

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 699 082**

51 Int. Cl.:

G01N 33/58 (2006.01)

C07C 271/22 (2006.01)

C07C 271/34 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.01.2015 PCT/EP2015/050555**

87 Fecha y número de publicación internacional: **23.07.2015 WO15107064**

96 Fecha de presentación y número de la solicitud europea: **14.01.2015 E 15703230 (1)**

97 Fecha y número de publicación de la concesión europea: **22.08.2018 EP 3094977**

54 Título: **Múltiples reacciones de cicloadición para el etiquetado de moléculas**

30 Prioridad:

14.01.2014 EP 14151175

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
07.02.2019

73 Titular/es:

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

72 Inventor/es:

**LEMKE, EDWARD;
SCHULTZ, CARSTEN;
PLASS, TILLMANN;
NIKIC, IVANA;
HOFFMANN, JAN-ERIK y
VALLE ARAMBURU, IKER**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 699 082 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

Múltiples reacciones de cicloadición para el etiquetado de moléculas

5 **Campo de la invención**

La invención se refiere a métodos para unir tetrazinas con dienófilos para establecer al menos dos enlaces realizando secuencialmente al menos dos reacciones de cicloadición. Los métodos en particular permiten establecer estrategias de etiquetado múltiple.

10

Antecedentes de la invención

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.

15

20

25

30

35

40

45

Las químicas potentes que se desarrollan eficientemente a temperaturas fisiológicas y en entornos biológicos altamente funcionalizados son la cicloadición de tipo Huisgen (3+2) catalizada por cobre (I) entre azidas lineales y alquinos, la cicloadición libre de cobre 3+2 entre azidas lineales y cicloalquinos en tensión, o la reacción de cicloadición de Diels-Alder (4+2) de demanda de electrones inversa entre un dienófilo en tensión tal como *trans*-cicloocteno o norborneno y una 1,2,4,5-tetrazina, ambas formas de química de clic (Blackman et al., J. Am. Chem. Soc. 2008, 130, 13518-13519; Kolb et al., Angew Chem Int Ed Engl 2001,40:2004; Devaraj et al., Angew Chem Int Ed Engl 2009, 48:7013; Devaraj et al., Bioconjugate Chem 2008, 19:2297; Devaraj et al., Angew Chem Int Ed Engl 2010, 49:2869; documento WO 2010/119389 A2; documento WO 2010/051530 A2). La cicloadición estándar (3+2) entre un alquino y una azida requiere un catalizador de cobre que sea tóxico para las bacterias y las células de los mamíferos, lo que reduce considerablemente la biocompatibilidad de este tipo de química de clic. Esta limitación ha sido superada por Bertozzi y colaboradores, quienes demostraron que la reacción de clic se desarrolla fácilmente sin la necesidad de un catalizador tóxico para las células cuando se utilizan alquinos en tensión anular como sustrato (Agard et al., J Am Chem Soc 2004, 126:15046; documento WO 2006/050262 A2). Desde entonces, la química de clic sin cobre ha encontrado un aumento de aplicaciones en el etiquetado de biomoléculas. Se usaron colorantes fluorescentes que comprendían grupos ciclooctinilo para etiquetar carbohidratos y proteínas que comprenden restos azida unidos enzimáticamente *in vivo* (Chang et al., Proc Natl Acad Sci USA 2010, 107:1821) y el etiquetado de ácido fosfatídico modificado con cicloalquino con fluoróforos de azido se describe en Neef y Schultz, Angew Chem Int Ed Engl 2009, 48:1498. No se requirió catalizador en estas aplicaciones. Karver et al. describen pares de reacciones bioortogonales y mutuamente excluyentes para imágenes simultáneas, selectivas y de múltiples dianas (Angew Chem Int Ed Engl 2012, 51:920-922).

50

55

60

Entre el repertorio en expansión de las químicas, las aplicaciones de química *in vivo* de las cicloadiciones de Diels-Alder inversas entre tetrazinas y dienófilos en tensión atraen un interés significativo, particularmente de aquellos interesados en realizar imágenes de células vivas y animales. Las ligaduras de tetrazina se benefician de una cinética rápida y ajustable, así como de la existencia de sondas fluorógenas. Las aplicaciones biomédicas de las cicloadiciones de tetrazina se han descrito ampliamente y se ha abordado la implementación de ligaduras de tetrazina en los diagnósticos de nanomateriales. Para todo esto, véase, por ejemplo, la revisión de Seckute y Devaraj, Current Opinion in Chemical Biology 2013, 17, 761 - 767, y las referencias citadas en el mismo. Además, se están desarrollando nuevas tetrazinas y métodos para sintetizarlas (véase, por ejemplo, el documento WO 2013/152359 A1). Más específicamente, el documento WO 2011/095336 A2 describe métodos y kits para la modificación post-sintética de ácidos nucleicos por la reacción inversa de Diels-Alder, y el documento WO 2013/029801 A1 describe métodos para el etiquetado ortogonal múltiple de oligonucleótidos realizando simultáneamente una reacción Diels-Alder inversa y una reacción de clic catalizada por cobre. El documento WO 2011/112970 A2 proporciona composiciones y métodos que utilizan reacciones de cicloadición de Diels-Alder de demanda de electrones inversa bioortogonal para un acoplamiento rápido y específico de compuestos orgánicos a puntos cuánticos (QD).

65

La modificación traduccional de proteínas por codificación genética directa de aminoácidos no naturales fluorescentes usando un par de ARNt ortogonal/aminoacil ARNt sintetasa ofrece especificidad exquisita, libertad de colocación dentro de la proteína diana y, en su caso, 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 UAA dansilalanina en *Saccharomyces cerevisiae*. En respuesta al codón de terminación ámbar TAG, la maquinaria de traducción del huésped incorporó inmediatamente dansilalanina.

Este enfoque se ha usado mientras tanto para codificar genéticamente varios colorantes pequeños y otros restos de interés. Por ejemplo, tirosil ARN^{t^{tyr}}/sintetasa modificado de *Methanococcus jannaschii*, leucil ARN^{t^{leu}}/sintetasa de *E. coli*, así como pirrolisina ARN^{t^{pyl}}/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; documento WO 2013/108044 A2; documento WO 2002/085923 A2; documento WO 2002/086075 A2; documento EP2192185 A1). Wu et al (Chembiochem 2012, 13: 1405-1408) y Chatterjee et al. (Biochemistry 2013, 52: 1828-1837) también se ocupan de la preparación traslacional de polipéptidos que contienen residuos de aminoácidos no naturales que pueden utilizarse para fines de etiquetado.

El poder de las técnicas de microscopía de súper resolución (SRM) depende en gran medida de las características de los fluoróforos. La mayoría de los tintes orgánicos tienen mejores propiedades fotofísicas y son normalmente más de 20 veces más pequeños que las proteínas fluorescentes ampliamente utilizadas. Con los avances recientes en la tecnología de supresión ámbar, ahora es posible dirigir fluoróforos pequeños, populares y disponibles comercialmente a residuos de proteínas específicas. Por medio de un par ortogonal ARN^t/aminoacil ARN^t sintetasa (ARN^t/RS) de *Methanosarcina mazei*, se incorporan genéticamente aminoácidos no naturales (UAA) que transportan cadenas laterales de alquino y alqueno en tensión en las posiciones codificadas por un codón de terminación ámbar (TAG) (A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. van Hest, F. L. van Delft, E. A.

Lemke, *Chembiochem* 2012, 13, 2094-2099; T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2011, 50, 3878-3881; T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2012, 51, 4166-4170; K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox, J. W. Chin, *Journal of the American Chemical Society* 2012, 134, 10317-10320; S. Schneider, M. J. Gattner, M. Vrabel, V. Flugel, V. Lopez-Carrillo, S. Prill, T. Carell, *Chembiochem* 2013, 14, 2114-2118; documento WO 2012/104422). Estas modificaciones añaden solo unos pocos átomos a la cadena lateral de aminoácidos y se pueden colocar libremente dentro de la proteína, lo que reduce el riesgo de impacto funcional. Posteriormente, los UAA de alquinos y alquenos en tensión pueden someterse a reacciones de cicloadición de alquino-azida promovida por cepas libres de catalizador (SPAAC) y cicloadición 4+2 de Diels-Alder [de demanda de electrones inversa promovida por cepas] (SPIEDAC) con fluoróforos orgánicos que contienen funcionalidades azida o tetrazina (Tet), respectivamente. Ambas reacciones son totalmente biocompatibles. Además, son ortogonales entre sí, ya que las azidas solo reaccionan con alquinos pero no con alquenos (Y. Liang, J. L. Mackey, S. A. Lopez, F. Liu, K. N. Houk, *Journal of the American Chemical Society* 2012, 134, 17904-17907; M. R. Karver, R. Weissleder, S. A. Hilderbrand, *Angew Chem Int Ed Engl* 2011.).

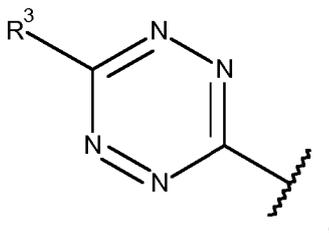
Mientras que la codificación de un solo UAA se ha vuelto relativamente sencilla y se ha descrito la incorporación de más de un UAA (documento US 2010/297693 A1; Han Xiao, et al., *Angew Chem Int Ed Engl* 2013, 52, 14080-14083), todavía existe una demanda de estrategias de etiquetado de múltiples colores robustas y eficientes en sistemas de mamíferos. Se pueden concebir al menos dos estrategias distintas para el etiquetado de dos colores basado en UAA y SRM, que sirven para diferentes diseños experimentales: i) Incorporación simultánea de dos UAA diferentes, que albergan dos químicas ortogonales (por ejemplo, SPAAC y SPIEDAC), reconociendo cada una un codón diferente en una sola proteína (por ejemplo, para la transferencia de energía por resonancia de Forster - estudios FRET) o en dos proteínas diferentes (por ejemplo, para la microscopía de localización de dos moléculas diferentes), ii) codificación secuencial de dos UAA diferentes, que albergan dos químicas ortogonales, en respuesta al mismo codón utilizando un solo sistema de ARN^t/RS. Esto se puede hacer en una forma de pulso-caza donde el primer UAA suministrado al medio de crecimiento sea perseguido entonces por el segundo UAA. Esto puede ayudar, por ejemplo, a visualizar la clasificación de proteínas.

A pesar de grandes esfuerzos, todavía existe una gran demanda de estrategias para facilitar el etiquetado específico de sitio de proteínas *in vitro* e *in vivo* y estrategias de etiquetado de múltiples colores fuertes en sistemas de mamíferos en particular. Por razones prácticas, sería útil que las reacciones de acoplamiento bioortogonales se desarrollaran con una cinética extremadamente rápida ($k > 10^2 \text{ M}^{-1} \text{ s}^{-1}$) y una alta especificidad. La mejora de la cinética minimizaría tanto el tiempo como la cantidad de agente de etiquetado requerido para mantener altos rendimientos de acoplamiento. Por lo tanto, fue un objetivo de la presente invención diseñar reacciones de acoplamiento bioortogonales extremadamente rápidas entre tetrazinas y dienófilos que permitieran establecer estrategias de etiquetado múltiple. Más específicamente, era un objetivo de la presente invención proporcionar aminoácidos o análogos de los mismos que pudieran incorporarse de forma traduccional en cadenas polipeptídicas y permitieran el etiquetado del polipéptido resultante *in vitro* e *in vivo*, así como establecer estrategias de etiquetado múltiple.

60 Sumario de la invención

La presente invención se refiere a métodos para formar uniones mediante reacciones de cicloadición, en donde el método comprende hacer reaccionar una primera tetrazina con un primer dienófilo seguido de hacer reaccionar una segunda tetrazina con un segundo dienófilo, en donde la reacción de la primera tetrazina con el primer dienófilo se desarrolla en presencia del segundo dienófilo, en donde

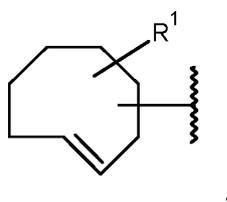
(i) la primera tetrazina comprende un grupo de la fórmula:



5 en la que

R^3 es alquilo C₁-C₃;

10 (ii) el primer dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:

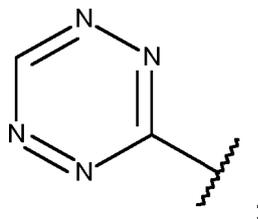


en la que

15 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

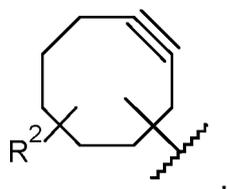
20 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

(iii) la segunda tetrazina comprende un grupo de la fórmula:



25 y

(iv) el segundo dienófilo comprende un grupo ciclooctinilo de la fórmula:



30 en la que

35 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

Por consiguiente, en el método de la invención, la primera tetrazina reacciona con el primer dienófilo que comprende el grupo *trans*-ciclooctenilo, y la segunda tetrazina reacciona con el segundo dienófilo que comprende el grupo

ciclooctinilo. Las constantes de velocidad de ambas reacciones son altas y, sin embargo, la primera reacción se desarrolla en presencia del segundo dienófilo sin que la primera tetrazina reaccione sustancialmente con el segundo dienófilo. Por lo tanto, la primera tetrazina se hace reaccionar con el primer dienófilo en condiciones que no permiten que la primera tetrazina reaccione sustancialmente con el segundo dienófilo. Dicho de otra manera, la primera tetrazina reacciona preferiblemente con el primer dienófilo en presencia del segundo dienófilo. Más específicamente, la primera tetrazina reacciona específicamente con el primer dienófilo en presencia del segundo dienófilo. Por lo tanto, los métodos de la invención permiten la formación de dos uniones diferentes a altas velocidades y aún con alta especificidad proporcionando primero un primer y un segundo dienófilo y luego poniendo en contacto el primer y segundo dienófilos con una primera tetrazina para que reaccione el primer dienófilo con la primera tetrazina en condiciones que no permiten que la primera tetrazina reaccione sustancialmente con el segundo dienófilo.

El tipo de reacciones realizadas de acuerdo con la invención permite conseguir constantes de velocidad k de $50 \text{ M}^{-1} \text{ s}^{-1}$ o más, $1 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ o más, o $2 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ o más para ambas reacciones.

De acuerdo con un aspecto de la invención, la primera reacción (de la primera tetrazina con el primer dienófilo) se desarrolla a constantes de velocidad k de $5 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ o más, $1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ o más, o $2 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ o más.

De acuerdo con un segundo aspecto de la invención, la segunda reacción (de la segunda tetrazina con el segundo dienófilo) se desarrolla a constantes de velocidad k de $50 \text{ M}^{-1} \text{ s}^{-1}$ o más, $1 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ o más, o $2 \times 10^2 \text{ M}^{-1} \text{ s}^{-1}$ o más.

De acuerdo con una realización de la invención, la primera y segunda reacciones se desarrollan a constantes de velocidad según dicho primer y segundo aspecto.

Para que la primera reacción (de la primera tetrazina con el primer dienófilo) se desarrolle de manera conveniente en presencia del segundo dienófilo, la primera reacción se produce a una velocidad más alta que la reacción de la primera tetrazina con el segundo dienófilo. Por consiguiente, la constante de velocidad k de la primera reacción suele ser al menos 100 veces mayor que la constante de velocidad k de la primera tetrazina con el segundo dienófilo. Por consiguiente, la reacción de la primera tetrazina con el segundo dienófilo usualmente se desarrolla a constantes de velocidad k de $5 \text{ M}^{-1} \text{ s}^{-1}$ o menos, $2 \text{ M}^{-1} \text{ s}^{-1}$ o menos, o $1 \text{ M}^{-1} \text{ s}^{-1}$ o menos. Es preferible que la constante de velocidad k de la primera reacción sea al menos 200 veces, 500 veces, 1000 veces, 5000 veces, o 10000 veces mayor que la constante de velocidad k de la primera tetrazina con el segundo dienófilo. Por consiguiente, la reacción de la primera tetrazina con el segundo dienófilo preferiblemente se desarrolla a constantes de velocidad k de $0,5 \text{ M}^{-1} \text{ s}^{-1}$ o menos, $0,1 \text{ M}^{-1} \text{ s}^{-1}$ o menos, $0,05 \text{ M}^{-1} \text{ s}^{-1}$ o menos, o $0,01 \text{ M}^{-1} \text{ s}^{-1}$ o menos.

Los métodos para determinar las constantes de reacción se conocen bien en la técnica y las constantes de reacción absolutas descritas en el presente documento se refieren en particular a la determinación descrita en el ejemplo J en el presente documento.

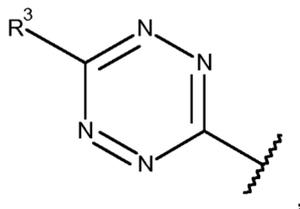
Ambas reacciones pueden realizarse en un amplio intervalo de temperaturas. En sistemas biológicos, temperaturas en el intervalo de $4 \text{ }^\circ\text{C}$ a $45 \text{ }^\circ\text{C}$, por ejemplo, en el intervalo de $15 \text{ }^\circ\text{C}$ a $25 \text{ }^\circ\text{C}$ tal como aproximadamente $22 \text{ }^\circ\text{C}$, o en el intervalo de $30 \text{ }^\circ\text{C}$ a $40 \text{ }^\circ\text{C}$ tal como aproximadamente $37 \text{ }^\circ\text{C}$, pueden usarse de manera conveniente.

La primera reacción normalmente se deja avanzar durante un tiempo suficiente para permitir que sustancialmente todos los grupos *trans*-ciclooctenilo (el primer dienófilo) reaccionen con la primera tetrazina antes de realizar la segunda reacción. Normalmente, los tiempos de reacción de unos pocos minutos, por ejemplo, 60, 45 o 30 minutos o menos, y en particular de 10 minutos o menos, son convenientes para que la primera reacción se desarrolle de manera conveniente. Sin embargo, se permite que la primera reacción se desarrolle solo en la medida en que no haya sustancialmente reacción de la primera tetrazina con el segundo dienófilo (es decir, los grupos ciclooctinilo). Normalmente, los tiempos de reacción de menos de 30 minutos, por ejemplo, menos de 25, 20 o 15 minutos, y en particular menos de 10 minutos, son convenientes para que la primera reacción se desarrolle sin sustancialmente ningún segundo dienófilo que reaccione con la primera tetrazina.

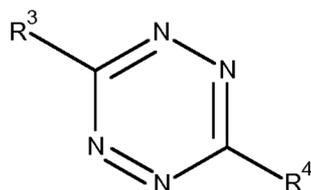
Normalmente, se permite que la segunda reacción se desarrolle durante un tiempo suficiente para permitir que sustancialmente todos los grupos ciclooctinilo (el segundo dienófilo) reaccionen con la segunda tetrazina. Normalmente, los tiempos de reacción de 2 horas o menos, por ejemplo, 1 hora o menos, 45 minutos o menos, o 30 minutos o menos, y en particular de 20 minutos o menos, son convenientes para que la segunda reacción se desarrolle completamente. Debido a que la segunda reacción generalmente se realiza solo una vez que todos los grupos *trans*-ciclooctenilo (el primer dienófilo) han reaccionado con la primera tetrazina, los tiempos de reacción de la segunda reacción no tienen que tener en cuenta otra reacción de cicloadición potencialmente competitiva adicional. Sin embargo, normalmente es conveniente que la segunda reacción se deje desarrollar durante menos de 1 hora, por ejemplo, menos de 50, 40 o 30 minutos, y en particular, durante menos de 20 minutos.

Los tiempos de reacción descritos en el presente documento se aplican en particular a las temperaturas de reacción descritas en el presente documento, en particular a reacciones realizadas a temperatura ambiente, tal como aproximadamente $37 \text{ }^\circ\text{C}$.

Si la primera o la segunda reacción no se completa, es posible hacer reaccionar cualquier primer o segundo dienófilo restante de una manera apropiada para convertir estos grupos a una forma que ya no sea reactiva con la primera o la segunda tetrazina. Esto es especialmente conveniente para que se complete la primera reacción, evitando así que el primer dienófilo restante reaccione con la segunda tetrazina. Por ejemplo, si la primera reacción no se desarrolla hasta completarse, quedando algunos grupos *trans*-ciclooctenilo sin reaccionar (el primer dienófilo), estos grupos pueden reaccionar con la primera tetrazina adicional (inactivador), en los que la primera tetrazina adicional comprende un grupo de fórmula:



pero es diferente de la primera tetrazina inicialmente utilizada. Por ejemplo, si la primera tetrazina utilizada inicialmente comprende una etiqueta, se puede usar una tetrazina correspondiente sin etiqueta como la primera tetrazina adicional. Por ejemplo, el inactivador puede ser un compuesto de la fórmula:



o una sal de adición de ácidos del mismo, en la que

R^3 es alquilo C_1-C_3 ; y

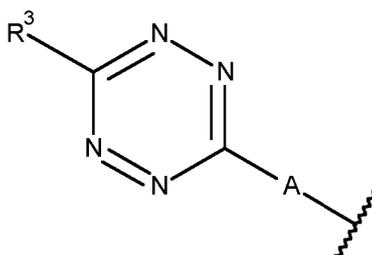
R^4 es un radical orgánico (tal como un grupo bencilo) que es diferente del resto en la posición correspondiente en la primera tetrazina usada inicialmente.

De acuerdo con una realización, el inactivador es (4-(6-metil-1,2,4,5-tetrazina-3-il)fenil)metanamina o una sal de adición de ácidos de la misma.

El desarrollo de esta manera puede ser ventajoso en situaciones en las que tener que la primera reacción desarrollada hasta completarse se asocia con desventajas (porque tendría que usarse una cantidad relativamente alta de la primera tetrazina o los tiempos de reacción convenientes no permiten que la reacción continúe hasta el final), y el uso de la primera tetrazina adicional se asocia con ventajas (porque ayuda a mantener la cantidad de la primera tetrazina a usar ser relativamente baja al tiempo que permite que la reacción avance hasta completarse dentro de los tiempos de reacción convenientes). Por ejemplo, aunque puede ser desventajoso usar una cantidad excesivamente alta de la primera tetrazina (por ejemplo, porque es costosa), la primera tetrazina adicional puede no tener las mismas desventajas y, por lo tanto, puede usarse en cantidades relativamente altas para completar la reacción dentro de tiempos de reacción convenientes). Por lo tanto, permitir que sustancialmente todos los grupos *trans*-ciclooctenilo (el primer dienófilo) reaccionen con la primera tetrazina significa que al menos el 95, por ejemplo, al menos el 96, 97, 98 o el 99 % de los grupos *trans*-ciclooctenilo presentes inicialmente han reaccionado con la primera tetrazina, mientras que los grupos *trans*-ciclooctenilo restantes pueden hacerse reaccionar opcionalmente con una primera tetrazina adicional para que todos los grupos *trans*-ciclooctenilo (es decir, al menos el 99,9 % de los grupos *trans*-ciclooctenilo inicialmente presentes) hayan reaccionado con la primera tetrazina (adicional).

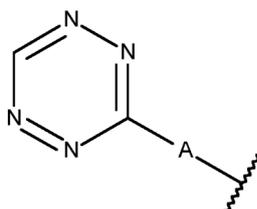
Puede ser conveniente eliminar la primera tetrazina sin reaccionar antes de realizar la segunda reacción. Esto se puede hacer de una manera conocida per se. Asimismo, puede ser conveniente eliminar la primera tetrazina (inactivador) sin reaccionar adicional antes de realizar la segunda reacción. Esto, además, puede hacerse de una manera conocida per se.

De acuerdo con un aspecto particular de la invención, la primera tetrazina comprende un grupo de fórmula:



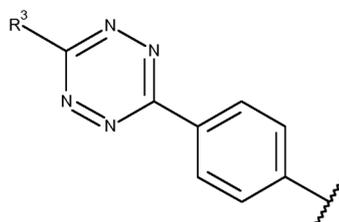
en la que R³ es alquilo C₁-C₃ y A es 1,4-fenileno o alquileno C₁-C₆.

- 5 De acuerdo con un aspecto particular adicional de la invención, la segunda tetrazina comprende un grupo de fórmula:



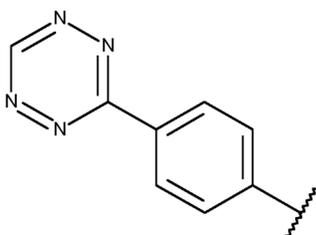
en la que A es 1,4-fenileno o alquileno C₁-C₆.

- 10 A es un grupo o parte de un grupo que une la tetrazina a un resto adicional, por ejemplo, una etiqueta. A puede ser el mismo grupo en la primera y la segunda tetrazina, o A de la primera tetrazina puede ser diferente de A de la segunda tetrazina. En particular, A es 1,4-fenileno.
- 15 Por lo tanto, de acuerdo con una realización particular de la invención, la primera tetrazina comprende un grupo de fórmula:



- 20 en la que R³ es alquilo C₁-C₃.

Asimismo, de acuerdo con una realización particular de la invención, la segunda tetrazina comprende un grupo de fórmula:

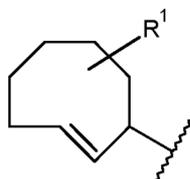


- 25 Debido a que la primera tetrazina lleva el sustituyente R³, la reactividad de la primera tetrazina con el primer dienófilo es suficientemente diferente de la reactividad de la primera tetrazina con el segundo dienófilo. Por consiguiente, R³ se selecciona para proporcionar reactividades suficientemente diferentes. De acuerdo con la invención, R³ es preferiblemente metilo o etilo. En particular, se prefiere metilo.
- 30

El punto de unión del grupo *trans*-ciclooctenilo o ciclooctinilo puede ser por un átomo de anillo en posición α, β o γ con respecto al doble o triple enlace.

De acuerdo con una realización particular, el grupo *trans*-ciclooctenilo o ciclooctinilo está unido por el átomo del anillo en la posición α con respecto al doble o triple enlace.

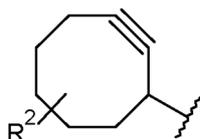
5 Por lo tanto, de acuerdo con un aspecto particular de la invención, el primer dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:



en la que

10 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 15 R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

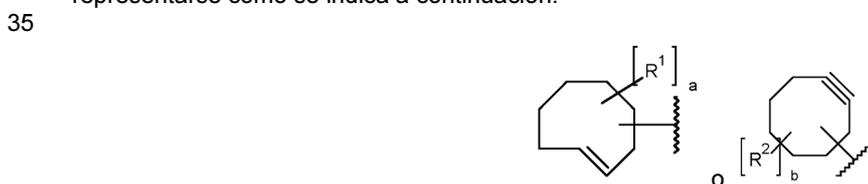
De acuerdo con un aspecto particular adicional de la invención, el segundo dienófilo comprende un grupo ciclooctinilo de la fórmula:



20 en la que

25 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 R^c , R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

30 El grupo *trans*-ciclooctenilo o ciclooctinilo puede estar sin sustituir (es decir, R^1 y R^2 son hidrógeno) o sustituido con uno o más de un radical R^1 o R^2 , respectivamente. Por lo tanto, puede haber uno o más de un sustituyente R^1 y/o R^2 . Más particularmente, puede haber hasta 5, por ejemplo, 1, 2 o 3, sustituyentes R^1 . Asimismo, puede haber hasta 5, por ejemplo, 1, 2 o 3, sustituyentes R^2 . Por lo tanto, los grupos *trans*-ciclooctenilo o ciclooctinilo de la invención pueden representarse como se indica a continuación:



en las que a es cero, 1, 2, 3, 4 o 5, y b es cero, 1, 2, 3, 4 o 5.

40 Si hay más de un radical R^1 , estos pueden ser radicales iguales o diferentes y dos radicales R^1 pueden estar enlazados al mismo átomo o a átomos diferentes. Asimismo, 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^1 y/o R^2 pueden ser dos átomos de flúor unidos a un átomo del anillo de carbono.

45 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino, siendo R^a y R^b independientemente hidrógeno o alcanoiloximetilo C₂-C₅.

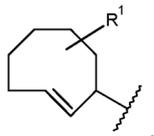
50 De acuerdo con una realización particular, R^1 es hidrógeno.

5 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarbonilo C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino, siendo R^c y R^d independientemente hidrógeno o alcanoiloximetilo C₂-C₅.

De acuerdo con una realización particular, R² es hidrógeno.

10 De acuerdo con una realización particular adicional, R¹ es hidrógeno y R² es hidrógeno.

De acuerdo con una realización particular adicional de la invención, el grupo *trans*-ciclooctenilo tiene la fórmula:

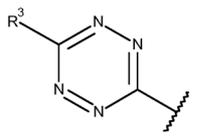


15 en la que

R¹ es hidrógeno;

y

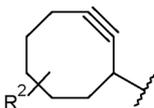
20 la primera tetrazina comprende un grupo de fórmula:



en la que

25 R³ es metilo.

De acuerdo con una realización particular adicional de la invención, el grupo ciclooctinilo tiene la fórmula:

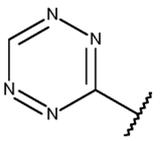


30 en la que

35 R¹ es hidrógeno;

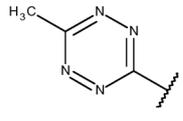
y

la segunda tetrazina comprende un grupo de fórmula:



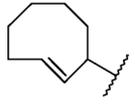
40 Por lo tanto, la presente invención se refiere en particular a métodos para formar uniones mediante reacciones de cicloadición, en donde el método comprende hacer reaccionar una primera tetrazina con un primer dienófilo seguido de hacer reaccionar una segunda tetrazina con un segundo dienófilo, en donde la reacción de la primera tetrazina con el primer dienófilo se desarrolla en presencia del segundo dienófilo, en donde

45 (i) la primera tetrazina comprende un grupo de la fórmula:



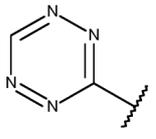
(ii) el primer dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:

5



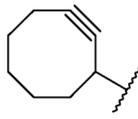
(iii) la segunda tetrazina comprende un grupo de la fórmula:

10



y
(iv) el segundo dienófilo comprende un grupo ciclooctinilo de la fórmula:

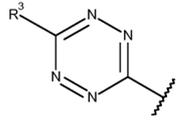
15



La presente invención también se refiere a kits que pueden usarse para realizar los métodos de la invención.

Los kits particulares de la invención incluyen:

20 (i) una primera tetrazina que comprende un grupo de la fórmula:



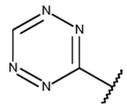
en la que

25

R³ es alquilo C₁-C₃;

y

30 (ii) una segunda tetrazina, que comprende un grupo de la fórmula:

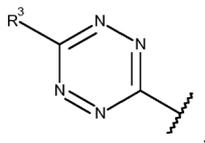


35

Dicha primera y segunda tetrazinas son en particular agentes que pueden usarse para etiquetar moléculas diana o composiciones de moléculas diana y, por lo tanto, se denominan en el presente documento también agentes de etiquetado. Estos incluyen, en particular, los agentes de etiquetado que se describen en el presente documento.

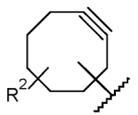
Otros kits particulares de la invención incluyen:

40 (i) un primer agente que comprende un grupo *trans*-ciclooctenilo de la fórmula:



en la que

- 5 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 10 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y (ii) un segundo agente que comprende un grupo ciclooctinilo de la fórmula:



en la que

- 15 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 20 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

Dicho primer y segundo agentes son en particular agentes que pueden usarse para modificar moléculas diana o composiciones de moléculas diana y, por lo tanto, se denominan en el presente documento también agentes de modificación. Estos incluyen en particular los aminoácidos no naturales y sus análogos como se describe en el presente documento.

Además, los kits de la invención pueden incluir la primera y la segunda tetrazina como se describe en el presente documento, así como el primer agente que comprende un grupo *trans*-ciclooctenilo y el segundo agente que comprende un grupo ciclooctinilo.

Los métodos de la invención son especialmente adecuados para unir una gran variedad de moléculas entre sí o ligadas entre sí. Por ejemplo, entre estas moléculas hay polipéptidos, oligonucleótidos, glicanos, lípidos, tintes, agentes terapéuticos, agentes de diagnóstico, agentes quelantes/formadores de complejo, superficies en fase sólida, nanopartículas, puntos cuánticos.

Los métodos de la invención son especialmente adecuados para el etiquetado múltiple, es decir, al menos doble, de moléculas uniendo secuencialmente la primera tetrazina al primer dienófilo seguido de la segunda tetrazina al segundo dienófilo. Por consiguiente, la primera tetrazina es un primer agente de etiquetado y la segunda tetrazina es un segundo agente de etiquetado.

Por lo tanto, la presente invención se refiere a métodos para etiquetar moléculas, comprendiendo el método poner en contacto una molécula diana o una composición de molécula diana con

- (i) un primer agente de etiquetado que comprende un grupo de la fórmula:

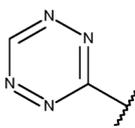


en la que

- 50 R^3 es alquilo C₁-C₃;

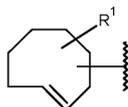
seguido de

- (ii) un segundo agente de etiquetado que comprende un grupo de la fórmula:



en la que la molécula diana comprende

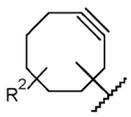
- 5 (i) un grupo *trans*-ciclooctenilo de la fórmula:



en la que

- 10 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 15 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

- (ii) un grupo ciclooctinilo de la fórmula:

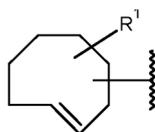


20 en la que

- R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 25 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

en la que la composición de moléculas diana comprende

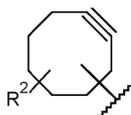
- 30 (i) un primera molécula diana que comprende un grupo *trans*-ciclooctenilo de la fórmula:



en la que

- 35 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 40 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

- (ii) una segunda molécula diana que comprende un grupo ciclooctinilo de la fórmula:



45 en la que

- R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃,

CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 R^c, R^d independientemente son hidrógeno o alcanoloximetilo C₂-C₅.

5 Los agentes de etiquetado de la invención pueden comprender cualquier etiqueta, siempre que la etiqueta del primer agente de etiquetado sea diferente de la etiqueta del segundo agente de etiquetado. Las etiquetas de la invención incluyen tintes (por ejemplo, tintes fluorescentes, luminiscentes o fosforescentes, tales como tintes de dansilo, cumarina, fluoresceína, acridina, rodamina, silicio-rodamina, BODIPY o cianina), cromóforos (por ejemplo, fitocromo, 10 ficobilina, bilirrubina, etc.), radiomarcadores (por ejemplo, formas radiactivas de hidrógeno, flúor, carbono, fósforo, azufre, o yodo, tal como tritio, flúor-18, carbono-11, carbono-14, fósforo-32, fósforo-33, azufre-33, azufre-35, yodo-123, o yodo-125), etiquetas de giro sensibles a la RM, etiquetas de afinidad (por ejemplo, biotina, etiqueta His, etiqueta Flag, etiqueta strep, azúcares, lípidos, esteroides, enzimas, grupos de PEG, bencilguaninas, bencilcitosinas, o co-factores), 15 grupos polietilenglicol (por ejemplo, un PEG ramificado, un PEG lineal, PEG de diferentes pesos moleculares, etc.), fotoreticuladores (como p-azidoyodoacetanilida), sondas de RMN, sondas de rayos X, sondas de pH, sondas IR, resinas, soportes sólidos y compuestos bioactivos (por ejemplo, fármacos sintéticos).

En algunas realizaciones, los tintes pueden incluir un agente de contraste NIR que emite fluorescencia en la región del infrarrojo cercano del espectro. Los fluoróforos de infrarrojo cercano a modo de ejemplo pueden incluir tintes y 20 otros fluoróforos con longitudes de onda de emisión (por ejemplo, longitudes de onda de emisión máximas) entre aproximadamente 630 y 1000 nm, por ejemplo, entre aproximadamente 630 y 800 nm, entre aproximadamente 800 y 900 nm, entre aproximadamente 900 y 1000 nm, entre aproximadamente 680 y 750 nm, entre aproximadamente 750 y 800 nm, entre aproximadamente 800 y 850 nm, entre aproximadamente 850 y 900 nm, entre aproximadamente 900 y 950 nm, o entre aproximadamente 950 y 1000 nm. Los fluoróforos con longitudes de onda de emisión (por ejemplo, 25 longitudes de onda de emisión máximas) superiores a 1000 nm también se pueden usar en los métodos descritos en el presente documento.

Los fluoróforos útiles en los presentes métodos incluyen: ácido 7-amino-4-metilcumarin-3-acético (AMCA), TEXAS RED™ (Molecular Probes, Inc., Eugene, Oreg.), 5-(y -6)-carboxi-X-rodamina, lisamina rodamina B, 5-(y -6)- 30 carboxifluoresceína, fluoresceína-5-isotiocianato (FITC), ácido 7-dietilaminocumarin-3-carboxílico, tetrametilrodamina-5-(y -6)-isotiocianato, 5-(y -6)-carboxitetrametilrodamina, ácido 7-hidroxycumarin-3-carboxílico, ácido 6-[fluoresceína 5-(y -6)-carboxamido]hexanoico, ácido N-(4,4-difluoro-5,7-dimetil-4-bora-3a,4a diaza-3-indacenopropiónico, eosina-5-isotiocianato, eritrosina-5-isotiocianato, y azul acetilazida CASCADE™ (Molecular Probes, Inc., Eugene, Oreg.) y tintes ATTO.

Los compuestos bioactivos que pueden unirse a un polipéptido de la invención incluyen compuestos citotóxicos (por ejemplo, compuestos quimioterapéuticos para el cáncer); compuestos antiviricos; modificadores de la respuesta biológica (por ejemplo, hormonas, quimiocinas, citocinas, interleucinas, etc.); agentes que afectan a los microtúbulos; 40 moduladores de hormonas; y compuestos esteroideos.

Una molécula diana de la invención puede ser cualquier molécula que se va a marcar. Las moléculas diana de la invención incluyen polipéptidos, oligonucleótidos, glicanos y lípidos.

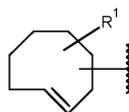
Los métodos de la invención se pueden usar para marcar la misma molécula diana (por ejemplo, un polipéptido), de 45 manera que la molécula diana comprenda ambas etiquetas primera y segunda, o se puedan marcar dos moléculas diana diferentes de manera que una primera molécula diana comprenda la primera etiqueta y una segunda molécula diana comprenden la segunda etiqueta. Si se marcan dos moléculas diana diferentes, las moléculas diana pueden ser del mismo tipo (por ejemplo, dos polipéptidos) o de diferentes tipos (por ejemplo, un polipéptido y un oligonucleótido). El etiquetado de dos moléculas diana diferentes del mismo tipo (por ejemplo, dos polipéptidos) incluye el etiquetado 50 de dos poblaciones del mismo polipéptido que son esencialmente idénticas, pero para el dienófilo. Dichos polipéptidos se pueden obtener de varias maneras, por ejemplo, incorporando el primer y el segundo dienófilo en diferentes poblaciones del polipéptido en diferentes puntos temporales.

Por lo tanto, de acuerdo con una realización de la invención, el grupo *trans*-ciclooctenilo está unido a un residuo de 55 aminoácido de un polipéptido diana, y el grupo ciclooctinilo está unido a un residuo de aminoácido de un polipéptido diana. De acuerdo con una realización particular de la invención, el grupo *trans*-ciclooctenilo y el grupo ciclooctinilo están unidos a residuos de aminoácidos de un polipéptido diana, en donde el resto de aminoácido al que está unido el grupo *trans*-ciclooctenilo es normalmente diferente del residuo de aminoácido al que está unido el grupo ciclooctinilo. De acuerdo con una realización particular adicional de la invención, el grupo *trans*-ciclooctenilo está unido a un residuo 60 de aminoácido de un primer polipéptido diana, y el grupo ciclooctinilo está unido a un residuo de aminoácido de un segundo polipéptido diana.

Si se va a marcar una primera y una segunda molécula diana, ambas forman parte de la composición que se somete a las reacciones de etiquetado. Las composiciones de la invención incluyen sistemas biológicos, tales como 65 organismos o muestras biológicas. De acuerdo con una realización de la invención, la muestra biológica comprende una célula.

Por lo tanto, la presente invención también se refiere a sistemas biológicos, por ejemplo, organismos o muestras biológicas, y en particular células, que comprenden

5 (i) un grupo *trans*-ciclooctenilo de la fórmula:



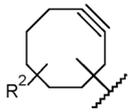
en la que

10 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

15 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

y

(ii) un grupo ciclooctinilo de la fórmula:



20 en la que

25 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

30 Los organismos de la invención incluyen todos los tipos de organismos, tales como procariotas y eucariotas. Por consiguiente, las células de la invención incluyen todos los tipos de células, tales como células procariotas o eucariotas. De acuerdo con una realización de la invención, es una célula eucariota, por ejemplo, una célula de mamífero.

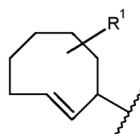
35 De acuerdo con un aspecto de la invención, los grupos *trans*-ciclooctenilo y ciclooctinilo pueden unirse a cualquier componente del sistema biológico, por ejemplo, componente celular, tales como polipéptidos, oligonucleótidos, glicanos y lípidos. Por consiguiente, las realizaciones particulares de la invención incluyen el grupo *trans*-ciclooctenilo que está unido a un polipéptido y el grupo ciclooctinilo que está unido a un polipéptido, en donde el grupo *trans*-ciclooctenilo y el grupo ciclooctinilo están unidos al mismo polipéptido, o el grupos *trans*-ciclooctenilo está unido a un primer polipéptido y el grupo ciclooctinilo está unido a un segundo polipéptido, el primer y el segundo polipéptido son polipéptidos diferentes.

De acuerdo con una realización particular, una célula de la invención se prepara por un método que comprende

45 a) proporcionar una célula que comprende:

- (i) una primera aminoacil ARNt sintetasa, o un polinucleótido que la codifica; y opcionalmente una segunda aminoacil ARNt sintetasa, o un polinucleótido que la codifica;
- 50 (ii) un primer ARNt que tiene un anticodón para un primer codón selector, o un polinucleótido que codifica dicho ARNt; y opcionalmente un segundo ARNt que tiene un anticodón para un segundo codón selector, o un polinucleótido que codifica dicho ARNt; y y
- (iii) un polinucleótido que codifica un polipéptido diana y que comprende uno o más de un primer y segundo codón o codones selectores; o un primer polinucleótido que codifica un primer polipéptido diana y que comprende uno o más de un primer codón selector y un segundo polinucleótido que codifica un segundo polipéptido diana y que comprende uno o más de un segundo codón selector,

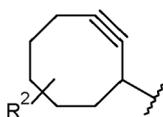
en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural o un análogo del mismo, que comprende un grupo *trans*-ciclooctenilo de fórmula:



en la que

- 5 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 10 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

y con un segundo aminoácido no natural o un análogo del mismo, que comprende un grupo ciclooctinilo de la fórmula:



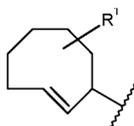
15

en la que

- 20 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

o

- 25 en la que dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural o un análogo del mismo, que comprende un grupo *trans*-ciclooctenilo de fórmula:



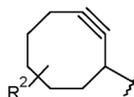
30

en la que

- 35 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

dicha segunda aminoacil ARNt sintetasa (i) es capaz de acilar el segundo ARNt (ii) con un segundo aminoácido no natural o un análogo del mismo, que comprende un grupo ciclooctinilo de la fórmula:

40



en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

b) poner en contacto la célula con el primer y el segundo aminoácido no natural o un análogo del mismo; y

c) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el primer y el segundo aminoácidos no naturales o los análogos de los mismos, en el polipéptido o polipéptidos diana en la posición o posiciones codificadas por el codón o codones selectores.

De acuerdo con una realización de la invención, la célula se pone en contacto con el primer y el segundo aminoácidos no naturales o los análogos de los mismos, secuencialmente. Por consiguiente, el método de la invención comprende:

a) poner en contacto la célula con el primer aminoácido no natural o el análogo del mismo; y

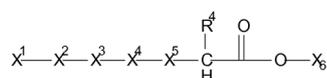
b) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el primer aminoácido no natural o el análogo del mismo en el polipéptido diana en la posición o posiciones codificadas por el codón o codones selectores;

c) poner en contacto la célula con el segundo aminoácido no natural o el análogo del mismo; y

d) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el segundo aminoácido no natural o el análogo del mismo en el polipéptido diana en la posición o posiciones codificadas por el codón o codones selectores.

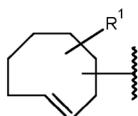
En esta realización, puede usarse un solo par aminoacil ARNt sintetasa/ARNt para incorporar el primer y el segundo aminoácido no natural o el análogo del mismo, en diferentes poblaciones de un polipéptido en diferentes puntos de tiempo.

De acuerdo con un aspecto de la invención, el primer aminoácido no natural o el análogo del mismo es un compuesto de la fórmula:



en la que:

X¹ tiene la fórmula:



R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquilenos C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

X⁵ es -(CH₂)_n- o fenileno-CH₂-;

X⁶ es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilulfanil C₂-C₇-alquilo C₁-C₂;

R⁴ es -OH o -NH₂;

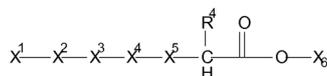
n es un número entero de 1 a 4;

m es un número entero de 1 a 6; y

p es un número entero de 1 a 6,

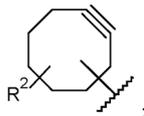
o una sal de adición de ácidos o bases del mismo.

De acuerdo con un aspecto adicional de la invención, el segundo aminoácido no natural o el análogo del mismo es un compuesto de la fórmula:



5 en la que:

X¹ tiene la fórmula:



10 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileno C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

20 X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

X⁵ es -(CH₂)_n- o fenileno-(CH₂)_n-;

X⁶ es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilsulfanil C₂-C₇-alquilo C₁-C₂;

R⁴ es -OH o -NH₂;

25 n es un número entero de 1 a 4;

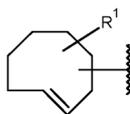
m es un número entero de 1 a 6; y

p es un número entero de 1 a 6,

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

30 La presente invención también se refiere a polipéptidos que comprenden

(i) un grupo *trans*-ciclooctenilo de la fórmula:



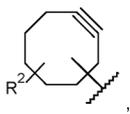
35 en la que

40 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

45 y

(ii) un grupo ciclooctinilo de la fórmula:



50 en la que

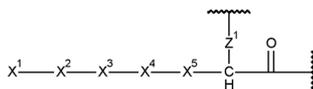
R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN,

hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

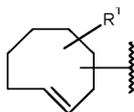
5

De acuerdo con un aspecto de la invención, los polipéptidos comprenden un residuo de la fórmula:



10 en la que:

X¹ tiene la fórmula



15

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

20 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

25 X⁵ es -(CH₂)_n- o fenileno-CH₂-;

Z¹ es -O- o -NH-;

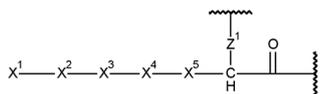
n es un número entero de 1 a 4;

m es un número entero de 1 a 6; y

p es un número entero de 1 a 6,

30

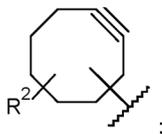
y un residuo de la fórmula:



en la que:

35

X¹ tiene la fórmula:



40 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

45 X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

50 X⁵ es -(CH₂)_n- o fenileno-CH₂-;

Z¹ es -O- o -NH-;

n es un número entero de 1 a 4;

m es un número entero de 1 a 6; y

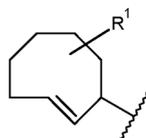
p es un número entero de 1 a 6.

Además, la presente invención también se refiere a un método para preparar el polipéptido de la invención, comprendiendo el método:

a) proporcionar un sistema de traducción que comprende:

- (i) una primera aminoacil ARNt sintetasa, o un polinucleótido que la codifica; y opcionalmente una segunda aminoacil ARNt sintetasa, o un polinucleótido que la codifica;
- (ii) un primer y un segundo aminoácido no natural o un análogo del mismo;
- (iii) un primer ARNt que tiene un anticodón para un primer codón selector, o un polinucleótido que codifica dicho ARNt; y opcionalmente un segundo ARNt que tiene un anticodón para un segundo codón selector, o un polinucleótido que codifica dicho ARNt; y y
- (iv) un polinucleótido que codifica un polipéptido diana y que comprende uno o más de un primer y segundo codón o codones selectores,

en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (iii) con el primer aminoácido no natural o el análogo del mismo (ii) que comprende un grupo *trans*-ciclooctenilo de la fórmula:

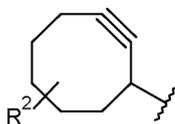


en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

dicha segunda aminoacil ARNt sintetasa (i) es capaz de acilar el segundo ARNt (iii) con el segundo aminoácido no natural o el análogo del mismo (ii) que comprende un grupo *trans*-ciclooctenilo de la fórmula:



en la que

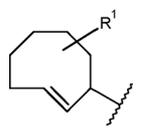
R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

b) permitir la traducción del polinucleótido o polinucleótidos (iv) incorporando de este modo el primer y el segundo aminoácidos no naturales o los análogos de los mismos, en el polipéptido en la posición o posiciones codificadas por el codón o codones selectores.

Los polipéptidos de la invención se pueden hacer reaccionar con una primera tetrazina seguida de una segunda tetrazina de acuerdo con los métodos descritos en el presente documento, por ejemplo, para introducir dos etiquetas diferentes en los sitios correspondientes a los aminoácidos no naturales que comprenden el grupo *trans*-ciclooctenilo y ciclooctinilo, respectivamente.

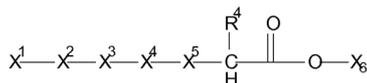
Aún adicionalmente, la presente invención se refiere a aminoácidos no naturales que comprenden un grupo *trans*-ciclooctenilo de la fórmula:



en la que

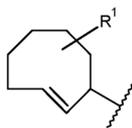
- 5 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 10 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,
o un análogo de dicho aminoácido no natural.

De acuerdo con un aspecto de la invención, dichos aminoácidos no naturales tienen la fórmula:



15 en la que

X^1 es un grupo *trans*-ciclooctenilo de la fórmula:



- 20 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 25 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.
 X^2 es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;
 X^3 es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p-, o un enlace sencillo;
 X^4 es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;
 X^5 es -(CH₂)_n- o fenileno-CH₂-;
30 X^6 es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilulfanil C₂-C₇-alquilo C₁-C₂;
 R^4 es -OH o -NH₂;
 n es un número entero de 0 a 4;
 m es un número entero de 1 a 6; y
35 p es un número entero de 1 a 6,

o una sal de adición de ácidos o bases de los mismos.

- 40 Con respecto a la capacidad de los aminoácidos no naturales de incorporarse traduccionalmente en una cadena polipeptídica, las variables $X^2, X^3, X^4, X^5, X^6, n, m, p$, y R^4 tienen, en particular, los siguientes significados que, cuando se toman en solitario o en combinación, representan realizaciones particulares de los aminoácidos no naturales descritos en el presente documento o cualquier otra fórmula en la que aparezcan estas variables.

45 X^2 es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NHC(O)- o -C(O)NH-.

Preferiblemente, X^2 es -O-.

X^3 es alquileo C₁-C₆, -(CH₂-CH₂-O)_m- o un enlace sencillo; y m es 1, 2, 3, 4, 5 o 6.

50 En relación con X^3 , alquileo C₁-C₆ se refiere preferiblemente a alquileo de cadena lineal.

Preferiblemente, X^3 es -CH₂-CH₂-O- o un enlace sencillo.

55 Como alternativa, X^3 es -(CH₂-O)_p-; y p es 1, 2, 3, 4, 5 o 6. De acuerdo con una realización particular, X^3 es -CH₂-O- (es decir, p es 1).

De acuerdo con un aspecto de la invención, 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 $-O-(CH_2)_2-O-$.

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 $-NH-C(NH)-NH-C(O)-$.

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

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

10 X^5 es $-(CH_2)_n-$ en la que n es como se define en el presente documento, o fenileno- CH_2- .

De acuerdo con una realización, X^5 es $-(CH_2)_n-$. n es un número entero de 1 a 4.

15 De acuerdo con un aspecto de la invención, n es 3 o 4.

De acuerdo con una realización preferida, n es 4.

De acuerdo con un aspecto adicional de la invención, n es 1.

20 De acuerdo con un aspecto adicional de la invención, X^5 es fenileno- CH_2- , en la que fenileno es en particular 1,3-fenileno o 1,4-fenileno. Se prefiere 1,4-Fenileno. El resto fenileno está unido a X^4 y el metileno al átomo de carbono que lleva R^4 (o Z^1).

25 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 las que n es preferiblemente 3 o 4, o n es 1.

30 De acuerdo con una realización preferida, $-X^4-(CH_2)_n-$ es $-C(O)-NH-(CH_2)_n-$, en las que n es preferiblemente 3 o 4, o n es 1.

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 $-C(O)-NH-C(NH)-NH-(CH_2)_3-$.

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

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

40 De acuerdo con una realización particular adicional, $-X^4$ -fenileno- CH_2- es $-NH$ -fenileno- CH_2- , $-NH$ -fenileno- CH_2- , $-NH-CH(NH_2)$ -fenileno- CH_2- , $-NH-C(NH)-NH$ -fenileno- CH_2- , $-C(O)-NH-CH(NH_2)$ -fenileno- CH_2- o $-C(O)-NH-C(NH)-NH$ -fenileno- CH_2- , en las que fenileno es preferiblemente 1,4-fenileno.

45 De acuerdo con un aspecto particular de la invención, $-X^2-X^3-X^4-$ comprende una funcionalidad 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 $-C(O)-NH-$).

50 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 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 $X^1-O-(CH_2)_2-O-C(O)-NH-(CH_2)_4-$.

55 De acuerdo con una realización preferida adicional, $X^1-X^2-X^3-X^4-(CH_2)_n-$ es $X^1-O-C(O)-NH-CH_2-$, $X^1-O-CH_2-O-C(O)-NH-CH_2-$ o $X^1-O-(CH_2)_2-O-C(O)-NH-CH_2-$.

60 De acuerdo con una realización preferida adicional, $X^1-X^2-X^3-X^4$ -fenileno- CH_2- es $X^1-O-C(O)-NH$ -fenileno- CH_2- , $X^1-O-CH_2-O-C(O)-NH$ -fenileno- CH_2- o $X^1-O-(CH_2)_2-O-C(O)-NH$ -fenileno- CH_2- , en las que fenileno es preferiblemente 1,4-fenileno.

X^6 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 .

65 De acuerdo con una realización particular, X^6 es hidrógeno, alcoximetilo C_1-C_6 , alcoxi-1-ilo C_1-C_6 (especialmente 1-

(alcoxi C₁-C₆)et-1-ilo), alcanoiloximetilo C₂-C₇ o alcanoilulfaniletilo C₂-C₇.

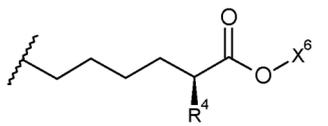
De acuerdo con una realización preferida, X⁶ es hidrógeno.

5 R⁴ es -OH o -NH₂. Preferiblemente, R⁴ es -NH₂.

Con respecto al átomo de carbono asimétrico que porta R⁴ (y Z¹) el aminoácido no natural o su análogo de la invención puede tener una configuración S o R (de acuerdo con las reglas de prioridad de Cahn-Ingold-Prelog), prefiriéndose la configuración S.

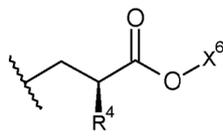
10

De acuerdo con una realización preferida, -X⁵-CHR⁴-C(O)O-X⁶ tiene la fórmula



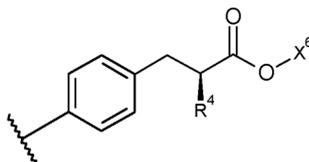
15 en la que R⁴ y X⁶ son como se definen en el presente documento y X⁶ es en particular hidrógeno.

De acuerdo con una realización preferida adicional, -X⁵-CHR⁴-C(O)O-X⁶ tiene la fórmula



20 en la que R⁴ y X⁶ son como se definen en el presente documento y X⁶ es en particular hidrógeno.

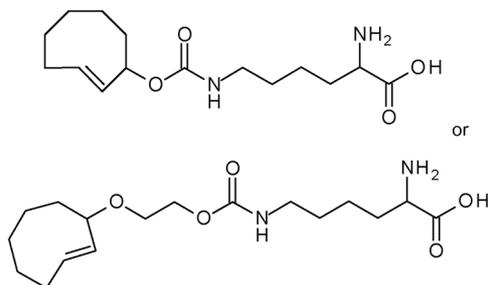
De acuerdo con una realización preferida adicional, -X⁵-CHR⁴-C(O)O-X⁶ tiene la fórmula



25

en la que R⁴ y X⁶ son como se definen en el presente documento y X⁶ es en particular hidrógeno.

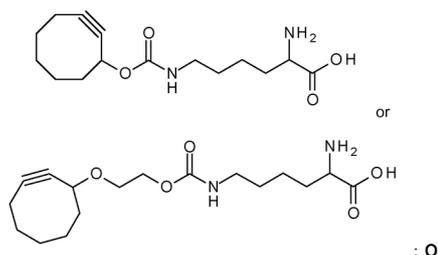
De acuerdo con una realización particular adicional, el primer aminoácido no natural es un compuesto de la fórmula:



30

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

35 De acuerdo con una realización particular adicional, el segundo aminoácido no natural es un compuesto de la fórmula:



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

5

Breve descripción de las figuras

La Figura 1 muestra: a) Estructuras de UAA; b) Gel de SDS-PAGE teñido con Coomassie de GFP^{TAG→UAA} en ausencia (-) o presencia (+) de UAA. La banda GFP (la flecha apunta a un marcador de peso molecular de 35 kDa) solo está presente cuando la mutación Y39TAG se suprime con éxito. Los isómeros TCO (TCO*, TCO[#]) muestran un mayor rendimiento de expresión que TCO; c) Los isómeros de TCO se mezclaron con clorhidrato de cisteamina y los espectros de ¹H RMN (solo se muestran las señales relevantes para observar la configuración *cis-trans*, el eje x es ppm) se midieron en múltiples puntos de tiempo y temperaturas (para más detalles, véase la **figura 8**). La gráfica muestran los datos de desplazamiento químico medidos a las 0 h, temperatura ambiente (rojo) y 24 h, 60 °C (azul). Los puntos de color negro indican las señales del doble enlace y los protones -CHO- de las formas *trans* de TCO* y TCO[#]. Las flechas de color negro indican las señales que pertenecen a los isómeros *cis* correspondientes que se forman en el tratamiento térmico con tiol. Mientras que los tres TCO se degradan con el tiempo en presencia de tioles, TCO* muestra la mayor estabilidad química (~80 % del isómero *trans* que queda después de 24 h); d) se hizo reaccionar GFP^{TAG→UAA} purificado (200 nM) con dos tetrazinas (15 μM, 20 min, 37 °C) y azida (45 μM, 10 h, 37 °C). Se muestra el resultado del gel de SDS-PAGE teñido con rayos UV así como teñido con Coomassie.

La Figura 2 muestra: a) Esquema del esquema de expresión y etiquetado empleado para el etiquetado de doble color del receptor de insulina que contiene una mutación ámbar (IR^{TAG}); b) Imágenes confocales de etiquetado de doble color de IR con diferentes combinaciones de UAA y tintes. Los paneles de la izquierda muestran una combinación de etiquetado SPIEDAC (TCO* + Me-Tet-Cy3 y SPAAC (BCN + azida-Cy5). Los paneles de la derecha muestran una combinación de SPIEDAC (TCO* + Me-Tet-Cy5) y seSPIEDAC entre SCO y H-Tet-Atto532; c) Etiquetado de doble color de partículas pseudovíricas (VLP) que muestra SPIEDAC entre TCO* y Me-Tet-Cy5 (parte superior) y seSPIEDAC entre SCO y H-Tet-Atto532 (parte inferior). Las barras de escala son de 20 μm.

La Figura 3 muestra imágenes SRM de IR y la partículas pseudovíricas (VLP) de influencia después del etiquetado SPIEDAC y seSPIEDAC, a) Imágenes de campo ancho (izquierda) y SRM (centro) de IR^{TAG} etiquetadas de acuerdo con la Figura 2a y la Figura 2b (Atto532 en color magenta, Cy5 en color cian). A la derecha, inserto del panel central y una gráfica de líneas (a través de la línea que se muestra en el panel central, que está resaltada por una flecha). El ancho de los picos marcados se da como anchura a media altura (FWHM); b) VLP etiquetadas análogas a (a) con campo amplio (izquierda), SRM (centro) y gráfico de líneas. Las imágenes de SRM se muestran a una resolución de 45 nm según lo determinado por la correlación del anillo de Fourier (FRC). Las barras de escala son de 1 μm.

La Figura 4 muestra compuestos que son útiles en el método de la presente invención (y se usaron en los ejemplos de trabajo): el Me-Tet-NH₂ no fluorescente que es útil para inactivar los grupos ciclooctenilo sin reaccionar; y los tintes fluorescentes azida-Cy3, H-Tet-Cy5, Me-Tet-Cy5, H-Tet-Atto532, Me-Tet-Cy3 y azida-Cy3.

La Figura 5 muestra la determinación de la cinética de reacción por FRET. El panel izquierdo muestra exploraciones de longitud de onda dependientes del tiempo de forma ilustrativa de GFP^{TAG→3(TCO*)} que reacciona con Me-Tet-Cy3 (fila superior) con H-Tet-Atto532 (fila inferior), respectivamente (color negro: primera exploración después de la adición de la sonda; a color gris: final de la reacción; norm. = normalizado a fluorescencia de GFP a "0 min"). El pico D y A se usaron para calcular FRET (FRET = A/(A+D)) y se representaron en el tiempo para todas las reacciones en el panel derecho (fila superior: Me-Tet-Cy3; fila inferior: H-Tet-Atto532). Para normalizar (norm. = normalizado) los datos de FRET, los experimentos se ajustaron por separado con el modelo monoexponencial descrito en el presente documento. Posteriormente, se calculó el primer punto de tiempo que alcanzó la amplitud máxima y se usó la relación de FRET correspondiente para normalizar los datos de FRET (para cada UAA y replicar de forma independiente). Por último, se promediaron los tres experimentos separados normalizados por UAA. Las barras de error indican la desviación estándar.

La Figura 6 muestra un ensayo FRET realizado durante 36 h para GFP^{TAG→3}. El panel izquierdo muestra las exploraciones de longitud de onda dependientes del tiempo de GFP^{TAG→3+Cy3} (norm. = normalizado a la fluorescencia de GFP en "0 min"). Los picos D y A se utilizaron para calcular FRET (FRET = A/(A+D)) y se

representaron en el tiempo en el panel derecho (norm. = normalizado a la amplitud máxima). La muestra no se agitó (pero ocasionalmente se mezcló en los puntos de tiempo indicados). Los valores de FRET fueron constantes en el tiempo.

5 La Figura 7 ilustra el análisis de GFP^{TAG→UAA+Cy3} etiquetado (con **1-4**) a través de FRET durante más de tres días. El panel izquierdo muestra las exploraciones de longitud de onda dependientes del tiempo a modo de ejemplo de GFP^{TAG→3+Cy3} (norm. = normalizado a la fluorescencia de GFP en "0 min"). Los picos D y A se usaron para calcular FRET ($FRET = A/(A+D)$) y se representaron en el tiempo para **1-4** en el panel derecho. Se ha de tener en cuenta que las muestras se mezclaron ocasionalmente. En particular, **1** y **3** dieron valores FRET consistentes a lo largo

10

La Figura 8 muestra a) espectros de ¹H RMN de los UAA **2** (TCO), **3** (TCO*), y **4** (TCO[#]) registrados en dPBS/dioxano-d8 (v/v 1:1) después de la incubación a temperatura ambiente (ta; líneas de color verde), 37 °C (líneas de color rojo), o 60 °C (líneas de color azul) durante 72 h. De acuerdo con estos datos, todos los compuestos permanecieron sin cambios en comparación con los espectros de ¹H RMN registrados directamente después de disolver los UAA en dPBS/dioxano-d8 (líneas de color morado). Los puntos de color negro indican las señales del doble enlace y los protones -CHO- de la forma *trans*; b) espectros de ¹H RMN de UAA **2** (TCO), **3** (TCO*), y **4** (TCO[#]) con dos equivalentes de cisteamina registrados en dPBS/dioxano-d8 (v/v 1:1) después de la incubación a temperatura ambiente (ta; líneas de color verde), 37 °C (líneas de color rojo), o 60 °C (líneas de color azul) durante 24 h. Los compuestos **2** (TCO) y **4** (TCO[#]) se convirtieron en más del 95 % en su forma *cis* después de 24 h a 60 °C en comparación con los espectros ¹H RMN registrados directamente después de la mezcla de los UAA con cisteamina (líneas superiores). Por el contrario, **3** (TCO*) permanecieron en más del 80 % en su forma *trans*. Los puntos de color negro indican las señales del doble enlace y los protones -CHO- de la forma *trans*. Las flechas de color negro indican las señales que pertenecen a los isómeros *cis* correspondientes que se forman en el tratamiento térmico con tiol; c) espectros de ¹H RMN de UAA **3** (TCO*) con dos equivalentes de cisteamina registrados en dPBS/dioxano-d8 (v/v 1:1) después de la incubación a 60 °C después de la mezcla (día 0; línea superior, panel izquierdo), un día (línea central, panel izquierdo), dos días (línea inferior, panel izquierdo), tres días (línea superior, panel derecho), siete días (línea central, panel derecho), o ten días (línea inferior, panel derecho). Los puntos de color negro indican las señales del doble enlace y los protones -CHO- de la forma *trans*. Las flechas de color negro indican las señales que pertenecen a los isómeros *cis* correspondientes que se forman en el tratamiento térmico con tiol.

15

20

25

30

La Figura 9 muestra el etiquetado SPAAC de BCN. canal GFP (izquierda) y canal Cy5 (derecha), a) etiquetado sin bloqueador de la endocitosis; b) etiquetado con bloqueador de la endocitosis. Barra de escala de 20 µm.

35

La Figura 10 muestra a) control de etiquetado de doble color de TCO* de IR sin (izquierda) y con inactivador (derecha); b) etiquetado de doble color de TCO* y BCN con el inactivador; c) control de etiquetado de doble color de TCO* para las VLP; d) etiquetado Me-Tet- Cy3/H-Tet-Atto532 de pEGFPN1_IR^{K676TAG}. Las células transfectadas se resaltan con guiones de color amarillo y las células no transfectadas (que no revelan ningún pegado o etiquetado inespecífico) con círculos de color rojo. Barra de escala de 20 µm.

40

La Figura 11 compara el espectro de ¹H RMN de 4-(6-metil-1,2,4,5-tetrazina-3-il)fenil)metanamina (línea inferior), el dienófilo indicado (línea central) y el producto de reacción registrado 10 minutos después de la mezcla de 4-(6-metil-1,2,4,5-tetrazina-3-il)fenil)metanamina con el dienófilo indicado (línea superior).

45

La Figura 12 muestra un gel de SDS-PAGE teñido con Coomassie de GFP^{TAG→UAA} purificado expresado en ausencia (control) o presencia de un UAA (**2**, **3**, **3a**, **3b** o **4**). Las bandas de sintetasa están ligeramente por encima del marcador de peso molecular de 55 kDa y las bandas de GFP están aproximadamente a la altura del marcador de peso molecular de 35 kDa (flecha). Las bandas de GFP solo están presentes cuando la mutación Y39TAG se suprime con éxito. El compuesto **3** (TCO*), **3a** (isómero axial de TCO*), **3b** (isómero ecuatorial de TCO*) y **4** (TCO[#]) muestran mayores rendimientos de expresión que el TCO (**4**).

50

La Figura 13 muestra: a) el aumento de la fluorescencia de Cy5 (aceptor) durante la reacción de etiquetado *in vitro* de GFP^{TAG→3} (A), GFP^{TAG→3a} (B) o GFP^{TAG→3b} (C) con H-Tet-Cy5; y b) ajustes lineales de las constantes de velocidad de reacción observadas k_{obs} . frente a la concentración de H-Tet-Cy5.

55

La Figura 14 muestra: a) los cambios de los perfiles de ¹H RMN del isómero axial **3a** y el isómero ecuatorial **3b** cuando se tratan con clorhidrato de cisteamina a 37 °C durante 0 h, 1 h, 3 h, 1 d, 2 d, 3 d, 4 d, 6 d, 13 d, 15 d, 21 d o 28 d; y b) la semivida de los compuestos **4** (TCO), **3a** (isómero axial de TCO*), **3b** (isómero ecuatorial de TCO*) y **1** (BCN).

60

La Figura 15 muestra imágenes confocales de células HEK293T que expresan una fusión recombinante de GFP del receptor de insulina (IR) y (C-terminal), en las que el compuesto **3**, **3a** o **3b** se ha incorporado en el sitio codificado por ámbar (es decir, expresando GFP-IR^{TAG→3}, GFP-IR^{TAG→3a} o GFP-IR^{TAG→3b}) que se marcaron con H-Tet-Cy5 o Me-Tet-Cy5, respectivamente. La fluorescencia de GFP (canal GFP) indica la expresión exitosa de la proteína GFP-IR^{TAG→UAA} recombinante. La fluorescencia de Cy5 (canal Cy5) indica un etiquetado exitoso de la

65

proteína GFP-IR^{TAG->UAA} recombinante. Barra de escala es de 20 μ m.

Descripción detallada de la invención

5 La invención se basa, entre otros, en el hallazgo de que las tetrazinas sustituidas con alquilo reaccionan preferiblemente con dienófilos en tensión que comprenden un grupo *trans*-ciclooctenilo en comparación con dienófilos en tensión que comprenden un grupo ciclooctinilo. Por lo tanto, en una situación competitiva en la que están presentes grupos *trans*-ciclooctenilo y ciclooctinilo, las tetrazinas sustituidas con alquilo reaccionarán preferiblemente con los grupos *trans*-ciclooctenilo. En este contexto, el término "reacciona preferiblemente" se refiere a una relación de constantes de velocidad k , cuya relación es de 100 o superior, 200 o superior, 500 o superior, o 1000 o superior, si la constante de velocidad k_1 para la reacción de las tetrazinas sustituidas con alquilo con los dienófilos en tensión que comprenden un grupo *trans*-ciclooctenilo se compara con la constante de velocidad k_2 para la reacción de las tetrazinas sustituidas con alquilo con los dienófilos en tensión que comprenden un grupo ciclooctinilo (es decir, la proporción es k_1/k_2). Por consiguiente, la reacción de la tetrazina sustituida con alquilo con el dienófilo en tensión que comprende un grupo *trans*-ciclooctenilo puede desarrollarse hasta un punto en el que sustancialmente todos los grupos *trans*-ciclooctenilo han reaccionado con la tetrazina sustituida con alquilo, mientras que sustancialmente ningún grupo ciclooctinilo ha reaccionado con la tetrazina sustituida con alquilo.

20 De acuerdo con ciertas realizaciones de la invención, las tetrazinas sustituidas con alquilo reaccionan específicamente con los dienófilos en tensión que comprenden un grupo *trans*-ciclooctenilo. Por lo tanto, en una situación competitiva en la que están presentes grupos *trans*-ciclooctenilo y ciclooctinilo, las tetrazinas sustituidas con alquilo reaccionarán preferiblemente con los grupos *trans*-ciclooctenilo. En este contexto, el término "reacciona específicamente" se refiere a una relación de constantes de velocidad k , cuya relación es de 2000 o superior, 5000 o superior, o 10000 o superior, si la constante de velocidad k_1 para la reacción de las tetrazinas sustituidas con alquilo con los dienófilos en tensión que comprenden un grupo *trans*-ciclooctenilo se compara con la constante de velocidad k_2 para la reacción de las tetrazinas sustituidas con alquilo con los dienófilos en tensión que comprenden un grupo ciclooctinilo (es decir, la proporción es k_1/k_2). Por consiguiente, la reacción de las tetrazinas sustituidas con alquilo con el dienófilo en tensión que comprende un grupo *trans*-ciclooctenilo puede desarrollarse hasta un punto en el que todos los grupos *trans*-ciclooctenilo han reaccionado con la tetrazina sustituida con alquilo, mientras que sustancialmente ningún grupo ciclooctinilo ha reaccionado con la tetrazina sustituida con alquilo.

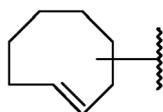
Las reacciones de la invención pueden realizarse *in vitro* o *in vivo*, dependiendo de las condiciones de reacción apropiadas. Debido a que solo el primer y segundo grupos dienófilos (por ejemplo, los aminoácidos no naturales que comprenden dichos grupos presentes en los polipéptidos diana) participan en la reacción de ligadura, los métodos de la invención se pueden usar de manera fiable para producir poblaciones homogéneas de conjugados bien definidos (por ejemplo, conjugados polipéptido-marcador diana que comprenden estequiometrías definidas y sitios de ligadura definidos) con alta eficacia y especificidad. Debido a que cualquiera de diversos primeros y segundos agentes modificadores reactivos (por ejemplo, aminoácidos no naturales) que comprenden el primer y segundo grupos dienófilos pueden incorporarse en diversas moléculas (por ejemplo, polipéptidos diana), la producción de conjugados (por ejemplo, conjugados de polipéptido diana) no se limita a moléculas específicas (por ejemplo, polipéptidos). Además, las tecnologías existentes permiten de manera beneficiosa la incorporación de aminoácidos no naturales en cualquier posición de aminoácidos en un polipéptido. Por lo tanto, la ubicación del primer y segundo aminoácidos no naturales químicamente reactivos en los polipéptidos diana, puede elegirse opcionalmente basándose, por ejemplo, en si la ubicación en esa ubicación cambiará, por ejemplo, las conformaciones, las actividades biológicas, las actividades farmacológicas, las estabildades, las biodisponibilidades, u otras propiedades, del polipéptido diana, o de los conjugados polipéptido-marcador diana resultantes.

La expresión "aminoácido no natural" se refiere a un aminoácido que no es uno de los 20 aminoácidos canónicos o selenocisteína o pirrolisina. El término también se refiere a análogos de aminoácidos, por ejemplo, en los que el grupo α -amino se reemplaza por un grupo hidroxilo; o en los que la función del ácido carboxílico forma un éster. Actualmente, cuando los aminoácidos no naturales de la invención o sales de los mismos, en los que X^6 es distinto de hidrógeno, se usan para la preparación de polipéptidos en un sistema de traducción, se cree que X^6 se elimina *in situ*, por ejemplo, enzimáticamente dentro del sistema de traducción elegido, antes de incorporarse en el polipéptido. Por consiguiente, X^6 se elige convenientemente para que sea compatible con la capacidad de un sistema de traducción para convertir aminoácidos no naturales de la invención o sales de los mismos, en una forma que sea reconocida y procesada por la aminoacil ARNt sintetasa.

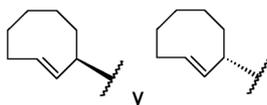
Los compuestos o sales de la invención poseen centro de asimetría y pueden existir en disposiciones espaciales diferentes o como diferentes tautómeros. Para la preparación de polipéptidos con grupos *trans*-ciclooctenilo o ciclooctinilo, se pueden usar mezclas enantioméricas, en particular racematos, 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.

Más específicamente, el grupo *trans*-ciclooctenilo de la fórmula:

65

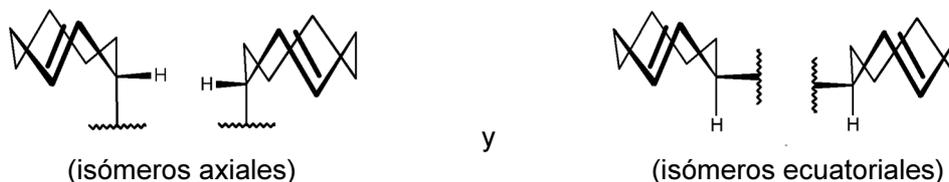


5 pretende incluir dos formas isoméricas que difieren entre sí por la configuración absoluta en el átomo de carbono donde el anillo *trans*-cicloocteno está unido al resto de la molécula. Por consiguiente, se pueden distinguir los dos enantiómeros que tienen una configuración S o R, por ejemplo, los enantiómeros de las fórmulas



10 Debido al atropisomerismo, los grupos *trans*-ciclooctenilo existen en cuatro formas estereoisoméricas diferentes. Con respecto al átomo de carbono donde se une el anillo *trans*-cicloocteno, se diferencian los isómeros axiales y los isómeros ecuatoriales (es decir, el resto de la molécula se une en posición axial o ecuatorial con respecto al anillo de *trans*-cicloocteno). Para cada una de dichas formas axiales y para cada una de dichas formas isoméricas ecuatoriales hay dos enantiómeros. Por consiguiente, los estereoisómeros del grupo *trans*-ciclooctenilo pueden representarse por las fórmulas

15



20 Salvo que se indique de otro modo, el término "isómero axial" se refiere a una mezcla de ambos enantiómeros (en particular un racemato) de la forma isomérica axial, y el término "isómero ecuatorial" se refiere a una mezcla de ambos enantiómeros (en particular un racemato) de la forma isomérica ecuatorial.

20

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.

25

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

30

Alquilo es un alquilo de cadena lineal o ramificada 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-dimetilbutilo, 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.

35

Alquenilo es un radical hidrocarburo individualmente insaturado que tiene 2, 3, 4, 5 o 6 átomos de carbono, por ejemplo, vinilo, alilo (2-propen-1-ilo), 1-propen-1-ilo, 2-propen-2-ilo, y metalilo (2-metilprop-2-en-1-ilo). Alquenilo C_3-C_5 es, en particular, alilo, 1-metilprop-2-en-1-ilo, 2-buten-1-ilo, 3-buten-1-ilo, metalilo, 2-penten-1-ilo, 3-penten-1-ilo, 4-penten-1-ilo, 1-metilbut-2-en-1-ilo, 2-etilprop-2-en-1-ilo, y 2-hexen-1-ilo.

40

Alquilenilo es un alquilenilo de cadena lineal o ramificada que tiene de 1 a 6, en particular, de 1 a 4 átomos de carbono. Los ejemplos incluyen metileno, etileno, 1,2-etileno, 1,3-propileno, isopropileno, 1,4-butileno, y 1,5-pentileno.

Alcoxi es un radical de la fórmula $R-O-$, en la que R es un grupo alquilo de cadena lineal o ramificada 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

Alquenoxi es un radical de la fórmula $R-O-$, en la que R es un grupo alquenilo de cadena lineal o ramificada 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.

50

Alcanoiloxi es un radical de la fórmula $R-(CO)-O-$, en la que R es un grupo alquilo de cadena lineal o ramificada 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.

Alquilaminocarboniloxi es un radical de la fórmula $R-NH-(CO)-O-$, en la que R es un grupo alquilo de cadena lineal o ramificada 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 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.

5 Alcanoilsulfanilo es un radical de la fórmula R-(CO)-S-, en la que R es un grupo alquilo de cadena lineal o ramificada 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.

Alquilamino es un radical de la fórmula R-NH- en la que R es un radical alquilo que tiene de 1 a 6, en particular, de 1 a 4 átomos de carbono como se define en el presente documento. Los ejemplos incluyen metilamino, etilamino, n-propilamino, iso-propilamino, n-butilamino, 2-butilamino, iso-butilamino, y *tert*-butilamino.

15 Dialquilamino es un radical de la fórmula RR'N- en la que R y R' son independientemente entre sí un radical alquilo que tiene de 1 a 6, en particular, de 1 a 4 átomos de carbono como se define en el presente documento. Los ejemplos incluyen dimetilamino, dietilamino, y N-metil-N-etilamino.

Alquenilamino es un radical de la fórmula R-NH- en la que R es un radical alquenilo que tiene de 2 a 6, en particular, de 2 a 4 átomos de carbono como se define en el presente documento. Los ejemplos incluyen vinilamino, alilamino (2-propen-1-il-amino), 1-propen-1-il-amino, 2-propen-2-il-amino, y metalilamino (2-metilprop-2-en-1-il-amino).

20 N-Alquil-N-alquenilamino es un radical de la fórmula RR'N- en la que R es un radical alquilo que tiene de 1 a 6, en particular, de 1 a 4 átomos de carbono como se define en el presente documento y R' un radical alquenilo que tiene de 2 a 6, en particular, de 2 a 4 átomos de carbono como se define en el presente documento. Los ejemplos incluyen N-metil-N-vinilamino, N-metil-N-alilamino (N-metil-N-2-propen-1-il-amino), N-metil-N-1-propen-1-il-amino, N-metil-N-2-propen-2-il-amino, y N-metil-N-metalilamino (N-metil-N-2-metilprop-2-en-1-il-amino).

25 Dialquenilamino es un radical de la fórmula RR'N- en la que R y R' son independientemente entre sí un radical alquilo que tiene de 2 a 6, en particular, de 2 a 4 átomos de carbono como se define en el presente documento. Los ejemplos incluyen divinilamino, dialilamino (di-(2-propen-1-il)-amino), y N-vinil-N-alil-amino.

30 Salvo que se indique de otro modo, el término "sustituido" significa que un radical está sustituido con 1, 2 o 3, especialmente 1 o 2, sustituyentes que se seleccionan en particular del grupo que consiste en halógeno, alquilo C₁-C₄, CN, CF₃, hidroxilo, -O-CF₃, alcoxi C₁-C₄, alcanoiloxi C₂-C₄, alquilaminocarboniloxi C₁-C₄ y alquiltio C₁-C₄.

35 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, y alcoholatos.

45 Los ácidos o bases fisiológicamente tolerados son, en particular, aquellos que son tolerados por el sistema usado para la incorporación del primer y segundo dienófilos (por ejemplo, un sistema biológico tal como un sistema de traducción usado para la preparación de polipéptidos con grupos *trans*-ciclooctenilo o ciclooctinilo), por ejemplo, que son sustancialmente no tóxicos para las células vivas.

50 Los compuestos y sales de la invención (por ejemplo, los dienófilos y las tetrazinas de la invención) pueden prepararse por analogía con los métodos que se conocen bien en la técnica. Los métodos adecuados para la preparación de compuestos de fórmula (I) se encuentran en las diversas publicaciones citadas en el presente documento. Algunos métodos se indican en el presente documento.

55 Los compuestos y sales de la invención se pueden usar para la preparación de polipéptidos que comprenden uno o más de un grupo análogo 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. Un polipéptido es cualquier oligómero de residuos de aminoácidos (naturales o no naturales, o una combinación de los mismos), de cualquier longitud, normalmente pero no exclusivamente, unidos por enlaces peptídicos covalentes. Un polipéptido puede ser de cualquier fuente, por ejemplo, un polipéptido de origen natural, un polipéptido producido por técnicas de genética molecular recombinante, un polipéptido de una célula o sistema de traducción, o un polipéptido producido por medios sintéticos libres de células. Un polipéptido se caracteriza por su secuencia de aminoácidos, por ejemplo, la estructura primaria de sus residuos de aminoácidos componentes. Como se usa en el presente documento, la secuencia de aminoácidos de un polipéptido no se limita a secuencias de longitud completa, sino que pueden ser secuencias parciales o completas. Además, no se pretende que un polipéptido esté limitado por poseer o no poseer ninguna actividad biológica particular. Como se usa en el presente documento, el término "proteína" es sinónimo de polipéptido. El término "péptido" se refiere a un pequeño polipéptido, por ejemplo, de 2-25 aminoácidos de longitud. Como se usa en el presente documento, "para incorporar un aminoácido no natural", por

ejemplo, en un polipéptido diana, se refiere a la adición directa de un aminoácido no natural a una cadena polipeptídica en crecimiento durante la construcción primaria del polipéptido diana, por ejemplo, a través de traducción o síntesis química.

- 5 El primer y segundo aminoácidos no naturales pueden incorporarse directamente en polipéptidos diana usando cualquiera de una serie de métodos conocidos en la técnica. Si bien muchas realizaciones utilizan sistemas de traducción ortogonales como la ruta de incorporación directa de los aminoácidos no naturales, se pueden usar, como alternativa, otros métodos de incorporación directa (por ejemplo, sistemas de traducción *in vitro*, síntesis en fase sólida, etc.). Se apreciará que en las realizaciones típicas en el presente documento, un aminoácido no natural se incorpora preferiblemente en el polipéptido diana, es decir, durante la construcción del polipéptido, y no se añade a través de la derivación química postraduccional.

15 En ciertas realizaciones descritas en el presente documento, los aminoácidos no naturales pueden incorporarse específicamente al sitio en un polipéptido diana con alta eficiencia y alta fidelidad utilizando pares ortogonales de ARNt/aminoacil-ARNt sintetasa. Las levaduras metilotróficas son candidatos atractivos para su uso como sistemas de expresión recombinante para proteínas heterólogas y terapéuticamente útiles. La organización subcelular eucariota de la levadura metilotrófica les permite realizar muchos de los eventos de plegamiento postraduccional, procesamiento y modificación necesarios para sintetizar polipéptidos portadores biológicamente activos y/o polipéptidos diana derivados de mamíferos. A diferencia de las proteínas expresadas en *S. cerevisiae*, las proteínas producidas por levaduras metilotróficas tales como *P. pastoris*, *P. methanolica*, *P. angusta* (también conocida como *Hansenula polymorpha*), *Candida boidinii*, y *Torulopsis spp.*, tienen menos probabilidades de contener estructuras de glicano con alto contenido de manosa que puedan obstaculizar el procesamiento aguas abajo de glucoproteínas expresadas de forma heteróloga. Además, los polipéptidos diana sintetizados en levadura metilotrófica están ventajosamente libres de compuestos pirogénicos y antigénicos característicos a menudo de las proteínas expresadas en *E. coli*. De manera más significativa, los sistemas de expresión de levaduras metilotróficas son particularmente útiles para la síntesis a gran escala. Por ejemplo, los sistemas de traducción ortogonales en levaduras metilotróficas pueden permitir la expresión de polipéptidos diana que comprenden aminoácidos no naturales en niveles de 10 a 100 veces más altos que en sistemas de *S. cerevisiae*, bacterianos, de insectos o mamíferos. Además, la levadura metilotrófica se puede cultivar fácilmente en un medio salino sencillo y definido, eliminando la necesidad de los costosos complementos de medio y equipos que se requieren para los sistemas de expresión de baculovirus.

25 El término "sistema de traducción" se refiere a los componentes necesarios para incorporar un aminoácido en una cadena polipeptídica en crecimiento (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasa, y ARNm.

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

40 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).

45 Preferiblemente, 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* o una levadura metilotrófica; una célula vegetal o una célula animal tal como una célula de insecto o una célula de mamífero, por ejemplo, una célula HEK o 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*.

55 De acuerdo con una realización particular adicional, el sistema de traducción es una célula de mamífero, por ejemplo, una célula HEK o HeLa.

60 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 ARNt que tiene un anticodón para un codón selector o un polinucleótido que codifica dicho ARNt; y un polinucleótido que codifica el polipéptido diana y que comprende uno o más de un codón selector.

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

65 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 los métodos de la invención es capaz de acilar un ARNt con un

aminoácido no natural de la invención.

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

5 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.

15 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.

20 Los pares de ARNt/RS ortogonales usados en procesos de la invención tienen preferiblemente las siguientes propiedades: el ARNt-O es preferiblemente 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 codón selector, por ejemplo, un codón de terminación ámbar, en el ARNm que codifica el polipéptido.

25 En algunos aspectos, el sistema de traducción comprende un segundo par ortogonal, por ejemplo, una segunda RS-O y un segundo ARNt-O que utilizan el segundo aminoácido no natural, de manera que el sistema ahora puede incorporar al menos dos aminoácidos no naturales en diferentes sitios seleccionados en un polipéptido. En esta realización, la segunda RS-O aminoacila preferiblemente el segundo ARNt-O con el segundo aminoácido no natural que es diferente del primer aminoácido no natural, y el segundo ARNt-O reconoce un codón selector que es diferente del codón selector reconocido por el primer ARNt-O. Los sistemas de traducción adecuados que comprenden dos pares de ARNt/RS ortogonales son conocidos en la técnica. Véase, por ejemplo, Han Xiao, et al., Angew Chem Int Ed Engl 2013, 52, 14080-14083.

35 En algunas realizaciones, el sistema de traducción comprende una célula por ejemplo, una célula de mamífero, insecto, una levadura, de bacteria o de E. coli. El tipo de célula utilizada no está particularmente limitado, siempre y cuando RS-O y ARNt-O conserven su ortogonalidad en el entorno de la célula.

40 El término "acila preferiblemente" se refiere a una eficiencia de, por ejemplo, aproximadamente el 50 % de eficacia, aproximadamente el 70 % de eficacia, aproximadamente el 75 % de eficacia, aproximadamente el 85 % de eficacia, aproximadamente el 90 % de eficacia, aproximadamente el 95 % de eficacia, o aproximadamente el 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 el 75 % de eficacia para un codón selector dado, a más de aproximadamente el 80 % de eficacia para un codón selector dado, a más de aproximadamente el 90 % de eficacia para un codón selector dado, a más de aproximadamente el 95 % de eficacia para un codón selector dado, o a más de aproximadamente el 99 % o más de eficacia para un codón selector dado.

50 El término "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 codón poco habitual.

Un anticodón tiene la secuencia complementaria inversa del codón correspondiente.

60 Un par ARNt-O/RS-O está compuesto por un ARNt-O, por ejemplo, un ARNt supresor, 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 preferiblemente 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. Preferiblemente, la célula usada como sistema de traducción es diferente de dichas especies fuente.
- 15 Por ejemplo, un ARNt-O ortogonal adecuado se puede derivar de una archaeobacterium, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-I, *Archaeoglobus fulgidus*, *Pirococcus furiosus*, *Pirococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pirobaculum aerophilum*, *Pirococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o una eubacterium, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, o *Bacillus stearothermophilus*, mientras que la RS-O ortogonal puede derivarse de un organismo o una combinación de organismos, por ejemplo, una archaeobacterium, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-J, *Archaeoglobus fulgidus*, *Pirococcus furiosus*, *Pirococcus horikoshii*, *Aeuropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Methanosarcina bakeri*; *Methanosarcina hafniense*; *Pirobaculum aerophilum*, *Pirococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o una eubacterium, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, o *Bacillus stearothermophilus*. En una realización, también pueden usarse fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales, por ejemplo, mamíferos, insectos, o artrópodos, como fuentes de ARNt-O y RS-O. Se describen métodos para desarrollar pares de ARNt/RS, por ejemplo, en los documentos WO 02/085923 y WO 02/06075.

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

- 35 La pirrolisil ARNt sintetasa usada en métodos de la invención puede ser una pyIRS de tipo silvestre o una modificada por ingeniería genética. Los ejemplos de pyIRS de tipo silvestre incluyen pyIRS de arqueobacterias y eubacterias, tales como *Methanosarcina mazei*, *Methanosarcina barkeri*, *Methanococcoides burtonii*, *Methanosarcina acetivorans*, *Methanosarcina thermophila*, y *Desulfotobacterium hafniense*.

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

- 45 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. mazei*.

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.

- 50 SEQ ID NO:1:

```
MDKKPLNLTLSATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL 60
RHHKYRKTCKRCRVSDLEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLE 120
NTEAAQAQPSGSKFSPAIPVSTQESVSVSPASVSTSISSISTGATASALVKGNTNPITSMS 180
APVQASAPALTKSQTDRLVLLNPKDEISLNSGKPFRELESELLSRKKDLQOIYAEERE 240
NYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM 300
LAPNLNYLRKLDRALPDPIKIFEIGPCYRKESDGEHLEEFMLNFCQMGSGCTRENLE 360
SIITDFLNHLGIDFKIVGDSVMVYGDTLDMHGDLELSSAVVGPIPLDREWGIDKPWIGA 420
GFGLERLLKVKHDFKNIKRAARSESYNGISTNL 454
```

- 55 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, preferiblemente seleccionadas de las sustituciones de aminoácidos Y306A e Y384F.

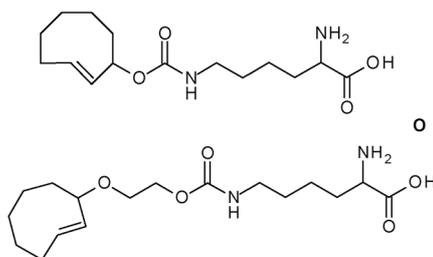
De acuerdo con una realización particular, la pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos de pirrolisil ARNt sintetasa de *M. maize* mutante expuesta en SEQ ID NO: 2, o un fragmento funcional de la misma.

SEQ ID NO:2:

5
 MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARAL 60
 RHHKYRKTCKRCRVSEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSVARAPKPLE 120
 NTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPITSMS 180
 APVQASAPALTKSQTDRLLEVLLNPKDEISLNSGKPFRELESELLSRKKDLQQIYAEERE 240
 NYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPM 300
 LAPNLANYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLE 360
 SIIITDFLNHLGIDFKIVGDSMVFGDTLDVMHGDLELSSAVVGP IPLDREWGIDKPWIGA 420
 GFGLERLLKVKHDFKNIKRAARSESYYNGISTNL 454

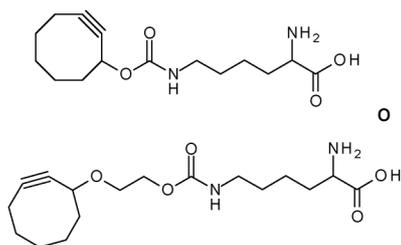
Cualquier aminoacil ARNt sintetasa descrita en el presente documento puede usarse para acilación de un ARNt con los aminoácidos no naturales de la invención.

10 De acuerdo con un aspecto de la invención, se usa pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre para acilación de un ARNt con un compuesto de fórmula



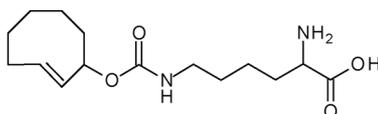
15 o una sal del mismo.

20 De acuerdo con un aspecto adicional de la invención, se usa pirrolisil ARNt sintetasa de *M. maize* de tipo silvestre para acilación de un ARNt con un compuesto de fórmula

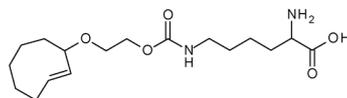


25 o una sal del mismo.

De acuerdo con otro aspecto de la invención, 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

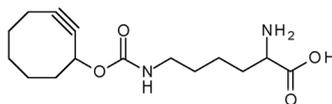


30 o

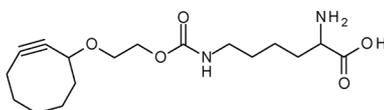


o una sal del mismo.

- 5 De acuerdo con otro aspecto de la invención, 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



o una sal del mismo.

- 15 El ARNt que se usa en combinación con la *pylRS* (ARNt^{pyl}) puede ser un ARNt de tipo silvestre o uno modificado por ingeniería genética. Los ejemplos de ARNt^{pyl} de tipo silvestre incluyen ARNt de arqueobacterias y eubacterias, tal como se ha mencionado anteriormente, que facilitan la incorporación en la traducción de restos de pirrolisilo.

- 20 De manera similar, pueden proporcionarse y usarse pares adecuados de ARNt^{Tyr}/TyrRS y ARNt^{Leu}/leucil-ARNt sintetasa.

- 25 Codones selectores utilizados en métodos 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, o al menos un codón de cuatro bases.

- 30 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.

- 35 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 ovocitos 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. En células de mamífero (células HEK), el codón ocre (TAA) y el codón ámbar (TAG) se han utilizado para incorporar dos aminoácidos no naturales diferentes en sitios distintos del mismo polipéptido. Véase, por ejemplo, Han Xiao, et al., *Angew Chem Int Ed Engl* 2013, 52, 14080- 14083.

- 40 En una realización, los métodos de la invención 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, preferiblemente el codón de terminación ámbar, y es acilado por una RS-O con un compuesto de la 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.

- 45 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 HEK 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^{Pyl} usado en procesos de la invención comprende el anticodón CUA para el codón de terminación ámbar.

5 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 compite con el ARNt^{Arg} 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*. Por
10 consiguiente, cualquier codón triplete no usado por el sistema de traducción aplicado en los procesos de la invención puede actuar como codón selector.

Una alternativa adicional para incorporar los UAA de acuerdo con la presente invención en polipéptidos es usar un codón cuadruplete (de cuatro bases) como un codón selector y un par ARNt-O/RS correspondiente.

15 El sistema de traducción se mantiene durante un tiempo adecuado en condiciones que permiten la formación del polipéptido de la invención por un ribosoma. El ribosoma se une a ARNm que codifica el polipéptido diana y comprende uno o varios codones selectores. Después, 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
25 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 de la invención puede recuperarse del sistema de traducción. Para este fin, los polipéptidos de la invención 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
35 hidrófoba, cromatografía de hidroxilapatita, cromatografía de lectina, y electroforesis en gel. Pueden usarse etapas de plegamiento de proteínas, según se desee, para preparar proteínas maduras correctamente plegadas. El análisis por cromatografía líquida de alto rendimiento (HPLC), cromatografía de afinidad u otros métodos adecuados pueden emplearse en etapas de purificación final cuando se desee alta pureza. Pueden usarse anticuerpos preparados contra el aminoácido no natural o los polipéptidos de la invención como reactivos de purificación, es decir, para la 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 los mismos.

45 Como se ha indicado, los expertos en la técnica 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 procariotas 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, en
50 ocasiones es deseable desnaturalizar y reducir los polipéptidos expresados y después hacer que los polipéptidos se replieguen en su configuración preferida. Por ejemplo, guanidina, urea, DTT, DTE y/o una chaperonina pueden añadirse a un producto de traducción de interés. Los métodos de reducción, desnaturalización y renaturalización de proteínas son bien conocidos por los expertos en la materia. Los polipéptidos pueden replegarse en un tampón redox que contiene, por ejemplo, glutatión oxidado y L-arginina.

Se apreciará que, si bien se detallan en el presente documento métodos particulares de construcción de polipéptidos diana de la invención que comprenden aminoácidos no naturales químicamente reactivos, por ejemplo, usando sistemas de traducción ortogonales, no deben interpretarse necesariamente como limitantes. Además, otros métodos, por ejemplo, no ortogonales, para construir polipéptidos diana que tienen aminoácidos no naturales también se incluyen en el presente documento en las diversas realizaciones. Dichos métodos se describen con más detalle en el presente documento.

En diferentes realizaciones de la invención, los polipéptidos diana de la invención pueden construirse a través de métodos de incorporación directa tales como un sistema de traducción ortogonal. Esto representa una realización preferida, debido a la capacidad de los sistemas ortogonales para producir altos rendimientos de polipéptidos

correctamente plegados y postraduccionalmente modificados con aminoácidos no naturales incorporados específicamente en el sitio. Como alternativa o adicionalmente, sin embargo, se pueden emplear otras estrategias para la incorporación directa de aminoácidos no naturales en una cadena polipeptídica para introducir el primer y segundo aminoácidos no naturales en los polipéptidos diana. Por ejemplo, un método biosintético *in vitro* general para incorporar aminoácidos no naturales en, por ejemplo, polipéptidos diana, durante la construcción primaria utiliza ARNt supresores sin sentido o de cambio de marco que se han acilado químicamente con el aminoácido no natural deseado y luego se añade a un extracto capaz de soportar la biosíntesis de proteínas y que incluye un gen que contiene una mutación sin sentido ámbar deseada. Esta estrategia se ha utilizado para incorporar específicamente en el sitio más de 100 aminoácidos no naturales en diversas proteínas de prácticamente cualquier tamaño, y se puede usar en el presente documento para crear polipéptidos diana que comprenden aminoácidos no naturales. En otras realizaciones, los aminoácidos no naturales pueden incorporarse directamente en polipéptidos diana más pequeños (que varían de 60 a 100 aminoácidos) a través de síntesis química. La síntesis de péptidos en fase sólida es un método que se usa ampliamente para sintetizar químicamente péptidos y proteínas pequeñas que comprenden aminoácidos no naturales (véase, por ejemplo, Merrifield (1963) "Solid Phase Peptide synthesis. I. The synthesis of a tetrapeptide". JACS 85: 2149-2154), y puede adaptarse para producir polipéptidos diana de la invención. Esta técnica comprende normalmente dos fases: La síntesis de péptidos en fase sólida (SPPS) de la primera etapa incluye el ensamblaje de una cadena peptídica utilizando derivados de aminoácidos protegidos en un soporte polimérico a través de ciclos repetidos de acoplamiento-desprotección. La amina N-terminal libre de un péptido unido en fase sólida se puede acoplar luego a una única unidad de aminoácido protegida en N, por ejemplo, un aminoácido no natural. Después, esta unidad se desprotege, revelando una nueva amina N-terminal a la que se puede unir un aminoácido adicional. Mientras que el péptido se sintetiza generalmente por métodos por etapas, todos los reactivos solubles pueden eliminarse de la matriz de soporte péptido-sólido por filtración y lavarse al final de cada etapa de acoplamiento. En la segunda fase de SPPS, el péptido se escinde del soporte y los grupos protectores de cadena lateral se eliminan para producir el péptido, por ejemplo, un polipéptido diana que comprende uno o más aminoácidos no naturales. Hay dos formas principales de síntesis de péptidos en fase sólida: Fmoc (Carpino et al. (1972) "9-Fluorenylmethoxycarbonyl amino-protecting group". J Org Chem 37: 3404-3409), en la que se usa un grupo protector alfa-amino lábil de base, y t-Boc, en la que se usa un grupo protector de ácido lábil. Cada método implica diferentes resinas y la protección de la cadena lateral de los aminoácidos y las consecuentes etapas de escisión/desprotección.

Por medio de los métodos de la invención, además, es posible introducir múltiples marcadores en oligonucleótidos obtenidos por síntesis. Las amiditas que componen los dienófilos de la invención requeridas para esto pueden prepararse fácilmente. El oligonucleótido puede tener cualquier longitud entre 3 y 10000 nucleótidos, preferiblemente entre 4 y 5000 nucleótidos, más preferiblemente entre 5 y 1000 nucleótidos o entre 10 y 500 nucleótidos, mucho más preferiblemente entre 10 y 200 nucleótidos. En una realización particular de la invención, el oligonucleótido a modificar por el método de la presente invención puede tener más de 50, preferiblemente más de 100 nucleótidos. El oligonucleótido de acuerdo con la invención puede ser de ADN o ARN monocatenario o bicatenario, así como análogos de ácidos nucleicos (por ejemplo, PNA, LNA) o quimera de estos con ADN y/o ARN.

Los azúcares no naturales que comprenden los dienófilos de la invención pueden incorporarse metabólicamente en glicanos. Por ejemplo, el ácido neuramínico o la N-acetilmanosaamina pueden modificarse para comprender un dienófilo de la invención y, por lo tanto, los métodos de la invención permiten el etiquetado múltiple de glicanos en los que se han incorporado los azúcares no naturales que comprenden los dienófilos de la invención.

En una aplicación adicional, los dienófilos de la invención pueden funcionalizarse para transportar grupos (por ejemplo, grupos trimetoxisililo) que pueden usarse para la síntesis de fases sólidas reactivas. Como resultado, las fases sólidas se vuelven accesibles y llevan tanto el primer como el segundo dienófilo.

Las aplicaciones siguientes van desde la tecnología de chip para oligonucleótidos, polipéptidos o glicanos hasta superficies catalíticas y reactivos de fase sólida.

Se entiende que los puntos cuánticos significan nanopartículas que están compuestas por compuestos tales como CdS o CdSe y tienen propiedades ópticas especiales. Excitados por los láseres, tienen una fluorescencia muy fuerte en función de su tamaño y, por lo tanto, se utilizan cada vez más en el campo del diagnóstico, especialmente porque permiten la detección de moléculas individuales. Sin embargo, una condición previa para esto es su dopaje con grupos funcionales, que se desarrolla a través de grupos SH y permite la detección de una interacción posterior con las moléculas.

Además, las nanopartículas de oro se consideran para investigaciones de biomoléculas de microscopía electrónica debido a sus propiedades especiales. El anclaje de las moléculas en la superficie también se logra a través de grupos SH.

Los métodos de la invención pueden usarse para unir moléculas a su superficie. Con este fin, se puede producir un dienófilo que contiene el grupo SH. Normalmente, se producen los disulfuros correspondientes y luego se puede preparar el compuesto mercapto a partir de los mismos por reducción con ditiotreitól. Los disulfuros como tales pueden anclarse a las superficies de oro. Por lo tanto, los dienófilos se pueden unir a la superficie de puntos cuánticos u otros metales. Por ejemplo, los anticuerpos, sacáridos o agentes terapéuticos pueden anclarse en la superficie de los puntos

cuánticos para fines diagnósticos o terapéuticos.

Los kits de la invención se pueden usar en particular para preparar polipéptidos de la invención. Con este fin, los kits pueden comprender uno o más medios para preparar un polipéptido. Dichos medios incluyen

- 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.

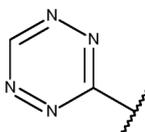
Tanto la aminoacil ARNt sintetasa como el ARNt pueden, por ejemplo, proporcionarse en forma de uno o varios vectores de expresión para dicha aminoacil ARNt sintetasa y ARNt correspondiente.

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.

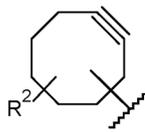
Además, dicho kit puede comprender medios adicionales para traducción de un polinucleótido que codifica dicho polipéptido, por ejemplo un sistema de traducción, tales como células de *E. coli*, células HeLa, extracto de *E. coli*, extracto de germen de trigo o lisado de reticulocitos de conejo, e instrucciones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

Los expertos en la técnica apreciarán que las reacciones de la invención deben realizarse en orden inverso si la reacción de la primera tetrazina con el primer dienófilo tiene lugar en presencia de la segunda tetrazina. Por lo tanto, la presente invención se refiere además a métodos para formar uniones mediante reacciones de cicloadición, en donde el método comprende hacer reaccionar una primera tetrazina con un primer dienófilo seguido de hacer reaccionar una segunda tetrazina con un segundo dienófilo, en donde la reacción de la primera tetrazina con el primer dienófilo se desarrolla en presencia de la segunda tetrazina, en donde

- (i) la primera tetrazina comprende un grupo de la fórmula:



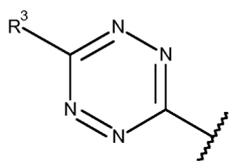
- (ii) el primer dienófilo comprende un grupo ciclooctinilo de la fórmula:



en la que

- R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquencilamino C₂-C₅, alquencil C₂-C₅-alquil C₁-C₄-amino o di-(alquencil C₂-C₅)amino; y
- R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅

- (iii) la segunda tetrazina comprende un grupo de la fórmula:

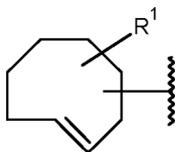


en la que

- R³ es alquilo C₁-C₃;

y

(iv) el segundo dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:



5 en la que

10 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 15 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

20 Los aspectos y realizaciones específicos para este método pueden derivarse directamente de la descripción en el presente documento que puede aplicarse de manera análoga a este método de orden inverso. Se observa que los métodos de la invención que comprenden las reacciones en orden inverso pueden ser particularmente adecuados para aplicaciones en las que se prefiere tener la primera y la segunda tetrazina en su lugar antes de hacerlas reaccionar con los dienófilos. Este puede ser el caso de ciertas aplicaciones en sistemas no biológicos, por ejemplo, el etiquetado de superficies, donde hay más libertad para elegir adecuadamente las condiciones para tener en cuenta las propiedades de las tetrazinas.

En resumen, la presente invención se refiere en particular a las siguientes realizaciones E1 a E110:

25 E1. Un método para formar uniones mediante reacciones de cicloadición, en el que el método comprende hacer reaccionar una primera tetrazina con un primer dienófilo seguido de hacer reaccionar una segunda tetrazina con un segundo dienófilo, en donde la reacción de la primera tetrazina con el primer dienófilo se desarrolla en presencia del segundo dienófilo, en donde

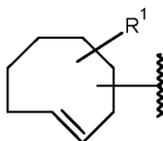
(i) la primera tetrazina comprende un grupo de la fórmula:



en la que

35 R³ es alquilo C₁-C₃;

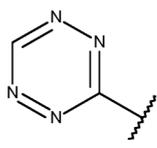
(ii) el primer dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:



40 en la que

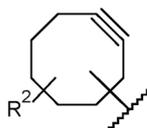
45 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

(iii) la segunda tetrazina comprende un grupo de la fórmula:



y

5 (iv) el segundo dienófilo comprende un grupo ciclooctinilo de la fórmula:



en la que

10 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcaniloxi C₂-C₅, alquilaminocarbonilo C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

15 R^c, R^d independientemente son hidrógeno o alcaniloximetilo C₂-C₅.

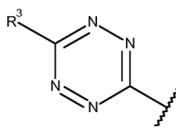
E2. El método de la realización E1, en el que la primera tetrazina reacciona preferiblemente con el primer dienófilo en presencia del segundo dienófilo.

20 E3. El método de la realización E1 o E2, en el que la constante de velocidad k de la primera tetrazina con el primer dienófilo es normalmente al menos 10² veces mayor que la constante de velocidad k de la reacción de la primera tetrazina con el segundo dienófilo.

25 E4. El método de una cualquiera de las realizaciones E1-E3, en el que la constante de velocidad k de la primera tetrazina con el primer dienófilo se deja avanzar durante 30 minutos o menos a una temperatura de aproximadamente 37 °C.

E5. El método de una cualquiera de las realizaciones E1-E4, en el que el método comprende poner en contacto una molécula diana o una composición de molécula diana con

30 (i) un primer agente de etiquetado que comprende un grupo de la fórmula:

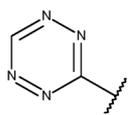


35 en la que

R³ es alquilo C₁-C₃;

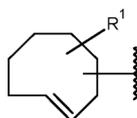
seguido de

40 (ii) un segundo agente de etiquetado que comprende un grupo de la fórmula:



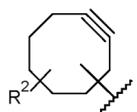
45 en la que la molécula diana comprende

(i) un grupo *trans*-ciclooctenilo de la fórmula:



en la que

- 5 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 10 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y (ii) un grupo ciclooctinilo de la fórmula:



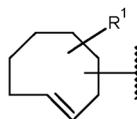
en la que

- 15 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 20 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

en la que la composición de moléculas diana comprende

- (i) un primera molécula diana que comprende un grupo *trans*-ciclooctenilo de la fórmula:

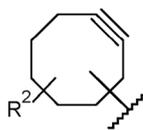
25



en la que

- 30 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 35 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

- (ii) una segunda molécula diana que comprende un grupo ciclooctinilo de la fórmula:

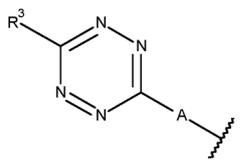


en la que

- 40 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 45 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

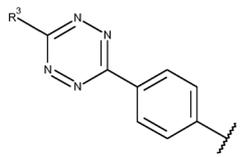
E6. El método de una cualquiera de las realizaciones E1-E5, en el que la primera tetrazina o el agente de etiquetado comprende un grupo de la fórmula:

50



en la que R³ es alquilo C₁-C₃ y A es 1,4-fenileno o alquileno C₁-C₆.

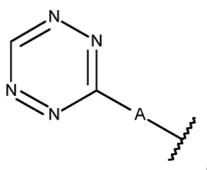
- 5 E7. El método de una cualquiera de las realizaciones E1-E6, en el que la primera tetrazina o el agente de etiquetado comprende un grupo de la fórmula:



10 en la que R³ es alquilo C₁-C₃.

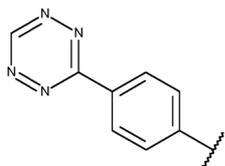
E8. El método de una cualquiera de las realizaciones E1-E7, en el que R³ es metilo.

- 15 E9. El método de una cualquiera de las realizaciones E1-E8, en el que la segunda tetrazina o el agente de etiquetado comprende un grupo de la fórmula:



20 en la que A es 1,4-fenileno o alquileno C₁-C₆.

E10. El método de una cualquiera de las realizaciones E1-E9, en el que la segunda tetrazina o el agente de etiquetado comprende un grupo de la fórmula:



- 25 E11. El método de una cualquiera de las realizaciones E1-E10, en el que el agente de etiquetado comprende un marcador seleccionado del grupo que consiste en tintes, radiomarcadores, etiquetas de giro sensibles a la RM, etiquetas de afinidad, grupos de pegilación y compuestos bioactivos.

- 30 E12. El método de la realización E11, en el que los tintes se seleccionan del grupo que consiste en tintes fluorescentes, luminiscentes y fosforescentes.

- E13. El método de la realización E11, en el que los tintes se seleccionan del grupo que consiste en dansilo, cumarina, fluoresceína, acridina, rodamina, silicio-rodamina, BODIPY, y cianina.

- 35 E14. El método de la realización E11, en el que las etiquetas de afinidad se seleccionan del grupo que consiste en biotina, etiqueta His, etiqueta Flag, etiqueta strep, azúcares, lípidos, esteroides, enlazadores de PEG, bencilguaninas, bencilcitosinas, y co-factores.

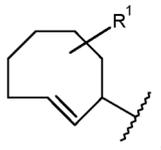
- 40 E15. El método de la realización E11, en el que los radiomarcadores se seleccionan del grupo que consiste en formas radiactivas de hidrógeno, flúor, carbono, fósforo, azufre, y yodo, incluyendo tritio, flúor-18, carbono-11, carbono-14, fósforo-32, fósforo-33, azufre-33, azufre-35, yodo-123, y yodo-125.

- 45 E16. El método de la realización E11, en el que los compuestos bioactivos se seleccionan de compuestos citotóxicos; compuestos antiviricos; modificadores de la respuesta biológica; agentes que afectan a los

microtúbulos; moduladores de hormonas; compuestos esteroideos.

E17. El método de una cualquiera de las realizaciones E1-E16, en el que el grupo *trans*-ciclooctenilo es un grupo de la fórmula:

5



en la que

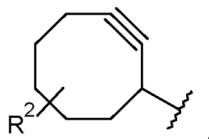
- 10 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 15 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

E18. El método de una cualquiera de las realizaciones E17, en el que R¹ es hidrógeno.

E19. El método de una cualquiera de las realizaciones E1-E18, en el que el grupo *trans*-ciclooctenilo está unido a un residuo aminoacídico.

20

E20. El método de una cualquiera de las realizaciones E1-E19, en el que el grupo ciclooctinilo es un grupo de la fórmula:



25

en la que

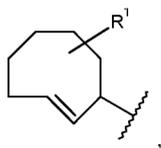
- 30 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 35 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

E21. El método de una cualquiera de las realizaciones E1-E20, en el que R² es hidrógeno.

E22. El método de una cualquiera de las realizaciones E1-E21, en el que el grupo ciclooctinilo está unido a un residuo aminoacídico.

40

E23. El método de una cualquiera de las realizaciones E1-E22, en el que el grupo *trans*-ciclooctenilo tiene la fórmula:



45

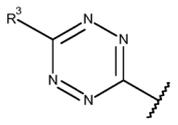
en la que

R¹ es hidrógeno;

y

la primera tetrazina o el agente de etiquetado comprende un grupo de la fórmula:

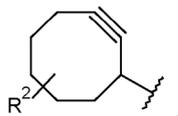
50



en la que

5 R³ es metilo.

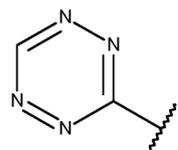
E24. El método de una cualquiera de las realizaciones E1-E23, en el que el grupo ciclooctinilo tiene la fórmula:



10 en la que

R² es hidrógeno;

15 y la segunda tetrazina o el agente de etiquetado comprende un grupo de la fórmula:



20 E25. El método de una cualquiera de las realizaciones E1-E24, en el que la primera tetrazina o agente de etiquetado reacciona con los grupos *trans*-ciclooctenilo.

25 E26. El método de una cualquiera de las realizaciones E1-E25, en el que la primera tetrazina o el agente de etiquetado se hacen reaccionar con el primer dienófilo o la molécula diana o la composición de moléculas diana en condiciones que no permiten que el primer agente de etiquetado reaccione sustancialmente con los grupos ciclooctinilo.

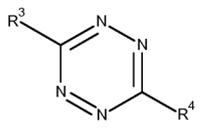
30 E27. El método de una cualquiera de las realizaciones E1-E26, en el que el primer dienófilo o la molécula diana o la composición de moléculas diana se hace reaccionar con la primera tetrazina o agente de etiquetado en condiciones que permiten que sustancialmente todos los grupos *trans*-ciclooctenilo reaccionen antes de reaccionar el segunda dienófilo o la molécula diana o composición de moléculas diana con la segunda tetrazina o agente de etiquetado.

35 E28. El método de una cualquiera de las realizaciones E1-E27, en el que el primer dienófilo o la molécula diana o la composición de moléculas diana se hacen reaccionar con un exceso molar de la primera tetrazina o agente de etiquetado basándose en los grupos *trans*-ciclooctenilo.

40 E29. El método de una cualquiera de las realizaciones E1-E28, en el que la primera tetrazina o el agente de etiquetado sin reaccionar se eliminan antes de hacer reaccionar el segundo dienófilo o la molécula diana o la composición de moléculas diana con la segunda tetrazina o el agente de etiquetado.

45 E30. El método de una cualquiera de las realizaciones E1-E29, en el que los grupos *trans*-ciclooctenilo sin reaccionar se hacen reaccionar con un inactivador antes de hacer reaccionar el segundo dienófilo o la molécula diana o la composición de moléculas diana con la segunda tetrazina o agente marcador.

E31. El método de la realización E30, en el que el inactivador es un compuesto de la fórmula:



o una sal de adición de ácidos del mismo, en la que

R³ es alquilo C₁-C₃; y
R⁴ es un radical orgánico.

5 E32. El método de la realización E31, en el que el inactivador es (4-(6-metil-1,2,4,5-tetrazina-3-il)fenil)metanamina o una sal de adición de ácidos de la misma.

10 E33. El método de una cualquiera de las realizaciones E30-E32, en el que el inactivador sin reaccionar se elimina de la composición antes de hacer reaccionar el segundo dienófilo o la molécula diana o la composición de moléculas diana con la segunda tetrazina o agente de etiquetado.

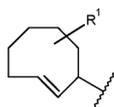
E34. El método de una cualquiera de las realizaciones E1-E33, en el que la composición es un sistema biológico.

15 E35. El método de la realización E34, en el que el sistema biológico es un organismo o una muestra biológica.

E36. El método de la realización E34 o E35, en el que la muestra biológica comprende una célula.

20 E37. El método de una cualquiera de las realizaciones E1-E36, en el que la molécula diana se selecciona del grupo que consiste en polipéptidos, oligonucleótidos, glicanos y lípidos.

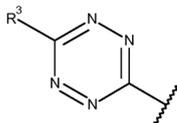
E38. El método de una cualquiera de las realizaciones E1-E37, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



en el que R¹ es como se define en una cualquiera de las realizaciones E1-E37.

30 E39. Un kit que comprende

(i) una primera tetrazina que comprende un grupo de la fórmula:

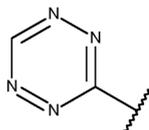


en la que

R³ es alquilo C₁-C₃;

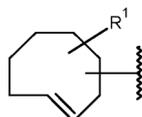
y

40 (ii) una segunda tetrazina que comprende un grupo de la fórmula:



E40. Un kit que comprende

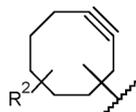
(i) un primer agente de modificación que comprende un grupo *trans*-ciclooctenilo de la fórmula:



en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

(ii) un segundo agente de modificación que comprende un grupo ciclooctinilo de la fórmula:



en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

E41. El kit de la realización E40, en el que el primer agente de modificación es un aminoácido no natural o un análogo del mismo.

E42. El kit de la realización E40 o E41, en el que el segundo agente de modificación es un aminoácido no natural o un análogo del mismo.

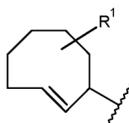
E43. El kit de la realización E40, en el que el primer agente de modificación es un nucleótido no natural o un análogo del mismo.

E44. El kit de la realización E40 o E43, en el que el segundo agente de modificación es un nucleótido no natural o un análogo del mismo.

E45. El kit de la realización E40, en el que el primer agente de modificación es un glicano no natural o un análogo del mismo.

E46. El kit de la realización E40 o E45, en el que el segundo agente de modificación es un glicano no natural o un análogo del mismo.

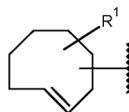
E47. El kit de una cualquiera de las realizaciones E40-E46, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



en el que R1 es como se define en una cualquiera de las realizaciones E40-E46.

E48. Una célula que comprende

(i) un grupo *trans*-ciclooctenilo de la fórmula:



en la que

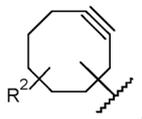
R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-

R^a, R^b alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

y

5

(ii) un grupo ciclooctinilo de la fórmula:



en la que

10

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarbonilo C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

15

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

E49. La célula de la realización E48, que es una célula de mamífero.

E50. La célula de la realización E48 o E49, en la que el grupo *trans*-ciclooctenilo está unido a un polipéptido.

20

E51. La célula de una cualquiera de las realizaciones E48-E50, en la que el grupo ciclooctinilo está unido a un polipéptido.

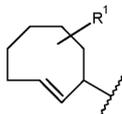
E52. La célula de una cualquiera de las realizaciones E48-E51, en la que el grupo ciclooctinilo y el grupo *trans*-ciclooctenilo están unidos al mismo polipéptido.

25

E53. La célula de una cualquiera de las realizaciones E48-E51, en la que el grupo ciclooctinilo está unido a un primer polipéptido y el grupo *trans*-ciclooctenilo está unido a un segundo polipéptido, el primer y el segundo polipéptido son polipéptidos diferentes.

30

E54. La célula de una cualquiera de las realizaciones E48-E53, en la que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



35

en la que R¹ es como se define en una cualquiera de las realizaciones E48-E53.

E55. Un método para preparar la célula de una cualquiera de las realizaciones E48-E53, que comprende

40

a) proporcionar una célula que comprende:

(i) una primera aminoacil ARNt sintetasa, o un polinucleótido que la codifica; y opcionalmente una segunda aminoacil ARNt sintetasa, o un polinucleótido que la codifica;

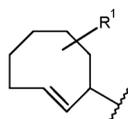
45

(ii) un primer ARNt que tiene un anticodón para un primer codón selector, o un polinucleótido que codifica dicho ARNt; y opcionalmente un segundo ARNt que tiene un anticodón para un segundo codón selector, o un polinucleótido que codifica dicho ARNt; y y

(iii) un polinucleótido que codifica un polipéptido diana y que comprende uno o más de un primer y segundo codón o codones selectores; o un primer polinucleótido que codifica un primer polipéptido diana y que comprende uno o más de un primer codón selector y un segundo polinucleótido que codifica un segundo polipéptido diana y que comprende uno o más de un segundo codón selector,

50

en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural o un análogo del mismo, que comprende un grupo *trans*-ciclooctenilo de fórmula:



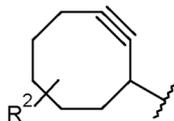
55

en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

y con un segundo aminoácido no natural o un análogo del mismo, que comprende un grupo ciclooctinilo de la fórmula:



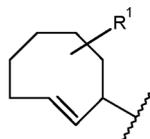
en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

o

en la que dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural o un análogo del mismo, que comprende un grupo *trans*-ciclooctenilo de fórmula:

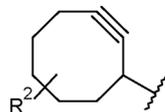


en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

dicha segunda aminoacil ARNt sintetasa (i) es capaz de acilar el segundo ARNt (ii) con un segundo aminoácido no natural o un análogo del mismo, que comprende un grupo ciclooctinilo de la fórmula:



en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

b) poner en contacto la célula con el primer y el segundo aminoácido no natural o un análogo del mismo; y

c) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el primer y el segundo aminoácidos no naturales o los análogos de los mismos, en el polipéptido o polipéptidos diana en la posición o posiciones codificadas por el codón o codones selectores.

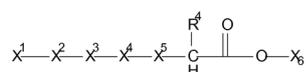
E56. El método de la realización E55, en el que la célula se pone en contacto con el primer y el segundo

aminoácidos no naturales o los análogos de los mismos, secuencialmente.

E57. El método de la realización E53, que comprende

- 5 a) poner en contacto la célula con el primer aminoácido no natural o el análogo del mismo; y
 b) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el primer aminoácido no natural o el análogo del mismo en el polipéptido diana en la posición o posiciones codificadas por el codón o codones selectores;
 10 c) poner en contacto la célula con el segundo aminoácido no natural o el análogo del mismo; y
 d) permitir la traducción del polinucleótido o polinucleótidos (iii) incorporando de este modo el segundo aminoácido no natural o el análogo del mismo en el polipéptido diana en la posición o posiciones codificadas por el codón o codones selectores.

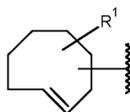
15 E58. El método de una cualquiera de las realizaciones E55-E57, en el que el primer aminoácido no natural o el análogo del mismo es un compuesto de la fórmula:



en la que:

20

X¹ tiene la fórmula:



25 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^a, R^b

independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

30

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

X⁵ es -(CH₂)_n- o fenileno-CH₂-;

35

X⁶ es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilsulfanil C₂-C₇-alquilo C₁-C₂;

R⁴ es -OH o -NH₂;

n es un número entero de 1 a 4;

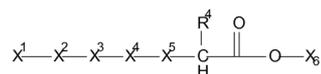
m es un número entero de 1 a 6; y

40

p es un número entero de 1 a 6,

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

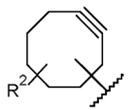
45 E59. El método de una cualquiera de las realizaciones E55-E58, en el que el segundo aminoácido no natural o el análogo del mismo es un compuesto de la fórmula:



en la que:

50

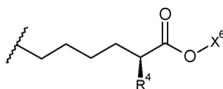
X¹ tiene la fórmula:



- 5 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^oO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 5 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;
- X^2 es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;
- X^3 es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;
- X^4 es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;
- 10 X^5 es -(CH₂)_n- o fenileno-CH₂-;
- X^6 es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilsulfanil C₂-C₇-alquilo C₁-C₂;
- R^4 es -OH o -NH₂;
- 15 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,

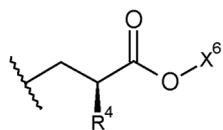
o una sal de adición de ácidos o bases del mismo.

- 20 E60. El método de la realización E58 o E59, en el que X^2 es -O-.
- E61. El método de una cualquiera de las realizaciones E58-E60, en el que X^3 es -CH₂-CH₂-O- o un enlace sencillo.
- 25 E62. El método de una cualquiera de las realizaciones E58-E61, en el que el elemento estructural - X^2 - X^3 - comprende de 1 a 6 átomos en la cadena principal.
- E63. El método de una cualquiera de las realizaciones E58-E62, en el que X^4 es -NH-, -C(O)-NH-, -NH-CH(NH₂)-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)- o -C(O)-NH-C(NH)-NH-.
- 30 E64. El método de una cualquiera de las realizaciones E58-E63, en el que X^4 es -C(O)-NH-.
- E65. El método de una cualquiera de las realizaciones E55-E64, en el que n es 3 o 4.
- 35 E66. El método de una cualquiera de las realizaciones E58-E65, en el que el elemento estructural - X^2 - X^3 - X^4 -(CH₂)_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.
- E67. El método de una cualquiera de las realizaciones E58-E66, en el que X^6 es hidrógeno, alcoximetilo C₁-C₆, alcoxi-1-ilo C₁-C₆, alcanoiloximetilo C₂-C₇ o alcanoilsulfaniletilo C₂-C₇.
- 40 E68. El método de una cualquiera de las realizaciones E58-E66, en el que X^6 es hidrógeno.
- E69. El método de una cualquiera de las realizaciones E58-E68, en el que el compuesto tiene la configuración S con respecto al átomo de carbono asimétrico que lleva R⁴.
- 45 E70. El método de una cualquiera de las realizaciones E58-E69, en el que - X^5 -CHR⁴-C(O)O- X^6 tiene la fórmula:



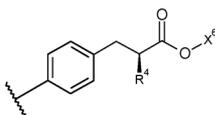
50 en la que R⁴ y X⁶ son como se definen en una cualquiera de las realizaciones E58-E69.

E71. El método de una cualquiera de las realizaciones E58-E69, en el que - X^5 -CHR⁴-C(O)O- X^6 tiene la fórmula:



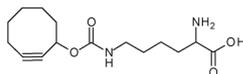
55 en la que R⁴ y X⁶ son como se definen en una cualquiera de las realizaciones E58-E69.

E72. El método de una cualquiera de las realizaciones E58-E69, en el que - X^5 -CHR⁴-C(O)O- X^6 tiene la fórmula:

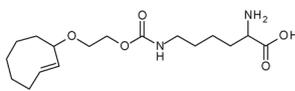


en la que R⁴ y X⁶ son como se definen en una cualquiera de las realizaciones E58-E69.

- 5 E73. El método de una cualquiera de las realizaciones E55-E57, en el que el primer aminoácido no natural es un compuesto de la fórmula:



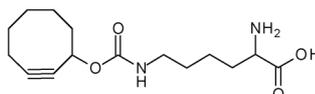
- 10 o



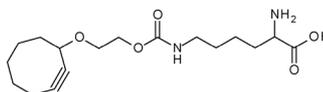
o una sal de adición de ácidos o bases del mismo.

- 15 E74. El método de la realización E73, en el que el primer aminoácido no natural es un isómero axial con respecto a su grupo *trans*-ciclooctenilo.

- 20 E75. El método de una cualquiera de las realizaciones E55-E57, E73 y E74, en el que el segundo aminoácido no natural es un compuesto de la fórmula:



- 25 o



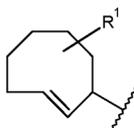
o una sal de adición de ácidos o bases del mismo.

- 30 E76. El método de una cualquiera de las realizaciones E55-E75, en el que dicho sistema de traducción es una célula que expresa dicha(s) aminoacil ARNt sintetasa(s).

E77. El método de la realización E76, en el que dicha aminoacil ARNt sintetasa es una pirrolisil ARNt sintetasa.

- 35 E78. El método de la realización E77, en el que dicha pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:1 o 2.

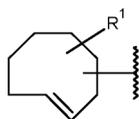
- 40 E79. El método de una cualquiera de las realizaciones E55-E72 y E75-E78, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



en la que R¹ es como se define en una cualquiera de las realizaciones E55-E72 y E75-E78.

- 45 E80. Un polipéptido que comprende

(i) un grupo *trans*-ciclooctenilo de la fórmula:

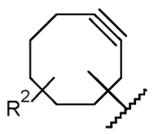


en la que

- 5 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 10 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

y

(ii) un grupo ciclooctinilo de la fórmula:

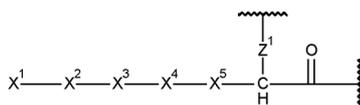


15

en la que

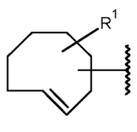
- 20 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
- 25 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

E81. El polipéptido de la realización E80, que comprende un residuo de la fórmula:



en la que:

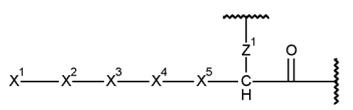
- 30 X^1 tiene la fórmula



- 35 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

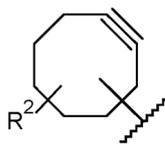
- 40 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;
- X^2 es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;
- X^3 es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;
- X^4 es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;
- 45 X^5 es -(CH₂)_n- o fenileno-CH₂-;
- 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,

50 y un residuo de la fórmula:



en la que:

5 X¹ tiene la fórmula:



10 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^oO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

15 X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

X⁵ es -(CH₂)_n- o fenileno-CH₂-;

Z¹ es -O- o -NH-;

20 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.

E82. El polipéptido de la realización E81, en el que X² es -O-.

25 E83. El polipéptido de la realización E81 o E82, en el que X³ es -CH₂-CH₂-O- o un enlace sencillo.

E84. El polipéptido de una cualquiera de las realizaciones E81-E83, en el que el elemento estructural -X²X³- comprende de 1 a 6 átomos en la cadena principal.

30 E85. El polipéptido de una cualquiera de las realizaciones E81-E84, en el que X⁴ es -NH-, -C(O)-NH-, -NH-CH(NH₂)-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)- o -C(O)-NH-C(NH)-NH-.

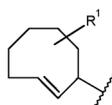
E86. El polipéptido de una cualquiera de las realizaciones E81-E84, en el que X⁴ es -C(O)-NH-.

35 E87. El polipéptido de una cualquiera de las realizaciones E81-E86, en el que n es 3 o 4.

E88. El polipéptido de una cualquiera de las realizaciones E81-E87, en el que el elemento estructural -X²-X³-X⁴-(CH₂)_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.

40 E89. El polipéptido de una cualquiera de las realizaciones E81-E88, en el que el compuesto tiene la configuración S con respecto al átomo de carbono asimétrico que lleva Z¹.

45 E90. El polipéptido de una cualquiera de las realizaciones E80-E89, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



50 en el que R¹ es como se define en una cualquiera de las realizaciones E80-E89.

E91. Un método para preparar el polipéptido de una cualquiera de las realizaciones E80-E89, comprendiendo el método:

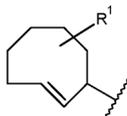
a) proporcionar un sistema de traducción que comprende:

55

- (i) una primera aminoacil ARNt sintetasa, o un polinucleótido que la codifica; y opcionalmente una segunda aminoacil ARNt sintetasa, o un polinucleótido que la codifica;
- (ii) un primer y un segundo aminoácido no natural o un análogo del mismo;
- (iii) un primer ARNt que tiene un anticodón para un primer codón selector, o un polinucleótido que codifica dicho ARNt; y opcionalmente un segundo ARNt que tiene un anticodón para un segundo codón selector, o un polinucleótido que codifica dicho ARNt; y y
- (iv) un polinucleótido que codifica un polipéptido diana y que comprende uno o más de un primer y segundo codón o codones selectores,

5

10 en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (iii) con el primer aminoácido no natural o el análogo del mismo (ii) que comprende un grupo *trans*-ciclooctenilo de la fórmula:



15 en la que

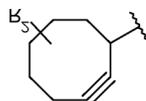
R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

20

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

dicha segunda aminoacil ARNt sintetasa (i) es capaz de acilar el segundo ARNt (iii) con el segundo aminoácido no natural o el análogo del mismo (ii) que comprende un grupo ciclooctinilo de la fórmula:

25



en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

30

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

35

b) permitir la traducción del polinucleótido o polinucleótidos (iv) incorporando de este modo el primer y el segundo aminoácidos no naturales o los análogos de los mismos, en el polipéptido en la posición o posiciones codificadas por el codón o codones selectores.

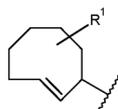
40 E92. El método de la realización E91, en el que dicho sistema de traducción es una célula que expresa dicha(s) aminoacil ARNt sintetasa(s).

E93. El método de la realización E92, en el que dicha aminoacil ARNt sintetasa es una pirrolisil ARNt sintetasa.

45 E94. El método de la realización E93, en el que dicha pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:1 o 2.

E95. El método de una cualquiera de las realizaciones E91-E94, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:

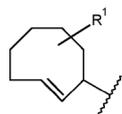
50



en el que R¹ es como se define en una cualquiera de las realizaciones E91-E94.

E96. Un aminoácido no natural que comprende un grupo *trans*-ciclooctenilo de la fórmula:

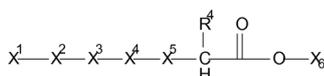
55



en la que

- 5 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- 10 R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,
o un análogo del aminoácido no natural.

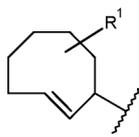
E97. el aminoácido no natural de la realización E96, que tiene la fórmula:



15

en la que

- 20 X^1 es un grupo *trans*-ciclooctenilo de la fórmula:



- 25 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
- R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.
- X^2 es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;
- 30 X^3 es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p-, o un enlace sencillo;
- X^4 es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;
- X^5 es -(CH₂)_n- o fenileno-CH₂-;
- X^6 es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilulfanil C₂-C₇-alquilo C₁-C₂;
- 35 R^4 es -OH o -NH₂;
- n es un número entero de 0 a 4;
- m es un número entero de 1 a 6; y
- p es un número entero de 1 a 6,

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

E98. El aminoácido no natural o sal de la realización E96 o E97, en el que R^2 es hidrógeno.

45 E99. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E98, en el que X^2 es -O-.

E100. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E99, en el que X^3 es -CH₂-CH₂-O- o un enlace sencillo.

50 E101. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E100, en el que el elemento estructural - X^2 - X^3 - comprende de 1 a 6 átomos en la cadena principal.

E102. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E101, en el que X^4 es -NH-, -C(O)-NH-, -NH-CH(NH₂)-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)- o -C(O)-NH-C(NH)-NH-.

55 E103. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E102, en el que X^4 es -C(O)-NH-

E104. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E103, en el que X^5 es $-(CH_2)_n-$ en la que n se define como en una cualquiera de las realizaciones E96-E103.

5

E105. El compuesto o sal de una cualquiera de las realizaciones E96-E104, en el que n es 3 o 4.

E106. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E105, en el que 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.

10

E107. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E106, en el que X^6 es hidrógeno, alcóximetilo C_1-C_6 , alcóxiet-1-ilo C_1-C_6 , alcanoiloximetilo C_2-C_7 o alcanoilsulfaniletilo C_2-C_7 .

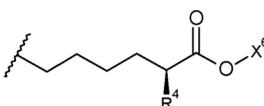
15

E108. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E106, en el que X^6 es hidrógeno.

E109. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E108 que tiene la configuración *S* con respecto al átomo de carbono asimétrico que lleva R^4 .

20

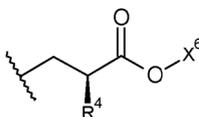
E110. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E109, en el que $-X^5-CHR^4-C(O)O-X^6$ tiene la fórmula:



25

en la que R^4 y X^6 son como se definen en una cualquiera de las realizaciones E96-E109.

E111. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E109, en el que $-X^5-CHR^4-C(O)O-X^6$ tiene la fórmula:

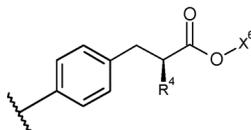


30

en la que R^4 y X^6 son como se definen en una cualquiera de las realizaciones E96-E109.

E112. El aminoácido no natural o sal de una cualquiera de las realizaciones E96-E109, en el que $-X^5-CHR^4-C(O)O-X^6$ tiene la fórmula:

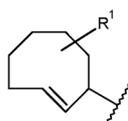
35



40

en la que R^4 y X^6 son como se definen en una cualquiera de las realizaciones E96-E109.

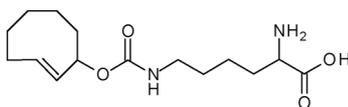
E113. El método de una cualquiera de las realizaciones E96-E112, en el que el grupo *trans*-ciclooctenilo es un isómero axial del grupo de la fórmula:



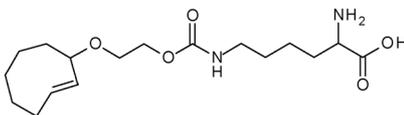
45

en la que R^1 es como se define en una cualquiera de las realizaciones E96-E112.

E114. El aminoácido no natural o sal de la realización E96 que es un compuesto de la fórmula:



o



5

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

10

E115. El aminoácido no natural o sal de adición de bases del mismo, de acuerdo con la realización E114 que es un isómero axial con respecto a su grupo *trans*-ciclooctenilo.

E116. Un método para preparar un polipéptido que comprende un grupo *trans*-ciclooctenilo, comprendiendo el método:

15

a) proporcionar un sistema de traducción que comprende:

20

- (i) una aminoacil ARNt sintetasa o un polinucleótido que la codifica;
- (ii) un aminoácido no natural o sal de una cualquiera de las realizaciones E96-E115;
- (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 el que la aminoacil ARNt sintetasa (i) es capaz de acilar 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.

30

E117. El método de la realización E116, en el que dicho sistema de traducción es una célula que expresa dicha aminoacil ARNt sintetasa.

E118. El método de la realización E117, en el que dicha aminoacil ARNt sintetasa es una pirrolisil ARNt sintetasa.

E119. El método de la realización E118, en el que dicha pirrolisil ARNt sintetasa comprende la secuencia de aminoácidos expuesta en la SEQ ID NO:1 o 2.

Ejemplos

35

Ejemplos de preparación

Materiales y métodos generales

40

A menos que se indique otra cosa, los materiales para las síntesis químicas se obtuvieron de proveedores comerciales (Acros, Alfa Aesar, Fox-Chemicals, GL Biochem, Sigma-Aldrich) con la pureza más alta disponible y se usaron sin purificación adicional. Se adquirieron disolventes secos en Sigma-Aldrich y Acros, se almacenaron sobre tamices moleculares y se usaron tal como se suministraron. Los disolventes usados para extracción y cromatografía se adquirieron en Acros, Fisher Scientific, y BDH Prolabo (VWR). Las tetrazinas y azidas de fluoróforos se adquirieron en Life Technologies (Darmstadt, Alemania), ATTO-TEC (Siegen, Alemania) o Jena Bioscience (Jena, Alemania). Los disolventes deuterados se obtuvieron en Deutero GmbH (Kastellaun, Alemania). La cromatografía ultrarrápida se realizó utilizando gel de sílice Macherey-Nagel 60 (0,04-0,063 mm, malla 230-400) y los sistemas de disolventes como se describe a continuación. La cromatografía en capa fina (TLC) se realizó en placas de gel de sílice recubiertas previamente, con recubrimiento de aluminio (gel de sílice Merck TLC 60 F254) con mezclas (en porcentaje en volumen) de C₆H₁₂/EtOAc, CH₂Cl₂/MeOH(/AcOH), o acetona/MeOH/H₂O como eluyentes. Las manchas se detectaron con una lámpara de mano UV a $\lambda = 254$ nm o $\lambda = 366$ nm o tinción con a) solución de tinción de anisaldehído (85 ml de EtOH, 10 ml de AcOH, 5 ml de H₂SO₄ concentrado, 0,5 ml de anisaldehído), a) una solución de tinción de KMnO₄ (3,0 g de KMnO₄, 20 g de K₂CO₃ en 300 ml NaOH acuoso al 5 %), o c) una solución de tinción de ninhidrina (250 ml de EtOH, 1,5 ml de AcOH, 5,0 g de ninhidrina) y posterior tratamiento térmico. El análisis por HPLC de fase inversa (RP) C18 se realizó en un sistema Waters (detector de absorbancia dual λ Waters 2487, bomba de HPLC binaria Waters 1525) usando un gradiente de concentración creciente de disolvente B (acetonitrilo con TFA al 0,1 %) a partir del 100 % de disolvente A (agua con TFA al 0,1 %). Los espectros de RMN se registraron a 25 °C utilizando un espectrómetro Bruker UltraShieldTM Advance 400 (400 MHz, ¹H; 100 MHz, ¹³C). Los desplazamientos químicos δ se refieren a señales de disolventes protonados residuales como patrón interno (por ejemplo, CDCl₃: $\delta = 7,26$ (¹H), 77,16 (¹³C) ppm)^[11]. Las

55

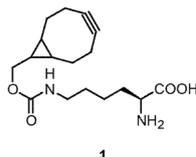
asignaciones de ^1H y ^{13}C se basan en APT y datos de espectroscopia de correlación bidimensional

(H,H-COSY). Las multiplicidades de señal $^3\text{J}(\text{H},\text{H})$ se abrevian como s (singlete), a (singlete ancho), d (doblete), dd (doblete de dobletes), dc (doblete de cuadrupletes), t (triplete), c (cuadruplete), dt (doblete de tripletes), o m (multiplete).

5 Los espectros de masas de alta resolución (HR) se registraron en la Universidad de Heidelberg utilizando espectrometría de masas de ionización por electropulverización (ESI) en un Bruker ApexQe 9.4 T FT-ICR híbrido o utilizando bombardeo de átomos rápidos (FAB) e ionización de electrones (EI, impacto de electrones), respectivamente, en un espectrómetro de masas de sector magnético JEOL JMS-700. Los productos se caracterizaron por RMN (^1H , ^{13}C) y/o MS/HR MS.

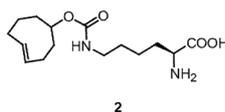
10

Ejemplos 1: N-ε-((1R,8S,9S)-biciclo[6.1.0]non-4-in-9-metiloxi)carbonil)-L-lisina



15 Se adquirió el aminoácido a base de lisina no natural **1** en Sirius Fine Chemicals SiChem (Bremen, Alemania). El compuesto **1** se usó como una mezcla del isómero *endo* y *exo*. El compuesto **1** también se denomina BCN.

Ejemplo 2: N-ε-((trans-Ciclooct-4-en-1-iloxi)carbonil)-L-lisina



20

El compuesto **2** se sintetizó como se describe en el documento WO 2012/051885 y T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2012, *51*, 4166-4170. El compuesto **2** también se denomina TCO.

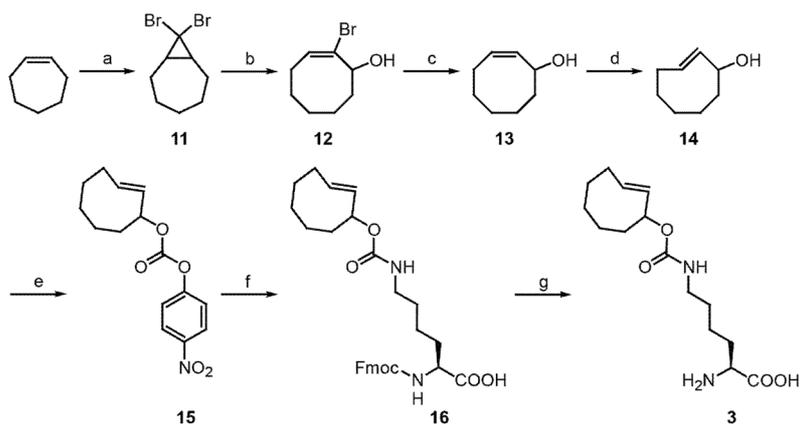
25

Ejemplo 3: N-ε-((trans-Ciclooct-2-en-1-iloxi)carbonil)-L-lisina

El Esquema S1 muestra la síntesis del derivado de lisina *trans*-ciclooct-2'-eno-funcionalizado **3** (también denominado TCO*).

30

Esquema S1:



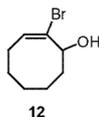
35 Reactivos y condiciones: a) KO^tBu, CHBr₃, pentano, 0 °C a ta, durante una noche; b) AgClO₄, acetona/H₂O, ta, 1 h; c) ^tBuLi en pentano, Et₂O, -78 °C a -20 °C, 4 h; d) h·ν (λ = 254 nm), benzoato de metilo, C₆H₁₂/EtOAc, ta, 8 h; e) cloroformato de 4-nitrofenilo, piridina, CH₂Cl₂, ta 2 h; f) Fmoc-Lys-OH·HCl, DIEA; DMSO, ta, durante una noche; g) piperidina, CH₂Cl₂, ta, 30 min.



40

El 8,8-dibromobicyclo[5.1.0]octano **11** conocido en la bibliografía se sintetizó partiendo de *cis*-ciclohepteno disponible en el mercado como se ha indicado anteriormente (A. B. Neef, C. Schultz, *Angewandte Chemie* 2009, *48*, 1498-1500) y se usó sin purificación cromatográfica.

^1H RMN (CDCl_3) δ = 2,32-2,21 (m, 2H), 1,94-1,78 (m, 3H), 1,77-1,66 (m, 2H), 1,45-1,30 (m, 2H), 1,27-1,10 (m, 3H) ppm.



Se sintetizó 2-bromociclooct-2-en-1-ol **12** conocido a partir de la bibliografía de acuerdo con protocolos descritos previamente (a) C. B. Reese, A. Shaw, *J Chem Soc Perk T 1* 1975, 2422-2434; b) H. J. J. Loozen, J. W. Dehaan, H. M. Buck, *Journal of Organic Chemistry* 1977, *42*, 418-422). En resumen, a una solución agitada de 8,8-dibromobicyclo[5.1.0]octano **11** (1,0 equiv.) en acetona (0,5 M) y agua (15,0 equiv.). Se añadió AgClO_4 (2,0 equiv.) en pequeñas porciones a *ta* durante 30 min y se agitó durante 60 min más. Se añadió 1 m de HCl hasta que la formación de un precipitado de color blanco se detuvo. La mezcla se filtró y el residuo se lavó con EtOAc. El filtrado se separó y la capa acuosa se extrajo con EtOAc (3 x). Las fracciones orgánicas combinadas se lavaron con salmuera, se secaron sobre Na_2SO_4 , se filtraron sobre gel de sílice, y se concentraron. El producto en bruto se purificó por cromatografía ultrarrápida (EtOAc al 20 % en C_6H_{12}) para producir **12** en forma de un líquido amarillo pálido.

F_r (EtOAc al 20 % en C_6H_{12}) = 0,4.

^1H RMN (CDCl_3) δ = 6,12 (dd, $^3\text{J}(\text{H,H})$ = 11,7, 4,2 Hz, 1H), 4,19 (dd, $^3\text{J}(\text{H,H})$ = 10,4, 5,3 Hz, 1H), 2,68 (dc, $^3\text{J}(\text{H,H})$ = 11,9, 5,5 Hz, 1H), 2,36-2,29 (m, 1H), 2,23-2,10 (m, 2H), 2,08-2,00 (m, 2H), 1,94-1,67 (m, 4H), 1,56-1,44 (m, 1H), 1,31-1,22 (m, 1H) ppm.



A una solución del 2-bromociclooct-2-en-1-ol **12** (1,0 equiv., 10,24 g, 49,7 mmol) en Et_2O seco (0,7 M, 71 ml) a -78°C se le añadió gota a gota $t\text{BuLi}$ (3,2 equiv., 159 mmol, 99,4 ml de una solución 1,6 M de $t\text{BuLi}$ en pentano) en una atmósfera de Ar durante 1 h. Después de la adición completa, la mezcla se agitó a -78°C durante 10 min más y después se dejó calentar a -20°C durante 3 h. La solución se inactivó mediante la adición de una solución ac. sat. de NaHCO_3 y se agitó durante 1 h a *ta*. Las capas se separaron y la capa orgánica se extrajo con EtOAc (3 x). A continuación, las fracciones orgánicas combinadas se lavaron con salmuera, se secaron sobre Na_2SO_4 , y se concentraron. El producto en bruto se purificó por FC (EtOAc al 20 % en C_6H_{12}) para producir el compuesto **13** (4,37 g, 34,6 mmol, rendimiento del 70 %) en forma de un líquido incoloro.

F_r (EtOAc al 20 % en C_6H_{12}) = 0,4.

^1H RMN (CDCl_3) δ = 5,71-5,50 (m, 2H), 4,70-4,60 (m, 0,3H), 4,32-4,22 (m, 0,7H), 2,45-2,35 (m, 0,7H), 2,20-2,04 (m, 1,3H), 2,02-1,34 (m, 6H), 1,10-1,03 (m, 0,6H), 0,94-0,84 (m, 0,7H), 0,81-0,70 (m, 0,7H) ppm (dos isómeros).

^{13}C RMN (CDCl_3) δ = 135,6, 134,9, 132,0, 128,7, 76,9, 69,5, 44,3, 38,9, 35,4, 29,1, 29,1, 27,7, 26,3, 25,9, 23,7 ppm (dos isómeros).

HR MS (El pos.) m/z : calc. para $\text{C}_8\text{H}_{14}\text{O}$ $[\text{M}+\text{e}]^+$: 126,1045, med.: 126,1040.

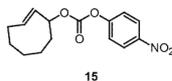


El compuesto **14** se preparó partiendo de su precursor *cis* **13** de acuerdo con un procedimiento descrito previamente (a) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand, R. Weissleder, *Angewandte Chemie* 2009, *48*, 7013-7016; b) M. Royzen, G. P. Yap, J. M. Fox, *Journal of the American Chemical Society* 2008, *130*, 3760-3761). La fotoisomerización con eliminación activa del isómero *trans* se realizó en un reactor Rayonet RPR-100 UV (Southern New England Ultraviolet Company, Branford, Connecticut, Estados Unidos) durante 8 h a *ta* (la temperatura en el interior del reactor era de aproximadamente 30°C). En resumen, el reactor UV se cargó con **13** (1,0 equiv., 1,89 g, 14,9 mmol), benzoato de metilo (1,5 equiv., 3,51 g, 2,80 ml, 22,4 mmol) y disolvente (C_6H_{12} al 20 % en Et_2O , 1000 ml). Después del tratamiento, se obtuvo un aceite de color amarillo (0,94 g, 7,42 mmol, 50 %) que se usó en la siguiente etapa sin purificación adicional. El análisis por RMN demostró la presencia de dos isómeros en la relación de 1,0:1,1. F_r (EtOAc al 20 % en C_6H_{12}) = 0,4.

^1H RMN (CDCl_3) δ = 6,00-5,91 (m, 0,5H), 5,71-5,62 (m, 0,5H), 5,61-5,51 (m, 1H), 4,62 (a, 0,5H), 4,27 (dt, $^3\text{J}(\text{H,H})$ = 9,5, 5,5 Hz, 0,5H), 2,52-2,45 (m, 0,5H), 2,43-2,36 (m, 0,5H), 2,20-2,12 (m, 0,5H), 2,08-1,92 (m, 2H), 1,90-1,79 (m, 1H), 1,77-1,66 (m, 2H), 1,64-1,37 (m, 2,5H), 1,16-1,06 (m, 0,5H), 0,94-0,84 (m, 0,5), 0,82-0,71 (m, 1H) ppm.

^{13}C RMN (CDCl_3) δ = 135,5, 135,1, 132,0, 130,5, 76,8, 71,2, 44,3, 43,2, 36,1, 35,8, 35,8, 35,4, 29,2, 29,0, 27,6, 23,1 ppm.

HR MS (El pos.) m/z : calc. para $\text{C}_8\text{H}_{14}\text{O}$ $[\text{M}+\text{e}]^+$: 126,1045, med.: 126,1044.



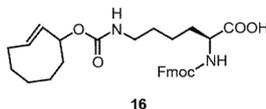
A una solución agitada del compuesto **14** (1,0 equiv., 0,94 g, 7,42 mmol) en CH₂Cl₂ (20 ml) se le añadió piridina (2,5 equiv., 1,50 ml, 18,6 mmol) en una atmósfera de Ar. Se añadió una solución de cloroforniato de 4-nitrofenilo (1,1 equiv., 1,65 g, 8,17 mmol) en CH₂Cl₂ (20 ml) a ta y la mezcla de reacción resultante se agitó durante 2 h. Se usó una solución ac. sat. de NH₄Cl (50 ml) para detener la reacción. Después de la separación en fases, la capa acuosa se extrajo con CH₂Cl₂ (2 x). Las capas orgánicas combinadas se lavaron con una solución ac. sat. de NaCl, se secaron sobre Na₂SO₄, y se concentraron. El producto en bruto se purificó por FC (EtOAc al 5 % en C₆H₁₂) para producir el compuesto **15** (1,63 g, 5,59 mmol, 75 %) en forma de una cera de color amarillo pálido (T_m= 74,6-76,8 °C).

F_r (EtOAc al 10 % en C₆H₁₂) = 0,4.

¹H RMN (CDCl₃) δ = 8,31-8,25 (m, 2H), 7,43-7,37 (m, 2H), 6,03-5,94 (m, 0,5H), 5,90-5,80 (m, 0,5H), 5,66 (dd, ³J(H,H)= 16,2, 9,5 Hz, 0,5H), 5,56 (dd, ³J(H,H)= 16,5, 2,2 Hz, 0,5H), 5,44 (a, 0,5H), 5,17 (dt, ³J(H,H)= 10,1, 5,6 Hz, 0,5H), 2,56-2,44 (m, 1H), 2,38-2,30 (m, 0,5H), 2,26-2,18 (m, 0,5H), 2,12-1,85 (m, 3H), 1,82-1,65 (m, 1,5H), 1,61-1,39 (m, 1,5H), 1,23-1,12 (m, 0,5H), 0,98-0,78 (m, 1,5H) ppm.

¹³C RMN (CDCl₃) δ = 155,6, 151,9, 151,7, 145,3, 134,8, 133,2, 130,1, 129,3, 125,3, 125,3, 121,8, 83,3, 78,8, 40,5, 40,4, 36,0, 35,9, 35,6, 35,5, 28,9, 28,8, 27,3, 24,0 ppm.

HR MS (FAB pos.) *m/z*: calc. para C₁₅H₁₇NO₅ [M+e]⁺: 291,1107, med.: 291,1116.



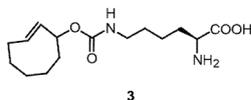
Se disolvió Fmoc-Lys-OH*HCl (2,0 equiv., 4,45 g, 11,0 mmol) en DIEA (3,0 equiv., 2,87 ml, 16,5 mmol) y DMSO anhidro (0,2 M, 27 ml) en una atmósfera de Ar. A continuación, se añadió gota a gota una solución transparente del compuesto **15** (1,0 equiv., 1,60 g, 5,49 mmol) en DMSO anhidro (0,2 M, 27 ml) a ta y en una atmósfera de Ar durante un periodo de 2 h. La mezcla de reacción se agitó durante una noche a ta. Se añadieron H₂O (50 ml) y EtOAc (150 ml) y el pH de la capa acuosa se ajustó a 1-3 con HCl conc. Las fases se separaron y la capa acuosa se extrajo con EtOAc (3 x). Las capas orgánicas combinadas se lavaron con salmuera (2*) y se secaron sobre Na₂SO₄. Todos los componentes volátiles se evaporaron a presión reducida y el producto en bruto se purificó por FC (MeOH al 10 % en CH₂Cl₂) para producir el compuesto **16** (1,45 g, 2,79 mmol, 51 %) en forma de una espuma de color blanco.

F_r (MeOH al 8 % y AcOH al 2 % en CH₂Cl₂) = 0,5.

¹H RMN (MeOD) δ = 7,79 (d, ³J(H,H)= 7,5 Hz, 2H), 7,67 (t, ³J(H,H) = 7,2 Hz, 2H), 7,39 (t, ³J(H,H) = 7,5 Hz, 2H), 7,31 (t, ³J(H,H) = 7,3 Hz, 2H), 5,88-5,64 (m, 1H), 5,58-5,43 (m, 1H), 5,22 (a, 0,5H), (4,98 (dt, ³J(H,H) = 9,9, 5,2 Hz, 0,5H), 4,37-4,32 (m, 2H), 4,22 (t, ³J(H,H) = 6,7 Hz, 1H), 4,13-4,04 (m, 1H), 3,09 (c, ³J(H,H) = 5,4 Hz, 2H), 2,45-2,34 (m, 1H), 2,16-2,08 (m, 0,5H), 2,05-1,34 (m, 10,5H), 1,19-1,07 (m, 0,5H), 0,92-0,75 (m, 1,5H) ppm.

¹³C RMN (MeOD) δ = 157,2, 144,0, 143,8, 143,8, 141,2, 132,4, 131,8, 131,5, 131,2, 127,4, 126,8, 124,9, 119,5, 78,5, 73,7, 66,5, 54,7, 47,1, 40,8, 40,3, 40,1, 40,0, 35,6, 35,4, 35,2, 34,9, 31,3, 29,1,28,7, 28,6, 27,0, 23,8, 27,8, 27,7 ppm.

HR MS (ESI neg.) *m/z*: calc. para C₃₀H₃₅N₂O₆ [M-H]⁻: 519,25006, med.: 519,25127.



El compuesto **16** (1,0 equiv., 1,43 g, 2,75 mmol) se disolvió en piperidina al 20 % en CH₂Cl₂ (40 ml v/v) y se agitó durante 30 min a ta. Después de la adición de H₂O (50 ml), todos los componentes volátiles se eliminaron a presión reducida y alto vacío. El producto en bruto se purificó a través de FC (acetona:MeOH:H₂O, 85:10:5 v/v/v para eliminar por lavado todas las impurezas seguido de 65:25:10 v/v/v para eluir el producto) sobre gel de sílice para producir el compuesto **3** (0,81 g, 2,70 mmol, 98 %) en forma de un polvo de color blanco. **2** purificado se descompuso con calentamiento (partiendo de 180 °C). El compuesto **3** también se denomina TCO* en esta publicación. El compuesto **3** era una mezcla de isómeros axiales de TCO* (denominados en el presente documento compuesto **3a** o TCO*^a) e isómeros ecuatoriales de TCO* (denominados en el presente documento **3b** o TCO*^e).

F_r (acetona:MeOH:H₂O, 65:25:10 v/v/v)= 0,6.

¹H RMN (DMSO-d₆) δ = 5,77-5,64 (m, 1H), 5,58-5,44 (m, 1H), 5,15 (a, 0,5H), 4,95 (dt, ³J(H,H)= 9,5, 5,5 Hz, 0,5H), 3,51 (t, ³J(H,H) = 5,7 Hz, 1H), 2,94 (c, ³J(H,H)= 5,4 Hz, 2H), 2,41- 2,29 (m, 1H), 2,13-2,04 (m, 0,5H), 1,97-1,84 (m, 2,5H), 1,79-1,24 (m, 8H), 1,05-0,94 (m, 0,5H), 0,88-0,71 (m, 1,5H) ppm.

¹³C RMN (DMSO-d₆) δ = 171,6, 155,9, 132,9, 132,7, 131,1, 77,9, 73,0, 53,3, 43,9, 41,4, 40,5, 36,1, 35,7, 35,3, 30,6, 29,5, 28,9, 27,3, 24,2, 22,6, 22,4, 22,2 ppm.

HR MS (ESI neg.) *m/z*: calc. para C₁₅H₂₅N₂O₄ [M-H]⁻: 297,18198, med.: 297,18285.

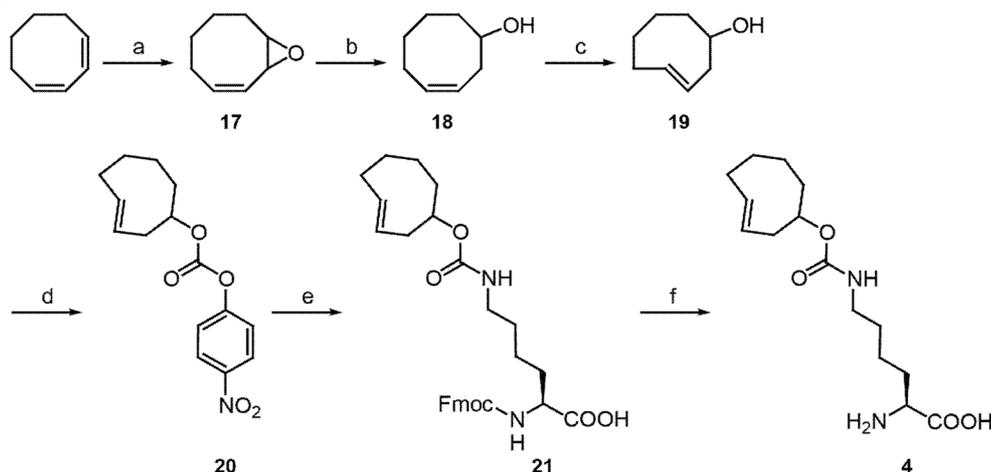
Ejemplo 3a: Preparación de *N*-ε-((*trans*-Ciclooct-2-en-1-ilo)carbonil)-L-lisina axial y ecuatorial

Después de la irradiación UV (véase la etapa d en el esquema S1 anterior), los isómeros axiales y ecuatoriales del compuesto **14** (*trans*-ciclooct-2-en-1-ol) se separaron por cromatografía ultrarrápida sobre sílice (eluyente: EtOAc/ciclohexano en una relación de 1:4; F_r TCO*^a = 0,38, F_r TCO*^e = 0,27), y después se acoplaron por separado a lisina (como se ha descrito para el compuesto **14** en el Ejemplo 3 anterior) para proporcionar el isómero axial **3a** y el isómero ecuatorial **3b** del compuesto **3**.

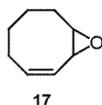
Ejemplo 4: *N*-ε-((*trans*-Ciclooct-3-en-1-ilo)carbonil)-L-lisina

El esquema S2 muestra la síntesis del derivado de lisina *trans*-ciclooct-2'-eno-funcionalizado **4** (también denominado TCO#).

Esquema S2:



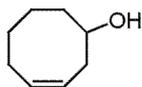
Reactivos y condiciones: a) ácido 3-cloroperbenzoico, CHCl_3 , ta, durante una noche; b) LiAlH_4 , THF, 0°C a ta, durante una noche; c) h- ν ($\lambda = 254$ nm), benzoato de metilo, $\text{C}_6\text{H}_{12}/\text{EtOAc}$, ta, 10 h; d) cloroformiato de 4-nitrofenilo, piridina, CH_2Cl_2 , ta 2 h; e) Fmoc-Lys-OH*HCl, DIEA; DMSO, ta, durante una noche; f) piperidina, CH_2Cl_2 , ta, 30 min.



El compuesto conocido en la bibliografía **17** se sintetizó partiendo de 1,3-ciclooctadieno disponible en el mercado de acuerdo con un protocolo descrito previamente (K. Zhang, M. A. Lackey, J. Cui, G. N. Tew, *Journal of the American Chemical Society* 2011, 133, 4140-4148). En resumen, se añadió gota a gota una solución de ácido 3-cloroperbenzoico (70-77 %, 0,8 equiv., 86,75 g, 327 mmol) en CHCl_3 (1100 ml) durante 2 h a 1,3-ciclooctadieno (1,0 equiv., 50,63 g, 468 mmol) con agitación. La mezcla de reacción se agitó a ta durante 15 h. Después de la filtración para retirar ácido 3-clorobenzoico, el filtrado se lavó sucesivamente con soluciones de bisulfito sódico (10 % en H_2O), NaHCO_3 (10 % en H_2O), y salmuera. A continuación, la capa orgánica se secó sobre Na_2SO_4 y se concentró. El análisis por FC (EtOAc al 5 % en C_6H_{12}) produjo **17** (48,28 g, 389 mmol, 83 %) en forma de un líquido transparente.

F_r (EtOAc al 10 % en C_6H_{12}) = 0,3.

^1H RMN (CDCl_3) $\delta = 5,81-5,74$ (m, 1H), 5,62-5,56 (m, 1H), 3,48-3,44 (m, 1H), 3,15-3,09 (m, 1H), 2,37-2,26 (m, 1H), 2,14-1,97 (m, 2H), 1,83-1,60 (m, 3H), 1,53-1,37 (m, 2H) ppm.



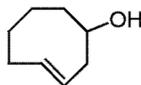
Una solución de LiAlH_4 (0,5 equiv., 81,5 ml de una solución 1,0 M en THF, 81,5 mmol) se añadió gota a gota a 0°C y en una atmósfera de Ar a una solución agitada de **17** (1,0 equiv., 20,23 g, 163 mmol) en THF (1,0 M, 163 ml). La mezcla de reacción se dejó calentar hasta ta y se agitó durante una noche. Se añadió cuidadosamente H_2O (aproximadamente 20 ml) para detener la reacción. La mezcla de reacción se filtró, se secó sobre Na_2SO_4 , y se concentró. El análisis por FC (EtOAc al 20 % en C_6H_{12}) produjo **18** (15,81 g, 125 mmol, 77 %) en forma de un líquido transparente.

F_r (EtOAc al 20 % en C_6H_{12}) = 0,4.

^1H RMN (CDCl_3) δ = 5,76-5,60 (m, 2H), 3,84-3,77 (m, 1H), 2,36 (dd, $^3J(\text{H,H})$ = 7,5, 6,3 Hz, 2H), 2,28-2,18 (m, 1H), 2,14-2,05 (m, 1H), 1,87-1,78 (m, 1H), 1,73-1,63 (m, 1H), 1,61-1,43 (m, 4H), 1,40-1,30 (m, 1H) ppm.

^{13}C RMN (CDCl_3) δ = 132,4, 125,0, 72,2, 35,1, 34,0, 28,3, 25,7, 21,2 ppm.

5



19

El compuesto **19** se preparó partiendo de su precursor *cis* **18** de acuerdo con un procedimiento descrito previamente (a) N. K. Devaraj, R. Upadhyay, J. B. Haun, S. A. Hilderbrand, R. Weissleder, *Angewandte Chemie* 2009, **48**, 7013-7016; b) M. Royzen, G. P. Yap, J. M. Fox, *Journal of the American Chemical Society* 2008, **130**, 3760-3761). La fotoisomerización con eliminación activa del isómero *trans* se realizó en un reactor Rayonet RPR-100 UV (Southern New England Ultraviolet Company, Branford, Connecticut, Estados Unidos) durante 10 h a *ta* (la temperatura en el interior del reactor era de aproximadamente 30 °C). En resumen, el reactor UV se cargó con **19** (1,0 equiv., 6,17 g, 48,9 mmol), benzoato de metilo (1,5 equiv., 9,20 ml, 73,3 mmol) y disolvente (C_6H_{12} al 30 % en Et_2O , 1000 ml). Después del tratamiento y la FC (EtOAc al 20 % en C_6H_{12}), se obtuvo un aceite transparente (4,25 g, 33,7 mmol, 69 %). El análisis por RMN mostró la presencia de dos isómeros en la relación de 1,0:1,2.

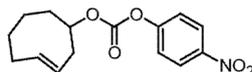
10

15

F_r (EtOAc al 20 % en C_6H_{12}) = 0,4.

^1H RMN (CDCl_3) δ = 5,85-5,76 (m, 0,45H), 5,61-5,44 (m, 1H), 5,32-5,23 (m, 0,55H), 4,44-4,38 (m, 0,45H), 3,74-3,65 (m, 0,55H), 2,87-2,79 (m, 0,55H), 2,48-2,34 (m, 1,45H), 2,21-2,12 (m, 0,45H), 2,10-1,78 (m, 3,56H), 1,56-1,29 (m, 2,9H), 1,24-1,06 (m, 1,55H), 0,81-0,71 (m, 0,55H) ppm.

20



20

A una solución agitada del compuesto **19** (1,0 equiv., 4,25 g, 33,7 mmol) en CH_2Cl_2 (44 ml) se le añadió piridina (2,5 equiv., 6,81 ml, 84,2 mmol) en una atmósfera de Ar. Se añadió una solución de cloroforniato de 4-nitrofenilo (1,1 equiv., 7,47 g, 37,1 mmol) en CH_2Cl_2 (30 ml) a *ta* y la mezcla de reacción resultante se agitó durante 2 h. Se usó una solución ac. sat. de NH_4Cl para detener la reacción. Después de la separación en fases, la capa acuosa se extrajo con CH_2Cl_2 (2 x). Las capas orgánicas combinadas se lavaron con salmuera, se secaron con Na_2SO_4 , y se concentraron. El producto en bruto se purificó por FC (EtOAc al 5 % en C_6H_{12}) para producir el compuesto **20** (8,51 g, 29,2 mol, 87 %) en forma de una cera de color amarillo pálido (T_m = 75,3-78,7 °C).

25

30

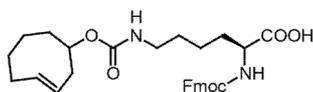
F_r (EtOAc al 5 % en C_6H_{12}) = 0,2.

^1H RMN (CDCl_3) δ = 8,31-8,25 (m, 2H), 7,42-7,35 (m, 2H), 5,80-5,55 (m, 1,5H), 5,35-5,30 (m, 0,5H), 5,28-5,19 (m, 0,5H), 4,72-4,63 (m, 0,5H), 3,04-2,96 (m, 0,5H), 2,74-2,65 (m, 0,5H), 2,53-2,19 (m, 2,5H), 2,11-1,83 (m, 3H), 1,61-1,47 (m, 1H), 1,42-1,12 (m, 2H), 0,89-0,77 (m, 0,5H) ppm.

35

^{13}C RMN (CDCl_3) δ = 155,7, 155,6, 152,0, 151,8, 145,3, 138,5, 137,0, 128,1, 125,3, 125,1, 121,8, 121,8, 85,9, 83,5, 40,5, 39,8, 36,0, 35,6, 35,6, 35,5, 32,7, 30,6, 23,5, 20,4 ppm.

HR MS (FAB pos.) m/z : calc. para $\text{C}_{15}\text{H}_{18}\text{NO}_5$ [$\text{M}+\text{H}$] $^+$: 291,1185, med.: 292,1189.



21

Se disolvió Fmoc-Lys-OH*HCl (1,5 equiv., 5,43 g, 20,0 mmol) en DIEA (3,0 equiv., 6,97 ml, 40,0 mmol) y DMSO anhidro (0,2 M, 67 ml) en una atmósfera de Ar. A continuación, se añadió gota a gota una solución transparente del compuesto **20** (1,0 equiv., 3,88 g, 13,3 mmol) en DMSO anhidro (0,5 M, 27 ml) a *ta* y en una atmósfera de Ar durante un periodo de 2 h. La mezcla de reacción se agitó durante una noche a *ta*. Se añadieron H_2O (80 ml) y EtOAc (170 ml). El pH de la capa acuosa se ajustó a 1-3 con HCl conc. Las fases se separaron y la capa acuosa se extrajo con EtOAc (3 x). Las capas orgánicas combinadas se lavaron con salmuera (2*) y se secaron sobre Na_2SO_4 . Todos los componentes volátiles se evaporaron a presión reducida y el producto en bruto se purificó por FC (MeOH al 10 % en CH_2Cl_2) para producir el compuesto **21** (4,79 g, 9,20 mmol, 71 %) en forma de una espuma de color blanco.

40

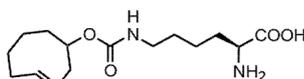
45

F_r (MeOH al 8 % y AcOH al 2 % en CH_2Cl_2) = 0,5.

^1H RMN (CDCl_3) δ = 7,76 (d, $^3J(\text{H,H})$ = 7,4 Hz, 2H), 7,64-7,55 (m, 2H), 7,40 (t, $^3J(\text{H,H})$ = 7,3 Hz, 2H), 7,31 (t, $^3J(\text{H,H})$ = 7,3 Hz, 2H), 5,73-5,44 (m, 2,5H), 5,27-5,18 (m, 0,5H), 4,52-4,33 (m, 3H), 4,27-4,18 (m, 1H), 3,24-3,12 (m, 2H), 2,60-0,73 (m, 14H) ppm.

50

HR MS (ESI neg.) m/z : calc. para $\text{C}_{30}\text{H}_{35}\text{N}_2\text{O}_6$ [$\text{M}-\text{H}$] $^-$: 519,25006, med.: 519,25114; calc. para $\text{C}_{30}\text{H}_{35}\text{N}_2\text{O}_7$ [$\text{M}-\text{OH}$] $^-$: 535,24498, med.: 535,24600.



4

El compuesto **21** (1,0 equiv., 4,16 g, 7,98 mmol) se disolvió en piperidina al 20 % en CH₂Cl₂ (25 ml) y se agitó durante 30 min a ta. Después de la adición de H₂O (30 ml), todos los componentes volátiles se eliminaron a presión reducida y alto vacío. El producto en bruto se purificó a través de FC (acetona:MeOH:H₂O, 85:10:5 v/v/v para eliminar por lavado todas las impurezas seguido de 65:25:10 v/v/v para eluir el producto) sobre gel de sílice para producir el compuesto **4** (2,02 g, 6,78 mmol, 85 %) en forma de un polvo de color blanco. **3** purificado se descompuso con calentamiento (partiendo de 180 °C). El compuesto **4** también se denomina TCO[#] en esta publicación.

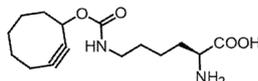
F_r (acetona:MeOH:H₂O, 65:25:10 v/v/v)= 0,6.

¹H RMN (MeOD) δ = 5,76-5,47 (m, 1,5H), 5,31-5,21 (m, 0,5H), 5,13 (a, 0,5H), 4,57-4,49 (m, 0,5H), 3,54-3,47 (m, 1H), 3,15-3,06 (m, 4H), 2,82-2,73 (m, 0,5H), 2,48-2,34 (m, 1,5H), 2,23-2,12 (m, 1H), 2,06-1,75 (m, 4,5H), 1,74-1,67 (m, 0,5H), 1,57-1,37 (m, 4H), 1,27-1,06 (m, 1,5H), 0,89-0,78 (m, 0,5H) ppm.

¹³C RMN (MeOD) δ = 150,7, 149,2, 137,4, 135,7, 132,3, 54,7, 45,5, 44,4, 40,8, 39,8, 39,7, 37,6, 35,0, 32,7, 30,5, 29,2, 24,0, 23,3, 22,5, 22,1 ppm.

HR MS (ESI neg.) *m/z*: calc. para C₁₅H₂₅N₂O₄ [M-H]⁻: 297,18198, med.: 297,18287.

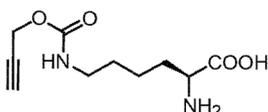
Ejemplos 5: N-ε-((Ciclooct-2-in-1-iloxi)carbonil)-L-lisina



5

Se adquirió el aminoácido a base de lisina no natural **5** en Sirius Fine Chemicals SiChem (Bremen, Alemania). El compuesto **5** también puede sintetizarse como se describe en el documento WO 2012/051885. El compuesto **5** también se denomina SCO.

Ejemplos 6: Propargil-lisina



6

Se adquirió el aminoácido a base de lisina no natural **6** en Sirius Fine Chemicals SiChem (Bremen, Alemania). El compuesto **6** también se denomina PrK.

MS (ESI) *m/z*: calculado para C₁₅H₂₇N₂O₄ [M+H]⁺: 299,19653, medido: 299,19656.

Ejemplos biológicos

Ejemplo A: Expresión de proteína recombinante en *E. coli*

Los plásmidos indicados previamente pEvolv ARNt^{Pyl}/PylIRS^{AF} y pBAD GFP^{Y39TAG} se usaron para expresar GFP^{TAG→UAA}, que codifica genéticamente el aminoácido no natural (UAA) respectivo (a) T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2011, *50*, 3878-3881; b) T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2012, *51*, 4166-4170). En resumen, los plásmidos se cotransformaron en *E. coli* Top10 (Invitrogen) y se cultivaron a 37 °C en presencia de ampicilina y cloranfenicol. Para la expresión a pequeña escala, 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 crecieron en 2 h hasta una DO de 0,2-0,3, momento en el que un compuesto seleccionado de los compuestos **1-5** (Figura 1a) y los compuestos **3a** y **3b** (soluciones madre 80 mM en NaOH 0,1 M), o una cantidad igual de NaOH 0,1 M (para experimentos de control) se añadieron a una concentración final de 1 mM. Los cultivos se dejaron crecer hasta una DO 0,4-0,6, cuando se indujo la expresión con arabinosa al 0,02 %. Los cultivos se recolectaron por centrifugación después de 6-8 h de agitación a 37 °C. Los sedimentos se resuspendieron en una solución salina tamponada con fosfato 4x (4 x PBS, pH 8,0) y las células se lisaron por sonicación. El sobrenadante se incubó con ~50 μl de Ni-NTA (Qiagen, Düsseldorf, Alemania). Las perlas de Ni se lavaron con imidazol 20 mM en PBS 4x, pH 8, y después se eluyeron con tampón que contenía imidazol 1 M.

El compuesto **3**, así como sus isómeros puros axiales y ecuatoriales **3a** y **3b** mostraron tasas de incorporación muy similares, se incorporaron más fácilmente que el compuesto **2** en aproximadamente un factor de dos (Figura 12) y

dieron rendimientos proteicos mejorados (0,4 mg/l).

Ejemplo B: Reactividad cruzada *in vitro* y ortogonalidad de las reacciones SPAAC y SPIEDAC

- 5 GFP^{TAG}→UAA se expresó en *E. coli* como se ha descrito anteriormente en presencia de **1**, **3**, y **5**. Posteriormente, la proteína purificada (Ni-NTA, véase anteriormente) se marcó con Cy3-azida (45 μM, 6 h, 37 °C), Me-Tet-Cy3 o H-Tet-Atto532 (ambos 15 μM, 20 min, 37 °C). Los tres tintes se obtuvieron en Jena Bioscience. Después del etiquetado, la muestras se cargaron directamente en un gel SDS-PAGE y se analizaron para determinar la fluorescencia en un sistema de documentación de gel disponible comercialmente (Alpha Innotech, CA) con excitación UV y ajustes de filtro de detección de etidio-bromo. Posteriormente, el gel se tiñó con Coomassie (**Figura 1d**). No se marcó ninguna otra proteína que fuera sea GFP^{TAG}→UAA, lo que demuestra la bioortogonalidad de las reacciones de etiquetado.

Ejemplo C: Construcciones para la expresión del receptor de insulina (IR) en células de mamífero

- 15 El mutante ámbar IR^{TAG} se generó a través de la mutagénesis de sitio dirigido basada en PCR de un plásmido pEGFPN1_IR, creando una mutación TAG en K676 en el gen IR. Esta posición fue elegida basándose en la estructura de IR disponible (PDB: 2DTG) (N. M. McKern, M. C. Lawrence, V. A. Streltsov, M. Z. Lou, T. E. Adams, G. O. Lovrecz, T. C. Elleman, K. M. Richards, J. D. Bentley, P. A. Pilling, P. A. Hoyne, K. A. Cartledge, T. M. Pham, J. L. Lewis, S. E. Sankovich, V. Stoichevska, E. Da Silva, C. P. Robinson, M. J. Frenkel, L. G. Sparrow, R. T. Fernley, V. C. Epa, C. W. Ward, *Nature* 2006, **443**, 218-221). El pEGFPN1_IR^{TAG} resultante se usó en la **Figura 2** y la **Figura 9 y 10**. La señal fluorescente de GFP fusionada en el extremo C solo estuvo presente después de una supresión de ámbar exitosa y, por lo tanto, se usó como lectura de la expresión de IR^{TAG} exitosa. Para evitar cualquier contaminación por la GFP en experimentos de SRM (**Figura 3**) se generó una construcción bicistrónica pCI-IR^{TAG}-IRES-CFP insertando el casete IR^{K676TAG}-IRES-CFP en un vector de expresión de mamífero pCI vacío (Promega, Madison, Estados Unidos). En el extremo C de IR^{TAG}, se introdujo una etiqueta Flag que fue seguida de un codón de terminación TGA antes de la secuencia IRES. Para la expresión de PyIRS^{AF}/ARNt^{PyI} en células de mamífero, se usó el plásmido pCMV ARNt^{PyI}/PyIRS^{AF} descrito previamente (a) T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2011, **50**, 3878-3881; b) T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2012, **51**, 4166-4170).

Ejemplo D: Cultivo celular y transfecciones con construcciones de IR

- Para experimentos con receptores de insulina, las células HEK293T se mantuvieron en medio de Eagle modificado por Dulbecco - DMEM (alto contenido de glucosa - 4,5 g/l) complementado con FBS al 10 % (Sigma L-glutamina al 1 % (Invitrogen, Palo Alto, Estados Unidos) y Pen-Strep al 1 % (Invitrogen) en una atmósfera de CO₂ al 5 % a 37 °C. Las células se pasaron cada 2-3 días hasta 15-20 pases. Para microscopía, las células se sembraron en recubrimientos de vidrio de borosilicato Lab-Tek#1.0 con cámaras de 4 pocillos (ThermoFisher, Ma, Estados Unidos) durante 15-24 h antes de las transfecciones. Las transfecciones se realizaron con reactivo JetPrime (PiqLab, Erlangen, Alemania) de acuerdo con las recomendaciones del fabricante. En pocas palabras, para las transfecciones dobles con vectores IR y ARNt^{PyI}/PyIRS^{AF} se usaron plásmidos en una relación 1:1 manteniendo la cantidad total de ADN en 1 μg. Los UAA se añadieron después de las transfecciones en dos pulsos (véase la **Figura 2a**). Los UAA se disolvieron en DMSO (solución madre 250 mM de PrK; solución madre 100 mM de TCO*) o ácido fórmico al 20 %-DMSO (solución madre 500 mM de BCN; solución madre 250 mM de SCO). Cada UAA se usó a una concentración final de 250 μM. Después del pulso 2º pulso de UAA, las células se aclararon con medio fresco y se mantuvieron durante una noche en DMEM.

Ejemplo E: Etiquetado de pulso-caza de IR

- A la mañana siguiente a las transfecciones, el medio se cambió a DMEM con alto contenido de glucosa sin suero complementado con FBS al 10 % y Pen-Strep, en el que también se realizaron todas las etapas posteriores de lavado y etiquetado. Las células transfectadas se marcaron entonces de acuerdo con los siguientes protocolos. En el caso de la combinación SPIEDAC-SPAAC, las células se incubaron con una solución 10 μM de sulfo-Cy5-azida (Lumiprobe, Hannover, Alemania) disuelta en DMEM sin suero durante 10 minutos a 37 °C, se lavaron con medio fresco, se incubaron con una solución 5 μM de Cy3-Me-Tet sulfonada (6-metil-tetrazina-Sulfo-Cy3, Jena Bioscience, Jena, Alemania) durante 10 minutos a 37 °C y después se lavaron con medio fresco y se mantuvieron a 37 °C antes de la fijación. Para obtener más detalles sobre el etiquetado SPAAC, véase el **Ejemplo M**.

- En el caso de SPIEDAC con una combinación seSPIEDAC, las células se incubaron primero con la solución 5 μM de Me-Tet-Cy5 sulfonada (6-Metil-Tetrazina-Sulfo-Cy5, Jena Bioscience) disuelta en DMEM sin suero durante 10 min a 37 °C. Después del etiquetado de Me-Tet-Cy5, las células se aclararon con medio fresco. Para inactivar cualquier TCO* sin reaccionar (véase la **Figura 10** para obtener más detalles sobre la elección del inactivador), las células se incubaron posteriormente con Me-Tet-NH₂ no fluorescente (50 μM, disuelto en DMEM sin suero). El clorhidrato de Me-Tet-NH₂ ((4-(6-Metil-1,2,4,5-tetrazin-3-il)fenil)-metanamina) se sintetizó de acuerdo con un procedimiento publicado (M. R. Karver, R. Weissleder, S. A. Hilderbrand, *Angew Chem Int Ed Engl* 2011). Después de esta etapa, las células se aclararon dos veces con DMEM sin suero y luego se marcaron con H-Tet-Atto532 (Jena Bioscience) incubando las células en una solución de 5 μM durante aproximadamente 10 min. Las soluciones madre de Me-Tet-Cy5 y H-Tet-

Atto532 se prepararon en DMSO a una concentración de 500 μM . Después del etiquetado, las células se aclararon de nuevo, el medio se intercambió y las células se mantuvieron en la incubadora durante 1-5 h antes de la etapa de fijación. La fijación se realizó con *para*-formaldehído al 2 % (PFA) en PBS a temperatura ambiente durante 10 min, antes de lo cual, las células se aclararon con PBS.

5

Ejemplo F: Microscopía de IR

Después de la fijación, las células se llevaron al microscopio, el mismo día o un máximo de 1-2 días después. Esto fue especialmente importante para las imágenes SRM. Todas las imágenes confocales se realizaron en un microscopio comercial LEICA TCS SP8 equipado con un objetivo HCX PL APO 100x/1,40 OIL (Mannheim, Alemania). Para obtener las imágenes que comparan diferentes niveles de etiquetado, se utilizaron los mismos ajustes de adquisición para todos los experimentos. Se obtuvieron imágenes de las células en PBS y se adquirieron imágenes de un solo plano con un tamaño de píxel de 114 nm. Toda la SRM se realizó en un microscopio Leica GSD comercial, equipado con un objetivo Leica HCX PL APO 100x/NA 1,47 Oil CORR TIRF PIFOC y ajustes de filtro de Cy3 y Cy5. Para la SRM, se usó un tampón de imágenes que contenía un sistema de captación de oxígeno (glucosa oxidasa con catalasa -GLOX) en presencia de tioles (2-aminoetanol 10 mM, MEA) producido de acuerdo con un protocolo publicado (G. T. Dempsey, J. C. Vaughan, K. H. Chen, M. Bates, X. Zhuang, *Nature methods* 2011, 8, 1027-1036). El tampón se preparó nuevo y se intercambia después de 1-3 h de imágenes o antes en caso de que no haya suficientes parpadeos. Todas las imágenes se adquirieron en modo epifluorescente, primero en el canal Cy5 y luego en Atto532. Para el canal Cy5, se usaron un láser de 642 nm, y para el canal Atto532, un láser de 532 nm, respectivamente, para bombear y obtener imágenes del parpadeo con tiempos de exposición de 10 ms. Para facilitar el retorno de los fluoróforos al estado fundamental, se usó un láser de 405 nm a intensidades muy bajas.

10

15

20

Ejemplo G: Construcciones para la expresión de VLP en células de mamífero

Se obtuvieron M1 (A/Hong Kong/68) y HA (A/Aichi/2/68) en el vector de expresión eucariota pCAGGS a partir de Mikhail Matrosovich (Marburg). La secuencia de HA se clonó en el vector de expresión pCI y se basó en la estructura disponible de HA (PDB: 1E08) (X. Morelli, M. Czjzek, C. E. Hatchikian, O. Bornet, J. C. Fontecilla-Camps, N. P. Palma, J. J. Moura, F. Guerlesquin, *The Journal of biological chemistry* 2000, 275, 23204-23210) e información sobre la conservación de secuencias entre diferentes variantes de la gripe, se seleccionaron diferentes posiciones de aminoácidos y se sustituyeron por el codón TAG ámbar con mutagénesis dirigida de sitio dirigido estándar. En este estudio, se usó la sustitución de aminoácidos HA342 ya que mostró una buena incorporación de UAA y eficiencia de etiquetado.

25

30

Ejemplo H: Cultivo celular y preparación de muestras para VLP

HEK293T se mantuvieron en DMEM (glucosa baja, piruvato, sin glutamina, sin rojo fenol, 11880-028 GIBCO/Invitrogen) complementado con suero fetal ternera al 10 % (FCS) y L-glutamina a 37 °C bajo CO₂ al 5 %.

40

45

50

55

Para fluorescencia y SRM, las células se sembraron en cubreobjetos de alta precisión bien limpiados ($\varnothing = 18$ mm, espesor = $0,17 \pm 0,005$ mm, CarlRoth) en placas de cultivo de tejidos de 6 pocillos. Las células se cultivaron hasta un 70-80 % de confluencia antes de transfectarse con los plásmidos apropiados en las relaciones 1:1 con reactivo de transfección JetPrime (PEQLAB) o 239Expresso (Excellgen) de acuerdo con las instrucciones del fabricante. 1 h después de la transfección, se añadieron 250 μM de SCO, PrK o TCO* al medio de crecimiento. Si la producción de partículas de virus filamentosos se indujo mediante la cotransfección de HA y M1, se añadió neuraminidasa bacteriana exógena (durante ambos pulsos UAA) además de los UAA a una concentración de 100 mU/ml (*Clostridium perfringens*; Sigma-Aldrich, Frankfurt, Alemania) para evitar la unión inmediata de los filamentos generados a las células productoras mediante la unión de HA al ácido siálico. Después de 10 h, el medio que contenía UAA se cambió a medio de crecimiento con el segundo UAA y se incubó durante 18 h. Después de esta detección, las células se incubaron durante 1 h en un medio de crecimiento normal antes de marcar el UAA incorporado con tintes modificados con tetrazina a 37 °C durante 20-30 min. Primero, TCO* se marcó con Me-Tet-Cy5 (Jena Bioscience, concentración final de 5 μM), el TCO* sin reaccionar después se inactivó con Me-Tet-NH₂ 50 μM durante 5 min y SCO se marcó con H-Tet-Atto532 (Jena Bioscience, concentración final de 5 μM). Antes de montar los portaobjetos en portaobjetos con cavidad (1,2-1,5 mm, Menzel-Glaeser/Thermo-Scientific), las células se lavaron una vez con medio de crecimiento normal durante 30 minutos y posteriormente se fijaron con PFA al 2 % en PBS durante 10 minutos. Para la SRM, se utilizaron las mismas configuraciones de adquisición que se han descrito anteriormente para IR. Para la microscopía confocal, la muestra se montó con ProLong Gold (Invitrogen) en un portaobjetos estándar y se curó durante 24 horas a temperatura ambiente antes de obtener imágenes con un microscopio Leica TCS SP8.

Ejemplo I: Análisis de imágenes

Las imágenes confocales se procesaron usando Fiji ImageJ (J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nature methods* 2012, 9, 676-682). Para las imágenes IR, se aplicó un filtro de mediana con un radio de dos píxeles. ImageJ también se usó para ajustar el contraste y el brillo (solo se aplicaron cambios lineales) para mostrar las imágenes de IR y VLP (Figuras 2, 9, 10). Los archivos tiff individuales de color se

65

fusionaron en una sola imagen de doble color alineando manualmente los dos canales.

El procesamiento por SRM se realizó en IgorPro (WaveMetrics, Portland OR, Estados Unidos). Se usó ImageJ para ajustes de contraste-brillo lineales para la visualización. Las moléculas de las películas de SRM se localizaron en cada fotograma utilizando el "Paquete localizador" (Dr. Peter Dedecker, Katholieke University, Louvain, Belgium) para IgorPro. Las moléculas individuales se equiparon con un gaussiano para determinar la posición (x,y) de cada evento de parpadeo de una sola molécula. Las posiciones localizadas se analizaron entonces mediante la correlación del anillo de Fourier para estimar la resolución del conjunto de datos, que después se usó para determinar la desviación puntual en las imágenes SRM reconstruidas (a) N. Banterle, K. H. Bui, E. A. Lemke, M. Beck, *J Struct Biol* 2013; b) R. P. Nieuwenhuizen, K. A. Lidke, M. Bates, D. L. Puig, D. Grunwald, S. Stallinga, B. Rieger, *Nature methods* 2013, 10, 557-562). Las imágenes finales se mostraron con un tamaño de pixel de 8 x 8 nm².

Ejemplo J: Determinación de la cinética de etiquetado a través de un ensayo FRET

Para la observación de las cinéticas de etiquetado, GFP^{TAG→UAA} se expresó con 1-5 (para estructuras de UAA véase la **Figura 1a**). El lisado de *E. coli* se ajustó a una concentración final de GFP de 200 nM basándose en los espectros de absorbancia. H-Tet-Atto532 **7** o Me-Tet-Cy3 **8** (véase la **Figura 4**), respectivamente, se añadieron a una concentración final de 2 μM y los espectros de fluorescencia (excitación a λ = 450 nm, emisión λ = 470-650 nm) se registraron en puntos de tiempo diferentes. Los experimentos se realizaron en un volumen total de 2 ml y agitación continua. Se usaron soluciones madre 1 mM de los tintes en DMSO. El etiquetado exitoso de GFP^{TAG→UAA} se controló mediante la transferencia de energía por resonancia de Forster (FRET) desde el cromóforo de GFP (que actúa como Donante, D) al fluoróforo sintético (Aceptor, A) cuando se une covalentemente. En los espectros individuales, esto es visible por una disminución de fluorescencia de GFP (aproximadamente λ = 503-506 nm) y un aumento simultáneo de Atto532- (λ = 551 nm) o Cy3-fluorescencia (λ = 563 nm), respectivamente, en el tiempo (mostrado a modo de ejemplo para GFP^{TAG→3} en el panel izquierdo de la **Figura 5**). El panel derecho de la **Figura 5** muestra la FRET representada en el tiempo (en minutos) para los cinco UAA diferentes y las dos sondas de tetrazina diferentes. Se ha de tener en cuenta que el punto de tiempo "0 min" se define como el punto de tiempo de aproximadamente 15 s después de pipetear el tinte a la solución de proteína.

La cinética de reacción resultante se ajustó con un modelo monoexponencial sencillo de acuerdo con

$$GFP^{UAA \rightarrow}(t) = A_0(1 - \exp(-k_Bt)),$$

donde A₀ corresponde a la amplitud del ajuste y es proporcional a la concentración de GFP inicial y B corresponde a la concentración de tinte dentro de la reacción. La constante de velocidad k de la reacción se obtiene del ajuste con la suposición de B constante durante la reacción (que es válida debido al gran exceso de tinte). Las constantes de velocidad aproximada en estas condiciones experimentales medidas a 20 °C se indican en la **Tabla 1** a continuación y se obtuvieron ajustando las trazas de FRET de tres experimentos independientes para cada UAA (diferentes expresiones de proteínas, diferentes días) y el promedio posterior.

Tabla 1: Cinética de reacción en condiciones experimentales probadas

	Me-Tet-Cy3 [M ⁻¹ ·s ⁻¹]	H-Tet-Atto532 [M ⁻¹ ·s ⁻¹]
GFP ^{TAG→1}	480 ± 190	16000 ± 2000
GFP ^{TAG→2}	2240 ± 340	21000 ± 4000
GFP ^{TAG→3}	1240 ± 250	13000 ± 2000
GFP ^{TAG→4}	3880 ± 470	10000 ± 1200
GFP ^{TAG→5}	no determinado	670 ± 180

Se observa que los derivados de Me-Tet-Cy3 se comparan con los derivados de H-Tet-Atto532 (debido a la disponibilidad comercial). Sin embargo, se puede esperar que las diferencias de tinte no contribuyan notablemente a las diferentes reactividades observadas.

Todos los UAA muestran una reactividad reducida con Me-Tet que con H-Tet. Sin embargo, la caída en la reactividad es más drástica para los alquinos que para los alquenos, lo que lleva básicamente a una reactividad no detectable de SCO en las condiciones experimentales probadas.

También se observa que los sustituyentes de las tetrazinas utilizadas en este trabajo sugieren que una fuerte demanda inversa de electrones no es una fuerza impulsora importante para la velocidad de las reacciones SPIEDAC descritas.

Como se muestra en la **Figura 5**, para GFP^{TAG→5} debido a la ausencia de reactividad, no se pudo observar una señal de FRET estable sobre el fondo en estas condiciones de medición y, por lo tanto, no se pudo intentar una aproximación de la cinética de la reacción.

Ejemplo J1: Reactividad de GFP^{TAG→3}, GFP^{TAG→3a} y GFP^{TAG→3b} con derivados de tetrazina marcados con Cy5

GFP^{TAG→UAA} se expresó en *E. coli* como se ha descrito anteriormente en presencia de **1**, **3**, **3a**, **3b** y **5**. Posteriormente, la proteína purificada (Ni-NTA, véase anteriormente) se marcó con H-Tet-Cy5 ("Tetrazina-C5" de Jena Bioscience) o Me-Tet-Cy5 ("6-Metil-Tetrazina-Sulfo-Cy5" de Jena Bioscience) como se describe a continuación. Los experimentos cinéticos se realizaron usando espectroscopia de flujo detenido (SFM-3000, Bio-Logic). La cinética de la reacción de etiquetado se controló detectando el aumento de la fluorescencia de Cy5 cuando se une covalentemente al UAA de la GFP respectiva debido a la transferencia de energía por resonancia de Forster (FRET) del cromóforo de GFP (que sirve como donante, D) al fluoróforo Cy5 (Aceptador, A). En los espectros, esto es visible por un aumento de la fluorescencia Cy5 en el tiempo (**Figura 13a**).

Las curvas de eficiencia de FRET no se analizaron debido a los efectos fotofísicos que se producen en la señal de GFP en presencia de un exceso de Cy5. Las mediciones se realizaron en condiciones de pseudo-primer orden en PBS pH 7,4 usando proteína GFP^{TAG→x} purificada 100 nM y un exceso (10-40 μM) de tinte Cy5 (H-Tet-Cy5 o Me-Tet-Cy5). El análisis de los datos se realizó ajustando los datos de fluorescencia Cy5 a la función mono-exponencial inversa $F(t) = A \cdot (1 - e^{-k_{obs}t}) + c$. El criterio de información de Akaike se usó para determinar el punto de tiempo de la curva hasta el cual una función mono-exponencial se ajustó mejor que una función bi-exponencial. Las diferentes constantes de velocidad de reacción observadas k_{obs} determinadas en función de los ajustes se representaron en función de la concentración del tinte Cy5 y las constantes de reacción K se obtuvieron haciendo un ajuste lineal forzándolo a 0 (**Figura 13b**, **Tabla 2**). Los experimentos se realizaron por triplicado.

Tabla 2: Constantes de reacción K para el etiquetado de GFP^{TAG→UAA} con tintes de tetrazina-Cy5

	Reacción con H-Tet-Cy5	Reacción con Me-Tet-Cy5
	K [M ⁻¹ ·s ⁻¹]	K [M ⁻¹ ·s ⁻¹]
GFP ^{TAG→1}	80400 ± 1100	390 ± 2
GFP ^{TAG→3}	30500 ± 600	700 ± 10
GFP ^{TAG→3a}	36500 ± 400	680 ± 10
GFP ^{TAG→3b}	12100 ± 200	no determinado*
GFP ^{TAG→5}	1180 ± 80	no determinado*
* no determinado porque la velocidad de reacción fue demasiado baja para obtener valores fiables		

Ejemplo K: Determinación de la estabilidad a largo plazo del enlace covalente en proteínas marcadas

Para probar la estabilidad de los enlaces covalentes formados entre TCO* y el tinte-Tet en una reacción SPIEDAC, se realizaron experimentos de FRET a largo plazo. Básicamente, el mismo ensayo FRET que se ha descrito anteriormente se usó para probar si los fluoróforos permanecen unidos covalentemente a la proteína durante un periodo de 36 h. GFP^{TAG→3} 200 nM se hicieron reaccionar por separado con Me-Tet-Cy3 2 μM a temperatura ambiente (ta) y la relación entre D y A fluoróforo se siguió a lo largo del tiempo. Ningún dato apoya una pérdida potencial del fluoróforo de molécula pequeña a lo largo del tiempo (**Figura 6**).

En otro experimento, GFP^{TAG→UAA} purificado con Ni-NTA (**1-4**) se marcó con Me-Tet-Cy3 (2 h, 37 °C). El exceso de Me-Tet-Cy3 se eliminó por filtración (~2 h). GFP^{TAG→UAA+Cy3} marcado con 200 nM (con **1-4**) se siguió seguidos a lo largo del tiempo por FRET como se ha descrito anteriormente. Se ha de tener en cuenta que el primer punto de datos se tomó aproximadamente 4 h después de añadir la tetrazina a las proteínas. No se observaron cambios significativos entre la relación de FRET medida después de 4 h y después de 76 h (**Figura 7**).

Ejemplo L: Determinación de la estabilidad de los diferentes isómeros de trans-ciclooctenilo a través de ¹H RMN

Estudios anteriores mostraron la degradación de compuestos con una alta tensión en el anillo, tal como **1** y **4**, en presencia de tioles (a) B. R. Varga, M. Kallay, K. Hegyi, S. Beni, P. Kele, *Chemistry* 2012, 18, 822-828; b) R. van Geel, G. J. Pruijn, F. L. van Delft, W. C. Boelens, *Bioconjugate chemistry* 2012, 23, 392-398). Por ejemplo, para los trans-ciclooctenos, se ha observado la isomerización de la forma trans en su forma cis correspondiente pero en órdenes de magnitud menos reactivos cuando se trata con L-cisteína y calor (J. Yang, J. Seckute, C. M. Cole, N. K. Devaraj, *Angew Chem Int Ed Engl* 2012, 51, 7476-7479).

El análisis de ¹H RMN mostró que los UAA **1-5** disueltos en NaOD 0,1 M en D₂O o DMSO-d₆ eran estables a temperatura ambiente (ta), 37 °C, así como a 60 °C durante un periodo de 72 h (datos no mostrados). Por lo tanto, NaOH 0,1 M (para células de *E. coli*) y DMSO (para células de mamífero) son disolventes adecuados para la preparación de soluciones madre para experimentos biológicos. Sin embargo, las soluciones madre se almacenaron a -20 °C y se descongelaron 30 min antes de su uso.

El análisis por ¹H RMN mostró que los UAA **2-4** disueltos en 1x PBS deuterado (dPBS) en D₂O (pH 7,4) mezclado con

1,4-dioxano-d8 (v/v 1:1) eran estables a temperatura ambiente (ta), 37 °C, así como a 60 °C durante un periodo de 72 h (**Figura 8a**). Teniendo en cuenta las propiedades de los compuestos purificados, estas condiciones se asemejan a las condiciones fisiológicas más cercanas posibles, ya que prevalecen en experimentos biológicos (células vivas, soluciones acuosas tamponadas de proteínas) ((a) B.R. Varga, M. Kallay, K. Hegyi, S. Beni, P. Kele, *Chemistry* 2012, 18, 822-828; b) R. van Geel, G. J. Pruijn, F. L. van Delft, W. C. Boelens, *Bioconjugate chemistry* 2012, 23, 392-398).

Para probar la estabilidad frente a tiores, los UAA **2-4** se combinaron por separado con dos equivalentes de clorhidrato de cisteamina en dPBS/dioxano-d8 (v/v 1:1), los tubos de RMN se sellaron herméticamente y se incubaron a ta, 37 °C, o 60 °C. Los espectros de ¹H RMN se midieron después de 24 h de incubación y en puntos temporales posteriores (en el caso de **3**). Los datos de ¹H RMN mostraron que todos los UAA se degradan con el tiempo en presencia de cisteamina de manera dependiente de la temperatura (**Figura 8b**) - pero con diferente velocidad. Para los tres UAA que contienen *trans*-cicloocteno, se observó la conversión de la forma *trans*-ciclooctenilo en su forma *cis*-ciclooctenilo correspondiente (**Figura 8b**). Mientras que **2** y **4** se isomerizan igualmente rápido (<5 % del isómero *trans* que queda después de 24 horas a 60 °C), se encontró que **3** era significativamente más estable (quedaba -80 % del isómero *trans* después de 24 h a 60 °C). Estos números se obtuvieron comparando las integrales de los protones de doble enlace y los protones -CHO- (si es posible/aplicable) antes y después del tratamiento térmico, como se hizo por otros para compuestos similares ((a) B.R. Varga, M. Kallay, K. Hegyi, S. Beni, P. Kele, *Chemistry* 2012, 18, 822- 828; b) R. van Geel, G. J. Pruijn, F. L. van Delft, W. C. Boelens, *Bioconjugate chemistry* 2012, 23, 392-398; J. Yang, J. Seckute, C. M. Cole, N. K. Devaraj, *Angew Chem Int Ed Engl* 2012, 51, 7476-7479). Las observaciones a diferentes temperaturas coincidieron con la expectativa de que cuanto mayor sea la temperatura, más rápida/fácil será la isomerización. La isomerización de la forma *cis* continuó en muestras mantenidas a ta o 37 °C después de 24 h (datos no mostrados).

Para mostrar que también **3** (TCO*) se puede convertir completamente a su forma *cis* en presencia de tiores a 60 °C, se continuó la incubación a 60 °C y se midieron los espectros de ¹H RMN más allá de los puntos de tiempo indicados en la **Figura 8c**. Incluso después de diez días (240 h), se dejó más cantidad de **3** (TCO*) en su forma *trans* que para **2** (TCO) y **4** (TCO#) después de 24 h de incubación a 60 °C con dos equivalentes de cisteamina.

Además, se encontró que el isómero axial de TCO* **3a** era más estable, es decir, menos propenso a la desintegración (principalmente por conversión en la forma *cis*) que el isómero ecuatorial de TCO* **3b**. Se determinó que semivida de **3a** era cercana a dos días a 60 °C y 34 d a 37 °C, mientras que **3b** se desintegró en estas condiciones en unas pocas horas (**Figuras 14a y 14b**).

Ejemplo M: Etiquetado SPAAC del receptor de insulina con/sin bloqueador de endocitosis

Las células HEK293T se transfectaron con los plásmidos pEGFPN1_IR^{K676TAG} y pCMV ARNt^{PyI}/PyIRS^{AF} como se ha descrito anteriormente. Después de las transfecciones, las células se incubaron durante una noche con BCN. A la mañana siguiente, las células se aclararon con DMEM libre de suero. 2 h más tarde se realizó el etiquetado. Como se muestra en la **Figura 2b**, el etiquetado con azida con 10 min de incubación y 10 μM del cinte no dio ningún resultado obvio en el etiquetado. Para optimizar la reacción de etiquetado con azida, se aumentó la concentración de la Cy5-azida a 50 μM y se incubó durante 2 horas a 37 °C. Después del etiquetado, las células se mantuvieron a 37 °C antes de que se fijaran y se llevaran al microscopio confocal (véase anteriormente para obtener más detalles). La **Figura 9a** muestra el etiquetado de membrana específico en el canal Cy5 (panel derecho) que se co-localiza con el canal GFP (panel izquierdo) del IR^{TAG}. Además, en el canal Cy5 se observó un fondo muy alto y un tinte inespecífico adherido en el interior de las células, probablemente debido a la prolongada etapa de incubación del tinte. Solo cuando se aplicó el bloqueador de endocitosis (hidrato de Dynasore, Sigma-Aldrich, Frankfurt, Alemania) (solución 80 μM en DMEM sin suero, 20-30 min antes del etiquetado y durante la incubación con Cy5-azida) se observaron acumulaciones de tinte menos inespecíficas (**Figura 9b**).

En un experimento adicional, las células transfectadas se incubaron con 250 μM de **3**, **3a** o **3b** durante 8 h. Al día siguiente, las células se marcaron con H-Tet-Cy5 5 μM o Me-Tet-Cy5 durante 10 minutos a 37 °C en DMEM. Después de cambiar el medio por DMEM nuevo, las células se incubaron a 37 °C durante 1-2 h más. Posteriormente, las células se fijaron y se analizaron a través de microscopía confocal (véase anteriormente para más detalles). Como se muestra en la **Figura 15**, la incubación con H-Tet-Cy5 5 μM produjo un etiquetado fiable de las tres proteínas de fusión, GFP-IR^{TAG->3}, GFP-IR^{TAG->3a} y GFP-IR^{TAG->3b}. Cuando las células se incubaron con el Me-Tet-Cy5 de reacción más lenta, no se detectó tinción para GFP-IR^{TAG->3b}, mientras que GFP-IR^{TAG->3} y GFP-IR^{TAG->3a} mostraron una tinción detectable que fue algo más débil que con el etiquetado de H-Tet-Cy5.

Ejemplo N: Controles de etiquetado de doble color de IR

Las células HEK293T se transfectaron con los plásmidos pEGFPN1_IR^{TAG} y pCMV ARNt^{PyI}/PyIRS^{AF} como se ha descrito anteriormente. Después de las transfecciones, las células se pulsaron en primer lugar con UAA1 y después se detectaron con UAA2 (véase la **Figura 2**).

Dado que TCO* puede reaccionar tanto con Me-Tet (aplicado durante la 1ª etapa de etiquetado) como con H-Tet (aplicado durante la 2ª etapa de etiquetado), fue necesario excluir la posibilidad de que si no todo el TCO* reacciona en la primera etapa, podría marcarse por H-Tet en la segunda etapa. El consumo total de TCO* en la primera etapa

de etiquetado es un requisito para un diseño de etiquetado doble ortogonal, ya que de lo contrario produciría un resultado ambiguo. Tal fenómeno podría producirse, por ejemplo, si las células expresan una alta concentración de TCO*, de manera que no se consuma completamente en la primera reacción de etiquetado con Me-Tet. Para probar esto, se realizaron los siguientes experimentos de control:

5 Después de las transfecciones, las células se pulsaron en primer lugar con UAA1(TCO*) y después se detectaron con UAA2(PrK), véase la **Figura 2**. La propargilsilina (PrK) es un alquino alifático y no participa en las reacciones SPAAC o SPIEDAC y funciona aquí como un UAA inerte que se incorpora por el mismo ARNt/RS^{AF} con rendimientos similares a TCO*, BCN y SCO (T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2011, 50, 3878-3881; T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, *Angew Chem Int Ed Engl* 2012, 51, 4166-4170; D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters, J. W. Chin, *Journal of the American Chemical Society* 2009, 131, 8720-8721; A. Borrmann, S. Milles, T. Plass, J. Dommerholt, J. M. Verkade, M. Wiessler, C. Schultz, J. C. van Hest, F. L. van Delft, E. A. Lemke, *Chembiochem* 2012, 13, 2094-2099). Después, el Me-Tet-Cy5 (5 µM en DMEM libre de suero, 10 min a 37 °C) se aplicó en primer lugar. En una segunda etapa de etiquetado, se aplicó H-Tet-Atto532 (compárese con la **Figura 2**). Como cabía esperar, se pudo ver una señal muy alta de Me-Tet-Cy5 (**Figura 10a**). Sin embargo, en el canal Atto532 también se puede observar alguna fluorescencia. Dado que solo el TCO* estaba presente como un UAA reactivo, indica que la mayor parte (>80 %), pero no todo el TCO*, se consumió en la primera etapa de etiquetado. Este fenómeno depende del nivel de expresión celular, la concentración de tinte, el tiempo de etiquetado y la temperatura. Para establecer un protocolo robusto que funcione en una amplia gama de parámetros, se encontró práctico introducir una etapa de inactivación. Para este propósito, se usó una alta concentración de Me-Tet (NH₂-Me-Tet) pequeño, que se compara con un tinte relativamente económico de obtener, que puede perfundirse rápidamente y eliminarse por lavado de las células. Se determinó que un pulso de 2 min (a TA) de 50 µM, disuelto en DMEM libre de suero, reaccionó con TCO* no consumido en todos los experimentos ensayados (**Figura 10a**). Por lo tanto, se continuó usando el inactivador en todos los experimentos de etiquetado SPIEDAC-seSPIEDAC posteriores.

También se realizó el etiquetado de doble color después de pulsar las células con UAA1 (TCO*) y detectarlas con UAA2 (BCN). El etiquetado de TCO* dirigido a Me-Tet-Cy5 (5 µM en DMEM libre de suero, 10 min a 37 °C) se siguió de inactivación como se ha descrito anteriormente. En la segunda etapa de etiquetado, se aplicó BCN dirigido a H-Tet-Atto532. Como se muestra en la **Figura 10b**, el TCO* se marcó con éxito, pero debido a la reactividad entre BCN y las altas concentraciones de Me-Tet del inactivador, no se observó señal en el canal Atto532. Esto demuestra que en las condiciones ensayadas, la reactividad restante de BCN frente a Me-Tet (en línea con la cinética de las **Figuras 5, 6 y 7**, el ensayo de gel en la **Figura 2**) no permite un enfoque de etiquetado de doble color ortogonal a la reacción de TCO* con Me-Tet.

También se muestra en la **Figura 10c**, las repeticiones del mismo experimento de control (como se ha descrito anteriormente para IR) con TCO* y PRK para partículas pseudovíricas (VLP), usando las mismas condiciones de etiquetado e inactivación (compárese con la **Figura 2c**).

Para probar la posibilidad de un etiquetado no específico en la **Figura 10d**, se muestra aquí el campo de visión más grande (compárese con la **Figura 2b**). Resulta evidente a partir de esta imagen que solamente las células positivas a pEGFPN1_IR^{K676TAG} están marcadas con H-Tet-Atto532/Me-Tet-Cy5.

Ejemplo O: Validación por espectrometría de masas

Para la validación espectrométrica de masas correspondiente a la proteína mostrada (compárese con la **Figura 1**), GFP^{TAG→3} y GFP^{TAG→4} expresados en *E. coli* que albergan el plásmido ARNt^{Pyl}/PylRS^{AF} se suprimieron y posteriormente se digirieron con tripsina (Tryp) siguiendo protocolos estándar para análisis de masas peptídicas de alta resolución. Los péptidos se analizaron utilizando un espectrómetro de masas Orbitrap (ThermoFisher, Estados Unidos). Los datos se analizaron mediante el algoritmo Mascot y los resultados que verifican la incorporación exitosa de **3 y 4** se resumen en la **Table 3**. Se analizó GFP^{TAG→5} como un control positivo.

Tabla 3: Validación espectrométrica de masas de la formación de GFP^{TAG→UAA}

Proteína	Masa monoisotópica (calc.) [Da]	Masa coincidente, hallada [Da]	Secuencia peptídica (X= sitio TAG ámbar)
GFP ^{TAG→3}	2075,07864	2075,07944	FSVSGEGEGDATXGKLTLLK
GFP ^{TAG→4}	2075,07864	2075,07724	FSVSGEGEGDATXGKLTLLK
GFP ^{TAG→5}	2073,06300	2073,06368	FSVSGEGEGDATXGKLTLLK

55 Las expresiones de ensayo con una construcción indicadora de GFP^{TAG}, que solo da una expresión de longitud completa y, por lo tanto, fluorescencia si se suprime la mutación ámbar en la posición Y39, muestran que TCO* y TCO[#] se aceptan por el par ARNt^{Pyl}/PylRS^{AF} aproximadamente tres veces mejor que el ciclooct-4-eno TCO original (**Figura 1b**, **Tabla 3** para datos de espectrometría de masas), produciendo aproximadamente 10 mg de un cultivo de expresión de 1 l de *E. coli*. Las **Figuras 5, 6 y 7** muestran que los tres TCO mantienen una reactividad similar en las

reacciones SPIEDAC. Sin embargo, se sabe que el *trans*-cicloocteno tiene problemas de estabilidad química y tiende a isomerizarse en la forma *cis* no reactiva, especialmente en presencia de tioles (J. Yang, J. Seckute, C. M. Cole, N. K. Devaraj, *Angew Chem Int Ed Engl* **2012**, *51*, 7476-7479). Dado que los tioles son abundantes en el citosol de las células de mamífero, esto puede conducir a la interferencia con la bioestabilidad de UAA durante las expresiones a largo plazo. Las mediciones de RMN mostradas en la **Figura 1c** y la **Figura 8** mostraron que el TCO* tiene una estabilidad al menos 10 veces mayor en presencia de tioles, lo que indica una protección eficiente del doble enlace *trans* hacia los tioles por la proximidad del enlace de carbamato.

El par mutante ARNt^{PyI}/PyIRS^{AF} permite codificar TCO* y BCN, que pueden experimentar reacciones SPIEDAC y SPAAC, respectivamente (**Figura 1d**). Para explorar el potencial de este par de UAA para el etiquetado de doble color de células vivas se usó para el etiquetado de pulso-caza del receptor de insulina (IR). La función y el reciclado de receptores de IR son temas de gran relevancia contemporánea debido a su papel central en la diabetes así como los nuevos roles reguladores genéticos (K. Siddle, *Journal of molecular endocrinology* **2011**, *47*, R1-10; R. Sarfstein, H. Werner, *Endocrinology* **2013**, *154*, 1672-1679). Se seleccionó una posición ubicada en el lado extracelular de la proteína (K676) para una mutación ámbar y se expresó el IR^{TAG} en presencia de un plásmido que codifica el ARNt^{PyI}/PyIRS^{AF} en células HEK293T.

Después, se realizó el experimento de pulso-caza como se describe en la **Figura 2a**, donde el medio de crecimiento se pulsó por primera vez durante 4 h con el primer UAA (TCO*), seguido de una detección de 4 h con el 2º UAA (BCN). IR^{TAG} se marcó en primer lugar con azida-Cy5 y más tarde con Me-Tet-Cy3, cada uno se incubó durante 10 min en células vivas. Como se puede observar en la **Figura 2b** (1er panel), la imagen confocal permitió visualizar la tinción de membrana de IR^{TAG}→TCO* con Me-Tet-Cy3 de la reacción SPIEDAC. El etiquetado corto de IR^{TAG}→BCN con azida-Cy5 no dio resultados. Como se detallada en la **Figura 9**, esto podría atribuirse a la velocidad de la reacción SPAAC, que es de tres a cuatro órdenes de magnitudes más lenta que la reacción SPIEDAC.

El derivado de ciclooctinil-lisina (SCO) (**Figura 1a**) es aceptado por el mismo par ARNt^{PyI}/PyIRS^{AF} con rendimientos similares a TCO*. Mientras que TCO* reacciona con H-Tet y Me-Tet con velocidades de reacción de >1000 1/Ms en ensayos cinéticos *in vitro* y experimentos de etiquetado, SCO no muestra ninguna reactividad sustancial con Me-Tet en las condiciones ensayadas (véase la **Figura 5**, **6** y **7** para cinética de reacción, **Figura 1d**). Sin embargo, SCO todavía reacciona en aproximadamente dos órdenes de magnitud más rápido en una reacción SPIEDAC con H-Tet que BCN en una reacción SPAAC con azida. El experimento de pulso-caza se repitió con TCO* y SCO (**Figura 2a**), seguido de etiquetado con Me-Tet-Cy5, y después con H-Tet-Atto532 durante 10 min. Como se muestra en la **Figura 2b**, esta combinación dio como resultado un etiquetado brillante del IR en la membrana plasmática para ambos canales. SCO reacciona selectivamente con H-Tet pero no con Me-Tet en la escala de tiempo de estos experimentos, lo que da como resultado una reacción que es ortogonal a SPIEDAC entre TCO* y Me-Tet. Este tipo de subreacción se conoce en el presente documento como "SPIEDAC con selectividad mejorada" (seSPIEDAC). Se observa que como TCO* es altamente reactivo con ambos, Me-Tet y H-Tet, se deben elegir las condiciones experimentales para asegurar que todo el TCO* se consuma antes de continuar con la segunda etapa de etiquetado (véase la **Figura 10** para obtener más detalles).

Además, se muestra en la **Figura 1d** y en la **Figura 10** que aumentar adicionalmente la velocidad de seSPIEDAC utilizando el BCN con alta tensión anular en lugar de SCO no es posible debido a la reactividad del BCN hacia Me-Tet (**Figuras 5**, **6** y **7**).

Dado que Cy5 y Atto532 se usan comúnmente para la SRM basada en la localización, se realizaron mediciones de SRM de doble color (M. Heilemann, S. van de Linde, M. Schuttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, M. Sauer, *Angew Chem Int Ed Engl* **2008**, *47*, 6172-6176; M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, *Science* **2007**, *317*, 1749-1753; J. Foiling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, C. Eggeling, S. W. Hell, *Nature methods* **2008**, *5*, 943-945). Las imágenes confocales (**Figura 2**) y de campo amplio mostraron una tinción de la membrana plasmática superpuesta de IR en ambos colores después del etiquetado de doble color de TCO* y SCO. Sin embargo, la SRM reveló una distribución heterogénea de grupos de IR en la membrana (**Figura 3a**). De forma destacable, también se ha observado el agrupamiento para otros receptores del factor de crecimiento usando estudios de SRM (S. Wilmes, M. Staufienbiel, D. Lisse, C. P. Richter, O. Beutel, K. B. Busch, S. T. Hess, J. Piehler, *Angew Chem Int Ed Engl* **2012**, *51*, 4868-4871).

Para demostrar la generalidad del enfoque de la invención, las partículas pseudovíricas (VLP) marcadas se ensamblaron mediante la coexpresión de hemaglutinina (HA) de las proteínas del virus de la influenza y la proteína de matriz 1 (M1) (para una revisión, véase J.S. Rossman, R.A. Lamb, *Virology* **2011**, *411*, 229-236). Los genomas virales son compactos y, a menudo, contienen genes superpuestos, lo que hace que la inserción de etiquetas codificadas genéticamente en proteínas virales sea un desafío particular. Se generó un mutante TAG de HA y se expresó junto con M1 y el ARNt^{PyI}/PyIRS^{AF} en células HEK293T. Se repitió el protocolo de pulso-caza usando TCO*, SCO y se marcaron con Me-Tet-Cy5 y H-Tet-Atto532. Como se muestra en la **Figura 2c**, las protuberancias filamentosas teñidas con Atto532 y Cy5, correspondientes a las VLP ensambladas, se hicieron visibles. La resolución mejorada de la SRM permite visualizar filamentos individuales (**Figura 3b**). De forma destacable, existe una importante superposición espacial entre los dos colores, lo que sugiere que las proteínas traducidas en diferentes momentos se incorporan en las mismas VLP de ensamblaje.

En resumen, la reacción SPIEDAC codificada genéticamente se sintonizó en dos reacciones SPIEDAC ortogonales mutuamente que se pueden usar para realizar un etiquetado rápido de proteínas en células vivas. Esto expande el repertorio existente de métodos de etiquetado de "clic" biocompatibles utilizando un código genético expandido de SPIEDAC y SPAAC a SPIEDAC y seSPIEDAC y SPAAC. Las dos reacciones rápidas de SPIEDAC permitieron experimentos de etiquetado de doble color compatibles con SRM en células de mamífero mientras que la reactividad lenta de SPAAC pareció insuficiente para el etiquetado rápido de alto contraste de células vivas.

TCO* es un TCO mejorado con mayor bioestabilidad y eficiencia de incorporación. El TCO* reacciona rápidamente con ambas tetrazinas ensayadas (Me-Tet, H-Tet), mientras que BCN tiene una reactividad comparativamente mucho más reducida hacia Me-Tet. En las condiciones experimentales realizadas, SCO solo reaccionó con H-Tet y no con Me-Tet. El etiquetado de doble color se logró usando un par de ARNt/RS promiscuo y un enfoque de pulso-caza. La etapa de etiquetado se realiza en células vivas, creando nuevas posibilidades para estudiar el destino de las proteínas con una resolución muy alta. La combinación del etiquetado de doble color de la invención con interruptores genéticos, tales como los mutantes sensibles a la temperatura y el control del promotor, podría permitir que se etiqueten distintas proteínas. Sin embargo, las químicas de etiquetado de la presente invención son generales y también pueden aplicarse directamente a la codificación específica a través de dos codones distintos.

El pequeño tamaño de la etiqueta UAA, en comparación con otras etiquetas fluorescentes codificadas genéticamente, es una ventaja importante, especialmente para los estudios de conjuntos de proteínas complejas tales como IR y HA, donde se pueden influir múltiples interacciones funcionales con otras proteínas y lípidos por etiquetas más grandes de manera impredecible. En particular, los genomas virales con frecuencia son extremadamente compactos y no toleran grandes modificaciones. La necesidad de cambiar solo un solo codón, por lo tanto, aumenta drásticamente la posibilidad de encontrar sitios permisivos que no alteren la función de la proteína.

Dado que las técnicas de la presente invención se basan en el mecanismo de ligadura genérico de dos reacciones SPIEDAC sintonizadas, por lo tanto, será compatible con cualquier desarrollo de tinte adecuado para el etiquetado intracelular vivo y también aplicable a una amplia gama de otras disciplinas para la instalación de etiquetas, tales como estudios de MRI y PET.

Ejemplo P: Ensayo de ¹H RMN para probar la ortogonalidad de las reacciones de seSPIEDAC y SPIEDAC

Se prepararon soluciones madre de los dienófilos en DMSO-d6 con una concentración final de 10 mM. Se preparó también una solución madre 20 mM de la sal del ácido fórmico de (4-(6-metil-1,2,4,5-tetrazina-3-il)fenil)metanamina (denominada metil tetrazina o Me-Tet-NH₂) en DMSO-d6. Como experimento de control, el dienófilo o el dieno (Me-Tet-NH₂) se mezclaron con DMSO-d6, 1:1 (v/v; 760 µl de volumen total) para producir una concentración final de 5 mM o 10 mM, respectivamente, en el tubo de RMN. Para confirmar si un dienófilo reacciona con el dieno (Me-Tet-NH₂), las soluciones madre correspondientes se mezclaron 1:1 (v/v; 760 µl de volumen total) para producir una concentración final de 5 mM del dienófilo y 10 mM del dieno en el tubo de RMN. Las muestras se incubaron durante 10 min a temperatura ambiente antes de analizarse por ¹H RMN.

La comparación de los espectros de ¹H RMN del dienófilo y el dieno con la mezcla de reacción mostró que el *trans*-ciclooct-2'-eno había reaccionado completamente con la tetrazina 10 min después de mezclar el *trans*-ciclooct-2'-eno con la tetrazina (**Figura 11a**). No se detectaron señales correspondientes al doble enlace *trans* (puntos de color negro en el espectro central) después de la incubación. Como cabía esperar, se detectó la tetrazina restante (puntos de color negro en el espectro inferior), ya que se usó en exceso. Se detectaron nuevas señales en la región aromática (espectro superior) correspondientes a los productos de reacción recién formados. Se encontró que la reacción de un isómero diferente de *trans*-ciclooct-2'-eno es más lenta que la reacción del *trans*-ciclooct-2'-eno (**Figura 11b**). Las flechas de color negro indican la disminución de la integral de los protones de doble enlace *trans* y la aparición de nuevas señales en la región aromática tras la reacción (espectro superior). La reacción se impulsó hasta finalizar extendiendo el tiempo de reacción, aumentando la temperatura o utilizando un mayor exceso de tetrazina.

Por el contrario, ni los restos 5-norborneno-2-ol (isómero *endo* y *exo*) ni ciclooctino reaccionaron en absoluto con la tetrazina (indicado por puntos de color negro en la **Figura 11c**, **Figura 11d** y la **Figura 11e**). No se detectaron nuevas señales en 10 minutos después de mezclar el dienófilo con la tetrazina. La reacción con el resto biclononino (dos isómeros), sin embargo, tubo lugar, aunque todavía no había reaccionado completamente con la tetrazina 10 minutos después de la mezcla (espectro superior de la **Figura 11f**). Por lo tanto, se encontró que la reacción de este par de reacción era más lenta que la reacción del par de reacción en la **Figura 11a**. Las flechas de color negro indican las nuevas señales que aparecen tras la reacción (espectro superior). La reacción se impulsó hasta finalizar extendiendo el tiempo de reacción, aumentando la temperatura o utilizando un mayor exceso de tetrazina.

Los experimentos similares realizados en 1x de PBS deuterado (dPBS) en D₂O (pH 7,4) mezclado con 1,4-dioxano-d8 (v/v 1:1) dieron los mismos resultados.

65 Abreviaturas

- AcF = p-acetilfenilalanina
 AcOH = ácido acético
 ac. = acuoso
 Ar = argón/atmósfera de gas inerte
 5 BCN = N-ε-((1R,8S,9S)-bicyclo[6,1,0]non-4-in-9-metiloxi)carbonil)-L-lisina (**Figura 1a: compuesto 1**)
 Boc-L-Lys-OH = N-α-*terc*-butiloxicarbonil-L-lisina
 salmuera = solución ac. sat. ac. de NaCl
 calc. = calculado
 cHex = ciclohexano
 10 conc. = concentrado
 DBU = 1,8-diazabicyclo[5.4.0]undec-7-eno
 DCM = diclorometano
 DIEA = N,N-diisopropiletilamina
 DMF = dimetilformamida
 15 DMSO = dimetilsulfóxido
 equiv. = equivalente(s)
 EtOH = etanol EtOAc = acetato de etilo
 FC = cromatografía ultrarrápida
 Fmoc-Lys-OH = N-α-(9-fluorenilmetiloxicarbonil)-L-lisina
 20 FRET = transferencia de energía por resonancia de fluorescencia, también denominada transferencia de energía por resonancia de Forster
 MeOH = metanol
 GFP = proteína verde fluorescente
 GFP^{WT} = GFP de tipo silvestre
 25 GFPTAQ = GFP codificada por una secuencia que comprende el codón de terminación ámbar TAG en el sitio permisivo 39
 GFP^{TAG→x} = GFP^{TAG} en la que el compuesto x se ha incorporado en el sitio codificado por ámbar, por ejemplo, GFP^{TAG→1} = GFP^{TAG} en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar
 h = hora(s)
 30 I_{GFP} = intensidad promedio de GFP
 IR = receptor de insulina
 GFP-IR^{TAG} = construcción de fusión GFP-IR codificada por una secuencia que comprende el codón de terminación ámbar TAG en el sitio permisivo 676
 GFP-IR^{TAG→x} = fusión GFP-IR^{TAG} en la que el compuesto x se ha incorporado en el sitio codificado por ámbar, por ejemplo, GFP-IR^{TAG→1} = fusión GFP-IR^{TAG} en la que el compuesto 1 se ha incorporado en el sitio 676 del IR
 35 MBP = proteína de unión a maltosa
 MBP^{TAG} = MBP codificada por una secuencia que comprendía el codón de terminación de ámbar TAG en el sitio permisivo 38 y un marcador de His C-terminal
 MBP^{TAG→1} = MBP^{TAG} en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar
 40 mCherry^{WT} = mCherry de tipo silvestre
 mCherry^{TAG→1} = mCherry en la que el compuesto 1 se ha incorporado en el sitio codificado por ámbar
 med. = medido
 min = minutos
 SLN = secuencia de localización nuclear
 45 DO₆₀₀ = densidad óptica a 600 nm
 o/n = durante una noche
 PBS = solución salina tamponada con fosfato
 PMSF = fluoruro de fenilmetilsulfonilo
 RS = aminoacil ARNt sintetasa
 50 ta = temperatura ambiente (20-25 °C)
 sat. = saturado
 SCO = N-ε-((Ciclooct-2-in-1-iloxi)carbonil)-L-lisina (**Figura 1a: compuesto 5**)
 DT = desviación típica
 SDS-PAGE = electroforesis en gel de poliacrilamida de dodecil sulfato sódico
 55 smFRET = observación de una única molécula de FRET TAMRA = tetrametilrodamina
 TB = caldo de cultivo Terrific
 TCO = N-ε-((*trans*-Ciclooct-4-en-1-iloxi)carbonil)-L-lisina (**Figura 1a: compuesto 2**)
 TCO* = N-ε-((*trans*-Ciclooct-2-en-1-iloxi)carbonil)-L-lisina (**Figura 1a: compuesto 3**)
 TCO# = N-ε-((*trans*-Ciclooct-3-en-1-iloxi)carbonil)-L-lisina (**Figura 1a: compuesto 4**)
 60 TEA = trietilamina
 TFA = ácido trifluoroacético
 THF = tetrahidrofurano
 TLC = cromatografía de capa fina
 UAA = aminoácido no natural
 65

LISTA DE SECUENCIAS

<110> European Molecular Biology Laboratory

5 <120> Múltiples reacciones de cicloadición para el etiquetado de moléculas

<130> M/54540-PCT

10 <150> EP14151175.8
<151> 14/01/2014

<160> 3

15 <170> PatentIn versión 3.5

<210> 1
<211> 454
<212> PRT
<213> M. mazei

20 <400> 1

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
145 150 155 160

Thr Gly Ala Thr Ala Ser Ala Leu Val Lys Gly Asn Thr Asn Pro Ile

ES 2 699 082 T3

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

ES 2 699 082 T3

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
 <213> Secuencia artificial

<220>
 <223> Pirrolisil ARNt sintetasa de M. maize mutante

<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
 145 150 155 160

Thr Gly Ala Thr Ala Ser Ala Leu Val Lys Gly Asn Thr Asn Pro Ile

5

10

ES 2 699 082 T3

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

ES 2 699 082 T3

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> 3
<211> 19
5 <212> PRT
<213> Secuencia artificial

<220>
10 <223> Secuencia peptídica parcial

<220>
<221> misc_feature
<222> (13)..(13)
15 <223> Xaa es un aminoácido sintético

<400> 3

Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Xaa Gly Lys Leu
1 5 10 15

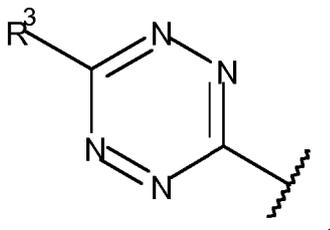
Thr Leu Lys

20

REIVINDICACIONES

1. Un método para formar uniones mediante reacciones de cicloadición, en donde el método comprende hacer reaccionar una primera tetrazina con un primer dienófilo seguido de hacer reaccionar una segunda tetrazina con un segundo dienófilo, en donde la reacción de la primera tetrazina con el primer dienófilo se desarrolla en presencia del segundo dienófilo, en donde

(i) la primera tetrazina comprende un grupo de la fórmula:



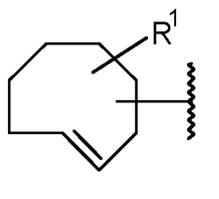
10

en la que

R^3 es alquilo C₁-C₃;

15

(ii) el primer dienófilo comprende un grupo *trans*-ciclooctenilo de la fórmula:



20

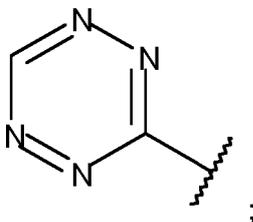
en la que

R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

25

R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

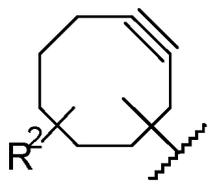
(iii) la segunda tetrazina comprende un grupo de la fórmula:



30

y

(iv) el segundo dienófilo comprende un grupo ciclooctinilo de la fórmula:



35

en la que

R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN,

hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y
R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

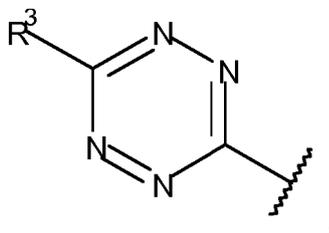
5 2. El método de la reivindicación 1, en el que la primera tetrazina reacciona preferiblemente con el primer dienófilo en presencia del segundo dienófilo.

10 3. El procedimiento de las reivindicaciones 1 o 2, en el que la constante de velocidad *k* de la primera tetrazina con el primer dienófilo es normalmente al menos 10² veces mayor que la constante de velocidad *k* de la reacción de la primera tetrazina con el segundo dienófilo.

15 4. El método de una cualquiera de las reivindicaciones 1-3, en el que la constante de velocidad *k* de la primera tetrazina con el primer dienófilo se deja avanzar durante 30 minutos o menos a una temperatura de aproximadamente 37 °C.

5. El método de una cualquiera de las reivindicaciones 1-4, en donde el método comprende poner en contacto una molécula diana o una composición de molécula diana con

20 (i) un primer agente de etiquetado que comprende un grupo de la fórmula:

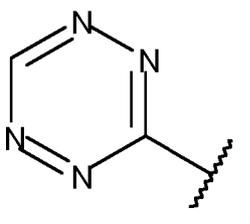


en la que

25 R³ es alquilo C₁-C₃;

seguido de

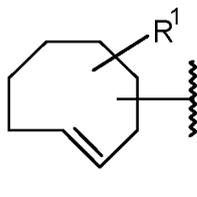
(ii) un segundo agente de etiquetado que comprende un grupo de la fórmula:



30

en la que la molécula diana comprende

35 (i) un grupo *trans*-ciclooctenilo de la fórmula:

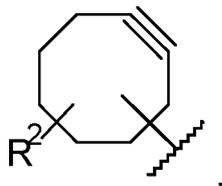


en la que

40 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

45 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

(ii) un grupo ciclooctinilo de la fórmula:

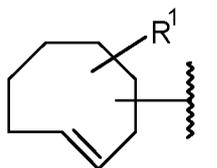


5 en la que

R^2 es hidrógeno, halógeno, alquilo C₁-C₄, $(R^cO)_2P(O)O$ -alquilo C₁-C₄, $(R^dO)_2P(O)$ -alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
 R^c , R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅,

en donde la composición de moléculas diana comprende

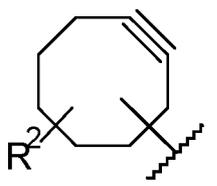
15 (i) un primera molécula diana que comprende un grupo *trans*-ciclooctenilo de la fórmula:



20 en la que

R^1 es hidrógeno, halógeno, alquilo C₁-C₄, $(R^aO)_2P(O)O$ -alquilo C₁-C₄, $(R^bO)_2P(O)$ -alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
 R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

(ii) una segunda molécula diana que comprende un grupo ciclooctinilo de la fórmula:



30

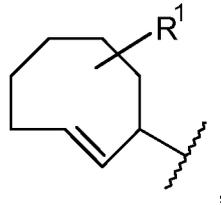
en la que

R^2 es hidrógeno, halógeno, alquilo C₁-C₄, $(R^cO)_2P(O)O$ -alquilo C₁-C₄, $(R^dO)_2P(O)$ -alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;
 R^c , R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

35

6. El método de una cualquiera de las reivindicaciones 1-5, en el que el grupo *trans*-ciclooctenilo tiene la fórmula:

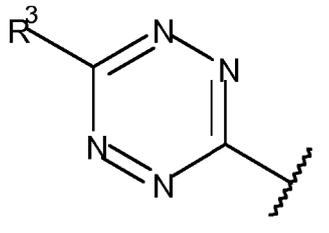
40



en la que

5 R¹ es hidrógeno;

la primera tetrazina o el agente de etiquetado comprenden un grupo de la fórmula:

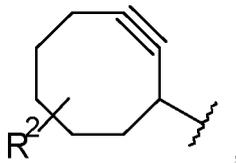


10

en la que

R³ es metilo;

15 el grupo ciclooctinilo tiene la fórmula:



en la que

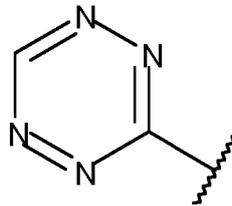
20

R² es hidrógeno;

y

la segunda tetrazina o el agente de etiquetado comprenden un grupo de la fórmula:

25

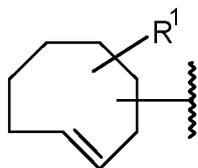


7. El método de una cualquiera de las reivindicaciones 1-6, en el que el primer dienófilo o la molécula diana o la composición de moléculas diana se hace reaccionar con la primera tetrazina o el agente de etiquetado en condiciones que permiten que sustancialmente todos los grupos *trans*-ciclooctenilo reaccionen antes de reaccionar el segunda dienófilo o la molécula diana o la composición de moléculas diana con la segunda tetrazina o el agente de etiquetado.

30

8. Una célula que comprende

35 (i) un grupo *trans*-ciclooctenilo de la fórmula:



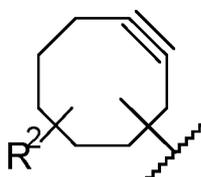
en la que

5 R^1 es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

10 R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

y

(ii) un grupo ciclooctinilo de la fórmula:



15

en la que

20 R^2 es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

R^c , R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

25 9. La célula de la reivindicación 8, en la que el grupo ciclooctinilo está unido a un primer polipéptido y el grupo *trans*-ciclooctenilo está unido a un segundo polipéptido, siendo diferentes el primer y el segundo polipéptidos.

10. Un método para preparar la célula de las reivindicaciones 8 o 9, que comprende

30 a) proporcionar una célula que comprende:

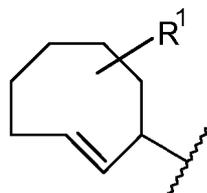
(i) una primera aminoacil ARNt sintetasa, o un polinucleótido que la codifica; y opcionalmente una segunda aminoacil ARNt sintetasa, o un polinucleótido que la codifica;

35 (ii) un primer ARNt que tiene un anticodón para un primer codón selector, o un polinucleótido que codifica dicho ARNt; y opcionalmente un segundo ARNt que tiene un anticodón para un segundo codón selector, o un polinucleótido que codifica dicho ARNt; y

(iii) un polinucleótido que codifica un polipéptido diana y que comprende uno o más de un primer y segundo codón o codones selectores; o un primer polinucleótido que codifica un primer polipéptido diana y que comprende uno o más de un primer codón selector(es) y un segundo polinucleótido que codifica un segundo polipéptido diana y que comprende uno o más de un segundo codón selector(es),

40

en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural que comprende un grupo *trans*-ciclooctenilo de fórmula:



45

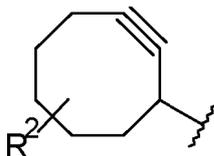
en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

5 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

en donde el grupo α-amino del primer aminoácido no natural se reemplaza opcionalmente por un grupo hidroxilo, y la función del ácido carboxílico del primer aminoácido no natural forma opcionalmente un éster;

10 y con un segundo aminoácido no natural que comprende un grupo ciclooctinilo de la fórmula:



en la que

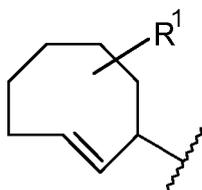
15 R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

20 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

en donde el grupo α-amino del segundo aminoácido no natural está reemplazado opcionalmente por un grupo hidroxilo, y la función del ácido carboxílico del segundo aminoácido no natural forma opcionalmente un éster;

o

25 en donde dicha primera aminoacil ARNt sintetasa (i) es capaz de acilar el primer ARNt (ii) con un primer aminoácido no natural que comprende un grupo *trans*-ciclooctenilo de fórmula:



en la que

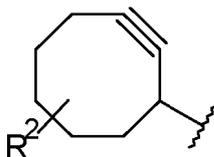
30

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

35 R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

en donde el grupo α-amino del primer aminoácido no natural está reemplazado opcionalmente por un grupo hidroxilo, y la función del ácido carboxílico del primer aminoácido no natural forma opcionalmente un éster; y

40 dicha segunda aminoacil ARNt sintetasa (i) es capaz de acilar el segundo ARNt (ii) con un segundo aminoácido no natural que comprende un grupo ciclooctinilo de la fórmula:



en la que

45

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y

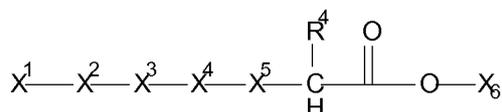
50 R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅, y

en donde el grupo α -amino del segundo aminoácido no natural está reemplazado opcionalmente por un grupo hidroxilo, y la función del ácido carboxílico del segundo aminoácido no natural forma opcionalmente un éster;

b) poner en contacto la célula con el primer y el segundo aminoácidos no naturales; y

5 c) permitir la traducción del polinucleótido o de los polinucleótidos (iii) incorporando de este modo el primer y el segundo aminoácidos no naturales o los análogos de los mismos en el polipéptido o los polipéptidos diana en la posición o las posiciones codificadas por el codón o los codones selector(es).

11. El método de la reivindicación 10, en el que el primer aminoácido no natural es un compuesto de la fórmula:

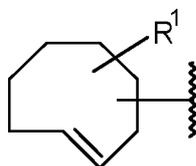


10

en la que:

15 X^1 tiene la fórmula:

15



;

20 R^1 es hidrógeno, halógeno, alquilo C_1-C_4 , $(R^aO)_2P(O)O$ -alquilo C_1-C_4 , $(R^bO)_2P(O)$ -alquilo C_1-C_4 , CF_3 , CN, hidroxilo, alcoxi C_1-C_4 , $-O-CF_3$, alquenoxi C_2-C_5 , alcanoiloxi C_2-C_5 , alquilaminocarboniloxi C_1-C_4 o alquiltio C_1-C_4 , alquilamino C_1-C_4 , di-(alquil C_1-C_4)amino, alquenilamino C_2-C_5 , alquenil C_2-C_5 -alquil C_1-C_4 -amino o di-(alquenil C_2-C_5)amino;

R^a , R^b independientemente son hidrógeno o alcanoiloximetilo C_2-C_5 ;

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

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

25

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 $-(CH_2)_n-$ o fenileno- CH_2- ;

X^6 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^4 es $-OH$ o $-NH_2$;

30

n es un número entero de 1 a 4;

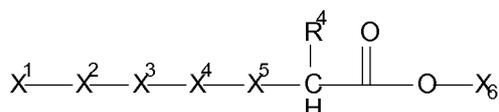
m es un número entero de 1 a 6; y

p es un número entero de 1 a 6,

o una sal de adición de ácidos o bases del mismo, y

35

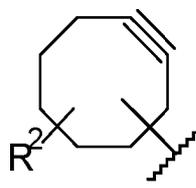
en donde el segundo aminoácido no natural es un compuesto de la fórmula:



40

en la que:

X^1 tiene la fórmula:



;

45

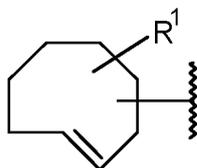
R^2 es hidrógeno, halógeno, alquilo C_1-C_4 , $(R^cO)_2P(O)O$ -alquilo C_1-C_4 , $(R^dO)_2P(O)$ -alquilo C_1-C_4 , CF_3 , CN, hidroxilo,

alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;
 X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;
 X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;
 X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;
 X⁵ es -(CH₂)_n- o fenileno-CH₂-;
 X⁶ es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilulfanil C₂-C₇-alquilo C₁-C₂;
 R⁴ es -OH o -NH₂;
 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.

12. Un polipéptido que comprende

(i) un grupo *trans*-ciclooctenilo de la fórmula:



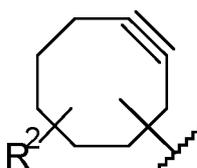
en la que

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

y

(ii) un grupo ciclooctinilo de la fórmula:

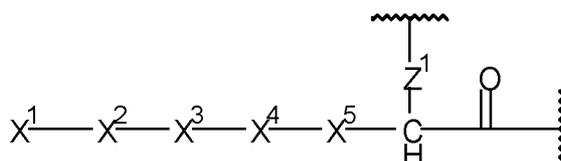


en la que

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

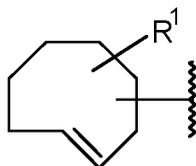
R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

13. El polipéptido de la reivindicación 12, que comprende un primer residuo de la fórmula:



en la que:

X¹ tiene la fórmula



5

R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileno C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

15 X⁵ es -(CH₂)_n- o fenileno-CH₂;

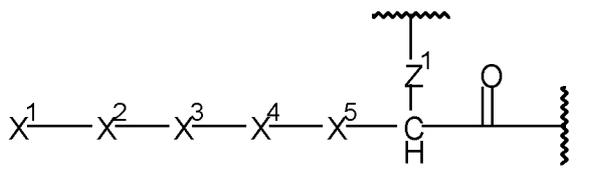
Z¹ es -O- o -NH-;

n es un número entero de 1 a 4;

m es un número entero de 1 a 6; y

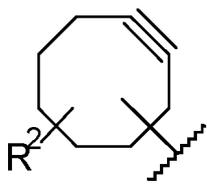
20 p es un número entero de 1 a 6,

y un segundo residuo de la fórmula:



25 en la que:

X¹ tiene la fórmula:



30

R² es hidrógeno, halógeno, alquilo C₁-C₄, (R^cO)₂P(O)O-alquilo C₁-C₄, (R^dO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino;

R^c, R^d independientemente son hidrógeno o alcanoiloximetilo C₂-C₅;

35 X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileno C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p- o un enlace sencillo;

X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

40 X⁵ es -(CH₂)_n- o fenileno-CH₂;

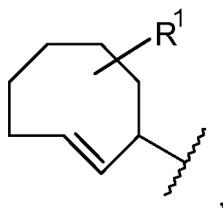
Z¹ es -O- o -NH-;

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.

45 14. Un aminoácido no natural que comprende un grupo *trans*-ciclooctenilo de la fórmula:

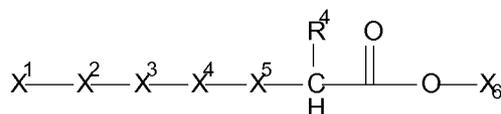


en la que

5 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅; y

10 en donde el grupo α-amino del aminoácido no natural está reemplazado opcionalmente por un grupo hidroxilo, y la función del ácido carboxílico del aminoácido no natural forma opcionalmente un éster.

15. El aminoácido no natural de la reivindicación 14, que tiene la fórmula:

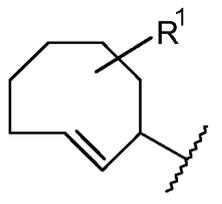


15

en la que

20 X¹ es un grupo *trans*-ciclooctenilo de la fórmula:

20



25 R¹ es hidrógeno, halógeno, alquilo C₁-C₄, (R^aO)₂P(O)O-alquilo C₁-C₄, (R^bO)₂P(O)-alquilo C₁-C₄, CF₃, CN, hidroxilo, alcoxi C₁-C₄, -O-CF₃, alquenoxi C₂-C₅, alcanoiloxi C₂-C₅, alquilaminocarboniloxi C₁-C₄ o alquiltio C₁-C₄, alquilamino C₁-C₄, di-(alquil C₁-C₄)amino, alquenilamino C₂-C₅, alquenil C₂-C₅-alquil C₁-C₄-amino o di-(alquenil C₂-C₅)amino; y R^a, R^b independientemente son hidrógeno o alcanoiloximetilo C₂-C₅.

X² es -CH₂-, -O-, -S-, -NH-, -C(O)-, -OC(O)-, -C(O)O-, -NH-C(O)- o -C(O)-NH-;

X³ es alquileo C₁-C₆, -(CH₂-CH₂-O)_m-, -(CH₂-O)_p-, o un enlace sencillo;

30 X⁴ es -NH-, -C(O)-NH-, -NH-C(O)-, -NH-CH(NH₂)-, -CH(NH₂)-NH-, -NH-C(NH)-NH-, -C(O)-NH-CH(NH₂)-, -C(O)-NH-C(NH)-NH-, NH-CH(NH₂)-C(O)- o -NH-C(NH)-NH-C(O)-;

X⁵ es -(CH₂)_n- o fenileno-CH₂-;

X⁶ es hidrógeno, alquilo C₁-C₆, alcoxi C₁-C₆-alquilo C₁-C₂, alcanoiloxi C₂-C₇-alquilo C₁-C₂ o alcanoilulfanil C₂-C₇-alquilo C₁-C₂;

R⁴ es -OH o -NH₂;

35 n es un número entero de 0 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.

40

Figura 1

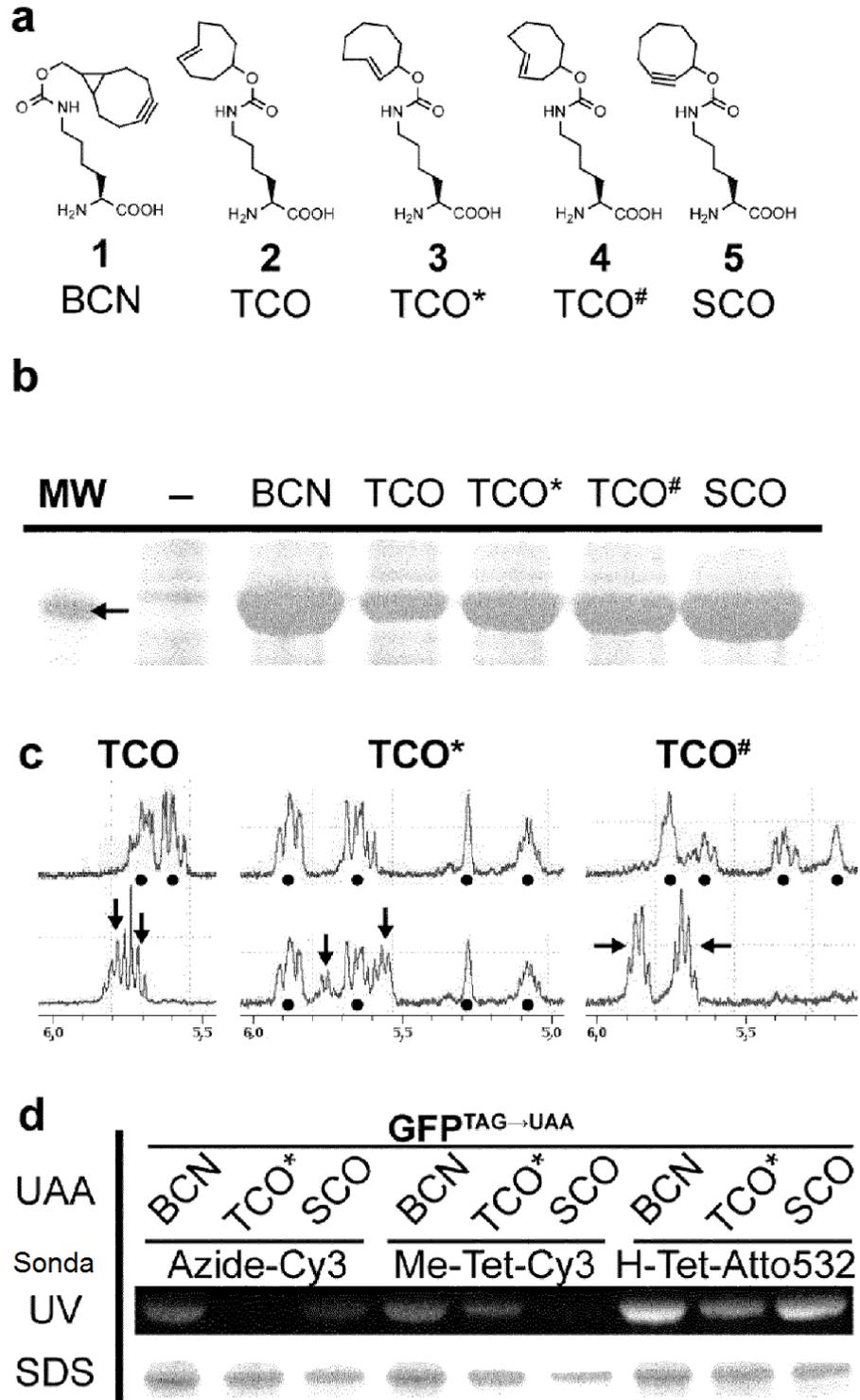


Figura 2

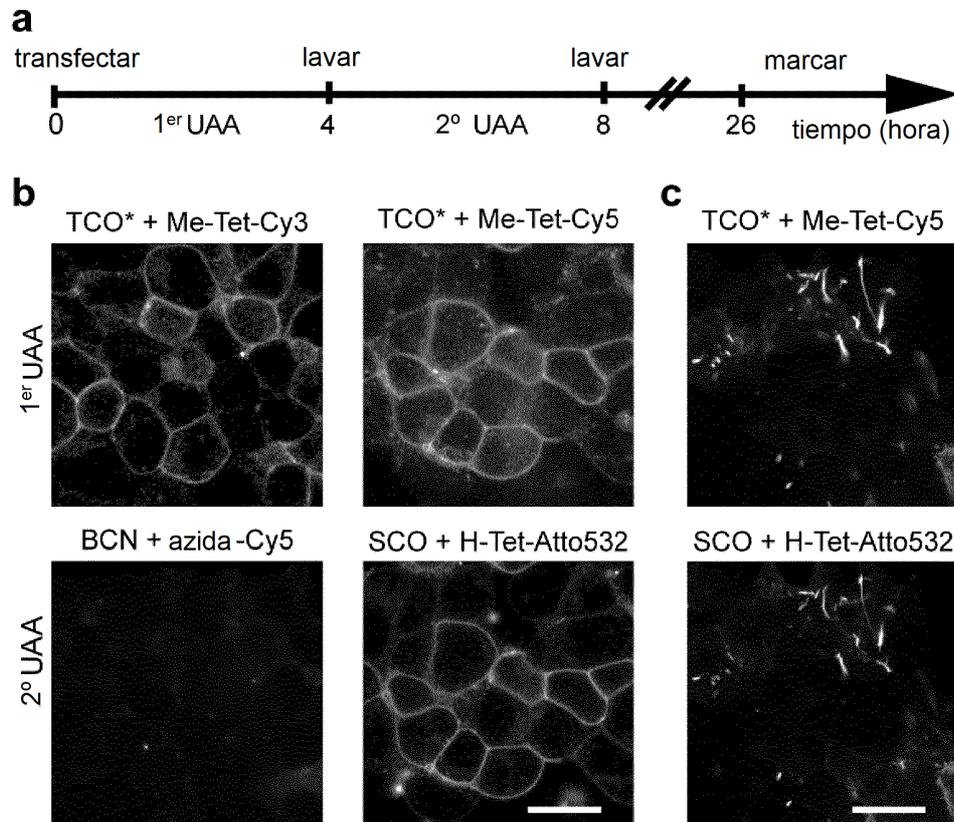


Figura 3

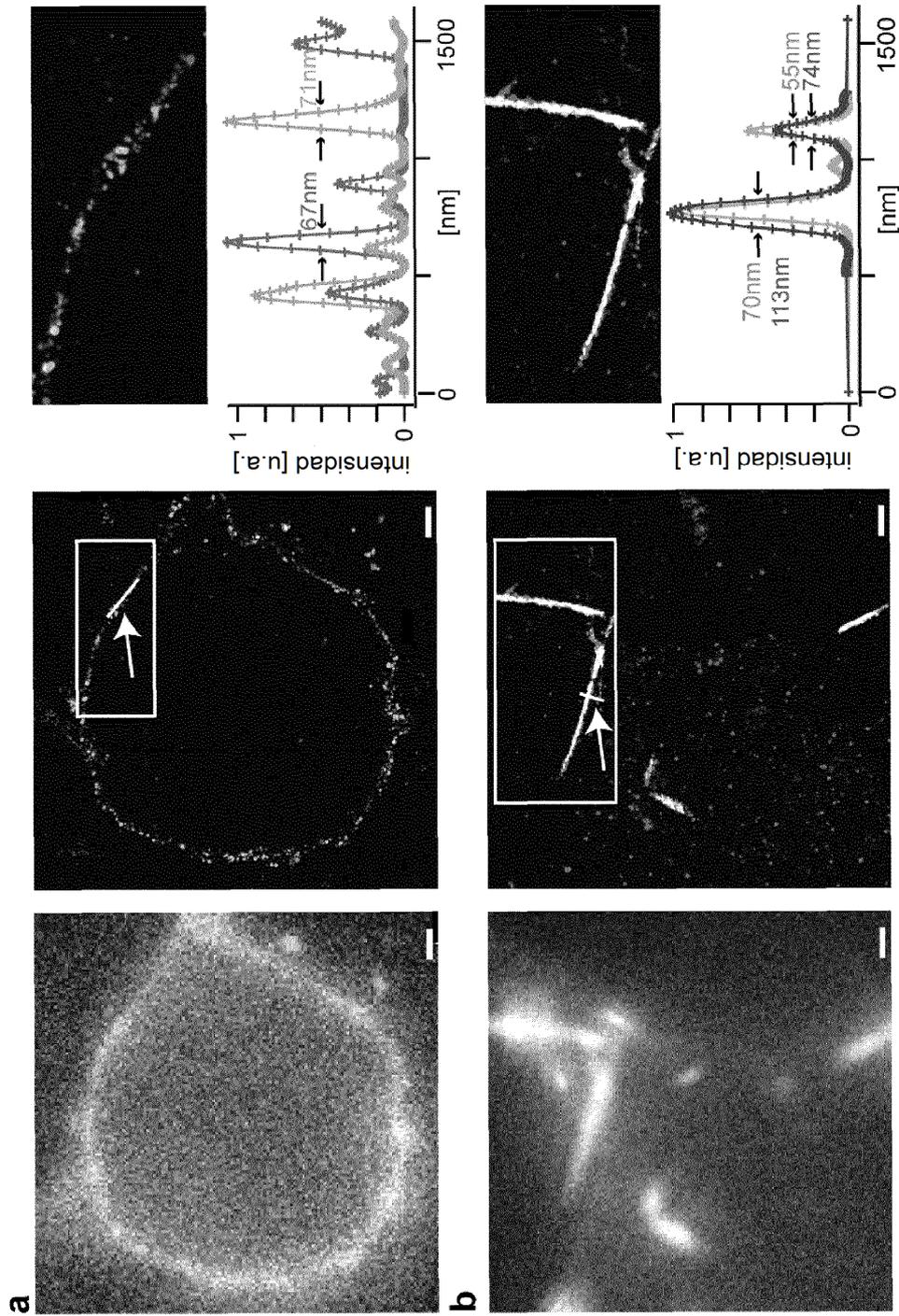
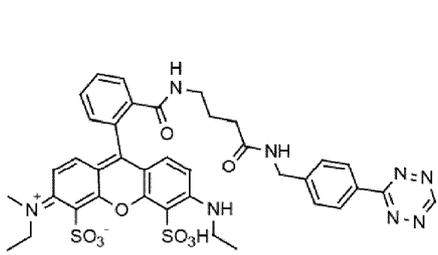
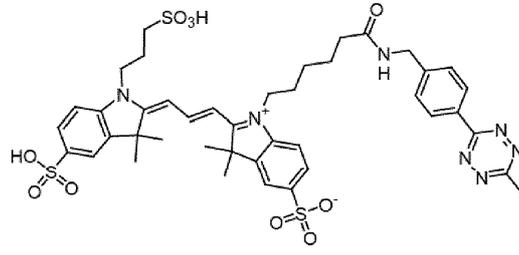


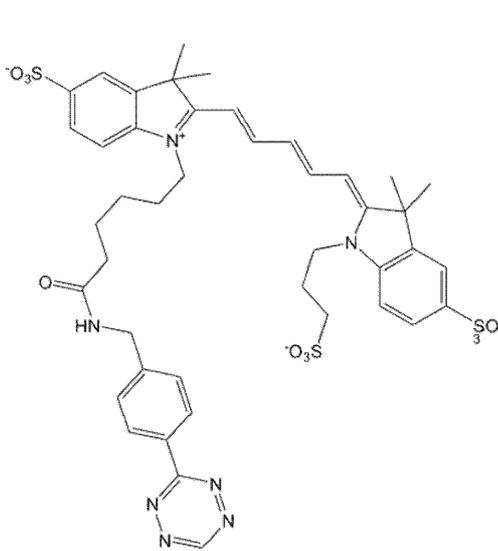
Figura 4



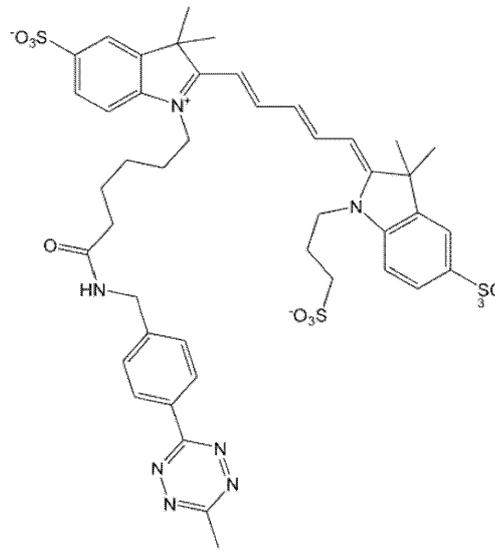
7 (H-Tet-Atto532)



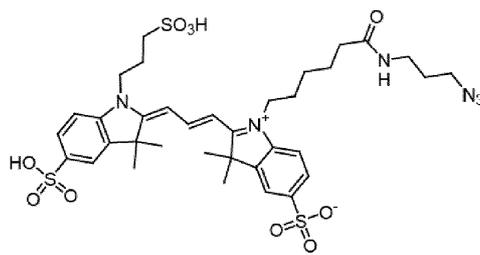
8 (Me-Tet-Cy3)



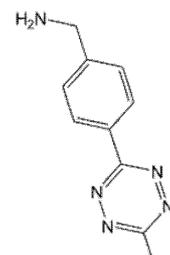
5 H-Tet-Cy5



6 Me-Tet-Cy5



9 (azida -Cy3)



10 Me-Tet-NH₂

Figura 5

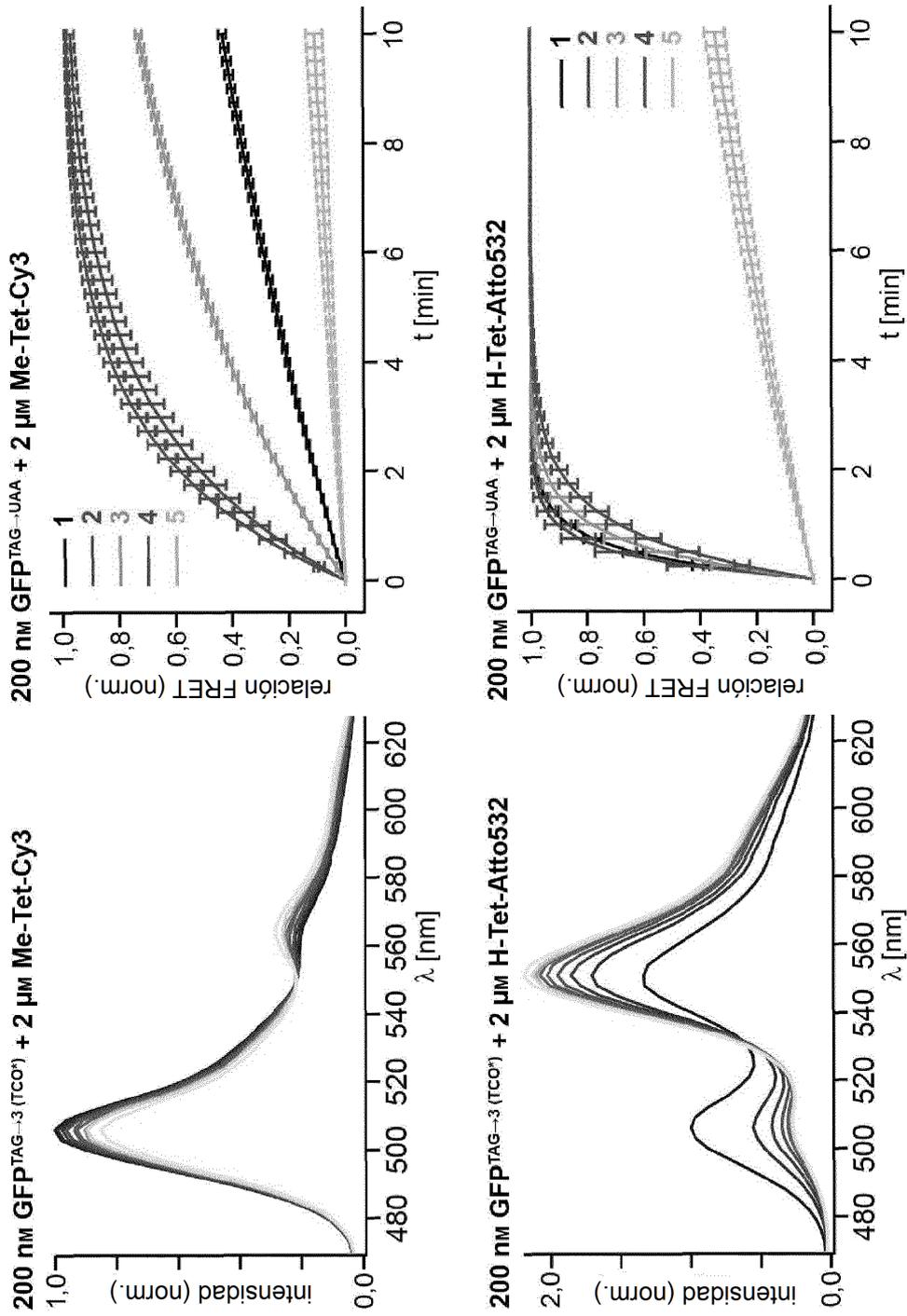


Figura 6

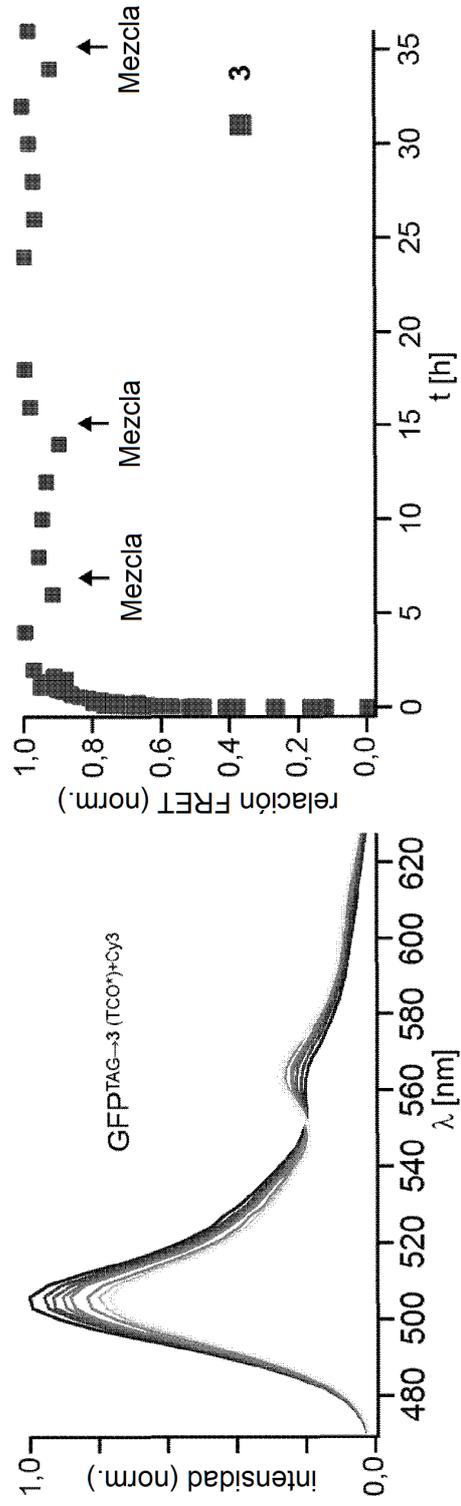


Figura 7

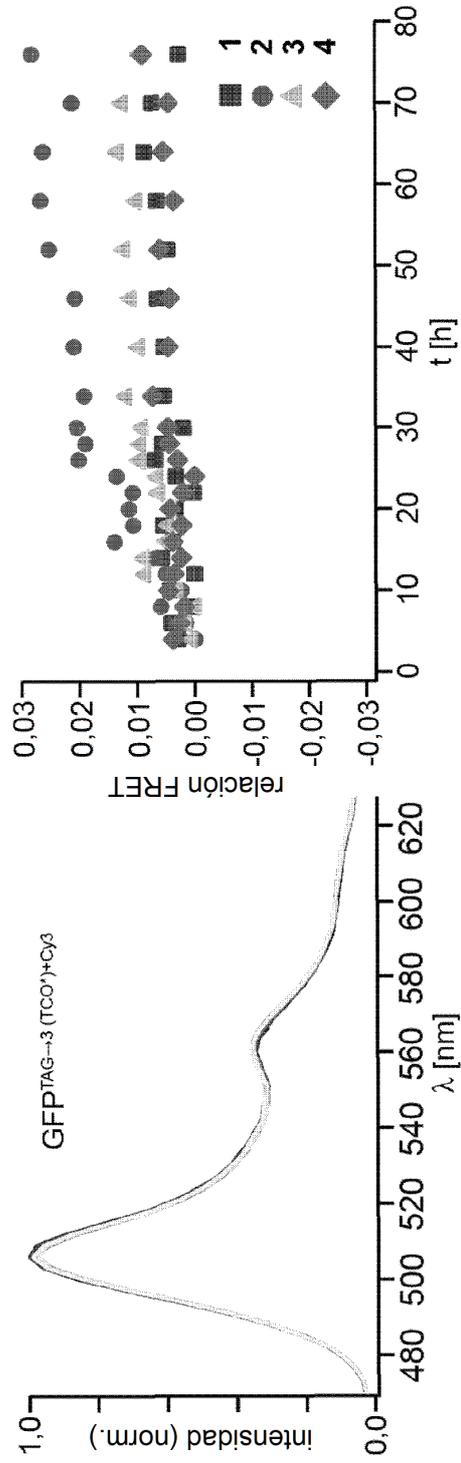
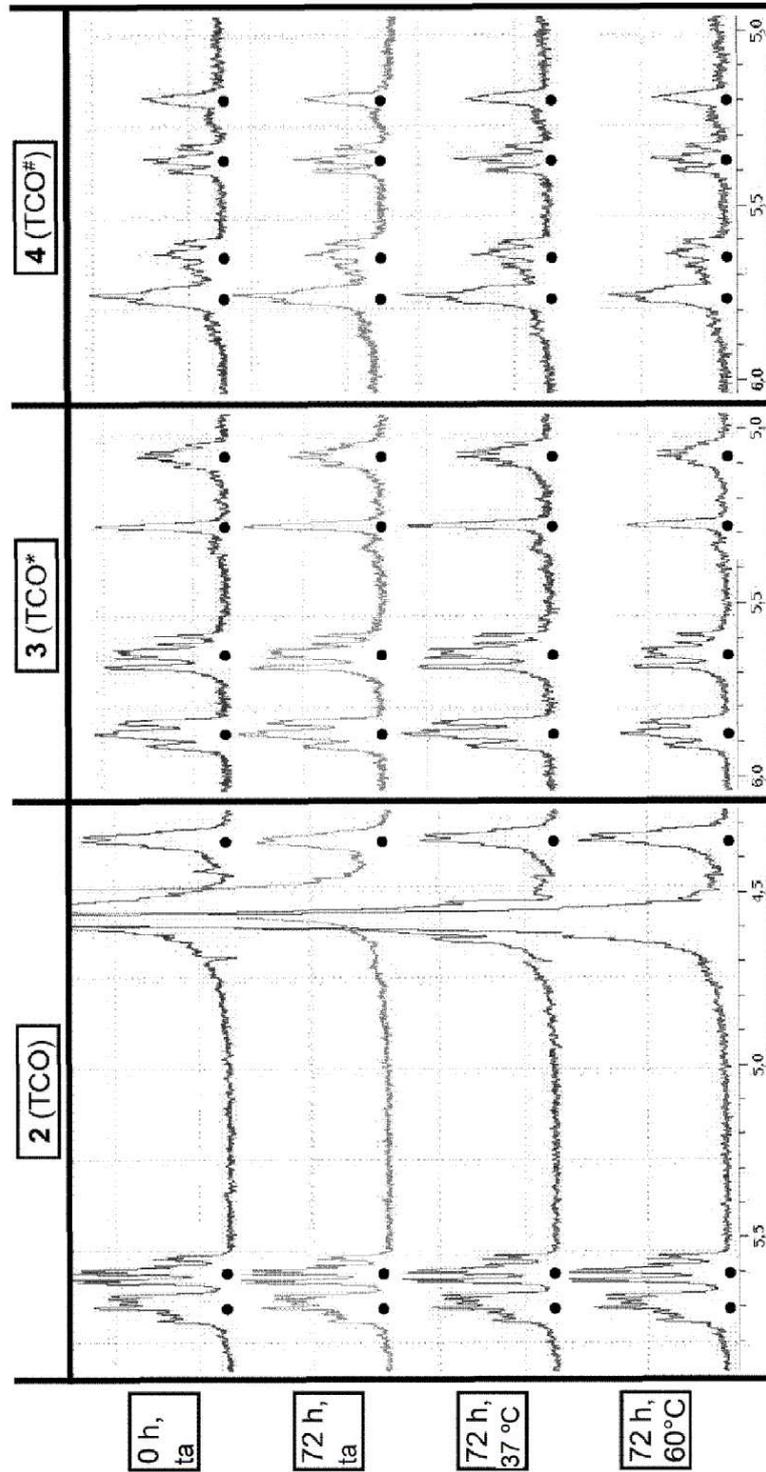
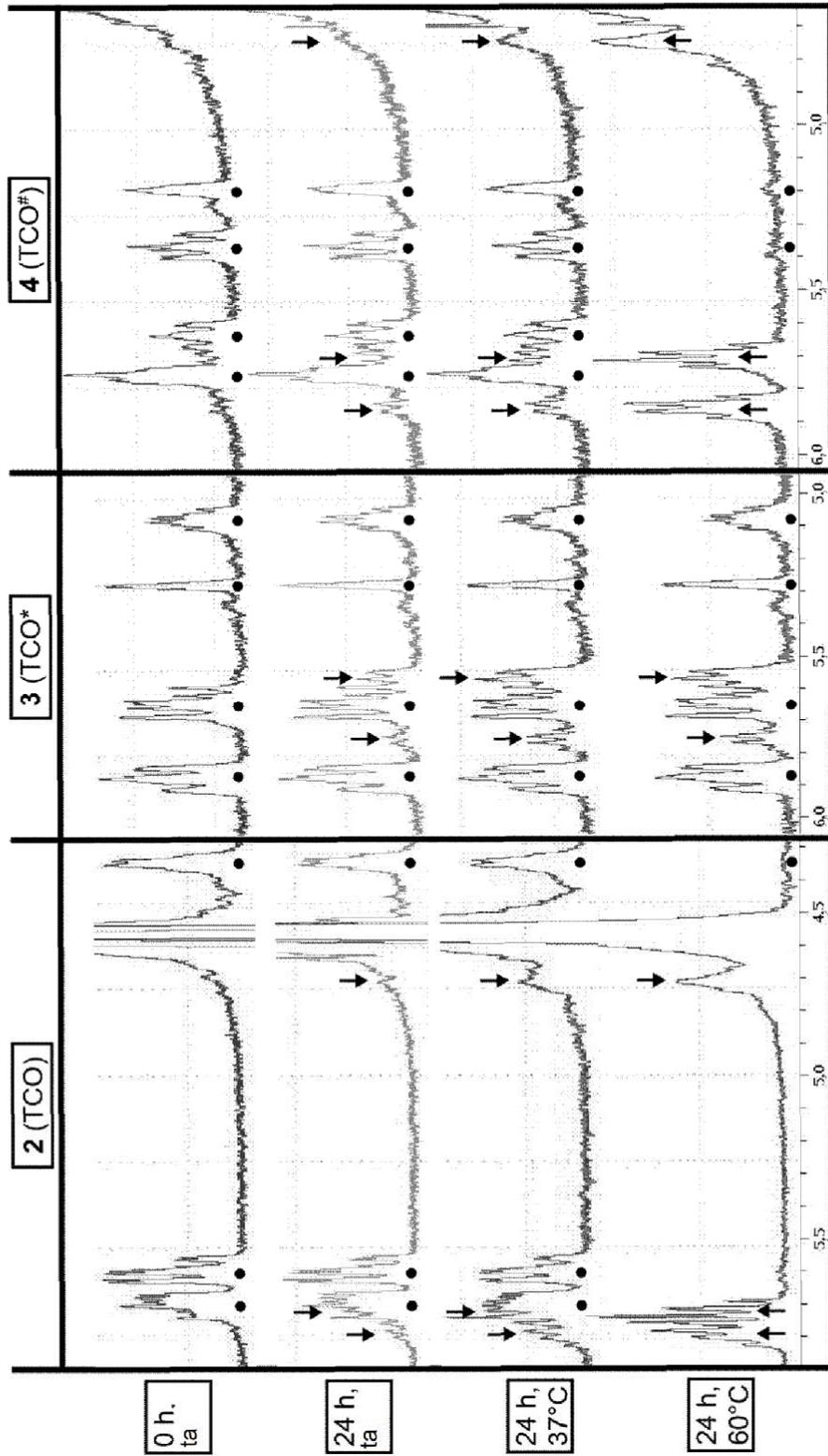


Figura 8

a



b



C

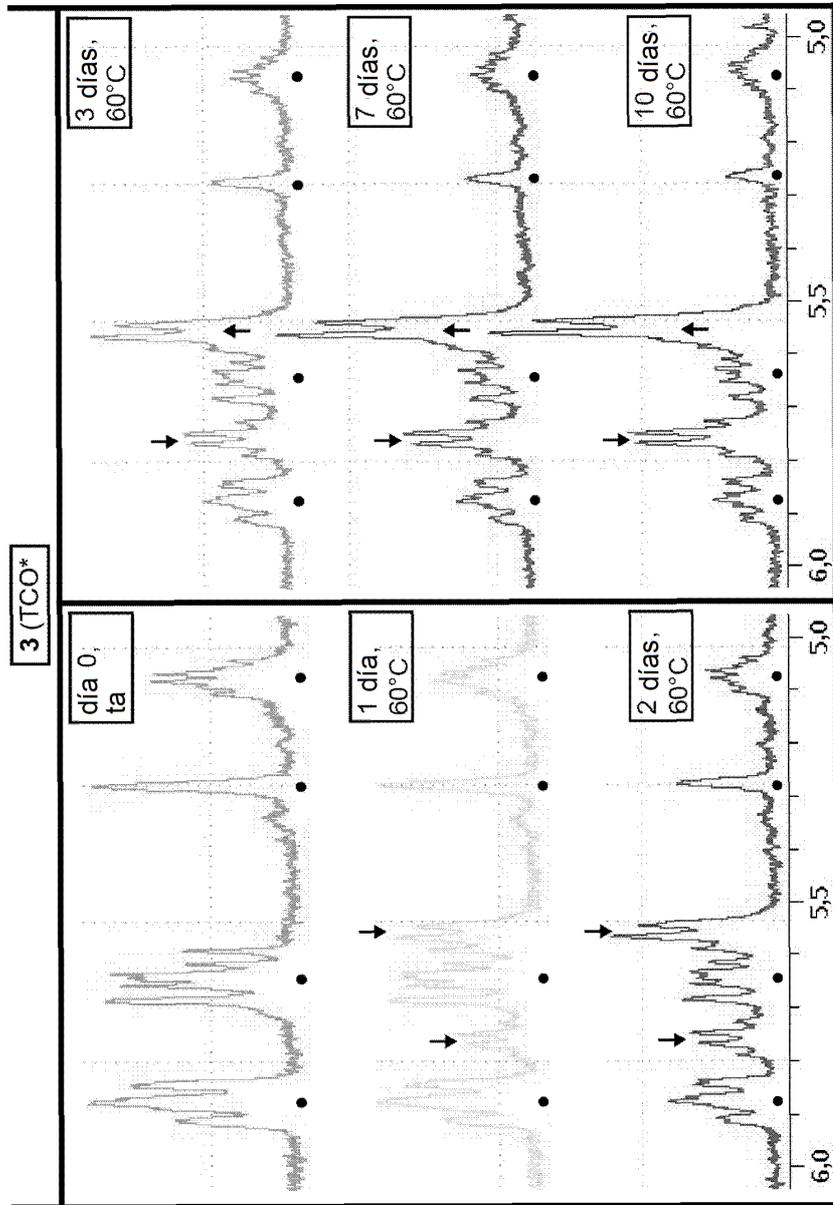


Figura 9

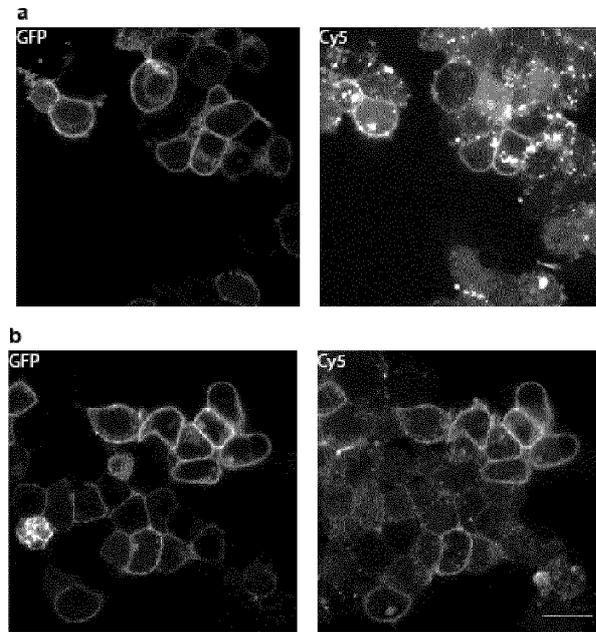
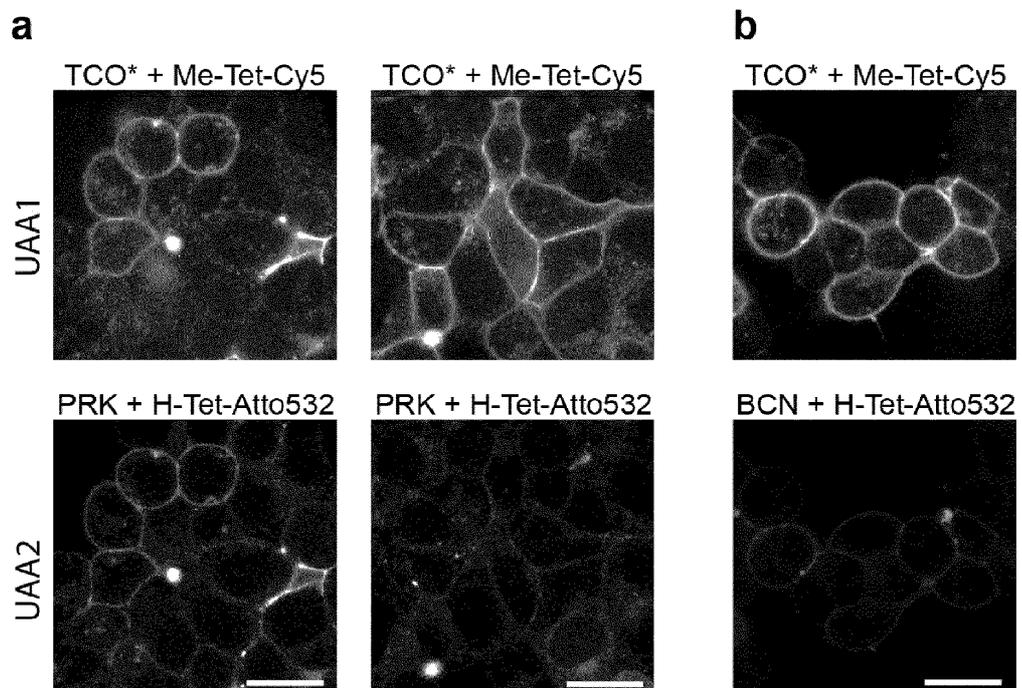
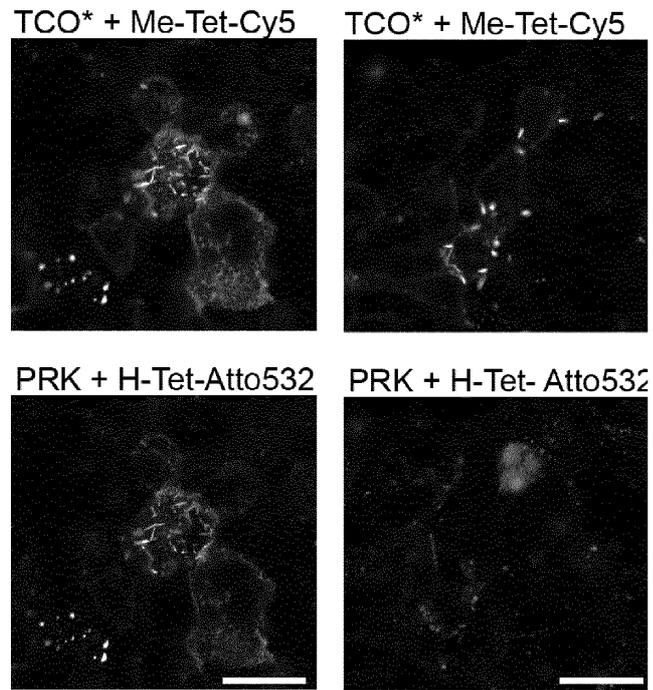


Figura 10



c



d

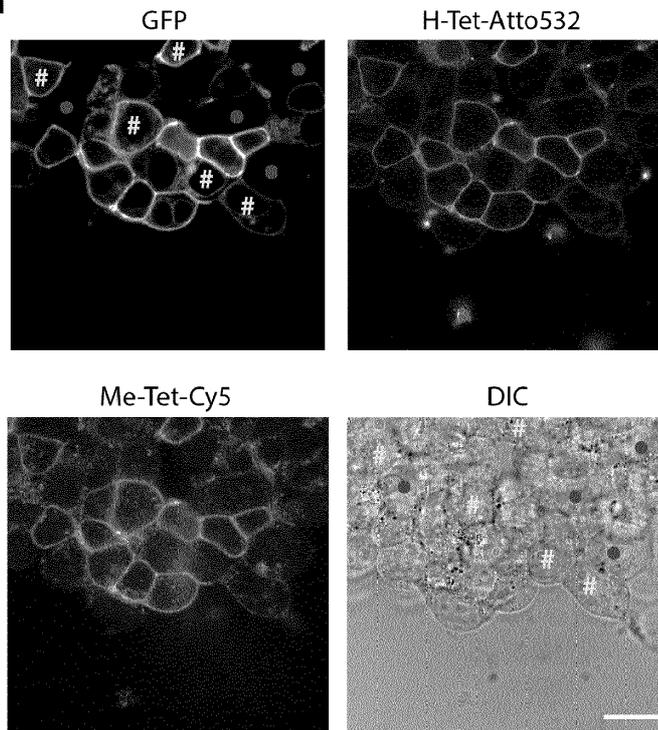
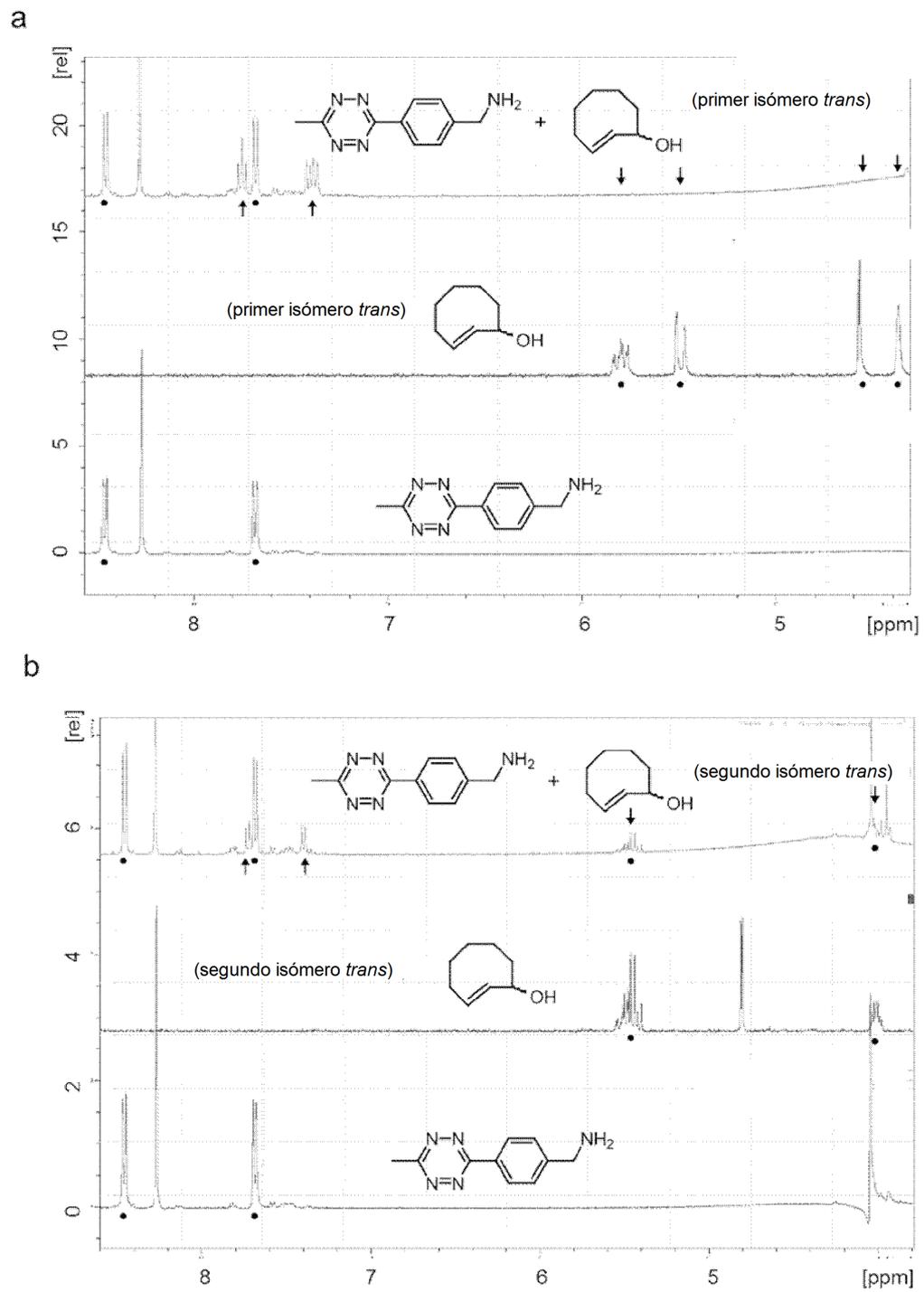
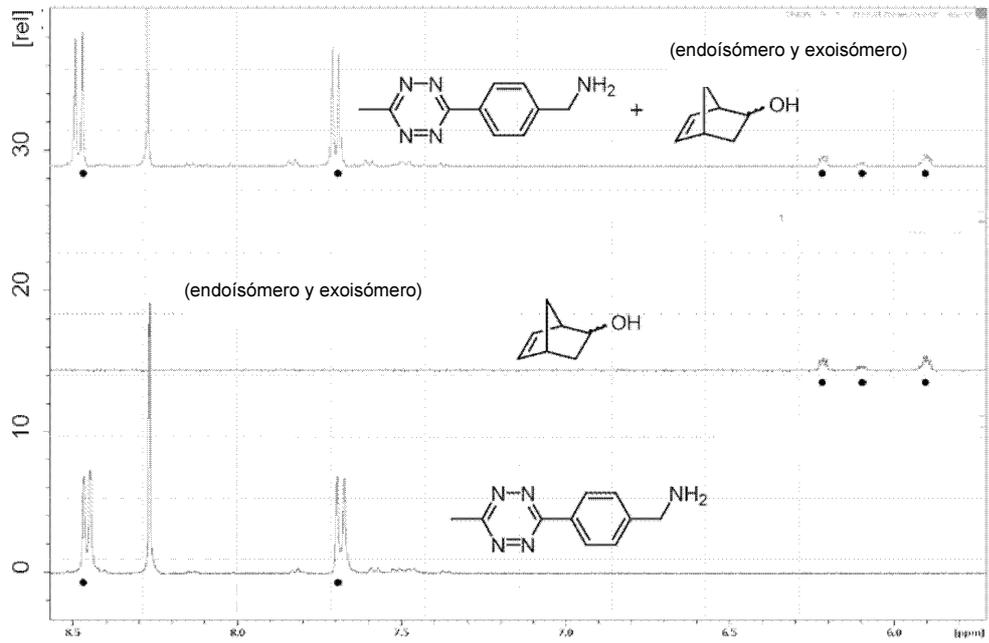


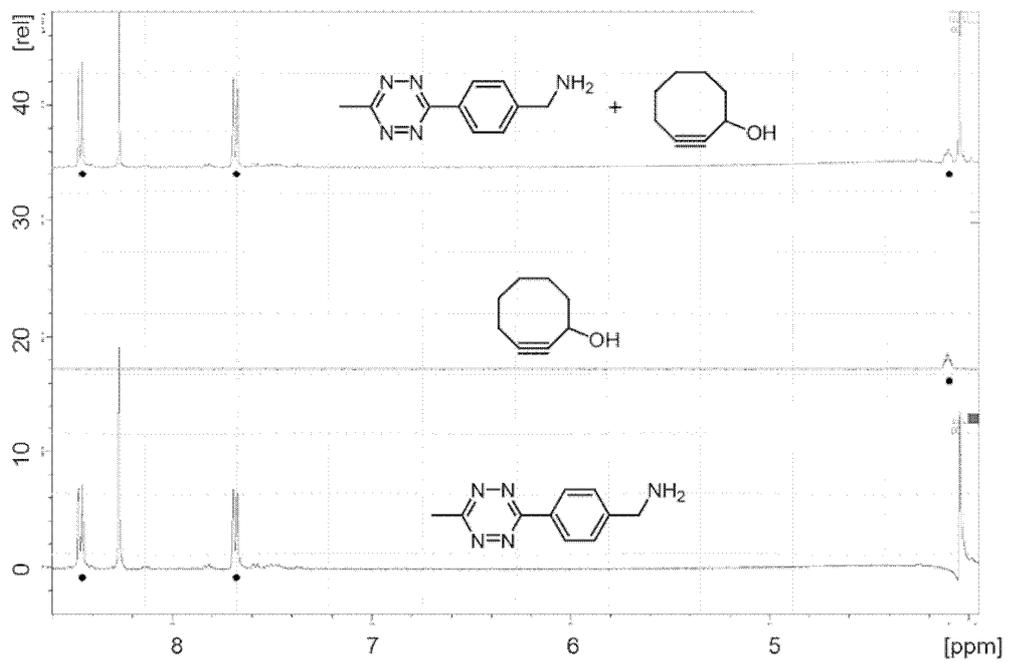
Figura 11



c



d



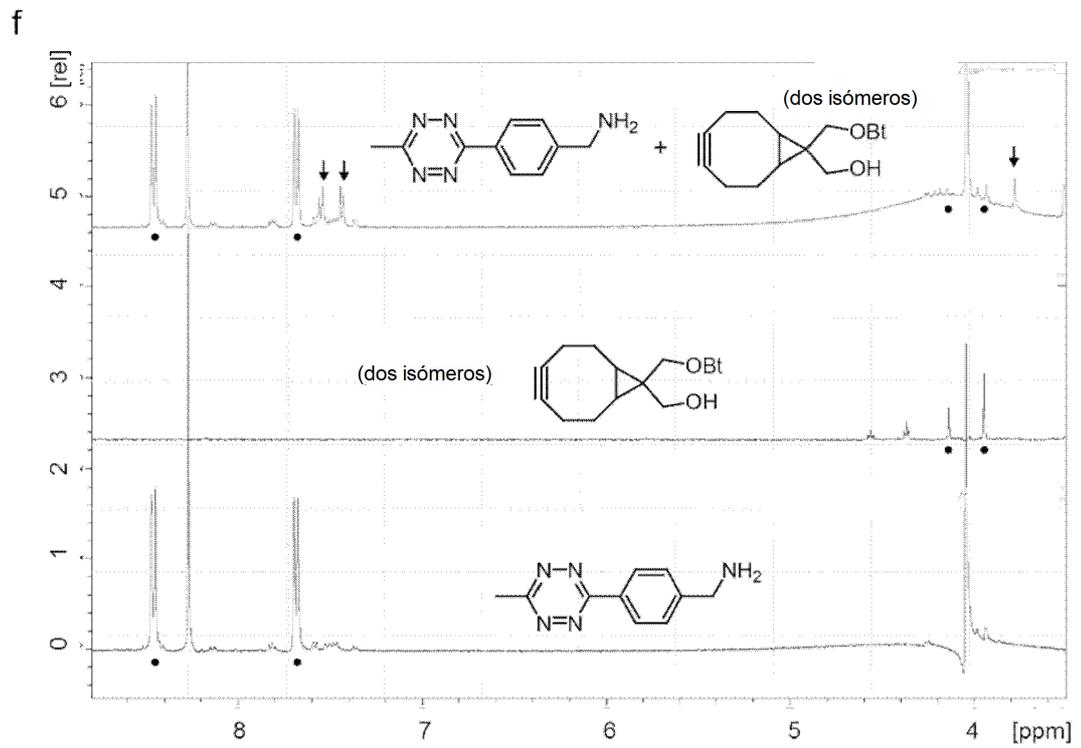
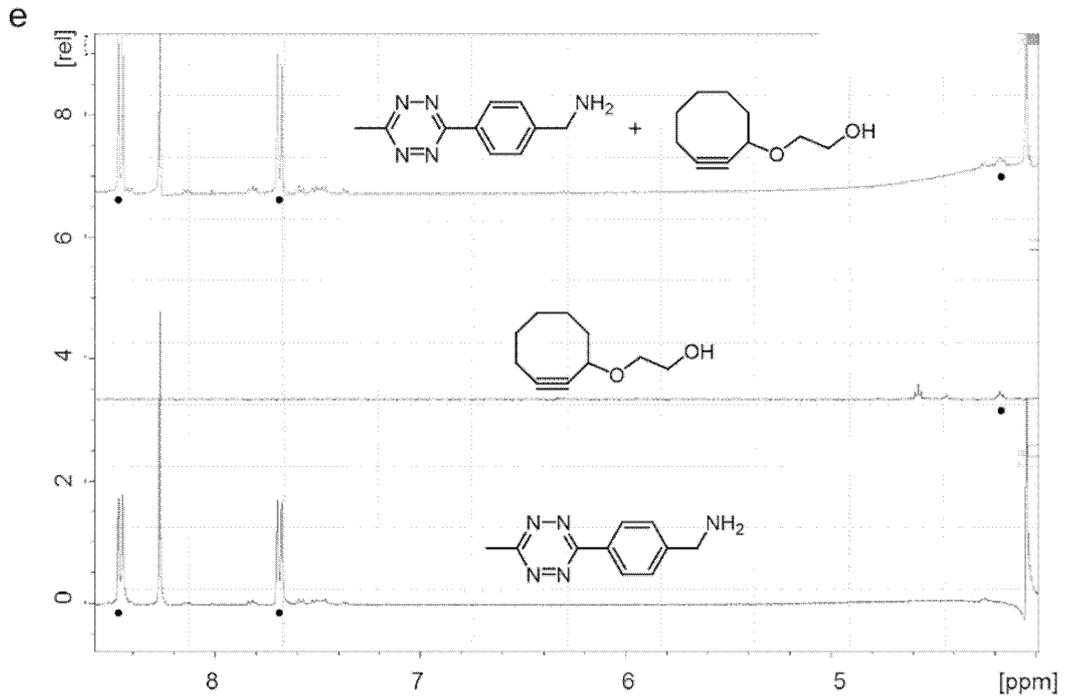


Figura 12

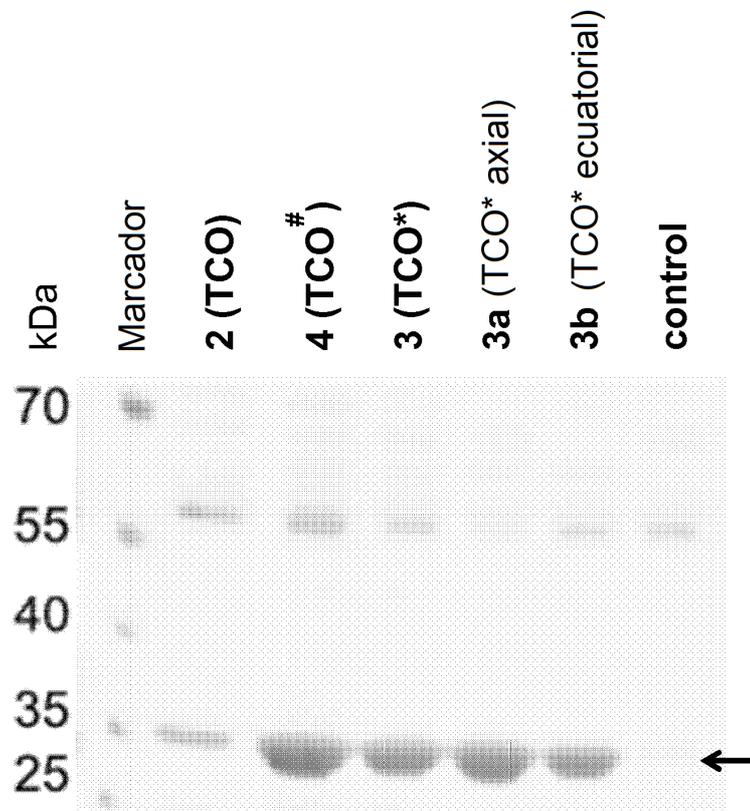


Figura 13

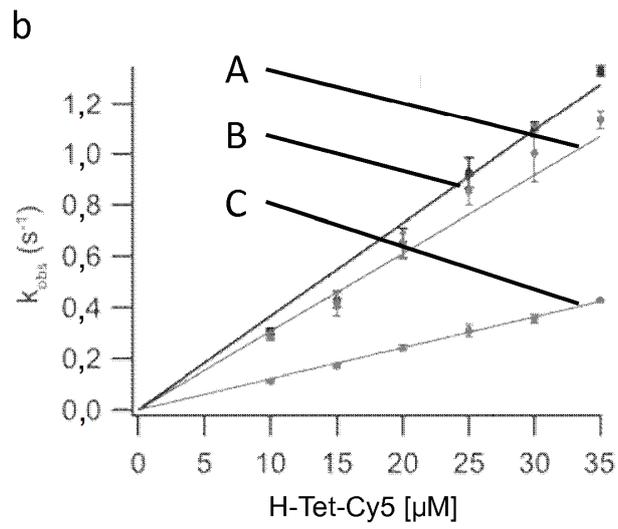
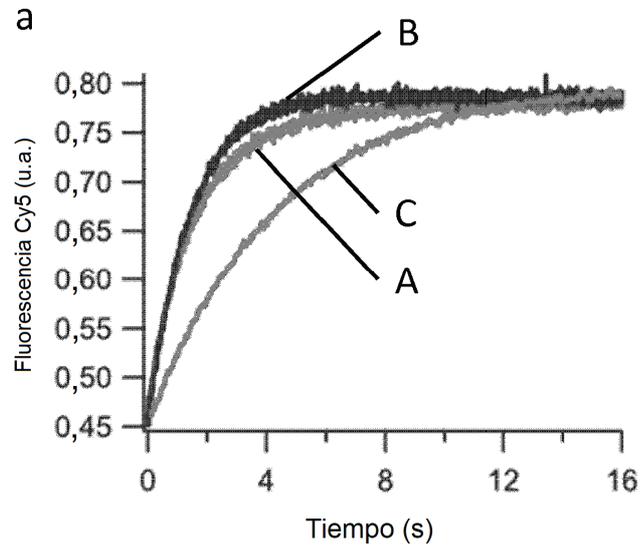
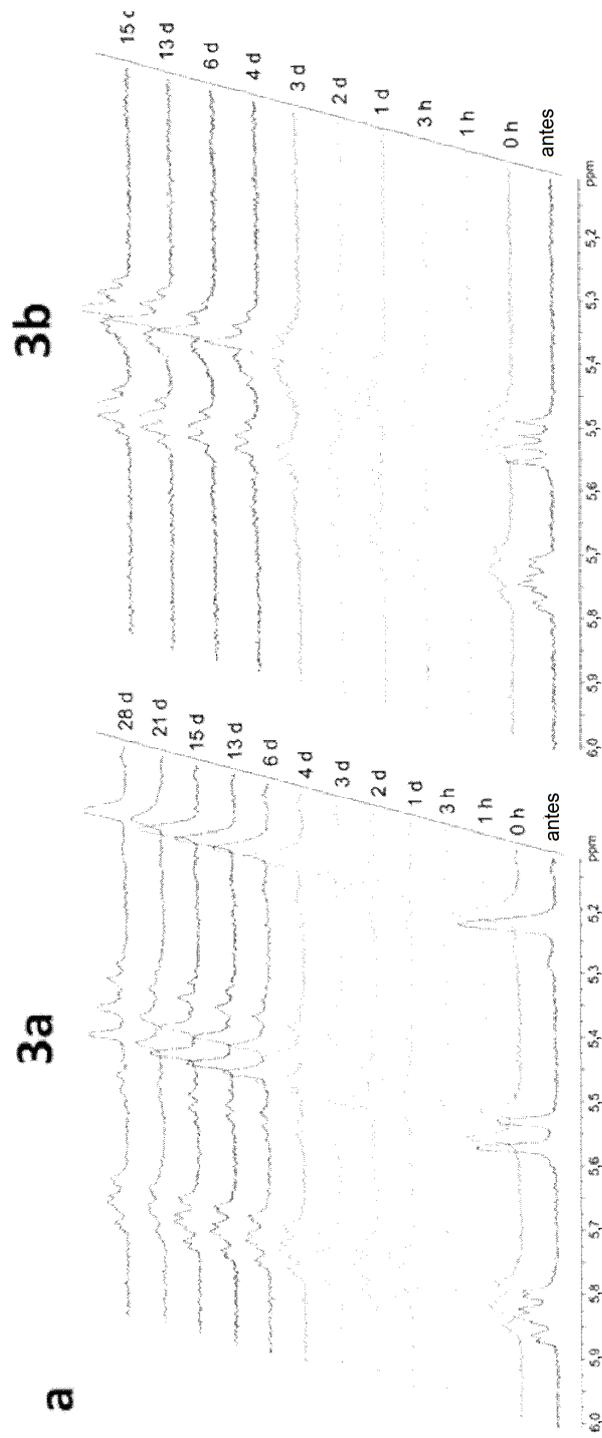


Figura 14



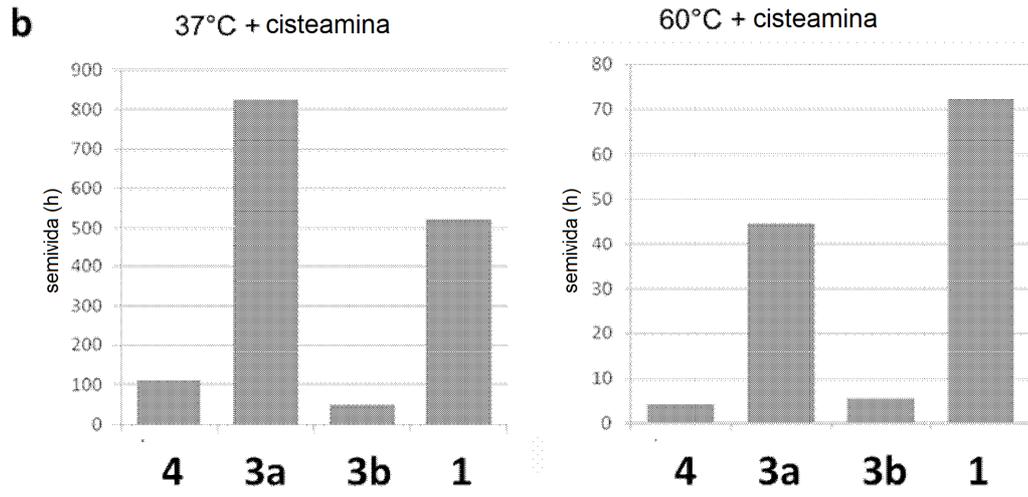


Figura 15

