducida y el residuo se secó durante 1 h a alto vacío, proporcionando el producto intermedio en forma de un aceite transparente que se usó sin purificación adicional en la siguiente etapa. Este residuo se recogió en THF seco (3,5 M, 10 ml) y se añadió lentamente a 0 °C a una solución de Boc-L-Lys-OH (9,93 g, 40,3 mmol, 1,2 equiv.) en NaOH 1,0 M/THF (2:1 v/v, 0,3 M, 112 ml). Después de la adición, la mezcla de reacción se dejó calentar hasta TA y se agitó durante 14 h más. Se añadió EtOAc (100 ml) y la capa acuosa se acidificó a pH < 4 con HCl concentrado. Las fases se separaron y la capa acuosa se extrajo con EtOAc (3x 70 ml). Las capas orgánicas combinadas se lavaron con una solución saturada de NaCl (100 ml) y se secó sobre  $\text{Na}_2\text{SO}_4$ .

55

Todos los componentes volátiles se evaporaron a presión reducida. El producto en bruto se purificó por cromatografía ultrarrápida (gel de sílice 60 de Macherey-Nagel, 0,04-0,063 mm, malla 230-400; DCM/MeOH/AcOH 96:3:1 v/v/v, coevaporación con tolueno) para producir el compuesto 19 (12,8 g, 32,3 mmol, 96 %) en forma de un sólido de color amarillo en una proporción *exo/endo* de 2:3 (determinada por RMN <sup>1</sup>H).

5 F<sub>r</sub> (DCM: MeOH: AcOH 90:8:2 v/v/v) = 0,44.

RMN <sup>1</sup>H (CDCl<sub>3</sub>) δ = 6,14 (dd, <sup>3</sup>J(H,H) = 4,9, 2,7 Hz, 0,6H), 6,11-6,06 (m, 0,8H), 5,94 (dd, <sup>3</sup>J(H,H) = 5,4, 2,6 Hz, 0,6H), 5,52-5,46 (m, 0,6H), 5,30-5,23 (m, 0,4 H), 4,34-4,24 (m, 1,4H), 4,19-4,07 (m, 1H), 3,99-3,80 (m, 1H), 3,69-3,57 (m, 0,6H), 3,23-3,10 (m, 3H), 2,86 (s, 0,6H), 2,84-2,78 (m, 1H), 2,70 (s, 0,4H), 1,92-1,65 (m, 6H), 1,61-1,10 (m, 12H), 1,19-1,11 (m, 0,4H), 0,58-0,50 (m, 0,6H) ppm.

10 RMN <sup>13</sup>C (DMSO-d<sub>6</sub>) δ = 179,5, 161,5, 161,5, 160,8, 142,5, 142,4, 141,4, 141,4, 83,2, 72,2, 71,6, 58,7, 54,1, 49,8, 48,6, 48,4, 46,9, 46,3, 43,4, 43,1, 38,8, 35,6, 34,2, 33,9, 33,4, 28,1 ppm. EM HR (IEN) m/z: calculado para C<sub>20</sub>H<sub>33</sub>N<sub>2</sub>O<sub>6</sub> [M+H]<sup>+</sup>: 397,23331, medido: 397,23405; calculado para C<sub>20</sub>H<sub>32</sub>N<sub>2</sub>NaO<sub>6</sub> [M+Na]<sup>+</sup>: 419,21526, medido: 419,21607; calculado para C<sub>20</sub>H<sub>32</sub>KN<sub>2</sub>O<sub>6</sub> [M+K]<sup>+</sup>: 435,18920, medido: 435,19006; calculado para C<sub>40</sub>H<sub>65</sub>N<sub>4</sub>O<sub>12</sub> [2M+H]<sup>+</sup>: 793,45935, medido: 793,46014.

15

### (b) iii) Síntesis del compuesto 17:

El Compuesto 19 (1,82 g, 4,60 mmol, 1,0 equiv.) se disolvió en ácido fórmico al 60 % en CHCl<sub>3</sub> (6:4 v/v, 0,2 M, 23 ml) y se agitó durante 24 h a TA. Se añadió DMF (0,2 M, 23 ml) y todos los componentes volátiles se retiraron a presión reducida. El residuo se recogió en HCl 50 mM y se liofilizó, proporcionando la sal pura de HCl del compuesto 17 (1,46 g, 4,37 mmol, 95 %) en forma de un sólido de color blanco en una proporción *exo/endo* de 2:3 (determinada por RMN <sup>1</sup>H).

20

F<sub>r</sub> (DCM: MeOH: AcOH 87:10:3 v/v/v) = 0,04.

25 RMN <sup>1</sup>H (DMSO-d<sub>6</sub>) δ = 7,11 (t, <sup>3</sup>J(H,H) = 5,4 Hz, 0,4H), 7,06 (t, <sup>3</sup>J(H,H) = 5,4 Hz, 0,6H), 6,15 (dd, <sup>3</sup>J(H,H) = 5,6, 3,0 Hz, 0,6H), 6,10-6,05 (m, 0,8H), 5,91 (dd, <sup>3</sup>J(H,H) = 5,6, 2,8 Hz, 0,6H), 4,02-3,95 (m, 0,4H), 3,86-3,79 (m, 0,4H), 3,69-3,62 (m, 0,6H), 3,58-3,40 (m, 2,6H), 2,92 (c, <sup>3</sup>J(H,H) = 5,2 Hz, 2H), 2,81-2,74 (m, 1,6H), 2,64 (s, 0,4H), 2,33-2,24 (m, 0,6H), 1,80-1,53 (m, 3,4H), 1,40-1,26 (m, 3,4H), 1,24-1,18 (m, 1,4H), 1,15-1,11 (m, 0,6H), 0,45 (ddd, <sup>3</sup>J(H,H) = 11,5, 4,1, 2,5 Hz, 0,6H) ppm.

30 RMN <sup>13</sup>C (DMSO-d<sub>6</sub>) δ = 171,3, 156,8, 156,7\*, 137,7, 137,2\*, 136,6\*, 132,6, 68,0\*, 67,5, 53,4, 49,4, 45,1, 43,9, 43,7\*, 42,2, 41,6\*, 38,6\*, 38,4, 30,6, 29,5, 29,1, 28,9\*, 22,4 ppm (\*=señales que pertenecen a isómero *exo*).

EM HR (IEN) m/z: calculado para C<sub>15</sub>H<sub>25</sub>N<sub>2</sub>O<sub>4</sub> [M+H]<sup>+</sup>: 297,18088, medido: 297,18102.

### Ejemplos biológicos

### 35 Ejemplo A: Expresión de GFP que comprende un resto de ciclooctinilo en *E. coli*

#### A.1 Plásmidos y construcciones de ADN

40 Un gen con codones optimizados para *E. coli* para pirrolisil ARNt sintetasa de tipo silvestre (pyIRS<sup>WT</sup>) y el ARNt correspondiente (tRNA<sup>pyl</sup>) de *M. maize* (obtenido de Mr Gene, Regensburg, Alemania) se usó para reemplazar las dos regiones codificantes para ARNt sintetasa de *M. jannaschii* y ARNt en el sistema de plásmido pEVOL descrito en Young et al. (J Mol Biol. 2010; 395:361) para producir el plásmido pEVOL ARNt<sup>pyl</sup>/pyIRS<sup>WT</sup>. Además, se preparó un plásmido pEVOL ARNt<sup>pyl</sup>/pyIRS<sup>AF</sup> que codifica una pirrolisil ARNt sintetasa mutante que comprende las sustituciones de aminoácidos Y306A e Y384F (pyIRS<sup>AF</sup>). Para este mutante doble, se realizaron dos ciclos de mutagénesis dirigida convencional para introducir Y306A e Y384F en el gen con codones optimizados. Con respecto al tipo silvestre (WT), se clonaron después dos copias de este gen en el plásmido pEvolv para generar el plásmido mutante pEVOL ARNt<sup>pyl</sup>/pyIRS<sup>AF</sup>.

45

#### A.2 Expresión y purificación de la proteína

50

Para expresión de la proteína diana GFP<sup>TAG</sup>, se preparó un plásmido pBAD (Invitrogen, Carlsbad, Estados Unidos) que albergaba una GFP marcada con FLAG N-terminal con una secuencia peptídica de 6-His C-terminal, en el que el sitio permisivo 39 contenía el codón de terminación ámbar (TAG). Se cotransformaron células *E. coli* Top10 con el vector de expresión GFP<sup>TAG</sup> y pEVOL ARNt<sup>pyl</sup>/pyIRS<sup>WT</sup> o pEVOL ARNt<sup>pyl</sup>/pyIRS<sup>AF</sup> y se cultivaron a 37°C en presencia de ampicilina y cloranfenicol. Normalmente, se usaron 0,5 ml de un cultivo de una noche para inocular 50 ml de medio de caldo de cultivo Terrific (TB) en un matraz de agitación. Los cultivos se dejaron crecer a 37°C en 1 - 2 horas hasta una DO<sub>600</sub> entre 0,2 y 0,3. Después el compuesto 1 o 2 (solución de reserva: 80 mM en NaOH 0,1 M) o una cantidad igual de NaOH 0,1 M (para experimentos de control) se añadió a una concentración final de 1 mM. Los cultivos se dejaron crecer adicionalmente hasta una DO<sub>600</sub> entre 0,4 y 0,6. Después se indujo expresión añadiendo arabinosa hasta una concentración final de 0,02 % (p/v). Después de 4 - 6 h de agitación a 37 °C, los cultivos se recogieron por centrifugación. Los sedimentos se resuspendieron en una solución salina tamponada con fosfato 4x (PBS 4x pH 8,0) que contenía fluoruro de fenilmetilsulfonilo 1 mM (PMSF) y las células se lisaron por ultrasonidos. El lisado se centrifugó durante 1 h a 14.000 g y el sobrenadante se incubó con -50 µl de perlas de Ni-NTA (Qiagen, Düsseldorf, Alemania). Las perlas se lavaron con imidazol 10 mM en PBS 4x y después se eluyeron con tampón que contenía imidazol 500 mM. Siempre que se mencione, el lavado y/o la elución también se llevaron a cabo en un tampón de PBS 4x desnaturalizante que contenía clorhidrato de guanidinio 6 M (pH 8,0). Enfoques de expresión

65

mayores usados para determinar los rendimientos con más precisión, se aumentaron de escala en consecuencia.

Se observó expresión de proteína GFP<sup>TAG</sup> en presencia del compuesto **1** y **2**, respectivamente, en comparación con un control negativo (Figura 1). La expresión de GFP<sup>TAG</sup> en células transfectadas con el vector de expresión para ARNt<sup>pyl</sup>/pyIRS<sup>AF</sup> fue mayor (producción absoluta de 10 mg de GFP<sup>TAG->1</sup> (GFP<sup>TAG</sup> que comprende el compuesto **1**) por litro de cultivo) que en células transfectadas con el vector de expresión para ARNt<sup>pyl</sup>/pyIRS<sup>WT</sup>. Las bandas del gel teñido con Coomassie mostradas en la Figura 1b se escindieron, se digirieron con tripsina y quimotripsina y los péptidos resultantes se analizaron por espectrometría de masas (espectrómetro de masas Orbitrap, Thermofisher, Estados Unidos; algoritmo Mascot). Este análisis confirmó la incorporación específica de sitio del compuesto **1** o **2**, respectivamente, en GFP<sup>TAG</sup> (Tabla 1).

Tabla 1: Espectrometría de masas de bandas de GFP<sup>TAG->1</sup> y GFP<sup>TAG->2</sup> de gel mostrada en la Figura 1b.

	proteasa	Masa monoisotópica (calc) [Da]	Masa coincidente, hallada [Da]	Δ masa	Puntuación iónica	Secuencia peptídica (X=sitio TAG ámbar)
GPP <sup>TAG-&gt;1</sup>	quimotripsina	1583,76788	1583,76788	(-) 0,00003	76	SVSGEGEGDATXGKL
GFP <sup>TAG-&gt;1</sup>	tripsina	1617,75224	1617,75224	(-) 0,00005	84	FVSGEGEGDATXGK
GFP <sup>TAG-&gt;2</sup>	quimotripsina	1627,79410	1627,79410	(-) 0,00007	31	SVSGEGEGDATXGKL
GFP <sup>TAG-&gt;2</sup>	tripsina	1661,77846	1661,77846	0,00008	39	FVSGEGEGDATXGK

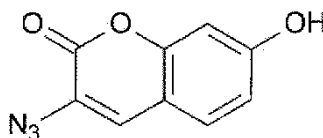
### Ejemplo B: Marcaje de fluorescencia de GFP<sup>TAG->1</sup> dentro de células *E. coli* vivas

15

#### B.1 Marcaje *in vivo* de ciclooctinilo que comprende GFP por azido cumarina fluorogénica

GFP<sup>TAG->1</sup> se expresó en *E. coli* que albergaba ARNt<sup>pyl</sup>/pyIRS<sup>AF</sup> en presencia de compuesto **1** 1 mM. De cuatro a seis horas después de la inducción de la expresión, se recogió una muestra de cultivo de 5 ml. Las células se lavaron dos veces con 12 ml de PBS, se resuspendieron en 12 ml de PBS, se incubaron durante 1 h a 4°C en oscuridad y se lavaron otras dos veces con 12 ml de PBS. Después las células se sedimentaron, se resuspendieron en 3 ml de PBS (DO<sub>600</sub> ~2 - 3) que contenía azido cumarina 50 μM (compuesto **3**; disponible en el mercado de Base Click; puede sintetizarse de acuerdo con Sivakumar et al., *Org Lett* 2004, 6:4603) y se incubaron con agitación a 37°C en oscuridad.

25



3

Se realizó un experimento de control repitiendo el mismo procedimiento de captura de imágenes con células que expresaban GFP<sup>WT</sup>. En esta construcción, la sintetasa y el ARNt aún están activos en el reconocimiento del compuesto **1**, pero GFP<sup>WT</sup> no contiene codón de terminación ámbar que permita la incorporación del compuesto **1**.

30

#### B.2 Análisis de GFP marcada mediante microscopía de fluorescencia

Después de 4 - 6 h de incubación con el compuesto **3**, se recogieron 5 ml de células, se lavaron dos veces con 1,5 ml de PBS, se resuspendieron en PBS, se incubaron durante 1 h a 4°C en oscuridad, se lavaron otras dos veces con 1,5 ml de PBS y después se dejaron reposar en un cubreobjetos. Se montaron células en un microscopio SP5 Leica que empleaba un objetivo de aceite de 1,4 NA (Leica, Mannheim, Alemania). Se adquirieron imágenes que contenían 1024\*1024 píxeles a una velocidad de exploración de 400 Hz y un factor de zoom de dos que produjo un tamaño de píxeles final de 120,1 nm x 120,1 nm. Además de una imagen DIC la muestra se excitó usando un láser de diodo azul que actúa a una longitud de onda de 405 nm, registrando simultáneamente la señal de fluorescencia en dos canales (azul = 415 - 470 nm y verde = 520 - 540 nm). La fluorescencia en el canal azul se originó del compuesto con clic **3**, mientras que la fluorescencia en el canal verde se originó de GFP que puede excitarse directamente también a 405 nm y posiblemente también mediante transferencia de energía de compuesto con clic **3** al cromóforo GFP. Los mismos canales de emisión se registraron durante excitación con un láser iónico de argón que actúa a λ = 488 nm, que solamente excita GFP. Se resumen intensidades de fluorescencia relativas en la Tabla 2. Cuando se excitaron a λ<sub>ex</sub> = 405 nm, las células que expresaban GFP<sup>TAG->1</sup> mostraron fluorescencia en el canal

45



azul así como en el canal verde lo que indica la presencia de compuesto con clic **3** así como del cromóforo GFP. En cambio, en las células de control que expresaban GFP<sup>WT</sup> solamente fue visible fluorescencia de fondo en el canal azul a  $\lambda_{ex} = 405$  nm, es decir no se detectó compuesto **3** en estas células. La fluorescencia de GFP observada en el canal verde fue más fuerte en las células de control que en células que expresaban GFP<sup>TAG->1</sup>, ya que GFP<sup>WT</sup> se expresa de forma natural mejor que el GFP<sup>TAG->1</sup> suprimido por ámbar. En resumen, la fluorescencia en el canal azul en células que expresaban GFP<sup>TAG->1</sup> era aproximadamente de dos a tres veces mayor que el fondo, verificando que se produjo acoplamiento con el compuesto **3** *in vivo*.

Tabla 2: Intensidad de fluorescencia de células *E. coli* que albergan GFP<sup>TAG->1</sup> marcado.

	$\lambda_{em} = 415 - 470$ nm	$\lambda_{em} = 510 - 540$ nm	
$\lambda_{ex} = 405$ nm	+++	+++	GFP <sup>TAG-&gt;1</sup>
$\lambda_{ex} = 488$ nm	-	+	
$\lambda_{ex} = 405$ nm	+	++++	GFP <sup>WT</sup> (control)
$\lambda_{ex} = 488$ nm	-	+++	

Intensidad de fluorescencia: - (sin fluorescencia), + (débil), ++ (moderado), +++ (fuerte), ++++ (muy fuerte)

10

### B.3 Análisis de GFP marcada mediante SDS-PAGE

El acoplamiento del compuesto **3** con ciclooctinilo que comprende GFP también se confirmó analizando la fluorescencia en el lisado celular mediante SDS-PAGE. Después de añadir el compuesto **3** a los cultivos de *E. coli* que expresaban GFP<sup>TAG->1</sup> o GFP<sup>WT</sup>, respectivamente, se tomaron muestras pequeñas después de 0, 15, 30, 60, 120, 180, 360 y 540 minutos, se diluyeron con PBS hasta un volumen total de 1,5 ml, se lavaron dos veces con 1,5 ml de PBS y se cargaron en un gel de poliacrilamida SDS para análisis de lisado celular completo. El gel se analizó con respecto a fluorescencia usando un sistema de documentación de gel disponible en el mercado (Alpha Innotech, CA) a una longitud de onda de excitación de 365 nm y detectando la emisión con un ajuste de filtro de bromuro de etidio (Figura 3a). GFP<sup>TAG->1</sup> marcada con el compuesto **3** ya era visible después de 15 min, lo que confirma que de hecho se produjo marcaje de GFP<sup>TAG->1</sup>.

15

20

### B. 4 Marcaje *in vivo* de ciclooctinilo que comprende mCherry por azido cumarina fluorogénica

Se realizó un experimento similar al de B.1/B.2 con cultivos de *E. coli* que expresaban mCherry, en el que el compuesto **1** se ha incorporado en un sitio codificado por ámbar (mCherry<sup>TAG->1</sup>), o mCherry de tipo silvestre (mCherry<sup>WT</sup>). Se recogieron 5 ml de dichos cultivos después de inducción durante una noche y se lavaron dos veces con 12 ml de PBS, se resuspendieron en 12 ml de PBS, se incubaron durante 1 h a 4°C en oscuridad y se lavaron otras dos veces con 12 ml de PBS. Las células se sedimentaron, se resuspendieron en 3 ml de PBS (DO<sub>600</sub> ~2 - 3) que contenía azido cumarina 50  $\mu$ M y se incubaron con agitación a 37°C en oscuridad. Las células se recogieron después de 3 - 4 h, se lavaron dos veces con 1,5 ml de PBS, se resuspendieron en 1,5 ml de PBS, se incubaron durante 1 h a 4°C en oscuridad, y se lavaron otras dos veces con 1,5 ml de PBS antes de dejar las células reposar en un cubreobjetos. Después se montaron células en un microscopio SP5 Leica que empleaba un objetivo de aceite de 1,4 NA (Leica, Mannheim, Alemania). Se adquirieron imágenes que contenían 512\*512 píxeles a una velocidad de exploración de 400 Hz y un factor de zoom de dos que produjo un tamaño de píxeles final de 160,5 nm x 160,5 nm. Además de una imagen DIC la muestra se excitó usando un láser de diodo azul que actúa a una longitud de onda de 405 nm, registrando simultáneamente la señal de fluorescencia en dos canales (azul/verde = 420 - 520 nm y rojo = 590 - 690 nm). La fluorescencia en el canal azul/verde se originó de cumarina con clic mientras que no fue visible fluorescencia en el canal rojo que se origina de mCherry. Los mismos canales de emisión se registraron también durante excitación con un láser DPSS que actúa a  $\lambda_{ex} = 561$  nm, que solamente excita mCherry. Como un experimento de control se repitió el mismo procedimiento de captura de imágenes con células que expresaban mCherry<sup>WT</sup>. En esta construcción, la sintetasa y el ARNt aún están activos en el reconocimiento del compuesto **1**, pero mCherry<sup>WT</sup> no contiene codón de terminación ámbar que permita la incorporación del compuesto **1**. Por lo tanto, en las células de control que expresaban mCherry<sup>WT</sup> solamente fue visible fluorescencia de fondo en el canal azul/verde, es decir no se detectó compuesto **3** en estas células. La fluorescencia de mCherry observada en el canal rojo fue más fuerte en las células de control que en células que expresaban mCherry<sup>TAG->1</sup>, ya que mCherry<sup>WT</sup> se expresa de forma natural mejor que el mCherry<sup>TAG->1</sup> suprimido por ámbar.

25

30

35

40

45

### Ejemplo C: Análisis de GFP<sup>TAG->1</sup> marcado por fluorescencia mediante TERFum

50

#### C. 1 Marcaje de GF<sup>TAG->1</sup> con colorante fluorescente

Para este fin, GFP<sup>TAG->1</sup> se expresó y purificó como se describe en el ejemplo A. Se incubó una solución 1 mM de GFP<sup>TAG->1</sup> en PBS 4x (pH 8,0) durante 12 h a 37°C con una solución 10 mM de At-to647N azida (Atto-Tec GmbH, Siegen, Alemania). La mezcla se incubó en perlas Ni-NTA y se lavó con tampón desnaturalizante leve (urea 2 M)

55

para retirar cualquier colorante unido de forma no específica de la proteína y después se eluyó en tampón de PBS 4x (pH 8,0). Las eficacias de marcaje típicas fueron de aproximadamente 50 %, como se determinó usando espectrometría de UV/Vis convencional y los coeficientes de extinción indicados para GFP, GFP desnaturalizada (si se mide en condiciones desnaturalizantes) y Atto647N.

5

### C.2 Observación de una única molécula de transferencia de energía por resonancia de fluorescencia (TERFum)

La GFP<sup>TAG->1,Atto647N</sup> resultante (GFP<sup>TAG</sup> marcada con Atto647N azida que comprende el compuesto **1**) se diluyó hasta una concentración de 50 pM y se analizó mediante espectrometría de una única molécula (um) de moléculas de difusión libre de forma similar a esquemas de mediciones previamente indicadas (Lemke et al., J Am Chem Soc 2009, 131:13610). Brevemente, la solución se montó en un microscopio confocal construido a medida centrado alrededor de un microscopio Olympus IX81 (Hamburgo, Alemania) equipado con un objetivo de agua 60x de 1,2 NA. La luz que se emitía de dos diodos de láser (LDH 485 y 660, Picoquant, Berlín, Alemania) se alternó a una frecuencia de pulso maestro de 56 MHz y se centró en la muestra. La luz de fluorescencia en ráfagas que se emitía de una única GFP<sup>TAG->1,Atto647N</sup> de difusión libre se filtró espacialmente usando un orificio de 100 μm y después se filtró espectralmente en canales fluorescentes donantes (D) y aceptores (A) (usando filtros de emisión 525/50, 700/75, dicroicos 500/660 y 560 de AHF, Tuebingen Alemania). Se detectaron fotones individuales usando MPD de Picoquant en el canal verde y APD (PerkinElmer, Vaudreuil Canadá) en el rojo. Las señales se contaron usando un Hydrharp (Picoquant) y se sometieron a análisis de excitación intercalado por pulsos rutinario (Müller et al., Biophys J 2005, 89:3508) después de agrupar la corriente de señal a una amplitud de grupo de 1 ms y aplicar un umbral de 30 recuentos por ráfaga de una única molécula. Por lo tanto, las ráfagas de emisión que surgen de moléculas de GFP<sup>TAG->1,Atto647N</sup> individuales podrían analizarse basándose en su estequiometría (S) y para la aparición de transferencia de energía (E<sub>TERF</sub>) de D a A.

Debido a las propiedades espectrales de las especies fluorescentes, el cromóforo GFP natural actuó como un donante (D), mientras que Atto647N actuó como el colorante aceptor (A). A una resolución de una única molécula usando una geometría de detección confocal (véase anteriormente) se observó GFP<sup>TAG->1,Atto647N</sup> de difusión libre (Figura 5). Se observaron dos poblaciones principales (Figura 4). Una población se centró alrededor de S = 1 y E<sub>TERF</sub> = 0, es decir tenía solamente fluorescencia donante. Esta especie de moléculas se observa casi siempre en experimentos de moléculas individuales (véase Lemke et al., J Am Chem Soc 2009, 131:13610) y se origina de las especies en las que A estaba fotofísicamente inactiva o no estaba presente. La segunda población se centró alrededor de S = 0,5 y E<sub>TERF</sub> = 1, es decir identifica claramente una especie de moléculas de GFP marcadas con Atto647N de modo que la transferencia de energía se produzca eficazmente. La alta eficacia de TERF observada con esta segunda población estuvo bastante de acuerdo con la estructura cristalina de GFP (Ormö et al., Science 1996, 273:1392), lo que indica que el colorante unido a la posición 39 se localizó a una distancia de 30 Å del cromóforo GFP. No se pudo observar ninguna señal de TERF cuando se desnaturalizó la proteína en clorhidrato de guanidinio 6 M y se hirvió durante 5 min a 95°C debido a la destrucción del cromóforo GFP.

### Ejemplo D: Expresión de GFP y MBP con AANN 13,16 y 17 incorporados en *E. coli*

40

#### D.1 GFP<sup>TAG</sup>

GFP<sup>TAG</sup> con AANN **13** incorporado, **16** o **17**, es decir GFP<sup>TAG->13</sup>, GFP<sup>TAG->16</sup> y GFP<sup>TAG->17</sup>, se preparó como se ha descrito para GFP<sup>TAG->1</sup> y GFP<sup>TAG->2</sup> (véase A.1) con las siguientes excepciones. Cuando los cultivos de *E. coli* alcanzaron una DO<sub>600</sub> entre 0,2 y 0,3, en lugar de los compuestos **1** o **2** el compuesto **13**, **16** o **17**, respectivamente, (solución de reserva 80 mM en NaOH 0,1 M) o una cantidad igual de NaOH 0,1 M (para experimentos de control negativo) se añadieron a una concentración final de 1 mM. Se purificó GFP<sup>TAG->13</sup> usando soporte de Macro-Prep HIC en lugar de perlas de Ni siguiendo el protocolo del fabricante (BIO-RAD, Múnich, Alemania).

Las células de *E. coli* se examinaron con respecto a fluorescencia de GFP, y las proteínas purificadas se separaron por SDS-PAGE y se tiñeron con Coomassie (Figura 6). Los resultados muestran incorporación eficaz de compuestos **1**, **13**, **16** y **17** por RS<sup>AF</sup>. La incorporación por RS<sup>WT</sup> fue menor con mayor expresión de GFP<sup>TAG</sup> en presencia del compuesto **16**.

#### D.2 MBP<sup>TAG</sup>

Una proteína de unión a maltosa (MBP) con un codón de terminación ámbar en el sitio permisivo 38 y un marcador His C terminal (MBP<sup>TAG</sup>) se usó para expresión por *E. coli* en presencia de AANN de forma análoga a GFP<sup>TAG</sup>. Se realizó expresión, lisis y purificación de proteínas como se describe en A.1 y D.1.

60

Se analizaron MBP<sup>TAG->1</sup>, MBP<sup>TAG->13</sup>, MBP<sup>TAG->16</sup> y MBP<sup>TAG->17</sup> purificadas de dicha *E. coli* usando un espectrómetro de masas en tándem de electropulverización de tiempo de vuelo de cuadrupolo (Q-ToF, Waters). Los resultados resumidos en la Tabla 3 confirman la incorporación de los AANN respectivos en la MBP<sup>TAG</sup>.

Tabla 3: Espectrometría de masas de MBP<sup>TAG->1</sup>, MBP<sup>TAG->13</sup>, MBP<sup>TAG->16</sup> y MBP<sup>TAG->17</sup> purificadas de *E. coli*. Los resultados se proporcionan como diferencias ( $\Delta$ ) con respecto a MBP<sup>TAG->AcF</sup>. AcF (p-acetilfenilalanina) es un AANN con una cadena principal de fenilalanina.

AANN			proteína	
	PM [Da] (calculado)	$\Delta$ PM [Da] con respecto a AcF (calculado)		$\Delta$ PM [Da] (observado) con respecto a MBP <sup>TAG-&gt;16</sup> (calculado)
AcF	207	0	MBP <sup>TAG-&gt;AcF</sup>	0
compuesto 1	296	89	MBP <sup>TAG-&gt;16</sup>	85
compuesto 13	298	91	MBP <sup>TAG-&gt;16</sup>	91
compuesto 16	282	75	MBP <sup>TAG-&gt;16</sup>	78
compuesto 17	296	89	MBP <sup>TAG-&gt;13</sup>	88

### 5 Ejemplo E: Marcaje con fluorescencia *in vivo* de GFP y MBP expresadas en *E. coli* con AANN incorporados

#### E.1 Visualización de proteínas GFP<sup>TAG</sup> marcadas *in vivo* separadas por SDS-PAGE

GFP<sup>TAG->1</sup>, GFP<sup>TAG->13</sup>, GFP<sup>TAG->16</sup> y GFP<sup>TAG->17</sup> se expresaron en *E. coli* como se describe en los ejemplos A y D y se marcaron con TAMRA-azida (Az) o TAMRA-tetrazina (Tet) (ambas 50  $\mu$ M durante 12 h a 37°C). Se usó GFP<sup>TAG</sup> expresada con el AANN propargilisina como un control negativo ya que los alquinos alifático solamente pueden realizar cicloadición de azida-alquino catalizada por cobre (I) y no reacciones azida-alquino promovidas por tensión. Las proteínas expresadas por *E. coli* se separaron por SDS-PAGE, y se visualizaron por tinción de Coomassie (Figura 7b) y exploración de fluorescencia (Figura 7a). La fluorescencia observada a la altura de la banda de GFP verificó el marcaje exitoso. Cada uno de GFP<sup>TAG->1</sup>, GFP<sup>TAG->13</sup>, GFP<sup>TAG->16</sup> y GFP<sup>TAG->17</sup> se marcó con éxito con TAMRA-tetrazina. Solamente la proteína que comprendía un resto de ciclooctinilo, es decir GFP<sup>TAG->1</sup>, se marcó con TAMRA-azida.

#### E.2 Visualización de MBP<sup>TAG</sup> marcada *in vivo* por microscopia de fluorescencia

Se cultivaron *E. coli* que expresaban MBP<sup>TAG</sup> en presencia del compuesto 1 o el compuesto 17 por separado y después se lavaron cuatro veces con PBS. Los dos cultivos de *E. coli* se mezclaron 1:1 (DO<sub>600</sub> ~2) y se incubaron con TAMRA-azida 50  $\mu$ M durante 4 h a 37°C. Después las *E. coli* se lavaron una vez con PBS antes de continuar la incubación con cumarina-tetrazina 10  $\mu$ M durante 4 h a 37°C. Como controles *E. coli* que expresaban MBP<sup>TAG->1</sup> y *E. coli* que expresaban MBP<sup>TAG->17</sup> se marcaron por separado con TAMRA-azida o cumarina-tetrazina, respectivamente. Después de marcar, todos los cultivos se lavaron cinco veces con PBS que contenía DMSO 5 % para eliminar el exceso de colorante. Las imágenes resultantes mostraban células fluorescentes verdes (MBP<sup>TAG->17</sup> marcada con cumarina-tetrazina) y rojas (MBP<sup>TAG->1</sup> marcada con TAMRA-azida). MBP<sup>TAG->17</sup> marcada con TAMRA-azida no mostró ninguna fluorescencia, mientras que MBP<sup>TAG->1</sup> marcada con cumarina-tetrazina mostró células verdes fluorescentes.

Tabla 4: Captura de imágenes *in vivo* de *E. coli* que expresan MBP<sup>TAG->1</sup> y MBP<sup>TAG->17</sup> marcadas con cumarina-tetrazina y TAMRA-azida. Fluorescencia de células indicada: cumarina = verde y TAMRA = rojo.

<i>E. coli</i> que expresan	cumarina-tetrazina	TAMRA-azida	superposición de cumarina / TAMRA
MBP <sup>TAG-&gt;1</sup> y MBP <sup>TAG-&gt;17</sup> (mezcladas)	verde	rojo	verde + rojo
MBP <sup>TAG-&gt;1</sup>	verde	sin fluorescencia	verde
MBP <sup>TAG-&gt;17</sup>	sin fluorescencia (no teñido)	solamente fluorescencia de fondo	solamente fluorescencia de fondo

#### E.3 Análisis cuantitativo de marcaje de TAMRA *in vivo* mediante TERF

Se expresó GFP<sup>TAG</sup> con compuesto 1, 13, 16 o 17 incorporado en *E. coli* como se describe en los ejemplos A y D. El lisado de *E. coli* se ajustó a una concentración de GFP final de aproximadamente 500 nM basándose en los espectros de absorbancia. Se añadió colorante 5  $\mu$ M (TAMRA-tetrazina o TAMRA-azida) y se registraron espectros de fluorescencia (excitación a 450 nm, emisión a 470-650 nm). Se supervisó el marcaje exitoso de GFP<sup>TAG</sup> mediante TERF del cromóforo GFP a TAMRA cuando se unió a la proteína. En los espectros individuales esto fue visible por una reducción de la fluorescencia de GFP (aproximadamente 505 nm) y un aumento simultáneo de la fluorescencia

de TAMRA (aproximadamente 575 nm) a lo largo del tiempo (gráfico de color oscuro a claro) como se muestra a modo de ejemplo para GFP<sup>TAG->13</sup> en la Figura 8).

5 Para evaluación de las trazas temporales correspondientes, todos los datos se corrigieron con respecto a excitación directa (es decir excitación de TAMRA por la luz de excitación) y filtración (emisión de GFP a la señal aceptora) usando el primer punto temporal (en el que la reacción aún no se ha producido sustancialmente). Para GFP<sup>TAG->13</sup> la reacción fue tan rápida que se usaron los valores de filtración y excitación directa de experimentos de control separados. Para observar reacciones más lentas y extraer constantes de velocidad, GFP<sup>TAG</sup> purificada con AANN se ajustó a una concentración final de aproximadamente 1 μM basándose en espectros de absorción. Se añadió TAMRA-tetrazina 5 μM y se registraron espectros de fluorescencia (excitación a 450 nm, emisión a 470-650 nm) durante varias horas. En el caso del compuesto 1 que reacciona con TAMRA-azida, la concentración de azida aumentó hasta 50 μM para conseguir marcaje en una escala temporal razonable. La cinética de reacción resultante se ajustó con un modelo monoexponencial sencillo de acuerdo con

$$15 \quad GFP^{AANN-TAMRA}(t) = A_0(1 - \exp(-k_Bt)),$$

donde A<sub>0</sub> corresponde a la amplitud del ajuste y es proporcional a la concentración de GFP inicial y B corresponde a la concentración de colorante dentro de la reacción. La constante de velocidad, k, de la reacción se obtuvo del ajuste con la suposición de B constante durante la reacción (que es válida debido al gran exceso de colorante). Se resumen constantes de velocidad aproximadas a 37°C en la Tabla 5.

Tabla 5: Cinética de reacción *in vitro* de GFP<sup>TAG</sup> con 1, 13, 16 o 17 incorporado.

	marcado con	
	TAMRA-tetrazina [1/s]	TAMRA-azida [1/s]
GFP <sup>TAG-&gt;1</sup>	~ 65	~ 1
GFPTAG->13	~ 40.000	sin reacción
GFPTAG->16	~ 6	sin reacción
GFPTAG->17	~ 8	sin reacción

### Ejemplo F: Expresión de GFP y MBP con AANN 1, 13, 16 y 17 incorporados en células de mamífero

25 F.1 Plásmidos y construcciones de ADN

Se generó un único plásmido de expresión para la expresión tanto de ARNt<sup>pyl</sup> como de la pirrolisina sintetasa pylRS reemplazando el ARNt y la sintetasa del plásmido de expresión de AANN de mamífero pSWAN (Liu et al., Nat Methods 2007, 4: 239) con un gen sintético obtenido del plásmido pEVOL ARNt<sup>pyl</sup>/pylRS<sup>WT</sup> o pEVOL ARNt<sup>pyl</sup>/pylRS<sup>AF</sup>. Esto dio como resultado la generación de los plásmidos pCMV ARNt<sup>pyl</sup>/pylRS<sup>WT</sup> y pCMV ARNt<sup>pyl</sup>/pylRS<sup>AF</sup> que se usaron para cotransfección de células de mamífero.

35 Para estudios de supresión de ámbar de mamífero, se generó una proteína de fusión SLN-mCherry-TAG<sup>GFP</sup>. Debido a la secuencia de localización nuclear (SLN) la proteína expresada se dirigió al núcleo. En consecuencia, la expresión de mCherry fue visible por fluorescencia naranja en el núcleo lo que significa la transfección exitosa del plásmido. La fluorescencia verde indicó expresión de GFP exitosa debido a la cotransfección exitosa del plásmido ARNt<sup>pyl</sup>/pylRS apropiado y la supresión del codón ámbar en la TAG<sup>GFP</sup> fusionada.

40 F.2 Procedimiento de microscopio automático para determinar la supresión de ámbar dependiente de AANN

Expresión de GFP dependiente de la concentración de AANN en presencia de AANN 1, 13, 16 o 17 y se analizó RS<sup>WT</sup> o RS<sup>AF</sup> usando el siguiente procedimiento de microscopio automático.

45 Se cultivaron células HeLa Kyoto en medio DMEM bajo en glucosa (1 g/l) (Sigma, Múnich, Alemania) con FBS 10 % (Sigma) y L-glutamina 1 %. Se sembraron 10-20X 10<sup>3</sup> células por pocillo en una cámara de 24 pocillos de fondo de vidrio y se cultivaron durante una noche. Al día siguiente, el medio de cultivo se intercambió con uno nuevo provisto de concentraciones crecientes de AANN (0, 1, 10, 100, 250, 1000 μM) y las células se cotransfectaron (relación 1:1) con plásmidos que portaban SLN-mCherry-TAG<sup>GFP</sup> y el par respectivo de ARNt<sup>pyl</sup>/pylRS (RS<sup>WT</sup> o RS<sup>AF</sup>) usando reactivo de transfección jetPRIME siguiendo el protocolo del fabricante (Polyplus-transfection SA, Illkirch, Francia). 24 h después de la transfección, las células se tiñeron con Hoechst 33342 (1 μg/ml, 10 min), se fijaron con paraformaldehído 2 % (15 min, TA) y se mantuvieron en PBS para captura de imágenes. Para cada concentración de AANN, el experimento se repitió dos veces y en dos cámaras de 24 pocillos preparadas de forma independiente.

55 Se realizó captura de imágenes por microscopía usando un microscopio Olympus ScanR de campo amplio

- automático (objetivo Uplan-Apo 20x, 0,70 NA, cámara Hamamatsu Orca R2 CCD) en tres canales (Hoechst, GFP, mCherry). Para cada pocillo se adquirieron al menos 25 imágenes (1344x1024 píxeles; 433x330  $\mu\text{m}$ , 12 bits) (tiempos de exposición: 30 ms Hoechst, 100 ms mCherry, 100 ms para GFP). La tinción de Hoechst permitió el uso de la opción de foco automático del sistema. Las imágenes se analizaron usando la macro ImageJ (<http://fiji.sc/wiki/index.php>) lo que permite la cuantificación de intensidad de fluorescencia de GFP en las células que muestran señal de fluorescencia de mCherry. Se usó tinción de Hoechst para actuar como umbral y seleccionar núcleos. Para cada núcleo detectado, se cuantificaron intensidades en los canales de mCherry y GFP. Se midieron intensidades de fondo en canales de GFP y mCherry usando un pocillo de cada cámara que contenía células que no estaban transfectadas pero se tiñeron con Hoechst. Los valores de fondo establecieron el umbral para diferenciar células que expresaban mCherry del fondo. Para cada pocillo se determinó la intensidad promedio de GFP ( $I_{\text{GFP}}$ ) en células positivas para señales de GFP y mCherry como una medida de la incorporación de AANN exitosa. Las intensidades de GFP se normalizaron por separado para cada AANN basándose en la señal observada máxima (para  $\text{RS}^{\text{WT}}$  o  $\text{RS}^{\text{AF}}$ ).
- Se resumen datos de  $I_{\text{GFP}}$  normalizados en la Tabla 6. Se detectó expresión de proteínas de fusión SLN-mCherry-TAG-GFP para todos los AANN de manera claramente dependiente de la concentración de AANN. La expresión de proteínas mostró normalmente un óptimo a aproximadamente 250  $\mu\text{M}$  de AANN.

Tabla 6: Intensidades de fluorescencia de GFP en núcleos de células HeLa cotransfectadas. Se indican las intensidades promedio de GFP ( $I_{\text{GFP}}$ ) y desviaciones típicas (DT).

Concentración de AANN [ $\mu\text{M}$ ]	compuesto 1				compuesto 13			
	$\text{RS}^{\text{WT}}$		$\text{RS}^{\text{AF}}$		$\text{RS}^{\text{WT}}$		$\text{RS}^{\text{AF}}$	
	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT
0	0,0	0,0	7,0	0,4	7,9	0,4	6,5	0,6
1	3,3	0,4	23,7	2,0	6,3	0,3	11,1	0,6
10	5,9	0,7	131,7	9,0	6,9	0,5	11,0	0,6
100	27,6	3,7	106,1	12,0	15,9	0,8	28,8	1,9
250	59,6	4,5	203,6	8,8	22,3	1,6	36,5	3,7
1000	57,5	8,0	118,4	12,0	33,2	1,9	20,6	2,2

Concentración de AANN [ $\mu\text{M}$ ]	compuesto 16				compuesto 17			
	$\text{RS}^{\text{WT}}$		$\text{RS}^{\text{AF}}$		$\text{RS}^{\text{WT}}$		$\text{RS}^{\text{AF}}$	
	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT	$I_{\text{GFP}}$	DT
0	1,8	0,4	11,6	1,3	5,7	0,5	3,8	0,2
1	1,5	0,4	5,7	0,7	4,2	0,3	10,9	1,2
10	6,9	0,8	46,1	11,3	2,6	0,2	57,5	4,6
100	9,4	2,2	114,8	9,0	4,3	0,5	76,6	10,0
250	171,6	10,5	100,4	22,4	5,7	0,5	163,5	8,8
1000	178,3	11,8	96,9	68,7	12,8	1,1	108,9	9,8

#### Abreviaturas

- AcF = p-acetilfenilalanina  
 AcOH = ácido acético  
 Boc-L-Lys-OH = *N*- $\alpha$ -*tert*-butiloxycarbonil-L-lisina  
 cHex = ciclohexano  
 DBU = 1,8-diazabicyclo[5.4.0]undec-7-eno  
 DCM = diclorometano  
 DMF = dimetilformamida  
 DMSO = dimetilsulfóxido

EtOH = etanol  
 EtOAc = acetato de etilo  
 CU = cromatografía ultrarrápida  
 TERF = transferencia de energía por resonancia de fluorescencia, también denominada transferencia de energía  
 5 por resonancia de Forster  
 MeOH = metanol  
 GFP = proteína verde fluorescente  
 GFP<sup>WT</sup> = GFP de tipo silvestre  
 GFP<sup>TAG</sup> = GFP codificada por una secuencia que comprende el codón de terminación ámbar TAG en el sitio  
 10 permisivo 39  
 GFP<sup>TAG->1</sup> = GFP<sup>TAG</sup> en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar  
 I<sub>GFP</sub> = intensidad promedio de GFP  
 MBP = proteína de unión a maltosa  
 MBP<sup>TAG</sup> = MBP codificada por una secuencia que comprendía el codón de terminación de ámbar TAG en el sitio  
 15 permisivo 38 y un marcador de His C-terminal  
 MBP<sup>TAG->1</sup> = MBP<sup>TAG</sup> en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar  
 mCherry<sup>WT</sup> = mCherry de tipo silvestre  
 mCherry<sup>TAG->1</sup> = mCherry en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar  
 20 SLN = secuencia de localización nuclear  
 DO<sub>600</sub> = densidad óptica a 600 nm  
 PBS = solución salina tamponada con fosfato  
 PMSF = fenilmetilsulfonilfluoruro  
 TA = temperatura ambiente  
 DT = desviación típica  
 25 SDS-PAGE = electroforesis en gel de poliacrilamida de dodecil sulfato sódico  
 TERF<sub>um</sub> = observación de una única molécula de TERF  
 TAMRA = tetrametilrodamina  
 TB = caldo de cultivo Terrific  
 TEA = trietilamina  
 30 THF = tetrahidrofurano  
 TLC = cromatografía de capa fina  
 AANN = aminoácido no natural

LISTA DE SECUENCIAS

35 <110> EMBL  
 <120> Aminoácidos no naturales que comprenden un grupo análogo de ciclooctinilo o trans-ciclooctenilo y usos  
 de los mismos  
 40 <130> M/51411-PCT  
 <150> US 61/462.477  
 <151> 03/02/2011  
 45 <150> US 61/453.358  
 <151> 16/03/2011  
 <160> 2  
 50 <170> PatentIn versión 3.5  
 <210> 1  
 <211> 454  
 55 <212> PRT  
 <213> M. maize  
 <400> 1

ES 2 661 095 T3

Met Asp Lys Lys Pro Leu Asn Thr Leu Ile Ser Ala Thr Gly Leu Trp  
 1 5 10 15

Met Ser Arg Thr Gly Thr Ile His Lys Ile Lys His His Glu Val Ser  
 20 25 30

Arg Ser Lys Ile Tyr Ile Glu Met Ala Cys Gly Asp His Leu Val Val  
 35 40 45

Asn Asn Ser Arg Ser Ser Arg Thr Ala Arg Ala Leu Arg His His Lys  
 50 55 60

Tyr Arg Lys Thr Cys Lys Arg Cys Arg Val Ser Asp Glu Asp Leu Asn  
 65 70 75 80

Lys Phe Leu Thr Lys Ala Asn Glu Asp Gln Thr Ser Val Lys Val Lys  
 85 90 95

Val Val Ser Ala Pro Thr Arg Thr Lys Lys Ala Met Pro Lys Ser Val  
 100 105 110

Ala Arg Ala Pro Lys Pro Leu Glu Asn Thr Glu Ala Ala Gln Ala Gln  
 115 120 125

Pro Ser Gly Ser Lys Phe Ser Pro Ala Ile Pro Val Ser Thr Gln Glu  
 130 135 140

Ser Val Ser Val Pro Ala Ser Val Ser Thr Ser Ile Ser Ser Ile Ser

ES 2 661 095 T3

145							150									155				160
Thr	Gly	Ala	Thr	Ala	Ser	Ala	Leu	Val	Lys	Gly	Asn	Thr	Asn	Pro	Ile					
				165					170					175						
Thr	Ser	Met	Ser	Ala	Pro	Val	Gln	Ala	Ser	Ala	Pro	Ala	Leu	Thr	Lys					
			180					185					190							
Ser	Gln	Thr	Asp	Arg	Leu	Glu	Val	Leu	Leu	Asn	Pro	Lys	Asp	Glu	Ile					
		195					200					205								
Ser	Leu	Asn	Ser	Gly	Lys	Pro	Phe	Arg	Glu	Leu	Glu	Ser	Glu	Leu	Leu					
	210					215					220									
Ser	Arg	Arg	Lys	Lys	Asp	Leu	Gln	Gln	Ile	Tyr	Ala	Glu	Glu	Arg	Glu					
	225				230					235										240
Asn	Tyr	Leu	Gly	Lys	Leu	Glu	Arg	Glu	Ile	Thr	Arg	Phe	Phe	Val	Asp					
				245					250					255						
Arg	Gly	Phe	Leu	Glu	Ile	Lys	Ser	Pro	Ile	Leu	Ile	Pro	Leu	Glu	Tyr					
			260					265					270							
Ile	Glu	Arg	Met	Gly	Ile	Asp	Asn	Asp	Thr	Glu	Leu	Ser	Lys	Gln	Ile					
		275					280					285								
Phe	Arg	Val	Asp	Lys	Asn	Phe	Cys	Leu	Arg	Pro	Met	Leu	Ala	Pro	Asn					
	290					295					300									
Leu	Tyr	Asn	Tyr	Leu	Arg	Lys	Leu	Asp	Arg	Ala	Leu	Pro	Asp	Pro	Ile					
	305				310					315					320					
Lys	Ile	Phe	Glu	Ile	Gly	Pro	Cys	Tyr	Arg	Lys	Glu	Ser	Asp	Gly	Lys					
				325					330					335						
Glu	His	Leu	Glu	Glu	Phe	Thr	Met	Leu	Asn	Phe	Cys	Gln	Met	Gly	Ser					
			340					345					350							
Gly	Cys	Thr	Arg	Glu	Asn	Leu	Glu	Ser	Ile	Ile	Thr	Asp	Phe	Leu	Asn					
		355					360					365								
His	Leu	Gly	Ile	Asp	Phe	Lys	Ile	Val	Gly	Asp	Ser	Cys	Met	Val	Tyr					
	370					375					380									
Gly	Asp	Thr	Leu	Asp	Val	Met	His	Gly	Asp	Leu	Glu	Leu	Ser	Ser	Ala					
	385				390					395					400					



ES 2 661 095 T3

Val Val Gly Pro Ile Pro Leu Asp Arg Glu Trp Gly Ile Asp Lys Pro  
 405 410 415

Trp Ile Gly Ala Gly Phe Gly Leu Glu Arg Leu Leu Lys Val Lys His  
 420 425 430

Asp Phe Lys Asn Ile Lys Arg Ala Ala Arg Ser Glu Ser Tyr Tyr Asn  
 435 440 445

Gly Ile Ser Thr Asn Leu  
 450

<210> 2

<211> 454

<212> PRT

5 <213> Secuencia artificial

<220>

<223> Pirrolisil ARNt sintetasa de M. maize mutante

10 <400> 2

Met Asp Lys Lys Pro Leu Asn Thr Leu Ile Ser Ala Thr Gly Leu Trp  
 1 5 10 15

Met Ser Arg Thr Gly Thr Ile His Lys Ile Lys His His Glu Val Ser  
 20 25 30

Arg Ser Lys Ile Tyr Ile Glu Met Ala Cys Gly Asp His Leu Val Val  
 35 40 45

Asn Asn Ser Arg Ser Ser Arg Thr Ala Arg Ala Leu Arg His His Lys  
 50 55 60

Tyr Arg Lys Thr Cys Lys Arg Cys Arg Val Ser Asp Glu Asp Leu Asn  
 65 70 75 80

Lys Phe Leu Thr Lys Ala Asn Glu Asp Gln Thr Ser Val Lys Val Lys  
 85 90 95

Val Val Ser Ala Pro Thr Arg Thr Lys Lys Ala Met Pro Lys Ser Val  
 100 105 110

Ala Arg Ala Pro Lys Pro Leu Glu Asn Thr Glu Ala Ala Gln Ala Gln  
 115 120 125

Pro Ser Gly Ser Lys Phe Ser Pro Ala Ile Pro Val Ser Thr Gln Glu  
 130 135 140

Ser Val Ser Val Pro Ala Ser Val Ser Thr Ser Ile Ser Ser Ile Ser

ES 2 661 095 T3

145					150						155					160
Thr	Gly	Ala	Thr	Ala	Ser	Ala	Leu	Val	Lys	Gly	Asn	Thr	Asn	Pro	Ile	
				165					170					175		
Thr	Ser	Met	Ser	Ala	Pro	Val	Gln	Ala	Ser	Ala	Pro	Ala	Leu	Thr	Lys	
			180					185					190			
Ser	Gln	Thr	Asp	Arg	Leu	Glu	Val	Leu	Leu	Asn	Pro	Lys	Asp	Glu	Ile	
		195					200					205				
Ser	Leu	Asn	Ser	Gly	Lys	Pro	Phe	Arg	Glu	Leu	Glu	Ser	Glu	Leu	Leu	
	210					215					220					
Ser	Arg	Arg	Lys	Lys	Asp	Leu	Gln	Gln	Ile	Tyr	Ala	Glu	Glu	Arg	Glu	
225					230					235						240
Asn	Tyr	Leu	Gly	Lys	Leu	Glu	Arg	Glu	Ile	Thr	Arg	Phe	Phe	Val	Asp	
				245					250					255		
Arg	Gly	Phe	Leu	Glu	Ile	Lys	Ser	Pro	Ile	Leu	Ile	Pro	Leu	Glu	Tyr	
			260					265					270			
Ile	Glu	Arg	Met	Gly	Ile	Asp	Asn	Asp	Thr	Glu	Leu	Ser	Lys	Gln	Ile	
		275					280					285				
Phe	Arg	Val	Asp	Lys	Asn	Phe	Cys	Leu	Arg	Pro	Met	Leu	Ala	Pro	Asn	
	290					295					300					
Leu	Ala	Asn	Tyr	Leu	Arg	Lys	Leu	Asp	Arg	Ala	Leu	Pro	Asp	Pro	Ile	
305					310					315					320	
Lys	Ile	Phe	Glu	Ile	Gly	Pro	Cys	Tyr	Arg	Lys	Glu	Ser	Asp	Gly	Lys	
				325					330					335		
Glu	His	Leu	Glu	Glu	Phe	Thr	Met	Leu	Asn	Phe	Cys	Gln	Met	Gly	Ser	
			340					345					350			
Gly	Cys	Thr	Arg	Glu	Asn	Leu	Glu	Ser	Ile	Ile	Thr	Asp	Phe	Leu	Asn	
		355					360					365				
His	Leu	Gly	Ile	Asp	Phe	Lys	Ile	Val	Gly	Asp	Ser	Cys	Met	Val	Phe	
	370					375					380					
Gly	Asp	Thr	Leu	Asp	Val	Met	His	Gly	Asp	Leu	Glu	Leu	Ser	Ser	Ala	
385					390					395					400	

ES 2 661 095 T3

Val Val Gly Pro Ile Pro Leu Asp Arg Glu Trp Gly Ile Asp Lys Pro  
405 410 415

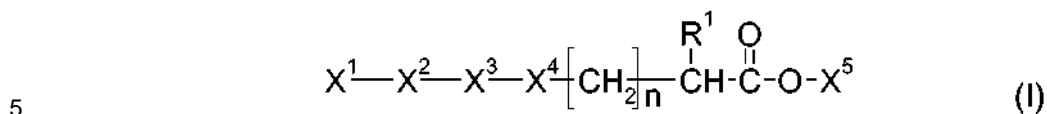
Trp Ile Gly Ala Gly Phe Gly Leu Glu Arg Leu Leu Lys Val Lys His  
420 425 430

Asp Phe Lys Asn Ile Lys Arg Ala Ala Arg Ser Glu Ser Tyr Tyr Asn  
435 440 445

Gly Ile Ser Thr Asn Leu  
450

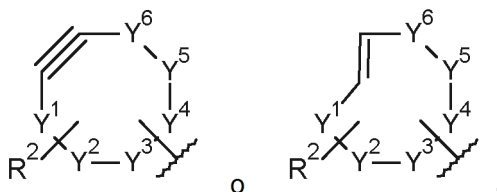
REIVINDICACIONES

1. Un compuesto de fórmula I



en la que:

10  $X^1$  tiene las fórmulas



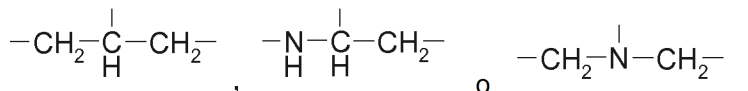
en las que:

15  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son independientemente  $-CH_2-$ ,  $-NH-$ ,  $-S-$  u  $-O-$  con la condición de que al menos 4 de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  sean  $-CH_2-$ ;

20  $R^2$  es hidrógeno, halógeno, alquilo  $C_1-C_4$ ,  $CF_3$ , CN, alcoxi  $C_1-C_4$ ,  $-O-CF_3$ , alcanoiloxi  $C_2-C_5$ , alquilaminocarboniloxi  $C_1-C_4$  o alquiltio  $C_1-C_4$ ;

$X^2$  es  $-CH_2-$ ,  $-O-$ ,  $-S-$ ,  $-NH-$ ,  $-C(O)-$ ,  $-OC(O)-$ ,  $-C(O)O-$ ,  $-NH-C(O)-$  o  $-C(O)-NH-$ , o  $X^2$  es  $>CH-$  o  $>N-$  en donde el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forman un anillo de 3 miembros, o

25  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en donde los dos átomos de carbono o el carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 4 miembros, o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,



30 en las que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos en el anillo de  $X^1$  adyacentes forman un anillo de 5 miembros;

35  $X^3$  es alquileo  $C_1-C_6$ ,  $-(CH_2-CH_2-O)_m-$ ,  $-(CH_2-O)_p-$  o un enlace sencillo;  $X^4$  es  $-NH-$ ,  $-C(O)-NH-$ ,  $-NH-C(O)-$ ,  $-NH-CH(NH_2)-$ ,  $-CH(NH_2)-NH-$ ,  $-NH-C(NH)-NH-$ ,  $-C(O)-NH-CH(NH_2)-$ ,  $-C(O)-NH-C(NH)-NH-$ ,  $NH-CH(NH_2)-C(O)-$  o  $-NH-C(NH)-NH-C(O)-$ ;

$X^5$  es hidrógeno, alquilo  $C_1-C_6$ , alcoxi  $C_1-C_6$ -alquilo  $C_1-C_2$ , alcanoiloxi  $C_2-C_7$ -alquilo  $C_1-C_2$  o alcanoilulfanil  $C_2-C_7$ -alquilo  $C_1-C_2$ ;

$R^1$  es  $-OH$  o  $-NH_2$ ;

$n$  es un número entero de 1 a 4;

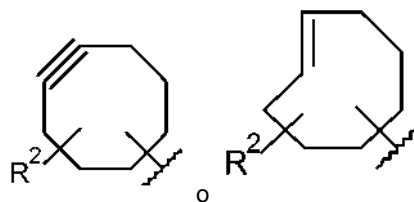
$m$  es un número entero de 1 a 6; y

40  $p$  es un número entero de 1 a 6,

o una sal de adición de ácidos o bases del mismo.

45 2. El compuesto o la sal de la reivindicación 1, en los que  $R^2$  es hidrógeno o halógeno; en particular  $R^2$  es flúor, tal como dos flúor enlazados a un átomo del anillo de carbono.

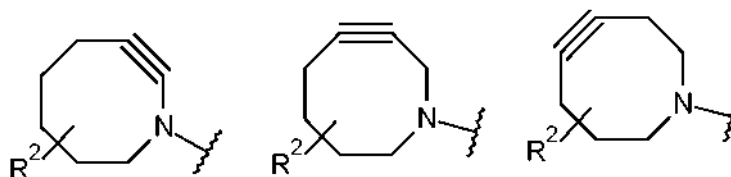
3. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que  $X^1$  tiene las fórmulas



en las que  $R^2$  es como se ha definido en las reivindicaciones 1 o 2.

5 4. El compuesto o la sal de una cualquiera de las reivindicaciones 1 a 3, en los que uno de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  es -NH- mientras que los cinco restantes de  $Y^1, Y^2, Y^3, Y^4, Y^5, Y^6$  son -CH<sub>2</sub>-) y  $R^2$  es como se ha definido en una cualquiera de las reivindicaciones 1 a 3.

10 5. El compuesto o la sal de las reivindicaciones 1 o 2, en los que  $X^1$  tiene una fórmula seleccionada entre



y  $R^2$  es como se ha definido en las reivindicaciones 1 o 2.

15 6. El compuesto o la sal de la reivindicación 1, en los que  $X^1$  es ciclooctinilo no sustituido; o  $X^1$  es ciclooctinilo sustituido con uno o dos átomos de halógeno, en particular ciclooctinilo sustituido con uno o dos átomos de flúor; o  $X^1$  es trans-ciclooctenilo no sustituido.

20 7. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que  $X^2$  es -O-.

8. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que  $X^3$  es -CH<sub>2</sub>-CH<sub>2</sub>-O- o un enlace sencillo.

25 9. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que el elemento estructural - $X^2$ - $X^3$ - comprende de 1 a 6 átomos en la cadena principal.

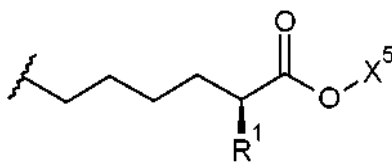
10. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que  $X^4$  es -NH-, -C(O)-NH-, -NH-CH(NH<sub>2</sub>)-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH<sub>2</sub>)- o -C(O)-NH-C(NH)-NH-.

30 11. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que n es 3 o 4.

12. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que el elemento estructural - $X^2$ - $X^3$ - $X^4$ -(CH<sub>2</sub>)<sub>n</sub>- comprende de 5 a 12 átomos en la cadena principal, tal como 6, 7, 8, 9, 10 u 11 átomos en la cadena principal.

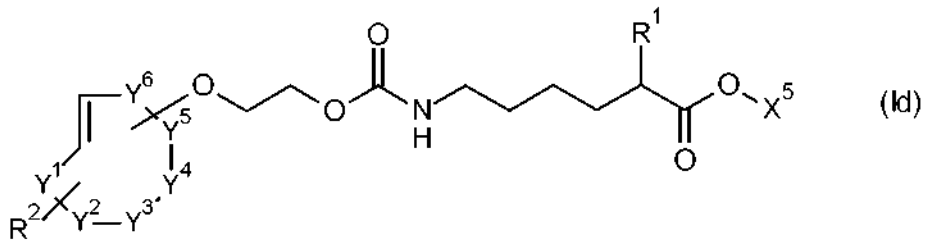
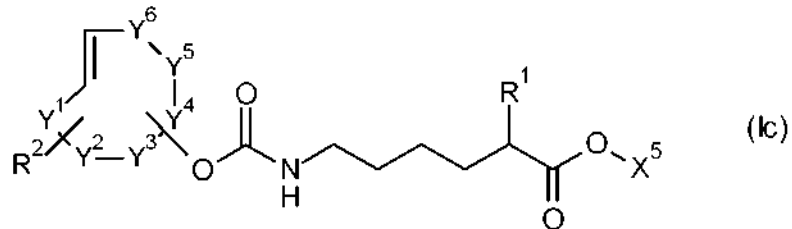
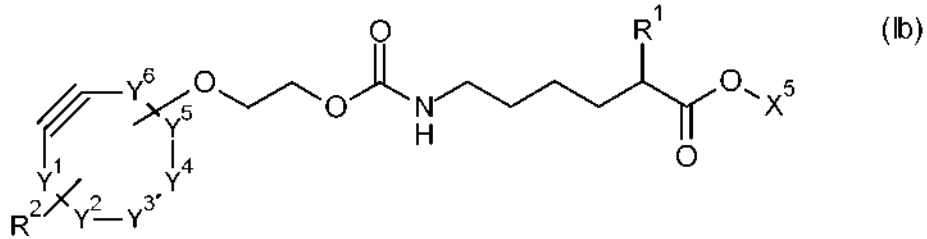
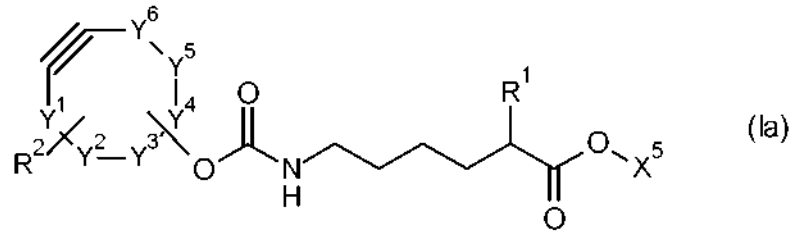
35 13. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores, en los que  $X^5$  es hidrógeno, alcoximetilo C<sub>1</sub>-C<sub>6</sub>, alcoxi-1-ilo C<sub>1</sub>-C<sub>6</sub>, alcanoiloximetilo C<sub>2</sub>-C<sub>7</sub> o alcanoilulfaniletilo C<sub>2</sub>-C<sub>7</sub>.

40 14. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores que tienen configuración S con respecto al átomo de carbono asimétrico que porta  $R^1$ , en particular en donde -(CH<sub>2</sub>)<sub>n</sub>-CHR<sup>1</sup>-C(O)O- $X^5$  tiene la fórmula



45 en la que  $R^1$  y  $X^5$  son como se han definido en una cualquiera de las reivindicaciones anteriores.

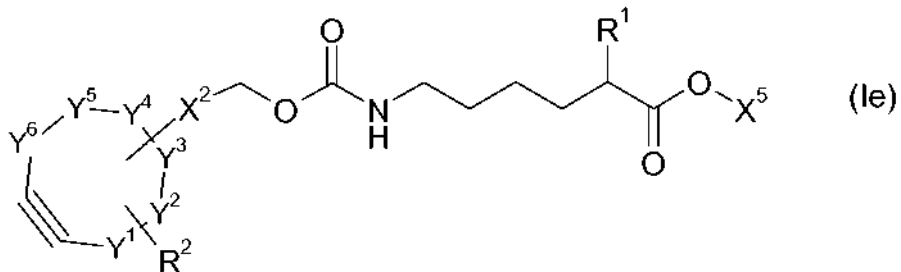
15. El compuesto o la sal de una cualquiera de las reivindicaciones anteriores que es un compuesto de una cualquiera de las fórmulas Ia, Ib, Ic e Id



en las que  $R^1$ ,  $R^2$ ,  $X^5$  e  $Y^1$  a  $Y^6$  son como se han definido en una de las reivindicaciones anteriores, o una sal de adición de ácidos o bases del mismo.

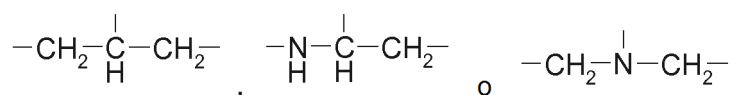
5

16. El compuesto o la sal de una cualquiera de las reivindicaciones 1 a 6 que es un compuesto de fórmula le



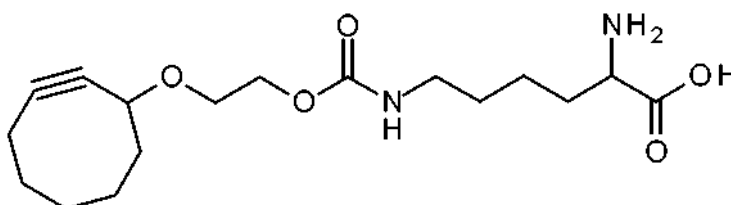
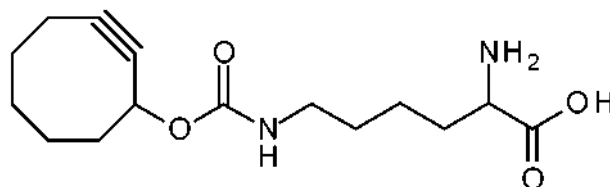
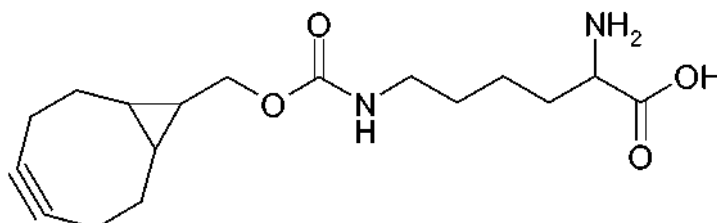
10 en la que  $R^1$ ,  $R^2$ ,  $X^5$  e  $Y^1$  a  $Y^6$  son como se definen en el presente documento y  $X^2$  es  $>CH-$  o  $>N-$  en donde el átomo de carbono o el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forman un anillo de 3 miembros; o  $X^2$  es  $-CH_2-CH<$ ,  $-NH-CH<$  o  $-CH_2-N<$  en donde los dos átomos de carbono o el átomo de carbono y el de nitrógeno junto con dos átomos adyacentes en el anillo de  $X^1$  forman un anillo de 4 miembros; o  $X^2$  es  $-CH_2-CH_2-CH<$ ,  $-NH-CH_2-CH<$ ,  $-CH_2-NH-CH<$ ,  $-CH_2-CH_2-N<$ ,

15



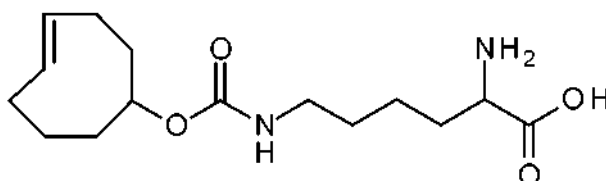
5 en las que los tres átomos de carbono o los dos átomos de carbono y el átomo de nitrógeno junto con dos átomos adyacentes en el anillo de X<sup>1</sup> forman un anillo de 5 miembros; o una sal de adición de ácidos o bases del mismo.

17. El compuesto o la sal de la reivindicación 1 que es un compuesto de fórmula



10 o

”



15 o una sal de adición de ácidos o bases del mismo.

18. Un proceso para preparar un polipéptido diana que tiene uno o varios grupos análogos de ciclooctinilo o trans-ciclooctenilo, comprendiendo el proceso:

20 a) proporcionar un sistema de traducción que comprende:

- (i) una aminoacil ARNt sintetasa o un polinucleótido que la codifica;
- ii) un compuesto o una sal de una cualquiera de las reivindicaciones 1 a 17;
- (iii) un ARNt que tiene un anticodón para un codón selector o un polinucleótido que codifica dicho ARNt; y
- (iv) un polinucleótido que codifica el polipéptido diana y que comprende uno o varios codones selectores,

25

en donde la aminoacil ARNt sintetasa (i) es capaz de acilar específicamente el ARNt (iii) con el compuesto o la sal (ii);

b) permitir la traducción del polinucleótido (iv) incorporando de este modo el compuesto (ii) al polipéptido diana en la posición o las posiciones codificadas por el codón o los codones selectores.

5 19. El proceso de la reivindicación 18, en el que dicho sistema de traducción es una célula que expresa dicha aminoacil ARNt sintetasa.

20. El proceso de las reivindicaciones 18 o 19, en el que dicha aminoacil ARNt sintetasa es una pirrolisil ARNt sintetasa, en particular una pirrolisil ARNt sintetasa que comprende la secuencia de aminoácidos expuesta en las SEQ ID NO: 1 o 2.

10

21. Uso de un compuesto de una cualquiera de las reivindicaciones 1 a 17 para preparar un polipéptido que tiene uno o varios grupos análogos de ciclooctinilo o trans-ciclooctenilo.



Figura 1

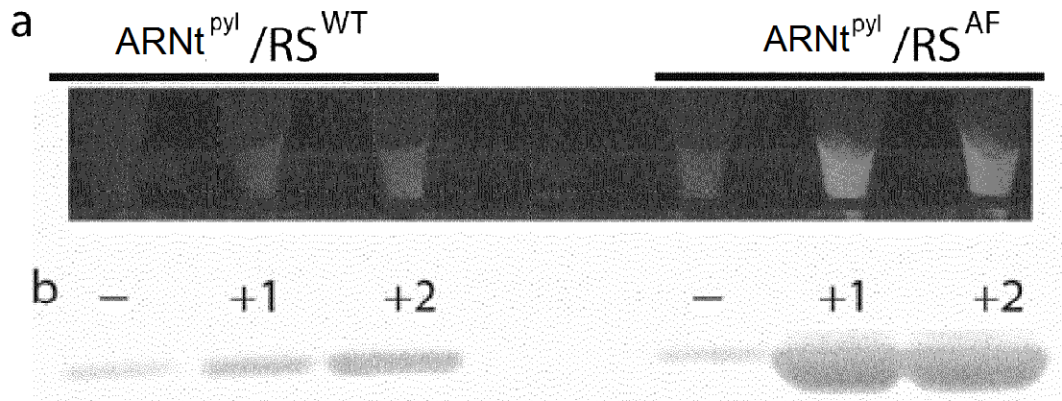


Figura 2

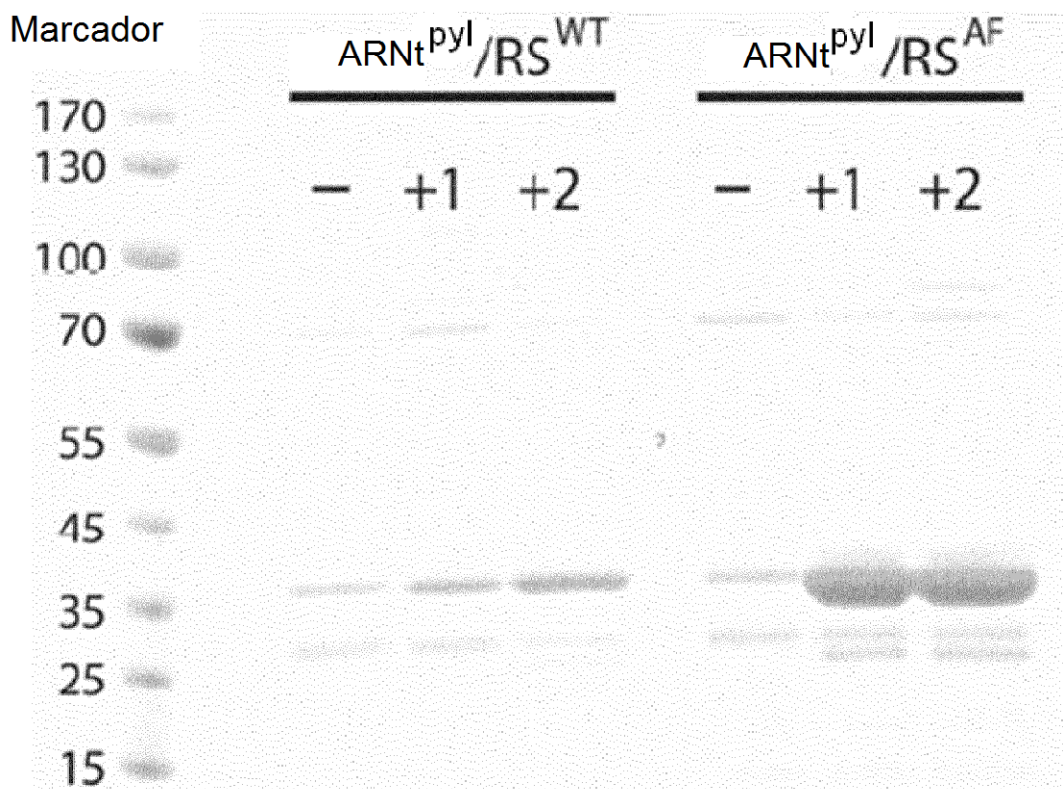


Figura 3

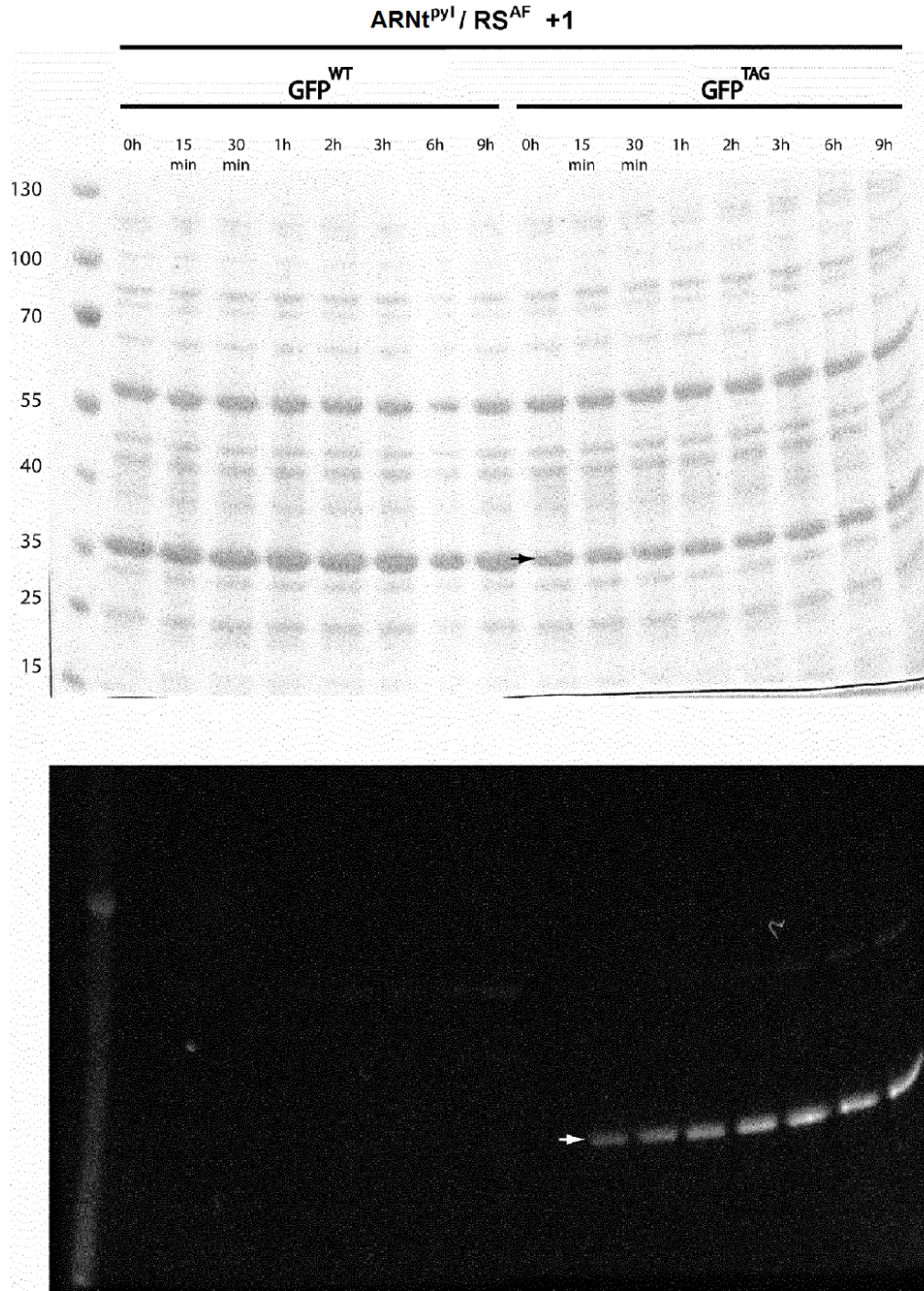


Figura 4

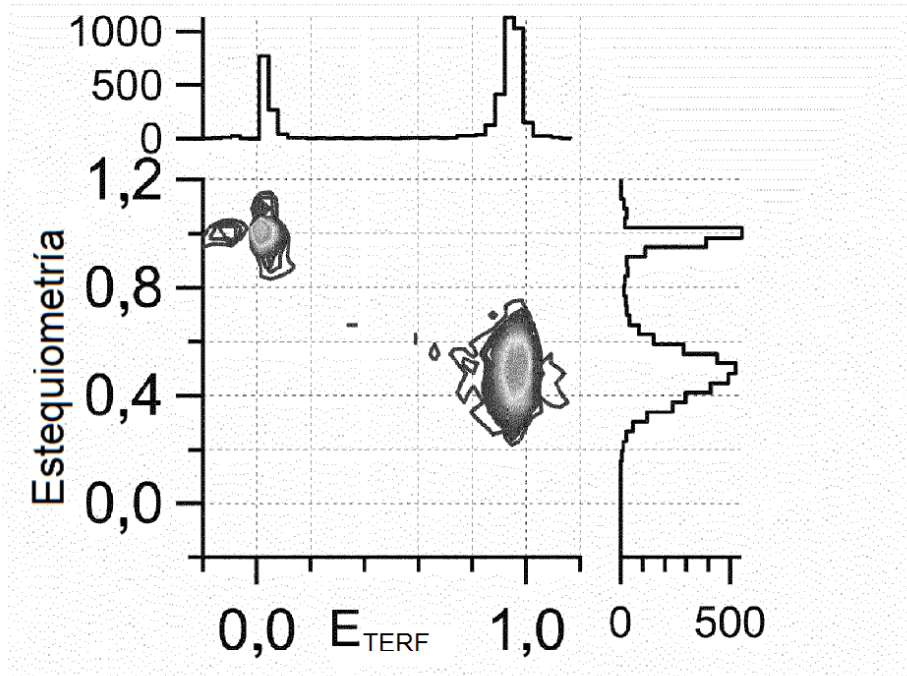


Figura 5

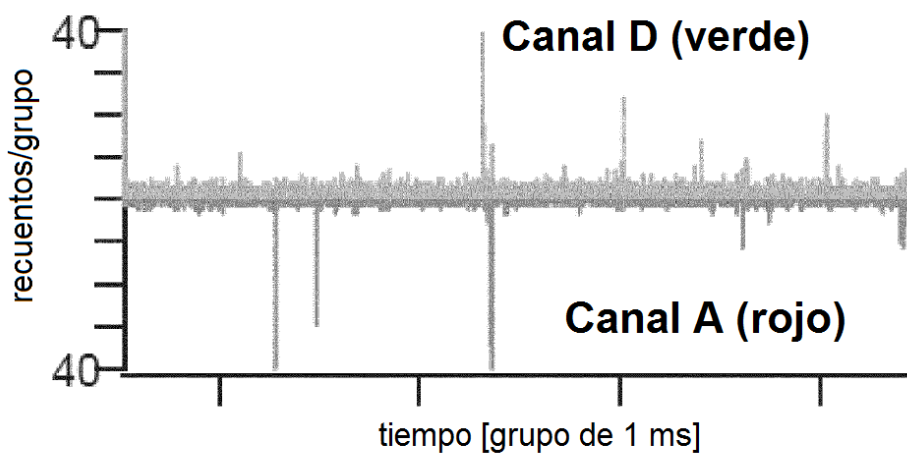


Figura 6

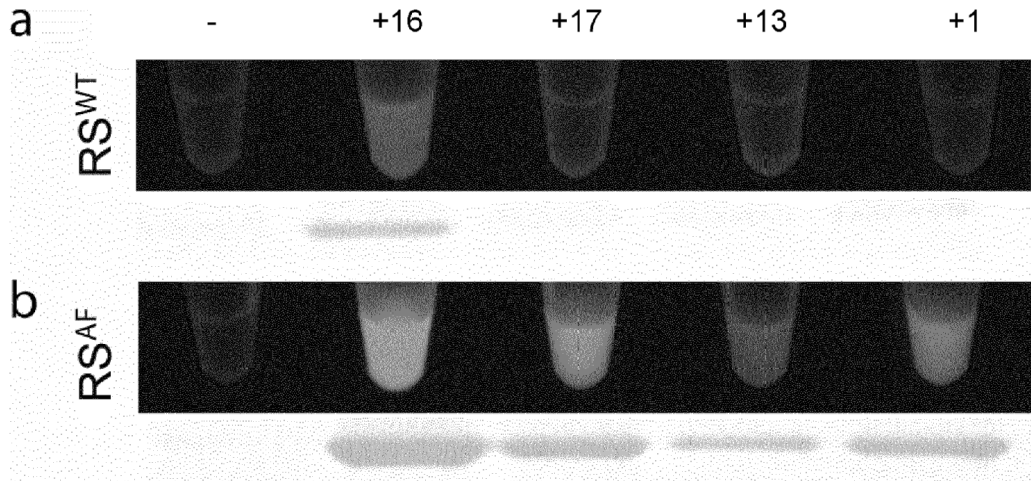


Figura 7

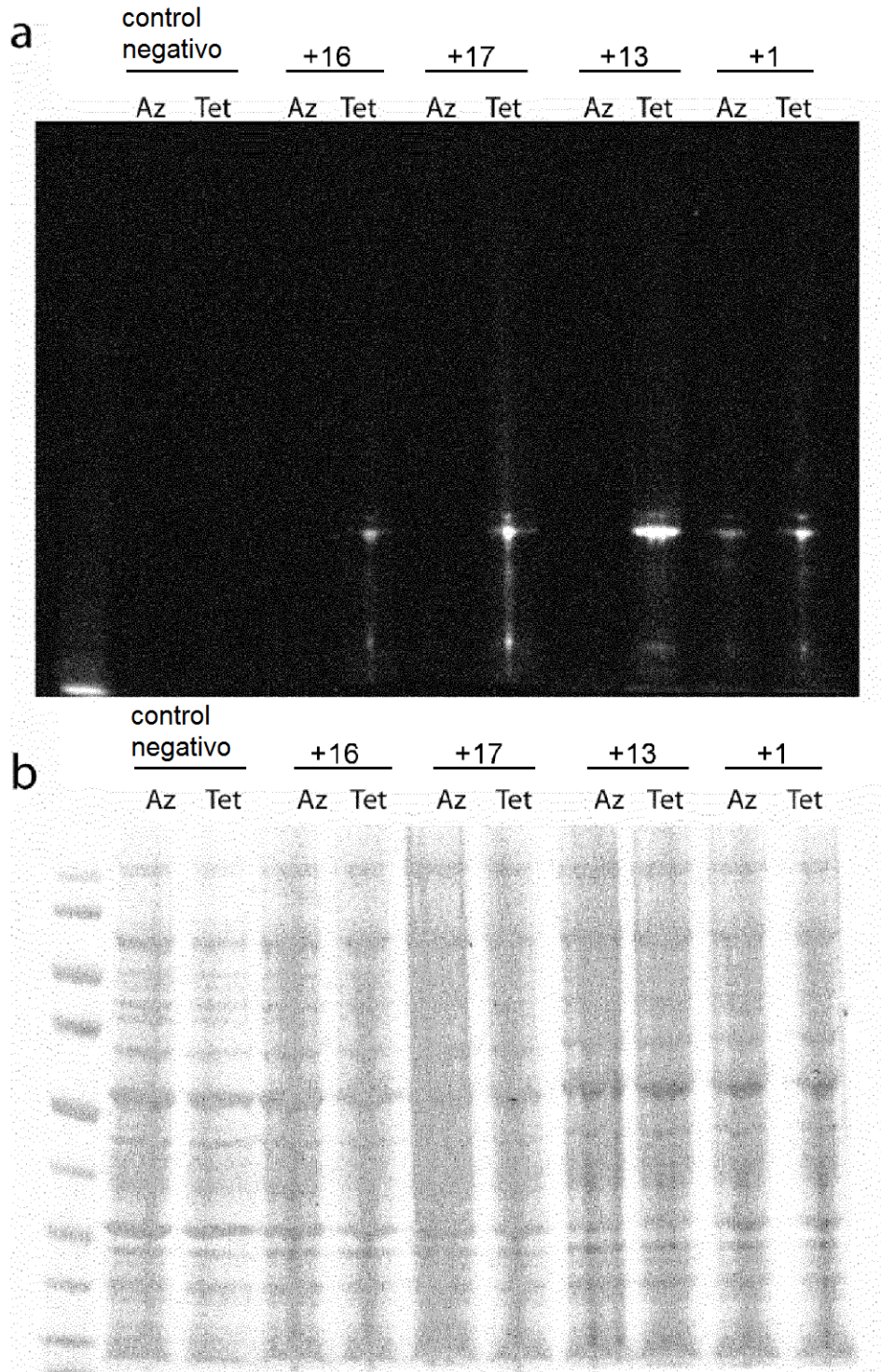


Figura 8

