

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 995 470**

51 Int. Cl.:

C07H 19/24	(2006.01)
C07H 1/00	(2006.01)
C07H 21/02	(2006.01)
C12N 15/10	(2006.01)
C12Q 1/68	(2008.01)
G01N 33/58	(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.06.2021 PCT/EP2021/067342**

87 Fecha y número de publicación internacional: **30.12.2021 WO21260107**

96 Fecha de presentación y número de la solicitud europea: **24.06.2021 E 21734353 (2)**

97 Fecha y número de publicación de la concesión europea: **25.09.2024 EP 4172170**

54 Título: **Análogos de citosina fluorescentes y su aplicación en la transcripción y la traducción**

30 Prioridad:

25.06.2020 GB 202009705

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

10.02.2025

73 Titular/es:

**STEALTH LABELS BIOTECH AB (100.00%)
c/o Chalmers Ventures AB Vera Sandbergs Allé 8
412 96 Gothenburg (Västra Götaland County), SE**

72 Inventor/es:

**WILHELMSSON, MARCUS;
ESBJÖRNER, ELIN y
BALADI, TOM**

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 995 470 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

Análogos de citosina fluorescentes y su aplicación en la transcripción y la traducción

Campo

5 Esta memoria descriptiva se refiere a un trifosfato de nucleobase fluorescente modificado que se puede utilizar para producir biosintéticamente ARN marcado, que a su vez es susceptible de traducción intracelular. Esto permite obtener imágenes de células vivas en las que se pueden visualizar simultáneamente el ARN mensajero marcado y su producto de traducción correspondiente (cuando se trata de una proteína de fusión fluorescente).

Antecedentes

10 El ARN juega un papel fundamental en la biología humana. Es el actor principal del dogma central de la bioquímica y un regulador crucial de la expresión genética a través de, por ejemplo, el ARN micro y pequeño interferente, así como a través de su actividad catalítica intrínseca. Por estas razones, también ha surgido como una nueva modalidad farmacológica muy prometedora y versátil: dado que las terapias de ARN tienen el potencial de modificar la función celular a nivel traduccional, pueden abrir nuevas oportunidades para abordar objetivos que antes no eran tratables.

15 Por lo tanto, un mayor conocimiento molecular y mecanicista de los procesos biológicos que involucran al ARN es vital para comprender las enfermedades y tratarlas. Por ejemplo, hay un creciente conjunto de evidencias que sugieren que la clave para liberar todo el potencial de los fármacos basados en ARN radica en comprender los procesos de captación celular y liberación endosómica (Dowdy, SF, Nat. Biotechnol. 35, 222-229, [2017]). Independientemente del mecanismo de endocitosis, el suministro de una carga de ácido nucleico al citoplasma
20 siempre depende del escape endosómico, cuya comprensión, a pesar de extensas investigaciones, sigue siendo difícil de alcanzar (Crooke, ST et al., Nat. Biotechnol. 35, 230 [2017]; Pei, D. y Buyanova, M., Bioconjugate Chem. [2018]). En este contexto, el seguimiento de los ARN endógenos y exógenos (terapéuticos) dentro de las células, incluida su translocación, localización, empalme y degradación, es de gran importancia.

25 Los avances recientes han dado como resultado el desarrollo de un amplio espectro de herramientas y sondas con las que se puede analizar y cuantificar el ARN, pero generalmente implican oligonucleótidos muy modificados con propiedades significativamente diferentes de las naturales, lo que potencialmente resulta en la pérdida de la capacidad de ser reconocidos y procesados por la maquinaria enzimática de las células. Por ejemplo, una desventaja de las tecnologías existentes basadas en fluorescencia para estudiar la localización celular del ARN es que se basan principalmente en construcciones fluorescentes externas altamente anfífilas y/o voluminosas que podrían perjudicar la motilidad y perturbar la localización del ARN y sus interacciones
30 moleculares con (por ejemplo) los componentes de la membrana. Además, la mayoría de estas tecnologías son incompatibles con la obtención de imágenes de células vivas (Li, Y., Ke, K. y Spitale, RC, Biochemistry 58, 379-386 [2019]).

35 Para superar estos problemas y proporcionar un método mejorado de investigación de los mecanismos del ARN, esta memoria descriptiva divulga un trifosfato de nucleobase modificado (compuesto (I), "tC⁰TP") que se puede utilizar para incorporar etiquetas internas y mínimamente perturbadoras ("tC⁰") en ARN mensajero funcional ("ARNm"), lo que le confiere utilidad en la obtención de imágenes de células vivas y en estudios de administración de fármacos. Una vez incorporado al ARN, la estructura de tC⁰ le permite conservar el apareamiento y apilamiento de bases, de modo que perturba mínimamente los procesos biológicos naturales (Fig. 1, donde la estructura esquemática superior es ARN marcado con tC⁰ en comparación con un ARN
40 marcado externamente convencional a continuación). Por lo tanto, este fluoróforo constituye una etiqueta alternativa de tipo nativo, abriendo nuevas posibilidades no sólo para rastrear el ácido nucleico de interés sino también para utilizar lecturas fluorescentes para obtener información detallada sobre la estructura y el comportamiento del ácido nucleico.

45 Para ejemplificar las posibles aplicaciones de tC⁰, la memoria descriptiva describe la transcripción in vitro exitosa y también efectiva en la traducción celular de un ARNm de longitud completa marcado internamente con este análogo de nucleobase fluorescente. Además, al utilizar una transcripción que codifica la proteína histona H2B fusionada a una proteína fluorescente verde ("H2B:GFP") que se localiza en el núcleo, la memoria descriptiva muestra que es posible visualizar las transcripciones de ARNm marcadas dentro de las células
50 mientras se registra simultáneamente la fluorescencia que emana de la proteína H2B:GFP expresada. Este enfoque debería aplicarse generalmente a cualquier proteína fluorescente.

55 Wilhelmsson, M. et al. Sci. Rep. 7, 2393 [2017] informa sobre la preparación de moléculas de ARN marcadas con tC⁰, pero sólo divulga oligómeros de ARN fluorescentes muy cortos preparados mediante síntesis de oligonucleótidos en fase sólida no enzimática, a diferencia de los ARN de longitud completa accesibles mediante biosíntesis. En virtud de su método de preparación, las moléculas de ARN marcadas no son susceptibles de análisis de células "vivas" de la transcripción, traducción o administración de ARN terapéuticos largos (que se basan en ARNm) y, por lo tanto, no permiten el mismo nivel de conocimiento mecanicista.

El documento WO2011/034895 se refiere a los métodos de marcado del ADN y del ARN. Se menciona un análogo de ribonucleótido fluorescente estructuralmente diferente, 1,3-diaza-2-oxofenotiazina-ribose-5'-trifosfato ("tCTP"), que se utiliza durante las reacciones de transcripción *in vitro* para preparar ARN marcado. Sin embargo, a diferencia de la presente memoria descriptiva, solo se preparó ARN marcado no codificante y no se divulga ninguna biosíntesis o visualización en células.

Los polímeros de ARN codificantes de longitud completa y marcados de forma diferente a los que se puede acceder mediante la tecnología divulgada en la presente memoria descriptiva también tienen propiedades ventajosas con respecto a los ARN marcados en el documento WO2011/034895, por ejemplo, 1) niveles mejorados de fluorescencia y fotoestabilidad de la etiqueta; 2) fidelidad de transcripción *in vitro* mejorada; y 3) niveles nativos de traducción intracelular de ARNm marcado con tC^o que da como resultado el producto proteico y la localización correctos.

En resumen, esta memoria descriptiva divulga una técnica de marcado que no solo permite la localización y el seguimiento del ARN etiquetado, sino que también facilita el análisis de la funcionalidad biológica y la eficacia de administración del ARNm, una importante modalidad futura de fármaco. Siempre que la etiqueta tC^o interna sea compatible con los procesos biológicos en los que participa el ARN y tenga un gran potencial para ser utilizada como una poderosa herramienta de obtención de imágenes en microscopía de células vivas, por ejemplo, en investigaciones detalladas de captación celular, liberación endosómica, carga y tráfico exosomal. Estos tienen un potencial significativo para dilucidar cómo funcionan estas vías de administración vitales y cómo pueden controlarse.

Resumen

Un objetivo principal de la presente memoria descriptiva es proporcionar un trifosfato de nucleobase modificado que pueda usarse para hacer que el ARN marcado sea especialmente adecuado para investigaciones mecanísticas *in vitro* e *in vivo*.

En consecuencia, esta memoria descriptiva describe, en parte, un compuesto de fórmula (I) o una sal del mismo de acuerdo con la reivindicación 1.

Esta memoria descriptiva también describe, en parte, un proceso para preparar un compuesto de fórmula (I) o una sal del mismo de acuerdo con la reivindicación 5.

Esta memoria descriptiva también describe, en parte, una composición para preparar una molécula de ARN marcada con tC^o que comprende un compuesto de fórmula (I) de acuerdo con la reivindicación 16.

Esta memoria descriptiva también describe, en parte, el uso de un compuesto de fórmula (I) o una sal del mismo para preparar enzimáticamente una molécula de ARN marcada con tC^o de acuerdo con la reivindicación 17.

Esta memoria descriptiva también describe, en parte, un proceso para preparar una molécula de ARN marcada con tC^o de acuerdo con la reivindicación 19.

Esta memoria descriptiva también describe, en parte, el uso de una molécula de ARNm marcada con tC^o para preparar una proteína codificada por el ARNm mediante traducción de acuerdo con la reivindicación 20.

Realizaciones ilustrativas

La invención detallada en esta memoria descriptiva no debe interpretarse como limitada a ninguna de las realizaciones o ejemplos citados. Otras realizaciones serán fácilmente evidentes para un lector experto en la materia.

Definiciones generales

"Un" o "uno, una" significa "al menos uno". En cualquier realización donde se utilizan "un" o "uno, una" para designar un material o elemento determinado, "un" o "uno, una" pueden significar uno. En cualquier realización donde se utilizan "un" o "uno, una" para denotar un material o elemento determinado, "un" o "uno, una" pueden significar 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 1000, 10000, 100000 o 1000000 (1 millón).

Cuando una realización incluye "una" característica X, las referencias posteriores a "la" característica X no implican que solo esté presente una de las características. En cambio, la interpretación anterior de "un" o "uno, una" continúa aplicándose, de modo que "el" también significa "al menos uno". En otras palabras, las realizaciones que comprenden "una característica X, donde la característica X es ..." deben interpretarse como "al menos una característica X, donde la al menos una característica X es ...".

"Que comprende" significa que una realización determinada puede contener otras características. Por ejemplo, en cualquier realización donde se menciona un material "que comprende" ciertos materiales o elementos, el material dado puede estar formado por al menos 10 % p/p, al menos 20 % p/p, al menos 30 % p/p o al menos 40 % p/p de los materiales o elementos (o combinación de materiales o elementos).

En cualquier realización donde se menciona "que comprende", "que comprende" también puede significar "que consiste en" (o "consiste en") o "que consiste esencialmente en" (o "consiste esencialmente en").

5 Con respecto a las realizaciones de un material, "que consiste en" o "consiste en" significa que el material o elemento está formado enteramente por el material o elemento (o combinación de materiales o elementos). En cualquier realización donde se menciona "que consiste en" o "consiste en", el material o elemento dado puede estar formado por un 100 % p/p del material o elemento.

10 Con respecto a las realizaciones de un material, "que consiste esencialmente en" o "consiste esencialmente en" significa que un material o elemento determinado consiste casi en su totalidad en ese material o elemento (o combinación de materiales o elementos). En cualquier realización donde se menciona "que consiste esencialmente en" o "consiste esencialmente en", el material o elemento dado puede estar formado por al menos 50 % p/p, al menos 60 % p/p, al menos 70 % p/p, al menos 80 % p/p, al menos 90 % p/p, al menos 95 % p/p o al menos 99 % p/p del material o elemento.

15 En cualquier realización donde se utiliza "es" o "puede ser" para definir un material o elemento, "es" o "puede ser" pueden significar que el material o elemento "consiste en" o "consiste esencialmente en" el material o elemento.

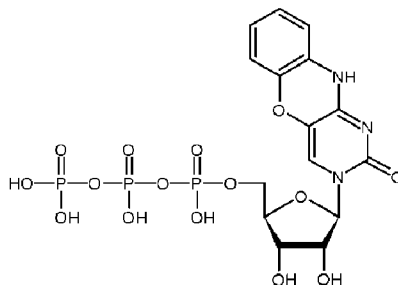
Cuando se menciona que "en algunas realizaciones ..." un determinado elemento puede estar presente, el elemento puede estar presente en una realización adecuada en cualquier parte de la memoria descriptiva, no solo en una realización adecuada en la misma sección o región textual de la memoria descriptiva.

20 Cuando se "selecciona desde" una característica de una lista, dicha característica se selecciona de una lista que consta de las alternativas especificadas (es decir, una lista de las alternativas especificadas y ninguna otra).

Las reivindicaciones son realizaciones.

Trifosfatos de nucleobases modificados

En una realización se proporciona un compuesto de fórmula (I) o una sal del mismo:



25

(I)

Los compuestos y sales descritos en esta memoria descriptiva pueden existir como una mezcla de tautómeros (isómeros estructurales resultantes de la migración de un átomo de hidrógeno que existen en equilibrio). Las realizaciones relevantes incluyen todos los tautómeros de compuestos de fórmula (I) o sales de los mismos.

30 Los átomos de los compuestos y sales descritos en esta memoria descriptiva pueden existir como sus isótopos. Las realizaciones incluyen todos los compuestos de fórmula (I) donde un átomo se reemplaza por uno o más de sus isótopos (por ejemplo, un compuesto de fórmula (I) donde uno o más átomos de carbono son un isótopo de carbono ^{11}C o ^{13}C , o donde uno o más átomos de hidrógeno son un isótopo de ^2H o ^3H).

35 Una sal adecuada de un compuesto de fórmula (I) es, por ejemplo, una sal de adición de base. Una sal de adición de base se forma poniendo el compuesto de fórmula (I) en contacto con una base orgánica o inorgánica adecuada. Se puede formar una sal de adición de base utilizando una base orgánica adecuada como una base de nitrógeno, por ejemplo, amoníaco o una trialkilamina como trietilamina. Una sal de adición de base también puede formarse, por ejemplo, utilizando una base inorgánica adecuada tal como un hidróxido de metal alcalino o de tierras raras, por ejemplo, hidróxido de potasio, hidróxido de sodio, hidróxido de magnesio o hidróxido de manganeso.

40 En una realización se proporciona un compuesto de fórmula (I) que es un ácido libre.

En una realización se proporciona un compuesto de fórmula (I) que es una sal.

En una realización, se proporciona un compuesto de fórmula (I) que es una sal de sodio, potasio, magnesio o amonio.

En una realización, se proporciona un compuesto de fórmula (I) que es una sal de sodio, potasio o amonio.

En una realización se proporciona un compuesto de fórmula (I) que es una sal de sodio o amonio.

En una realización, se proporciona un compuesto de fórmula (I) que es una sal monosódica, disódica, trisódica, tetrasódica, monoamónica, diamónica, triamónica o tetraamónica.

- 5 En una realización se proporciona un compuesto de fórmula (I) que es una sal monosódica, disódica, trisódica o tetrasódica.

En una realización se proporciona un compuesto de fórmula (I) que es una sal monosódica.

En una realización se proporciona un compuesto de fórmula (I) que es una sal disódica.

En una realización se proporciona un compuesto de fórmula (I) que es una sal trisódica.

- 10 En una realización se proporciona un compuesto de fórmula (I) que es una sal de monoamonio, diamonio, triamonio o tetraamonio.

En una realización se proporciona un compuesto de fórmula (I) que es una sal de monoamonio.

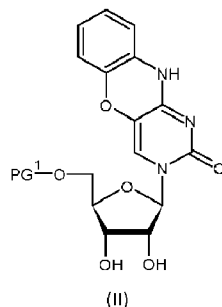
En una realización se proporciona un compuesto de fórmula (I) que es una sal de diamonio.

En una realización se proporciona un compuesto de fórmula (I) que es una sal de triamonio.

- 15 Procesos de síntesis

En una realización se proporciona un proceso para preparar un compuesto de fórmula (I) o una sal del mismo que comprende:

i. Proporcionar un compuesto de fórmula (II) o una sal del mismo:



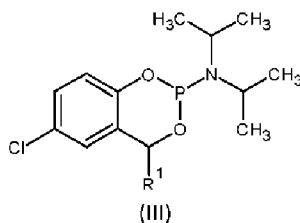
- 20 Donde PG¹ es un grupo protector adecuado;

ii. Inmovilizar el compuesto de fórmula (II) o una sal del mismo uniendo uno de sus grupos de alcohol secundario a un soporte adecuado;

iii. Proteger cualquier grupo de alcohol secundario restante con un grupo protector PG² adecuado;

iv. Eliminación del grupo protector PG¹;

- 25 v. Hacer reaccionar el grupo de alcohol primario expuesto con un compuesto de fórmula (III):



Donde R¹ se selecciona entre un grupo hidrógeno y un grupo alquilo C₁₋₃;

vi. Oxidar el compuesto de fósforo (III) resultante a un compuesto de fósforo (V);

vii. Hacer reaccionar el compuesto de fósforo (V) con un pirofosfato de tetraalquilamonio para generar un trifosfato;

viii. Eliminar el grupo protector PG²; y

5 ix. Escindir el trifosfato resultante del soporte para generar un compuesto de fórmula (I) o una sal del mismo.

10 Un grupo protector ("PG", por ejemplo, PG¹ y PG²) es cualquier grupo adecuado para proteger temporalmente un centro reactivo, por ejemplo, un grupo hidroxilo. Los grupos protectores adecuados para los centros reactivos divulgados en este documento se pueden encontrar, por ejemplo, en "Greene's Protective Groups in Organic Synthesis, Fourth Edition", Greene T. W., Wuts P. G. M.; John Wiley & Sons, Inc. 2007, doi: 10.1002/0470053488).

Un grupo "hidro" es equivalente a un átomo de hidrógeno. Los átomos con un grupo hidro unido a ellos pueden considerarse no sustituidos.

Un "grupo alquilo C₁₋₃" es un grupo alquilo saturado de cadena lineal o ramificada con el número indicado de carbonos. Los ejemplos de grupos alquilo C₁₋₃ incluyen metilo, etilo, propilo e isopropilo.

15 En la etapa iii) anterior, los alcoholes secundarios que se deben proteger pueden ser aquellos en la parte de la ribosa de la molécula.

Este proceso global es una preparación ventajosa del compuesto de fórmula (I) por varias razones:

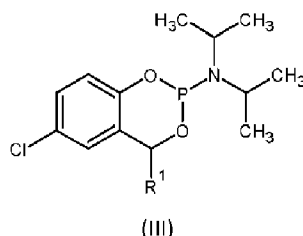
- No es necesaria ninguna protección previa de los grupos de alcohol secundario en el nucleósido protegido de partida;
- 20 - Después de la etapa 1, se puede recuperar el nucleósido que no reaccionó, lo que minimiza la pérdida de material;
- Los nucleósidos en soporte sólido se pueden almacenar hasta 3 meses sin degradarse;
- La ruta es compatible con la síntesis automatizada;
- El proceso es robusto y tiene buena reproducibilidad general;
- 25 - Se generan productos crudos de fosforilación limpios que son fáciles de purificar; y
- Proporciona altos rendimientos de fosforilación (normalmente aproximadamente 60 % a partir de la resina cargada de nucleósidos).

En algunas realizaciones R¹ puede ser un grupo hidro.

30 En algunas realizaciones R¹ puede ser un grupo alquilo C₁₋₃. Se ha observado que cuando R¹ es un grupo alquilo C₁₋₃, la preparación del reactivo de fosforamidita es más fácil y de mayor rendimiento, pero funciona al menos tan bien en la etapa v anterior como cuando R¹ es un grupo hidro.

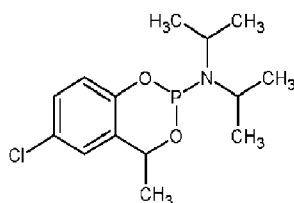
En algunas realizaciones R¹ puede ser metilo.

En una realización se proporciona un compuesto de fórmula (III):



35 Donde R¹ es un grupo alquilo C₁₋₃.

En una realización se proporciona un compuesto de fórmula (IIIa):



(IIIa)

En algunas realizaciones, el soporte puede ser un polímero sólido.

En algunas realizaciones, el soporte puede ser un polímero sólido seleccionado entre vidrio de porosidad controlada y poliestireno.

5 En algunas realizaciones el soporte puede ser poliestireno.

En algunas realizaciones, el soporte puede ser vidrio de porosidad controlada.

En algunas realizaciones, el soporte puede estar funcionalizado con un grupo amino primario. Éste puede formar el punto reactivo de unión al soporte.

10 En algunas realizaciones, el soporte puede ser vidrio de porosidad controlada funcionalizado con un grupo amino primario (por ejemplo, Amino-SynBase™).

En algunas realizaciones PG¹ puede seleccionarse entre tritilo, dimetoxitritilo y trimetoxitritilo.

En algunas realizaciones PG² puede seleccionarse entre acetilo, benzoilo, 2,2,2-tricloroetilcarbonilo, parametoxibencilo, metilo, tetrahidropiranilo, trietilsililo, triisopropilsililo, trimetilsililo, terbutildimetilsililo y metoxietilo.

15 En algunas realizaciones PG² puede ser acetilo. Cuando una molécula inmovilizada es lábil a la base, esto permite una síntesis eficiente en la que se elimina el grupo PG² y la escisión de la resina se puede lograr en una sola etapa.

En algunas realizaciones PG¹ puede ser dimetoxitritilo y PG² puede ser acetilo.

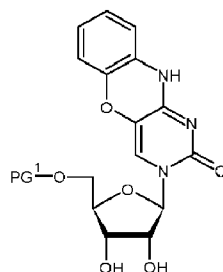
20 En algunas realizaciones, la inmovilización del compuesto de fórmula (II) en la etapa i) puede ocurrir principalmente en la posición 2'-hidroxi.

Cuando la inmovilización ocurre principalmente en la posición 2'-hidroxi, esta puede ser >50 %, >60 %, >70 %, >80 %, >90 % o 100 % de la inmovilización total (es decir, la unión covalente total de ambos grupos hidroxilo secundarios al soporte).

En algunas realizaciones, el pirofosfato de tetraalquilamonio puede ser pirofosfato de tetrabutilamonio.

25 En una realización se proporciona un proceso para preparar un compuesto de fórmula (I) o una sal del mismo que comprende:

i. Proporcionar un compuesto de fórmula (II) o una sal del mismo:



(II)

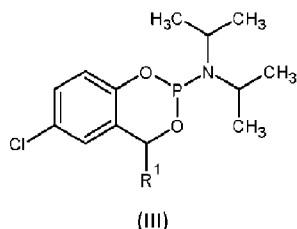
Donde PG¹ se selecciona entre tritilo, dimetoxitritilo y trimetoxitritilo;

30 ii. Inmovilizar el compuesto de fórmula (II) o una sal del mismo mediante la unión de uno de sus grupos de alcohol secundario a un soporte de vidrio de porosidad controlada;

iii. Proteger cualquier grupo de alcohol secundario restante con un grupo protector PG² seleccionado entre acetilo, benzoilo, 2,2,2-tricloroetilcarbonilo, parametoxibencilo, metilo, tetrahidropiraniolo, trietilsililo, trisopropilsililo, trimetilsililo, tercbutildimetilsililo y metoxietilo;

iv. Eliminación del grupo protector PG¹;

5 v. Hacer reaccionar el grupo de alcohol primario expuesto con un compuesto de fórmula (III):



Donde R¹ es un grupo alquilo C₁₋₃;

vi. Oxidar el compuesto de fósforo (III) resultante a un compuesto de fósforo (V);

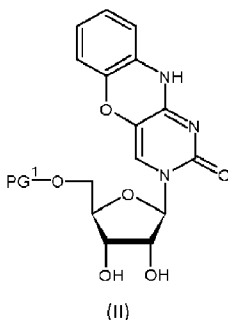
vii. Hacer reaccionar el compuesto de fósforo (V) con pirofosfato de tetrabutilamonio para generar un trifosfato;

10 viii. Eliminar el grupo protector PG²; y

ix. Escindir el trifosfato resultante del soporte para generar un compuesto de fórmula (I) o una sal del mismo.

En una realización se proporciona un proceso para preparar un compuesto de fórmula (I) o una sal del mismo que comprende:

i. Proporcionar un compuesto de fórmula (II) o una sal del mismo:



15

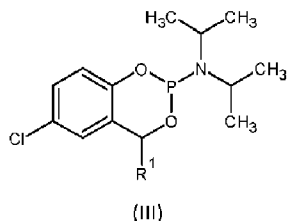
Donde PG¹ es dimetoxitritilo;

ii. Inmovilizar el compuesto de fórmula (II) o una sal del mismo mediante la unión de uno de sus grupos de alcohol secundario a un soporte de vidrio de porosidad controlada;

iii. Proteger cualquier grupo de alcohol secundario restante con un grupo protector PG² que es acetilo;

20 iv. Eliminar el grupo protector PG¹;

v. Hacer reaccionar el grupo de alcohol primario expuesto con un compuesto de fórmula (III):



Donde R¹ es un grupo metilo;

vi. Oxidar el compuesto de fósforo (III) resultante a un compuesto de fósforo (V);

25 vii. Hacer reaccionar el compuesto de fósforo (V) con pirofosfato de tetrabutilamonio para generar un trifosfato;

viii. Eliminar el grupo protector PG²; y

ix. Escindir el trifosfato resultante del soporte para generar un compuesto de fórmula (I) o una sal del mismo.

Las condiciones y reactivos adecuados para llevar a cabo cada una de las etapas i) a ix) anteriores son conocidos por el experto en la materia o se pueden encontrar en la Descripción detallada.

5 En algunas realizaciones, la inmovilización del compuesto de fórmula (II) o una sal del mismo en la etapa ii) se puede lograr mediante un reactivo de acoplamiento (por ejemplo, anhídrido succínico catalizado por dimetilaminopiridina cuando el soporte está funcionalizado con un grupo amino primario).

En algunas realizaciones, la reacción del grupo de alcohol primario expuesto con un compuesto de fórmula (III) se puede lograr utilizando un activador (por ejemplo, activador BTT o Activador 42[®]).

10 En algunas realizaciones, el compuesto de fósforo (III) en la etapa vi) puede oxidarse a un compuesto de fósforo (V) usando piridina acuosa y yodo.

15 En algunas realizaciones, la escisión del trifosfato del soporte se puede lograr utilizando condiciones básicas (por ejemplo, mediante tratamiento con AMA). Cuando hay un soporte lábil a las bases y se elige un grupo protector lábil a las bases para PG², el uso de estas condiciones permite la desprotección y escisión simultáneas.

Síntesis de ARN marcado

En una realización se proporciona una composición para preparar una molécula de ARN marcada con tC[○] que comprende un compuesto de fórmula (I) y un trifosfato de ribonucleótido natural.

20 Un "trifosfato de ribonucleótido natural" comprende el ribonucleósido natural apropiado con un grupo trifosfato unido a la posición hidroxilo 5'. Es equivalente a un trifosfato de ribonucleósido natural. En algunas realizaciones, se puede seleccionar un trifosfato de ribonucleótido natural entre 5'-trifosfato de citidina, 5'-trifosfato de uridina, 5'-trifosfato de adenosina y 5'-trifosfato de guanidina. Una composición de trifosfatos de ribonucleótidos naturales (es decir, una que comprende un trifosfato de ribonucleótidos como se define en este documento) puede comprender combinaciones de cantidades variables de estos bloques de construcción, en cantidades
25 suficientes para construir la molécula de ARN objetivo (por ejemplo, como se proporciona en la mezcla NTP).

En una realización se proporciona el uso de un compuesto de fórmula (I) o una sal del mismo para preparar enzimáticamente una molécula de ARN marcada con tC[○].

30 Una molécula de ARN marcada con tC[○] comprende al menos un residuo de tC[○], pero por lo demás es similar a la molécula de ARN natural (es decir, una con un residuo de citosina no modificado en la misma ubicación que el residuo de tC[○]).

En algunas realizaciones, una molécula de ARN marcada con tC[○] puede comprender >10 %, >20 %, >30 %, >40 %, >50 %, >60 %, >70 %, >80 %, >90 % o 100 % de residuos de tC[○] en lugar de residuos de citosina no modificados.

35 En algunas realizaciones, una molécula de ARN marcada con tC[○] puede comprender entre un 10 % - 20 %, un 10 % - 30 %, un 10 % - 40 %, un 20 % - 50 %, un 30 % - 60 %, un 40 % - 70 %, un 50 % - 80 % o un 50 % - 90 % de residuos de tC[○] en lugar de residuos de citosina no modificados.

40 En una realización se proporciona un proceso para preparar una molécula de ARN marcada con tC[○] que comprende proporcionar una plantilla de ADN a una composición que comprende un compuesto de fórmula (I) y un trifosfato de ribonucleótido natural (por ejemplo, una combinación de cantidades variables de 5'-trifosfato de citidina, 5'-trifosfato de uridina, 5'-trifosfato de adenosina y/o 5'-trifosfato de guanidina en cantidades suficientes para construir la molécula de ARN objetivo, por ejemplo como se proporciona en una mezcla de NTP), y luego tratar la mezcla resultante con una ARN polimerasa.

En algunas realizaciones, la molécula de ARN marcada con tC[○] puede ser un ARNm marcado con tC[○].

45 En algunas realizaciones, la molécula de ARN marcada con tC[○] puede codificar una proteína fusionada a una proteína fluorescente. Los ejemplos de proteínas fluorescentes fusionables incluyen la proteína fluorescente verde (GFP) y las proteínas de la familia mFruit. Cuando la proteína objetivo es fluorescente (ya sea de manera inherente o debido a una etiqueta), es posible visualizar simultáneamente tanto la molécula de ARN marcada como la proteína que se está utilizando para sintetizar, lo que proporciona un mayor grado de conocimiento mecanicista.

50 En algunas realizaciones, la molécula de ARN marcada con tC[○] puede codificar una proteína seleccionada entre H2B, calmodulina, H2B:GFP y calmodulina-3:GFP.

La terminología ".GFP" significa que la proteína objetivo que precede a los dos puntos está fusionada a una proteína de la familia de la proteína fluorescente verde (GFP).

En algunas realizaciones, la molécula de ARN marcada con tC^o puede codificar H2B:GFP.

En algunas realizaciones, la molécula de ARN marcada con tC^o puede codificar calmodulina-3.

5 En algunas realizaciones, la ARN polimerasa puede seleccionarse entre la polimerasa T7 y la polimerasa SP6.

En algunas realizaciones, un proceso para preparar una molécula de ARN marcada con tC^o puede llevarse a cabo en presencia de un tampón de transcripción (por ejemplo, tampón de transcripción 5X), sal de magnesio (por ejemplo, cloruro de magnesio (II)) y/o un inhibidor de la RNasa (por ejemplo, Ribolock).

10 En algunas realizaciones, un proceso para preparar una molécula de ARN marcada con tC^o puede llevarse a cabo en presencia de un tampón de transcripción (por ejemplo, tampón de transcripción 5X), sal de magnesio (por ejemplo, cloruro de magnesio (II)) y un inhibidor de la RNasa (por ejemplo, Ribolock). En algunas realizaciones, un proceso para preparar una molécula de ARN marcada con tC^o puede llevarse a cabo sustancialmente como se describe en la sección experimental (por ejemplo, como se detalla en la sección titulada "Transcripción y purificación de ARN de H2B:GFP").

15 En una realización se proporciona un kit para preparar una molécula de ARN marcada con tC^o que comprende:

- i. Un compuesto de fórmula (I);
- ii. Una composición de trifosfatos de ribonucleótidos naturales;
- iii. Una ARN polimerasa; opcionalmente
- iv. Una plantilla de ADN; y opcionalmente

20 v. Instrucciones de uso.

Traducción de ARN marcado

En una realización se proporciona el uso de una molécula de ARNm marcada con tC^o para preparar una proteína codificada por el ARNm mediante traducción.

25 "Traducción" se refiere al proceso biológico central mediante el cual el ARNm se decodifica en un ribosoma para producir un polipéptido específico, que puede plegarse en una proteína activa antes de realizar sus funciones en una célula.

30 En una realización se proporciona el uso de una molécula de ARNm marcada con tC^o para preparar una proteína codificada por el ARNm mediante traducción *in vitro* (por ejemplo, sustancialmente como se describe en la parte de la sección experimental (por ejemplo, como se detalla bajo el encabezado "traducción libre de células").

En algunas realizaciones, la traducción *in vitro* se puede realizar utilizando lisados bacterianos de *E. coli* y/o sistema mini de expresión libre de células Expressway[®].

35 En una realización se proporciona el uso de una molécula de ARNm marcada con tC^o para preparar una proteína codificada por el ARNm mediante la traducción celular (por ejemplo, sustancialmente como se describe en las partes de la sección experimental (por ejemplo, como se detalla bajo los encabezados "cultivo celular" y "electroporación o transfección química").

En algunas realizaciones, la traducción intracelular puede realizarse en células de neuroblastoma humano (por ejemplo, células SH-SY5Y).

En una realización se proporciona la traducción de una molécula de ARN marcada con tC^o en una proteína.

40 En una realización se proporciona la traducción *in vitro* de una molécula de ARN marcada con tC^o en una proteína.

En una realización se proporciona la traducción en la célula de una molécula de ARN marcada con tC^o en una proteína.

45 En algunas realizaciones, la proteína codificada puede fusionarse a una proteína fluorescente (por ejemplo, una proteína de la familia GFP o mFruit). Cuando este es el caso, es posible visualizar simultáneamente tanto la molécula de ARN marcada como la proteína que se está utilizando para la síntesis, lo que proporciona un mayor grado de conocimiento mecanicista.

En algunas realizaciones, el ARNm marcado con tC° y la proteína codificada pueden analizarse simultáneamente en el espacio y el tiempo utilizando microscopía confocal (por ejemplo, microscopía confocal de fluorescencia). Esta es una forma conveniente de monitorear simultáneamente un ARN y una proteína marcados, y el ARN que contiene un análogo de base fluorescente nunca antes se ha utilizado en dicha visualización de células vivas.

Figuras

Figura 1: Esquema que muestra ARN marcado con tC° mínimamente perturbado en comparación con un ARN común marcado externamente.

Figura 2: Esquema de síntesis básico para la preparación del compuesto (I).

Figura 3: Incorporación de tC° en ARNm de longitud completa mediante transcripción *in vitro* asistida por la ARN polimerasa T7. Geles de agarosa blanqueadores desnaturizantes que muestran transcripciones de ARN formadas en cinco diferentes relaciones de TP/CTP de tC° (0-100 %). Visualización directa de fluorescencia de tC° (a) y después de la tinción con bromuro de etidio (b). Las muestras de ARN se desnaturalizaron por calor (65 °C durante 5 minutos, 1.5 % de blanqueador en el gel) antes de la carga. (c) Las mismas transcripciones de ARN tras una desnaturalización más severa (70 °C durante 10 min, 2 % de blanqueador en el gel). Se utilizó la escalera de ARN de alto rango RiboRuler.

Figura 4: Incorporación de tC° en ARNm de longitud completa mediante transcripción *in vitro* asistida por la ARN polimerasa SP6 y T7, (a) Geles de agarosa blanqueadores desnaturizantes que muestran transcripciones de ARN formadas en cinco diferentes relaciones de TP/CTP de tC° (0-100 %). El ARN producido se visualizó directamente mediante fluorescencia de tC° (imagen izquierda) y después de la tinción con bromuro de etidio (imagen derecha). Las muestras de ARN se desnaturalizaron por calor (65 °C durante 5 min, 1.5 % de blanqueador en el gel) antes de cargarlas en el gel. (b) Geles blanqueadores desnaturizantes de las mismas transcripciones de ARN que en (A) pero en condiciones desnaturizantes más fuertes (70 °C durante 10 min, 2 % de blanqueador en el gel). Se utilizó la escalera de ARN de alto rango RiboRuler.

Figura 5: Caracterización espectroscópica de transcripciones de ARN modificadas con tC° sintetizado *in vitro*. Se realizaron cuatro reacciones cargadas con diferentes fracciones molares de $TC^{\circ}TP + CTP$ en el grupo total de trifosfato de citosina ($TC^{\circ}TP + CTP$). Las transcripciones del producto se purificaron para lavar los trifosfatos que no reaccionaron antes de la caracterización. Todas las reacciones se realizaron como duplicados independientes y los resultados se presentan como la media \pm desviación estándar. a) Espectros de absorción UV-vis normalizados a $A = 1$ en la banda de ARN, aproximadamente 260 nm con mayor absorción de tC° centrada en 360 nm que crece con una relación creciente de tC° a C. Recuadro: Absorción de $TC^{\circ}TP$ normalizada a $A = 1$ en la banda $\lambda_{m\acute{a}x}$ de tC° (360 nm). b) Barras simples: Fracción de tC° incorporado (en relación con la cantidad total de citosinas incorporadas, es decir, $tC^{\circ} + C$) en las transcripciones. Barras a cuadros: Relación de las constantes de velocidad de reacción de primer orden para CTP vs. El consumo de $TC^{\circ}TP$. c) Líneas continuas: Espectros de absorción UV-vis (normalizados a $A = 1$ en la banda de ARN, aproximadamente 260 nm) que muestra la banda de tC° centrada en 368-369 nm. Líneas discontinuas: Espectros de emisión normalizados a $I = 1$ a $\lambda_{m\acute{a}x}$ (457 nm y 459 nm para la transcripción del 25 % y 100 %, respectivamente). Para mayor claridad, se omitieron los espectros de emisión para las reacciones del 50 % y 75 %. d) Barras simples: Rendimientos cuánticos de fluorescencia. Barras rayadas: duración de la fluorescencia.

Figura 6: Traducción libre de células de calmodulina-3. a) Tinción de Coomassie y b) transferencia Western (WB) de las reacciones de traducción *in vitro*. NTC: sin control de plantilla; +: kit de control de ADN de plantilla. Se utilizó el sistema escalera de proteínas previamente tenidas PageRuler. c) Cuantificación por WB y análisis densitométrico (media de 4 réplicas).

Figura 7: Eficiencia de la traducción de construcciones de ARN modificadas en células humanas y validación de tC° como sonda de seguimiento intracelular. La proteína codificada H2B:GFP se observó mediante microscopía confocal y se cuantificó mediante citometría de flujo para cada una de las construcciones de ARN que incorporaban tC° . Imágenes representativas (ampliadas 3x, barras de escala: 10 μ m), diagramas de dispersión e histogramas (muestran la distribución de la señal en células vivas individuales (a, b) 24 h después de la electroporación o (c) 48 h después de la transfección química. Los diagramas de caja muestran las intensidades de fluorescencia media de GFP (MFI GFP) hasta 72 h a partir de 3 experimentos independientes realizados por triplicado. (d) Las células que sobreexpresaban mRFP-Rab5 (biomarcador de endosoma temprano) se transfectaron con 75 % de tC° ARNm y seguimiento a lo largo del tiempo para validar tC° como sonda de seguimiento intracelular (flechas blancas) que no altera la traducción, barras de escala: 10 μ m. (e) Las células se analizaron 24 h después de la electroporación o después de la transfección con ARNm codificantes de eGFP no marcados (NL) o marcados con Cyanine5 (Cy5) (TriLink®), barras de escala: 10 μ m. (f) El impacto de la incorporación de tC° o Cy5 se expresó como la relación entre MFI GFP con respecto al ARN no marcado para todas las construcciones.

Figura 8: Eficiencia de la traducción de las construcciones de ARN modificadas en células humanas y evaluación de la citotoxicidad. Imágenes confocales representativas (vista grande, barra de escala: 10 μm) de construcciones de ARN-tC^o y ARNm de TriLink[®] transfectadas por (a, e) electroporación o (b, f) transfección química. (c) Porcentajes de células positivas para H2B:GFP a las 24 h, 48 h y 72 h después de la transfección con construcciones de ARN-tC^o. (d) Histograma representativo de la distribución de la señal de GFP en células vivas individuales 48 h después de la transfección química. La evaluación de la citotoxicidad se realizó 24 h (g) después de la electroporación o (h) después de la transfección química utilizando el ensayo de integridad de la membrana celular LDH.

Descripción detallada

10 Ejemplo 1: Síntesis de trifosfatos de nucleobases modificados

El compuesto (I) puede prepararse de acuerdo con el esquema que se muestra en la Figura 2. A menos que se indique lo contrario, los reactivos estaban disponibles comercialmente y se utilizaron sin purificación adicional. Los siguientes reactivos utilizados para la trifosforilación fueron adquiridos a través de Sigma-Aldrich: solución de desbloqueo con DCA para ÅKTA , CAP A para ÅKTA , CAP B1 y B2 para ÅKTA , Activador de BTT. Los espectros de RMN de ¹H (500 MHz) y ¹³C (126 MHz) se registraron a 300 K en un sistema Bruker de 500 MHz equipado con un CryoProbe. Los espectros de RMN de ³¹P (202 MHz) se registraron a 300 K en un sistema Bruker de 500 MHz. Todos los cambios se registran en ppm en relación con el disolvente deuterado (DMSO-d₆, CDCl₃ o D₂O).

20 3-((2R,3R,4S,5R)-5-((bis(4-metoxifenil)(fenil)metoxi)metil)-3,4-dihidroxitetrahidrofuran-2-il)-3H-benzo[b]pirimidio[4,5-e][1,4]oxazin-2(10H)-ona 1

El compuesto 1 se preparó de acuerdo con la literatura (Füchtbauer, AF et al., Sci. Rep. 7, 2393 [2017])

MS (ESI-) [M-H]⁻ = 634,5. RMN de ¹H (500 MHz, DMSO-d₆) δ 10.61 (bs, 1H), 7.42 (d, J = 7.7 Hz, 2H), 7.27 - 7.35 (m, 7H), 7.22 (t, J = 7.1 Hz, 1H), 6.90 (dd, J = 8.6, 4.2 Hz, 4H), 6.75 - 6.87 (m, 3H), 6.46 (d, J = 7.8 Hz, 1H), 5.71 (d, J = 3.6 Hz, 1H), 5.49 (bs, 1H), 5.18 (bs, 1H), 4.08 (d, J = 5.3 Hz, 1H), 4.04 (s, 1H), 3.94 (s, 1H), 3.71 (s, 3H), 3.70 (s, 3H), 3.29 (d, J = 4.8 Hz, 1H), 3.16 (d, J = 9.1 Hz, 1H).

Soporte sólido de CPG 3

Amino-SynBase[™] CPG 500/110 (LCAA) de LinkTech (Nu. 1397-C025, 1 g, 0.08 mmol) se activó mediante agitación en ácido tricloroacético al 3 % en DCE (8 mL, 0.08 mmol) durante 18 h. A continuación, el soporte activado se filtró y se lavó con trietilamina:diisopropiltilamina 9:1 (20 mL), diclorometano (20 mL) y éter dietílico (20 mL). El soporte activado se secó al vacío durante 2 días antes de su uso. Posteriormente, el soporte (1 g, 0.08 mmol), anhídrido succínico (0.345 g, 3.44 mmol) y N,N-dimetilpiridin-4-amina (0.070 g, 0.57 mmol) se suspendieron en piridina seca (3 mL) bajo atmósfera de N₂. A continuación, la mezcla de reacción se agitó suavemente a temperatura ambiente durante 4 h. Después de 4 h, se filtró el disolvente y se lavó el soporte sucesivamente con piridina (20 mL), diclorometano (20 mL), éter dietílico (20 mL) y se secó al aire. La prueba de ninhidrina negativa en una pequeña porción de soporte demostró una succinilación completa. El CPG succinilado podría posteriormente conservarse a temperatura ambiente durante varios meses.

(2R,3R,4R,5R)-5-((bis(4-metoxifenil)(fenil)metoxi)metil)-4-hidroxi-2-(2-oxo-2,10-dihidro-3H-benzo[b]pirimidio [4,5-e][1,4]oxazin-3-il)tetrahidrofuran-3-ilo 4 soportado por CPG

En una jeringa de 10 mL con filtro de PTFE se suspendieron el soporte succinilado 3 (1.420 g, 82 $\mu\text{mol/g}$, 0.12 mmol), DMAP (0.028 g, 0.23 mmol), DIC (719 μL , 4.64 mmol), 1 (0.076 g, 0.12 mmol) y trietilamina (49 μL , 0.35 mmol) en piridina (5 mL). La mezcla se agitó suavemente durante 18 h a temperatura ambiente. Después de 18 h, se purgó la jeringa y se lavó el soporte con piridina (5 mL), diclorometano (5 mL) y éter dietílico. Posteriormente, en la misma jeringa, se añadieron al soporte DMAP (0.028 g, 0.23 mmol), diisopropilcarbodiimida (719 μL , 4.64 mmol), trietilamina (49 μL , 0.35 mmol) y 2,3,4,5,6-pentaclorofenol (0.309 g, 1.16 mmol) y se suspendieron en piridina (4 mL). La mezcla se agitó suavemente durante 4 h a temperatura ambiente antes de agregar una solución de piperidina (2 mL, 20 % en DMF - para proteger los ácidos carboxílicos que no reaccionaron en el soporte) durante 1 minuto (un tiempo de exposición más prolongado reducirá la carga ya que la piperidina escinde los enlaces éster con el nucleósido), luego se lavó rápidamente con DMF (3 x 5 mL), diclorometano (5 mL) y éter dietílico (5 mL). Finalmente, la resina se agitó en una mezcla CAP A + CAP B (50/50 v/v) durante 2 horas bajo atmósfera de argón (carga final: 13 $\mu\text{mol/g}$ - determinado mediante la lectura de la densidad óptica de una solución de DMT escindida de una cantidad pesada de soporte - E=70000 M⁻¹.cm⁻¹ a 498 nm). La carga final se puede aumentar realizando un segundo acoplamiento con 1 en las mismas condiciones antes de la protección (carga típica después del segundo acoplamiento 20-25 $\mu\text{mol/g}$). Concentrar la mezcla de reacción y lavar el residuo varias veces con agua y éter dietílico permite recuperar casi el 85 % del nucleósido 1 sin reaccionar.

6-cloro-N,N-diisopropil-4H-benzo[d][1,3,2]dioxafosfinin-2-amina 5

El compuesto 5 se preparó de acuerdo con la literatura (Ducho, C. et al., J. Med. Chem. 50, 1335-1346 [2007]). Brevemente, el ácido 5-clorosalicílico se redujo con LAH (0.5 equiv.) a -20 °C y el alcohol 5-clorosalicílico resultante se ciclizó en 2,6-dicloro-4H-benzo[d][1,3,2]dioxafosfinina usando PCl₃ (1.2 equiv.) y trietilamina (2.3 equiv.) a -20 °C bajo atmósfera de argón. La baja temperatura y el uso de trietilamina como base fueron decisivos para evitar un reordenamiento rápido y cuantitativo de Arbuzov del producto deseado en el más estable 2-óxido de 2,5-dicloro-3H-benzo[d][1,2]oxafosfol. La 2,6-dicloro-4H-benzo[d][1,3,2]dioxafosfinina cruda se trató posteriormente con diisopropilamina (3 equiv.) durante 2 h a temperatura ambiente. Luego la mezcla se filtró bajo atmósfera de argón, se concentró a sequedad y se absorbió en diisopropilamina al 20 % en heptano. La filtración rápida sobre un pequeño tapón de gel de sílice permitió obtener el compuesto 5 deseado en forma de aceite incoloro, que cristalizó con el tiempo a -20 °C. Cualquier intento de realizar una cromatografía en columna más exhaustiva sobre el compuesto 5 conduciría a un reordenamiento cuantitativo de Arbuzov.

RMN de ¹H (500 MHz, DMSO-d₆) δ = 7.23 (dd, J = 8.6, 2.6 Hz, 1H), 7.20 (d, J = 2.4 Hz, 1H), 6.92 (d, J = 8.6 Hz, 1H), 5.06 (dd, J = 14.7, 5.2 Hz, 1H), 4.89 (dd, J = 19.6, 14.8 Hz, 1H), 3.53 - 3.63 (m, 2H), 1.15 - 1.19 (dd, J = 8.0, 7.0 Hz, 12H). RMN de ³¹P (202 MHz, DMSO-d₆) δ = 136.00 (s, 1P).

Dihidrógeno difosfato de *bis*(tetrabutilamonio) 6

El compuesto 6 se preparó de acuerdo con la literatura (Warnecke, S. y Meier, C., J. Org. Chem. 74, 3024-3030 [2009]).

RMN de ¹H (500 MHz, D₂O) δ 3.04 - 3.13 (m, 16H), 1.53 (bs, 16H), 1.24 (h, J = 7.3, 16H), 0.83 (t, J = 7.4, 24H). RMN de ³¹P (202 MHz, D₂O) δ = -10.78 (s, 2P).

Trifosfato de ((2R,3S,4R,5R)-3,4-dihidroxi-5-(2-oxo-2,10-dihidro-3H-benzo[b]pirimido[4,5-e][1,4]oxazin-3-il) tetrahydrofuran-2-il)metilo 7

Las reacciones se realizaron en una jeringa de 5 mL con filtro de PTFE cargado con 4 (800 mg, 0.016 mmol) bajo una atmósfera de argón y con agitación.

Las etapas se realizaron de la siguiente manera:

a. Eliminación de 5'-DMT: El soporte se lavó con un flujo de solución de desbloqueo con DCA hasta que el filtrado fue incoloro, luego se lavó con ACN (5 × 5 mL).

b. Acoplamiento: La N,N-diisopropil-4H-benzo[d][1,3,2]dioxafosfinin-2-amina 5 (345 mg, 1.36 mmol) se disolvió en 4.8 mL de ACN y se hizo reaccionar en porciones con el soporte (3 acoplamientos iguales con tiempos de reacción de 60 s - 60 s - 90 s respectivamente). A cada acoplamiento también se le añadió un activador (por ejemplo, activador de BTT (2.4 mL) o Activador 42). Posteriormente se lavó el soporte con ACN (3 × 5 mL).

c. Oxidación: piridina/agua/yodo (9/1/12.7 v/v/v, 5 mL) durante 45 s, seguido de lavado con ACN (3 × 5 mL) y secado del soporte en flujo de argón.

d. Trifosforilación: Dos inyecciones de dihidrógeno difosfato de *bis*(tetrabutilamonio) 6 (0.5 M, 5 mL) durante 15 minutos y 18 horas, respectivamente. Posteriormente, el soporte se enjuagó con DMF (5 mL), agua (3 × 5 mL), ACN (5 mL) y luego se secó en un flujo de argón.

e. Escisión y purificación: La escisión del trifosfato se realizó en 2 h a temperatura ambiente con AMA (mezcla 50/50 v/v de NH₄OH acuoso al 23 % y metilamina acuosa al 40 %, 5 mL). Después de 2 horas, el filtrado de AMA se purgó en un matraz de fondo redondo y el soporte se enjuagó 3 veces con solución acuosa de NH₄OH al 23 %. Después de la liofilización de la mezcla, se realizó su purificación por HPLC (columna Waters Acquity HSS T3, 2.1 × 50 mm, 0.4 mL/min, NH₄OAc 50 mM del 2 a 99 % en agua 80:20 de EtOH) para obtener el compuesto 7 (5.6 mg, 62.0 % determinado a partir de la absorbancia UV) como un sólido de color amarillo claro (sal de amonio). El mismo nivel de pureza se podría lograr con HPLC de intercambio iónico utilizando una columna Dionex DNAPac PA100 semipreparativa (9 × 250 mm) en un sistema de HPLC ÄKTA pure 25 utilizando un gradiente de agua con NH₄HCO₄ 1 M al 20 % (pH 7.8) en 30 min a un caudal de 4 mL/min.

HRMS (ESI-TOF) m/z calculado para C₁₅H₁₈N₃O₁₅P₃ [M+H]⁺: 574.0029, encontrado: 574.0013; m/z calculado para C₁₅H₁₈N₃O₁₅P₃ [M-H]⁻: 571.9878, encontrado: 571.9872. RMN de ¹H (500 MHz, D₂O) δ 7.44 (s, 1H), 6.84 - 6.94 (m, 3H), 6.79 (dd, J = 7.5, 1.7 Hz, 1H), 5.91 (d, J = 4.9 Hz, 1H), 4.36 (t, J = 4.8 Hz, 1H), 4.29 (t, J = 5.1 Hz, 1H), 4.25 (d, J = 4.1 Hz, 3H). RMN de ¹³C (126 MHz, D₂O) δ 155.8, 154.8, 142.4, 129.4, 124.9, 124.3, 122.3, 116.6, 88.8, 82.8, 73.4, 69.7, 64.5. RMN de ³¹P (202 MHz, D₂O) δ -10.89 (d, J = 18.5 Hz, 1P), -11.46 (d, J = 19.7 Hz, 1P), -23.21 (t, J = 19 Hz, 1P).

El compuesto (I) también se puede preparar mediante una ruta ligeramente modificada en la que la etapa de acoplamiento (b anterior) se lleva a cabo con una fosforamida modificada tal como 6-cloro-N,N-diisopropil-4-metil-4H-benzo[d][1,3,2]dioxafosfinin-2-amina 8 (compuesto (IIIa) anterior). Se ha descubierto que este reactivo

se prepara más fácilmente y el compuesto 8 se puede obtener con un rendimiento del 60 % en comparación con alrededor del 3-10 % para la preparación del compuesto 5 en las condiciones de esta memoria descriptiva.

6-cloro-*N,N*-diisopropil-4-metil-4*H*-benzo[d][1,3,2]dioxafosfinin-2-amina 8

5 Se hizo reaccionar 5-cloro-2-hidroxibenzaldehído con bromuro de metilmagnesio (2.5 equiv.) a -20 °C y el 4-cloro-2-(1-hidroxietil)fenol resultante se ciclizó en 2,6-dicloro-4-metil-4*H*-benzo[d][1,3,2]dioxafosfinina usando PCl_3 (1.2 equiv.) y trietilamina (2.3 equiv.) a -20 °C bajo atmósfera de argón. La 2,6-dicloro-4-metil-4*H*-benzo[d][1,3,2]dioxafosfinina cruda se trató posteriormente con diisopropilamina (3 equiv.) durante 2 h a temperatura ambiente. Luego la mezcla se filtró bajo atmósfera de argón, se concentró a sequedad y se absorbió en diisopropilamina al 20 % en heptano. La filtración rápida sobre un pequeño tapón de gel de sílice proporcionó el compuesto 8 deseado en forma de aceite incoloro.

10 RMN de ^1H (500 MHz, DMSO- d_6) δ = 6.96 (d, J = 8.5 Hz, 1H), 6.87 (d, J = 8.5 Hz, 1H), 6.74 (d, J = 8.4 Hz, 1H), 5.19 - 5.26 (m, 1H), 5.16 (dq, J = 10.4, 6.6 Hz, 1H), 3.57 (tdt, J = 13.6, 10.6, 6.8 Hz, 2H), 1.63 (d, J = 6.6 Hz, 3H), 1.55 (d, J = 6.4 Hz, 2H), 1.16 - 1.19 (m, 24H). RMN de ^{31}P (202 MHz, DMSO- d_6) δ = 137.63 (s, 1P), 127.90 (s, 1P).

15 **Ejemplo 2: Transcripción *in vitro* sin células**

La utilidad del compuesto (I) en el marcaje de ARN se demostró mediante su transcripción *in vitro* sin células para producir ARN mensajero (ARNm) de longitud completa fluorescente, a partir de una plantilla de ADN que codifica la proteína histona H2B fusionada a GFP (H2B:GFP).

20 La plantilla fue optimizada en cuanto a codones para limitar el número de repeticiones de C, evitando la autoextinción y mejorando el brillo. La transcripción eficiente y la incorporación de tC° se observaron utilizando dos ARN polimerasas de bacteriófagos diferentes, T7 y SP6 en proporciones de $\text{tC}^\circ\text{TP}/\text{CTP}$ canónico varían de 0 a 100 % (reemplazo completo), como se demuestra mediante electroforesis en gel de agarosa blanqueado (Fig. 3a para T7 y Fig. 4a para SP6).

25 Todas las transcripciones de ARN corren como una sola banda en los geles, con un tamaño correspondiente al producto de ARNm esperado de 1247 nt (H2B:GFP), lo que demuestra que se forma el ARNm de longitud completa. Las bandas de ARNm que contienen tC° se pudieron visualizar directamente con una excitación de 302 nm (Fig. 4a); las intensidades crecientes de banda con el aumento de la relación de reacción $\text{tC}^\circ\text{TP}/\text{CTP}$ apoyaron la incorporación exitosa de tC° dependiente de la concentración. La nueva visualización del gel después de la tinción con bromuro de etidio (Fig. 3b) proporcionó otra indicación cualitativa de que la incorporación de tC° no reduce el rendimiento de la reacción.

30 Además, no se observaron transcripciones más cortas, lo que sugiere que T7 procesa tC°TP correctamente y sin aborto prematuro. Se observan bandas de orden superior en todos los carriles del gel (Fig. 3b y 4b), pero se eliminaron tras la desnaturalización térmica, lo que sugiere la presencia de estructuras secundarias de ARN. Cabe destacar que esta característica aparece independientemente de la relación $\text{CTP}/\text{tC}^\circ\text{TP}$, que indica que el efecto no es específico de la base de citosina modificada.

35 Por lo tanto, estos resultados demuestran que tC° se puede incorporar con éxito en transcripciones de ARN de longitud completa incluso en condiciones en las que todo el CTP canónico se reemplaza con TC°TP (0 % de CTP; es decir, 100 % de eficiencia de marcado con C).

40 **Ejemplo 3: Caracterización espectroscópica de transcripciones sintetizadas *in vitro* de ARN modificado con tC°**

Se utilizó un enfoque espectroscópico para cuantificar la eficiencia de incorporación de tC°TP , comparado con el CTP canónico. Para permitir esto, todas las transcripciones de ARN se purificaron utilizando un kit Monarch RNA Cleanup, lo que garantiza la eliminación completa del tC°TP sin reaccionar. Los espectros de absorción (Fig. 5a) mostraron la aparición de una banda centrada en aproximadamente 370 nm en muestras con el análogo de citosina incorporado tC° , consistente con el perfil espectral de este análogo de base fluorescente (Fig. 5a).

45 Relacionando la absorción de las transcripciones de ARN purificadas a 260 nm, que refleja su concentración total, con la absorción a 370 nm (que emana exclusivamente de tC°), las constantes de velocidad relativas para la incorporación de CTP y TC°TP (k_C y k_{tC° , respectivamente, véase más adelante para más detalles). Los cocientes calculados k_C/k_{tC° (Fig. 5b) tienen valores cercanos o ligeramente superiores a la unidad (0.96-1.4, Fig. 5b), lo que demuestra que la polimerasa T7 no muestra una preferencia sustancial por el CTP canónico sobre el tC°TP en las reacciones *in vitro*. Esto respalda que la modificación química tricíclica de la citosina perturba mínimamente el proceso de transcripción.

55 El comportamiento emisor del tC° también se investigó en las transcripciones de ARNm explorando la relación con la fracción tC°TP añadida a la mezcla de reacción inicial. Se observó una disminución sustancial en el

rendimiento cuántico de fluorescencia (de 0.18 a 0.09, Fig. 5d) con el aumento de la incorporación de tC^o. Esto estuvo acompañado por una disminución en la vida útil de la fluorescencia (de 4.3 ns a 3.2 ns) y un ligero desplazamiento hacia el rojo del espectro de emisión (aproximadamente 4 nm, Figura 5c).

- 5 Esto puede atribuirse a la interacción electrónica (acoplamiento) de los estados moleculares del fluoróforo tC^o y el efecto autoextintor causado por la concentración creciente esperada de tC^o vecinales. Es importante destacar que este efecto de extinción con las fracciones de tC^o alto se equilibran con el gran número total de incorporaciones y no impiden la visualización del ARNm, incluso para las transcripciones donde todas las C se reemplazan por tC^o.

Ejemplo 4: Traducción de ARNm marcado con tC^o en lisados bacterianos

- 10 Para verificar la funcionalidad de las transcripciones de ARNm marcado con tC^o se investigó la traducción de un ARNm de calmodulina-3 marcado con tC^o en condiciones libres de células utilizando lisados bacterianos. Los ARNm marcados que codifican la proteína de 17 kDa se transcribieron a partir de un plásmido de plantilla de ADN de calmodulina-3 comercial utilizando las mismas relaciones de tC^oTP/CTP que para el ARNm codificante de H2B:GFP (0 a 100 % de tC^oTP). Después de la purificación del ARN y la traducción sin células,
- 15 se confirmó la presencia de calmodulina-3 mediante tinción de Coomassie (Fig. 6a) así como mediante transferencia Western (Fig. 6b). Satisfactoriamente observamos niveles de expresión estables al aumentar el contenido de tC^o de las transcripciones, que varían de 80 % - 137 % de los obtenidos utilizando un control de plantilla de ADN no marcado disponible comercialmente (Figura 6c).

Ejemplo 5: Eficiencia de traducción de ARNm marcado con tC^o en células humanas

- 20 Se utilizó electroporación para introducir transcripciones de ARNm marcadas con tC^o transcritas *in vitro* en células SH-SY5Y de neuroblastoma humano. Aprovechando que codifican para una proteína de fusión fluorescente con localización nuclear (H2B:GFP), se detectó la traducción por fluorescencia (Figs. 7 y 8). Para mejorar la estabilidad y reducir la degradación citosólica, los ARNm se protegieron con un análogo 5'-Cap 0 y 3' protegido mediante poliadenilación (mediante aproximadamente 300 nt).
- 25 La microscopía confocal de células vivas y la citometría de flujo mostraron que la fluorescencia de GFP en los núcleos celulares se podía detectar en el 32, 25, 18 y 12 % de las células 24 horas después de la electroporación para ARNm que contenían 25, 50, 75 y 100 % de tC^o, respectivamente (Fig. 7a y Fig. 8a). En comparación, la eficiencia de transfección con ARNm no modificado fue del 46 %. Esto proporciona la primera observación de que una transcripción de ARN modificada con un análogo de base fluorescente puede ser
- 30 traducida de manera precisa y eficiente por las maquinarias ribosómicas humanas, dando como resultado un producto proteico correctamente localizado y plegado.

- Mediante citometría de flujo se cuantificaron los niveles de fluorescencia de H2B:GFP en las células (Fig. 7b), observándose una disminución en la intensidad media de fluorescencia de H2B:GFP celular al aumentar el porcentaje de tC^o en la transcripción (aproximadamente un orden de magnitud de diferencia entre el 0 % y el
- 35 100 % de tC^o (Fig. 7a y 7f). Esto sugiere que, en estas condiciones, la traducción, a diferencia de la transcripción, se ve algo impedida por la modificación de tC^o, especialmente en la fracción de incorporación más alta.

- Es importante destacar que no se encontró evidencia de toxicidad celular inducida por ARNm 24 horas después de la electroporación (Fig. 8g). Es de destacar que la intensidad media de fluorescencia de la proteína traducida en cultivos celulares electroporados con un ARNm comercial codificante de GFP mejorado y marcado con Cy5 a través de conjugación con UTP (aproximadamente 25 % de todas las posiciones U de este ARNm están
- 40 marcadas) fue solo el 17 % de eso en los cultivos celulares electroporados con la secuencia no marcada correspondiente (Fig. 7f). Esto es comparable al efecto de una transcripción de tC^o marcada un 50 % - 75 %, lo que sugiere que las modificaciones de Cy5 afectan la capacidad de traducción del ribosoma más de lo que lo hace tC^o.
- 45

- De las imágenes de las figuras 7a y 7e se desprende claramente que ni el ARNm marcado con tC^o fluorescente ni el ARNm marcado con Cy5 pueden detectarse dentro de las células después de la electroporación debido a su concentración citoplasmática baja/difundida. Por lo tanto, se investigó la administración de las transcripciones H2B:GFP utilizando un reactivo de transfección química (lipofectamina), que también es más
- 50 relevante desde la perspectiva de la administración de fármacos. Esto dio como resultado la producción exitosa de H2B:GFP (Figs. 7c, 8b y 8d) independientemente del contenido de tC^o, pero, aunque con eficiencias de administración más bajas (3.8 - 4.6 % de células positivas para H2B:GFP para transcripciones marcadas con tC^o frente a la electroporación posterior del 12-32 %). Esto refleja la transfectabilidad relativamente pobre de las células SH-SY5Y en comparación con muchas otras líneas celulares en lugar de la transfectabilidad de
- 55 construcciones marcadas con tC^o, como lo respalda además el hallazgo de que los cultivos transfectados con ARNm no modificado mostraron una respuesta prácticamente idéntica (4.5 % de células que expresaban H2B:GFP).

Se encontró que la fluorescencia de H2B:GFP aumentaba gradualmente con el tiempo entre 24 h y 72 h (Fig. 7c), lo que es un comportamiento contrastante en comparación con las células electroporadas (Fig. 7a), lo que sugiere que los complejos lipofectamina-ARNm se internalizan continuamente y, potencialmente, que los complejos endocitados liberan progresivamente más transcripciones con el tiempo, contrarrestando el efecto degradativo en el citosol.

Cuando se administra usando lipofectamina, se descubrió que los ARNm marcados con tC^o promovían una traducción de H2B:GFP muy similar en comparación con el ARNm no marcado correspondiente, como lo indican los niveles de fluorescencia en la Fig. 7f. Por otra parte, el ARNm etiquetado con Cy5 produce un nivel de fluorescencia promedio que es un 80 % inferior al de su transcripción no marcada correspondiente. Esto sugiere que tC^o no afecta la capacidad del ARNm de ser procesado por los ribosomas tras la transfección química, posiblemente debido a una ausencia de carga y un impedimento estérico reducido, mientras que Cy5, actualmente el fluoróforo comercial más común para el marcado del ARNm, afecta considerablemente el proceso de traducción y/o interfiere con un proceso de captación similar al nativo del ARNm, ya que introduce una anfifilicidad significativa en el ARNm y, por lo tanto, posiblemente interacciones no nativas entre el CY5 - ARNm y los componentes de la membrana lipofílica.

Es importante destacar que la complejación del ARNm marcado con tC^o con lipofectamina permitió su visualización directa dentro de las células mediante microscopía confocal de células vivas (Fig. 7c). Esto representa la primera observación de ácidos nucleicos marcados con análogos de bases fluorescentes dentro de células vivas. También se descubrió que era posible visualizar simultáneamente, en tiempo real, la captación y la posterior traducción de un ARNm modificado con un análogo de base fluorescente mediante registros en lapsos de tiempo. Observamos la localización conjunta de la señal de tC^o con una proteína Rab5 marcada con mRFP, destacando así que el ARNm transita a través del endosoma temprano (Fig. 7d). En consecuencia, esta tecnología permite el seguimiento tanto de las transcripciones de ARNm marcadas intrínsecamente como de sus productos de traducción en vivo, para recopilar información espaciotemporal sobre el producto de traducción, incluso con un 50 % de contenido de tC^o. Esto demuestra la flexibilidad y versatilidad de este nuevo enfoque de marcado donde el ajuste fino del contenido de tC^o se puede utilizar para optimizar el ARNm para aplicaciones específicas de administración de fármacos.

Ejemplo 6: Métodos bioquímicos

Generación de la plantilla de ADN de H2B:GFP

La secuencia codificante original para H2B:GFP se tomó del plásmido pCS2-H2B:GFP (Addgene, plásmido # 53744, optimizado manualmente por codones para minimizar la aparición de tramos de poli-Cn ($n < 3$), ensamblado *in silico* con un promotor T7 adicional y otras características deseadas (secuencias de consenso Shine-Dalgarno/Kozak para mejorar la traducción y un 3xStop, respectivamente en los extremos 5' y 3' de la secuencia codificante en sí, más los sitios de restricción HindIII/SnaBI necesarios, para generar los extremos pegajosos propensos a la ligadura) y se ordenó a través de Twist Bioscience como un bloque de genes sintéticos. La secuencia obtenida se amplificó luego por PCR, utilizando un Phusion Hot Start High-Fidelity Taq (Thermo Scientific), y se subclonó en una cadena principal pCS2 vacía digerida con HindIII/SnaBI (enzimas Fast Digest, Thermo Scientific). Después de la ligadura con ligasa T4 durante 1 h a temperatura ambiente (Roche), las células competentes de *E. coli* DH5 α (Invitrogen) se transformaron siguiendo el protocolo recomendado y las colonias obtenidas se analizaron mediante PCR de colonias. Luego, la colonia seleccionada se inoculó en un volumen a escala midiprep de medio de crecimiento Luria-Bertani líquido (VWR) y se aisló el ADN plasmídico utilizando un kit de purificación de plásmido PureLink Fast Low-Endotoxin Midi (Thermo Scientific). El plásmido purificado fue finalmente digerido nuevamente con HindIII/SnaBI y purificado en gel, para generar la plantilla de transcripción con el tamaño deseado.

Cebadores (Eurofins Genomics):

Giro-H2B.F: GAAGTGCCATTCCGCCTGAC

Giro-H2B.R: CACTGAGCCTCCACCTAGCC

Transcripción y purificación del ARN de H2B:GFP

Las reacciones de transcripción *in vitro* para las polimerasas T7 y SP6 (Thermo Scientific) se ensamblaron según lo recomendado por los protocolos correspondientes, con algunas modificaciones que dieron como resultado un rendimiento consistentemente mayor en todas las condiciones:

1. Tampón de transcripción 5X - 10 μ L
2. Mezcla NTP, 10 mM cada una (concentración final de 2 mM) - volumen según lote para tC^oTP
3. Plantilla de ADN linealizada 1 μ g - volumen en función de la concentración

4. Inhibidor de la RNasa RiboLock - 1.25 μ L (50 U)

5. ARN polimerasa T7/T3/SP6 - 3 μ L (60 U, el doble de lo recomendado)

6. $MgCl_2$ concentración final de 4 mM (aumentada según lo recomendado por Thomen, P. et al. Biophys. J. 95, 2423-2433 [2008])

5 7. Agua tratada con DEPC cantidad suficiente para 50 μ L

10 Las transcripciones *in vitro* siempre se realizaron a 20 °C durante 14 h, luego los ARN se purificaron utilizando un kit Monarch RNA Cleanup (NEB), o tampones equivalentes caseros y columnas regeneradas siguiendo el mismo razonamiento. Fue posible recuperar parcialmente el TC^oTP que no reaccionó de las mezclas de transcripción por HPLC para reutilizar en ensayos posteriores. Para los estudios celulares, a cada lote de ARN se le agregó enzimáticamente una cola de poliA (con una polimerasa Poli(A), protocolo NEB # M0276 con incubación extendida a 1 hora) y un análogo de Cap 0 (utilizando un sistema de protección Vaccinia, protocolo NEB # M2080), siguiendo los procedimientos recomendados.

Geles desnaturalizantes de lejía y agarosa

15 Para una verificación cualitativa de todos los ARN sintetizados *in vitro*, se corrió un gel de agarosa desnaturalizante, en presencia de lejía al 1,5 % (Sigma Aldrich), como se recomienda en Aranda, PS, LaJoie, DM y Jorcyk, CL, Electrophoresis 33, 366-369 [2012]. Los ARN se mezclaron primero con un colorante de carga de ADN 6x (Invitrogen) y luego se desnaturalizaron por calor a 70 °C durante 10 minutos en un bloque calefactor, luego se transfirieron inmediatamente y se mantuvieron en hielo. La escalera de ARN de alto rango RiboRuler (Thermo Scientific) se sometió al mismo tratamiento; se cargaron 2 μ L de escalera de ARN a lo largo de las muestras y el gel se corrió a voltaje constante (70 V) durante 1 h y luego se tomaron imágenes bajo transiluminación UV (302 nm) utilizando un ChemiDoc Touch (BioRad). Para contrateñir todo el gel, y especialmente los carriles sin muestras que contenían tC^oTP, finalmente se realizó una tinción estándar con bromuro de etidio a temperatura ambiente durante 10 min y agitación suave, seguido de dos lavados en TAE y luego un lavado final en agua destilada (10 min cada uno).

25 Cultivo celular

30 Las células de neuroblastoma humano SH-SY5Y (Sigma-Aldrich) se cultivaron en una mezcla 1:1 de medio esencial mínimo (HyClone) y mezcla de nutrientes F-12 Ham (Sigma-Aldrich) suplementada con 10 % de suero bovino fetal (FBS), 1 % de aminoácidos no esenciales (Lonza) y L-glutamina 2 mM. Para los experimentos de seguimiento, se cultivó un modelo generado internamente de células hepáticas humanas Huh-7 que sobreexpresaban de forma estable mRFP-Rab5 en DMEM/GlutaMax/Alto en glucosa (Gibco) suplementado con 10 % de FBS. Las células se desprenden con tripsina-EDTA 0.05 % (Gibco) y se pasan dos veces por semana.

Electroporación o transfección química

35 Las células se electroporaron con 9.7 μ g de TC^oTP (para experimentos de incorporación *in vitro*) o 100 ng de ARNm marcado con tC^o por cada 105 células (para traducción *in vitro*, evaluación de citotoxicidad, análisis de citometría de flujo y microscopía confocal), utilizando un sistema de transfección Neon (Invitrogen, Carlsbad, CA, EE. UU.) y siguiendo el protocolo para 10 μ L de Neon Tip proporcionado por el fabricante, con un pulso triple de 1200 V y un ancho de pulso de 20 ms. Para la transfección química, las células SH-SY5Y se sembraron un día antes de la transfección a una densidad de 0.8 $\cdot 10^6$ células/mL, en placas de 48 pocillos o placas de cultivo con fondo de vidrio para análisis por citometría de flujo o microscopía confocal, respectivamente. Se utilizó Lipofectamina MessengerMAX como reactivo químico para la transfección de acuerdo con las instrucciones del fabricante. Brevemente, el reactivo se diluyó y se incubó durante 10 minutos a temperatura ambiente en medio Opti-MEM. Se agregaron construcciones de tC^o-ARNm al reactivo para alcanzar una relación final de reactivo-ARNm de 1:1 (v/p), seguido de una incubación de 5 minutos a temperatura ambiente que permitió que se formara el complejo ARNm-lípido. Las células se incubaron con este complejo hasta 72 h. Para abordar el impacto de la incorporación del colorante en la traducción del ARN, se describe aquí que se electroporaron o transfectaron químicamente las células SH-SY5Y con ARNm codificantes de eGFP no marcados (NL) o marcados con Cianina5 (Cy5) disponibles comercialmente (Trilink®).

Evaluación de citotoxicidad

50 La integridad de la membrana celular se determinó utilizando el kit de ensayo de citotoxicidad Pierce™ LDH (Invitrogen) de acuerdo con las instrucciones del fabricante. Brevemente, LDH liberada en los sobrenadantes de las células 24 h después de la electroporación o de la transfección con ARNm marcado con tC^o o Cy5-ARNm, se midió con un ensayo enzimático acoplado que da como resultado la conversión de una sal de tetrazolio en un producto de formazán rojo. La absorbancia se registró a 490 nm y 680 nm. La toxicidad se expresó como el porcentaje de liberación de LDH en el sobrenadante en comparación con la liberación máxima

de LDH (lisado celular del sobrenadante). Los datos son las medias \pm DE de tres experimentos realizados por triplicado.

Citometría de flujo

5 Después de la electroporación de ARNm marcado tC^o, las células se sembraron en una placa de 48 pocillos (2.105 células/pocillo) y la expresión de H2B:GFP en las células se cuantificó mediante citometría de flujo. Brevemente, 24 h, 48 h o 72 h después de la electroporación o después de la transfección con tC^o-ARNm, ARNm no marcado o Cy5-ARNm se recolectaron y analizaron las células en un citómetro de flujo Guava EasyCyte 8HT (Millipore). Los datos son intensidades de fluorescencia medias \pm DE de células vivas individuales seleccionadas de tres experimentos realizados por triplicado. Las intensidades de fluorescencia promedio se corrigieron al valor inicial restando la señal de las células electroporadas o transfectadas con agua libre de RNasa. Todos los datos de citometría de flujo se analizaron en el software Flowing (versión 2.5.1) y se mostraron usando R (<http://www.R-project.org/>). H2B:GFP: Excitación 488 nm; Emisión 525-530 nm.

Microscopía confocal

15 Después de la electroporación, las células se sembraron en placas de cultivo con fondo de vidrio (placas de cultivo con fondo de vidrio MatTek o en placas CELLview subdivididas en 4 sectores; 2.105 células/cámara). Para el experimento de seguimiento, las células Huh-7 que sobreexpresaban de forma estable mRFP-Rab5 se incubaron con el complejo lipofectamina/tC^o-ARNm y se registró el lapso de tiempo hasta 20 h después de la transfección química. Las imágenes confocales se adquirieron en un microscopio confocal Nikon C2+ equipado con una unidad detectora C2-DUVB GaAsP y utilizando un objetivo Nikon APO de inmersión en aceite de 60 \times 1.4 (Nikon Instruments, Ámsterdam, Países Bajos). Los datos fueron procesados con el software Fiji. H2B:GFP: Excitación 488 nm; Emisión 495-558 nm. ARNm marcado con tCO: Excitación 405 nm; Emisión 447-486 nm. ARNm marcado con Cy5: Excitación 640 nm; Emisión 652-700 nm. mRFP-Rab5: Excitación 561 nm; Emisión 565-720 nm.

Traducción sin células

25 Se realizaron reacciones de traducción sin células utilizando lisados bacterianos de *E. coli* y un sistema de expresión libre de células Expressway™ Mini (Thermo Scientific). La proteína similar a calmodulina 3 se proporciona como un plásmido de control positivo (pEXP5-NT/CALML3) en el propio kit; este vector de ADN contiene un promotor de polimerasa T7 y una etiqueta 6xHis, por lo que primero se transcribió *in vitro* en presencia de las concentraciones deseadas de TC^oTP (véase más arriba). Los ARN obtenidos, una vez purificados, se utilizaron como plantillas para la reacción de traducción sin células de acuerdo con las recomendaciones del fabricante: Extracto de *E. coli* slyD - 20 μ L; tampón de reacción de *E. coli* IVPS 2.5X (-A.A.) - 20 μ L; 50 mM de aminoácidos (-Met) - 1.25 μ L; 75 mM de metionina* -1 μ L; Mezcla de enzimas T7 -1 μ L (se omite cuando se utilizan ARN marcados con tC^o); plantilla de ADN - 1 μ g (al probar los ARN marcados con tC^o, se añadió la misma cantidad de ARN en su lugar); agua destilada libre de DNasa/RNasa cantidad suficiente para 50 μ L.

Tinción de Coomassie y transferencias Western

40 Las muestras de proteínas de los experimentos de traducción *in vitro* se cuantificaron con un kit de ensayo de proteínas Qubit (Thermo Scientific), se mezclaron con un tampón reductor SDS LaemmLi 6x (Alfa Aesar), luego se desnaturalizaron por calor a 85 °C durante 10 minutos y se mantuvieron a temperatura ambiente hasta que fueron necesarias. Las muestras generalmente se procesaron en geles de poliacrilamida al 4-20 % Novex MES/SDS de 1 mm (Thermo Scientific) y utilizando un minitanque de gel, con la fuente de alimentación configurada a voltaje constante (200 V). Para la tinción de Coomassie, el gel se lavó luego tres veces en agua hirviendo, para eliminar el exceso de SDS, en un agitador de mesa; se agregó una solución de tinción no tóxica de Coomassie 1x al gel y se calentó en microondas hasta la ebullición inicial.

45 Finalmente, el gel se lavó después del tiempo de incubación adecuado, para eliminar el exceso de ruido de fondo, en agua destilada y se fotografió utilizando un ChemiDoc Touch. Para la transferencia Western, los geles se transfirieron a membranas activadas con etanol PVDF LF (BioRad) con un sistema semiseco TransBlot (BioRad), de acuerdo con las recomendaciones del fabricante (configuraciones para geles de 1 mm de espesor y proteínas de peso mixto). Luego, las membranas de PVDF se lavaron durante 5 minutos en TBS-T (TBS y 0.1 % de Tween-20, Sigma Aldrich), se bloquearon en leche al 5 % en TBS durante 1 hora a temperatura ambiente y se incubaron con las diluciones apropiadas de anticuerpos primarios.

55 Después de 3 lavados de 5 min en TBST y una incubación de 1 h con los correspondientes anticuerpos secundarios conjugados con HRP, la membrana se lavó nuevamente tres veces en TBS-T, una vez en TBS y una vez más en agua destilada. Finalmente, las membranas se incubaron con un volumen mínimo de SuperSignal West Pico PLUS (Thermo Scientific) y se obtuvieron imágenes con un ChemiDoc Touch. Anticuerpos primarios: etiqueta anticuerpo monoclonal de ratón anti-6xHistidina (Invitrogen) y anticuerpo monoclonal de ratón anti-GAPDH (ref. 437000, Invitrogen), ambos diluidos 1:1000 en BSA al 3 %/TBS-T.

Anticuerpos secundarios: IgG de cabra anti-Ms y anti-Rb conjugado con HRP de adsorción cruzada (H + L) (ref. A16072 y A16104, Invitrogen), utilizado en una dilución de 1:10000 en TBS-T.

Ejemplo 8: Métodos espectroscópicos

5 Los productos de tC^o-ARN de las reacciones de transcripción libres de células (antes de la poliadenilación y protección, ver Métodos: Bio para más detalles) se midieron tal como se recibieron, es decir, en agua Milli Q libre de RNasa. Todas las mediciones se realizaron a temperatura ambiente (aproximadamente 22 °C) en una cubeta de cuarzo con una longitud de trayectoria de 3.0 mm, con un volumen de muestra de aproximadamente 60 µL.

Absorción en estado estacionario

10 Los espectros de absorción se registraron en un espectrofotómetro Cary 5000 (Varian Technologies) con un intervalo de longitud de onda de 1.0 nm, un tiempo de integración de 0.1 s y un ancho de banda espectral (SBW) de 1 nm. Todos los espectros se corrigieron según la línea base restando la absorción correspondiente únicamente del solvente. Se aplicó un filtro de suavizado polinomial de segundo orden Savitzky-Golay (cinco puntos) a todos los espectros. Para las muestras que exhibieron una dispersión significativa, como lo demuestra la absorción característica en la región de longitud de onda larga (aquí para $\lambda > 475$ nm), se aplicó una corrección adicional. La contribución de la dispersión ($A_{dispersión}$) a la absorción se ajustó en tales casos (utilizando la absorción a 550-475 nm como entrada) a la función de dispersión de Rayleigh (ecuación S1), donde c es una constante de proporcionalidad y A_0 una constante y luego se resta para todas las longitudes de onda.

$$20 \quad A_{dispersión}(\lambda) = \log\left(\frac{c}{1 - c\lambda^{-4}}\right) + A_0 \quad (S1)$$

Emisión en estado estacionario

25 Los espectros de emisión se registraron en un fluorímetro SPEX Fluorolog (Jobin Yvon Horiba) con excitación a 356 nm. La emisión se recogió en ángulo recto con un tiempo de integración de 0.1 s y un intervalo de longitud de onda de 1 nm. Las rendijas del monocromador se ajustaron para lograr una salida de señal óptima, lo que produjo SBW en el intervalo de 1.5-2.5 nm tanto en el lado de excitación como en el de emisión. Los espectros de emisión se corrigieron para la dispersión Raman restando la emisión correspondiente de una muestra que contenía solo disolvente. Se aplicó un filtro de suavizado polinomial de segundo orden Savitzky-Golay (cinco puntos) a todos los espectros.

Determinación del rendimiento cuántico de fluorescencia

30 Los rendimientos cuánticos de fluorescencia de muestra (Φ_F) se determinaron en relación con una solución de sulfato de quinina (Sigma) en H₂SO₄ 0.5 M ($\Phi_{F,REF} = 0.546$) y calculado de acuerdo con la ecuación S2.

$$\Phi_F = \Phi_{F,REF} \times \frac{\int_{\lambda_i}^{\lambda_f} I_S(\lambda) d\lambda}{\int_{\lambda_i}^{\lambda_f} I_{REF}(\lambda) d\lambda} \times \frac{A_{REF}}{A_S} \times \frac{\eta_S^2}{\eta_{REF}^2} \quad (S2)$$

35 Los espectros de emisión de la muestra, $I_S(\lambda)$ y referencia, $I_{REF}(\lambda)$, se integraron entre $\lambda_i = 365$ nm y $\lambda_f = 700$ nm. La absorción a la longitud de onda de excitación (356 nm) para la muestra (A_S) y referencia (A_{REF}) estaban en el intervalo de 0.05-0.11 para todas las muestras. Los índices de refracción de disolventes adoptados para las muestras (agua) y la referencia (H₂SO₄ 0.5 M) eran $\eta_S = 1.333$ y $\eta_{REF} = 1.339$, respectivamente. Todos los rendimientos cuánticos se presentan como la media \pm la desviación estándar de dos reacciones de transcripción independientes libres de células.

Emisión resuelta en el tiempo

40 Los tiempos de vida de la fluorescencia se determinaron utilizando el recuento de fotones individuales correlacionado en el tiempo (TCSPC). Las muestras se excitaron utilizando un diodo láser pulsado LDH-P-C-375 (PicoQuant) con emisión centrada a 377 nm (el ancho de pulso FWHM fue de 1 nm y 70 ps con respecto a la longitud de onda y el tiempo, respectivamente), operado con un controlador láser PDL 800-B (PicoQuant) a una frecuencia de repetición de 10 MHz. La emisión de la muestra (458 nm, SBW = 10 nm) se recolectó en ángulo recto, a través de un polarizador de emisión ajustado a 54.9° (detección de ángulo mágico). Los recuentos de fotones se registraron en una placa PMT de 50 microcanales R3809U (Hamamatsu) y se introdujeron en un analizador multicanal LifeSpec (Edinburgh Analytical Instruments) con 2048 canales activos (24.4 ps/canal), hasta que se cumplió la condición de detención de 104 recuentos en el canal superior. La función de respuesta del instrumento (IRF) se determinó utilizando un inserto modular de vidrio esmerilado (dispersión) mientras se observaba la emisión a 377 nm (SBW = 10 nm).

Ajuste de los tiempos de vida de la fluorescencia

Las caídas de intensidad se ajustaron mediante reconvolución IRF al modelo multiexponencial que se muestra en la ecuación S3.

$$I(t) = \int_0^t IRF(t') \sum_{i=1}^n \alpha_i e^{-\frac{t-t'}{\tau_i}} dt' \quad (S3)$$

- 5 El procedimiento de ajuste de reconvolución de mínimos cuadrados se llevó a cabo utilizando el software DecayFit (<http://www.fluorotools.com/software/decayfit>). Todas las caídas se ajustaron a un modelo triexponencial ($n = 3$). Los tiempos de vida presentados son tiempos de vida promedio ponderados por amplitud (τ), calculados utilizando los factores preexponenciales α_i y tiempos de vida (τ_i) de acuerdo con la ecuación S4. Los parámetros de ajuste para las desintegraciones se muestran en la Tabla 2.

$$\bar{\tau} = \sum_{i=1}^n \alpha_i \tau_i \quad (S4)$$

- 10 Tabla 2: Parámetros de tiempo de vida ajustados para los experimentos de TCSPC. El valor χ^2 (Chi-cuadrado) se evaluaron para indicar la bondad del ajuste.

	transcripción	α_i	τ_i (ns)	α_i	τ_i (ns)	α_i	τ_i (ns)	$\bar{\tau}$ (ns)	χ^2
Conjunto 1	25%	0.19	0.67	0.49	3.2	0.31	5.6	3.5	1.09
	50%	0.28	0.71	0.43	2.8	0.30	5.2	2.9	1.06
	75%	0.33	0.66	0.45	2.6	0.23	5.2	2.5	1.12
	100%	0.33	0.48	0.44	2.1	0.23	4.8	2.2	1.00
Conjunto 2	25%	0.21	0.63	0.40	2.8	0.34	5.3	3.3	0.99
	50%	0.26	0.64	0.41	2.7	0.30	5.2	3.0	1.03
	75%	0.29	0.50	0.42	2.1	0.22	4.8	2.4	1.02
	100%	0.37	0.47	0.40	1.9	0.39	4.5	2.0	1.01

Cinética de la reacción de transcripción sin células

- 15 La relación entre las constantes de velocidad de la citosina y la incorporación de tC^O (k_C/k_{tC^O}) se calculó utilizando los espectros de absorción de las transcripciones de tC^O -ARN (A_{260} y A_{369}) y concentraciones iniciales de trifosfato ($[CTP]_0$ y $[tC^O TP]_0$) como entrada. Las ecuaciones S5 y S6 se deducen de la suposición de una cinética de reacción de primer orden con respecto a las especies de trifosfato $[CTP]$ y $[tC^O TP]$.

$$\frac{d[CTP]}{dt} = -k_C \times [CTP] \quad (S5)$$

$$\frac{d[tC^O TP]}{dt} = -k_{tC^O} \times [tC^O TP] \quad (S6)$$

- 20 Resolviendo S5 y S6 para las constantes de velocidad respectivas se obtiene la ecuación S7, en la que $[C]$ y $[tC^O]$ denotan la concentración de C y tC^O incorporados, respectivamente.

$$\frac{k_C}{k_{tC^O}} = \frac{\ln\left(\frac{[CTP]_0}{[CTP]_0 - [C]}\right)}{\ln\left(\frac{[tC^O TP]_0}{[tC^O TP]_0 - [tC^O]}\right)} \quad (S7)$$

Utilizando la ley de Lambert-Beer, la absorción está relacionada con la concentración de nucleobase de acuerdo con las ecuaciones S8 y S9. Las siguientes absorptividades molares (unidad: $M^{-1}cm^{-1}$) fueron adoptadas:

$$\epsilon_{260}^{tC^O} = 12200, \epsilon_{260}^C = 7400, \epsilon_{260}^G = 11800, \epsilon_{260}^U = 9300, \epsilon_{260}^A = 15300, \epsilon_{369}^{tC^O} = 9370.$$

- 25 $A_{260} = 0.9 \times l \times (\epsilon_{260}^{tC^O} \times [tC^O] + \epsilon_{260}^C \times [C] + \epsilon_{260}^G \times [G] + \epsilon_{260}^U \times [U] + \epsilon_{260}^A \times [A]) \quad (S8)$

$$A_{369} = l \times \epsilon_{369}^{tC^O} \times [tC^O] \quad (S9)$$

Suponiendo que el ARN producto es uniforme en tamaño (1247 nucleótidos), su composición de bases (A: 408, U: 272, G: 307, C: 260) permite las ecuaciones S10 - S13.

$$[tC^O] + [C] = [ARN] \times 260 \quad (S10)$$

- 30 $[A] = [ARN] \times 408 \quad (S11)$

$$[U] = [ARN] \times 272 \quad (S12)$$

$$[G] = [ARN] \times 307 \quad (S13)$$

5 La solución del sistema de ecuaciones compuesto por S7 a S13 permite cuantificar k_c/k_{tC^O} , $[tC^O]$, $[ARN]$, $[C]$, $[A]$, $[U]$, y $[G]$. El grado de incorporación de tC^O (θ_{tC^O}) de cadena promedio se puede calcular entonces de acuerdo con la ecuación S14.

$$\theta_{tC^O} = \frac{[tC^O]}{[C] + [tC^O]} \quad (S14)$$

Utilizando el volumen de la reacción libre de células (50 μ L) y la solución del producto resultante (100 μ L), se aplicó la ecuación S15 para calcular el rendimiento de incorporación de tC^O (η_{tC^O}).

$$\eta_{tC^O} = \frac{[tC^O] \times 100 \mu\text{L}}{[tC^O]_{TP} \times 50 \mu\text{L}} \quad (S15)$$

10 En consecuencia, el rendimiento de ARN (η_{ARN}) se calculó de acuerdo con la ecuación S16.

$$\eta_{ARN} = \frac{[ARN] \times 100 \mu\text{L}}{[ARN]_{TP} \times 50 \mu\text{L}} \quad (S16)$$

Conclusiones

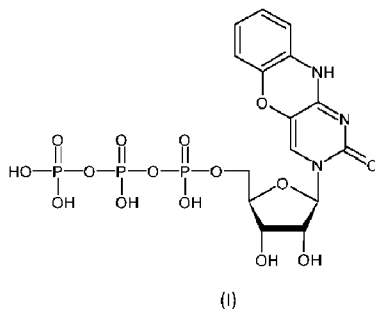
15 Esta memoria descriptiva demuestra que un análogo artificial de la citosina, de tamaño expandido, asume el papel de la citosina natural y es reconocido correctamente por varias maquinarias enzimáticas, incluido el ribosoma. Este análogo de base fluorescente, tC^O , se ha demostrado que es una etiqueta de imagen intrínseca adecuada para ARN de diferentes tamaños que altera mínimamente las propiedades nativas y es compatible con los procesos de marcado enzimático.

20 Las transcripciones modificadas no son tóxicas y son traduccionalmente activas tanto en lisados bacterianos como en sistemas eucariotas, independientemente de su grado de incorporación de tC^O . Esto permite convenientemente el monitoreo simultáneo de la captación y traducción de ARNm en H2B:GFP en microscopía confocal de células vivas utilizando excitación selectiva, un enfoque que debería ser aplicable a la traducción de cualquier proteína etiquetada de manera similar con una proteína de la familia GFP.

25 Por lo tanto, las metodologías de marcado de ARN con fluorescencia intrínseca que se describen en este documento son excelentes formas no invasivas de dilucidar, en tiempo real, los mecanismos de tráfico celular, como el escape endosómico o la formación de exosomas, ambos de fundamental importancia para las aplicaciones farmacéuticas. Como tal, la tecnología para obtener imágenes de células vivas debería permitir estrategias de administración nuevas y mejoradas para medicamentos basados en ácidos nucleicos de próxima generación.

REIVINDICACIONES

1. Un compuesto de fórmula (I), también conocido como tC^oTP, o una sal del mismo:



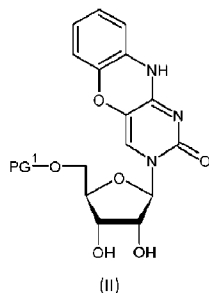
2. Un compuesto de fórmula (I) de acuerdo con la reivindicación 1.

5 3. Un compuesto de fórmula (I) de acuerdo con la reivindicación 1, que es una sal de sodio, potasio o amonio.

4. Un compuesto de fórmula (I) de acuerdo con la reivindicación 3, que es una sal monosódica, disódica, trisódica, monoamónica, diamónica o triamónica.

5. Un proceso para preparar un compuesto de fórmula (I) o una sal del mismo de acuerdo con cualquiera de las reivindicaciones anteriores que comprende:

10 i. Proporcionar un compuesto de fórmula (II) o una sal del mismo:



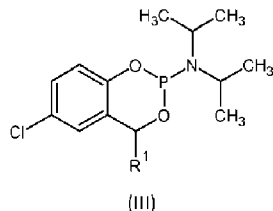
Donde PG¹ es un grupo protector adecuado;

ii. Inmovilizar el compuesto de fórmula (II) o una sal del mismo uniendo uno de sus grupos de alcohol secundario a un soporte adecuado;

15 iii. Proteger cualquier grupo de alcohol secundario libre con un grupo protector PG² adecuado;

iv. Eliminación del grupo protector PG¹;

v. Hacer reaccionar el grupo de alcohol primario expuesto con un compuesto de fórmula (III):



Donde R¹ se selecciona entre un grupo hidro y un grupo alquilo C₁₋₃;

20 vi. Oxidar el compuesto de fósforo (III) resultante a un compuesto de fósforo (V);

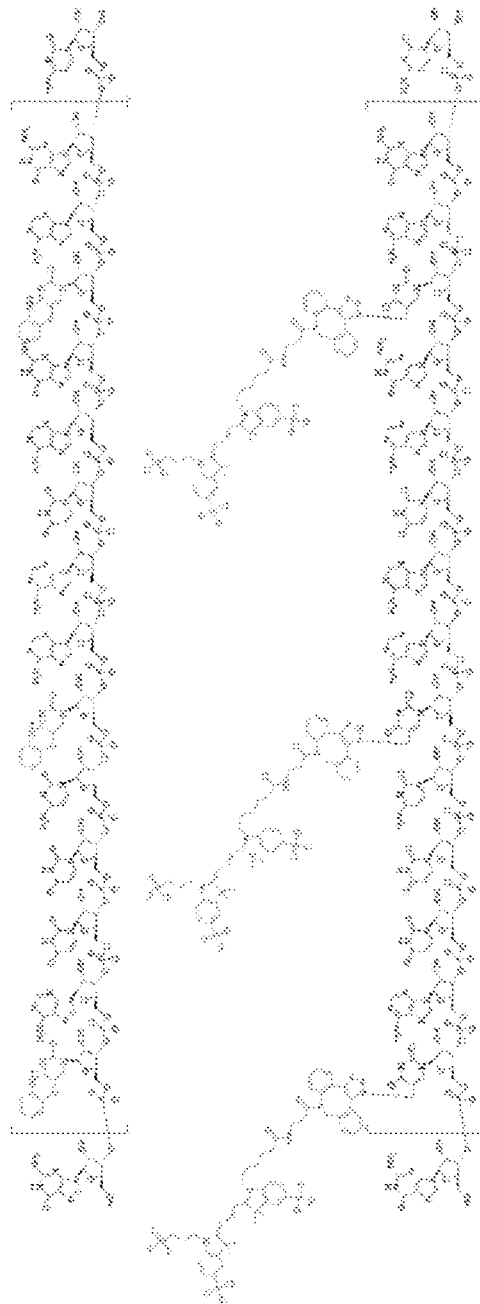
vii. Hacer reaccionar el compuesto de fósforo (V) con un pirofosfato de tetraalquilamonio para generar un trifosfato;

viii. Eliminación del grupo protector PG²; y

ix. Escindir el trifosfato resultante del soporte para generar un compuesto de fórmula (I) o una sal del mismo.

6. Un proceso para preparar un compuesto de fórmula (I) o una sal del mismo de acuerdo con la reivindicación 5, donde R^1 es un grupo metilo.
7. Un proceso para preparar un compuesto de fórmula (I) o una sal del mismo de acuerdo con la reivindicación 5 o la reivindicación 6, donde el soporte es un polímero sólido.
- 5 8. Un proceso para preparar un compuesto de fórmula (I) o una sal del mismo de acuerdo con cualquiera de las reivindicaciones 5 a 7, donde PG^1 es dimetoxitritilo y PG^2 es acetilo.
9. Una composición para preparar una molécula de ARN marcada con tC^O que comprende un compuesto de fórmula (I) de acuerdo con cualquiera de las reivindicaciones 1 a 4 y un trifosfato de ribonucleótido natural.
- 10 10. El uso de un compuesto de fórmula (I) o una sal del mismo de acuerdo con cualquiera de las reivindicaciones 1 a 4 para preparar enzimáticamente una molécula de ARN marcada con tC^O .
11. El uso de un compuesto de fórmula (I) o una sal del mismo de acuerdo con la reivindicación 10, donde la molécula de ARN es ARNm.
12. Un proceso para preparar una molécula de ARN marcada con tC^O que comprende proporcionar una plantilla de ADN a una composición que comprende un compuesto de fórmula (I) y un trifosfato de ribonucleótido natural, luego tratar la mezcla resultante con una ARN polimerasa.
- 15 13. El uso de una molécula de ARNm marcada con tC^O para preparar una proteína codificada por el ARNm mediante traducción.
14. El uso de una molécula de ARNm marcada con tC^O de acuerdo con la reivindicación 13, en la que la proteína codificada está fusionada a una proteína fluorescente.
- 20 15. El uso de una molécula de ARNm marcada con tC^O de acuerdo con la reivindicación 14, donde el ARNm marcado con tC^O y la proteína codificada se analizan simultáneamente espaciotemporalmente utilizando microscopía confocal.

Figure 1



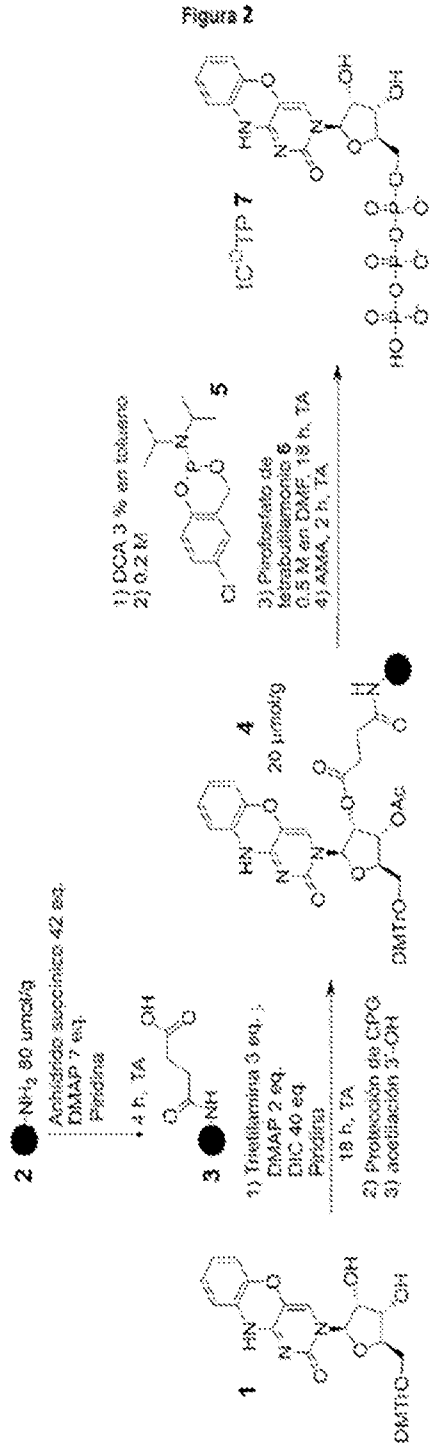


Figura 3a

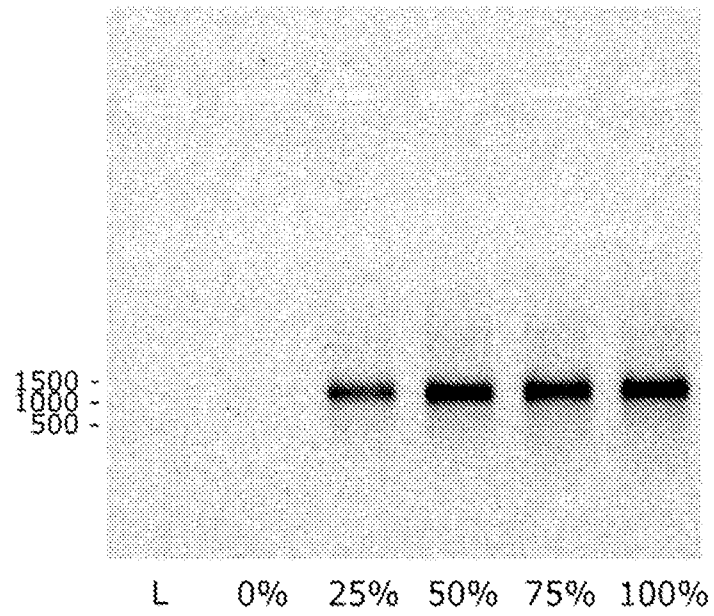


Figura 3b

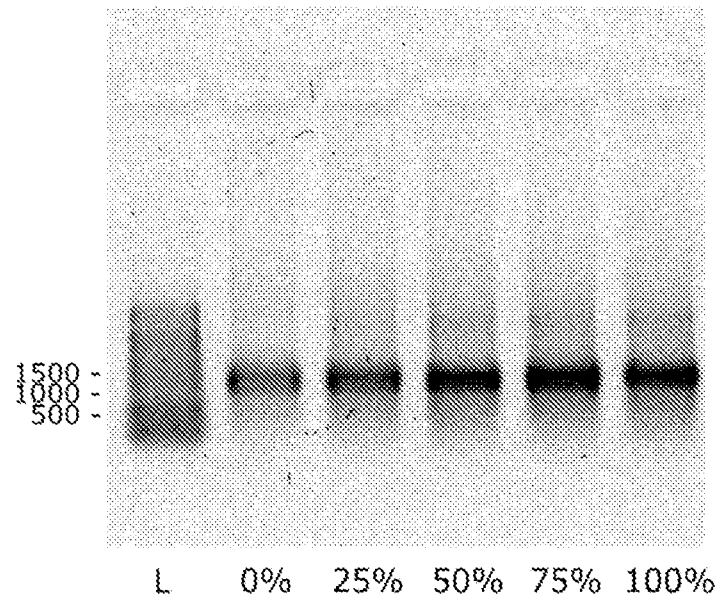


Figura 3c

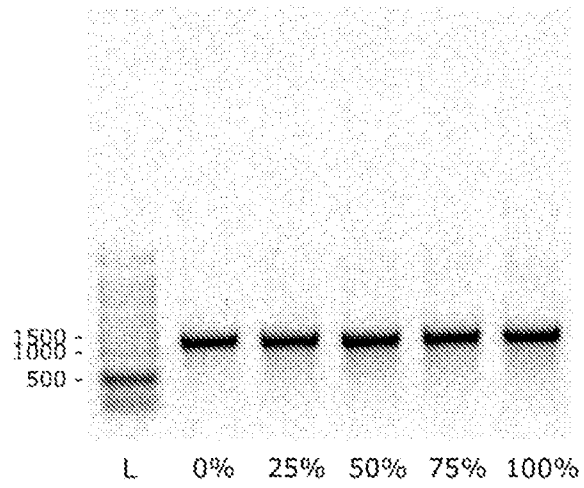


Figura 4a

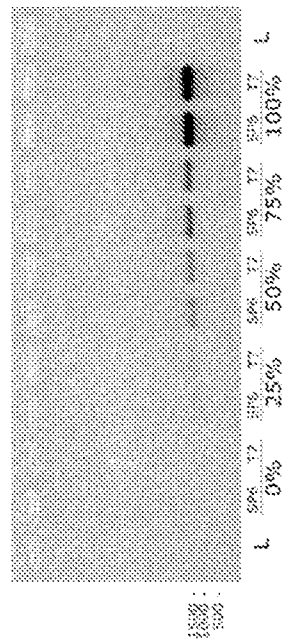
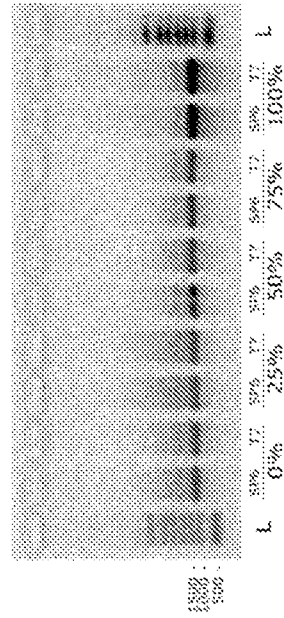


Figura 4b

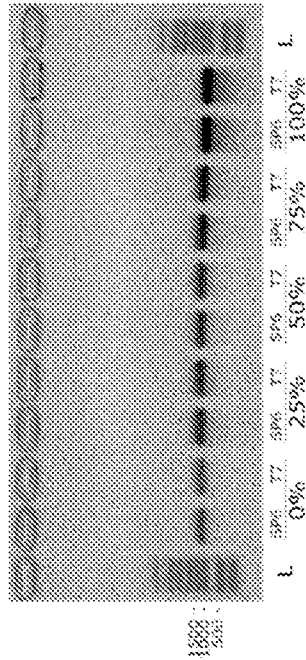


Figura 5a

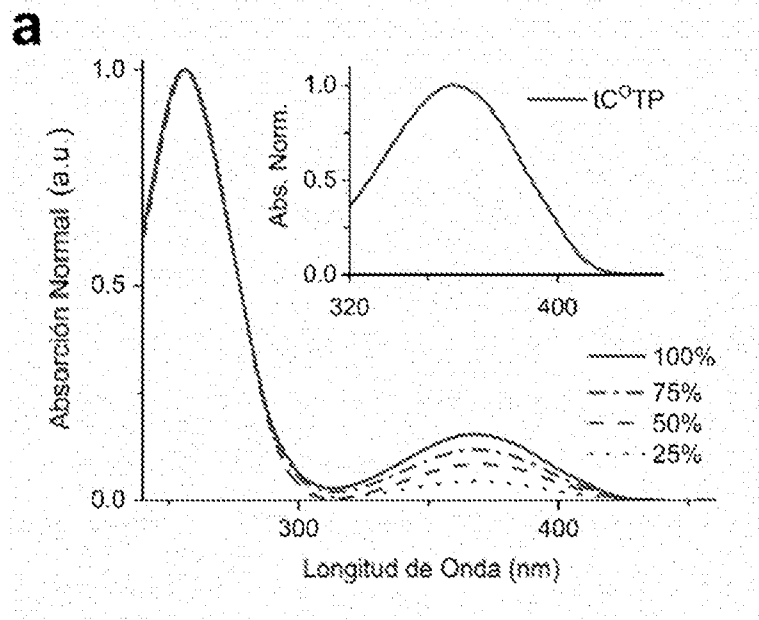


Figura 5b

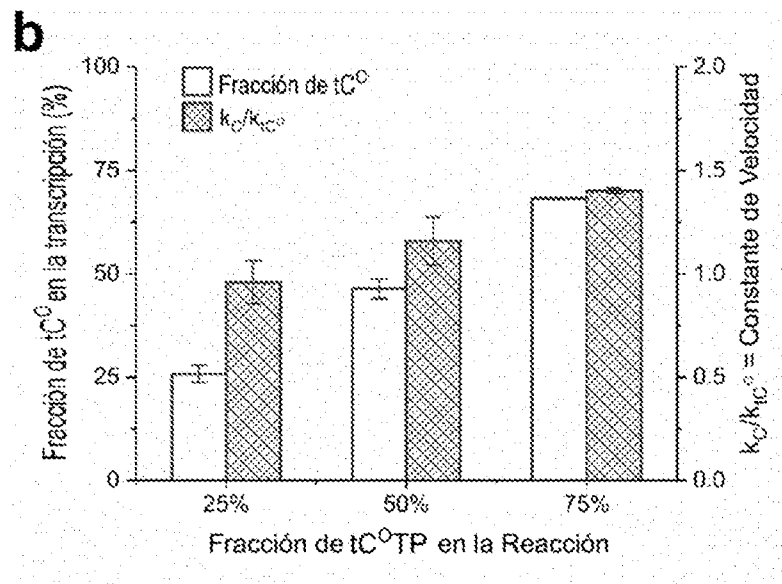


Figura 5c

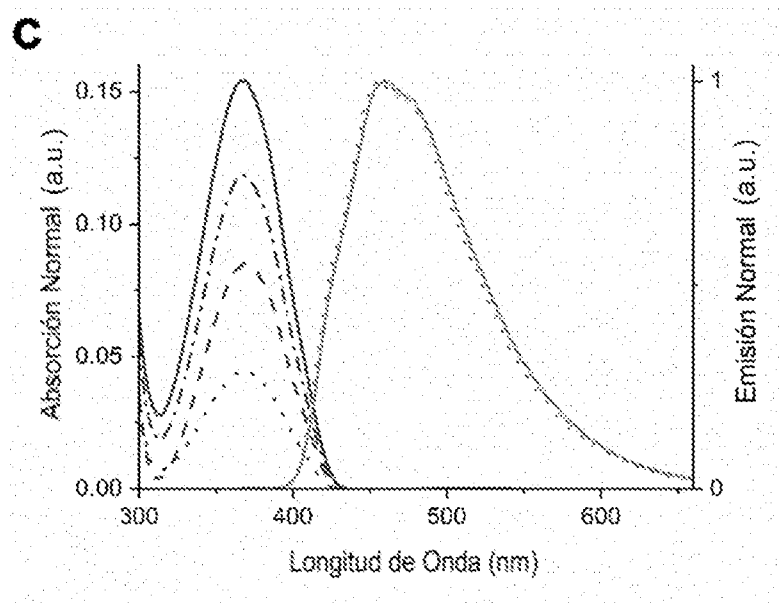


Figura 5d

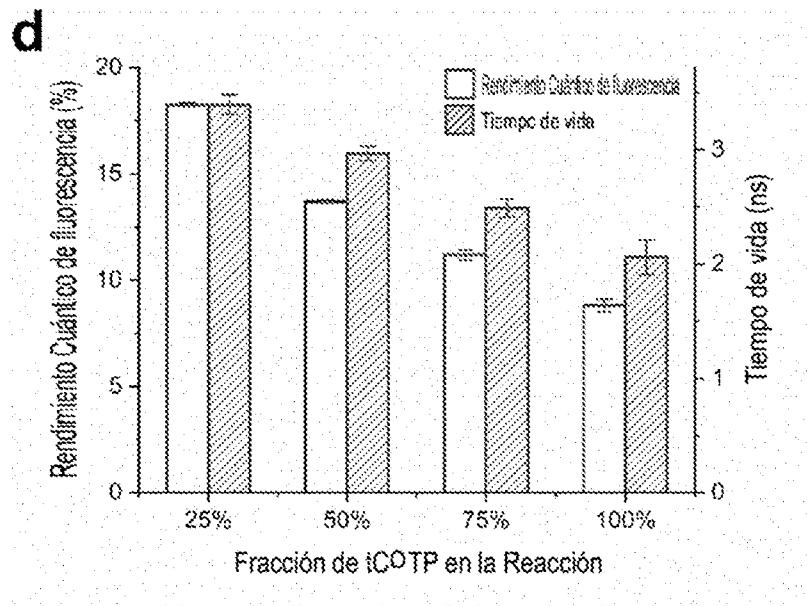


Figura 6a

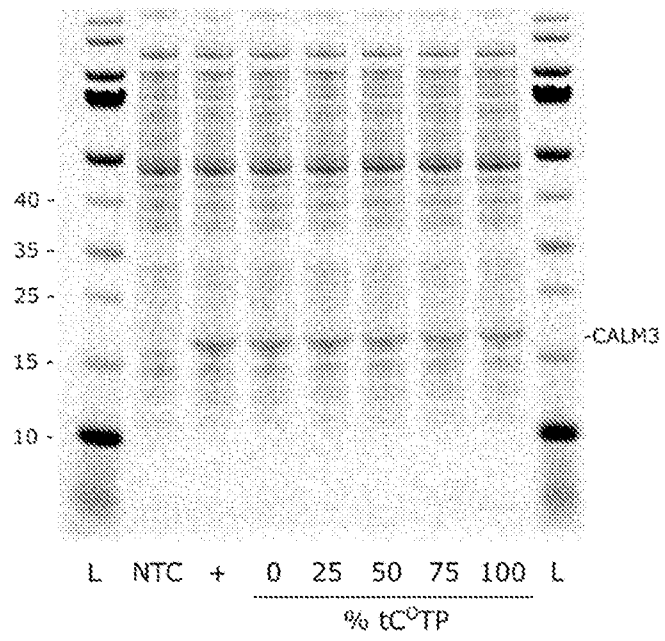


Figure 6b

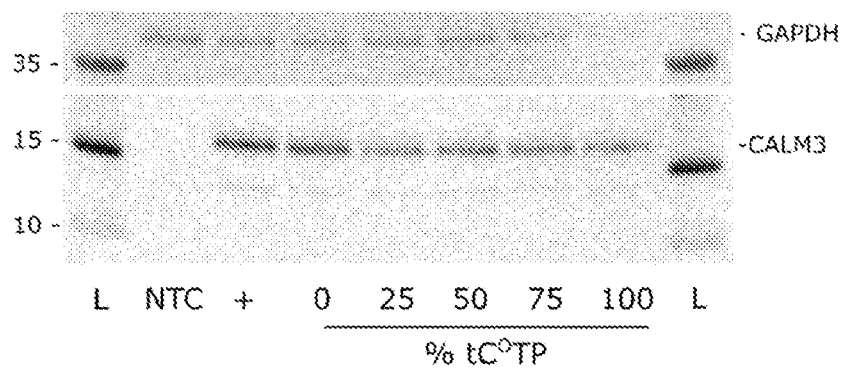


Figura 6c

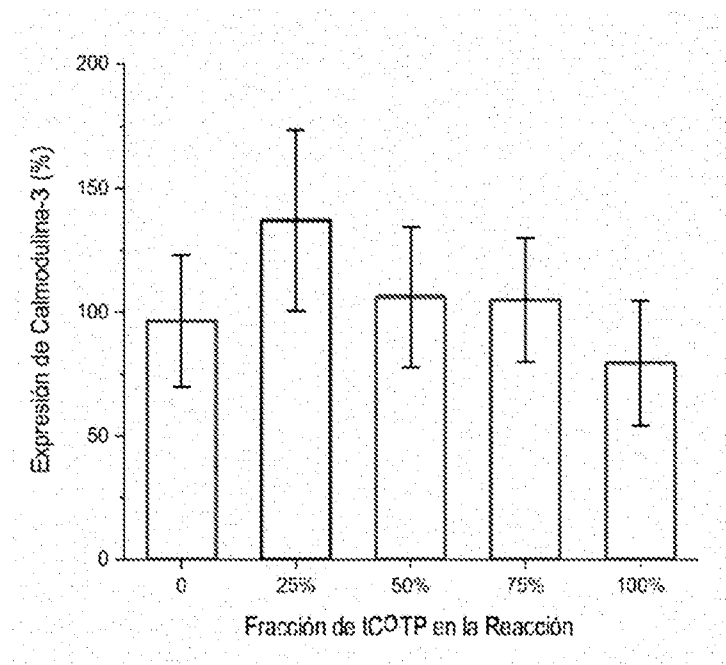


Figura 7a

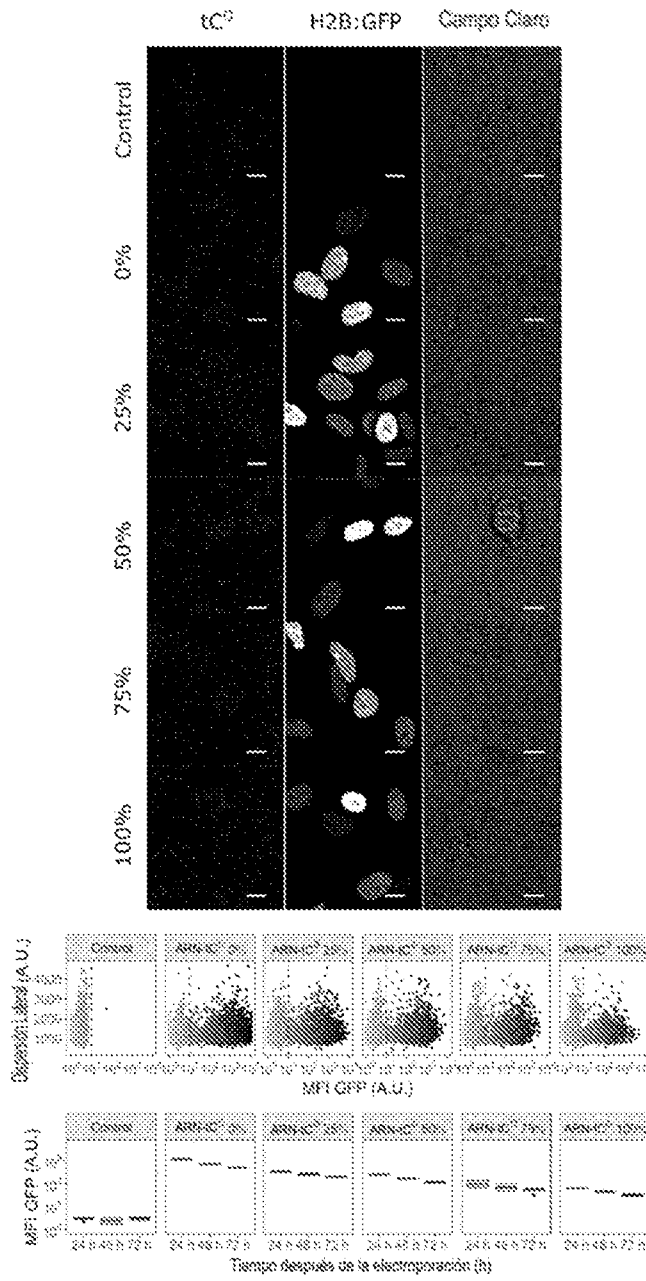


Figura 7b

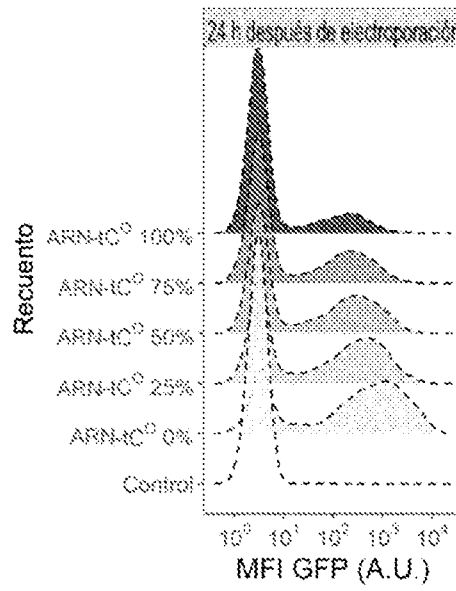


Figura 7c

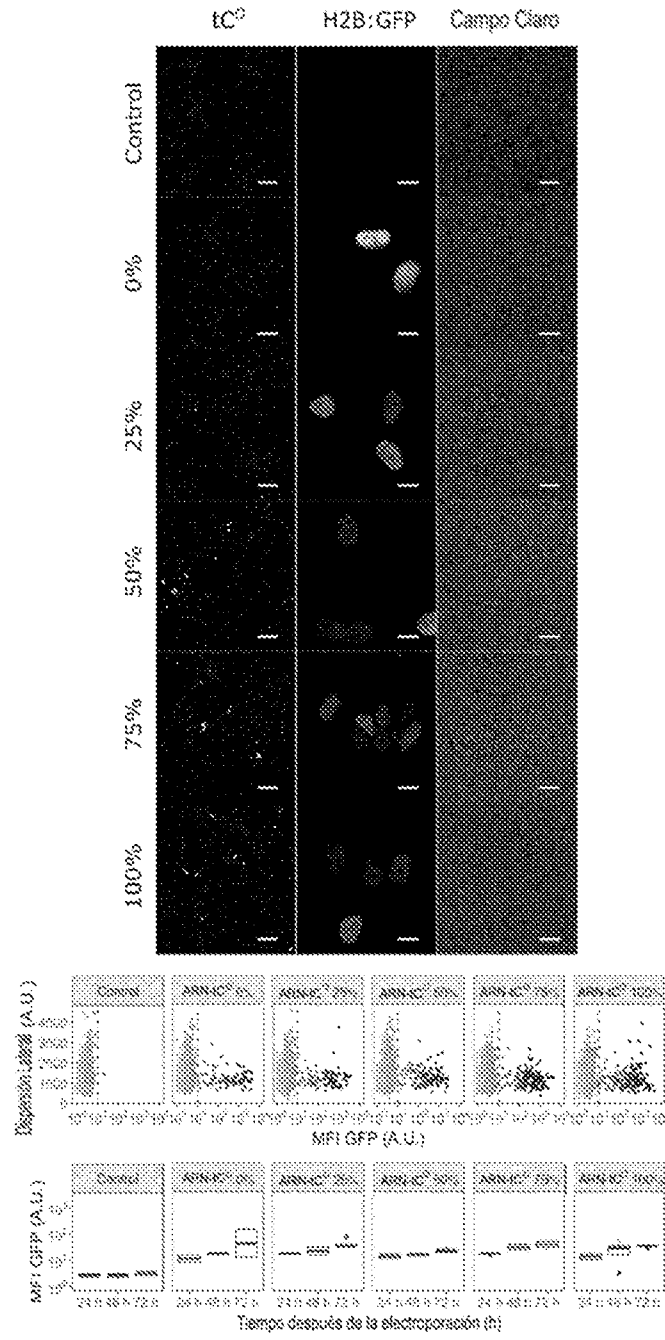


Figure 7d

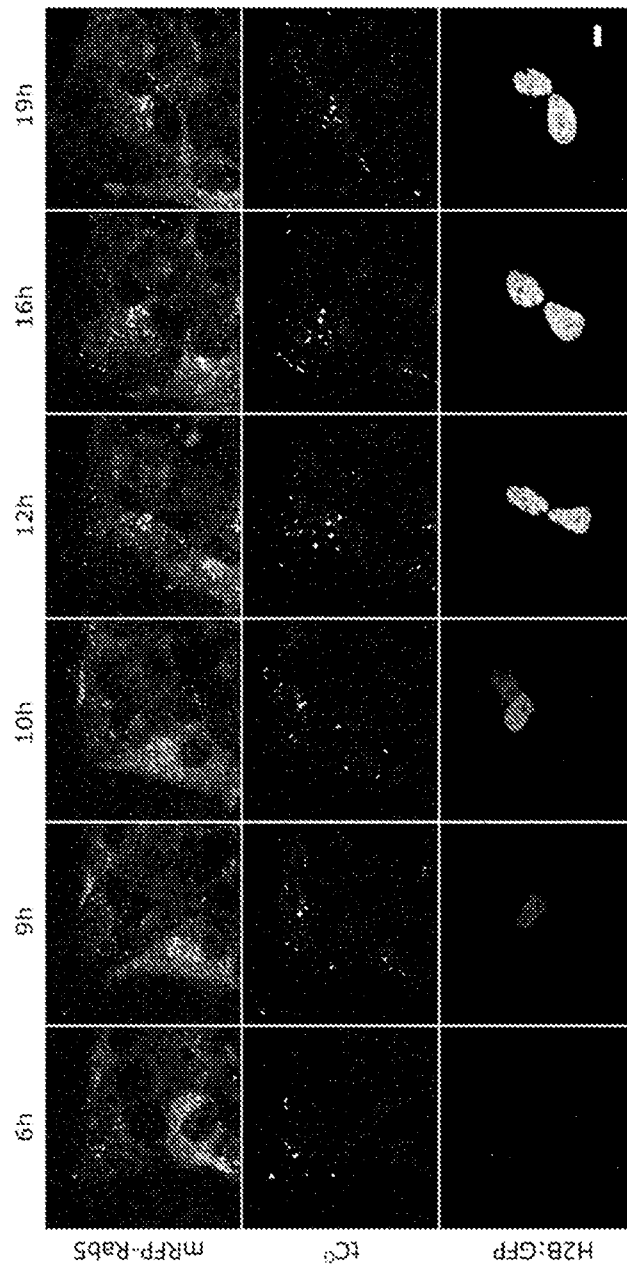


Figura 7e

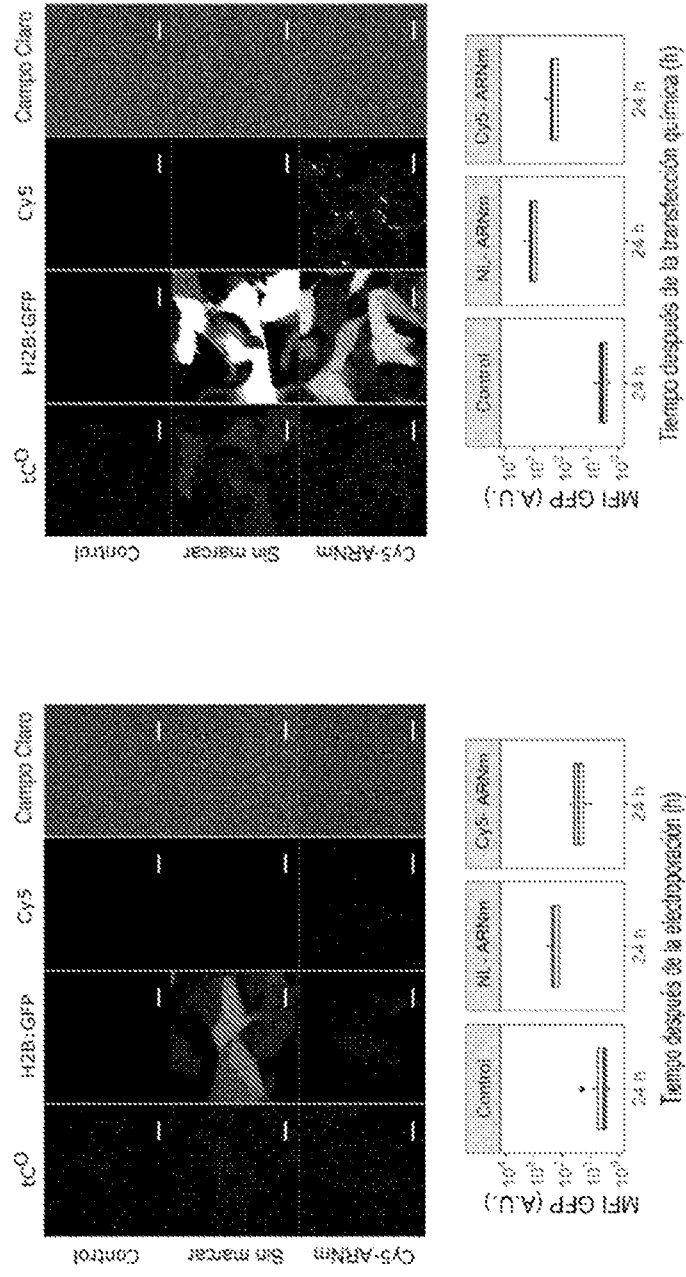
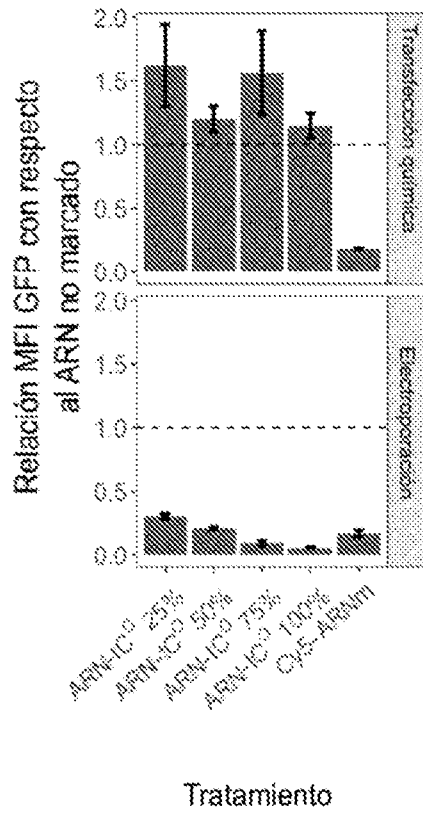
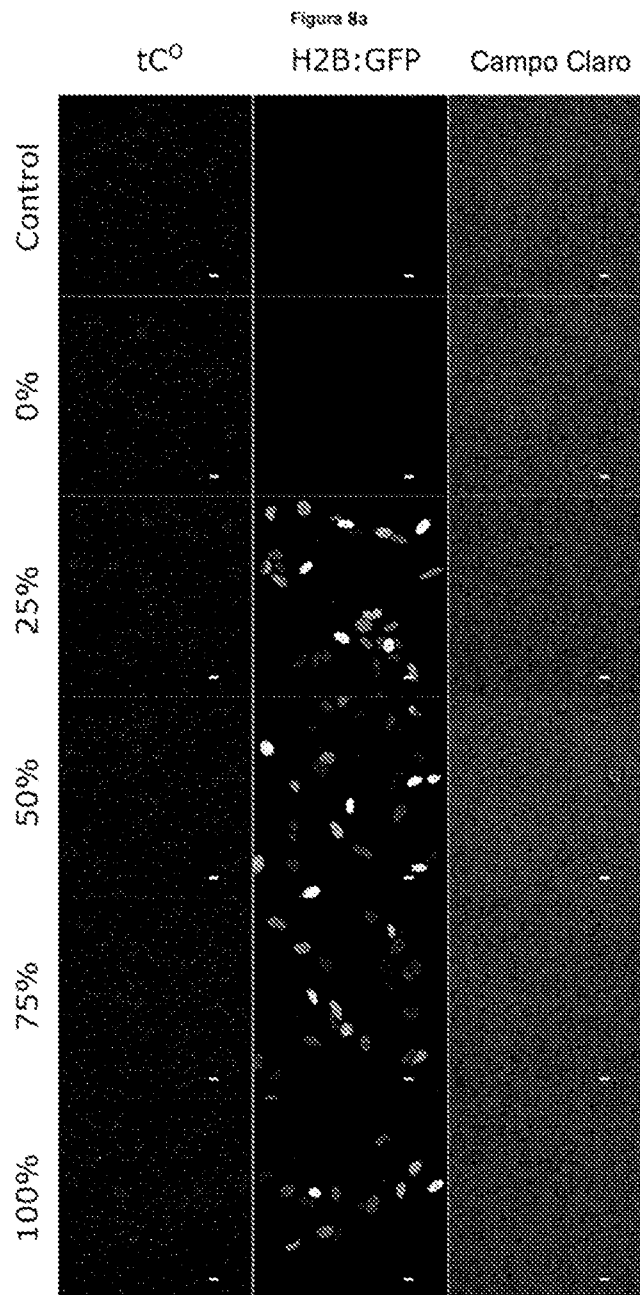


Figura 7f





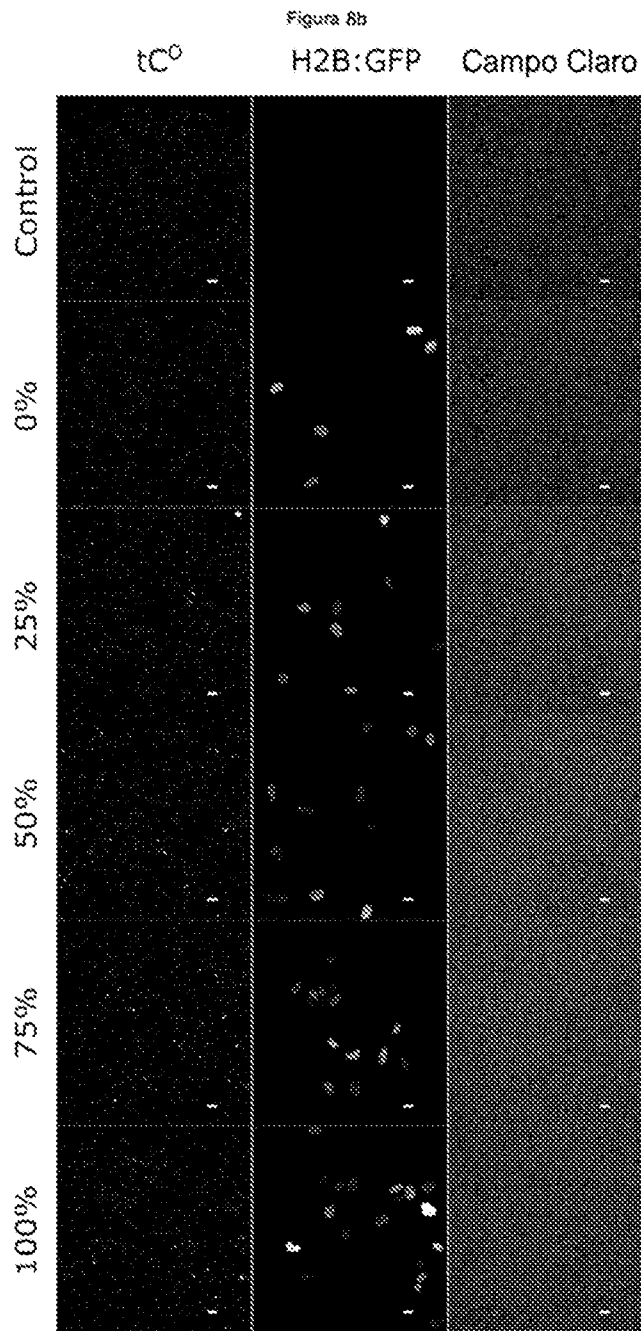


Figura 8c

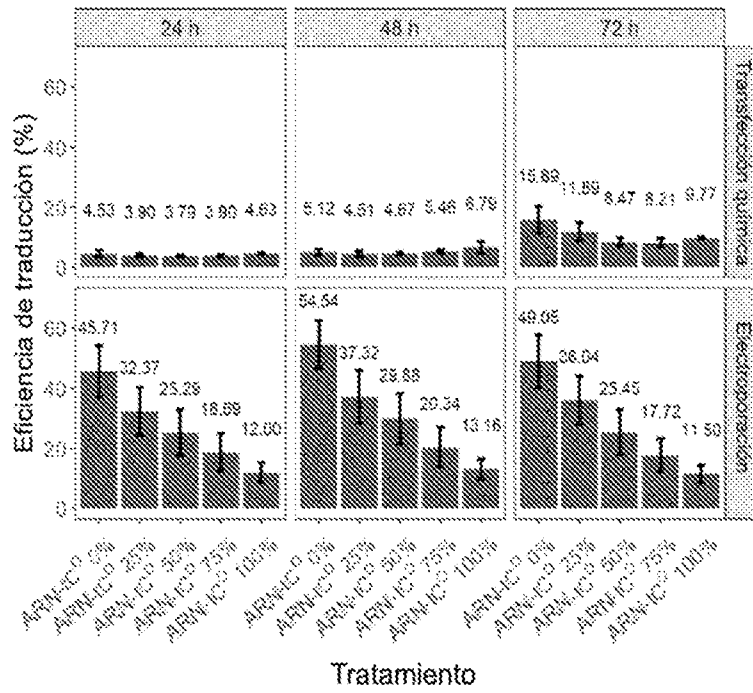


Figura 8d

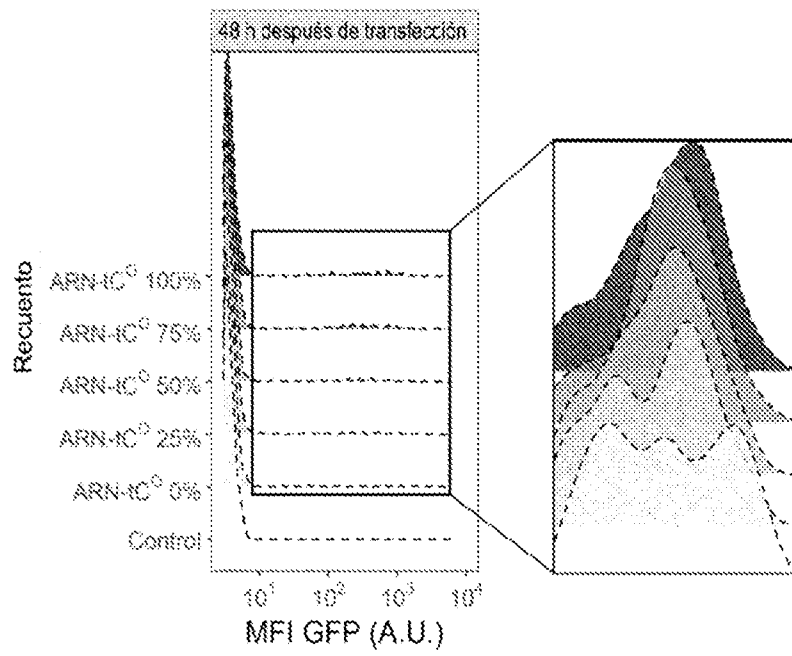


Figura 8e

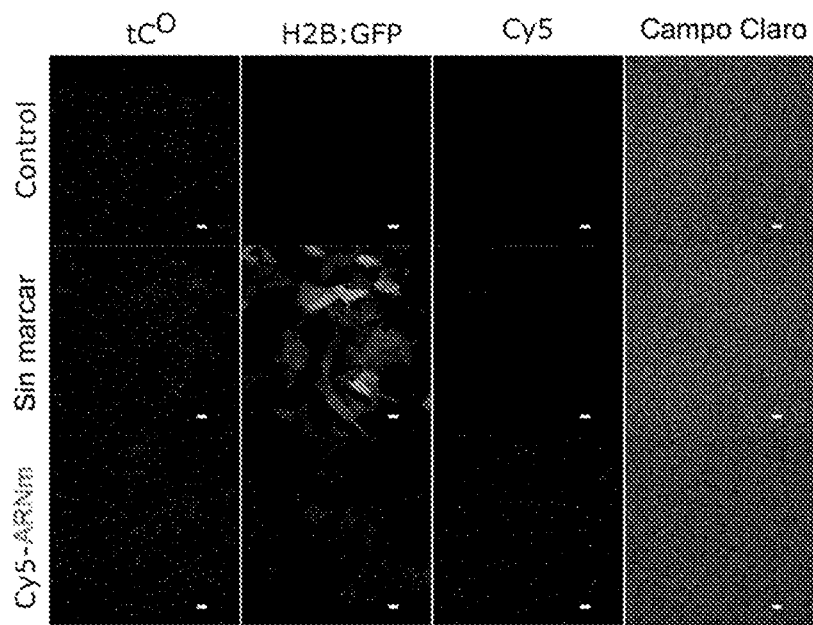


Figura 8f

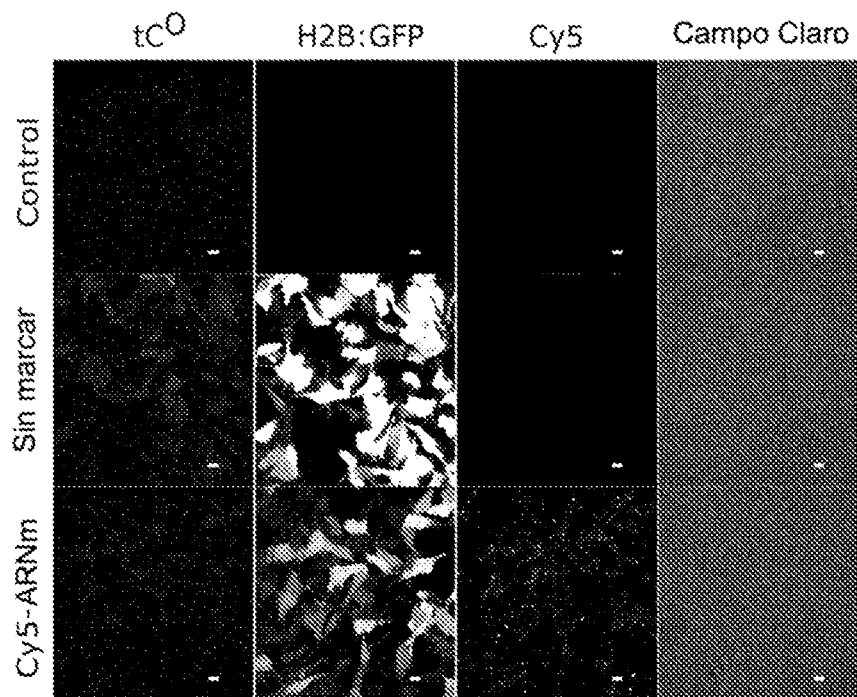


Figura 8g

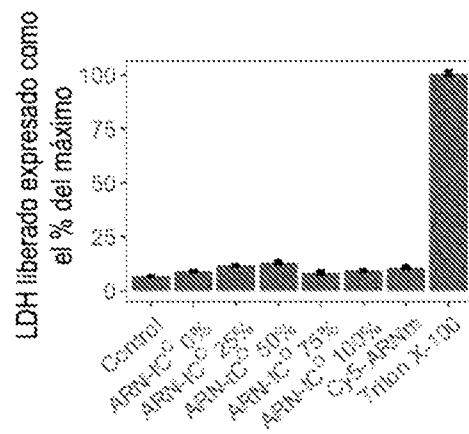


Figura 8h

