(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 733 030**

(51) Int. CI.:

C07D 235/18 (2006.01)
C07D 403/14 (2006.01)
C07H 11/00 (2006.01)
C40B 40/08 (2006.01)
G06E 1/00 (2006.01)
G06N 3/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **08.08.2014 PCT/US2014/050399**

(87) Fecha y número de publicación internacional: **12.02.2015 WO15021415**

(96) Fecha de presentación y número de la solicitud europea: **08.08.2014 E 14752771 (7)**

(97) Fecha y número de publicación de la concesión europea: **27.03.2019 EP 3030548**

(54) Título: **Diseño de transcriptoma completo de moléculas pequeñas bioactivas, selectivas que seleccionan como diana ARN**

(30) Prioridad:

09.08.2013 US 201361864081 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.11.2019

(73) Titular/es:

THE SCRIPPS RESEARCH INSTITUTE (100.0%)
10550 North Torrey Pines Road
La Jolla, CA 92037, US

(72) Inventor/es:

DISNEY, MATTHEW D. y
VELAGAPUDI, SAI

(74) Agente/Representante:

ZUAZO ARALUZE, Alexander

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

Diseño de transcriptoma completo de moléculas pequeñas bioactivas, selectivas que seleccionan como diana ARN

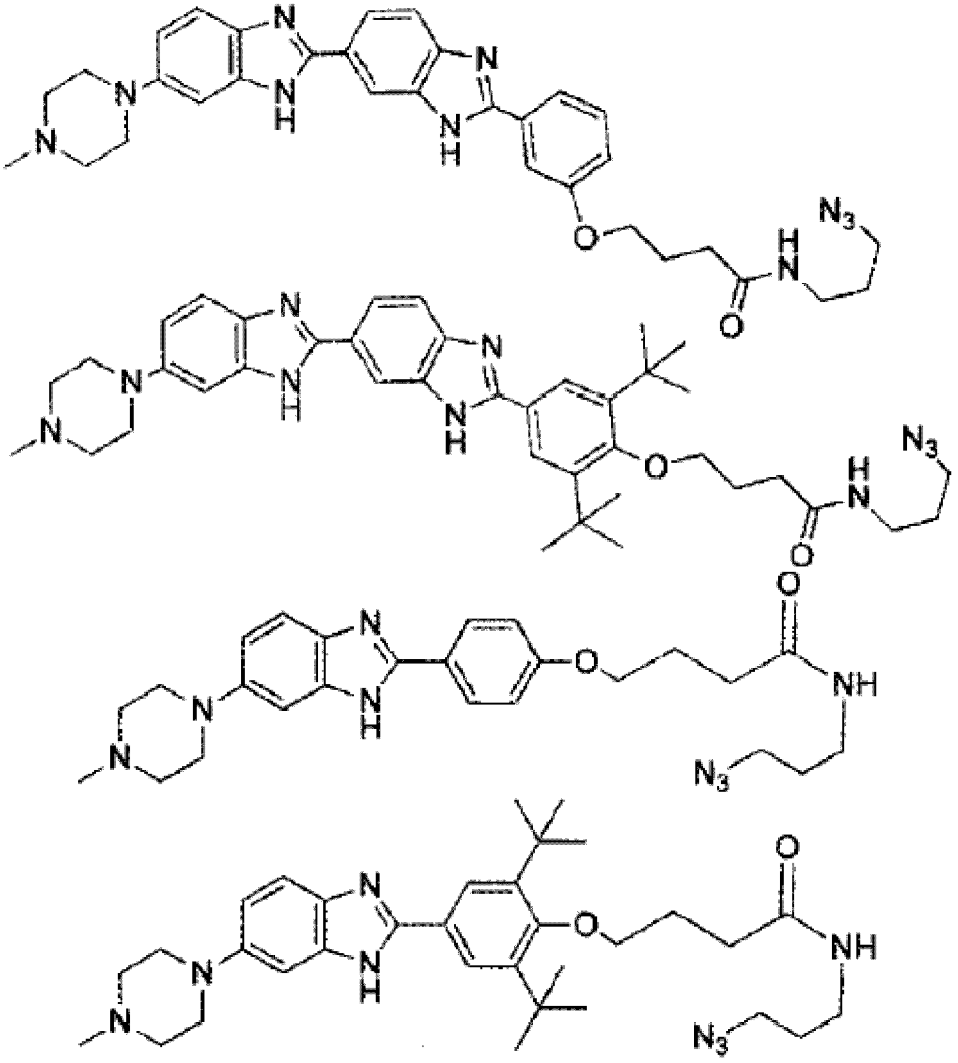
Campo de la invención

Esta invención se refiere a compuestos y a composiciones para su uso en el tratamiento de cáncer en un sujeto.

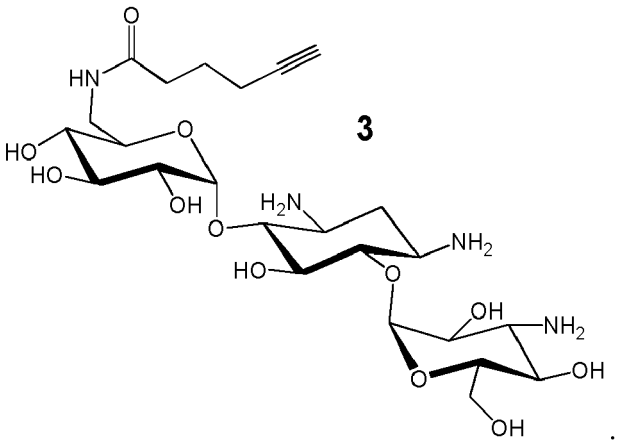
Antecedentes de la invención

El ARN tiene funciones celulares esenciales y es una diana muy deseable para moduladores de molécula pequeña de función. El desarrollo de compuestos bioactivos que seleccionan como diana ARN constituye un desafío, sin embargo, debido a la percepción de que el ARN no tiene “capacidad farmacológica” (Guan & Disney, ASC Chem Biol 7: 73-86 (2012); Thomas & Hergenrother, Chem Rev 108: 1171-1224 (2008)). La falta de éxito en esta área puede atribuirse a una falta fundamental de comprensión sobre los motivos estructurales secundarios del ARN que son los sitios de unión preferidos de moléculas pequeñas y sobre los tipos de moléculas pequeñas que se unen a motivos de ARN con alta afinidad y especificidad. Si pudieran diseñarse de manera adecuada moléculas pequeñas para seleccionar como diana el ARN, podrían diseñarse agentes terapéuticos más eficaces, al igual que estudios de unión de agentes antibacterianos al ribosoma fueron útiles para dilucidar las complejidades de la maquinaria traduccional (Yoshizawa *et al.*, Science 285: 1722-25 (1999): Carter *et al.*, Nature 407: 340-348. (2000)).

Velagapudi *et al.*, Defining the RNA Internal Loops Preferred by Bencimidazole Derivatives via 2D Combinatorial Screening and Computational Analysis, J. Am. Chem. Soc., 2011, 133 (26), págs. 10111-10118 describen la identificación de los bucles internos de ARN derivados de una biblioteca de bucles de nucleótidos 3x3 de 4096 miembros que son los agentes de unión más específicos y con la mayor afinidad a una serie de cuatro bencimidazoles de tipo fármaco, de diseño

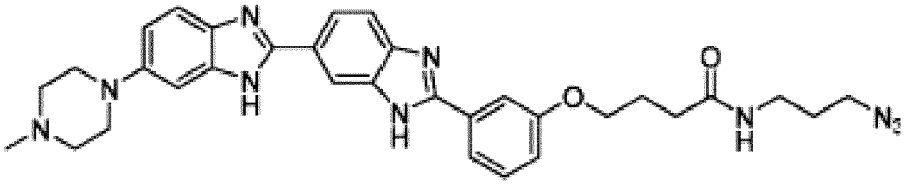


Lee *et al.*, Influencing uptake and localization of aminoglycoside-functionalized peptides, Mol. Biosyst, 2011, 7, 2441-1451 describen peptoides ensamblados de manera modular que portan como carga 6'-N-5-hexinoato de kanamicina A (compuesto 3 a continuación) que se localiza en el citoplasma y la región perinuclear de líneas celulares adherentes C2C12 y A 549 y en la totalidad de células HeLa [véase también el documento WO 2008/103489 (Disney)].



3

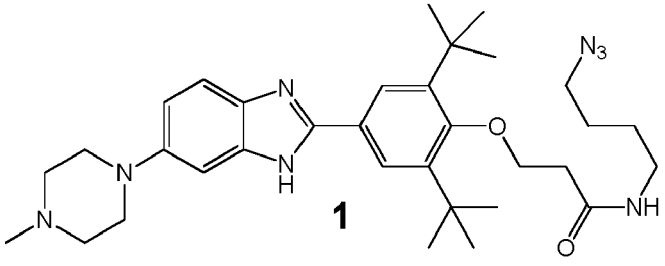
Guan *et al.*, Recent Advances in Developing Small Molecules Targeting RNA, ACS Chem. Biol. 2012, 7, 73-8 analizan trastornos de diferentes repeticiones de nucleótidos expandidas incluyendo repeticiones de r(CUG) y r-(CCUG) expandidas. Se notifica que el compuesto 3 anterior inhibe la interacción r(CUG)-MBNL1 con una CI50 en el rango micromolar y se prefiere que se una a bucles internos ricos en pirimidina tales como los presentes en los ARN que producen DM1 y DM2. Se notifica además un derivado de Hoechst que presenta azida y de fórmula



El derivado anterior se ensambló de manera modular en una estructura principal peptoide para proporcionar una estructura general nH-m en la que n indica el número de residuos de derivados de Hoechst en la molécula y m representa el número de moléculas de propilamina de separación, siendo la separación óptima cuatro moléculas de propilamina que proporcionan 2H-4. Se notificó que 2H-4 se une a repeticiones de r(CUG) con una afinidad >10 veces mayor y con una especificidad >10 veces mayor que MBNL1. Estas propiedades de unión se consideraron importantes, ya que una molécula pequeña tiene que competir con MBNL1 para unirse a repeticiones de r(CUG), es decir los ligandos deben ser o bien de mayor afinidad que MBNL1 o bien deben presentar una permeabilidad celular suficiente como para estar presentes a concentraciones mucho mayores que MBNL1 en un sistema biológico.

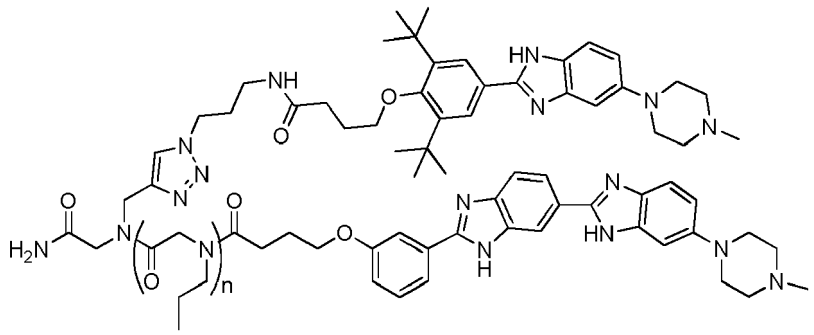
Sumario de la invención

La invención proporciona el siguiente compuesto:



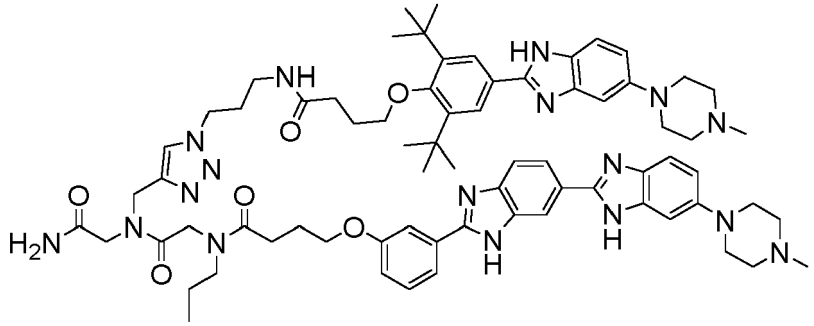
1

y también compuestos de fórmula:

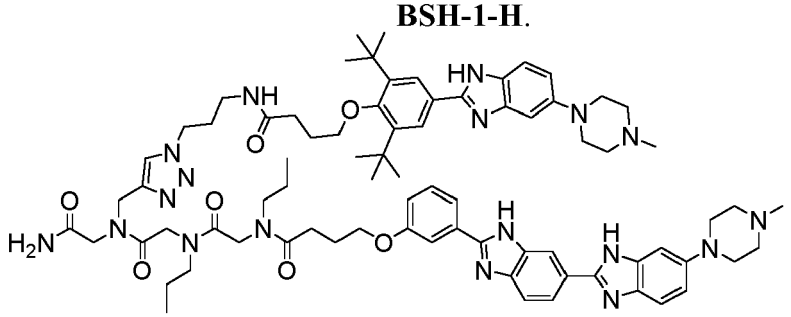


BSH-n-H,

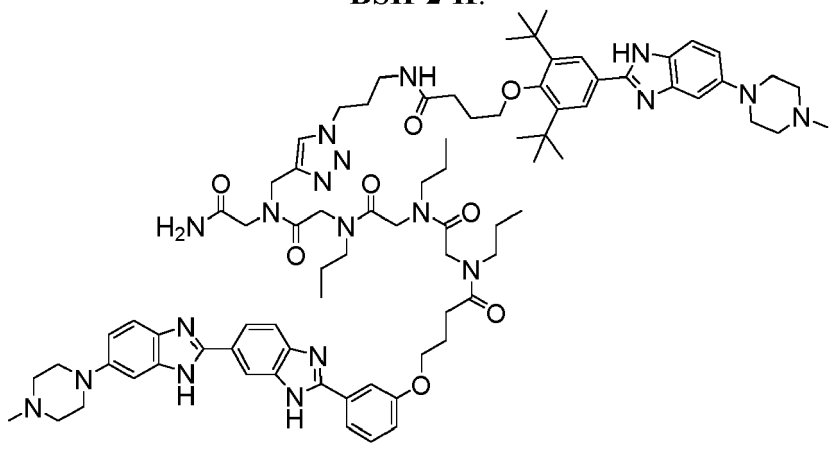
en la que n es un número entero desde 1 hasta 10, o cualquier combinación de los mismos, y los siguientes compuestos individuales:



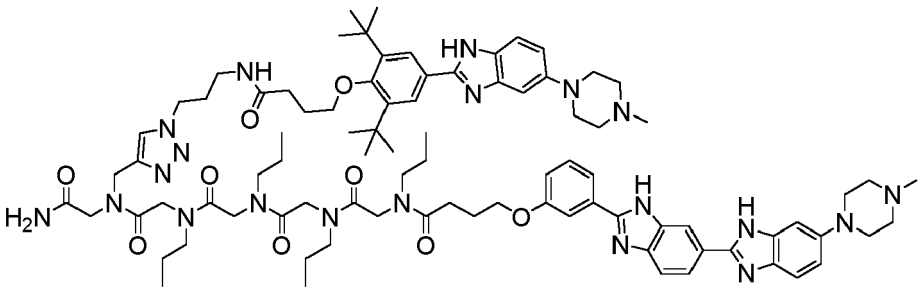
BSH-1-H.



BSH-2-H.

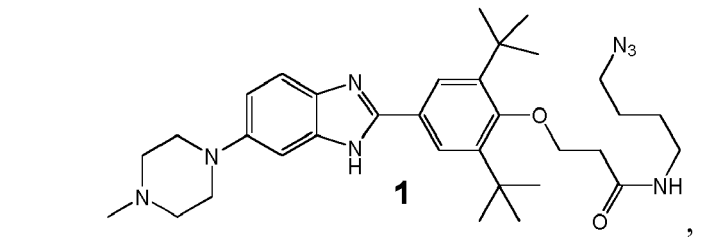


BSH-3-H

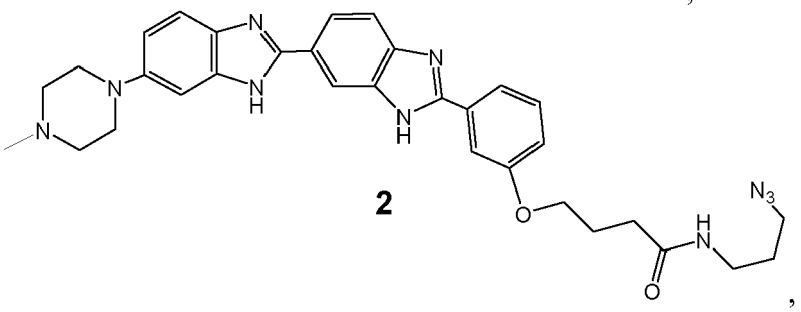


BSH-4-H.

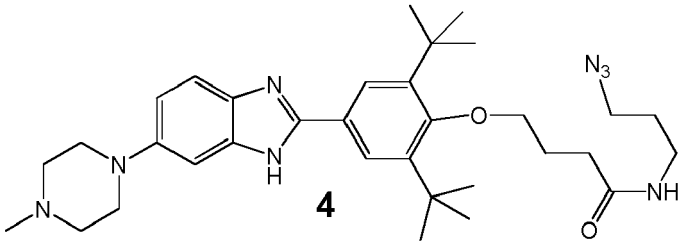
La invención proporciona además un compuesto para su uso en el tratamiento de cáncer en un sujeto, en el que dicho compuesto es cualquiera de



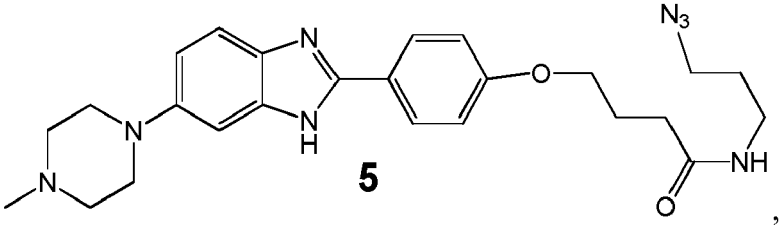
1,



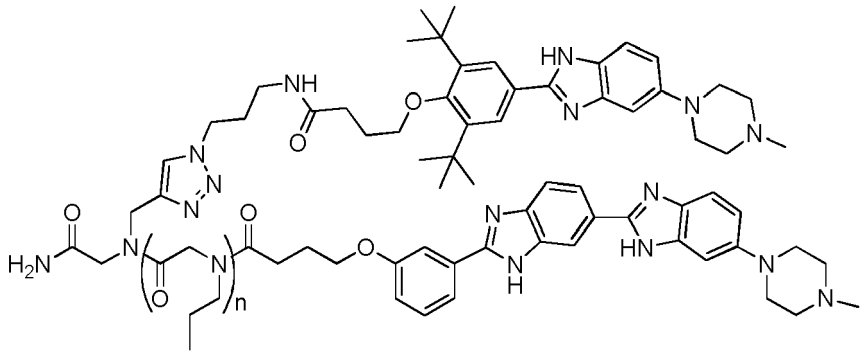
2,



4,



5,



BSH-n-H,

o cualquier combinación de los mismos; en las que n es un número entero desde 1 hasta 10.

Los compuestos anteriores pueden ser para modular la función o actividad de microARN mediante contacto con el microARN que puede ser un pre-microARN-96, un pre-microARN-210 o un pre-microARN-182.

La invención también proporciona una composición que comprende un portador y uno o más de los compuestos anteriores o cualquier combinación de los mismos.

El documento EP-A-3293292 se divide a partir de esta solicitud y reivindica un método para identificar un compuesto que se une a un ARN, que comprende:

(a) comparar un conjunto de datos de consulta de estructuras secundarias de ARN del ARN, con un conjunto de datos de pares de moléculas pequeñas y motivos de ARN unidos identificados, para generar un listado de salida de pares de estructuras secundarias de ARN identificadas y la molécula pequeña identificada que se une a las mismas;

(b) obtener un ARN con una estructura secundaria de ARN identificada y la molécula pequeña identificada que se une a la misma;

(c) determinar una afinidad de unión de la estructura secundaria de ARN identificada con la molécula pequeña identificada que se une a la misma mezclando un ARN que tiene la estructura secundaria de ARN con la molécula pequeña identificada y midiendo la afinidad de unión, y/o determinando si la molécula pequeña identificada reduce o aumenta las cantidades de un ARN que tiene la estructura secundaria de ARN en una célula de mamífero;

para identificar de ese modo un compuesto que se une al ARN.

El documento EP-A-3293292 reivindica además un sistema informático para identificar una molécula que se une a un ARN que comprende: uno o más procesadores y almacenamiento informáticos configurados para comparar un conjunto de datos de consulta estructurado que describe estructuras secundarias de ARN del ARN, y un conjunto de datos estructurado de pares de moléculas pequeñas y motivos de ARN identificados, para identificar de este modo una molécula que se une al ARN y también reivindica uno o más almacenamientos de hardware legibles por ordenador que tienen instrucciones utilizables por ordenador incorporadas en los mismos para realizar un método de comparación de un conjunto de datos de consulta estructurado que describe las estructuras secundarias de ARN del ARN, con un conjunto de datos de pares de moléculas pequeñas y motivos de ARN identificados para identificar una molécula que se une al ARN.

En la siguiente descripción, el contenido de la solicitud divisional se mantiene con fines de referencia y no forma parte de la presente invención.

Descripción de las figuras

La figura 1 muestra estructuras secundarias del casete de horquilla de ARN C1 (GGGAGAGGGUUUAAUUACGAAAGUAAUUGGAUCCGCAAGG, SEQ ID NO: 1) usado como región de entramado para presentar la biblioteca de bucles internos de nucleótidos 3x3 (BBI 3x3; GGGAGAGGGUUUAAUNNNUACGAAAGUANNN AUUGGAUCCGCAAGG; SEQ ID NO:2) de motivos. También se muestran los oligonucleótidos competidores de ARN C2 (GGGAGA), C3 (GGGUUUAAUUAC; SEQ ID NO: 3), C4 (GUAAUUGGAUCC; SEQ ID NO: 4), C5 (GCAAGG) y C6 (CGCGAAAGCG; SEQ ID NO: 5). Se usaron los desoxoligonucleótidos C7 desoxi$(AT)_{11}$ (SEQ ID NO: 6) y C8 desoxi$(GC)_{11}$ (SEQ ID NO: 7) para impedir interacciones inespecíficas con el ADN.

La figura 2 es un diagrama esquemático y una tabla para el cálculo de las puntuaciones de ΣZ para el bucle interno

de miR-96 y el compuesto 1. Hay 51 motivos diferentes que se muestran en la tabla. La puntuación Z para cada motivo se muestra debajo de la secuencia del motivo. La mayor puntuación Z (8,24) se encuentra en la casilla superior izquierda de esta tabla, y la suma de puntuaciones Z para estos 51 motivos es de 228.

Las figuras 3A-3E ilustran el enfoque Informa para diseñar moléculas pequeñas que seleccionan como diana ARN tal como se aplica a los precursores de microARN (miARN) humanos. La figura 3A ilustra esquemáticamente una base de datos de información sobre pares de unión de moléculas pequeñas - motivos de ARN. La figura 3B ilustra el proceso Informa. Las secuencias de todos los precursores de miARN en el transcriptoma humano se descargaron de miRBase (Griffiths-Jones *et al*., Nucleic Acids Res. 36, D154-158 (2008)) y sus estructuras secundarias se predijeron a través de la estructura de ARN (Mathews *et al*., Proc. Natl Acad Sci. U.S.A. 101, 7287-7292 (2004)). Informa extrajo luego los elementos estructurales secundarios de cada ARN de consulta y se compararon esas estructuras secundarias con una base de datos de interacciones de moléculas pequeñas y motivos de ARN identificadas mediante análisis combinatorio bidimensional (2DCS). Las figuras 3C y 3D ilustran los resultados de minería de Informa de las interacciones de moléculas pequeñas y precursores de miARN. La figura 3C es una representación gráfica de los precursores de miARN para los cuales se predijo que una molécula pequeña se uniría a los sitios de procesamiento enzimático Dicer o Drosha en función de la puntuación de idoneidad de la molécula pequeña. Puntuaciones de idoneidad más altas indican una interacción de afinidad mayor. La figura 3D muestra las estructuras de las moléculas pequeñas 1, 2, 3 y 4 que se unen a sitios de procesamiento en precursores de miARN. El compuesto 5 es estructuralmente similar al compuesto 1 y se empleó en diversos ensayos. La figura 3E muestra las estructuras de las otras moléculas pequeñas indicadas en la figura 3C, incluyendo el compuesto 6 con nombre químico 6'-N-5-hexinoato de neamina (triángulo invertido); ek compuesto 7 con el nombre químico 5-O-(2-azidoetil)-neamina (estrella); el compuesto 8 con nombre químico 6'-azido-tobramicina (hexágono); y el compuesto 9 con nombre químico 5'-azidoneomicina B (rombo).

Las figuras 4A-4J ilustran el análisis de StARTS para compuestos y dianas de miARN mostrando representaciones gráficas de idoneidad, estructuras secundarias de ARN y curvas de unión representativas. La figura 4A ilustra el análisis de StARTS mostrando representaciones gráficas de idoneidad para diversos compuestos que se predice que se unirán a miARN. Se muestra el análisis de StARTS para los compuestos 1 (círculos), 2 (cuadrados) y 3 (triángulos) y todas las posibles dianas de miARN. En cada caso, el compuesto inhibe la biogénesis de su diana de miARN predicha (figura 5B). La figura 4B ilustra el análisis de StARTS para los compuestos 1 (círculos), 2 (cuadrados) y 4 (X) y todas las posibles dianas de miARN. El análisis indica que sólo se espera que el compuesto 1 se una al precursor de miR-96 con alta afinidad y selectividad. Mediante este análisis no se predice que los compuestos 2 y 4 se unan al precursor de miR-96. La figura 4C muestra estructuras secundarias de ARN para las que se midieron las afinidades de moléculas pequeñas. La estructura secundaria de 5'UUU/3'AUA (con la secuencia GGGAGAGGGUUUAA UUUUACGAAAGUAAUAUUGGAUCCGCAAGG; SEQ ID NO: 8) contiene el bucle y los pares de bases de cierre presentes cerca del sitio de procesamiento del pre-microARN-96. La estructura secundaria de 5'CGAUUU/3'GGUAUA (con la secuencia GGGAGAGGGUUUAAUCCGAUUUUACGAAA GUAAUAUGGGAUUGGAUCCGCAAGG: SEQ ID NO: 9) es una región expandida del precursor de miR-96 cerca del sitio de procesamiento Drosha (que contiene dos bucles internos separados por dos pares de bases). La estructura secundaria de 5'GUA/3'UCU (con la secuencia GGGAGAGGGUUUAAUGGUAGUACGAAA GUACUCUCAUUGGAUCCGCAAGG; SEQ ID NO: 10) es el bucle que puede seleccionarse como diana en el precursor de miR-449c. La figura 4D muestra una curva de unión representativa para 1-FI y 5'UUU/3'AUA. La figura 4E muestra una curva de unión representativa para 1-FI y 5'-CGAUUU/3'GGUAUA. La figura 4F muestra una curva de unión representativa para 1-FI a C1. La figura 4G muestra una curva de unión representativa para 2-FI y 5'UUU/3'AUA. La figura 4H muestra una curva de unión representativa para 4-FI y 5'UUU/3'AUA. La figura 4H muestra una curva de unión representativa para 5-FI y 5'UUU/3'AUA. La figura 4J ilustra que los compuestos 2, 4 y 5 (que son químicamente similares a 1) no afectan a la expresión de miR-96, miR-182 o miR-183 según se determina mediante qRT-PCR. Se trataron células MCF7 con 40 μM de los diferentes compuestos durante 20 h. Entonces se aisló el ARN total y se sometió a qRT-PCR.

Las figuras 5A-5C ilustran que las moléculas pequeñas de diseño que seleccionan como diana precursores de miARN identificados a través de Informa son bioactivas y selectivas. Estos compuestos inhiben selectivamente la biogénesis de su miARN precursor diana por las células MCF7 de una manera dependiente de la dosis. La figura 5A muestra las estructuras secundarias de los precursores de miARN estudiados, en los que el precursor de miR-96 tiene la SEQ ID NO: 11; el precursor de miR-182 tiene la SEQ ID NO: 12; el precursor de miR-183 tiene SEQ ID NO: 13; y el precursor de miR-210 tiene la SEQ ID NO: 14. Los compuestos 1, 2 y 3 se unen a las regiones indicadas con recuadros en los precursores de miR-96, miR-182 y miR-210, respectivamente. Debe tenerse en cuenta que miR-96, miR-182 y miR-183 se transcriben como una agrupación (*cluster*). La figura 5B ilustra gráficamente la expresión relativa de microARN-96 (barra derecha), microARN-182 (barra intermedia), microARN-183 (barra derecha), cuando se incuban diversas concentraciones de compuesto 1 o compuesto 3 con células MCF7. El gráfico de la izquierda ilustra gráficamente la expresión relativa de microARN-210 cuando se incuban diversas concentraciones de compuesto 2 con células MCF7. Tal como se muestra, los compuestos de diseño modulan la biogénesis de los microARN en diferentes grados y con selectividades variables. La figura 5C muestra que la selección como diana de oligonucleótidos de miR-96 no es tan selectiva como para 1. Las secuencias de los miARN maduros que se estudiaron se muestran a la izquierda. El miR-96 tiene la secuencia UUUGGCACUAGCACAUUUUUGCU (SEQ ID NO: 15). El miR-182 tiene la secuencia UUUGGCAAUGGUAGAACUCACACU (SEQ ID NO: 16). El miR-183 tiene la

secuencia UAUGGCACUGGUAGAAUUCACU (SEQ ID NO: 17). Las regiones subrayadas indican sitios de unión a oligonucleótidos diseñados; los apareamientos erróneos están dentro de recuadros. El gráfico de la derecha ilustra gráficamente los niveles de ARN cuantitativos de miR-96, miR-182 y miR-183 después del tratamiento con un oligonucleótido de LNA que selecciona como diana miR-96. Tal como se ilustra, los oligonucleótidos de LNA son incapaces de discriminar entre miR-96 y miR-182 en cualquier concentración sometida a prueba. En cambio, el compuesto 1 silencia selectivamente miR-96 (figura 5B). El símbolo "*" indica un valor de p <0,05 mientras que "**" indica un valor de p <0,01 determinado mediante una prueba de la t de Student bilateral.

Las figuras 6A-6E ilustran los efectos de las moléculas pequeñas sobre dianas posteriores de miR-96. La figura 6A ilustra gráficamente el efecto de los compuestos 1, 2, 4 ó 5 sobre la expresión de FOXO1 según se evalúa mediante un sistema modelo de luciferasa, que está regulado negativamente por miR-96. Los compuestos inhiben la maduración de miR-96 en células MCF7 y aumentan la actividad luciferasa de forma dependiente de la dosis (barras de color oscuro). Por ejemplo, el compuesto 1 (40 μM) aumenta la actividad luciferasa en ~2,2 veces. No se observa ningún efecto sobre la expresión de luciferasa cuando la región semilla de miR-96 se muta de manera que no responde al microARN-96 (barras de color claro). En cambio, los compuestos 2, 4 y 5 no tienen efecto sobre la expresión de luciferasa. La figura 6B ilustra gráficamente que el compuesto 1 aumenta la expresión de la proteína FOXO1 endógena tal como se determina mediante inmunotransferencia de tipo Western (inserción). La figura 6C ilustra gráficamente que el compuesto 1 induce apoptosis tal como se determina mediante los ensayos TUNEL y de anexina V/PI (véase también la figura 8) al aumentar la expresión de FOXO1. Tal como se ilustra, el compuesto 1 estimula la apoptosis cuando se tratan células MCF7 con compuesto 1 40 μM, mientras que los compuestos 2, 4 y 5 no estimulan (también a 40 μM) la apoptosis. La figura 6D ilustra gráficamente el efecto de diferentes concentraciones de compuesto 1 sobre el porcentaje de células que expresan anexina (sección superior de cada barra), el porcentaje de células sanas (segunda sección desde la parte superior de cada barra), el porcentaje de células que expresan anexina y tinción con yoduro de propidio (la tercera sección desde la parte inferior de cada barra) y el porcentaje de células teñidas con yoduro de propidio solo (la sección inferior de cada barra). La tinción con yoduro de propidio es un indicador de muerte celular. La figura 6E confirma que el compuesto 1 induce apoptosis a través del aumento de la expresión de FOXO1. Se usó ARN de interferencia pequeño (ARNip) para atenuar la expresión de FOXO1. Si la expresión de otra proteína se ve afectada por el compuesto 1, entonces el tratamiento todavía debe inducir apoptosis. La capacidad del compuesto 1 para inducir apoptosis se reduce en un 75% con el tratamiento con ARNip de FOXO1. Así, el modo de acción del compuesto 1 es a través de la inhibición de la maduración de premicroARN-96 y la inducción concomitante de la expresión de FOXO1. El símbolo "**" indica un valor de p <0,01 tal como se determina mediante una prueba t de Student bilateral.

Las figuras 7A-7B ilustran gráficamente cómo los microARN asociados a enfermedad se ven afectados por la adición de 40 μM de compuesto 1. La figura 7A es una representación gráfica de idoneidad para el compuesto 1 en el que se indican posibles efectos fuera de diana. Los círculos grandes y sombreados indican sitios Drosha; los círculos grandes y claros indican los sitios Dicer. La figura 7B ilustra que entre 149 microARN, solo la producción de miR-96 maduro se ve afectada significativamente por el compuesto 1. Los símbolos por encima de la curva representan microARN que están activados; y los símbolos por debajo de la curva representan microARN que se inhiben por el compuesto 1. Estos estudios demuestran la selectividad de la molécula pequeña para la diana y la ruta pretendidas.

Las figuras 8A-8B ilustran que la tinción con anexina V/yoduro de propidio y los ensayos TUNEL confirman que el compuesto 1 induce apoptosis y no necrosis. La anexina es un marcador temprano para la apoptosis. La tinción con yoduro de propidio es un indicador de muerte celular. La figura 8A muestra un análisis de cuadrantes de citometría de flujo de células MCF7 teñidas con anexina V/yoduro de propidio, tratadas o no tratadas con compuesto 1, que muestra que el porcentaje de células en el cuadrante 3 aumenta hasta aproximadamente el 66,4% cuando la concentración de compuesto 1 aumenta hasta 40 μM. La figura 8B ilustra gráficamente la cuantificación del análisis de citometría de flujo de células teñidas con anexina V/yoduro de propidio, que muestra el porcentaje de células que se tiñen solo con anexina V (sección superior de cada barra). El número de células teñidas con anexina V aumenta a medida que la concentración de compuesto 1 aumenta, mientras que la concentración de células sanas (segunda sección desde la parte superior de cada barra) disminuye. El número de células que se tiñen tanto con anexina V como con yoduro de propidio también aumenta a medida que la concentración de compuesto 1 aumenta (tercera sección desde la parte superior de cada barra). Las células teñidas solo con yoduro de propidio (un indicador de muerte celular) tienden a disminuir a medida que la concentración de compuesto 1 aumenta (sección inferior de cada barra). El símbolo "**" indica p <0,01 tal como se determina mediante una prueba t de Student bilateral.

Las figuras 9A-9C ilustran que el compuesto 1 se une al sitio de procesamiento Drosha e inhibe la maduración de microARN *in vitro* e *in vivo*. La figura 9A muestra la estructura de pre-microARN-96 (SEQ ID NO: 231) y autorradiogramas representativos que ilustran que G6, U8, U9 y U10 de GGG-pre-microARN-96 están protegidos frente a escisión mediante ARNasa III y T1 por el compuesto 1. El GGG-pre-microARN-96 se marcó radiactivamente y se incubó con ARNasa III en presencia de 0, 1 y 10 μM de compuesto 1 (indicado por el triángulo, con la mayor concentración en el extremo más grande del triángulo). "L" indica una escalera de hidrólisis; el "ARN de control" no se trata con nucleasa o compuesto; y "S1" indica GGG-pre-microARN-96 escindido con la nucleasa S1 (escinde regiones monocatenarias). Los círculos indican nucleótidos que están protegidos frente a la escisión, dentro o adyacente al bucle interno que predice Inforna que se unirá al compuesto 1. La figura 9B muestra un

autorradiograma de gel representativo que ilustra que el compuesto 1 inhibe la escisión de Drosha *in vitro* (izquierda) con la cuantificación de los datos en el gráfico a la derecha. La figura 9C ilustra gráficamente que el compuesto 1 inhibe el procesamiento Drosha de pri-miR-96 *in vivo*. Se trataron las células con compuesto 1 seguido de extracción de ARN total, que se sometió a qRT-PCR. Se observó un aumento de la cantidad de pri-miR-96 (barra más a la derecha) con disminuciones concomitantes de la expresión de pre-miR-96 (barra intermedia) y miR-96 maduro (barra más a la izquierda). El símbolo "**" indica p <0,01 tal como se determina mediante una prueba t de Student bilateral.

La figura 10A ilustra la permeabilidad celular de los compuestos 1, 2, 4 y 5 en células MCF7. Se trataron las células con 10 μM de molécula pequeña durante 20 h. La figura 10B proporciona imágenes de células después del tratamiento con 40 μM de los compuestos indicados 1, 2, 4 ó 5. Se adquirieron imágenes a un aumento de 20X.

Las figuras 11A-11B muestran compuestos usados en los ensayos de unión a microARN descritos en los ejemplos. Las figuras 11A1-11A2 muestran las estructuras de los compuestos usados para estudiar la afinidad de unión. La figura 11B muestra el método de síntesis usado para fabricar producir el compuesto fluorescente 5-FI.

Las figuras 12A-12B muestran las estructuras de los compuestos y los resultados de un ensayo de luciferasa para evaluar si los compuestos inhiben la biogénesis de miR-96. Las figuras 12A1-A8 muestran las estructuras de los compuestos examinados para inhibir la biogénesis de miR-96 usando el sistema modelo de luciferasa. La figura 12B ilustra gráficamente los resultados del ensayo de luciferasa para cada compuesto. Ninguno de estos compuestos es capaz de aumentar la producción de luciferasa, lo que indica que no inhiben la biogénesis de miR-96.

Las figuras 13A-13B muestran diagramas esquemáticos que ilustran aspectos de los métodos descritos en el presente documento. La figura 13A es un diagrama esquemático de las interacciones de moléculas pequeñas y motivos de ARN en el sistema que se ha construido. El sistema puede usarse para guiar el diseño racional de moléculas pequeñas que seleccionan como diana un ARN de interés. La figura 13B es un esquema del flujo de datos para el motor/algoritmo de búsqueda que consulta la base de datos de interacciones de moléculas pequeñas y motivos de ARN.

Las figuras 14A-14B son diagramas esquemáticos que ilustran un método para identificar posibles dianas de ARN para moléculas pequeñas y las moléculas pequeñas que se unen a esas dianas de ARN. La figura 14A es un diagrama de un método que muestra que el proceso de comparar e identificar (por ejemplo, Inforna) puede recibir información de diversos conjuntos de datos y proporcionar una salida de pares de molécula pequeña - motivo de ARN que probablemente se unirán con especificidad. La figura 14B es un diagrama en el que se puede incluir un proceso de predicción (tal como StARTS), por ejemplo, para refinar adicionalmente el proceso Inforna y proporcionar una salida de pares de molécula pequeña - motivo de ARN con una mayor probabilidad de especificidad de unión.

La figura 15 es un diagrama de bloques de una máquina en la forma de ejemplo de un sistema informático.

Las figuras 16A-16C ilustran el diseño de moléculas diméricas que seleccionan como diana miARN-96 precursor (SEQ ID NO: 11). La figura 16A es un diagrama esquemático de un proceso para diseñar moléculas diméricas pequeñas que y ese miARN-96 precursor diana (SEQ ID NO: 11). Los dímeros contienen el compuesto 1 (símbolo circular) y una nueva molécula (H, molécula en forma de rombo) unidos entre sí. Se utilizó Inforna para identificar las moléculas pequeñas de H que se pueden emparejar con el compuesto 1 para seleccionar como diana un sitio de procesamiento Drosha y un bucle interno adyacente en el precursor de horquilla de miARN-96 (SEQ ID NO: 11). Los dímeros de diseño óptimo eran 400 veces más potentes que los compuestos monoméricos que se unen al sitio de procesamiento Drosha. La figura 16B muestra la estructura general de la biblioteca de dímeros peptoides diseñados, en la que se usa el símbolo circular para designar la mitad de compuesto 1 del dímero, y se usa el símbolo en forma de rombo para designar la mitad de compuesto genérico H de la molécula dimérica. La figura 16C muestra la estructura del dímero óptimo, BSH-2-H, que era activo a 50 nM como modulador de la biogénesis del precursor de horquilla microARN-96 celular.

La figura 17 muestra las estructuras secundarias de los ácidos nucleicos usados para evaluar las afinidades de las moléculas diméricas pequeñas identificadas en el examen descrito en el ejemplo 10. C1 es el ARN de casete en el que se insertaron los bucles para realizar ensayos de unión (GGGAGAGGGUUUAAUUACGAAAGUAA UUGGAUCCGCAAGG, SEQ ID NO: 1). El ARN 5'UUU/3'AUA tiene el bucle y los pares de bases de cierre presentes cerca del sitio de procesamiento del precursor de miARN-96. El ARN 5'UUU/3'AUA tiene una región de entramado similar a la región de entramado del género de motivos de BBI 3x3 pero con un bucle 5'UUU/3'AUA y la siguiente secuencia (SEQ ID NO: 227): GGGAGAGGGUUUAAUUUUU ACGAAAGUAAUAAUUGGAUCCGCAAGG. El ARN 5'CGA/3'GGU tiene un bucle cerca de un sitio de procesamiento de Drosha separado de 5'UUU/3'AUA por dos pares de bases. Por tanto, el ARN 5'CGA/3'GGU tiene una secuencia similar al ARN 5'UUU/3'AUA pero con un bucle 5'CGA/3'GGU en lugar del bucle 5'UUU/3'AUA. El ARN 5'CGA/3'GGU tiene la siguiente secuencia (SEQ ID NO: 228): GGGAGAGGGUUUAAUCCGAUACGAAAGUAU GGGAUUGGAUCCGCAAGG. El ARN 5'CGAUU/3'GGUAUA es una molécula de ARN dimérica que contiene ambos bucles 5'UUU/3'AUA y 5'CGA/3'GGU. El ARN 5'CGAUU/3'GGUAUA tiene la siguiente secuencia (SEQ ID NO: 229): GGGAGAGGGUUUAAUCCGAUUUUACGAAAGUAAUAU GGGAUUGGAUCCGCAAGG. La horquilla de ADN tiene

la siguiente secuencia (SEQ ID NO: 230): CGCGAATTCGCGTTTTCGCGAATTCGCG. La horquilla de ADN (horquilla H) es la horquilla de ADN que tiene una alta afinidad hacia H.

La figura 18 ilustra gráficamente las cantidades de miARN-96 maduro, pre-miARN-96 y pri-miARN-96 en presencia de moléculas diméricas diseñadas como una medida de la capacidad de estas moléculas diméricas para modular la biogénesis de precursor de horquilla de microARN-96. Las cantidades de miARN-96 maduro (barras a la izquierda), pre-miARN-96 (barras intermedias) y pri-miARN-96 (barras a la derecha) en las células MCF7 se detectaron mediante qRT-PCR después de incubar las células MCF7 con las moléculas diméricas diseñadas (a 1 μM, 0,5 μM y 0,05 μM) durante 24 horas. Los compuestos diméricos BSH-2-H y BSH-4-H redujeron los niveles de miARN-96 maduro de forma dependiente de la dosis. El compuesto BSH-2-H tuvo el efecto más significativo sobre el pri-miARN 96 (un aumento de aproximadamente 2,3 veces) pero el dímero BSH-4-H también aumentó los niveles de pri-miARN 96. Los monómeros BSH (compuesto 1) y H se sometieron a prueba como controles y no tuvieron efecto a 0,05 μM.

La figura 19 ilustra gráficamente el efecto del compuesto dimérico BSH-2-H sobre la expresión de FOXO1 en células MCF7 tal como se determina mediante inmunotransferencia de tipo Western. Tal como se muestra, BSH-2-H a 50 nM aumenta el nivel de expresión de FOXO1 endógeno en 2,5 veces.

Las figuras 20A-C ilustran gráficamente la apoptosis en los números de células cancerosas cuando las células se incuban con los compuestos diméricos BSH-1-H, BSH-2-H, BSH-3-H y/o BSH-4-H. La figura 20A ilustra gráficamente los efectos de dímeros peptoides 50 nM sobre la apoptosis en células MDA MB 231 (una línea celular de cáncer de mama triple negativa) después de la incubación durante 72 horas. Tal como se muestra, los dímeros peptoides BSH-2-H y BSH-4-H inducen apoptosis en un grado variable tal como se determina mediante un ensayo de anexina V/PI. BSH-2-H induce apoptosis en aproximadamente el 75% de las células, mientras que BSH-4-H induce apoptosis en aproximadamente el 40% de las células. La cantidad de apoptosis se indica mediante la sección superior de color claro de cada barra. La figura 20B ilustra gráficamente los efectos de dímeros peptoides 50 nM sobre la apoptosis en MCF10A (células de la mama sanas) después de la incubación durante 72 horas. Ninguno de los dímeros peptoides de diseño inducen una apoptosis significativa en MCF10A, células de mama humanas sanas en condiciones en las que se induce apoptosis en MDA MB 231. Tal como se ilustra, solo se observan pequeñas cantidades de apoptosis, como lo indica la sección superior de color claro de cada barra. La figura 20C ilustra gráficamente los efectos del dímero BSH-2-H 50 nM sobre la apoptosis en células cancerosas MDA MB 231 en comparación con células MCF10A de mama humanas sanas después de la incubación durante 72 horas. Tal como se ilustra, BSH-2-H induce la apoptosis (sección de color claro y superior de cada barra) en al menos el 70% de las células cancerosas MDA MB 231, pero tiene poco o ningún efecto sobre las células de cáncer de mama sanas.

Descripción detallada

Se describe en el presente documento un método denominado Informa que proporciona un enfoque simplificado para diseñar moléculas pequeñas bioactivas que seleccionan como diana estructuras secundarias de la secuencia de ARN (por ejemplo, motivos estructurales). Una gran cantidad de estudios genómicos y funcionales están proporcionando rápidamente información sobre genes asociados a enfermedad, incluidos ARN no codificantes (Kramer y Cohen, Nat. Rev. Drug Discov. 3, 965-972 (2004). Los métodos Informa proporcionan una ruta agilizada para identificar moléculas pequeñas que seleccionan como diana el producto de ARN de esos genes. Los métodos Informa no solo aceleran el descubrimiento de fármacos, sino que también identifican con mayor precisión candidatos a fármaco que tienen una mayor probabilidad de tener una actividad útil. Por ejemplo, tal como se describe con más detalle a continuación, el uso de los métodos Informa identificó múltiples compuestos bioactivos a partir de solo un pequeño conjunto de datos de interacciones de moléculas pequeñas y motivos de ARN.

Los métodos Informa proporcionados en el presente documento utilizan y comparan conjuntos de datos de información, proporcionando una salida de qué estructuras secundarias estructurales de ARN probablemente se unirán a qué molécula pequeña. Esos conjuntos de datos incluyen (a) un conjunto de datos de estructuras secundarias de ARN que deben consultarse; y (b) un conjunto de datos de interacciones de moléculas pequeñas-motivos de ARN identificadas (por ejemplo, tal como se identifica mediante examen combinatorio bidimensional (2DCS)). En general, el término “motivo” se refiere a una estructura de ARN que ya se ha identificado. El término “estructura secundaria” es un término más general que se refiere a las estructuras que pueden formarse cuando una molécula de ARN se repliega sobre sí misma.

La salida de secuencias de ARN y estructuras secundarias que probablemente se unirán a una molécula pequeña pueden analizarse adicionalmente mediante otros procesos de predicción y mediante ensayos químicos y biológicos (por ejemplo, ensayos de unión). Por ejemplo, se puede utilizar un método estadístico StARTS para refinar más las predicciones. El método estadístico StARTS predice las afinidades y selectividades de las interacciones de moléculas pequeñas-motivos de ARN comparando la tasa de aparición de pequeñas características estructurales (una guanina adyacente a una adenina, por ejemplo) en motivos de ARN seleccionados con su tasa de aparición en toda la biblioteca de ARN. Por tanto, el método StARTS facilita la identificación de cuáles estructuras secundarias y motivos de ARN son los más únicos o distintivos en las poblaciones de moléculas de ARN.

Por ejemplo, las figuras 13 a 14 son diagramas esquemáticos que ilustran cómo pueden realizarse los conjuntos de

datos y los análisis. La etapa del método y los conjuntos de datos que se pueden emplear se describen con más detalle a continuación.

Puede generarse un conjunto de datos de las estructuras secundarias de ARN que van a consultarse a partir de una o más secuencias de ARN solas. Por ejemplo, las estructuras secundarias de ARN pueden identificarse como las estructuras secundarias de menor energía libre formadas por un ARN cuando se repliega sobre sí mismo para formar regiones bicatenarias, así como bucles monocatenarios y "burbujas" con apareamiento erróneo en las regiones bicatenarias. Tales estructuras secundarias de baja energía libre pueden predecirse mediante programas tales como *RNAstructure* (Mathews *et al*., Proc. Natl Acad Sci. U. S. A. 101, 7287-7292 (2004); Mathews *et al*., J. Mol. Biol. 288, 911-940 (1999), que se incorporan específicamente como referencia en el presente documento en su totalidad.

Las regiones de estructura secundaria que tienen más probabilidad de unirse a moléculas pequeñas son aberturas en regiones bicatenarias ("bucles internos"), extremos monocatenarios de moléculas de ARN y otras regiones monocatenarias. Por ejemplo, la figura 1 ilustra algunos tipos de bucles y "burbujas" (bucles internos) en diversos motivos de estructura secundaria de ARN. Las estructuras secundarias y los motivos también pueden incluir regiones de ARN bicatenario, aunque en algunas realizaciones estas regiones de estructura secundaria bicatenaria se excluyen de los conjuntos de datos porque puede ser menos probable que se unan a una molécula pequeña con especificidad.

Puede evaluarse un listado de una o varias secuencias de ARN seleccionadas para proporcionar un conjunto de datos de las estructuras secundarias de ARN. Por ejemplo, el conjunto de datos de consulta de las estructuras secundarias de ARN puede incluir una o más estructuras secundarias de ARN de un solo ARN de interés particular, que podría ser una diana para el diseño de fármacos. Por tanto, el conjunto de datos puede incluir cada estructura secundaria de ARN dentro del ARN seleccionado. En otras situaciones, el conjunto de datos de consulta de las estructuras secundarias de ARN puede generarse a partir de una familia de especies de ARN que tienen funciones relacionadas, o a partir de una serie de especies de ARN (por ejemplo, aquellas que se cree que están involucradas en el inicio, mantenimiento o progresión de la enfermedad). Por ejemplo, el conjunto de datos de consulta de las estructuras secundarias de ARN puede ser de solo una molécula de ARN, o de 1-5 moléculas de ARN, o de 1-10 moléculas de ARN, o de 1-50 moléculas de ARN, o de 1-100 moléculas de ARN, o de 1-500 moléculas de ARN, o de 1-1000 moléculas de ARN, o de 1-5000 moléculas de ARN, o de 1-10.000 moléculas de ARN. El conjunto de datos de consulta de las estructuras secundarias de ARN también puede incluir, por ejemplo, estructuras de solo una molécula de ARN, o de 2-5 moléculas de ARN, o de 2-10 moléculas de ARN, o de 5-50 moléculas de ARN, o de 10-100 moléculas de ARN, o de 20-500 moléculas de ARN, o de 100-1000 moléculas de ARN, o de 5000-5000 moléculas de ARN, o de 100-10.000 moléculas de ARN.

En algunos casos, los métodos descritos en el presente documento pueden emplear una biblioteca de motivos de ARN, o pueden emplear una biblioteca de motivos de ARN como un conjunto de datos de consulta. Por ejemplo, la biblioteca de motivos de ARN puede ser una biblioteca de bucles internos de ARN cuyos miembros difieren entre sí (i) en la identidad de las bases en el bucle interno de ARN y/o (ii) en la identidad de los pares de bases adyacentes al bucle interno de ARN (los llamados pares de bases de cierre de bucle). La biblioteca de motivos de ARN puede ser, por ejemplo, una biblioteca de bucles internos simétrica, una biblioteca de bucles internos asimétrica, una biblioteca de bucles internos 1x1, una biblioteca de bucles internos 1x2, una biblioteca de bucles internos 1x3, una biblioteca de bucles internos 2x2, una biblioteca de bucles internos 2x3, una biblioteca de bucles internos 2x4, una biblioteca de bucles internos 3x3, una biblioteca de bucles internos 3x4, una biblioteca de bucles internos 4x4, una biblioteca de bucles internos 4x5, una biblioteca de bucles internos 5x5, una biblioteca de protuberancias de 1 base, una biblioteca de protuberancias de 2 bases, una biblioteca de protuberancias de 3 bases, una biblioteca de protuberancias de 4 bases, una biblioteca de protuberancias de 5 bases, una biblioteca de bucles de horquilla de 4 bases, una biblioteca de bucles de horquilla de 5 bases, una biblioteca de bucles de horquilla de 6 bases, una biblioteca de bucles de horquilla de 7 bases, una biblioteca de bucles de horquilla de 8 bases, una biblioteca de bucles de horquilla de 9 bases, una biblioteca de bucles de horquilla de 10 bases, una biblioteca de bucles de ramificación múltiple, una biblioteca de pseudonudos, etc. Pueden usarse o evaluar combinaciones de estas y otras bibliotecas de motivos de ARN. Por ejemplo, la biblioteca de motivos de ARN puede tener dos o más bucles internos o terminales, protuberancias, tallos, horquillas u otros elementos estructurales.

Para completar, puede ser deseable emplear o evaluar una biblioteca de motivos de ARN que incluya todas las combinaciones posibles de bases (por ejemplo, una biblioteca de bucles internos 3x3 que contiene 1600 bucles internos 3x3 diferentes). Los miembros de la biblioteca de motivos de ARN pueden incluir además (es decir, además de la región variable de motivo de ARN) regiones de ARN que no varían de un miembro a otro (por ejemplo, regiones de tallo invariantes, regiones de bucle de horquilla invariantes, etc.). Las bibliotecas de motivos de ARN adecuadas se pueden preparar mediante técnicas de transcripción convencionales (por ejemplo, las que emplean la ARN polimerasa T7, tal como se describe, por ejemplo, en Milligan *et al*., "Synthesis of Small RNAs Using T7 RNA Polymerase", Methods Enzymol., 180: 51-62 (1989), que se incorpora en el presente documento mediante referencia a partir de moldes de ADN, tales como moldes de ADN que están disponibles comercialmente en Integrated DNA Technologies (Coralville, Iowa)).

Los ejemplos ilustran los análisis de numerosos microARN humanos (miARN). Los microARN tienen importantes papeles funcionales en la regulación de la transcripción y la traducción. Usando los métodos descritos en el presente documento, pueden identificarse moléculas pequeñas que pueden modular la función de tales microARN. Sin embargo, los métodos también son útiles para modular la función de otros tipos de ARN, tales como pri-miARN, ARNm, ARNt y ARNr. Las moléculas de ARN que se evalúan pueden ser pequeñas o grandes.

La secuencia completa de cada molécula de ARN de consulta puede proporcionarse en el conjunto de datos de las estructuras secundarias de ARN, o la secuencia completa de cada molécula de ARN de consulta puede vincularse a sus estructuras secundarias de ARN de consulta, de modo que pueda identificarse la fuente de las estructuras secundarias dentro del conjunto de datos de consulta.

El conjunto de datos de las interacciones de moléculas pequeñas y motivos de ARN identificadas es distinto del conjunto de datos de las estructuras secundarias de ARN que van a consultarse. El conjunto de datos de interacciones de moléculas pequeñas y motivos de ARN es un conjunto de datos de los motivos de ARN y las moléculas que se sabe que se unen a esos motivos de ARN. Un conjunto de datos de este tipo puede generarse mediante los procedimientos de examen combinatorio bidimensional (2DCS, por sus siglas en inglés) desarrollados previamente por los inventores (véase, por ejemplo, la publicación de solicitud de patente estadounidense 20080188377 de Disney & Childs-Disney; Childs-Disney *et al.*, ACS Chem. Biol. 2, 745-754 (2007); Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008), cada uno de los cuales se incorpora específicamente mediante referencia en el presente documento en su totalidad).

El método 2DCS implica explorar una biblioteca de compuestos con una biblioteca de motivos de ARN y luego identificar qué motivos de ARN se unen a qué compuestos. De este modo, la experimentación por vía húmeda y la manipulación física se realizan en las moléculas pequeñas y en la biblioteca de motivos RNS. La biblioteca de motivos de ARN empleados para 2DCS puede incluir una región de entramado de ARN sintético con al menos una región estructural definida (por ejemplo, una o más regiones de 'tallo' bicatenarias de secuencia conocida), y una región variable (por ejemplo, un bucle monocatenario, o una "burbuja" monocatenaria de secuencia variable que puede ser un ARN de sección no hibridado flanqueado por segmentos de ARN bicatenario). Por ejemplo, la biblioteca de motivos de ARN puede ser una biblioteca de bucles internos de ARN de nucleótidos 3x3, donde las regiones definidas de la molécula de ARN de BBI 3x3 se muestran en la figura 1 junto con dos segmentos de secuencia variable, y donde cada segmento variable tiene una longitud de tres nucleótidos (N-N-N). La figura 2 muestra la biblioteca de combinaciones de motivos de ARN de tres nucleótidos que se usó en algunos de los experimentos descritos en el presente documento. En otro ejemplo, la biblioteca de motivos de ARN puede tener más de un segmento variable o más de un par de segmentos variables.

Para el método 2DCS, la biblioteca de compuestos puede inmovilizarse sobre un microalineamiento de modo que se conozca la dirección de cada compuesto. El alineamiento de compuestos se puede incrustar dentro de un gel (por ejemplo, agarosa o poliacrilamida) para facilitar la localización y el procesamiento de pares de unión de moléculas pequeñas y motivos de ARN. Una mezcla de moléculas de ARN, que incluye una biblioteca completa de motivos de ARN, opcionalmente con cualquier ARN de control deseado, y/o oligonucleótidos competidores se incuba con cada compuesto en el alineamiento. Después de la incubación en condiciones que permiten la unión, las moléculas de ARN unidas por los compuestos se recogen, amplifican y secuencian por separado.

Por ejemplo, en experimentos de 2DCS descritos en el presente documento, se conjugó una biblioteca de moléculas pequeñas sobre una superficie de microalineamiento de agarosa. Luego se analizó con sonda el microalineamiento de moléculas pequeñas con una biblioteca de motivos de ARN pequeños que probablemente se encontrarán como componentes de los ARN de interés. Por ejemplo, el ARN de BBI 3x3 con un bucle interno variable que se muestra en la figura 1 se usó como plataforma para presentar la biblioteca de motivos de ARN pequeños que se muestra en la figura 2. La incubación de los compuestos con la biblioteca de motivos de ARN se puede realizar en presencia de oligonucleótidos competidores, tales como los oligonucleótidos C2-C8 (figura 1) para garantizar que solo las moléculas pequeñas unidas a la región aleatorizada del ARN de BBI 3x3 se conservarán, mientras que las moléculas pequeñas que podrían unirse a regiones comunes a todos los miembros de la biblioteca se unirán a los oligonucleótidos. C2-C8 (véase, Childs-Disney *et al.*, ACS Chem. Biol. 2, 745-754 (2007); Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008), cuyo contenido se incorpora específicamente en el presente documento mediante referencia en su totalidad).

Los ARN unidos a los compuestos inmovilizados se recogen, amplifican y secuencian por separado. Los motivos de ARN en los ARN unidos se identifican y el conjunto de datos de interacciones de moléculas pequeñas y motivos de ARN identificadas se prepara enumerando las moléculas pequeñas y los motivos de ARN que están unidos por la(s) molécula(s) pequeña(s). El conjunto de datos puede incluir información estructural primaria y secundaria para cada motivo de ARN, así como información estructural para cada molécula pequeña (por ejemplo, nombre, fórmula química, estructura química, estructura tridimensional y similares).

StARTS es un enfoque estadístico que puede combinarse con Inforna para evaluar adicionalmente la afinidad de unión de las estructuras secundarias de ARN para la o las parejas de molécula pequeña identificadas mediante Inforna. StARTS identifica características en los motivos de ARN que contribuyen positiva y negativamente a la

unión (véase, Velagapudi *et al*., Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al*., J. Am. Chem. Soc. 133, 10111-10118 (2011); Paul *et al*., Nucleic Acids Res. 37 (17): 5894-5907 (2009), cada uno de los cuales se incorpora en el presente documento mediante referencia en su totalidad).

En el enfoque de StARTS, se compilan las secuencias de una o más estructuras secundarias de ARN identificadas como que se unen a una molécula pequeña, y la tasa de aparición de cada característica de secuencia en las estructuras secundarias de ARN se compara con la tasa de aparición de esa característica en una población más grande de motivos de ARN. Una característica de secuencia es cualquier secuencia corta de ARN (por ejemplo, un tramo de 5'GC) que puede o no ser diferente de las características de secuencia que están presentes en una población más grande de secuencias de ARN. Sin embargo, las características de la secuencia son aquellas secuencias que están presentes en la población de estructuras secundarias de ARN que se unen a una molécula pequeña. Al comparar estas dos poblaciones, se puede calcular el enriquecimiento relativo de una característica específica en la estructura secundaria de ARN para la unión a una molécula pequeña. Por tanto, el método StARTS identifica qué características de secuencia son más prevalentes en una población seleccionada de secuencias de ARN que en una población más grande de secuencias de ARN.

A las características de la secuencia más distintivas se les asigna una significación estadística, o una puntuación Z y un valor de p bilateral correspondiente. Las puntuaciones Z pueden determinarse mediante análisis estadístico utilizando un programa de predictor de espacio privilegiado de ARN (RNA-PSP) que determina qué características aparecen en las estructuras secundarias de ARN seleccionadas con más del 95% de confianza (véase, Paul *et al*., Nucleic Acids Res. 37 (17): 5894-5907 (2009), que se incorpora en el presente documento como referencia en su totalidad). Los intervalos de confianza están asociados con una puntuación Z, donde un valor mayor corresponde a un nivel de confianza más alto. Cada estructura secundaria de ARN puede tener múltiples características que contribuyen a que sea diferente de una población más grande de motivos de ARN y se puede calcular una suma de las puntuaciones Z para todas las características en una estructura secundaria de ARN (ΣZ) como indicador de la distinción estructural total de un motivo de ARN.

Para completar el análisis de StARTS, las puntuaciones Z pueden representarse gráficamente frente a las afinidades de unión medidas de la estructura secundaria de ARN para un compuesto, y esta relación se puede ajustar a una ecuación inversa de primer orden, lo que permite la predicción de la afinidad de un compuesto para un miembro de la biblioteca de ARN.

El programa informático RNA-PSP ha sido desarrollado previamente por los inventores para abordar la necesidad de un análisis estadístico rápido y preciso de los ARN seleccionados. RNA-PSP se desarrolló en una plataforma de Microsoft Visual Basic 2008 y permite la entrada directa de archivos de secuencia de cualquier selección. El archivo de secuencia introducido se analiza para extraer las secuencias de la región variable para cada miembro de la biblioteca seleccionada. Para la extracción automatizada de secuencias seleccionadas, los usuarios especifican las regiones constantes y variables de la biblioteca, lo que permite que RNA-PSP clasifique un archivo de secuenciación e identifique los ARN incorporados de la selección. Véase, Paul *et al*., Nuc. Acids Res. 37 (17): 5894-5907 (2009). Una vez que el programa extrae las estructuras seleccionadas, genera todas las combinaciones posibles de secuencias de la biblioteca original y almacena los resultados. Por ejemplo, en la biblioteca de horquillas de nucleótidos 3x3 que se muestra en la figura 2, solo hay 51 casillas en la tabla, pero 4096 posibles motivos porque cada secuencia de tres nucleótidos que se muestra en la tabla tiene al menos un nucleótido variable.

RNA-PSP clasifica la característica más estadísticamente significativa para las diversas secuencias de motivos de ARN al realizar una prueba Z que genera puntuaciones Z usando las ecuaciones (I) y (II):

$$\varphi = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} \qquad \text{I}$$

$$Z_{obs} = \frac{(p_1 - p_2)}{\sqrt{\varphi(1-\varphi)\left(\left({}^{1}/_{n_1}\right)+\left({}^{1}/_{n_2}\right)\right)}} \qquad \text{II}$$

dónde

n_1 es el tamaño de la población 1 (por ejemplo, la estructura o estructuras secundarias de ARN seleccionadas),

n_2 es el tamaño de la población 2 (por ejemplo, una biblioteca de motivos de ARN),

p1 es la proporción observada de la población 1 que muestra la característica, y

p2 es la proporción observada para la población 2 que muestra la característica.

La salida de RNA-PSP es una puntuación Z_{obs}, y se asigna un valor de p bilateral correspondiente para reflejar un nivel de confianza de que una característica estructural es distinta de las de la población en general. Las puntuaciones Z son las puntuaciones Z_{obs} con límites de confianza de más del 95%. La tabla de la figura 2 muestra

la salida del análisis de RNA-PSP para las 51 características indicadas (puntuaciones Z).

La población 1 puede ser cualquier subconjunto seleccionado de secuencias de ARN, como la salida de un análisis de Inforna, o un subconjunto de la salida de Inforna, tal como una o más estructuras secundarias de ARN que se predice que se unirán a una molécula pequeña, o que parecen contribuir a la unión de una molécula pequeña mediante pruebas por vía húmeda. Una familia o un género de secuencias de ARN relacionadas o motivos que comparten algunas características estructurales se evalúa normalmente para evaluar si las características estructurales comunes contribuyen a la unión con una molécula pequeña. Por ejemplo, la población 1 podría incluir todas las secuencias definidas por la estructura de bucle genérica en la primera casilla de la tabla que se muestra en la figura 2. Esta tabla define un género de estructuras de ARN porque hay un primer segmento de tres nucleótidos (5'ANU) y un segundo de tres nucleótidos (5'UNU) en ese bucle, donde cada uno de estos segmentos de tres nucleótidos tiene un nucleótido variable que puede ser cualquiera de los cuatro ribonucleótidos (A, C, G, U). Así que la primera casilla en la tabla que se muestra en la figura 2 describe un género de bucles, no solo uno. La población 2 en este ejemplo podrían ser todas de las 4096 secuencias posibles en la tabla de la figura 2. Alternativamente, la población 1 podría ser todas las secuencias en la tabla de la figura 2 que se unen a una molécula pequeña específica, según se evalúa mediante Inforna, mientras que la población 2 podría ser todas las secuencias en la figura 2, o todas las secuencias de microARN.

El género de estructuras o motivos secundarios (por ejemplo, los bucles) puede tener una multitud de características de secuencia diferentes y una multitud de puntuaciones Z diferentes. Cuando un género dado contiene la secuencia con la mayor puntuación Z, esa secuencia de puntuación alta y ese género podrían ser una buena diana para la unión a una molécula pequeña. Sin embargo, cada secuencia dentro del género de secuencias (las descritas por la primera casilla en la figura 2) también comparte características estructurales comunes con todas las demás secuencias del género. Por tanto, una suma de todas las puntuaciones Z (ΣZ) para todas las características de secuencia en el género de secuencias puede ser un indicador más útil del potencial de un motivo estructural para la unión específica a una molécula pequeña.

La afinidad de una molécula pequeña para los diversos miembros de la biblioteca de ARN puede esclarecerse cuando las puntuaciones Z se representan gráficamente frente a las afinidades de unión medidas y esta representación gráfica se ajusta a una ecuación inversa de primer orden. Dicho gráfico puede usarse para predecir la afinidad de la molécula pequeña para cualquier miembro de la biblioteca de ARN (véase, por ejemplo, Velagapudi *et al.*, Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011)). La combinación de evaluación estadística y experimental hace de StARTS una herramienta valiosa para aclarar cuáles de las estructuras secundarias identificadas mediante Inforna son las mejores dianas farmacológicas.

Por tanto, StARTS predice las afinidades y selectividades de las interacciones ARN-molécula pequeña al comparar la tasa de aparición de una característica en los motivos de ARN seleccionados (una guanina adyacente a una adenina, por ejemplo) con su tasa de aparición en toda la biblioteca de ARN. A la confianza de que una característica seleccionada no se produjo aleatoriamente se le asigna una puntuación Z y un valor de p bilateral correspondiente. Solo se consideran las características que son estadísticamente significativas ($p \leq 0,05$ o confianza ≥95%). El análisis identifica las características que contribuyen positivamente (puntuación Z positiva) y negativamente (puntuación Z negativa) a la unión, lo que facilita la predicción de qué ARN se unen y qué ARN no.

La figura 13 es un diagrama esquemático de un sistema para el análisis de interacciones de moléculas pequeñas y motivos de ARN que se ha construido. El sistema puede usarse para guiar el diseño racional de moléculas pequeñas que seleccionan como diana un ARN de interés. Las moléculas pequeñas se denominan ligandos en la figura 13. El sistema puede basarse en un conjunto de datos de todas las interacciones de moléculas pequeñas y motivos de ARN identificadas mediante 2DCS o por otros métodos. Dicho conjunto de datos de motivo de ARN-ligando se puede almacenar dentro del sistema junto con el Inforna programa, o el conjunto de datos de moléculas pequeñas y motivos de ARN se puede mantener independientemente de la Inforna programa y accederse o usarse como entrada para Inforna cuando se desee.

El sistema puede incluir una serie de entradas. Por ejemplo, a cada entrada se le pueden asignar los siguientes parámetros, que pueden estar presentes en varias tablas que están vinculadas para búsquedas fáciles: (i) un identificador de ligando único (molécula pequeña); (ii) una estructura secundaria de ARN o un identificador de motivo único; (iii) la estructura secundaria de ARN o el tipo de motivo; (iv) la estructura secundaria de ARN o el tamaño de motivo; (v) la estructura secundaria de ARN o secuencia del motivo; (vi) el/los par(es) de bases de cierre del ARN; (vii) la puntuación de idoneidad para la estructura o el motivo secundario de ARN (que indica la idoneidad global de las interacciones de molécula pequeña y motivo de ARN y está altamente correlacionada con la afinidad de unión); (viii) la constante de disociación, K_d, del par de molécula pequeña-motivo de ARN si se mide; y, (ix) otras notas que incluyen, por ejemplo, el número de referencia de identificación de PubMed de las moléculas de ARN que son la fuente de la(s) estructura(s) secundaria(s) y el/los motivo(s).

La figura 13 se refiere a ID INT, que crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla. El término VARCHAR (#) indica que se puede enumerar una cadena de texto de longitud variable, donde "#" indica el número máximo de caracteres con la cadena.

La tabla de motivos mostrada a la izquierda en la figura 13A tiene una serie de componentes. El ID de motivo (identificador de motivo) en la tabla de motivos asigna a cada motivo de ARN un identificador numérico. El conjunto de datos actual tiene alrededor de 1500 parejas de motivo de ARN-molécula pequeña. El identificador de motivo VARCHAR asigna un nombre a cada motivo.

En la figura 13A, la tabla de motivos define la secuencia completa del motivo de ARN, que incluye:

- Secuencia 5': la secuencia 5' del motivo de ARN con el par de bases de cierre
- Secuencia 3': la secuencia 3' del motivo de ARN con el par de bases de cierre
- Secuencia con par de bases: la secuencia completa del motivo de ARN y los pares de bases de cierre.
- Nucleótidos de bucle: las secuencias de cada motivo de ARN excluyendo los pares de bases de cierre
- Pares de cierre (5', 3'): secuencia de los pares de bases de cierre en 5' y 3'.

La tabla de moléculas pequeñas a la derecha en la figura 13A identifica el nombre de molécula pequeña (también denominada ligando) a la que se une el motivo de ARN.

El componente INT de tamaño de motivo en la tabla de tamaño de motivo de la FIG 13A se vincula con la tabla de motivos para rellenar automáticamente la tabla y asignar un identificador único a cada fila; esto crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.

INT de ID de molécula Pequeña en la tabla a la derecha en la figura 13A se vincula a la tabla de motivos para rellenar automáticamente la tabla, creando una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.

El valor variable de la puntuación de idoneidad en la tabla de motivos a la izquierda en la figura 13A es la idoneidad de un motivo de ARN para unirse a una molécula pequeña específica (ligando). La constante de unión (Kd nM error nm) en la tabla de motivos de la figura 13A es la afinidad de unión o CI_{50} si se determina, y se puede mostrar en la salida. El PMID (VARCHAR) es el número de identificación de PubMed del ARN que es la fuente del/de los motivo(s).

La tabla de tamaño de motivo en el lado derecho de la figura 13 está vinculada a la tabla de motivos. La tabla de tamaño de motivo incluye una función INT de ID de tamaño de motivo que crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla. Tal como se explicó anteriormente, VARCHAR (#) representa cadenas de texto de longitud variable donde "#" indica el número máximo de caracteres con la cadena. Cada tipo de motivo se representa mediante un identificador numérico dentro del ID del tamaño de motivo, y este identificador se anota dentro de esta tabla. El "tamaño de motivo" puede tener diferentes formas funcionales. Por ejemplo, para las horquillas y protuberancias, el tamaño de motivo es simplemente un número que indica el número de nucleótidos en el bucle. La forma funcional para los bucles internos es "AxB", donde A indica el número de nucleótidos no apareados en 5' y B el número de nucleótidos no apareados en 3'. La forma funcional para el tamaño de motivo de un bucle de ramificación múltiple puede tener múltiples formas, tales como "AxBxC" o "AxBxCxD", lo que indica una unión de 3 ó 4 vías, respectivamente.

La tabla de moléculas pequeñas mostrada a la derecha en la figura 13A también tiene una serie de componentes. La tabla de moléculas pequeñas está vinculada a la tabla de motivos tal como se ilustra en la figura 13A. El INT de ID crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla. La ID de molécula pequeña en la tabla de moléculas pequeñas asigna un identificador numérico a cada molécula pequeña a partir de 1. La base de datos actual tiene 24 moléculas pequeñas (véase, por ejemplo, la figura 12). El nombre de molécula pequeña es el nombre asignado a cada molécula pequeña para identificarlo. SMILES, o sistema de entrada de línea de entrada molecular simplificado, es un texto que describe la estructura de la molécula pequeña. El texto de SMILES se puede introducir en varios programas para reconstituir la estructura del ligando. Puede haber una carpeta separada de las estructuras con archivos (por ejemplo, archivos JPEG) que se puede generar para cada búsqueda.

La tabla de motivos que tienen ligandos que se muestra a la derecha en la figura 13A también tiene una serie de componentes. La tabla de motivos que tienen ligandos está vinculada a la tabla de motivos tal como se muestra en la figura 13. La tabla de motivos que tienen ligandos correlaciona el ID del motivo con una identificación de molécula pequeña. La INT de ID crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla. El ID de molécula pequeña es un identificador numérico asignado a cada molécula pequeña a partir de 1. La base de datos actual tiene 24 moléculas pequeñas. El ID de motivo es un identificador numérico para cada tipo de motivo, que se anota en la tabla de motivos.

La figura 13B es un diagrama esquemático que ilustra un ejemplo del flujo de datos de un motor de búsqueda durante los métodos descritos en el presente documento. El software Informa acepta un archivo .CT (un archivo de texto simple que describe la estructura secundaria de un ARN) con dos opciones de búsqueda: nucleótidos de bucle de búsqueda SIN par de bases de cierre y nucleótidos de bucle de búsqueda CON pares de base de cierre. El usuario puede seleccionar un archivo .CT o un archivo zip que contenga varios archivos .CT. Después de elegir el archivo .CT y enviar la opción de búsqueda, el proceso primero llama a una función para analizar el archivo .CT. Esta función aplica un algoritmo de análisis, que crea un parámetro de búsqueda de base de datos. Una vez que se completa la función de análisis de archivos .CT, se crea otro parámetro de búsqueda según la opción de búsqueda seleccionada. Cuando se completan estas funciones, otra función convierte estos parámetros de búsqueda en una instrucción de lenguaje de consulta estructurado (SQL) utilizada para consultar la base de datos. Esta instrucción SQL consiste en los campos en la base de datos que se consultan, las tablas dentro de la base de datos que contienen los campos consultados y los criterios de búsqueda, que filtran el conjunto de resultados según las opciones seleccionadas del usuario.

A cada registro en la base de datos se le asigna un ID de motivo, identificador de motivo, tipo de motivo, tamaño de motivo, par de bases de cierre, puntuación z suma para el motivo, que indica significación estadística y está altamente correlacionada con la afinidad, puntuación de idoneidad, constante de disociación (K_d) si se mide y el ID de referencia de publicación PMID. Hay dos funciones definidas en el sistema que se usan para analizar los archivos CT. Dependiendo de los criterios de búsqueda, el sistema no devolverá coincidencias o un conjunto de registros que coincidan con los criterios de búsqueda. Este conjunto de registros puede hacerse pasar entonces a la interfaz de usuario donde se procesa adicionalmente para aplicar los cambios de formato. Una vez completado esto, este conjunto de registros rellena la cuadrícula de la interfaz de usuario. Los valores se presentan en el siguiente orden: nombre de archivo CT, estructura de compuesto que es una imagen que visualiza el campo SMILES, motivo de consulta, motivo en ARN diana, nucleótidos de bucle, puntuación de idoneidad, identificador de bucle, constante de disociación (K_d) si se mide y el enlace de referencia de publicación PMID. Dado que los resultados de la búsqueda pueden ser bastante grandes, se aplica un límite de 200 filas para reducir la carga del servidor y la demora entre las presentaciones de búsqueda. Si el usuario desea ver todos los registros, está disponible una opción de exportación a Excel. Esta opción no está limitada por el número de registros.

Por tanto, el proceso Informa genera una salida después de comparación/comparaciones/consultas de conjuntos de datos, pudiendo incluir la salida la estructura de la(s) molécula(s) de ARN consultadas (por ejemplo, como un archivo .ct), la estructura de la(s) molécula(s) pequeña(s) que se unen a un motivo identificado en el ARN consultado que incluye un enlace al archivo SMILES correspondiente, el/los motivo(s) dentro del conjunto de datos de motivo de ARN - molécula pequeña que es/son similar(es) o coincide(n) exactamente con el/los motivo(s) en el ARN consultado que se une a la molécula pequeña, el/los motivo(s) en el ARN consultado que se predice que se unirá(n) a un ligando de molécula pequeña, los nucleótidos (de bucle) en el/los motivo(s) del ARN consultado, la puntuación de idoneidad de cada par de molécula pequeña - motivo de ARN recién descubierto, una constante de unión (K_d o CI_{50}), un identificador de bucle para el motivo en la base de datos que es similar o que coincide exactamente con el/los motivo(s) en el ARN consultado (tal como se hace referencia en las publicaciones), y la identificación de PubMed en la que se informa del motivo en la base de datos.

Las funciones o los algoritmos descritos en el presente documento pueden implementarse en software o una combinación de software y procedimientos implementados por un ser humano, por ejemplo. El software puede consistir en instrucciones ejecutables por ordenador almacenadas en medios legibles por ordenador, tales como la memoria u otro tipo de dispositivos de almacenamiento. Además, tales funciones corresponden a módulos, que son software, hardware, firmware o cualquier combinación de los mismos. Pueden realizarse múltiples funciones en uno o más módulos según se desee, y las realizaciones descritas son meramente ejemplos. El software puede ejecutarse en un procesador de señales digitales, ASIC, microprocesador u otro tipo de procesador que funcione en un sistema informático, tal como un ordenador personal, un servidor u otro sistema informático. En una realización, múltiples sistemas informáticos de este tipo se utilizan en una red distribuida para implementar múltiples análisis, obtener información de fuentes distribuidas o facilitar el uso basado en transacciones. Puede usarse una arquitectura orientada a objetos, orientada a servicios u otra arquitectura para implementar tales funciones y comunicarse entre los múltiples sistemas y componentes.

Por ejemplo, el ordenador puede operar en un entorno en red usando una conexión de comunicación para conectarse a una o más ordenadores remotos, tales como servidores de bases de datos. El ordenador remoto puede incluir un ordenador personal (PC), servidor, enrutador, PC de red, un dispositivo del mismo nivel u otro nodo de red común, o similar. La conexión de comunicación puede incluir una red de área local (LAN), una red de área amplia (WAN) u otras redes.

Pueden almacenarse instrucciones legibles por ordenador (por ejemplo, para Informa) en un medio legible por ordenador y pueden ejecutarse por una unidad de procesamiento del ordenador. Un disco duro, CD-ROM y RAM son algunos ejemplos de artículos que incluyen un medio legible por ordenador no transitorio. Por ejemplo, un programa de ordenador vinculado a, o que incluye, los programas Informa puede proporcionar una técnica genérica para realizar una verificación de control de acceso para el acceso a datos y/o para realizar una operación en uno de

los servidores en un sistema basado en un modelo de objetos componentes (COM), o pueden incluirse en un CD-ROM y cargarse desde el CD-ROM en un disco duro. Las instrucciones legibles por ordenador le permiten al ordenador proporcionar controles de acceso genéricos en un sistema de red informática basado en COM que tiene múltiples usuarios y servidores.

La figura 15 es un diagrama de bloques de la máquina en la forma de ejemplo de un sistema 1100 informático dentro del cual pueden ejecutarse instrucciones 1124 para hacer que la máquina realice una o más de las metodologías comentadas en el presente documento. En realizaciones alternativas, la máquina funciona como un dispositivo independiente o puede conectarse (por ejemplo, en red) a otras máquinas. En una implementación en red, la máquina puede operar en la capacidad de un servidor o una máquina cliente en un entorno de red servidor-cliente, o como una máquina del mismo nivel en un entorno de red del mismo nivel (o distribuido). La máquina puede ser un ordenador personal (PC), un PC de tipo tableta, un decodificador (STB), una PDA, un teléfono celular, un dispositivo web, un enrutador de red, un conmutador o puente, o cualquier máquina capaz de ejecutar instrucciones. (secuenciales o de otro tipo) que especifique las acciones que debe realizar esa máquina. Además, aunque solo se ilustra una máquina, el término “máquina” también debe considerarse que incluye cualquier conjunto de máquinas que ejecuten individual o conjuntamente un conjunto (o múltiples conjuntos) de instrucciones para realizar una o más de las metodologías comentadas en el presente documento.

El sistema 1100 informático de ejemplo incluye un procesador 1102 (por ejemplo, una unidad central de procesamiento (CPU), una unidad de procesamiento de gráficos (GPU) o ambas), una memoria 1104 principal y una memoria 1106 estática, que se comunican entre sí a través de un bus 1108. El sistema 1100 informático puede incluir además una unidad 1110 de visualización de vídeo (por ejemplo, una pantalla de cristal líquido (LCD) o un tubo de rayos catódicos (CRT)). El sistema 1100 informático también incluye un dispositivo 1112 de entrada alfanumérico (por ejemplo, un teclado), un dispositivo 1114 de control del cursor (por ejemplo, un dispositivo de navegación de interfaz de usuario (UI) o un ratón de ordenador), una unidad 1116 de disco, un dispositivo 1118 de generación de señales (por ejemplo, un altavoz) y un dispositivo 1120 de interfaz de red.

La unidad 1116 de disco incluye un medio 1122 legible por máquina en el que se almacenan uno o más conjuntos de estructuras de datos e instrucciones 1124 (por ejemplo, software) que se incorporan o utilizan por una o más de las metodologías o funciones descritas en el presente documento. Las instrucciones 1124 también pueden residir, total o al menos parcialmente, dentro de la memoria 1104 principal, la memoria 1106 estática y/o dentro del procesador 1102 durante su ejecución por el sistema 1100 informático, la memoria 1104 principal y el procesador 1102 que también constituyen medios de lectura legibles por máquina.

Las instrucciones 1124 también pueden transmitirse o recibirse a través de una red 1126 de comunicaciones utilizando un medio de transmisión. Las instrucciones 1124 pueden transmitirse utilizando el dispositivo 1120 de interfaz de red y uno cualquiera de una serie de protocolos de transferencia conocidos (por ejemplo, HTTP). Los ejemplos de redes de comunicación incluyen una LAN, una WAN, Internet, redes de telefonía móvil, redes de telefonía convencional (POTS) y redes de datos inalámbricas (por ejemplo, redes Wi-Fi y WiMax). Debe considerarse que el término “medio de transmisión” incluye cualquier medio intangible que sea capaz de almacenar, codificar o transportar instrucciones para la ejecución de la máquina, e incluye señales de comunicaciones digitales o analógicas u otros medios intangibles para facilitar la comunicación de dicho software.

Métodos de tratamiento

Tal como se ilustra en el presente documento, los presentes métodos pueden identificar compuestos que modulan la función del ARN. Por ejemplo, se identificaron varios compuestos que modulan la función de microARN, incluidos los compuestos 1, 2 y 3 (véanse, las figuras 3 y 5; el ejemplo 4). Así, el compuesto 1 reduce el nivel de expresión de miR-96 en un 90% a 40 μM; el compuesto 2 reduce la formación de miR-210 en un 60% a 500 nM; y el compuesto 3 reduce la producción de miR-182 en un 40% a 200 μM. En otro ejemplo, un compuesto dimérico denominado en el presente documento BSH-2-H mostró una selectividad significativa para un ARN que contiene dos sitios de bucle. La molécula dimérica BSH-2-H también tenía una afinidad superior a 30 veces mayor para el ARN de dos sitios que los compuestos monoméricos que componen BSH-2-H. El compuesto 1 constituye la mitad de la molécula BSH-2-H, mientras que la otra mitad era una molécula identificada mediante los experimentos descritos en el ejemplo 10. La incubación del compuesto BSH-2-H con células MCF7 condujo a una reducción significativa de la producción del microARN-96 maduro a una concentración de 50 nM, mientras que también inhibió la producción del pre-microARN-96 y reforzó la producción del pri-microARN-96. Además, el compuesto BSH-2-H a una concentración de 50 nM indujo apoptosis en aproximadamente el 75% de las células de cáncer de mama MDA MB 231, pero no afectó de manera adversa a las células de mama sanas a concentraciones similares.

Tal como se explicó anteriormente, la invención proporciona los compuestos 1, 2, 4, 5 y BSH-nH, donde n es cualquier número entero desde 1 hasta 10, incluidos los compuestos BSH-1-H, BSH-2-H, BSH-3-H y BSH-4-H y cualquier combinación de los mismos.

Tal como se ilustra en el presente documento, los compuestos descritos en el presente documento pueden aumentar la apoptosis. Por ejemplo, los compuestos 1, BSH-2-H y BSH-4-H inducen apoptosis mediante modulación

de la ruta de regulación de miR-96-FOXO1 en células de cáncer de mama (células de cáncer de mama MCF7 o MDA MB 231). Tal como se muestra en la figura 6E, la adición del compuesto 1 aumenta drásticamente el porcentaje de células MCF7 positivas para TUNEL al menos 10 veces. Tal como se muestra en la figura 20, el compuesto BSH-4-H induce apoptosis en aproximadamente el 40% de las células de cáncer de mama, mientras que el compuesto de BSH-2-H induce apoptosis en aproximadamente el 75% de las células de cáncer de mama.

En algunas realizaciones, los compuestos identificados mediante los métodos descritos en el presente documento pueden aumentar la apoptosis en al menos el 10%, o el 20%, o el 40%, o el 50%, o el 70%, o el 75%, o el 100%, o el 150%, o el 200%, o el 300%, o el 400%, o el 500%, o el 700%, o el 1000%. Algunos compuestos pueden aumentar la apoptosis al menos 2 veces, o 3 veces, o 4 veces, o 5 veces, o 6 veces, o 7 veces, o 8 veces, o 9 veces, o en al menos 10 veces.

Los métodos y las composiciones descritos en el presente documento pueden usarse para tratar una variedad de cánceres y tumores, por ejemplo, leucemia, sarcoma, osteosarcoma, linfomas, melanoma, glioma, feocromocitoma, hepatoma, cáncer de ovario, cáncer de piel, cáncer de testículo, cáncer gástrico, cáncer de páncreas, cáncer renal, cáncer de mama, cáncer de próstata, cáncer colorrectal, cáncer de cabeza y cuello, cáncer cerebral, cáncer de esófago, cáncer de vejiga, cáncer adrenocortical, cáncer de pulmón, cáncer de endometrio, cáncer de nasofaringe, cáncer cervicouterino o de hígado y cáncer en un sitio primario desconocido.

Composiciones

La invención también se refiere a composiciones farmacéuticas que contienen cualquiera de los compuestos 1, 2, 4, 5 y BSH-nH, donde n es cualquier número entero desde 1 hasta 10, incluidos los compuestos BSH-1-H, BSH-2-H, BSH-3-H y BSH-4-H y cualquier combinación de los mismos y un portador farmacéuticamente aceptable. Por "farmacéuticamente aceptable" se entiende un portador, diluyente, excipiente y/o sal que es compatible con los demás componentes de la formulación y no es perjudicial para el receptor del mismo.

Los compuestos pueden administrarse en una cantidad terapéuticamente eficaz, es decir, una cantidad suficiente para obtener el efecto fisiológico deseado, por ejemplo, tratamiento de un estado, trastorno, enfermedad y similares, o reducción de los síntomas del estado, trastorno, enfermedad y similares. Por ejemplo, los agentes terapéuticos pueden administrarse para tratar un estado, trastorno o una enfermedad tal como cáncer, infección viral, infección bacteriana y/o infección microbiana.

Para lograr el/los efecto(s) deseado(s), un compuesto tal como se expuso anteriormente o una combinación de los mismos, puede administrarse como dosificaciones únicas o divididas. Por ejemplo, uno o más de los compuestos pueden administrarse en dosificaciones de al menos aproximadamente 0,001 mg/kg a aproximadamente 500 a 750 mg/kg, de al menos aproximadamente 0,01 mg/kg a aproximadamente 300 a 500 mg/kg, de al menos aproximadamente 0,1 mg/kg a aproximadamente 100 a 300 mg/kg o de al menos aproximadamente 1 mg/kg a aproximadamente 50 a 100 mg/kg de peso corporal, aunque otras dosificaciones pueden proporcionar resultados beneficiosos. La cantidad administrada variará dependiendo de diversos factores incluyendo, pero sin limitarse a, la molécula, el polipéptido, anticuerpo o ácido nucleico elegido para la administración, la enfermedad, el peso, el estado físico, la salud y la edad del mamífero. Tales factores pueden determinarse fácilmente por el médico empleando modelos animales u otros sistemas de prueba que están disponibles en la técnica.

La administración de las moléculas pequeñas como agente terapéutico puede realizarse en una dosis única, en dosis múltiples, de manera continua o intermitente, dependiendo, por ejemplo, de la condición fisiológica del receptor, si el propósito de la administración es terapéutico o profiláctico, y otros factores conocidos por los profesionales cualificados. La administración de las moléculas pequeñas y composiciones puede ser esencialmente continua durante un periodo de tiempo preseleccionado o puede ser en una serie de dosis espaciadas. Se contempla tanto la administración local como sistémica.

Para preparar la composición, las moléculas pequeñas (y otros agentes, si se desea) se sintetizan o se obtienen de otra manera, se purifican según sea necesario o se desee. Estas moléculas pequeñas (y otros agentes, si se desea) pueden suspenderse en un portador farmacéuticamente aceptable y/o liofilizarse o estabilizarse de otra manera. Estas moléculas pequeñas (y los agentes seleccionados, si corresponde) pueden ajustarse a una concentración apropiada y, opcionalmente, combinarse con otros agentes. El peso absoluto de una molécula pequeña dada (y otros agentes opcionales) incluidos en una dosis unitaria puede variar ampliamente. Por ejemplo, pueden administrarse de aproximadamente 0,01 a aproximadamente 2 g, o de aproximadamente 0,1 a aproximadamente 500 mg, de al menos una molécula pequeña. Alternativamente, la dosificación unitaria puede variar entre aproximadamente 0,01 g y aproximadamente 50 g, entre aproximadamente 0,01 g y aproximadamente 35 g, entre aproximadamente 0,1 g y aproximadamente 25 g, entre aproximadamente 0,5 g y aproximadamente 12 g, entre aproximadamente 0,5 g y aproximadamente 8 g, entre aproximadamente 0,5 g y aproximadamente 4 g, o entre aproximadamente 0,5 g y aproximadamente 2 g.

Las dosis diarias de la molécula pequeña también pueden variar. Tales dosis diarias pueden oscilar, por ejemplo, entre aproximadamente 0,1 g/día y aproximadamente 50 g/día, entre aproximadamente 0,1 g/día y

aproximadamente 25 g/día, entre aproximadamente 0,1 g/día y aproximadamente 12 g/día, entre aproximadamente 0,5 g/día y aproximadamente 8 g/día, entre aproximadamente 0,5 g/día y aproximadamente 4 g/día, y entre aproximadamente 0,5 g/día y aproximadamente 2 g/día.

Se apreciará que la cantidad de moléculas pequeñas para uso en el tratamiento variará no solo con el portador particular seleccionado, sino también con la vía de administración, la naturaleza del estado que esté tratándose y la edad y el estado del paciente. En última instancia, el proveedor de atención médica puede determinar la dosificación adecuada. Además, una composición farmacéutica puede formularse como una forma de dosificación unitaria individual.

Por tanto, una o más formas de dosificación unitaria adecuadas que comprenden las moléculas pequeñas pueden administrarse por una variedad de vías incluyendo las vías parenteral (incluyendo subcutánea, intravenosa, intramuscular e intraperitoneal), oral, rectal, dérmica, transdérmica, intratorácica, intrapulmonar e intranasal (respiratoria). Las moléculas pequeñas también pueden formularse para una liberación sostenida (por ejemplo, usando microencapsulación, véase el documento WO 94/07529 y la patente estadounidense n.º 4.962.091). Las formulaciones pueden presentarse, cuando sea apropiado, convenientemente en formas de dosificación unitarias discretas y pueden prepararse mediante cualquiera de los métodos bien conocidos en las técnicas farmacéuticas. Tales métodos pueden incluir la etapa de mezclar el agente terapéutico con portadores líquidos, matrices sólidas, portadores semisólidos, portadores sólidos finamente divididos o combinaciones de los mismos, y luego, si es necesario, introducir o conformar el producto en el sistema de administración deseado.

Las composiciones pueden prepararse en muchas formas que incluyen disoluciones acuosas, suspensiones, comprimidos, cápsulas de gelatina dura o blanda, y liposomas y otras formulaciones de liberación lenta, tales como geles poliméricos conformados. Sin embargo, la administración de moléculas pequeñas también puede implicar la administración parenteral o local en una disolución acuosa o vehículo de liberación sostenida.

Las composiciones farmacéuticas líquidas pueden estar en forma de, por ejemplo, suspensiones acuosas u oleosas, disoluciones, emulsiones, jarabes o elixires, polvos secos para la constitución con agua u otro vehículo adecuado antes de su uso. Tales composiciones farmacéuticas líquidas pueden contener aditivos convencionales tales como agentes de suspensión, agentes emulsionantes, vehículos no acuosos (que pueden incluir aceites comestibles) o conservantes.

Un compuesto o molécula pequeña puede formularse para administración parenteral (por ejemplo, mediante inyección, por ejemplo, inyección en bolo o infusión continua) y puede presentarse en forma de dosificación unitaria en ampollas, jeringas precargadas, recipientes de infusión de pequeño volumen o recipientes de dosis múltiples con un conservante añadido. Las composiciones farmacéuticas pueden adoptar formas tales como suspensiones, disoluciones o emulsiones en vehículos aceitosos o acuosos, y pueden contener agentes de formulación tales como agentes de suspensión, estabilizantes y/o dispersantes. Los portadores adecuados incluyen solución salina y otros materiales comúnmente utilizados en la técnica.

Las composiciones también pueden contener otros componentes tales como agentes quimioterápicos, agentes antivirales, agentes antibacterianos, agentes antimicrobianos y/o conservantes. Los ejemplos de agentes terapéuticos adicionales que pueden usarse incluyen, pero no se limitan a: agentes alquilantes, tales como mostazas de nitrógeno, alquilsulfonatos, nitrosoureas, etileniminas y triazenos; antimetabolitos, tales como antagonistas de folato, análogos de purina y análogos de pirimidina; antibióticos, tales como antraciclinas, bleomicinas, mitomicina, dactinomicina y plicamicina; enzimas, tales como L-asparaginasa; inhibidores de la farnesil-proteína transferasa; agentes hormonales, tales como glucocorticoides, estrógenos/antiestrógenos, andrógenos/antiandrógenos, progestágenos y antagonistas de hormona liberadora de hormona luteinizante, acetato de octreotida; agentes disruptores de microtúbulos, tales como ecteinascidinas o sus análogos y derivados; agentes estabilizadores de microtúbulos como paclitaxel (Taxol®), docetaxel (Taxotere®) y epotilonas A-F o sus análogos o derivados; productos derivados de plantas, tales como alcaloides de la vinca, epipodofilotoxinas, taxanos; e inhibidores de la topoisomerasa; inhibidores de prenil-proteína transferasa; y agentes varios tales como, hidroxiurea, procarbazina, mitotano, hexametilmelamina, complejos de coordinación de platino tales como cisplatino y carboplatino; y otros agentes utilizados como agentes anticancerígenos y citotóxicos, tales como modificadores de la respuesta biológica, factores de crecimiento; moduladores inmunitarios y anticuerpos monoclonales. Los compuestos también pueden usarse junto con la radioterapia.

Definiciones

Los términos “molécula pequeña” y “compuesto” tienen el mismo significado y se usan indistintamente.

“Ácido nucleico”, tal como se usa en el presente documento, se refiere a ARN y ADN. “ARN”, tal como se usa en el presente documento, pretende referirse a moléculas de ácido ribonucleico y oligómeros. “ADN”, tal como se usa en el presente documento, pretende referirse a moléculas de ácido desoxirribonucleico y oligómeros.

“Motivo de ARN”, tal como se usa en el presente documento, se refiere a un bucle interno que puede seleccionarse

como diana, bucle de horquilla, protuberancia u otros motivos estructurales de ácido nucleico que pueden seleccionarse como diana, por ejemplo, tal como se describe en Batey *et al.*, "Tertiary Motifs en RNA Structure y Folding", Angew. Chem. Int. Ed., 38: 2326-2343 (1999), que se incorpora al presente documento como referencia. Los ejemplos de motivos de ARN incluyen bucles internos simétricos, bucles internos asimétricos, bucles internos 1x1, bucles internos 1x2, bucles internos 1x3, bucles internos 2x2, bucles internos 2x3, bucles internos 2x4, bucles internos 3x3, bucles internos 3x4, bucles internos 4x4, bucles internos 4x5, bucles internos 5x5, protuberancias de 1 base, protuberancias de 2 bases, protuberancias de 3 bases, protuberancias de 4 bases, protuberancias de 5 bases, bucles de horquilla de 4 bases, bucles de horquilla de 5 bases, bucles de horquilla de 6 bases, bucles de horquilla de 7 bases, bucles de horquilla de 8 bases, bucles de horquilla de 9 bases, bucles de horquilla de 10 bases, bucles de ramificación múltiple, pseudonudos, etc. Los motivos de ARN tienen estructuras conocidas.

"Interacciona", tal como se usa en el presente documento, pretende referirse a la unión u otra asociación estabilizada entre una molécula pequeña y un motivo de ARN. La asociación puede estabilizarse termodinámicamente o estabilizarse cinéticamente o ambas cosas, y la interacción puede ser el resultado de un enlace covalente, enlace de hidrógeno, interacciones de van der Waals, interacciones electrostáticas o combinaciones de estos y/u otros tipos de interacciones.

Los siguientes ejemplos no limitativos describen algunos de los experimentos realizados para desarrollar y validar aspectos de la invención y los de la solicitud divisional.

Ejemplo 1: Materiales y métodos

Este ejemplo describe algunos de los materiales y métodos utilizados en el desarrollo de la invención.

Análisis de StARTS

Los inventores han desarrollado previamente una estrategia basada en la selección denominada examen combinatorio bidimensional (2DCS; Childs-Disney *et al.*, ACS Chem. Biol. 2, 745-754 (2007); Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008)) así como un método para analizar estadísticamente datos de selección denominados relaciones estructura-actividad a través de secuenciación (StARTS) para identificar y anotar (puntuar) interacciones motivo de ARN-molécula pequeña (Velagapudi *et al.*, Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011), cuyo contenido se incorpora específicamente en el presente documento mediante referencia en su totalidad).

Para 2DCS, se conjuga una biblioteca de moléculas pequeñas sobre una superficie de microalineamiento de agarosa. Entonces se examina con sonda el microalineamiento para determinar la unión a una biblioteca de motivos de ARN pequeños que probablemente se encuentren como componentes de ARN celulares más grandes, por ejemplo, el motivo de ARN de BBI 3x3 mostrado en la figura 1. Se completa la incubación en presencia de oligonucleótidos competidores, tales como los oligonucleótidos C2-C8 (figura 1) para garantizar que las moléculas pequeñas se unan a la región aleatorizada (la región con dos series de nucleótidos 'N' en el motivo de ARN de BBI 3x3 mostrado en la figura 1), y no a regiones comunes a todos los miembros de la biblioteca. (Véase, Childs-Disney *et al.*, ACS Chem. Biol. 2, 745-754 (2007); Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008), cuyo contenido se incorpora específicamente en el presente documento como referencia en su totalidad. La figura 2 muestra las secuencias de tres nucleótidos (N) aleatorizadas que pueden estar en el motivo de ARN de BBI 3x3 mostrado en la figura 1. Los ARN unidos se recogen, amplifican y secuencian. Las secuencias de ARN seleccionados se analizan mediante StARTS.

StARTS es un enfoque estadístico que identifica características en los motivos de ARN que contribuyen positiva y negativamente a la unión (Velagapudi *et al.*, Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011). StARTS predice las afinidades y selectividades de interacciones ARN-molécula pequeña al comparar la tasa de aparición de una característica en motivos de ARN seleccionados (una guanina adyacente a una adenina, por ejemplo) con su tasa de aparición en toda la biblioteca de ARN. A la confianza de que una característica seleccionada no apareció aleatoriamente se le asigna una puntuación Z y un valor de p bilateral correspondiente (Velagapudi *et al.*, Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011). Solo se consideran las características que son estadísticamente significativas ($p \leq 0,05$ o confianza ≥95%). Este análisis identifica características que contribuyen positivamente (puntuación Z positiva) y negativamente (puntuación Z negativa) a la unión, lo que permite predecir qué ARN se unen y cuáles no. Cada motivo de ARN tiene muchas características estadísticamente significativas. Por tanto, las puntuaciones Z para cada característica se suman para proporcionar una puntuación de ΣZ (figura 2).

Al combinar parámetros estadísticos de StARTS con afinidades de unión determinadas experimentalmente, puede predecirse la afinidad y selectividad de cada motivo de ARN presentado en una biblioteca. La selectividad de un ARN seleccionado para diferentes moléculas pequeñas puede predecirse comparando su puntuación de ΣZ para una molécula pequeña con su puntuación de ΣZ para otra. Por ejemplo, un ARN que tiene una alta puntuación de ΣZ para la molécula pequeña A y una pequeña puntuación de ΣZ para la molécula pequeña B es selectivo para la molécula pequeña A. Las puntuaciones de idoneidad se normalizan con respecto al motivo de ARN con la mayor

puntuación de ΣZ.

Se usaron una base de datos de interacciones de motivo de ARN-molécula pequeña y sus correspondientes análisis de StARTS para identificar moléculas pequeñas de partida que pueden modular la función de los microARN. La versión actual de la base de datos de interacciones de motivo de ARN-molécula pequeña consiste en 794 motivos de ARN y 11 moléculas pequeñas.

Todos los miARN precursores humanos (1.048) se descargaron de miRBase (v. 16) (Griffiths-Jones *et al*., Nucleic Acids Res. 34, D140-144 (2006); Griffiths-Jones *et al*., Nucleic Acids Res. 36, D154-158 (2008)), y se predijeron sus estructuras secundarias usando el programa de minimización de energía libre RNAstructure (Mathews *et al*., Proc. Natl Acad Sci. U.S.A. 101, 7287-7292 (2004)). Debe tenerse en cuenta que la predicción de la estructura secundaria se considera la determinación de la estructura para microARN (Ambros *et al*., RNA 9, 277-279 (2003)). Inforna analizó los motivos estructurales secundarios y los comparó con la base de datos de interacciones de motivo de ARN-molécula pequeña. Se obtuvieron un total de 1.668 coincidencias de motivo de ARN-molécula pequeña. De esas coincidencias, 26 motivos son bucles internos ubicados en sitios de procesamiento Drosha o Dicer de miARN que están implicados en enfermedad y que se han validado para la modulación de la enfermedad por oligonucleótidos. Solo se enumeran los motivos de ARN que se encuentran en sitios de procesamiento (por Drosha o Dicer) de los miARN asociados con enfermedades. Inforna proporcionó una salida de motivos que pueden seleccionarse como diana en cada ARN, y las moléculas pequeñas correspondientes que se unen a esos ARN. Se proporciona un subconjunto de los resultados en la tabla 1, en la que la columna 1 representa el número de serie de las coincidencias de microARN (1-22), la columna 4 (etiquetada como “Pequeña”) identifica la molécula pequeña mediante un número (1-9), la columna 6 es la referencia de Pubmed para la asociación con enfermedad del microARN, la columna 7 (etiquetada como “Incremento o Disminución”) muestra si el microARN está regulado por incremento o por disminución de la enfermedad indicada en la columna 5. Los números “4” y “6 “en las dos últimas filas de la tabla 1, indican que hay dos moléculas pequeñas” 4 “y” 6 “que seleccionan como diana pre-miR-885.

Tabla 1: Coincidencias obtenidas de la búsqueda de motivos de ARN en miRBase (v. 16) para el solapamiento con la base de datos de interacciones de motivo de ARN-molécula pequeña usando Inforna

	microARN	Bucle	Pequeña	Enfermedad	ID de PubMed	Incremento o disminución
1	pre-miR-34c	5’GCU/3’CCA	6	cáncer de páncreas	19714243	Disminución
				enfermedad de Parkinson	21558425	Disminución
				enfermedad de Alzheimer	21946562	Disminución
				melanoma	22102694	Disminución
				cáncer de mama	22074923	Disminución
2	pre-miR-92a-2	5’CCU/3’GUG	6	linfoma	21383985	Disminución
				carcinoma hepatocelular	16331254	Incremento
				glioblastoma	22895567	Incremento
				neoplasia colorrectal	21826996	Incremento
3	pre-miR-96	5’UUU/3’AUA	1	neoplasia de mama	19574223	Incremento
				neoplasia de mama	21203424	Incremento
				neoplasia de vejiga urinaria	21166959	Incremento
				neoplasia hepatocelular	22160187	Incremento
				neoplasia de próstata	22045813	Incremento
				neoplasia colorrectal	22844381	Incremento
4	pre-miR-130b	5’UAC/3’ACG	6	neoplasias de estómago	20176475	Incremento
				leucemia mieloide crónica	21638198	Incremento
5	pre-miR-181c	5’GAG/3’C_C (abultamiento)	9	leucemia mieloide aguda	22251480	Incremento
6	pre-miR-182	5’UUUU/3’AUCA	3	melanoma	19188590	Incremento
				neoplasias de próstata	19267923	Incremento
				glioma	20406893	Incremento

				neoplasia de mama	19574223	Incremento
				leucemia aguda	20227111	Incremento
				neoplasia endometrial	20028871	Incremento
				cáncer de pulmón	21904633	Incremento
				neoplasia de próstata	22045813	Incremento
				neoplasia ovárica	22322863	Incremento
7	pre-miR-210	5'ACU/3'UCA	2	cardiopatía isquémica	20837903	Incremento
			7	carcinoma renal	21465485	Incremento
8	pre-miR-301a	5'UAC/3'ACG	6	cáncer de páncreas	22628193	Incremento
				cáncer colorrectal	23393589	Incremento
9	pre-miR-301b	5'UCU/3'AAA	1	cáncer de mama	21393507	Incremento
10	pre-miR-320c	5'UCU/3'AAA	1	cáncer de mama	21393507	Incremento
11	pre-miR-320d-1	5'UCU/3'AAA	1	cáncer de mama	21393507	Incremento
				sarcoma de Kaposi	23418466	Disminución
12	pre-miR-378	5'GGC/3'CGG	8	neoplasias de estómago	19175831	Disminución
				leucemia	19022373	Disminución
				melanoma	20529253	Disminución
				neoplasias (regulador myc)	21242960	Incremento/ disminución
				cáncer de pulmón de células no pequeñas	22052152	Incremento
				neoplasia colorrectal	22469014	Disminución
13	pre-miR-433	5'UUA/3'ACU	1	enfermedad de Parkinson	18252210	Incremento
14	pre-miR-449c	5'GUA/3'UCU	1	neoplasia de estomago	21418558	Disminución
15	pre-miR-515-1	5'UUC/3'GCG	7	neoplasia de células escamosas	21244772	Incremento
16	pre-miR-515-2	5'UCA/3'AUU	4	neoplasia de células escamosas	21244772	Incremento
			7	neoplasia de estomago	22112324	Incremento
17	pre-miR-517c	5'CCC/3'GUG	6	carcinoma hepatocelular	22027761	Incremento
			9			
18	pre-miR-518e	5'CCC/3'GUG	9	melanoma	20529253	Disminución
19	pre-miR-519d	5'CCC/3'GUG	9	neoplasia de mama	20331864	Incremento
				neoplasia hepatocelular	21524841	Disminución
				neoplasia hepatocelular	22262409	Incremento
20	pre-miR-525	5'CUC/3'GCG	9	carcinoma hepatocelular	22362728	Incremento
21	pre-miR-661	5'AGG/3'UGC	6	neoplasia de mama	20543867	Incremento
				neoplasia de estomago	22112324	Disminución
22	pre-miR-885	5'UCU/3'AUA	4	neoplasia de células escamosas	22071691	Disminución
			6			

Se usaron análisis de StARTS para determinar la idoneidad de una molécula pequeña para unir la diana de ARN de interés identificada como salida mediante Inforna (tabla 1; figuras 3C-3D). Basándose en estos análisis, se seleccionaron tres dianas de pre-microARN: pre-microARN-96 (se une a 1), pre-microARN-210 (se une a 2) y pre-microARN-182 (se une a 3). Todas las demás posibles dianas de pre-microARN tienen puntuaciones de ΣZ bajas que indican que se unen débilmente a la molécula pequeña correspondiente (figuras 3C y 3E). Así, las moléculas pequeñas 6-9 probablemente se unirán más estrechamente a otros ARN celulares y no serían específicas para los motivos de microARN con los que están emparejados en la tabla 1.

Se predijo mediante Informa que el compuesto 1 se unirá a otros pre-microARN, incluyendo pre-microARN-301b, 320c, 320d-1, 433 y 449c, lo que puede indicar que el compuesto no es selectivo. Sin embargo, el análisis de StARTS predice que 1 se une al bucle que puede seleccionarse como diana en el pre-microARN-449c con baja afinidad (figuras 4A-4B), lo que se confirmó midiendo la afinidad en disolución (figuras 4D-4I). Tomados junto con los datos *in vivo* mostrados en las figuras 5-7, StARTS predice con precisión las dianas de ARN que se unirán o no y la selectividad potencial de una molécula pequeña.

Se evaluaron varios compuestos tal como se describe en el presente documento, incluidos, por ejemplo, los compuestos 1, 2, 3, 4 y 5 anteriores: mientras que los compuestos 1, 2, 4 y 5 comparten algunas características estructurales, la estructura del compuesto 3 es bastante diferente de los otros que se mostraron anteriormente.

Se completó la selección de 2DCS para tres compuestos que son químicamente similares al compuesto 1: los compuestos 2, 4 y 5 (Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011). La inspección visual sugiere que 2 y 4 podrían unirse a ARN con mayor afinidad que 1 debido a sus áreas superficiales más grandes que podrían apilarse sobre bases de ARN y debido a la presencia de donadores y aceptores de enlaces de hidrógeno adicionales. Sin embargo, el análisis de StARTS predice que 2 y 4 se unen débilmente al bucle que puede seleccionarse como diana en el precursor de miR-96, lo que concuerda de forma excelente con los datos *in vivo* (figuras 4A-4B). (Debe tenerse en cuenta que se intentó una selección de 2DCS para el compuesto 5; sin embargo, no pudieron seleccionarse ARN debido a su débil afinidad (*id.*).)

Moldes de ADN y amplificación por PCR

Los motivos de ARN (bucles internos) usados en estos estudios se incrustaron en un casete de horquilla, C1 (figura 1). Los moldes de ADN correspondientes (adquiridos de Integrated DNA Technologies (IDT) y usados sin purificación adicional) se amplificaron por PCR en tampón de PCR IX (Tris 10 mM, pH 9,0, KCl 50 mM y Triton X-100 al 0,1% (v/v)), cebador directo 2 μM: 5'-GGCCGGATCCTAATACGACTCACTATAGGGAGAGGGTTTAAT (SEQ ID NO: 18), cebador inverso 2 μM: 5'-CCTTGCGGATCCAAT (SEQ ID NO: 19), $MgCl_2$ 4,25 mM, DNTP 330 μM y 1 μl de ADN polimerasa Taq en una reacción de 50 μl. Las condiciones de ciclado usadas para la PCR fueron de 95ºC durante 30 s, 50ºC durante 30 s y 72ºC durante 1 min.

El precursor de miR-96 usado en los ensayos de protección frente a nucleasas se modificó para contener una proyección 5'-GGG para facilitar la transcripción usando ARN polimerasa de T7, o GGG-pre-microARN-96. No hubo cambios en la estructura secundaria de energía libre más baja predicha por RNAstructure (Mathews *et al.*, Proc. Natl Acad Sci. U. S. A. 101, 7287-7292 (2004)), y los ensayos de protección con nucleasas confirman que la estructura predicha se adopta en disolución (figura 9). El molde de ADN para GGG-pre-microARN-96 se amplificó por PCR tal como se describió anteriormente, excepto en que los cebadores fueron: 5'-GGCCGGATCCTAATACGACTCACTATA GGGTGGCCGATTTTGGC (SEQ ID NO: 20, directo) y 5'-TTTCCC ATATTGGCA (SEQ ID NO: 21, inverso) y las condiciones de ciclado fueron de 95ºC durante 30 s, 55ºC durante 30 s y 72ºC durante 1 min. Los moldes de ADN usadas para la PCR para producir ADN bicatenarios adecuados para la transcripción fueron:

C1: 5'-GGGAGAGGGTTTAATTACGAAAGTAATTGGATCCGCAAGG (SEQ ID NO:1);

5'UUU/3'AUA: 5'-GGGAGAGGGTTTAA**TTT**TACGAAAGTA**ATA**TT

GGATCCGCAAGG (SEQ ID NO:8);

5'CGAUUU/3'GGUAUA: 5'-GGGAGAGGGTTTAATC**CGATTT**T

ACGAAAG TA**ATATGG**GATTGGATCCGCAAGG (SEQ ID NO:9);

y

GGG-pre-microRNA-96:

5'**GGG**TGGCCGATTTTGGCACTAGCACATTTTTGCTTGTGTCTCTCCGC

TCTGAGCAATCATGTGCAGTGCCAATATGGGAAA (SEQ ID NO:22).

Transcripción de ARN

Se transcribieron *in vitro* oligonucleótidos de ARN mediante ARN polimerasa de T7 en tampón de transcripción IX (Tris HCl 40 mM, pH 8,1, espermidina 1 mM, Triton X-100 al 0,001% (v/v) y DTT 10 mM) (McKenna *et al.*, Nat. Protoc. 2, 3270-3277 (2007)) que contiene 2,25 mM de cada rNTP y $MgCl_2$ 5 mM a 37ºC durante la noche. Los ARN transcritos se purificaron en un gel de poliacrilamida al 15% desnaturalizante y se aislaron tal como se describió anteriormente por Childs-Disney *et al.*, ACS Chem. Biol. 2, 745-754 (2007). Se determinaron las concentraciones mediante la absorbancia a 260 nm y el coeficiente de extinción correspondiente. Se calcularon los coeficientes de extinción usando el servidor HyTher (Peyret *et al.*, Biochemistry 38, 3468-3477 (1999); SantaLucia *et al.*, Proc. Natl Acad Sci. U. S. A. 95, 1460-1465 (1998)), que usa parámetros basados en los coeficientes de extinción de los vecinos más próximos al ARN (Puglisi & Tinoco, Methods Enzymol. 180, 304-325 (1989)).

Mediciones de afinidad de unión

Se determinaron las constantes de disociación usando un ensayo en disolución, basado en fluorescencia (Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008).). El ARN de interés se plegó en tampón de ensayo IX (Na_2HPO_4 8 mM, pH 7,0, NaCl 190 mM, EDTA 1 mM y BSA 40 μg/ml) calentando a 60ºC durante 5 min y enfriando lentamente hasta temperatura ambiente. Se añadió un compuesto marcado con fluorescencia seleccionado a una concentración final de 50 nM para los compuestos 1-Fl, 4-Fl y 5-Fl o 500 nM para compuesto 2. Se completaron entonces diluciones en serie (1:2) en tampón de ensayo IX complementado con 50 nM de los compuestos 1-Fl, 4-Fl o 5-Fl o 500 nM de 2. Se incubaron las disoluciones durante 30 min a temperatura ambiente y luego se transfirieron a una placa de 96 pocillos y se midió la intensidad de fluorescencia. Se ajustó el cambio en la intensidad de fluorescencia en función de la concentración de ARN a la siguiente ecuación (Wang y Rando, Chem. Biol. 2, 281-290. (1995).):

$$I = I_0 + 0{,}5\Delta\varepsilon(([FL]_0 + [ARN]_0 + K_t) - (([FL]_0 + [ARN]_0 + K_t)^2 - 4[FL]_0[ARN]_0)^{0,5})$$

donde I e I_0 son la intensidad de fluorescencia observada en presencia y ausencia de ARN respectivamente, $\Delta\varepsilon$ es la diferencia entre la intensidad de fluorescencia en ausencia de ARN y en presencia de una concentración de ARN infinita, $[FL]_0$ y $[ARN]_0$ son las concentraciones de molécula pequeña y ARN, respectivamente, y K_t es la constante de disociación.

Los bucles que pueden seleccionarse como diana (motivos) se incrustaron en C1 (figura 1) para que pudieran completarse las mediciones de afinidad. Las estructuras secundarias de los ARN tal como se predicen por RNAstructure (Mathews *et al.*, Proc. Natl Acad Sci. U. S. A. 101, 7287-7292 (2004).) y curvas de unión representativas se muestran en (figuras 4D-4I).

Ensayos de protección frente a nucleasas

Se etiquetó en el extremo 5' GGG-pre-microARN-96 con ^{32}P tal como se describió anteriormente por Disney *et al.* (Biochemistry 39, 14269-14278 (2000)).

Para la escisión con endorribonucleasa específica de ARN bicatenario, se plegó el ARN en tampón de reacción IX (Ambion) calentando a 60ºC durante 5 minutos y enfriando lentamente hasta temperatura ambiente. Entonces se añadió endorribonucleasa específica de ARN bicatenario (ARNasa III de *Escherichia coli*; Ambion) a una concentración final de 0,15 unidades/μl seguido de la adición de concentraciones diluidas en serie de compuesto 1. Se incubó la disolución a 37ºC durante 2 h, y se separaron los productos de escisión en un gel de poliacrilamida al 15% desnaturalizante.

Para la escisión con ARNasa T1, se plegó el ARN en tampón de secuenciación de ARN IX (Ambion; condiciones desnaturalizantes) o en tampón de estructura de ARN IX (Ambion; condiciones nativas) incubándolo a 55ºC durante 10 minutos, seguido de un enfriamiento lento hasta temperatura ambiente. Se añadió ARNasa T1 a una concentración final de 0,1 unidades/μl seguido de la adición de concentraciones diluidas en serie de 1. Se incubó la disolución a temperatura ambiente durante 15 minutos, y se separaron los productos escindidos en un gel de poliacrilamida al 15% desnaturalizante.

Preparación de extractos celulares que contienen Drosha

Drosha es una enzima de la ARNasa III de clase 2 que inicia el procesamiento de microARN (miARN), o moléculas de ARN cortas expresadas naturalmente por las células.

Se mantuvieron células HEK 293T en DMEM complementado con FBS al 10% en un matraz T-75. Una vez que las células alcanzaron el 70% de confluencia, se transfectaron con Drosha-cmyc51, obtenido de Addgene (plásmido de Addgene 10828) usando Lipofectamine 2000 (Invitrogen) según el protocolo del fabricante. Aproximadamente 48 h después de la transfección, se recogieron las células rascándolas en 1 ml de DPBS IX enfriado con hielo, seguido de

centrifugación a 6000 rpm durante 5 min a 4ºC. Se resuspendieron las células en 500 µl de tampón de lisis 1X (Tris HCl 20 mM, pH 8,0, KCl 100 mM y EDTA 0,2 mM) y se sonicaron durante 30 segundos. Se sedimentaron los residuos celulares mediante centrifugación (12000 rpm durante 15 min a 4ºC) y se transfirió el sobrenadante que contenía Drosha a un nuevo tubo.

Inhibición de la escisión de Drosha *in vitro*

El molde de ADNc para pri-miARN-96 se amplificó por PCR a partir del ADN genómico de MCF7 usando los siguientes cebadores:

cebador directo: 5'-GGCCGAATTCTAATACGACTCACTATAGGC ACCAGTGCCATCTGCTT (SEQ ID NO: 23); y cebador inverso: 5'-CGCAGCTGCGGGTCCT (SEQ ID NO: 232). El cebador directo contiene un promotor de T7 que se empleó para producir pri-miR-96 a través de la transcripción no iniciada (*run-off*) tal como se describe en el presente documento. El pri-miR-96 etiquetado internamente se transcribió usando α-^{32}P-ATP y se purificó usando un gel de poliacrilamida al 10% desnaturalizante. Para determinar si el compuesto 1 inhibe la escisión de Drosha *in vitro* se incubaron 2 µl de pri-miR-96 marcado con ^{32}P (aprox. 10.000 cuentas) y 40 µM de compuesto 1 en $MgCl_2$ 6,4 mM (30 µl de volumen total) a temperatura ambiente durante 10 min. (Los controles no tratados incluyeron DMSO al 0,04%, la misma concentración que en las muestras tratadas). Luego, se añadió 1 µl de lisado de Drosha-cmyc y se incubaron las muestras a 37ºC durante 3 h. Se extinguieron las reacciones mediante extracción con fenol-cloroformo seguida de precipitación con etanol. Se disolvió el sedimento resultante en 10 µl de tampón de carga de gel 2X (urea 8 M, EDTA 50 mM, azul de bromofenol al 0,05% (p/v), xileno-cianol al 0,05% (p/v)) y se separaron los productos de reacción en un gel de acrilamida al 10% desnaturalizante.

Cultivo de células

Las células (por ejemplo, de la línea celular de cáncer de mama MCF7) se cultivaron en medio de Eagle modificado por Dulbecco/F12 (DMEM/F12) complementado con FBS al 10% (medio de crecimiento completo) a 37ºC y el 5% de CO_2.

Plásmidos

Los constructos de luciferasa son aquellos descritos por Guttilla & White (J. Biol. Chem. 284, 23204-23216 (2009)).

Aislamiento de ARN y reacción en cadena de polimerasa en tiempo real cuantitativa (qRT-PCR) de miARN

Se hicieron crecer células MCF7 en placas de 6 pocillos o de 12 pocillos, y se extrajo el ARN total usando el reactivo TRIzol LS (Ambion) según el protocolo del fabricante. Se usaron aproximadamente 200 ng de ARN total en las reacciones de transcripción inversa (RT), que se completaron usando un kit Taqman MicroRNA RT (Applied Biosystems) o un kit miScript II RT (Qiagen) según el protocolo del fabricante. Se realizó qRT-PCR en un sistema de PCR en tiempo real 7900HT Fast (Applied Biosystem) usando mezcla maestra para PCR Taqman Universal o mezcla maestra power SYBR green (Applied Biosystems). Todos los conjuntos de cebadores para miARN maduros se adquirieron de Applied Biosystems. El nivel de expresión de miARN maduros se normalizó con respecto a ARN nuclear pequeño U6.

Se analizaron los productos de RT-PCR de miR-96 pre, pri y maduros en un gel de poliacrilamida al 10% desnaturalizante teñido con bromuro de etidio. Se normalizaron los niveles de expresión con respecto a GAPDH. Los cebadores usados para pre y pri-miR-96 y GAPDH se adquirieron de IDT y tienen las secuencias que se muestran en la tabla 2.

Tabla 2: Secuencias de cebadores directos usados para un sistema de SYBR qRT-PCR para perfilar el efecto del compuesto 1 sobre otros miARN y la secuencia del cebador inverso universal. También se enumeran los cebadores directo e inverso para pri-miR-96, pre-miR-96 y GAPDH.

MicroARN	Secuencia de cebador	SEQ ID
hsa-miR-212	TAACAGTCTCCAGTCACGGCC	24
hsa-miR-346	TGTCTGCCCGCATGCCTGCCTCT	25
hsa-miR-194	TGTAACAGCAACTCCATGTGGA	26
hsa-miR-196b	TAGGTAGTTTCCTGTTGTTGGG	27
hsa-miR-196a	TAGGTAGTTTCATGTTGTTGGG	28
hsa-miR-106b	TAAAGTGCTGACAGTGCAGAT	29
hsa-miR-210	CTGTGCGTGTGACAGCGGCTGA	30

hsa-miR-214	ACAGCAGGCACAGACAGGCAGT	31
hsa-miR-17	CAAAGTGCTTACAGTGCAGGTAG	32
hsa-miR-514	ATTGACACTTCTGTGAGTAGA	33
hsa-miR-29c	TAGCACCATTTGAAATCGGTTA	34
hsa-miR-410	AATATAACACAGATGGCCTGT	35
hsa-miR-198	GGTCCAGAGGGGAGATAGGTTC	36
hsa-let-7i	TGAGGTAGTAGTTTGTGCTGTT	37
hsa-miR-188-5p	CATCCCTTGCATGGTGGAGGG	38
hsa-miR-92a	TATTGCACTTGTCCCGGCCTGT	39
hsa-miR-126	TCGTACCGTGAGTAATAATGCG	40
hsa-miR-181c	AACATTCAACCTGTCGGTGAGT	41
hsa-miR-30b	TGTAAACATCCTACACTCAGCT	42
hsa-miR-183	TATGGCACTGGTAGAATTCACT	43
hsa-miR-30c	TGTAAACATCCTACACTCTCAGC	44
hsa-miR-421	ATCAACAGACATTAATTGGGCGC	45
hsa-miR-193b	AACTGGCCCTCAAAGTCCCGCT	46
hsa-miR-651	TTTAGGATAAGCTTGACTTTTG	47
hsa-miR-15a	TAGCAGCACATAATGGTTTGTG	48
hsa-miR-301b	CAGTGCAATGATATTGTCAAAGC	49
hsa-miR-18b	TAAGGTGCATCTAGTGCAGTTAG	50
hsa-miR-15b	TAGCAGCACATCATGGTTTACA	51
hsa-miR-10a	TACCCTGTAGATCCGAATTTGTG	52
hsa-miR-26b	TTCAAGTAATTCAGGATAGGT	53
hsa-let-7a	TGAGGTAGTAGGTTGTATAGTT	54
hsa-miR-506	TAAGGCACCCTTCTGAGTAGA	55
hsa-let-7g	TGAGGTAGTAGTTTGTACAGTT	56
hsa-miR-25	CATTGCACTTGTCTCGGTCTGA	57
hsa-miR-433	ATCATGATGGGCTCCTCGGTGT	58
hsa-miR-144	TACAGTATAGATGATGTACT	59
hsa-miR-32	TATTGCACATTACTAAGTTGCA	60
hsa-miR-200c	TAATACTGCCGGGTAATGATGGA	61
hsa-miR-29a	TAGCACCATCTGAAATCGGTTA	62
hsa-miR-497	CAGCAGCACACTGTGGTTTGT	63
hsa-miR-548b-3p	CAAGAACCTCAGTTGCTTTTGT	64
hsa-miR-24	TGGCTCAGTTCAGCAGGAACAG	65
hsa-miR-187	TCGTGTCTTGTGTTGCAGCCGG	66
hsa-miR-18a	TAAGGTGCATCTAGTGCAGATAG	67
hsa-miR-222	AGCTACATCTGGCTACTGGGT	68
hsa-miR-9	TCTTTGGTTATCTAGCTGTATGA	69
hsa-miR-125b	TCCCTGAGACCCTAACTTGTGA	70
hsa-miR-365	TAATGCCCCTAAAAATCCTTAT	71
hsa-miR-10b	TACCCTGTAGAACCGAATTTGTG	72
hsa-let-7c	TGAGGTAGTAGGTTGTATGGTT	73

hsa-miR-185	TGGAGAGAAAGGCAGTTCCTGA	74
hsa-miR-648	AAGTGTGCAGGGCACTGGT	75
hsa-miR-206	TGGAATGTAAGGAAGTGTGTGG	76
hsa-miR-124	TAAGGCACGCGGTGAATGCC	77
hsa-miR-132	TAACAGTCTACAGCCATGGTCG	78
hsa-miR-519b-5p	CTCTAGAGGGAAGCGCTTTCTG	79
hsa-miR-148b	TCAGTGCATCACAGAACTTTGT	80
hsa-miR-125a-5p	TCCCTGAGACCCTTTAACCTGTGA	81
hsa-miR-20b	CAAAGTGCTCATAGTGCAGGTAG	82
hsa-miR-512-3p	AAGTGCTGTCATAGCTGAGGTC	83
hsa-miR-186	CAAAGAATTCTCCTTTTGGGCT	84
hsa-miR-149	TCTGGCTCCGTGTCTTCACTCCC	85
hsa-miR-622	ACAGTCTGCTGAGGTTGGAGC	86
hsa-miR-429	TAATACTGTCTGGTAAAACCGT	87
hsa-miR-23a	ATCACATTGCCAGGGATTTCC	88
hsa-miR-485-3p	GTCATACACGGCTCTCCTCTCT	89
hsa-miR-19b	TGTGCAAATCCATGCAAAACTGA	90
hsa-miR-141	TAACACTGTCTGGTAAAGATGG	91
hsa-miR-103	AGCAGCATTGTACAGGGCTATGA	92
hsa-miR-96	TTTGGCACTAGCACATTTTTGCT	93
hsa-miR-23b	ATCACATTGCCAGGGATTACC	94
hsa-miR-152	TCAGTGCATGACAGAACTTGG	95
hsa-miR-610	TGAGCTAAATGTGTGCTGGGA	96
hsa-miR-203	GTGAAATGTTTAGGACCACTAG	97
hsa-miR-127-3p	TCGGATCCGTCTGAGCTTGGCT	98
hsa-miR-134	TGTGACTGGTTGACCAGAGGGG	99
hsa-let-7f	TGAGGTAGTAGATTGTATAGTT	100
hsa-miR-122	TGGAGTGTGACAATGGTGTTTG	101
hsa-miR-142-3p	TGTAGTGTTTCCTACTTTATGGA	102
hsa-miR-34a	TGGCAGTGTCTTAGCTGGTTGT	103
hsa-miR-24	TGGCTCAGTTCAGCAGGAACAG	104
hsa-miR-187	TCGTGTCTTGTGTTGCAGCCGG	105
hsa-miR-18a	TAAGGTGCATCTAGTGCAGATAG	106
hsa-miR-222	AGCTACATCTGGCTACTGGGT	107
hsa-miR-9	TCTTTGGTTATCTAGCTGTATGA	108
hsa-miR-125b	TCCCTGAGACCCTAACTTGTGA	109
hsa-miR-365	TAATGCCCCTAAAAATCCTTAT	110
hsa-miR-10b	TACCCTGTAGAACCGAATTTGTG	111
hsa-let-7c	TGAGGTAGTAGGTTGTATGGTT	112
hsa-miR-185	TGGAGAGAAAGGCAGTTCCTGA	113
hsa-miR-648	AAGTGTGCAGGGCACTGGT	114
hsa-miR-206	TGGAATGTAAGGAAGTGTGTGG	115
hsa-miR-124	TAAGGCACGCGGTGAATGCC	116

hsa-miR-132	TAACAGTCTACAGCCATGGTCG	117
hsa-miR-519b-5p	CTCTAGAGGGAAGCGCTTTCTG	118
hsa-miR-148b	TCAGTGCATCACAGAACTTTGT	119
hsa-miR-125a-5p	TCCCTGAGACCCTTTAACCTGTGA	120
hsa-miR-20b	CAAAGTGCTCATAGTGCAGGTAG	121
hsa-miR-512-3p	AAGTGCTGTCATAGCTGAGGTC	122
hsa-miR-186	CAAAGAATTCTCCTTTTGGGCT	123
hsa-miR-149	TCTGGCTCCGTGTCTTCACTCCC	124
hsa-miR-622	ACAGTCTGCTGAGGTTGGAGC	125
hsa-miR-429	TAATACTGTCTGGTAAAACCGT	126
hsa-miR-23a	ATCACATTGCCAGGGATTTCC	127
hsa-miR-485-3p	GTCATACACGGCTCTCCTCTCT	128
hsa-miR-19b	TGTGCAAATCCATGCAAAACTGA	129
hsa-miR-141	TAACACTGTCTGGTAAAGATGG	130
hsa-miR-103	AGCAGCATTGTACAGGGCTATGA	131
hsa-miR-96	TTTGGCACTAGCACATTTTTGCT	132
hsa-miR-23b	ATCACATTGCCAGGGATTACC	133
hsa-miR-152	TCAGTGCATGACAGAACTTGG	134
hsa-miR-610	TGAGCTAAATGTGTGCTGGGA	135
hsa-miR-203	GTGAAATGTTTAGGACCACTAG	136
hsa-miR-127-3p	TCGGATCCGTCTGAGCTTGGCT	137
hsa-miR-134	TGTGACTGGTTGACCAGAGGGG	138
hsa-let-7f	TGAGGTAGTAGATTGTATAGTT	139
hsa-miR-122	TGGAGTGTGACAATGGTGTTTG	140
hsa-miR-142-3p	TGTAGTGTTTCCTACTTTATGGA	141
hsa-miR-34a	TGGCAGTGTCTTAGCTGGTTGT	142
hsa-miR-19a	TGTGCAAATCTATGCAAAACTGA	143
hsa-miR-130b	CAGTGCAATGATGAAAGGGCAT	144
hsa-miR-128a	TCACAGTGAACCGGTCTCTTT	145
hsa-miR-182	TTTGGCAATGGTAGAACTCACACT	146
hsa-miR-135b	TATGGCTTTTCATTCCTATGTGA	147
hsa-miR-34c-5p	AGGCAGTGTAGTTAGCTGATTGC	148
hsa-miR-1	TGGAATGTAAAGAAGTATGTAT	149
hsa-miR-143	TGAGATGAAGCACTGTAGCTC	150
hsa-miR-338-3p	TCCAGCATCAGTGATTTTGTTG	151
hsa-miR-138	AGCTGGTGTTGTGAATCAGGCCG	152
hsa-miR-27a	TTCACAGTGGCTAAGTTCCGC	153
hsa-miR-27b	TTCACAGTGGCTAAGTTCTGC	154
hsa-miR-320	AAAAGCTGGGTTGAGAGGGCGA	155
hsa-miR-16	TAGCAGCACGTAAATATTGGCG	156
hsa-miR-100	AACCCGTAGATCCGAACTTGTG	157
hsa-miR-140-3p	TACCACAGGGTAGAACCACGG	158
hsa-miR-590-5p	GAGCTTATTCATAAAAGTGCAG	159

hsa-miR-135a	TATGGCTTTTTATTCCTATGTGA	160
hsa-miR-378	ACTGGACTTGGAGTCAGAAGG	161
hsa-miR-139-5p	TCTACAGTGCACGTGTCTCCAG	162
hsa-miR-331-5p	CTAGGTATGGTCCCAGGGATCC	163
hsa-miR-342-3p	TCTCACACAGAAATCGCACCCGT	164
hsa-miR-99a	AACCCGTAGATCCGATCTTGTG	165
hsa-miR-517a	ATCGTGCATCCCTTTAGAGTGT	166
hsa-miR-489	GTGACATCACATATACGGCAGC	167
hsa-miR-20a	TAAAGTGCTTATAGTGCAGGTAG	168
hsa-miR-218	TTGTGCTTGATCTAACCATGT	169
hsa-miR-101	TACAGTACTGTGATAACTGAA	170
hsa-miR-205	TCCTTCATTCCACCGGAGTCTG	171
hsa-miR-98	TGAGGTAGTAAGTTGTATTGTT	172
hsa-let-7b	TGAGGTAGTAGGTTGTGTGGTT	173
hsa-miR-181b	AACATTCATTGCTGTCGGTGGGT	174
hsa-miR-30a	TGTAAACATCCTCGACTGGAAG	175
hsa-miR-21	TAGCTTATCAGACTGATGTTGA	176
hsa-miR-140-5p	CAGTGGTTTTACCCTATGGTAG	177
hsa-miR-302a	TAAGTGCTTCCATGTTTTGGTGA	178
hsa-miR-29b	TAGCACCATTTGAAATCAGTGTT	179
hsa-miR-590-3p	TAATTTTATGTATAAGCTAGT	180
hsa-let-7d	AGAGGTAGTAGGTTGCATAGTT	181
hsa-miR-519c-3p	AAAGTGCATCTTTTTAGAGGAT	182
hsa-miR-181a	AACATTCAACGCTGTCGGTGAGT	183
hsa-miR-181d	AACATTCATTGTTGTCGGTGGGT	184
hsa-miR-301a	CAGTGCAATAGTATTGTCAAAGC	185
hsa-miR-150	TCTCCCAACCCTTGTACCAGTG	186
hsa-miR-298	AGCAGAAGCAGGGAGGTTCTCCCA	187
hsa-miR-137	TTATTGCTTAAGAATACGCGTAG	188
hsa-miR-127-5p	CTGAAGCTCAGAGGGCTCTGAT	189
hsa-miR-133a	TTTGGTCCCCTTCAACCAGCTGAA	190
hsa-miR-153	TTGCATAGTCACAAAAGTGATC	191
hsa-miR-193a-3p	AACTGGCCTACAAAGTCCCAGTAA	192
hsa-miR-216a	TAATCTCAGCTGGCAACTGTGAAA	193
hsa-miR-217	TACTGCATCAGGAACTGATTGGAA	194
hsa-miR-224	CAAGTCACTAGTGGTTCCGTT	195
hsa-miR-199a-5p	CCCAGTGTTCAGGCTACCTGTTC	196
hsa-miR-200a	TAACACTGTCTGGTAACGATGTAA	197
hsa-miR-302a*	TAAACGTGGATGTACTTGCT	198
hsa-miR-326	TGGGCCCTTCCTCCAGAA	199
hsa-miR-33b	GTGCATTGCTGTTGCATTGC	200
hsa-miR-139-3p	GGAGACGCGGCCCTGTTGGAGT	201
hsa-miR-451	CCGTTACCATTACTGAGTTAA	202

hsa-miR-324-3p	ACTGCCCCAGGTGCTGCTGG	203
hsa-miR-509-3-5p	GGGTACTGCAGACGTGGCAATCAT	204
hsa-miR-483-3p	TCACTCCTCTCCTCCCGTCTT	205
hsa-miR-146b-3p	CTGTGGACTCAGTTCTGGAA	206
hsa-miR-34b	AATCACTAACTCCACTGCCATC	207
hsa-miR-371-5p	CAAACTGTGGGGGCACTAA	208
hsa-miR-488*	CCCAGATAATGGCACT	209
hsa-miR-513a-3p	TTTCACCTTTCTGAGAAGGAA	210
hsa-miR-513b	TTCACAAGGAGGTGTCATTTAT	211
hsa-miR-449b	GCCAGCAGGCAGTGTATTGTTAGC	212
hsa-miR-486-5p	TCCTGTACTGAGCTGCCCCGAG	213
hsa-miR-1246	AATGGATTTTTGGAGCAGG	214
hsa-miR-1245	AAGTGATCTAAAGGCCTACAT	215
SNORD48	CCCCAGGTAACTCTTGAGTGT	216
SNORD47	TAATGATATCACTGTAAAACC	217
SNORD44	ATGCTGACTGAACATGAAGGTC	218
RNU6	ACACGCAAATTCGTGAAGCGTTC	219
Inverso universal	GAATCGAGCACCAGTTACGC	220
Pre-miARN-96-F	ATTTTGGCACTAGCACATTTTTGCT	221
Pre-miARN-96-R	CCATATTGGCACTGCACATGATT	222
Pri-miARN-96-F	AGAGAGCCCGCACCAGT	223
Pri-miARN-96-R	CTTGAGGAGGAGCAGGCT	224
GAPDH-F	AAGGTGAAGGTCGGAGTCAA	225
GAPDH-R	AATGAAGGGGTCATTGATGG	226

Obtención de perfiles de transcriptoma completo de microARN maduros por qRT-PCR

Se hicieron crecer células MCF7 y se extrajo el ARN total tal como se describió anteriormente. Se usó aproximadamente 1 μg de ARN total en las reacciones de RT, que se completaron usando el kit miScript II RT (Qiagen) según el protocolo del fabricante. Se realizó qRT-PCR usando la mezcla maestra power SYBR green (Applied Biosystems) en un sistema de PCR en tiempo real 7900HT Fast. Todos los cebadores directos se adquirieron de Life Technologies, mientras que el cebador inverso universal se adquirió de IDT (tabla 2). Se normalizaron los niveles de expresión usando RNU6, SNORD44, SNORD47 y SNORD48.

Ensayo de luciferasa dual

Se hicieron crecer células MCF7 en placas de 96 pocillos hasta ~80% de confluencia en medio de crecimiento completo. Se transfectaron las células de manera transitoria con 100 ng de plásmido que codificaba para la diana de 3' UTR de miR-96 o una diana de UTR 3' mutada usando Lipofectamine 2000 según el protocolo del fabricante. Aproximadamente 5 h después de la transfección, se añadió la molécula pequeña de interés en medio de crecimiento completo y se incubaron las células durante otras 20 h. Entonces se midió la actividad luciferasa usando un sistema de análisis de luciferasa Dual Glo (Promega) según el protocolo del fabricante. Los valores notificados son el promedio de al menos tres mediciones, y los errores son las desviaciones estándar correspondientes.

Inmunotransferencia de tipo Western

Se hicieron crecer células MCF7 en placas de 6 pocillos hasta ~80% de confluencia en medio de crecimiento completo. Entonces se incubaron las células con 40 μM de 1 durante 20 h. Se extrajo la proteína total usando el reactivo de extracción de proteínas de mamífero M-PER (Pierce Biotechnology) usando el protocolo del fabricante. Se cuantificó la proteína total extraída usando un kit de ensayo de proteínas Micro BCA (Pierce Biotechnology). Aproximadamente 40 μg de proteína total se resolvieron en un gel de poliacrilamida-SDS al 8% y luego se

transfirieron a una membrana de PVDF. Se lavó la membrana en resumen con solución salina tamponada con Tris IX (TBS) y luego se bloqueó en leche al 5% en TBST IX (TBS IX que contenía Tween-20 al 0,1%) durante 1 h a temperatura ambiente. Entonces se incubó la membrana en anticuerpos primarios FOXO1 1:1000 (Cell Signaling Technology) en TBST IX que contenía BSA al 3% durante la noche a 4ºC. Se lavó la membrana con TBST IX y se incubó con un conjugado de anticuerpo secundario con peroxidasa del rábano de IgG anti-conejo 1:2000 (Cell Signaling Technology) en TBS IX durante 1 hora a temperatura ambiente. Después de lavar con TBST IX, se cuantificó la expresión de proteína usando el sustrato quimioluminiscente SuperSignal West Pico (Pierce Biotechnology) según el protocolo del fabricante. Luego se separó la membrana usando un tampón de separación IX (glicina 200 mM, pH 2,2 y SDS al 0,1%) seguido de lavado en TBS IX. Se bloqueó la membrana y se trató con sonda en busca de GAPDH siguiendo el mismo procedimiento descrito anteriormente usando anticuerpos primarios GAPDH 1:2000 (Abcam). Se usó el software Image J de los Institutos Nacionales de Salud para cuantificar las intensidades de banda.

Tratamiento de ARNip

Se adquirieron ARNip de FOXO1 ON-TARGETplus SMARTpool y reserva de control de GAPD ONTARGETplus sde Dharmacon (Thermo Scientific) y se usaron a una concentración final de 100 nM. Se sometieron a transfección inversa células MCF7 en una placa de 6 pocillos usando el reactivo Lipofectamine RNAiMAX (Invitrogen) según el protocolo del fabricante.

Ensayo TUNEL de APO BrdU

Se hicieron crecer células MCF7 en placas de 6 pocillos hasta una confluencia del 70-80%. Para los experimentos en los que se usó iARN para atenuar la expresión de FOXO1 o GAPDH, se transfectaron las células en primer lugar tal como se describió anteriormente. Se trataron las células con moléculas pequeñas durante 20 h seguido por que se completa un ensayo TUNEL de APO BrdU (Molecular Probes) según el protocolo del fabricante. Se realizó citometría de flujo con un instrumento BD LSRII (BD Biosciences). Se usaron al menos 10.000 eventos para el análisis.

Ensayos de anexina V/PI

Se usó la tinción con anexina V como indicador de apoptosis. En células normales, la fosfatidilserina se ubica en la superficie citoplasmática de la membrana celular. En células apoptóticas, la fosfatidilserina se transloca a la cara externa de la membrana plasmática. Dicha translocación expone la fosfatidilserina, de modo que puede teñirse con anexina V, que tiene una alta afinidad por la fosfatidilserina.

Se hicieron crecer células MCF7 en placas de 6 pocillos hasta una confluencia del 70-80%. Se incubaron las células con 40 μM de compuesto 1 durante 20 h y luego se separaron de la superficie usando Accutase. Se lavaron dos veces, cada una con DPBS IX enfriado con hielo y tampón de unión a anexina IX (Hepes 50 mM (pH 7,4), NaCl 700 mM y $CaCl_2$ 12,5 mM). Se resuspendieron las células en 100 μl de tampón de unión a anexina IX, y luego se añadieron 5 μl de anexina V-APC (eBioscience). Se incubó la disolución durante 10 min a temperatura ambiente, seguido de lavado con tampón de unión a anexina IX. Luego se tiñeron las células con yoduro de propidio 1 μg/ml en 300 μl de tampón de unión a anexina IX durante 15 min a temperatura ambiente. Se realizó citometría de flujo con un instrumento BD LSRII (BD Biosciences). Se usaron al menos 10.000 eventos para el análisis.

Permeabilidad celular

Se hicieron crecer células MCF7 en placas de 6 pocillos hasta ~80% de confluencia en medio de crecimiento completo. Luego, se añadieron 10 μM de la molécula pequeña indicada y se incubaron las células durante 20 h. Las células se tripsinizaron de la placa, se lavaron dos veces con DPBS IX y se tiñeron con PI 1 μg/ml durante 30 min en hielo. Se cuantificó la permeabilidad celular mediante citometría de flujo usando un citómetro de flujo BD LSRII (BD Biosciences) y filtros Hoechst. Se usaron al menos 10.000 eventos para los análisis (figura 10).

Evaluación de la morfología celular.

Se hicieron crecer células MCF7 en placas de 6 pocillos hasta ~80% de confluencia en medio de crecimiento completo. Luego, se añadieron al pocillo 40 μM de la molécula pequeña indicada y se incubaron las células durante 20 h. Se tomaron imágenes de las células usando un microscopio Olympus IX71 (figura 10).

Síntesis de 1, 2, 4, 5 y sus conjugados con fluoresceína

Se sintetizaron los compuestos 1, 2, 4 y 5 y los conjugados con fluoresceína de 1 y 4 (1-Fl y 4-Fl, respectivamente (figura 11) tal como se describió anteriormente por Velagapudi *et al.* (J. Am. Chem. Soc. 133, 10111-10118 (2011)). Los compuestos 1-Fl, 4-Fl y 5-Fl se usaron para medir las afinidades de unión; el compuesto 2 se usó directamente para medir las afinidades sin conjugación con fluoresceína.

Síntesis de 5-FI

El esquema de síntesis para preparar el compuesto 5-FI se muestra en la figura 11. Se añadió una muestra de 3,4 μmol de N-(2-propinil)-5-fluoresceincarboxamida en metanol a una disolución que contenía 10 μmol de compuesto 5, 34 μmol de $CuSO_4$, 72 μmol de ácido ascórbico recién disuelto, y 0,3 μmol de TBTA. El volumen final se llevó a 1,5 ml con metanol. Se transfirió la mezcla de reacción a un recipiente de reacción para microondas, se añadió una barrita de agitación magnética y se selló el matraz con un septo de teflón y una tapa engarzada de aluminio. El recipiente de reacción se colocó en un sistema EmrysTM Optimizer (Biotage) y se mantuvo la reacción a 110ºC durante 4 h con agitación. Se purificó la mezcla de reacción en bruto mediante HPLC de fase inversa usando un gradiente lineal de disolvente B a del 20% al 100% en el disolvente A durante 60 min. El disolvente A fue TFA al 0,1% (v/v) en agua, mientras que B fue TFA al 0,1% (v/v) en metanol. La pureza del producto se evaluó en una columna Waters Symmetry® C18 de 5 μm 4,6x150 mm usando una bomba de HPLC binaria Waters 1525 equipada con un sistema de detector de absorbancia de λ dual Waters 2487. Las separaciones se completaron a temperatura ambiente usando un caudal de 1 ml/min y un gradiente lineal de disolvente B a del 0% al 100% en el disolvente A durante 50 min. La absorbancia se controló a 220 y 254 nm. t_R = 35 min; rendimiento aislado = 70% (determinado mediante la absorbancia a 496 nm en PBS IX, pH 7,4, usando un coeficiente de extinción de 45000 $M^{-1}cm^{-1}$). EMAR, masa calculada: 890,3580 ($M+H^+$); masa observada: 890,3590 ($M+H^+$).

Ejemplo 2: Compuestos identificados para 22 precursores de miARN relacionados con enfermedad

Este ejemplo describe el desarrollo de un enfoque para identificar moléculas pequeñas iniciales que inhiben la biogénesis de microARN, donde los compuestos se identifican únicamente a partir de las secuencias de ARN.

Con el fin de desarrollar un enfoque computacional para diseñar moléculas pequeñas que se unan al ARN a partir de la secuencia, se desarrolló un algoritmo para analizar las estructuras secundarias de ARN en motivos. Luego, estos motivos se comparan con nuestra base de datos de interacciones de motivo de ARN-molécula pequeña para identificar solapamiento (figuras 3A-3B). La salida son los motivos estructurales de ARN que pueden seleccionarse como diana y las moléculas pequeñas iniciales correspondientes para un ARN de interés. Los compuestos iniciales pueden someterse a prueba para determinar la modulación de la función biológica.

Se validó el enfoque computacional usando microARN (miARN), una clase de ARN que regulan muchos procesos biológicos. Los microARN se transcriben como precursores que se procesan para dar ARN de 21-25 nucleótidos que regulan negativamente la expresión génica a través de la represión de la traducción o la escisión de un ARNm diana (figura 3B) (véase, por ejemplo, Bartel, Cell 136, 215-233 (2009) para una descripción de las funciones reguladoras y de reconocimiento de dianas de microARN).

El objetivo de estos estudios fue usar Inforna para identificar una molécula pequeña que se una a un miARN precursor e inhibe su maduración (véanse las figuras 3A-3B para un diagrama esquemático del proceso Inforna). Las secuencias de todos los miARN precursores conocidos en el transcriptoma humano (Griffith-Jones *et al.*, Nucleic Acids Res. 36, D154-158 (2008)) se descargaron y modelaron en estructuras secundarias (Mathews *et al.*, Proc. Natl Acad Sci. U.S.A. 101, 7287-7292 (2004)). Los miARN precursores se pliegan para dar pequeñas estructuras secundarias de horquilla (figura 3B) que se predicen con precisión a partir de la secuencia (Ambros *et al.*, RNA 9, 277-279 (2003). El conjunto completo de estructuras secundarias se analizó luego mediante Inforna, lo que genera una salida de los motivos que pueden seleccionarse como diana en cada ARN y las moléculas pequeñas correspondientes que se unen a esos motivos. Al extraer todos los miARN precursores en el transcriptoma humano para determinar su solapamiento con la base de datos de moléculas pequeñas-motivos de ARN, Inforna investigó más de 5.400.000 interacciones potenciales (los motivos contenidos en 1.048 precursores de miARN (aproximadamente 6.850) con 792 interacciones de motivo de ARN-molécula pequeña alojadas en la base de datos de los inventores). En este estudio, los inventores requerían que el motivo que puede seleccionarse como diana y sus pares de bases de cierre fueran coincidencias exactas para los motivos en la base de datos. Anteriormente, se ha demostrado que la identidad de los pares de cierre del bucle interno puede afectar drásticamente a la estructura del bucle y, por tanto, el reconocimiento por parte de una molécula pequeña (Wu & Turner, Biochemistry 35, 9677-9689 (1996); SantaLucia & Turner, Bioquímica 32, 12612-12623 (1993)).

A continuación, los inventores refinaron las interacciones principales basándose en los siguientes criterios:

(i) el motivo que puede seleccionarse como diana debe estar en un sitio de procesamiento Drosha o Dicer, que se escinden para producir pre-miARN y miARN maduros, respectivamente. El procesamiento es necesario para que se produzca el miARN maduro y activo (Bartel, Cell 116, 281-297 (2004)); y

(ii) el miARN debe ser causante de enfermedad.

Para validarse como miARN regulado por incremento en la enfermedad, la reversión del fenotipo con oligonucleótidos (es decir, antagomires) se estableció previamente (Krutzfeldt *et al.*, Nature 438, 685-689 (2005); Ebert & Sharp, RNA 16, 2043-2050 (2010); Obad *et al.*, Nat. Genet. 43, 371-378 (2011). Para los miARN

subexpresados en una enfermedad, la validación se estableció previamente añadiendo el miARN a las células y observando la mejora de defectos asociados con la enfermedad. En resumen, se identificaron moléculas pequeñas iniciales para veintidós (22) precursores de miARN diferentes que son causantes de enfermedades incluyendo cánceres de próstata, mama, ovario y páncreas, enfermedad de Parkinson y enfermedad de Alzheimer (figuras 3C-3E y tabla 1).

Ejemplo 3: Análisis de StARTS de compuestos iniciales

Este ejemplo describe las relaciones estructura-actividad a través de secuenciación (StARTS, por sus siglas en inglés) para investigar computacionalmente la idoneidad de las interacciones de moléculas pequeñas y precursores de miARN.

El análisis de StARTS de los compuestos coincidentes se usó para evaluar la idoneidad de cada molécula pequeña inicial para la unión del motivo de ARN correspondiente en la diana de miARN precursor (Velagapudi *et al.*, Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011); véase también el ejemplo 1). La figura 2 muestra las posibles secuencias aleatorizadas (N) que pueden estar en el motivo de ARN de BBI 3x3 mostrado en la figura 1, mientras que la figura 3C ilustra gráficamente que las puntuaciones de idoneidad para los diferentes compuestos varían según con qué miARN estén interaccionando.

En resumen, StARTS determina la idoneidad de un agente de unión de molécula pequeña asignando significación estadística para cada característica en un motivo (como un tramo de GC) que contribuye positiva o negativamente a la unión. La significación estadística se expresa como una puntuación Z; las puntuaciones Z se suman para obtener una puntuación de ΣZ para cada motivo (figura 3). Las puntuaciones de ΣZ predicen con precisión la afinidad y la selectividad de una molécula pequeña para todos los miembros de una biblioteca de ARN (como una biblioteca de bucles internos de nucleótidos 3x3, que tiene 4.096 miembros (figuras 1-2) con una afinidad que se correlaciona positivamente con la puntuación de ΣZ. Se calcula la puntuación de idoneidad normalizando la puntuación de ΣZ con respecto a la puntuación de ΣZ más alta del análisis de StARTS. Por ejemplo, una puntuación de idoneidad de 100 indica que una interacción de motivo de ARN-molécula pequeña dada es el/los motivo(s) de ARN de mayor afinidad que se une(n) a una molécula pequeña dada de una biblioteca de motivos de ARN. Asimismo, una puntuación de idoneidad de 80 significa que una interacción de ARN-molécula pequeña es el 80% tan adecuada como la interacción óptima (puntuación de idoneidad = 100).

El análisis de StARTS determinó que las interacciones más idóneas eran: el compuesto 1 con el precursor de miR-96 (puntuación de idoneidad = 100), el compuesto 2 con el precursor de miR-210 (puntuación de idoneidad = 95), y el compuesto 3 con el precursor de miR-182 (puntuación de idoneidad = 75) (figuras 3C, 3D, 5A). Se debe tener en cuenta que las afinidades y selectividades de estas interacciones de motivo de ARN-molécula pequeña y sus puntuaciones de idoneidad se midieron en estudios previos (Disney *et al.*, J. Am. Chem. Soc. 130, 11185-11194 (2008); Velagapudi *et al.*, J. Am. Chem. Soc. 133, 10111-10118 (2011)). Debido a la alta idoneidad de estas interacciones, los compuestos 1 - 3 se sometieron a prueba para determinar la inhibición de la biogénesis de microARN en líneas celulares primarias en las que los microARN se expresan altamente y provocan fenotipos asociados con enfermedad. Específicamente, se sometieron a prueba los compuestos 1 y 3 en la línea celular de cáncer de mama MCF7 (Guttilla & White, J. Biol. Chem. 284, 23204-23216 (2009)) mientras que se sometió a prueba el compuesto 2 en la línea celular de cáncer renal ACHN (Redova *et al.*, Tumor Biol. 34, 481-491 (2013)). Las otras interacciones principales identificadas en la figura 3C también podrían aprovecharse para diseñar moléculas bioactivas dirigidas a otros precursores de miARN. Las puntuaciones de idoneidad más bajas de estas interacciones en relación con las seleccionadas para el seguimiento sugieren que pueden requerir optimización, lo que podría lograrse mediante enfoques de ensamblaje modular (Childs-Disney *et al.*, ACS Chem. Biol. 7, 856-862 (2012); Pilch *et al.*, Proc. Natl Acad Sci. U. S. A. 93, 8306-8311 (1996); Pushechnikov *et al.*, J. Am. Chem. Soc. 131, 9767-9779 (2009); Lee *et al.*, ACS Chem. Biol. 4, 345-355 (2009)) o búsqueda de similitud química (Parkesh *et al.*, J. Am. Chem. Soc. 134, 4731-4742 (2012); Kumar *et al.*, ACS Chem. Biol. 7, 496-505 (2012); Pinto *et al.*, J. Med. Chem. 51, 7205-7215 (2008); Disney *et al.*, ACS Chem. Biol. 7, 1711-1718 (2012); Luzhkov *et al.*, Bioorg. Med. Chem. 15, 7795-7802 (2007)).

Ejemplo 4: Bioactividad y selectividad de compuestos que se predice que seleccionan como diana precursores de microARN mediante Inforna

Este ejemplo ilustra la bioactividad y selectividad de las moléculas pequeñas que se predice que seleccionan como diana precursores de microARN mediante Inforna Inicialmente, los compuestos 1, 2 y 3 se sometieron a prueba para determinar la inhibición de la producción del miARN maduro a partir de sus precursores diana en líneas celulares primarias mediante RT-PCR cuantitativa en tiempo real (figuras 5A-5B). Cada compuesto inhibe la biogénesis de su miARN precursor diana, aunque en grados variables (figura 5B): el compuesto 1 reduce el nivel de expresión de miR-96 en un 90% a 40 μM; el compuesto 2 reduce la formación de miR-210 en un 60% a 500 nM; y el compuesto 3 reduce la producción de miR-182 en un 40% a 200 μM.

El tratamiento de células primarias con mayores concentraciones de compuesto 2 conduce a una potencia reducida,

tal vez debido a una falta de selectividad en estas concentraciones elevadas. Estas diferencias en la bioactividad podrían deberse a diferencias en la afinidad, selectividad, permeabilidad y localización celular. Es importante destacar que estos estudios demuestran que las moléculas pequeñas que seleccionan como diana sitios o bien Dicer (compuesto 2) o bien Drosha (compuestos 1 y 3) en miARN precursores pueden inhibir la biogénesis en cultivo celular. Es importante destacar que estos datos muestran que pueden diseñarse moléculas pequeñas bioactivas que seleccionan como diana ARN a partir de una única secuencia de una manera de transcriptoma completo sin sesgo de diana.

Curiosamente, los precursores de miR-96, -182 y -183 (figura 5A) se transcriben como un solo transcrito (Xu *et al.*, J. Biol. Chem. 282, 25053-25066 (2007)). Por tanto, la biogénesis de un microARN puede servir como control interno para los demás. Por tanto, se estudiaron el compuesto 1 (que selecciona como diana el miR-96 precursor) y el compuesto 3 (que selecciona como diana el miR-182 precursor) para determinar la disminución de la producción de los otros dos microARN en la línea celular MCF7 mediante qRT-PCR. Aunque el compuesto 3 disminuye la producción de la diana deseada, miR-182, en aproximadamente un 40% cuando se les dosifica a las células compuesto 200 μM, se observó un efecto menor, pero significativo, sobre la expresión de miR-96 que indica una selectividad subóptima (figura 5B). En cambio, el compuesto 1 silencia de manera eficiente y selectiva la producción de miR-96 a 40 μM mientras que no afecta a miR-182 o -183 (figura 5B). Así, el compuesto 1 proporciona una mayor atenuación del microARN diana que los compuestos 2 y 3 y es más selectivo que 3.

Para confirmar que el sitio de unión del compuesto 1 en el precursor de miR-96 fue predicho con precisión mediante Inforna Los inventores completaron los ensayos de protección frente a nucleasas. Estos estudios confirman que el compuesto 1 se une al sitio de procesamiento Drosha (figura 9A). Por otra parte, el compuesto 1 inhibe la escisión por Drosha de pri-miR-96 *in vitro* (figura 9B) e *in vivo*, como lo demuestra un aumento de la cantidad de pri-miARN y una reducción de los miARN pre y maduros en las células tratadas, tal como se espera si el compuesto 1 se une al sitio Drosha (figura 9C). A continuación, los inventores compararon la selectividad del compuesto 1 a un oligonucleótido de ácido nucleótido bloqueado (LNA) que es complementario a los nucleótidos 2-9 en la región de semilla de microARN-96 (figura 5C). El oligonucleótido de LNA dirigido a miR-96 se estudió para determinar el silenciamiento de miR-96, miR-182 y miR-183; los complejos miR-182:LNA y miR-183:LNA contienen, cada uno, un único apareamiento erróneo. Curiosamente, el LNA solo discrimina de forma moderada entre miR-96 y miR-183 en todas las concentraciones sometidas a prueba (1 - 200 nM) (figura 5C). A una concentración de 50 nM, el LNA silencia alrededor del 90% de la expresión de miR-96 y el ~50% de la expresión de miR-183. Se han observado previamente efectos no selectivos de los oligonucleótidos sobre el silenciamiento de miARN específicos (Stenvang *et al.*, Silence 3, 1 (2012)). En cambio, a una concentración de compuesto 1 que silencia el 90% de la expresión de miR-96, el microARN-182 se ve afectado solo en el 15% y el microARN-183 no se ve afectado (figura 5C). En conjunto, las moléculas pequeñas que seleccionan como diana la estructura secundaria de miARN precursores pueden ser moduladores más selectivos de la función que los oligonucleótidos que seleccionan como diana la secuencia de ARN, tal vez debido a la degeneración energética en la secuencia de selección como diana (incluso con apareamientos erróneos) que no se produce cuando se selecciona como diana la estructura con una molécula pequeña. Otra ventaja importante de las moléculas pequeñas es su permeabilidad celular, una propiedad no innata de los oligonucleótidos.

Ejemplo 5: Los efectos posteriores del compuesto 1

Este ejemplo ilustra el efecto del compuesto 1 sobre las dianas posteriores de miR-96. El microARN-96 está regulado por incremento en el cáncer y está vinculado con la transformación oncogénica mediante el silenciamiento de *FOXO1* (proteína de la caja de cabeza de tenedor (*forkhead box*) O1) a través de la represión traslacional (Guttilla & White, J. Biol. Chem. 284, 23204-23216 (2009); Xie, *et al.*, Blood 119, 3503-3511 (2012)). Los factores de transcripción FOXO funcionan como reguladores de la progresión del ciclo celular, incluida la apoptosis (Dansen & Burgering, Trends Cell Biol. 18, 421-429 (2008).

Se sometió a prueba el compuesto 1 para determinar su capacidad para aumentar la expresión de la proteína FOXO1 usando un sistema modelo de luciferasa. Se insertó una secuencia diana totalmente complementaria a miR-96 se insertó de manera posterior al gen de la luciferasa de *Renilla* (Guttila *et al*. J. Biol. Chem. 284, 23204-23216 (2009)). Una molécula pequeña que inhibe la maduración de miR-96 aumentará la expresión de luciferasa. Se transfectaron células MCF7 con un plásmido que expresa el sistema modelo de luciferasa, seguido de un tratamiento con 1. De manera acorde a la disminución de la producción de miR-96 observada en los experimentos de qRT-PCR (figura 5C), el compuesto 1 estimula la producción de luciferasa, lo que indica la inhibición de la maduración de miR-96 (figura 6A). Específicamente, el compuesto 1 aumenta la producción de luciferasa en aproximadamente 2,2 veces cuando se tratan las células con 40 μM de compuesto. Es importante destacar que el compuesto 1 no afecta a la producción de luciferasa cuando se usa un plásmido que codifica para un 3' UTR de *FOXO1* que no responde a miR-96, lo que confirma que miR-96, y no la UTR de *FOXO1*, es la diana del compuesto 1 (figura 6a).

A continuación, los inventores sometieron a prueba el compuesto 1 por su capacidad para aumentar los niveles endógenos de la proteína FOXO1 mediante inmunotransferencia de tipo Western (figura 6B). Cuando se trataron

células MCF7 con 40 μM de compuesto 1, se observó un aumento aproximado de 2,5 veces en los niveles de proteína FOXO1, que concuerda con los experimentos con luciferasa. No se observó ningún efecto sobre la expresión de un control de GAPDH (figura 6B). La regulación positiva de FOXO1 estimula la apoptosis (Huang & Tindall, Future Oncol. 2, 83-89. (2006) 37). Por tanto, los inventores determinaron si el compuesto 1 estimula la apoptosis tardía a través de un ensayo TUNEL. Cuando se tratan células MCF7 con 40 μM de compuesto 1, aproximadamente el 40% de las células son positivas a TUNEL (figura 6C). Como prueba secundaria de apoptosis, se emplearon ensayos de anexina V/I de propidio, ya que pueden distinguir la necrosis de la apoptosis temprana. Como era de esperar, estos estudios muestran que el compuesto 1 estimula la apoptosis temprana, no la muerte celular global por necrosis (figura 6C; figuras 8A-8B).

Ejemplo 6: el compuesto 1 puede estimular la apoptosis

Este ejemplo ilustra que el compuesto 1 induce apoptosis mediante modulación de la ruta de regulación de miR-96-*FOXO1*.

Tal como se muestra en la figura 6E, la adición de compuesto 1 aumenta drásticamente el porcentaje de células positivas para TUNEL (debe tenerse en cuenta que la segunda barra es mucho mayor que la primera barra). Si el compuesto 1 induce la apoptosis seleccionando como diana selectivamente miR-96, entonces la eliminación de *FOXO1* debe afectar a la apoptosis. Se redujo la expresión de *FOXO1* a través de ARNip (figura 6E) para ilustrar los efectos de la reducción de la expresión de *FOXO1* sobre la apoptosis. Tal como se muestra, la atenuación de la expresión de *FOXO1* aumenta el porcentaje de células positivas para TUNEL, lo que indica que la atenuación de *FOXO1* también aumenta o induce apoptosis. Cuando se aplica ARNip de *FOXO1* a células que luego se tratan con compuesto 1, se observa una reducción del 70% de la apoptosis en comparación con las células transfectadas con un ARNip contra *GAPDH* (control) y tratadas con compuesto 1 (figura 6E). No es de extrañar que el uso de un ARNip para ablacionar el ARNm de *FOXO1* de las células no elimina completamente el efecto apoptótico del compuesto 1 porque los miARN pueden seleccionar como diana muchos ARNm diferentes, y pueden silenciar diferentes ARNm simultáneamente (Lewis *et al.*, Cell 120, 15-20 (2005)). Estos estudios, sin embargo, demuestran que la ruta del ARNm de miR-96-*FOXO1* está regulada por el compuesto 1, ilustrando que el compuesto 1 es un inductor específico de apoptosis. Esta selectividad se remonta a la modulación de un oncogén más que afectar inespecíficamente a la función celular. La mayoría de los agentes terapéuticos contra el cáncer, tales como cis-platino y clorambucilo, atacan las biomoléculas de manera inespecífica, dando lugar a efectos secundarios (Wolkenberg y Boger, Chem. Rev. 102, 2477-2495 (2002)).

Ejemplo 7: Selectividad del compuesto 1 para microARN-96

Se predijo que el compuesto 1 se uniría a los sitios de procesamiento Dicer y Drosha en precursores de miARN distintos del precursor de miR-96. Sin embargo, estas interacciones de Dicer y Drosha predichas parecen ser menos idóneas que entre el compuesto 1 y el precursor de miR-96 (figura 7A).

Por tanto, se evaluó la selectividad del compuesto 1 para modular la expresión de 149 miARN asociados con enfermedad y altamente abundantes (tablas 1 y 2) mediante qRT-PCR (figura 7B). Tal como se muestra en la figura 7, el único miARN que se ve afectado significativamente por el compuesto 1 es miR-96.

Estos estudios confirman la selectividad del compuesto 1 para la diana diseñada a nivel de transcriptoma completo. Por tanto, parece que el compuesto 1 proporciona un nivel de selectividad sin precedentes para una molécula pequeña que modula la función del ARN y tiene una selectividad que puede ir más allá de la observada con algunos oligonucleótidos que seleccionan como diana miARN, como lo demuestran los estudios en la figura 5C.

Ejemplo 8: Comparación de Inforna A con enfoques de química medicinal tradicionales

El sistema modelo de luciferasa descrito anteriormente proporciona un ensayo robusto para someter a prueba otras moléculas pequeñas para modular la ruta de miR-96-*FOXO1*. Por tanto, este sistema se usó para comparar el diseño de moléculas pequeñas a través de Inforna con los enfoques de química médica más tradicionales: selección y optimización inicial mediante la síntesis de derivados de compuestos. En primer lugar, se sometió a prueba una biblioteca de moléculas pequeñas construida previamente que está sesgada por la unión a ARN (28 compuestos totales; figura 12A) (Tran y Disney, Nat. Commun. 3, 1125 (2012)). Ninguno de los compuestos estimuló la producción de luciferasa a 40 μM (figura 12B), lo que indica que: (i) los compuestos no son bioactivos; (ii) los compuestos de examen son menos efectivos que diseñar compuestos con Inforna incluso cuando la biblioteca química está sesgada para la unión a ARN; y (iii) la ruta de miARN-*FOXO1* no tiene capacidad farmacológica fácilmente.

Un método empleado comúnmente para la optimización de fármacos es la síntesis de derivados de compuestos, es decir, moléculas pequeñas químicamente similares. Por tanto, se sometieron a prueba tres compuestos que son químicamente similares a 1, los compuestos 2, 4 y 5 (figura 3D), tal como se determina mediante sus coeficientes de Tanimoto de forma (Hawkins *et al.*, J. Med. Chem. 50, 74-82 (2007). Los coeficientes de Tanimoto de forma determinan cuantitativamente la similitud tridimensional de dos compuestos; los valores oscilan entre 0 (sin similitud)

y 1 (similitud completa) (*id*.). Los coeficientes de Tanimoto de forma para los compuestos 2, 4 y 5 en comparación con 1 son de 0,94, 0,89 y 0,80, respectivamente, lo que ilustra la similitud cuantitativa entre estos compuestos. Aunque los cuatro compuestos se basan en un andamiaje de bencimidazol, la inspección visual de la estructura sugiere que los compuestos 2 y 5 podrían unirse al ARN con mayor afinidad que el compuesto 1 debido a la mayor área superficial y a los donadores y aceptores de enlaces de hidrógeno adicionales. Cada uno de estos factores sugiere que estos compuestos podrían modular la maduración de miR-96. A pesar de sus similitudes, los compuestos 2, 4 y 5 tienen representaciones gráficas de idoneidad muy diferentes para la unión al sitio diana de 1 en el precursor de miR-96 (figura 4). De hecho, Inforna predice que ninguno de los compuestos debe unirse con avidez. La falta de afinidad de unión de estos compuestos para el precursor de miR-96 se confirmó experimentalmente tal como se muestra en las figuras 4C-4I.

Se estudiaron las actividades de los compuestos 2, 4 y 5 en la línea celular MCF7. Los compuestos son inactivos a una concentración de 40 μM en todos los ensayos: (i) no afectan a los niveles de expresión de miR-96 (figura 4J); (ii) no estimulan la producción de luciferasa en el sistema modelo de microARN-96-*FOXO1* (figura 6a); y (iii) no inducen apoptosis (figura 6C). La inactividad de 2, 4 y 5 no se debe a diferencias en la permeabilidad celular (figura 10). Así, 2DCS, StARTS e Inforna proporcionan una identificación fiable de moléculas pequeñas que son capaces e incapaces de seleccionar como diana los ARN, y estas predicciones son más precisas que los enfoques de química médica convencionales, tales como el examen y la búsqueda de similitud química. Además, la tasa de coincidencia de Inforna es superior al examen de alto rendimiento y es más rápida que los enfoques computacionales, tales como el diseño basado en estructuras y el acoplamiento. Por ejemplo, un importante estudio utilizó una molécula pequeña que se acoplaba a un conjunto dinámico de ARN para identificar compuestos que se unen al ARN de TAR del VIH y modular la función (Stelzer *et al.*, Nat. Chem. Biol. 7, 553-559 (2011)). Sin embargo, los inventores han demostrado que Inforna permite el diseño fiable de moléculas pequeñas bioactivas a partir de la secuencia sola y sin tener que completar estudios estructurales o de acoplamiento a menudo laboriosos.

Ejemplo 9: Proceso Inforna

Este ejemplo describe los aspectos computacionales del proceso Inforna.

Muchas interacciones de motivo de ARN-molécula pequeña se han identificado mediante el uso de una biblioteca de moléculas pequeñas frente a una plataforma de examen de bibliotecas de ARN desarrollada por los inventores y sus colaboradores, actualmente de alrededor 1.500. En general, se necesitan herramientas computacionales para ensamblar y procesar esta información de manera efectiva, de modo que las moléculas de ARN (especialmente los ARN celulares más grandes) puedan convertirse en dianas farmacológicas efectivas. Para habilitar la búsqueda programática fácil de estas interacciones frente a un ARN diana, los inventores construyeron un conjunto de datos de interacciones de motivo de ARN-molécula pequeña e Inforna que puede ser una interfaz web para buscar en la base de datos.

Descripción de la base de datos

Un esquema de la base de datos se muestra en la figura 13A. La base de datos contiene una lista de todas las interacciones de motivo de ARN-molécula pequeña identificadas mediante 2DCS o mediante otros métodos. A cada entrada se le asignan los siguientes parámetros, que comprenden cuatro tablas que están vinculadas para búsquedas fáciles: (i) un identificador único de molécula pequeña; (ii) un identificador de motivo único; (iii) el tipo de motivo; (iv) el tamaño de motivo; (v) la secuencia del motivo; (vi) los pares de bases de cierre; (vii) la puntuación de idoneidad para el motivo (indica la idoneidad global de las interacciones motivo de ARN-molécula pequeña y está altamente correlacionada con la afinidad); (viii) la constante de disociación, K_d, si se mide; y, (ix) otras notas, incluidos los números de referencia de publicación de identificación PubMed.

Tabla de motivos:

- id INT crea una columna “id” que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.
- VARCHAR (#) indica cadenas de texto de longitud variable donde “#” indica el número máximo de caracteres con la cadena.
- DOBLE es un tipo de datos numérico aproximado que puede consistir en un número entero, una fracción o ambos e indica que el contenido de la columna es de naturaleza numérica.

(i) Identificador de motivo: a cada motivo de ARN se le asigna un identificador numérico. La base de datos actual tiene ~1500 parejas de motivo de ARN-molécula pequeña.

(ii) ID del motivo: a cada motivo se le asigna un número.

(iii) Secuencia con pares de cierre: secuencia completa del motivo de ARN

- Secuencia 5': la secuencia 5' del motivo de ARN con el par de bases de cierre.
- Secuencia 3': la secuencia 3' del motivo de ARN con el par de bases de cierre
- Secuencia con par de bases: la secuencia completa del motivo de ARN y los pares de bases de cierre.
- Nucleótidos de bucle: secuencia del motivo de ARN excluyendo pares de bases de cierre
- Pares de cierre (5', 3'): secuencia de los pares de bases de cierre en 5' y 3'

(iv) ID de molécula pequeña: nombre de molécula pequeña (también conocida como ligando) que se une al motivo de ARN

(v) INT de tamaño: enlaces a la tabla Tipo y tamaño de motivo para rellenar automáticamente la tabla y asignar un identificador único a cada fila; crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.

(vi) Tipo de motivo: a cada tipo de motivo (horquilla, bucle interno de nucleótidos 3 x 3, etc.) se le asigna un identificador numérico único.

(vii) Puntuación de idoneidad: representa la idoneidad del motivo de ARN para la unión a la molécula pequeña (molécula pequeña).

(viii) Len: indica la longitud (solo valores enteros) de la región aleatorizada o nucleótidos de bucle.

(viii) Kd nM error nm: afinidad de unión (constante de disociación; K_d) o CI_{50} si se determina se muestra en la salida.

(x) PMID: ID de PubMed

Tabla de tamaño de motivo. La tabla de tamaño de motivo está vinculada a la tabla de motivos tal como se muestra en la figura 13A.

- id INT crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.
- VARCHAR (#) indica cadenas de texto de longitud variable donde "#" indica el número máximo de caracteres con la cadena.

(i) ID del tamaño de motivo: cada tipo de motivo se representa con un identificador numérico, que se anota dentro de esta tabla. El "tamaño de motivo" puede tener diferentes formas funcionales. Por ejemplo, para las horquillas y protuberancias, el tamaño de motivo es simplemente un número que indica el número de nucleótidos en el bucle. La forma funcional para los bucles internos es "AxB", donde A indica el número de nucleótidos no apareados en 5' y B el número de nucleótidos no apareados en 3'. La forma funcional para el tamaño de motivo de un bucle de ramificación múltiple puede tener múltiples formas, como "AxBxC" o "AxBxCxD", lo que indica una unión de 3 ó 4 vías, respectivamente.

Tabla de moléculas pequeñas. La tabla de moléculas pequeñas está vinculada a la tabla de motivos tal como se muestra en la figura 13A.

- id INT crea una columna "id" que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.
- VARCHAR (#) indica

(i) ID de molécula pequeña: un identificador numérico asignado a cada molécula pequeña a partir de 1. La base de datos actual tiene 24 moléculas pequeñas.

(ii) Nombre de molécula pequeña: nombre asignado a cada molécula pequeña para identificarla.

(iii) SMILES: SMILES, o sistema de entrada de línea de entrada molecular simplificada, es un texto

que describe la estructura de la molécula pequeña. El texto de SMILES puede introducirse en varios programas para reconstituir la estructura de la molécula pequeña. Hay una carpeta separada de las estructuras con archivos JPEG que se emiten para cada búsqueda.

Tabla de motivos que tienen ligandos. La tabla de motivos que tienen ligandos está vinculada a la tabla de motivos tal como se muestra en la figura 13A. Esta tabla correlaciona el ID de motivo con el ID de molécula pequeña.

- id INT crea una columna “id” que se incrementará automáticamente cada vez que se añada una nueva entrada a la tabla.
- VARCHAR (#) indica cadenas de texto de longitud variable donde “#” indica el número máximo de caracteres con la cadena.

(i) ID de molécula pequeña: un identificador numérico asignado a cada molécula pequeña a partir de 1. La base de datos actual tiene 24 moléculas pequeñas.

(ii) ID de motivo: cada tipo de motivo se representa con un identificador numérico, que se anota en la tabla de motivos.

Descripción del algoritmo/motor de búsqueda

En la figura 13B se muestra un esquema del flujo de datos del motor de búsqueda. El software Inforna acepta un archivo .CT (un archivo de texto simple que describe la estructura secundaria de un ARN) con dos opciones de búsqueda: nucleótidos de bucle de búsqueda SIN pares de bases de cierre y nucleótidos de bucle de búsqueda CON pares de base de cierre. El usuario puede seleccionar un archivo .CT o un archivo zip que contenga varios archivos .CT. Después de elegir el archivo .CT y presentar la opción de búsqueda, la aplicación primero llama a una función para analizar el archivo .CT. Esta función aplica un algoritmo de análisis, que crea un parámetro de búsqueda de base de datos. Una vez que se completa la función de análisis de archivos .CT, se crea otro parámetro de búsqueda según la opción de búsqueda seleccionada. Cuando se completan estas funciones, otra función convierte estos parámetros de búsqueda en una instrucción SQL usada para consultar la base de datos. Esta instrucción SQL consiste en los campos en la base de datos que se consultan, las tablas dentro de la base de datos que contienen los campos consultados y los criterios de búsqueda, que filtran el conjunto de resultados según las opciones seleccionadas del usuario.

A cada registro en la base de datos se le asigna un ID de motivo, identificador de motivo, tipo de motivo, tamaño de motivo, par de bases de cierre, puntuación z suma para el motivo, que indica significación estadística y está altamente correlacionada con la afinidad, puntuación de idoneidad, constante de disociación (K_d) si se mide y el ID de referencia de publicación PMID. Hay dos funciones definidas en la base de datos que se usan para analizar los archivos CT. Dependiendo de los criterios de búsqueda, la base de datos no devolverá coincidencias o un conjunto de registros que coincidan con los criterios de búsqueda. Este conjunto de registros se hace pasar ahora a la interfaz de usuario donde se procesa para aplicar cualquier cambio de formato. Una vez completado esto, este conjunto de registros rellena la cuadrícula de la interfaz de usuario. Los valores se muestran en el siguiente orden: nombre de archivo CT, estructura de compuesto que es una imagen que visualiza el campo SMILES, motivo de consulta, motivo en ARN diana, nucleótidos de bucle, puntuación de idoneidad, identificador de bucle, constante de disociación (K_d) si se mide y el enlace de referencia de publicación PMID (figura 3). Dado que los resultados de la búsqueda pueden ser bastante grandes, se aplica un límite de 200 filas para reducir la carga del servidor y la demora entre las presentaciones de búsqueda. Si el usuario desea ver todos los registros, está disponible una opción de exportación a Excel. Esta opción no se limita a búsquedas con más de 200 registros devueltos.

Salida de una consulta de base de datos. El resultado de una consulta de base de datos incluye (figura 3):

- la estructura del ARN que se consultó (como un archivo .ct)
- estructura de la molécula pequeña que se une a un motivo en el ARN consultado. El texto asociado de SMILES está disponible haciendo clic en la estructura de la molécula pequeña.
- motivo dentro de la base de datos que es similar o coincide exactamente con el/los motivo(s) en el ARN consultado que se une a la molécula pequeña
- motivo en el ARN consultado que se predice que se une a la molécula pequeña, molécula pequeña
- nucleótidos de bucle en el motivo de ARN consultado.
- puntuación de idoneidad

- identificador de bucle
- constante de disociación (K_d o CI_{50})
- PMID que está vinculado a la base de datos en ncbi.nlm.nih.gov/pubmed/.

Ejemplo 10: moléculas pequeñas de diseño que seleccionan como diana el precursor del microARN-96 y desencadenan apoptosis selectivamente en líneas celulares de cáncer de mama.

Este ejemplo describe la identificación de moléculas pequeñas que pueden seleccionar como diana bucles de dímeros en las estructuras secundarias de moléculas de ARN, y la capacidad de tales moléculas pequeñas para reducir los niveles de microARN-96 y desencadenar apoptosis en células cancerosas

Materiales y métodos

Productos químicos: se adquirieron resina de amida Rink protegida por Fmoc, diisopropilcarbodiimida (DIC) y 1-hidroxi-7-azabenztriazol (HOAT), de Advanced ChemTech. Se adquirieron 1-etil-3-(3-dimetilaminopropil)carbodiimida (EDC), propilamina y ácido bromoacético de Acros Organics. Se adquirió propargilamina de Combi-Blocks. Se adquirieron N,N-dimetilformamida (DMF), diclorometano y metanol (grado ACS y grado HPLC) de Fisher Scientific. Se adquirió N,N-dimetilformamida seca (dDMF) de EMD. Se adquirieron N,N-diisopropropiletilamina (DIEA) y ácido trifluoroacético (TFA) de Alfa Aesar. Se adquirió piperidina de Sigma-Aldrich.

Espectros de masas: Los espectros de masas se registraron en un analizador MALDI TOF/TOF 4800 plus.

HPLC preparativa: Los peptoides se purificaron usando una columna de 5 μM C18 Atlantis Prep T3 de fase inversa o una columna Sunfire Prep C18 5 μM 19x150 mm. Se completaron las separaciones de HPLC usando una bomba de HPLC binaria Waters 1525 equipada con un sistema de detector de absorbancia dual Waters 2487. Se empleó un gradiente lineal de B a del 20% al 100% en A durante 60 min y un caudal de 5 ml/min. (A: agua + TFA al 0,1% (v/v); B: metanol + TFA al 0,1% (v/v)).

HPLC analítica: Se evaluó la pureza en una columna Waters Symmetry C18 de 5 μm 4,6x150 mm de fase inversa a temperatura ambiente. Se aplicó un caudal de 1 ml/min y un gradiente lineal de B a del 0% al 100% en A durante 60 min. La absorbancia se controló a 220 y 345 nm.

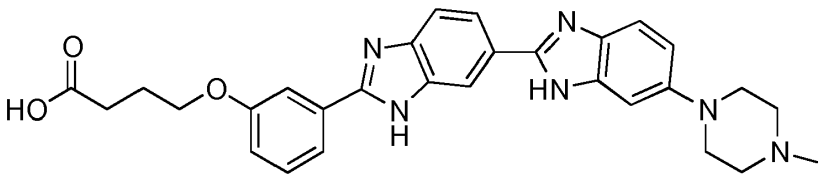
Protocolo general para la síntesis de peptoides: Se sintetizaron peptoides mediante un protocolo de oligomerización con soporte de resina convencional[1]. Se permitió que se hinchase la resina de amida Rink protegida con Fmoc (200 mg, 138 μmol) con un nivel de sustitución de 0,69 mmol/g durante 5 min cada uno en DCM y DMF con agitación. La resina se desprotegió con piperidina al 20% en DMF (3 ml, 2x 20 min) a temperatura ambiente.

Etapa de acoplamiento: entonces se lavó la resina con DMF (3x 5 min). Se acopló el ácido bromoacético a la amina unida a la resina en presencia de 5 equivalentes de ácido bromoacético y 5 equivalentes de DIC en 3 ml de dDMF. Se calentó la mezcla de reacción en un microondas de Panasonic a una potencia del 10% (70 vatios) (2x30 s). Entonces se lavó la resina con dDMF (3x 5 min).

Etapa de desplazamiento: (a) Introducción del homólogo de clic: entonces se trató la resina con 10 equivalentes de propargilamina (1,38 mmol, 88 μl) en 3 ml de DMF en un microondas de Panasonic a una potencia del 10% (70 vatios) (1x30 s) y se agitó a temperatura ambiente durante 2 h. (b) Extensión de cadena con espaciador (propilamina): se repitió el acoplamiento con ácido bromoacético después de la introducción de propargilamina. Luego, se trató la resina con 10 equivalentes de propilamina (1,38 mmol, 113 μl) en 3 ml de DMF en un microondas de Panasonic a una potencia del 10% (70 vatios) (1x30 s) y se agitó a temperatura ambiente durante 20 min. Entonces se lavó la resina se con dDMF y la reacción con propilamina se repitió una vez más. Se repitió la etapa (b) hasta que se obtuvo la longitud requerida (n = de 1 a 4) del peptoide.

Conjugación de los módulos de ligando al andamiaje de peptoide: Se empleó el siguiente procedimiento:

(a) se acopló un carboxilato de Ht con la siguiente estructura al extremo del esqueleto de peptoide.



Se trató directamente el peptoide (50 mg, 34,5 μmol) de la etapa anterior con una disolución de carboxilato

de Ht (27 mg, 52 μmol), EDC (8 mg, 52 μmol), HOAT (7 mg, 52 μmol) y DIEA (66 μL, 345 μmol) en 2 ml de dDMF en un vial para microondas en un instrumento Biotage Initiator+ a 75ºC durante 2 h. Se lavó la resina luego con DMF (3x 5 min).

(b) Se añadió azida de BSH en el peptoide de la etapa previa mediante la reacción de cicloadición dipolar de Huisgen (HDCR). Se trató directamente el peptoide (50 mg, 34,5 μmol) de la etapa previa con una disolución de azida de BSH (7 mg, 8,6 μmol), catalizador de Cu (I) (1,5 mg, 2,6 μmol) y DIEA (66 μl, 345 μmol) en 2 ml de dDMF en un vial para microondas a 120ºC durante 2 h en un instrumento Biotage Initiator +. Se lavó la resina luego con DMF (3 x 5 min) seguido de DCM (3 x 5 min) antes de escindir el peptoide de la resina en TFA: DCM (1:1) durante 15 min a temperatura ambiente. Se eliminó el disolvente a vacío y se purificó el producto en bruto mediante HPLC tal como se describe en los métodos generales.

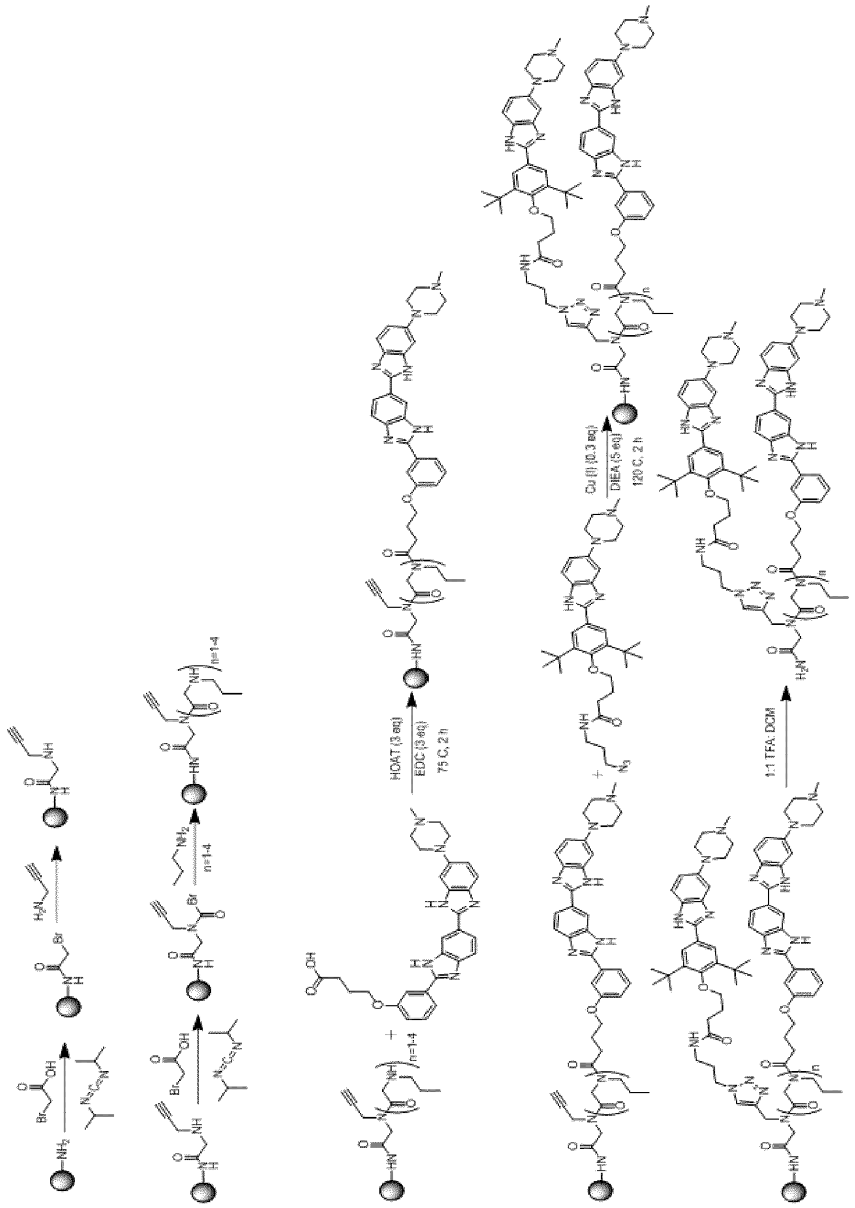
Métodos generales para cultivo celular: Se hicieron crecer células MDA MB 231 en medio de Roswell Park Memorial Institute 1640 (RPMI 1640) (Cellgro) complementado con FBS al 10% (Cellgro) y penicilina-estreptomicina (MP Biomedicals). Se cultivaron células MCF-10A en medio de Eagle modificado por Dulbecco/F12 (DMEM/F12) (Cellgro) complementado con FBS al 10%, EGF 20 ng/ml, hidrocortisona 0,5 μg/ml (Pfaltz and Bauer Inc.), toxina del cólera 100 ng/ml (Sigma), insulina 10 μg/ml (Gemini Bio-Products) y penicilina-estreptomicina.

Ensayos de anexina V/PI: se hicieron crecer células MDA MB 231 o MCF-10A en placas de 6 pocillos hasta una confluencia del 40-50%. Se incubaron las células con 50 nM de molécula pequeña durante 72 h y luego se desprendieron de la superficie usando Accutase. Se lavaron dos veces cada una con DPBS 1X enfriado cpn hielo y tampón de unión a anexina 1X (Hepes 50 mM (pH 7,4), NaCl 700 mM y $CaCl_2$ 12,5 mM). Se resuspendieron las células en 100 μl de tampón de unión a anexina 1X, y luego se añadieron 5 μl de anexina V-APC (eBioscience). Se incubó la disolución durante 10 minutos a temperatura ambiente, seguido de lavado con tampón de unión a anexina 1X. Luego se tiñeron las células con yoduro de propidio 1 μg/ml en 300 μl de tampón de unión a anexina 1X durante 15 min a temperatura ambiente. Se realizó citometría de flujo con un instrumento BD LSRII (BD Biosciences). Se usaron al menos 10.000 eventos para el análisis.

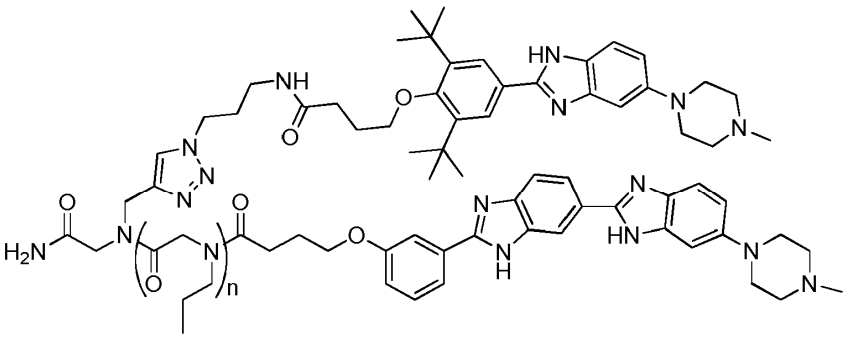
Resultados

La molécula pequeña inicial 1 se optimizó usando el enfoque de ensamblaje modular habilitado mediante Inforna. En resumen, se extrajo la estructura secundaria del precursor de horquilla de microARN-96 usando los métodos descritos en el presente documento para identificar compuestos que podrían unirse a sitios diana en la estructura del ARN (figura 16). Este procedimiento identificó el módulo H (símbolo de rombo en la figura 16) como una molécula de unión al bucle interno GG de nucleótidos 1x1. Al sintetizar los compuestos diméricos espaciados apropiadamente, tanto el bucle interno GG de nucleótidos 1x1 como el bucle interno UU de nucleótidos 1x1 en el sitio Drosha (bucle rodeado con un círculo en la figura 16) podrían unirse simultáneamente por una única molécula pequeña, lo que daría como resultado un aumento de la afinidad de unión y también una potencia celular suficiente para desencadenar la apoptosis por los compuestos.

Los compuestos diméricos se sintetizaron usando un enfoque de ensamblaje modular que se muestra en el siguiente esquema de síntesis. Se empleó un andamiaje de peptoide en el que un derivado de azida del compuesto 1 era el primer módulo, y el segundo era un derivado acilado del compuesto H donde la acilación estaba en un sitio amino en un esqueleto peptoide. Se unió un esqueleto peptoide al H compuesto, que luego se unió al derivado de azida del compuesto 1 a través de una reacción de cicloadición dipolar de Huisgen entre la azida en el compuesto 1 y un alquino en el peptoide H. La separación entre los dos módulos se varió insertando espaciadores de propilamina entre los módulos de unión a ARN.



Los compuestos se denominan BSH-n-H en los cuales el BSH se refiere al compuesto 1 y H se refiere al módulo 2, ambos conjugados al esqueleto. El número entre estos BSH y H es el número de espaciadores de propilamina entre el compuesto 1 y los módulos de unión a ARN. Por tanto, BSH-1-H se refiere a un compuesto con un espaciador de propilamina entre los dos módulos de unión de ARN. La estructura del género BSH-n-H se muestra a continuación.

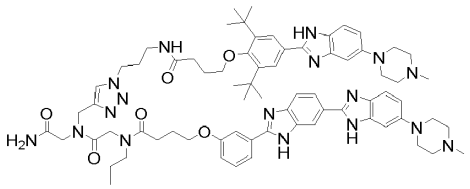
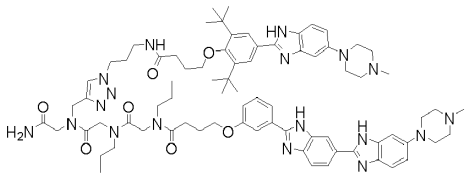
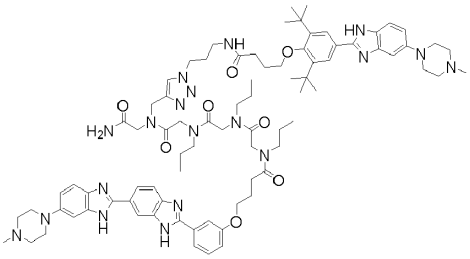
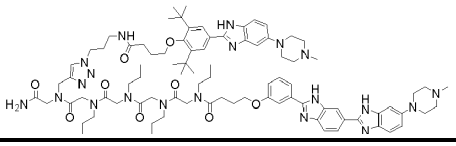


BSH-n-H

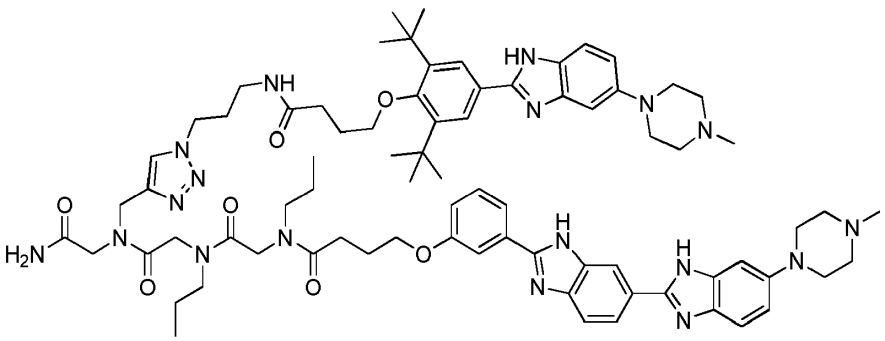
Se sintetizó una biblioteca de estos compuestos diméricos con separación variada de acuerdo con las rutas mostradas anteriormente. Estos compuestos diméricos se caracterizaron usando los métodos descritos

anteriormente. Los resultados se muestran en la tabla 3.

Tabla 3: Caracterización de compuestos diméricos de diseño, BSH-n-H

Nombre de compuesto	Estructura	Masa esperada/observada ($M+H^+$)	t_R (tiempo de retención de HPLC)
BSH-1-H		1292/1292	33 min (metanol al 55%)
BSH-2-H		1392/1392	37 min (metanol al 61%)
BSH-3-H		1490/1490	38 min (metanol al 63%)
BSH-4-H		1590/1590	40 min (metanol al 67%)

Esta biblioteca de dímeros se examinó luego para procesar la inhibición de los precursores de horquilla de microARN-96 mediante el uso de qRT-PCR para cuantificar la abundancia de los microARN maduros, pre y pri. Los datos se muestran en la figura 18. Estos estudios identificaron BSH-2-H como el compuesto más eficaz. El compuesto BSH-2-H inhibió significativamente la producción del microARN-96 maduro a una concentración de 50 nM, mientras que también inhibió la producción del pre-microARN-96 y aumentó la producción del pri-microARN-96. La estructura del dímero óptimo, BSH-2-H, se muestra a continuación.



BSH-2-H

La afinidad de unión del dímero BSH-2-H óptimo se sometió a prueba entonces frente a una variedad de constructos que desplazan el sitio de unión del bucle interno GG de nucleótidos 1x1, el sitio de unión del bucle interno UU de nucleótidos 1x1, o ambos. La unión también se evaluó frente a una horquilla de ADN. Estos estudios demostraron

que BSH-2-H tiene una selectividad significativa para el ARN que contiene ambos sitios y es un agente de unión de afinidad > 30 veces mayor que los compuestos monoméricos a los sustratos de ARN diseñados, 5'CGAUU/3'GGUAUA (figura 17 y tabla 4).

Tabla 4: Constantes de disociación de BSH-FL (derivado con fluoresceína de BSH, compuesto 1), H y BSH-2-H hacia los diferentes ácidos nucleicos con las constantes de unión se informan en nanomolar.

Molécula pequeña	5'UUU/3'AUA	5'CGA/3'GGU	5'CGAUU/ 3'GGUAUA	C1	Horquilla H
BSH-FL	1270 ± 140 nM	9400 ± 1400 nM	3400 ± 100 nM	> 30.000 nM	> 30.000 nM
H	> 20.000 nM	1450 ± 150 nM	2680 ± 267 nM	> 2000 nM	200 nM
BSH-2-H	1167 ± 279 nM	900 ± 200 nM	85 ± 11 nM	> 4000 nM	> 50.000 nM

El compuesto BSH-2-H se sometió a prueba para evaluar si afectaba a la producción de proteína FOXO1 en células cancerosas cultivadas. El ARNm de FOXO1 se selecciona como diana por miR-96, y se reprime por la traducción. Estos estudios demostraron que la aplicación de 50 nM de BSH-2-H provocó un aumento de la cantidad de proteína FOXO1 expresada en las células en 2,5 veces (figura 19).

Si los niveles de FOXO1 se elevan en las células, debe desencadenarse la apoptosis. Por tanto, el compuesto BSH-2-H se sometió a prueba para determinar la inducción de apoptosis en una variedad de líneas celulares de cáncer de mama (células MCF7 y MDA MB 231 triple negativas), así como en tejido de mama sano (MCF-10A). Tal como se muestra en las figuras 20A-20C, cuando se añade BSH-2-H a las líneas celulares de cáncer de mama, se induce una apoptosis significativa. Sin embargo, cuando se añade el compuesto BSH-2-H a las células de mama sanas, no hay efecto sobre la apoptosis.

Bibliografía:

1. Guan, L. & Disney, M.D. Recent advances in developing small molecules targeting RNA. ACS Chem. Biol. 7, 73-86 (2012).

2. Thomas, J.R. & Hergenrother, P.J. Targeting RNA with small molecules. Chem. Rev. 108, 1171-1224 (2008).

3. Yoshizawa, S., Fourmy, D. & Puglisi, J.D. Recognition of the codon-anticodon helix by ribosomal RNA. Science 285, 1722-1725 (1999).

4. Carter, A.P. *et al.* Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407, 340-348. (2000).

5. Mathews, D.H. *et al.* Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. U. S. A. 101, 7287-7292 (2004).

6. Batey, R.T., Rambo, R.P. & Doudna, J.A. Tertiary Motifs in RNA Structure and Folding. Angew. Chem. Int. Ed. Engl. 38, 2326-2343 (1999).

7. Spahn, C.M. *et al.* Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. Science 291, 1959-1962 (2001).

8. Childs-Disney, J.L., Wu, M., Pushechnikov, A., Aminova, O. & Disney, M.D. A small molecule microarray platform to select RNA internal loop-ligand interactions. ACS Chem. Biol. 2, 745-754 (2007).

9. Disney, M.D. *et al.* Two-dimensional combinatorial screening identifies specific aminoglycoside-RNA internal loop partners. J. Am. Chem. Soc. 130, 11185-11194 (2008).

10. Velagapudi, S.P., Seedhouse, S.J. & Disney, M.D. Structure-activity relationships through sequencing (StARTS) defines optimal and suboptimal RNA motif targets for small molecules. Angew. Chem. Int. Ed. Engl. 49, 3816-3818 (2010).

11. Velagapudi, S.P., Seedhouse, S.J., French, J. & Disney, M.D. Defining the RNA internal loops preferred by benzimidazole derivatives via 2D combinatorial screening and computational analysis. J. Am. Chem. Soc. 133, 10111-10118 (2011).

12. Jiang, Q. *et al.* miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 37, D98-104 (2009).

13. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154-158 (2008).

14. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233 (2009).

15. Ambros, V. *et al.* A uniform system for microRNA annotation. RNA 9, 277-279 (2003).

16. Wu, M. & Turner, D.H. Solution structure of (rGCGGACGC)2 by two-dimensional NMR and the iterative relaxation matrix approach. Biochemistry 35, 9677-9689 (1996).

17. SantaLucia, J., Jr. & Turner, D.H. Structure of (rGGCGAGCC)2 in solution from NMR and restrained molecular dynamics. Biochemistry 32, 12612-12623 (1993).

18. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297 (2004).

19. Krutzfeldt, J. *et al.* Silencing of microRNAs in vivo with 'antagomirs'. Nature 438, 685-689 (2005).

20. Ebert, M.S. & Sharp, P.A. MicroRNA sponges: progress and possibilities. RNA 16, 2043-2050 (2010).

21. Obad, S. *et al.* Silencing of microRNA families by seed-targeting tiny LNAs. Nat. Genet. 43, 371-378 (2011).

22. Guttilla, I.K. & White, B.A. Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cáncer cells. J. Biol. Chem. 284, 23204-23216 (2009).

23. Redova, M. *et al.* MiR-210 expression in tumor tissue and in vitro effects of its silencing in renal cell carcinoma. Tumour Biol. 34, 481-491 (2013).

24. Childs-Disney, J.L., Hoskins, J., Rzuczek, S.G., Thornton, C.A. & Disney, M.D. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potently bioactive. ACS Chem. Biol. 7, 856-862 (2012).

25. Pilch, D.S. *et al.* Binding of a hairpin polyamide in the minor groove of DNA: sequence-specific enthalpic discrimination. Proc. Natl. Acad. Sci. U. S. A. 93, 8306-8311 (1996).

26. Pushechnikov, A. *et al.* Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. J. Am. Chem. Soc. 131, 9767-9779 (2009).

27. Lee, M.M., Pushechnikov, A. & Disney, M.D. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. ACS Chem. Biol. 4, 345-355 (2009).

28. Parkesh, R. *et al.* Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif-ligand database & chemical similarity searching. J. Am. Chem. Soc. 134, 4731-4742 (2012).

29. Kumar, A. *et al.* Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. ACS Chem. Biol. 7, 496-505 (2012).

30. Pinto, I.G., Guilbert, C., Ulyanov, N.B., Stearns, J. & James, T.L. Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. J. Med. Chem. 51, 7205-7215 (2008).

31. Disney, M.D. *et al.* A small molecule that targets r(CGG)exp and improves defects in fragile X-associated tremor ataxia syndrome ACS Chem. Biol. 7, 1711-1718 (2012).

32. Luzhkov, V.B. *et al.* Virtual screening and bioassay study of novel inhibitors for dengue virus mRNA cap (nucleoside-2'O)-methyltransferase. Bioorg. Med. Chem. 15, 7795-7802 (2007).

33. Xu, S., Witmer, P.D., Lumayag, S., Kovacs, B. & Valle, D. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. J. Biol. Chem. 282, 25053-25066 (2007).

34. Stenvang, J., Petri, A., Lindow, M., Obad, S. & Kauppinen, S. Inhibition of microRNA function by antimiR oligonucleotides. Silence 3, 1 (2012).

35. Xie, L. *et al.* FOXO1 is a tumor suppressor in classical Hodgkin lymphoma. Blood 119, 3503-3511 (2012).

36. Dansen, T.B. & Burgering, B.M. Unravelling the tumor-suppressive functions of FOXO proteins. Trends Cell Biol.

18, 421-429 (2008).

37. Huang, H. & Tindall, D.J. FOXO factors: a matter of life and death. Future Oncol. 2, 83-89. (2006).

38. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15-20 (2005).

39. Wolkenberg, S.E. & Boger, D.L. Mechanisms of in situ activation for DNA-targeting antitumor agents. Chem. Rev. 102, 2477-2495 (2002).

40. Tran, T. & Disney, M.D. Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. Nat. Commun. 3, 1125 (2012).

41. Hawkins, P.C., Skillman, A.G. & Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. J. Med. Chem. 50, 74-82 (2007).

42. Stelzer, A.C. *et al.* Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. Nat. Chem. Biol. 7, 553-559 (2011).

43. Kramer, R. & Cohen, D. Functional genomics to new drug targets. Nat. Rev. Drug Discov. 3, 965-972 (2004).

44. Bevilacqua, J.M. & Bevilacqua, P.C. Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. Biochemistry 37, 15877-15884 (1998).

45. McKenna, S.A. *et al.* Purification and characterization of transcribed RNAs using gel filtration chromatography. Nat. Protoc. 2, 3270-3277 (2007).

46. Peyret, N., Seneviratne, P.A., Allawi, H.T. & SantaLucia, J., Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. Biochemistry 38, 3468-3477 (1999).

47. SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. U. S. A. 95, 1460-1465 (1998).

48. Puglisi, J.D. & Tinoco, I., Jr. Absorbance melting curves of RNA. Methods Enzymol. 180, 304-325 (1989).

49. Wang, Y. & Rando, R.R. Specific binding of aminoglycoside antibiotics to RNA. Chem. Biol. 2, 281-290 (1995).

50. Disney, M.D., Gryaznov, S.M. & Turner, D.H. Contributions of individual nucleotides to tertiary binding of substrate by a Pneumocystis carinii group I intron. Biochemistry 39, 14269-14278 (2000).

51. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. Curr. Biol. 14, 2162-2167 (2004).

52. Jang, H.; Fafarman, A.; Holub, J. M.; Kirshenbaum, K. Org Lett, 7: 1951 (2005).

53. Disney, M. D.; Labuda, L. P.; Paul, D. J.; Poplawski, S. G.; Pushechnikov, A.; Tran, T.; Velagapudi, S. P.; Wu, M.; Childs-Disney, J. L. J Am Chem Soc 130: 11185 (2008).

Lista de secuencias

<110> Disney, Matthew D.
Velagapudi, Sai
The Scripps Research Institute

<120> DISEÑO DE TRANSCRIPTOMA COMPLETO DE MOLÉCULAS PEQUEÑAS BIOACTIVAS, SELECTIVAS QUE SELECCIONAN COMO DIANA ARN

<130> Documento 1361.196WO1

<150> Documento US 61/864.081
<151> 09-08-2013

<160> 232

<170> FastSEQ para Windows Versión 4.0

<210> 1
<211> 40
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 1
gggagagggu uuaauuacga aaguaauugg auccgcaagg 40

<210> 2
<211> 46
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<221> misc_feature
<222> (1)...(46)
<223> n = A, U, C o G

<400> 2
gggagagggu uuaaunnnua cgaaaguann nauuggaucc gcaagg 46

<210> 3
<211> 12
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 3
ggguuuaauu ac 12

<210> 4
<211> 12
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 4
guaauuggau cc 12

<210> 5
<211> 10
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 5
cgcgaaagcg 10

<210> 6
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Una secuencia de deoxioligonucleótidos sintética

<400> 6
atatatatat atatatatat at 22

<210> 7
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Una secuencia de deoxioligonucleótidos sintética

<400> 7
gcgcgcgcgc gcgcgcgcgc gc 22

<210> 8
<211> 44
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 8
gggagagggu uuaauuuuac gaaaguaaua uuggauccgc aagg 44

<210> 9
<211> 54
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 9
gggagagggu uuaauccgau uuuacgaaag uaauauggga uuggauccgc aagg 54

<210> 10
<211> 50
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 10
gggagagggu uuaaugguag uacgaaagua cucucauugg auccgcaagg 50

<210> 11
<211> 78
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 11
uggccgauuu uggcacuagc acauuuuugc uugugucucu ccgcucugag caaucaugug 60
cagugccaau augggaaa 78

<210> 12
<211> 110
<212> ARN
<213> Secuencia artificial

<220>

<223> Una secuencia de ARN sintética

<400> 12
gagcugcuug ccucccccccg uuuuuggcaa ugguagaacu cacacuggug agguaacagg 60
auccgguggu ucuagacuug ccaacuaugg ggcgaggacu cagccggcac 110

<210> 13
<211> 110
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 13
ccgcagagug ugacuccugu ucuguguaug gcacugguag aauucacugu gaacagucuc 60
agucagugaa uuaccgaagg gccauaaaca gagcagagac agauccacga 110

<210> 14
<211> 109
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 14
acccggcagu gccuccaggc gcagggcagc cccugcccac cgcacacugc gcugcccagg 60
cccacugugc gugugacagc ggcugaucug ugccugggca gcgcagccc 109

<210> 15
<211> 23
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 15
uuuggcacua gcacauuuuu gcu 23

<210> 16
<211> 24
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 16
uuuggcaaug guagaacuca cacu 24

<210> 17
<211> 22
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintética

<400> 17
uauggcacug guagaauuca cu 22

<210> 18
<211> 42
<212> ADN

<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 18
ggccggatcc taatacgact cactataggg agagggttta at 42

<210> 19
<211> 15
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 19
ccttgcggat ccaat 15

<210> 20
<211> 44
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 20
ggccggatcc taatacgact cactataggg tggccgattt tggc 44

<210> 21
<211> 15
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 21
tttcccatat tggca 15

<210> 22
<211> 81
<212> ADN
<213> Secuencia artificial

<220>
<223> Una secuencia de oligonucleótidos sintética

<400> 22

gggtggccga ttttggcact agcacatttt tgcttgtgtc tctccgctct gagcaatcat 60
gtgcagtgcc aatatgggaa a 81

<210> 23
<211> 47
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 23
ggccgaattc taatacgact cactataggc accagtgcca tctgctt 47

<210> 24

<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 24
taacagtctc cagtcacggc c 21

<210> 25
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 25
tgtctgcccg catgcctgcc tct 23

<210> 26
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 26
tgtaacagca actccatgtg ga 22

<210> 27
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 27
taggtagttt cctgttgttg gg 22

<210> 28
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 28
taggtagttt catgttgttg gg 22

<210> 29
<211> 21
<212> PRT
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 29

Thr Ala Ala Ala Gly Thr Gly Cys Thr Gly Ala Cys Ala Gly Thr Gly
1 5 10 15
Cys Ala Gly Ala Thr
20

<210> 30
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 30
ctgtgcgtgt gacagcggct ga 22

<210> 31
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 31
acagcaggca cagacaggca gt 22

<210> 32
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 32
caaagtgctt acagtgcagg tag 23

<210> 33
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 33
attgacactt ctgtgagtag a 21

<210> 34
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 34
tagcaccatt tgaaatcggt ta 22

<210> 35
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 35
aatataacac agatggcctg t 21

<210> 36
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 36
ggtccagagg ggagataggt tc 22

<210> 37
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 37
tgaggtagta gtttgtgctg tt 22

<210> 38
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 38
catcccttgc atggtggagg g 21

<210> 39
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 39
tattgcactt gtcccggcct gt 22

<210> 40
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 40
tcgtaccgtg agtaataatg cg 22

<210> 41
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 41
aacattcaac ctgtcggtga gt 22

<210> 42
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 42
tgtaaacatc ctacactcag ct 22

<210> 43
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 43
tatggcactg gtagaattca ct 22

<210> 44
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 44
tgtaaacatc ctacactctc agc 23

<210> 45
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 45
atcaacagac attaattggg cgc 23

<210> 46
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 46
aactggccct caaagtcccg ct 22

<210> 47
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 47
tttaggataa gcttgacttt tg 22

<210> 48
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 48
tagcagcaca taatggtttg tg 22

<210> 49
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 49
cagtgcaatg atattgtcaa agc 23

<210> 50
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 50
taaggtgcat ctagtgcagt tag 23

<210> 51
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 51
tagcagcaca tcatggttta ca 22

<210> 52
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 52
taccctgtag atccgaattt gtg 23

<210> 53
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 53
ttcaagtaat tcaggatagg t 21

<210> 54
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 54
tgaggtagta ggttgtatag tt 22

<210> 55
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 55
taaggcaccc ttctgagtag a 21

<210> 56
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 56
tgaggtagta gtttgtacag tt 22

<210> 57
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 57
cattgcactt gtctcggtct ga 22

<210> 58
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 58
atcatgatgg gctcctcggt gt 22

<210> 59
<211> 20
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 59
tacagtatag atgatgtact 20

<210> 60
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 60
tattgcacat tactaagttg ca 22

<210> 61
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 61
taatactgcc gggtaatgat gga 23

<210> 62
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 62
tagcaccatc tgaaatcggt ta 22

<210> 63
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 63
cagcagcaca ctgtggtttg t 21

<210> 64
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 64
caagaacctc agttgctttt gt 22

<210> 65
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 65
tggctcagtt cagcaggaac ag 22

<210> 66
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 66
tcgtgtcttg tgttgcagcc gg 22

<210> 67
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 67
taaggtgcat ctagtgcaga tag 23

<210> 68
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 68
agctacatct ggctactggg t 21

<210> 69
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 69
tctttggtta tctagctgta tga 23

<210> 70
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 70
tccctgagac cctaacttgt ga 22

<210> 71
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 71
taatgcccct aaaaatcctt at 22

<210> 72
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 72
taccctgtag aaccgaattt gtg 23

<210> 73
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 73
tgaggtagta ggttgtatgg tt 22

<210> 74
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 74
tggagagaaa ggcagttcct ga 22

<210> 75
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 75
aagtgtgcag ggcactggt 19

<210> 76
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 76
tggaatgtaa ggaagtgtgt gg 22

<210> 77
<211> 20
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 77
taaggcacgc ggtgaatgcc 20

<210> 78
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 78
taacagtcta cagccatggt cg 22

<210> 79
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 79
ctctagaggg aagcgctttc tg 22

<210> 80
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 80
tcagtgcatc acagaacttt gt 22

<210> 81
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 81
tccctgagac cctttaacct gtga 24

<210> 82
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 82
caaagtgctc atagtgcagg tag 23

<210> 83
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 83
aagtgctgtc atagctgagg tc 22

<210> 84
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 84
caaagaattc tccttttggg ct 22

<210> 85
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 85
tctggctccg tgtcttcact ccc 23

<210> 86
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 86
acagtctgct gaggttggag c 21

<210> 87
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 87
taatactgtc tggtaaaacc gt 22

<210> 88
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 88
atcacattgc cagggatttc c 21

<210> 89
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 89
gtcatacacg gctctcctct ct 22

<210> 90
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 90
tgtgcaaatc catgcaaaac tga 23

<210> 91
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 91
taacactgtc tggtaaagat gg 22

<210> 92
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 92
agcagcattg tacagggcta tga 23

<210> 93
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 93
tttggcacta gcacattttt gct 23

<210> 94
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 94
atcacattgc cagggattac c 21

<210> 95
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 95
tcagtgcatg acagaacttg g 21

<210> 96
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 96
tgagctaaat gtgtgctggg a 21

<210> 97
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 97
gtgaaatgtt taggaccact ag 22

<210> 98
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 98
tcggatccgt ctgagcttgg ct 22

<210> 99
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 99
tgtgactggt tgaccagagg gg 22

<210> 100
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 100
tgaggtagta gattgtatag tt 22

<210> 101
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 101
tggagtgtga caatggtgtt tg 22

<210> 102
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 102
tgtagtgttt cctactttat gga 23

<210> 103
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 103
tggcagtgtc ttagctggtt gt 22

<210> 104
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 104
tggctcagtt cagcaggaac ag 22

<210> 105
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 105
tcgtgtcttg tgttgcagcc gg 22

<210> 106
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 106
taaggtgcat ctagtgcaga tag 23

<210> 107
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 107
agctacatct ggctactggg t 21

<210> 108
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 108
tctttggtta tctagctgta tga 23

<210> 109
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 109
tccctgagac cctaacttgt ga 22

<210> 110
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 110
taatgcccct aaaaatcctt at 22

<210> 111
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 111
taccctgtag aaccgaattt gtg 23

<210> 112
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 112
tgaggtagta ggttgtatgg tt 22

<210> 113
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 113
tggagagaaa ggcagttcct ga 22

<210> 114
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 114
aagtgtgcag ggcactggt 19

<210> 115
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 115
tggaatgtaa ggaagtgtgt gg 22

<210> 116
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 116
taaggcacgc ggtgaatgcc 20

<210> 117
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 117
taacagtcta cagccatggt cg 22

<210> 118
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 118
ctctagaggg aagcgctttc tg 22

<210> 119
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 119
tcagtgcatc acagaacttt gt 22

<210> 120
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 120
tccctgagac cctttaacct gtga 24

<210> 121
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 121
caaagtgctc atagtgcagg tag 23

<210> 122
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 122
aagtgctgtc atagctgagg tc 22

<210> 123
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 123
caaagaattc tccttttggg ct 22

<210> 124
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 124
tctggctccg tgtcttcact ccc 23

<210> 125
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 125
acagtctgct gaggttggag c 21

<210> 126
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 126
taatactgtc tggtaaaacc gt 22

<210> 127
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 127
atcacattgc cagggatttc c 21

<210> 128
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 128
gtcatacacg gctctcctct ct 22

<210> 129
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 129
tgtgcaaatc catgcaaaac tga 23

<210> 130
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 130
taacactgtc tggtaaagat gg 22

<210> 131
<211> 23
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 131
agcagcattg tacagggcta tga 23

<210> 132
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 132
tttggcacta gcacattttt gct 23

<210> 133
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 133
atcacattgc cagggattac c 21

<210> 134
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 134
tcagtgcatg acagaacttg g 21

<210> 135
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 135
tgagctaaat gtgtgctggg a 21

<210> 136
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 136
gtgaaatgtt taggaccact ag 22

<210> 137
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 137
tcggatccgt ctgagcttgg ct 22

<210> 138
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 138
tgtgactggt tgaccagagg gg 22

<210> 139
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 139
tgaggtagta gattgtatag tt 22

<210> 140
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 140
tggagtgtga caatggtgtt tg 22

<210> 141
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 141
tgtagtgttt cctactttat gga 23

<210> 142
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 142
tggcagtgtc ttagctggtt gt 22

<210> 143
<211> 23
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 143
tgtgcaaatc tatgcaaaac tga 23

<210> 144
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 144
cagtgcaatg atgaaagggc at 22

<210> 145
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 145
tcacagtgaa ccggtctctt t 21

<210> 146
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 146
tttggcaatg gtagaactca cact 24

<210> 147
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 147
tatggctttt cattcctatg tga 23

<210> 148
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 148
aggcagtgta gttagctgat tgc 23

<210> 149
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 149
tggaatgtaa agaagtatgt at 22

<210> 150
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 150
tgagatgaag cactgtagct c 21

<210> 151
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 151
tccagcatca gtgattttgt tg 22

<210> 152
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 152
agctggtgtt gtgaatcagg ccg 23

<210> 153
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 153
ttcacagtgg ctaagttccg c 21

<210> 154
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 154
ttcacagtgg ctaagttctg c 21

<210> 155
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 155
aaaagctggg ttgagagggc ga 22

<210> 156
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 156
tagcagcacg taaatattgg cg 22

<210> 157
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 157
aacccgtaga tccgaacttg tg 22

<210> 158
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 158
taccacaggg tagaaccacg g 21

<210> 159
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 159
gagcttattc ataaaagtgc ag 22

<210> 160
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 160
tatggctttt tattcctatg tga 23

<210> 161
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 161
actggacttg gagtcagaag g 21

<210> 162
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 162
tctacagtgc acgtgtctcc ag 22

<210> 163
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 163
ctaggtatgg tcccagggat cc 22

<210> 164
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 164
tctcacacag aaatcgcacc cgt 23

<210> 165
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 165
aacccgtaga tccgatcttg tg 22

<210> 166
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 166
atcgtgcatc cctttagagt gt 22

<210> 167
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 167
gtgacatcac atatacggca gc 22

<210> 168
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 168
taaagtgctt atagtgcagg tag 23

<210> 169
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 169
ttgtgcttga tctaaccatg t 21

<210> 170
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 170
tacagtactg tgataactga a 21

<210> 171
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 171
tccttcattc caccggagtc tg 22

<210> 172
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 172
tgaggtagta agttgtattg tt 22

<210> 173
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 173
tgaggtagta ggttgtgtgg tt 22

<210> 174
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 174
aacattcatt gctgtcggtg ggt 23

<210> 175
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 175
tgtaaacatc ctcgactgga ag 22

<210> 176
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 176
tagcttatca gactgatgtt ga 22

<210> 177
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 177
cagtggtttt accctatggt ag 22

<210> 178
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 178
taagtgcttc catgttttgg tga 23

<210> 179
<211> 23
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 179
tagcaccatt tgaaatcagt gtt 23

<210> 180
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 180
taattttatg tataagctag t 21

<210> 181
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 181
agaggtagta ggttgcatag tt 22

<210> 182
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 182
aaagtgcatc tttttagagg at 22

<210> 183
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 183
aacattcaac gctgtcggtg agt 23

<210> 184
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 184
aacattcatt gttgtcggtg ggt 23

<210> 185
<211> 23
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 185
cagtgcaata gtattgtcaa agc 23

<210> 186
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 186
tctcccaacc cttgtaccag tg 22

<210> 187
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 187
agcagaagca gggaggttct ccca 24

<210> 188
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 188
ttattgctta agaatacgcg tag 23

<210> 189
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 189
ctgaagctca gagggctctg at 22

<210> 190
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 190
tttggtcccc ttcaaccagc tgaa 24

<210> 191
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 191
ttgcatagtc acaaaagtga tc 22

<210> 192
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 192
aactggccta caaagtccca gtaa 24

<210> 193
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 193
taatctcagc tggcaactgt gaaa 24

<210> 194
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 194
tactgcatca ggaactgatt ggaaa 25

<210> 195
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 195
caagtcacta gtggttccgt t 21

<210> 196
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 196
cccagtgttc aggctacctg ttc 23

<210> 197
<211> 24
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 197
taacactgtc tggtaacgat gtaa 24

<210> 198
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 198
taaacgtgga tgtacttgct 20

<210> 199
<211> 18
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 199
tgggcccttc ctccagaa 18

<210> 200
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 200
gtgcattgct gttgcattgc 20

<210> 201
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 201
ggagacgcgg ccctgttgga gt 22

<210> 202
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 202
ccgttaccat tactgagtta a 21

<210> 203
<211> 20
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 203
actgccccag gtgctgctgg 20

<210> 204
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 204
gggtactgca gacgtggcaa tcatg 25

<210> 205
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 205
tcactcctct cctcccgtct t 21

<210> 206
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 206
ctgtggactc agttctggaa 20

<210> 207
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 207
aatcactaac tccactgcca tc 22

<210> 208
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 208
caaactgtgg gggcactaa 19

<210> 209
<211> 16
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 209
cccagataat ggcact 16

<210> 210
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 210
tttcaccttt ctgagaagga a 21

<210> 211
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 211
ttcacaagga ggtgtcattt at 22

<210> 212
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 212
gccagcaggc agtgtattgt tagctggc 28

<210> 213
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 213
tcctgtactg agctgccccg ag 22

<210> 214
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 214
aatggatttt tggagcagg 19

<210> 215
<211> 21
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 215
aagtgatcta aaggcctaca t 21

<210> 216
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 216
ccccaggtaa ctcttgagtg t 21

<210> 217
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 217
taatgatatc actgtaaaac c 21

<210> 218
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 218
atgctgactg aacatgaagg tc 22

<210> 219
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 219
acacgcaaat tcgtgaagcg ttc 23

<210> 220
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 220
gaatcgagca ccagttacgc 20

<210> 221
<211> 25
<212> ADN
<213> Secuencia artificial

<220>

<223> Un cebador sintético

<400> 221
attttggcac tagcacattt ttgct 25

<210> 222
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 222
ccatattggc actgcacatg att 23

<210> 223
<211> 17
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 223
agagagcccg caccagt 17

<210> 224
<211> 18
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 224
cttgaggagg agcaggct 18

<210> 225
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 225
aaggtgaagg tcggagtcaa 20

<210> 226
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

<400> 226
aatgaagggg tcattgatgg 20

<210> 227
<211> 46
<212> ARN
<213> Secuencia artificial

<220>

<223> Una secuencia de ARN sintético

<400> 227
gggagagggu uuaauuuuua cgaaaguaau aauuggaucc gcaagg 46

<210> 228
<211> 48
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintético

<400> 228
gggagagggu uuaauccgau acgaaaguau gggauuggau ccgcaagg 48

<210> 229
<211> 54
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintético

<400> 229
gggagagggu uuaauccgau uuuacgaaag uaauauggga uuggauccgc aagg 54

<210> 230
<211> 28
<212> ADN
<213> Secuencia artificial

<220>
<223> Una secuencia de horquilla de ADN sintético

<400> 230
cgcgaattcg cgttttcgcg aattcgcg 28

<210> 231
<211> 81
<212> ARN
<213> Secuencia artificial

<220>
<223> Una secuencia de ARN sintético

<400> 231

ggguggccga uuuuggcacu agcacauuuu ugcuuguguc ucuccgcucu gagcaaucau 60
gugcagugcc aauaugggaa a 81

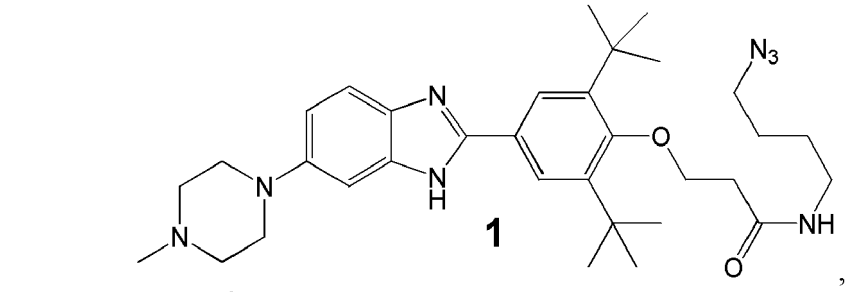
<210> 232
<211> 16
<212> ADN
<213> Secuencia artificial

<220>
<223> Un cebador sintético

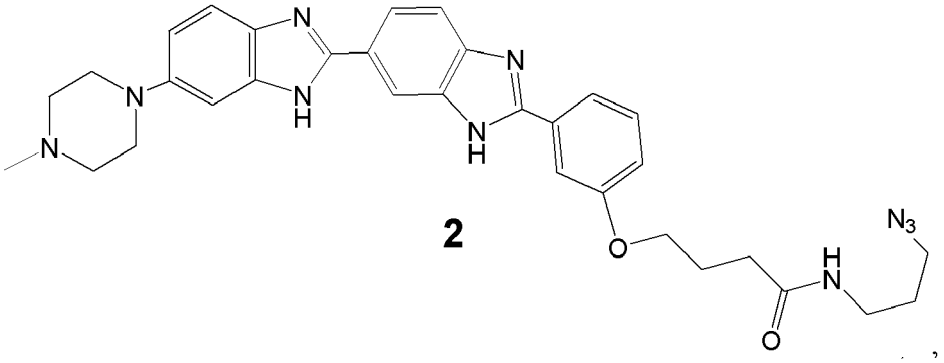
<400> 232
cgcagctgcg ggtcct 16

REIVINDICACIONES

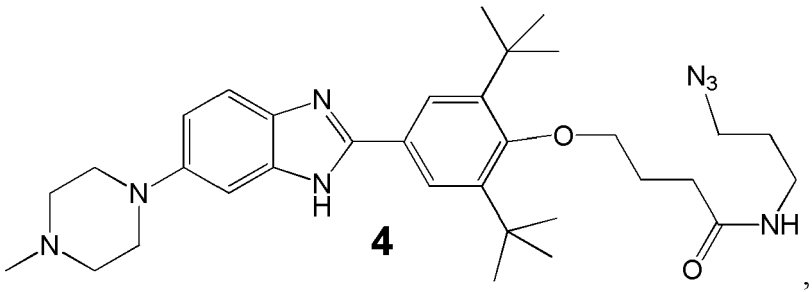
1. Compuesto para su uso en el tratamiento de cáncer en un sujeto, en el que dicho compuesto es cualquiera de



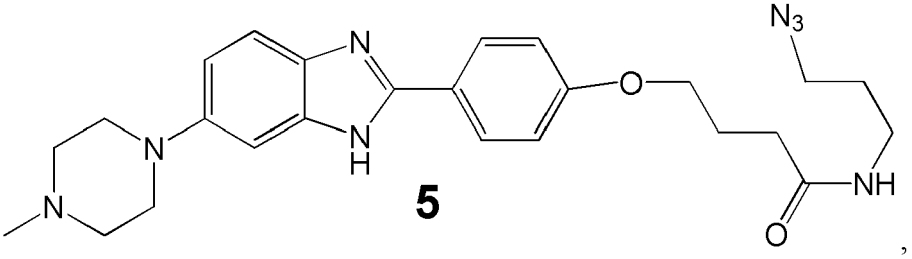
1,



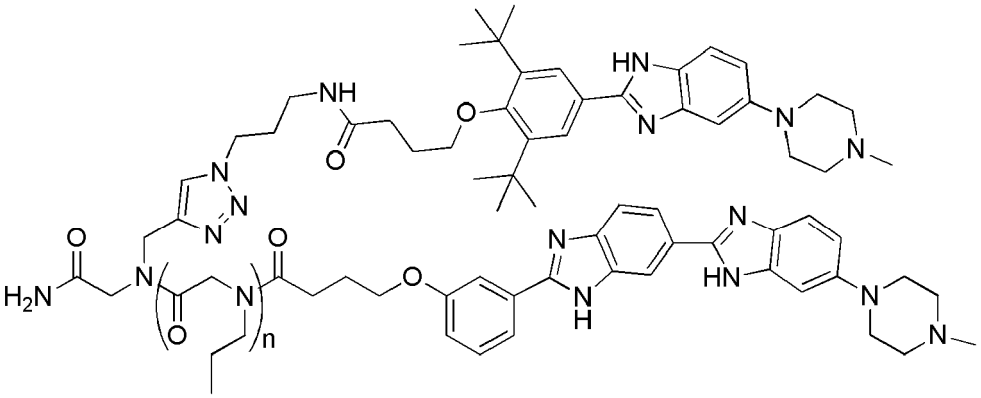
2,



4,



5,

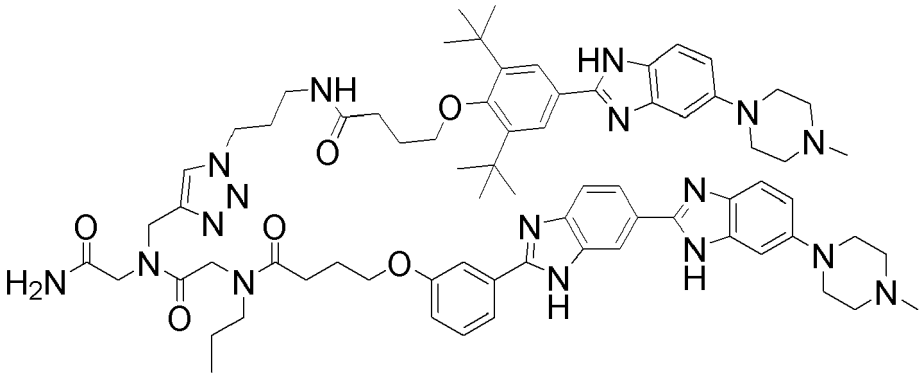


BSH-n-H,

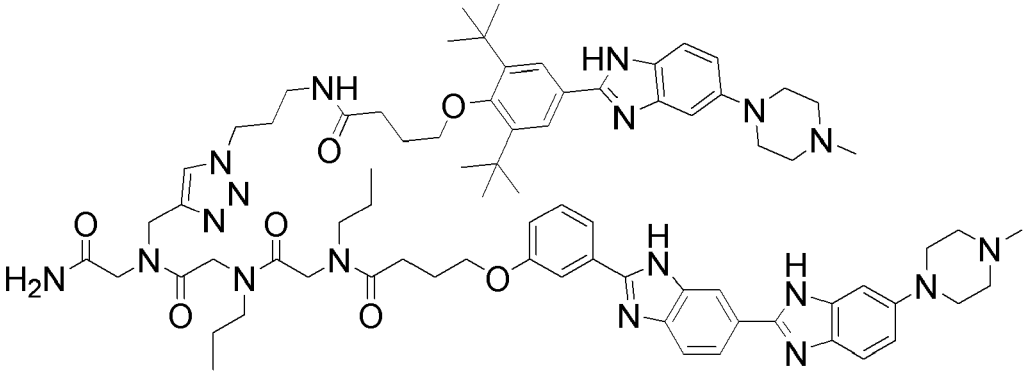
o cualquier combinación de los mismos;

en las que n es un número entero desde 1 hasta 10.

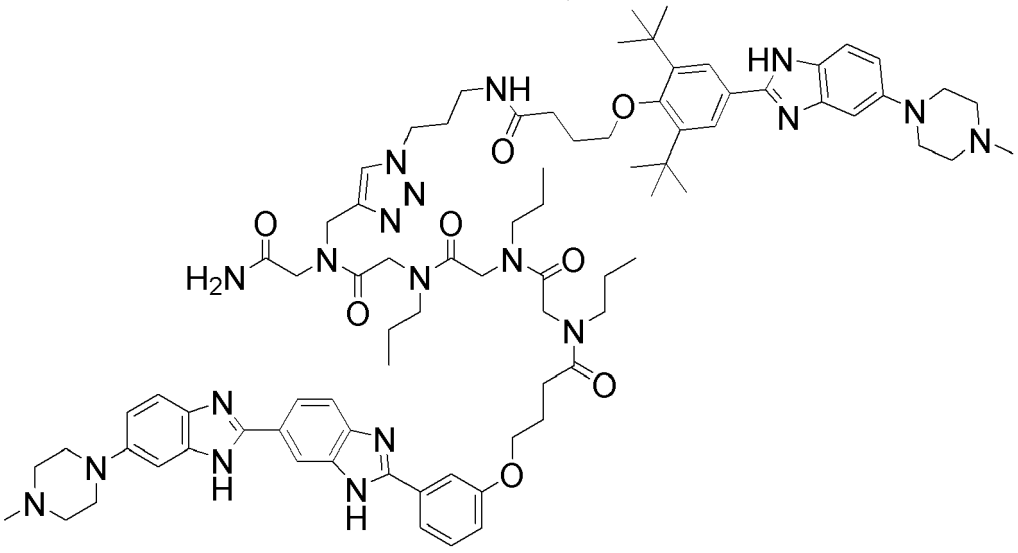
2. Compuesto para su uso según la reivindicación 1, que es cualquiera de



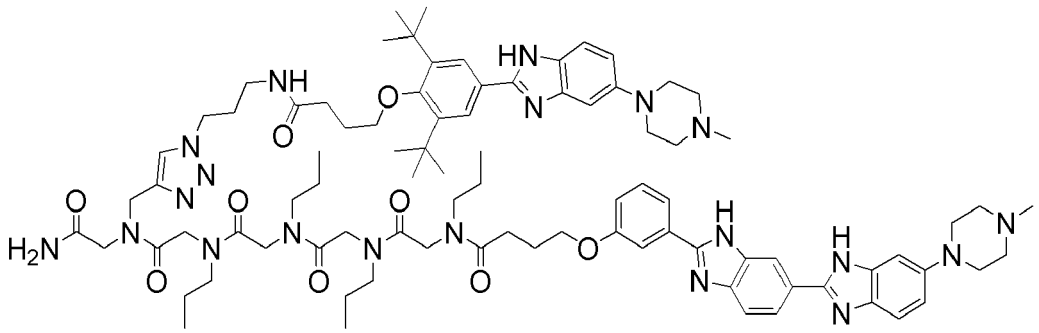
BSH-1-H,



BSH-2-H,



BSH-3-H,



BSH-4-H,

o cualquier combinación de los mismos.

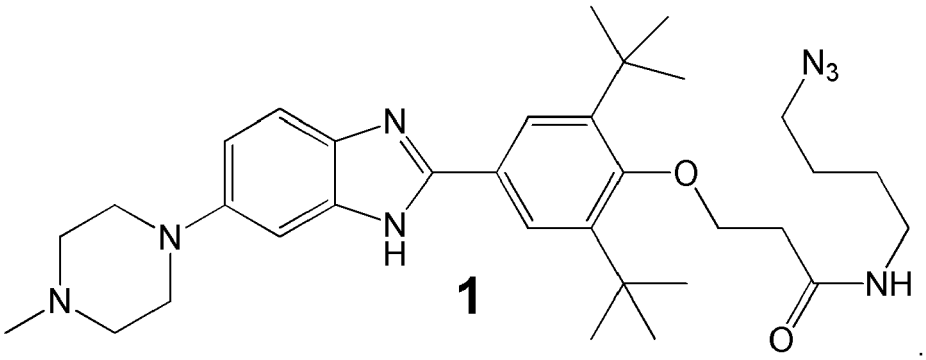
3. Compuesto para su uso según la reivindicación 1 ó 2, para modular la función o actividad de microARN mediante contacto con el microARN.

4. Compuesto para su uso según la reivindicación 3, en el que el microARN es un pre-microARN-96.

5. Compuesto para su uso según la reivindicación 3, en el que el microARN es un pre-microARN-210.

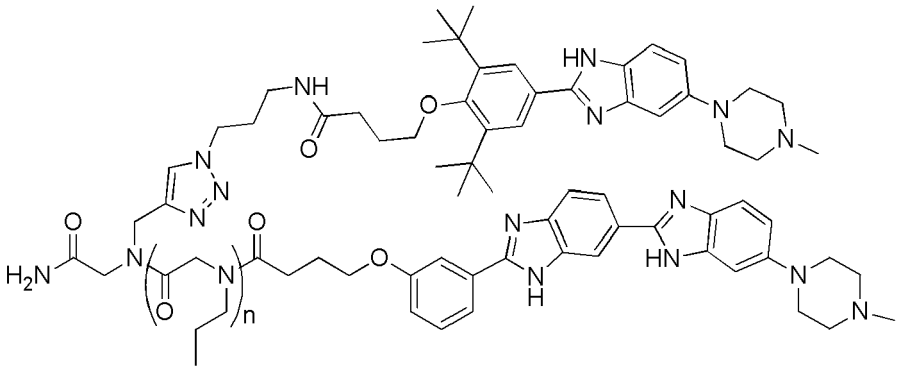
6. Compuesto para su uso según la reivindicación 3, en el que el microARN es un pre-microARN-182.

7. Compuesto de fórmula:



1

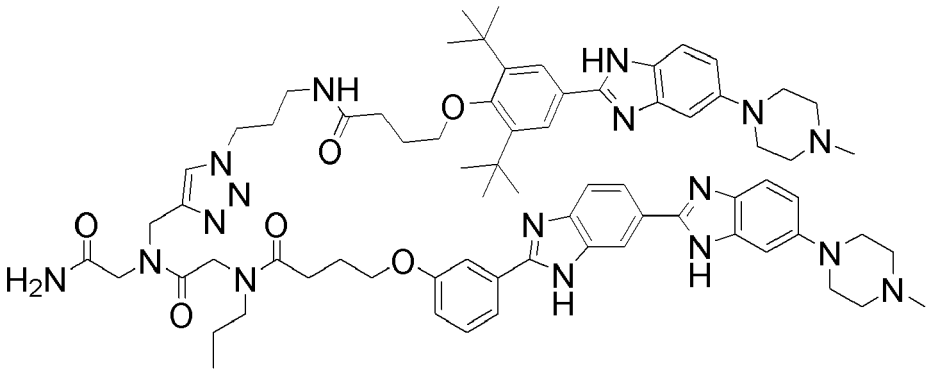
8. Compuesto de fórmula:



BSH-n-H,

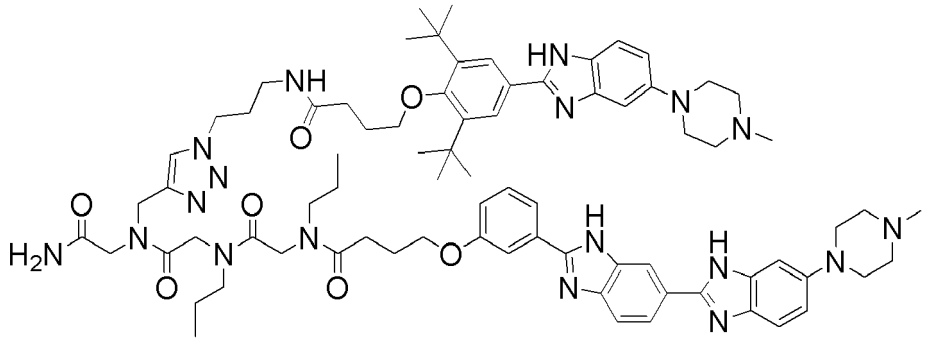
en la que n es un número entero desde 1 hasta 10, o cualquier combinación de los mismos.

9. Compuesto de fórmula:



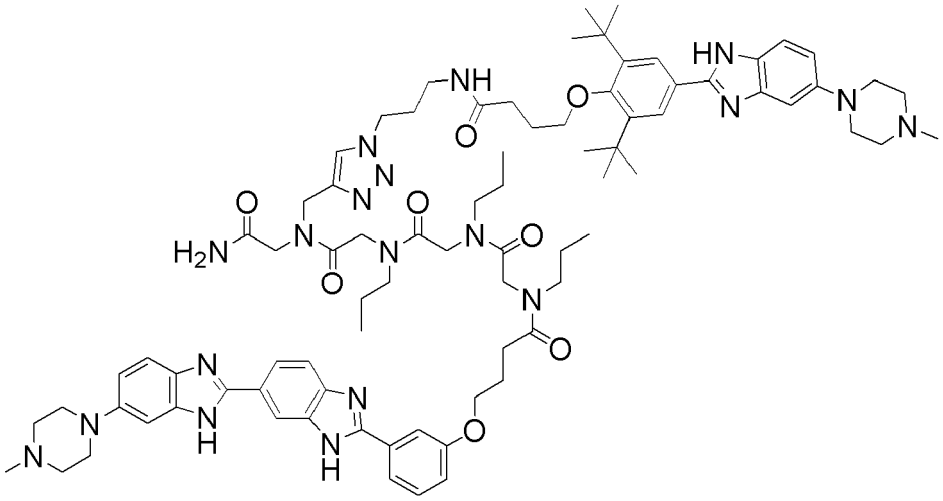
BSH-1-H.

10. Compuesto de fórmula:



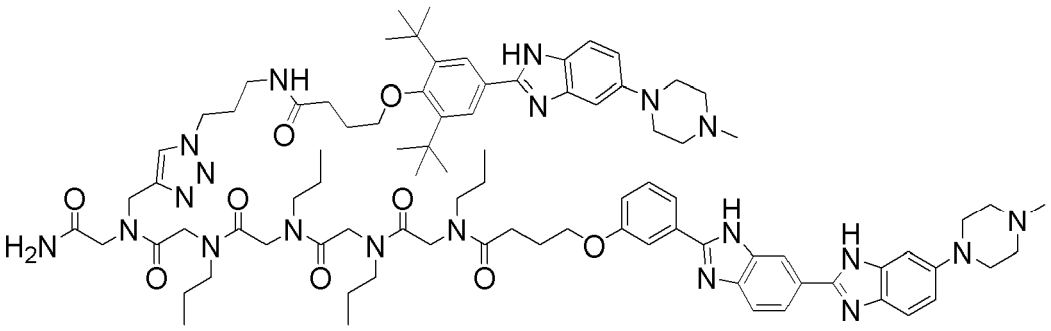
BSH-2-H.

11. Compuesto de fórmula:



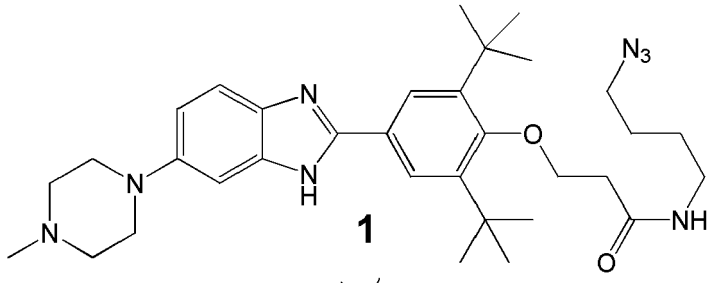
BSH-3-H.

12. Compuesto de fórmula:

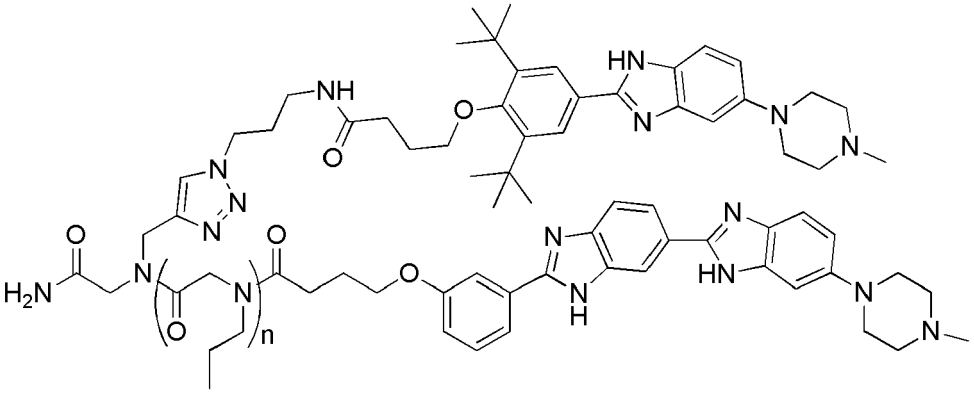


BSH-4-H.

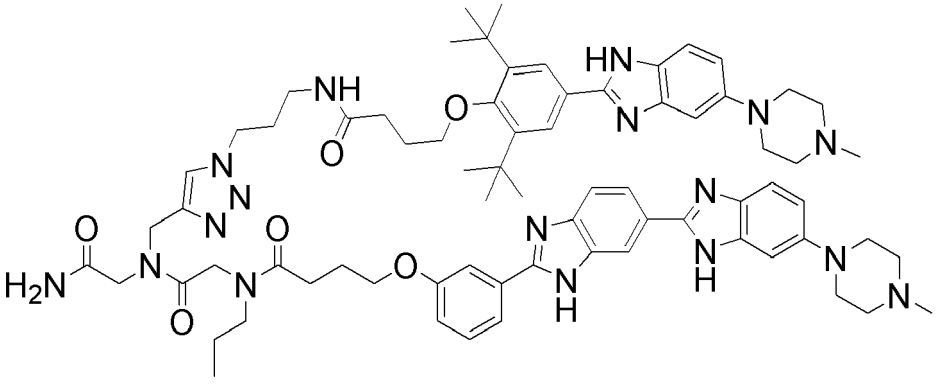
13. Composición que comprende un portador y uno o más de los siguientes compuestos o cualquier combinación de los mismos, siendo n un número entero desde 1 hasta 10:



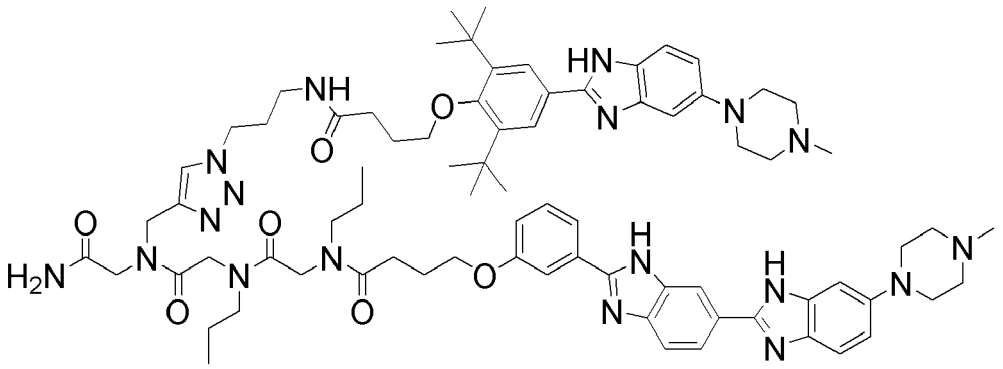
1



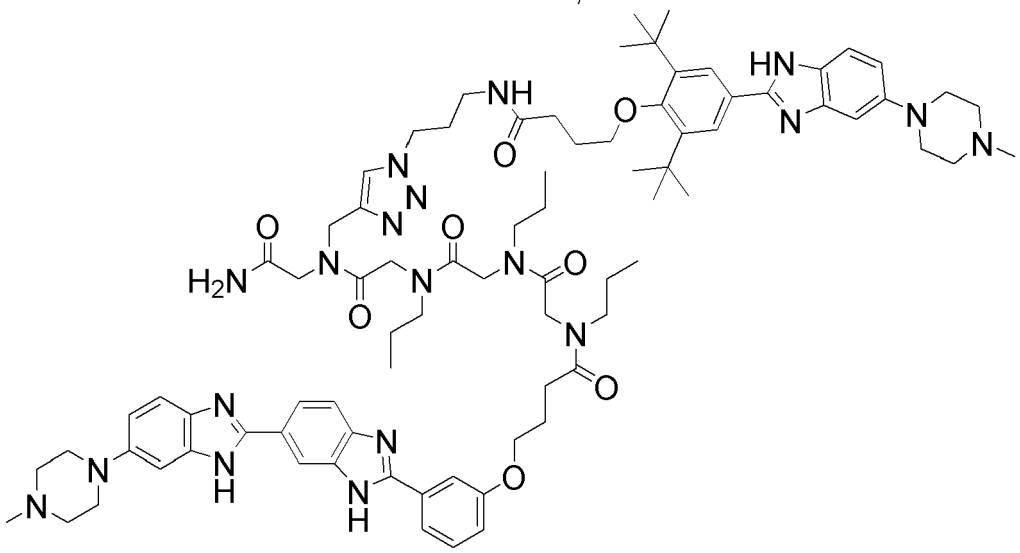
BSH-n-H,



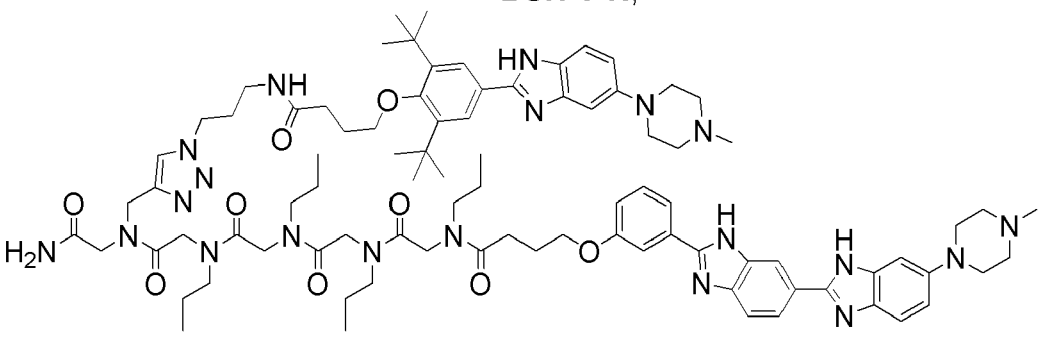
BSH-1-H,



BSH-2-H,



BSH-3-H,



BSH-4-H,

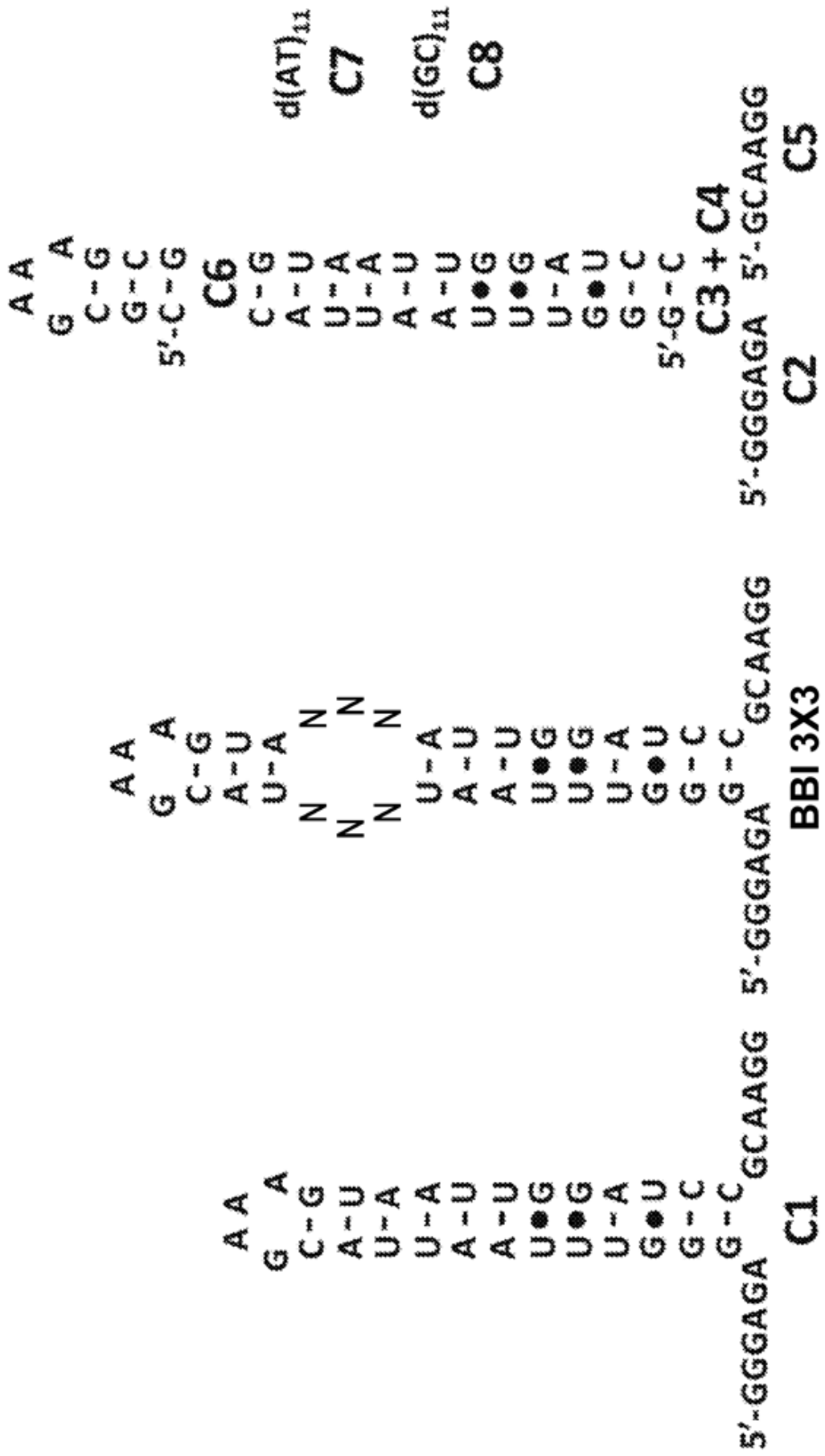
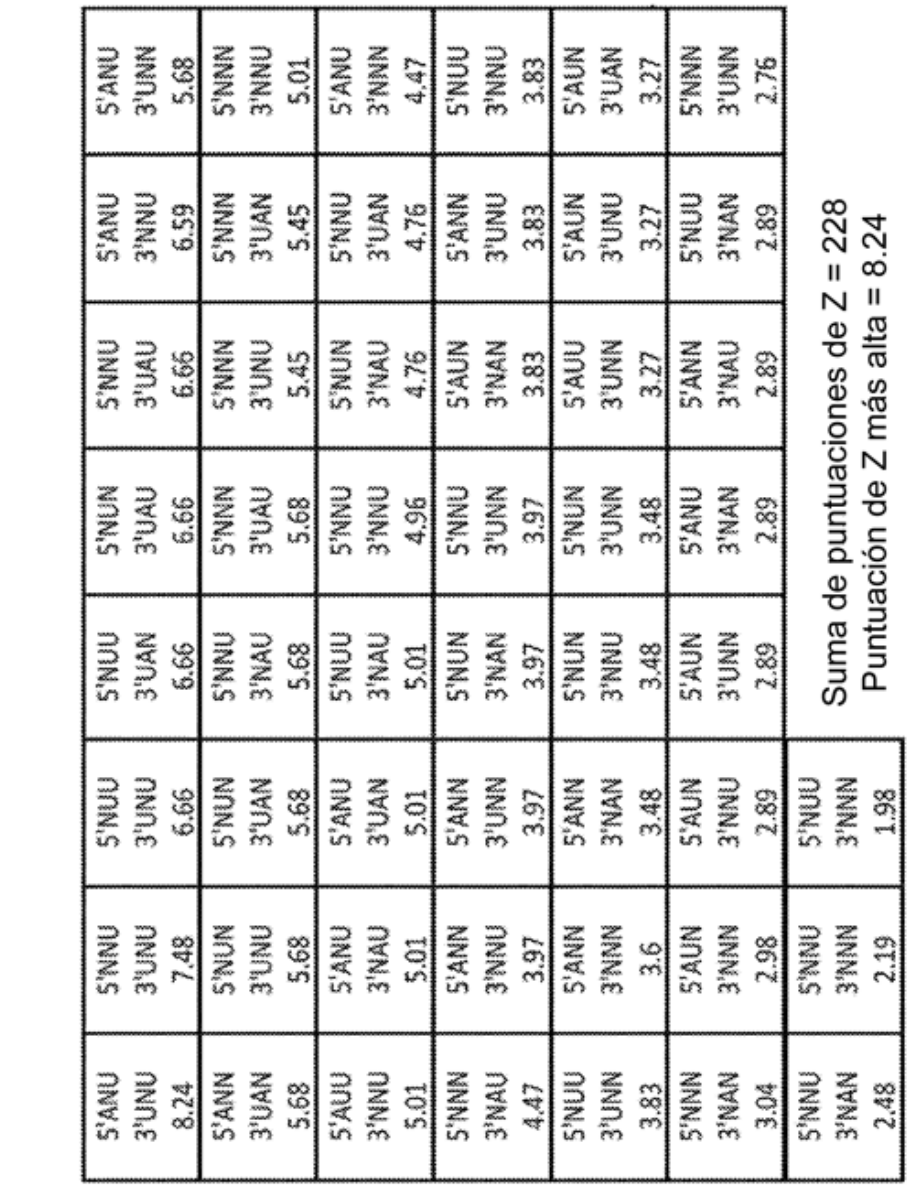

C1
BBI 3X3
C2
C3 + C4
C5
C6
C7
C8
d(AT)11
d(GC)11
5'-GGGAGA
GCAAGG
5'-GCAAGG

Fig. 1



5'ANU 3'UNU 8.24	5'NNU 3'UNU 7.48	5'NUU 3'UNU 6.66	5'NUU 3'UAN 6.66	5'NUN 3'UAU 6.66	5'NNU 3'UAU 6.66	5'ANU 3'NNU 6.59	5'ANU 3'UNN 5.68
5'ANN 3'UAN 5.68	5'NUN 3'UNU 5.68	5'NUN 3'UAN 5.68	5'NNU 3'NAU 5.68	5'NNN 3'UAU 5.68	5'NNN 3'UNU 5.45	5'NNN 3'UAN 5.45	5'NNN 3'NNU 5.01
5'AUU 3'NNU 5.01	5'ANU 3'NAU 5.01	5'ANU 3'UAN 5.01	5'NUU 3'NAU 5.01	5'NNU 3'NNU 4.96	5'NUN 3'NAU 4.76	5'NNU 3'UAN 4.76	5'ANU 3'NNN 4.47
5'NNN 3'NAU 4.47	5'ANN 3'NNU 3.97	5'ANN 3'UNN 3.97	5'NUN 3'NAN 3.97	5'NNU 3'UNN 3.97	5'AUN 3'NAN 3.83	5'ANN 3'UNU 3.83	5'NUU 3'NNU 3.83
5'NUU 3'UNN 3.83	5'ANN 3'NNN 3.6	5'ANN 3'NAN 3.48	5'NUN 3'NNU 3.48	5'NUN 3'UNN 3.48	5'AUU 3'UNN 3.27	5'AUN 3'UNU 3.27	5'AUN 3'UAN 3.27
5'NNN 3'NAN 3.04	5'AUN 3'NNN 2.98	5'AUN 3'NNU 2.89	5'AUN 3'UNN 2.89	5'ANU 3'NAN 2.89	5'ANN 3'NAU 2.89	5'NUU 3'NAN 2.89	5'NNN 3'UNN 2.76
5'NNU 3'NAN 2.48	5'NNU 3'NNN 2.19	5'NUU 3'NNN 1.98					

Fig.2

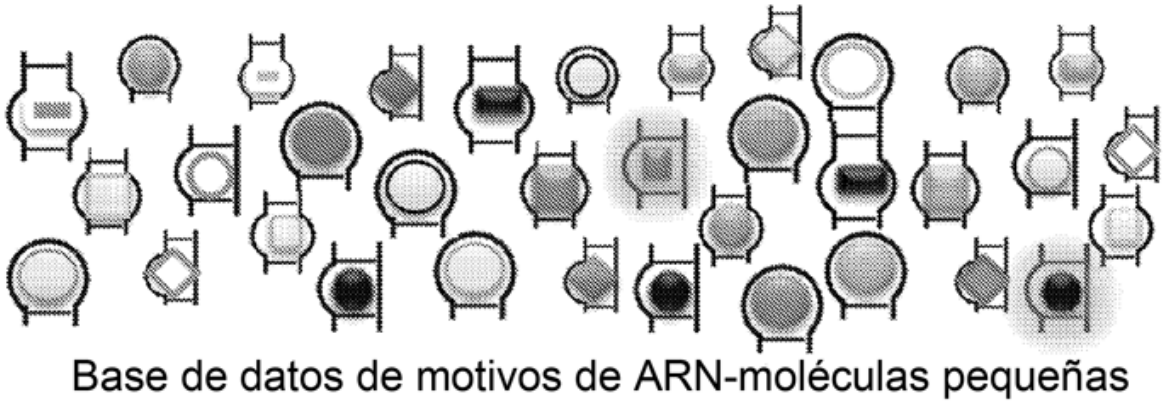

Base de datos de motivos de ARN-moléculas pequeñas

Fig. 3A

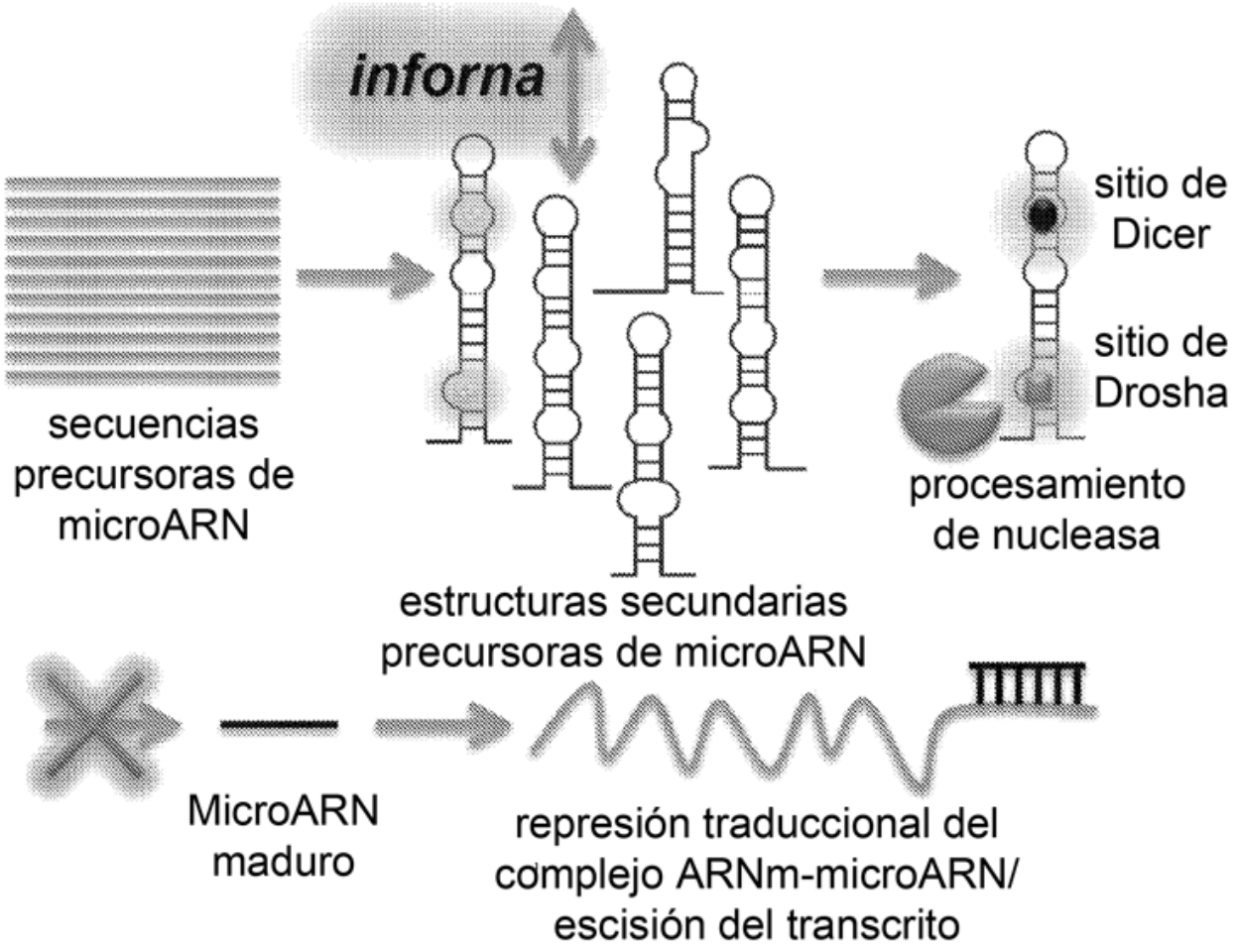

inforna
secuencias precursoras de microARN
estructuras secundarias precursoras de microARN
sitio de Dicer
sitio de Drosha
procesamiento de nucleasa
MicroARN maduro
represión traduccional del complejo ARNm-microARN/ escisión del transcrito

Fig. 3B

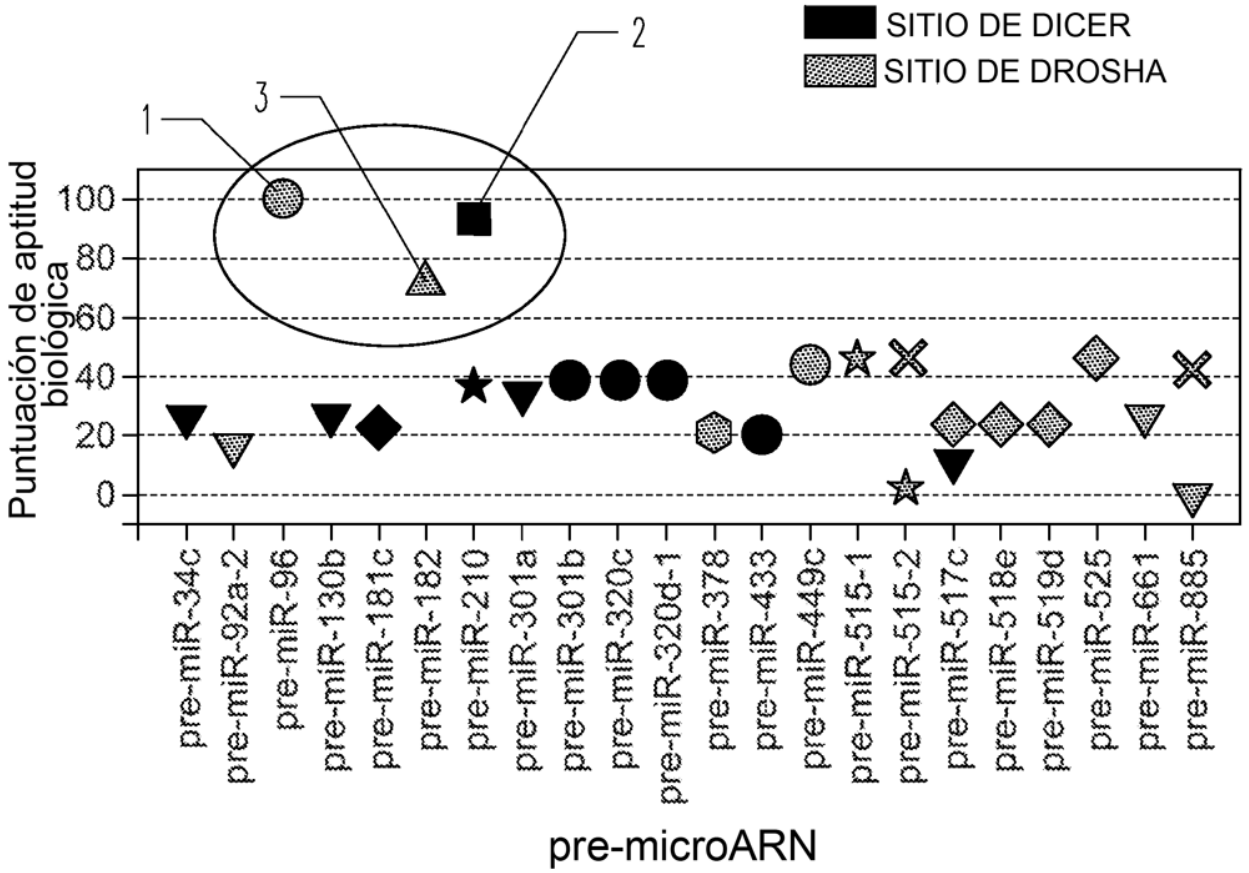

SITIO DE DICER
SITIO DE DROSHA
1
2
3
Puntuación de aptitud biológica
100
80
60
40
20
0
pre-miR-34c
pre-miR-92a-2
pre-miR-96
pre-miR-130b
pre-miR-181c
pre-miR-182
pre-miR-210
pre-miR-301a
pre-miR-301b
pre-miR-320c
pre-miR-320d-1
pre-miR-378
pre-miR-433
pre-miR-449c
pre-miR-515-1
pre-miR-515-2
pre-miR-517c
pre-miR-518e
pre-miR-519d
pre-miR-525
pre-miR-661
pre-miR-885
pre-microARN

Fig. 3C

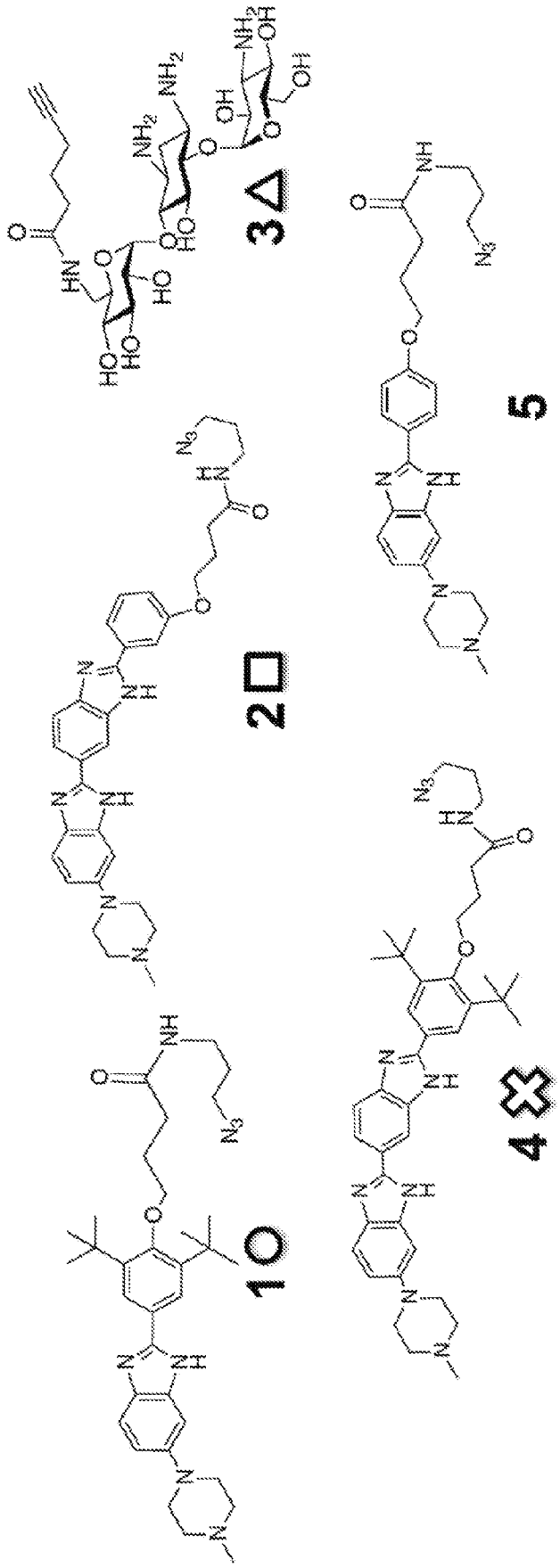
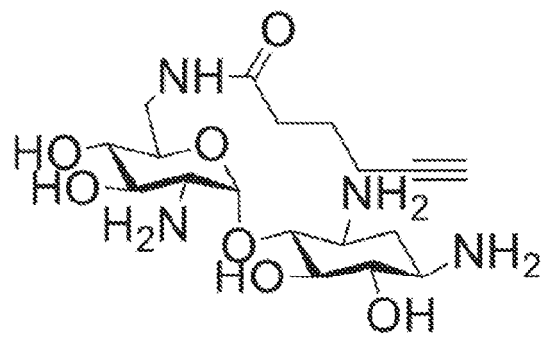
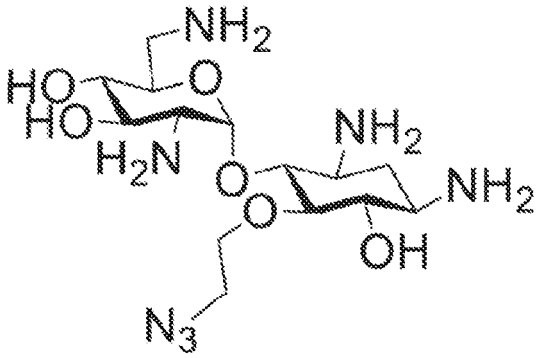


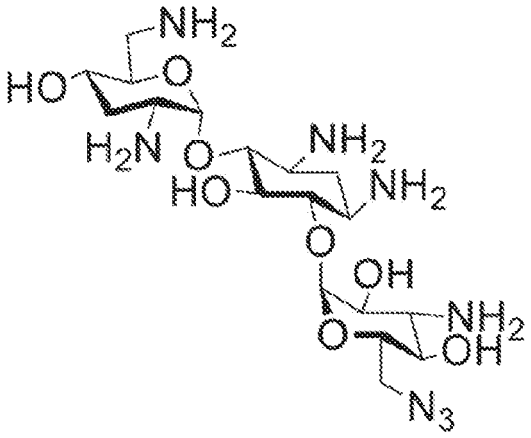
Fig. 3D



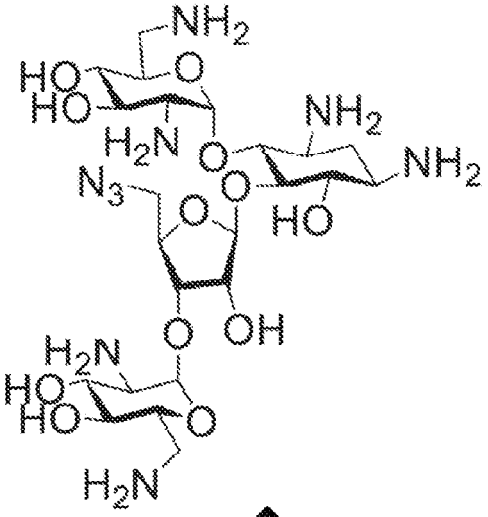
6 ▽



7 ☆



8 ⬡



9 ◇

Fig. 3E

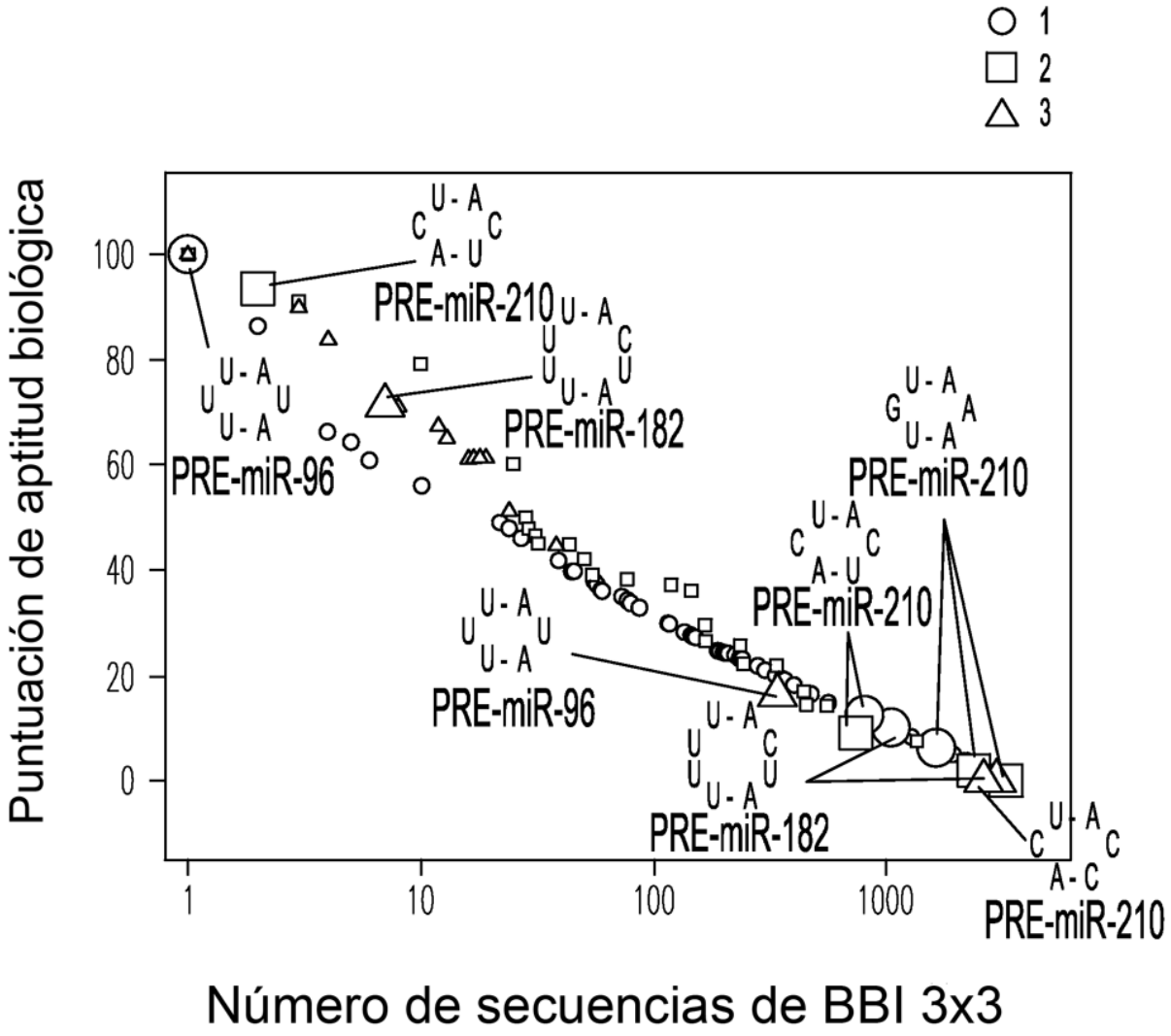

1
2
3
Puntuación de aptitud biológica
100
80
60
40
20
0
1
10
100
1000
PRE-miR-210
PRE-miR-182
PRE-miR-96
PRE-miR-210
PRE-miR-210
PRE-miR-96
PRE-miR-182
PRE-miR-210
Número de secuencias de BBI 3x3

Fig. 4A

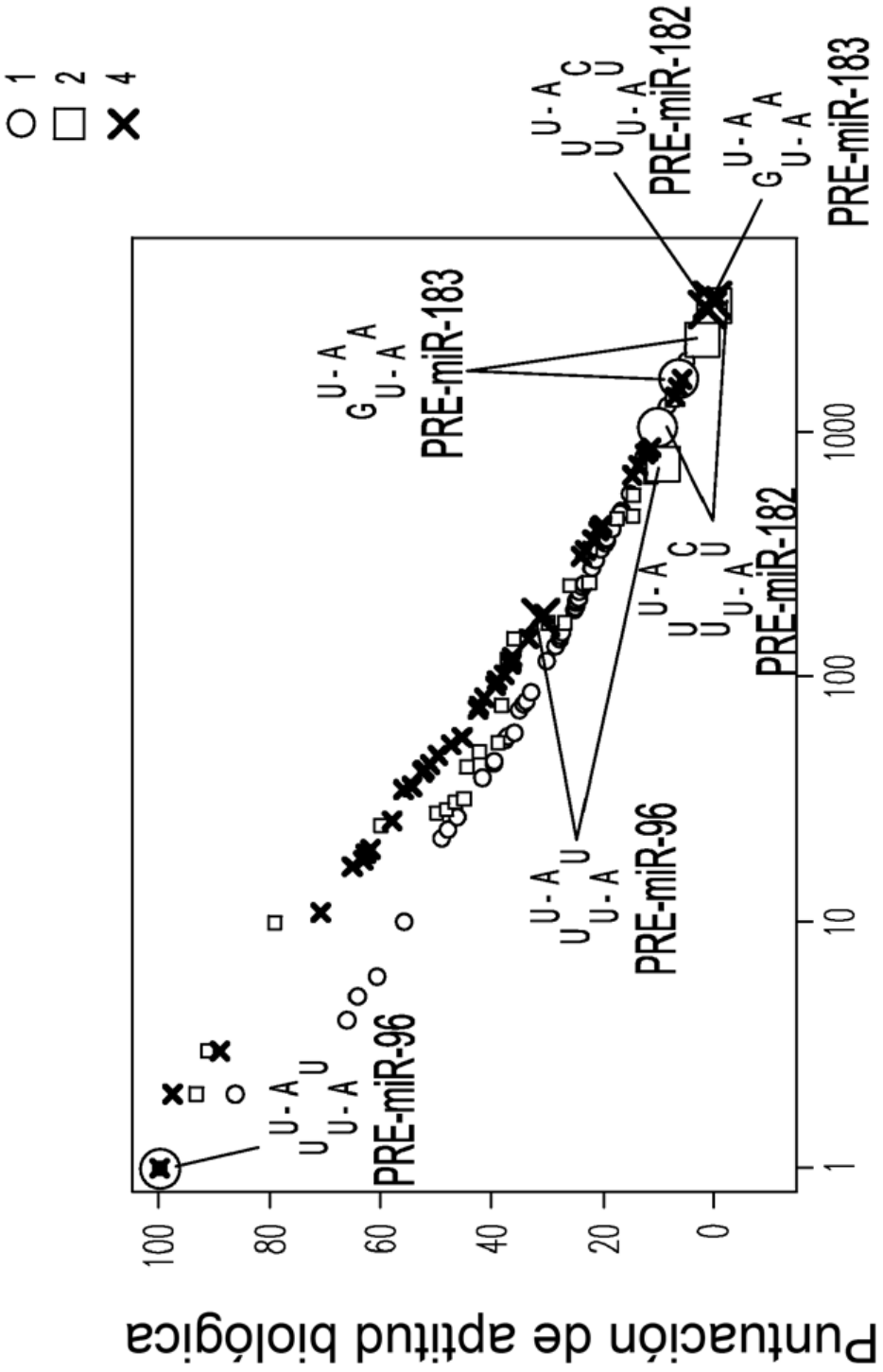

1
2
4
Puntuación de aptitud biológica
100
80
60
40
20
0
1
10
100
1000
Número de secuencias de BBI 3x3
PRE-miR-96
PRE-miR-183
PRE-miR-182
PRE-miR-96
PRE-miR-182
PRE-miR-183

Fig. 4B

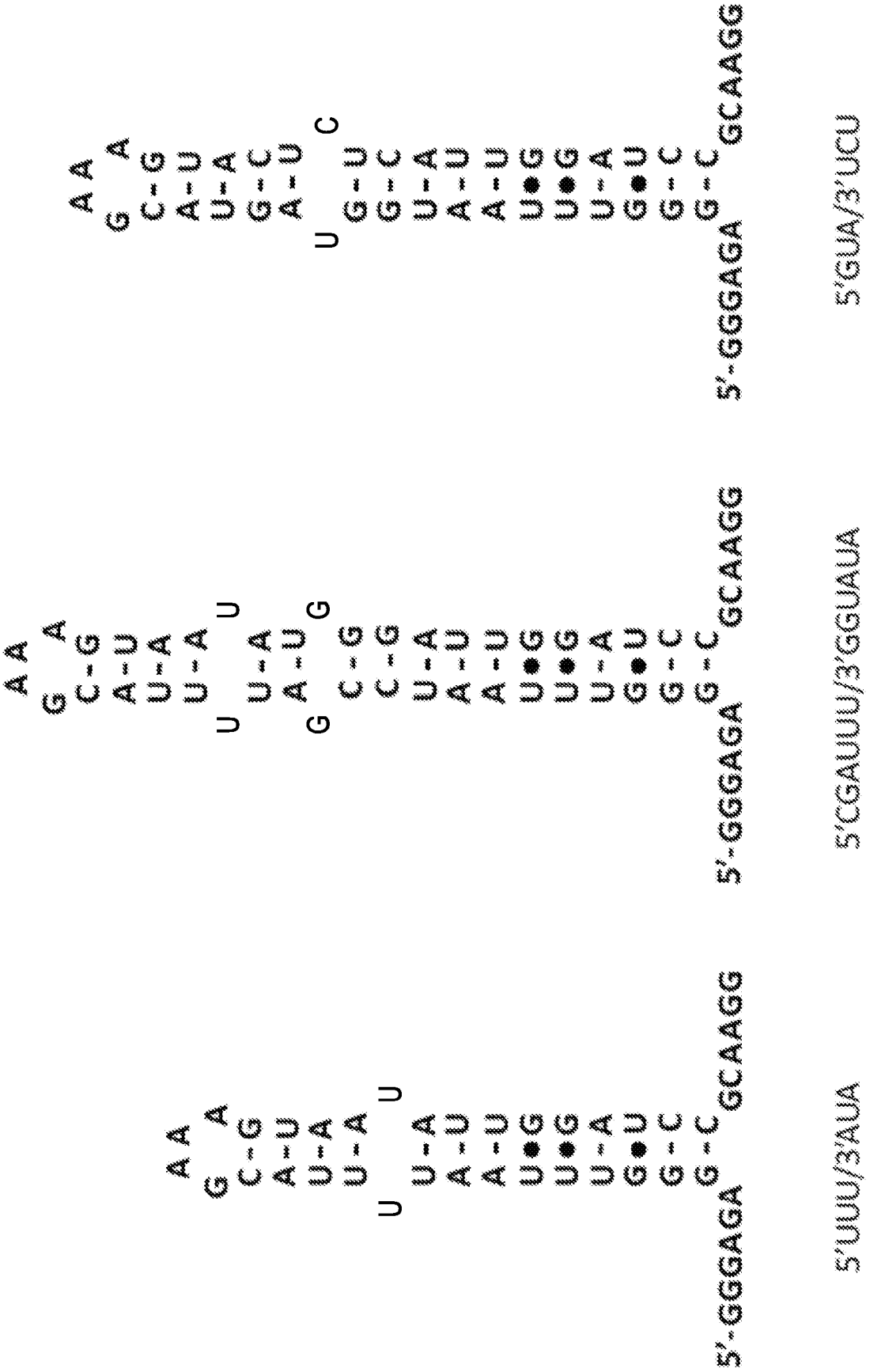

5'-GGGAGA
GCAAGG
5'UUU/3'AUA
5'-GGGAGA
GCAAGG
5'CGAUUU/3'GGUAUA
5'-GGGAGA
GCAAGG
5'GUA/3'UCU

Fig. 4C

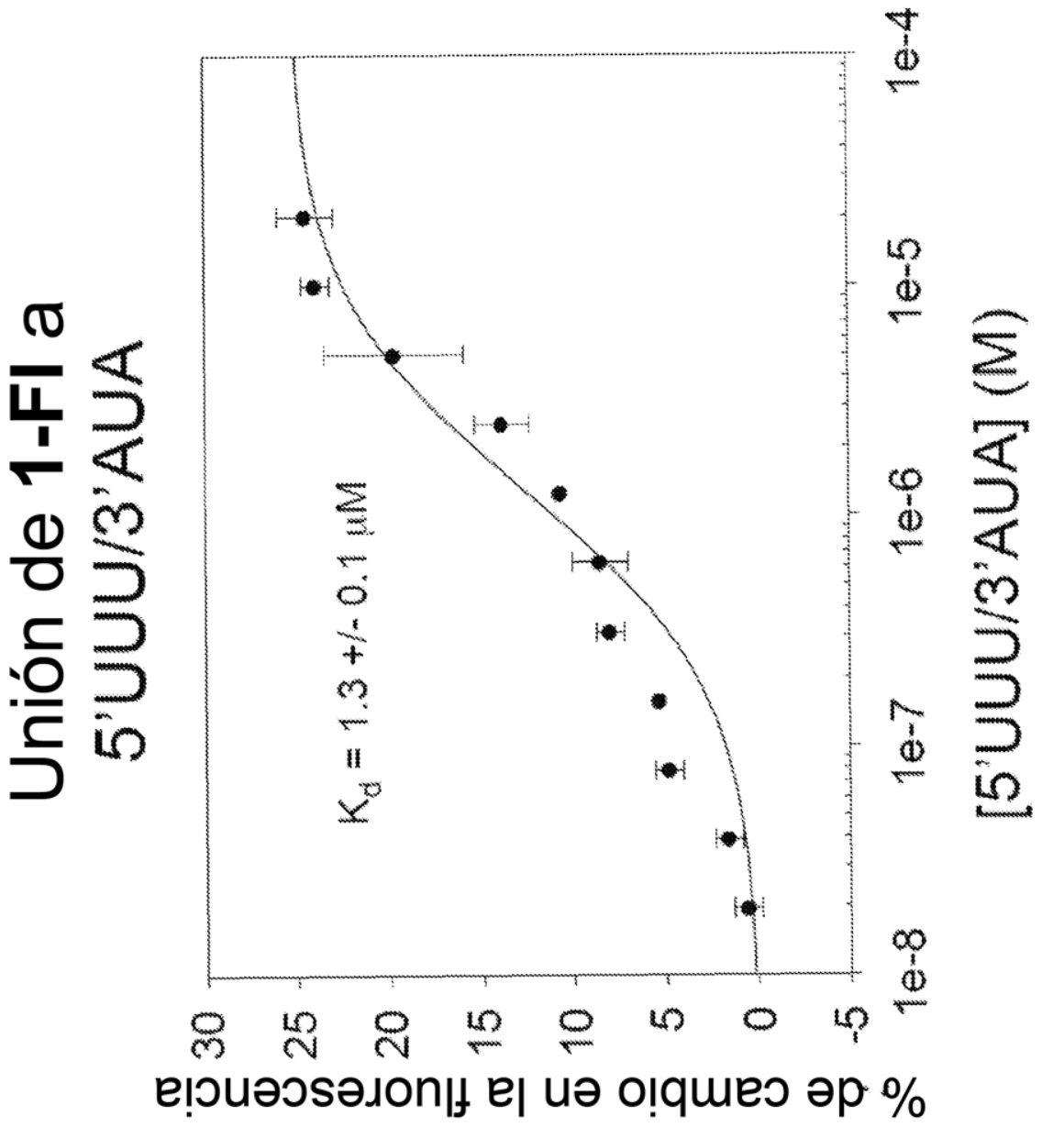

Unión de 1-FI a
5'UUU/3'AUA
% de cambio en la fluorescencia
30
25
20
15
10
5
0
-5
1e-8
1e-7
1e-6
1e-5
1e-4
[5'UUU/3'AUA] (M)
K_d = 1.3 +/- 0.1 µM

Fig. 4D

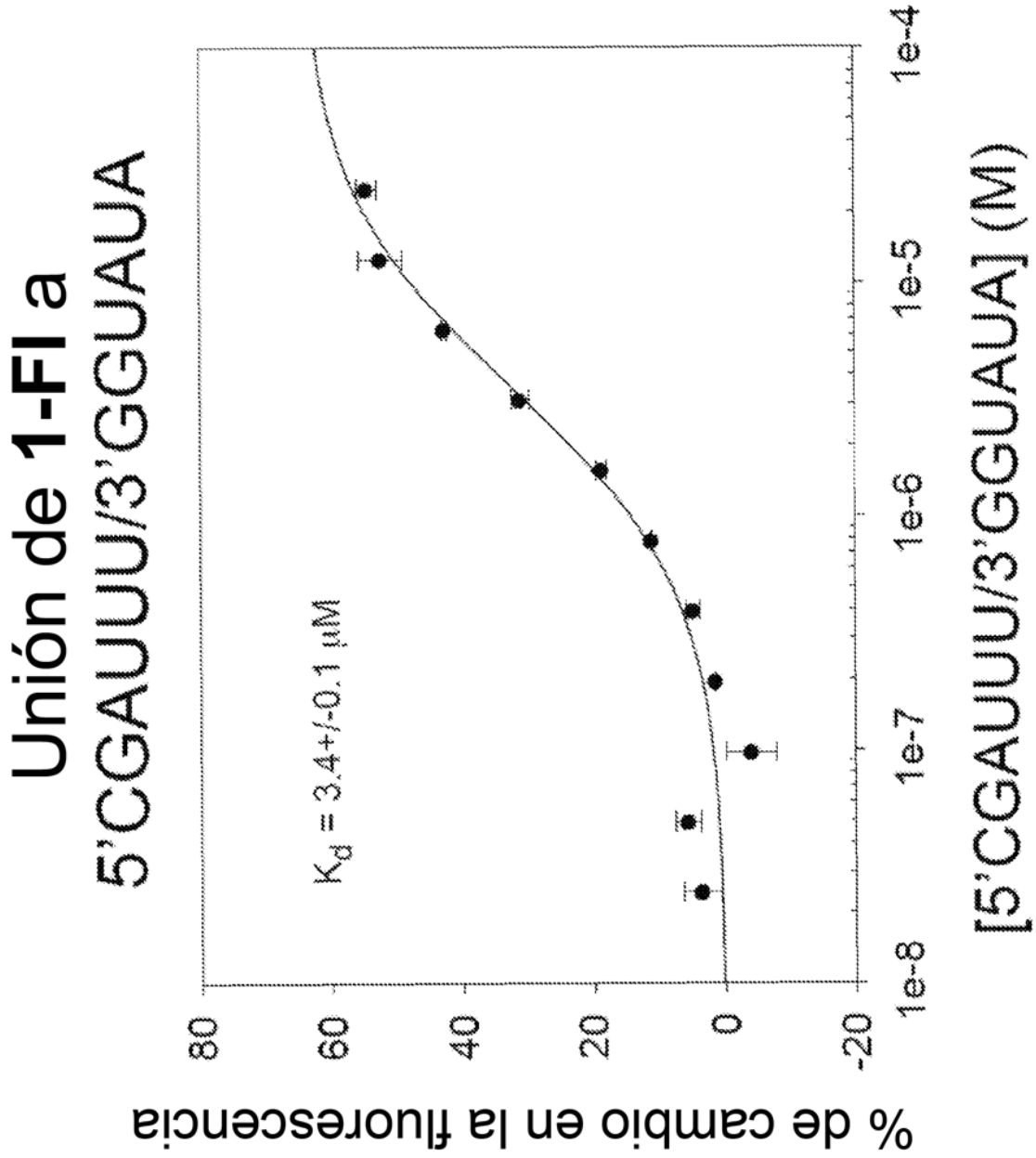

Unión de 1-FI a
5'CGAUUU/3'GGUAUA
K_d = 3.4+/-0.1 µM
% de cambio en la fluorescencia
80
60
40
20
0
-20
1e-8
1e-7
1e-6
1e-5
1e-4
[5'CGAUUU/3'GGUAUA] (M)

Fig. 4E

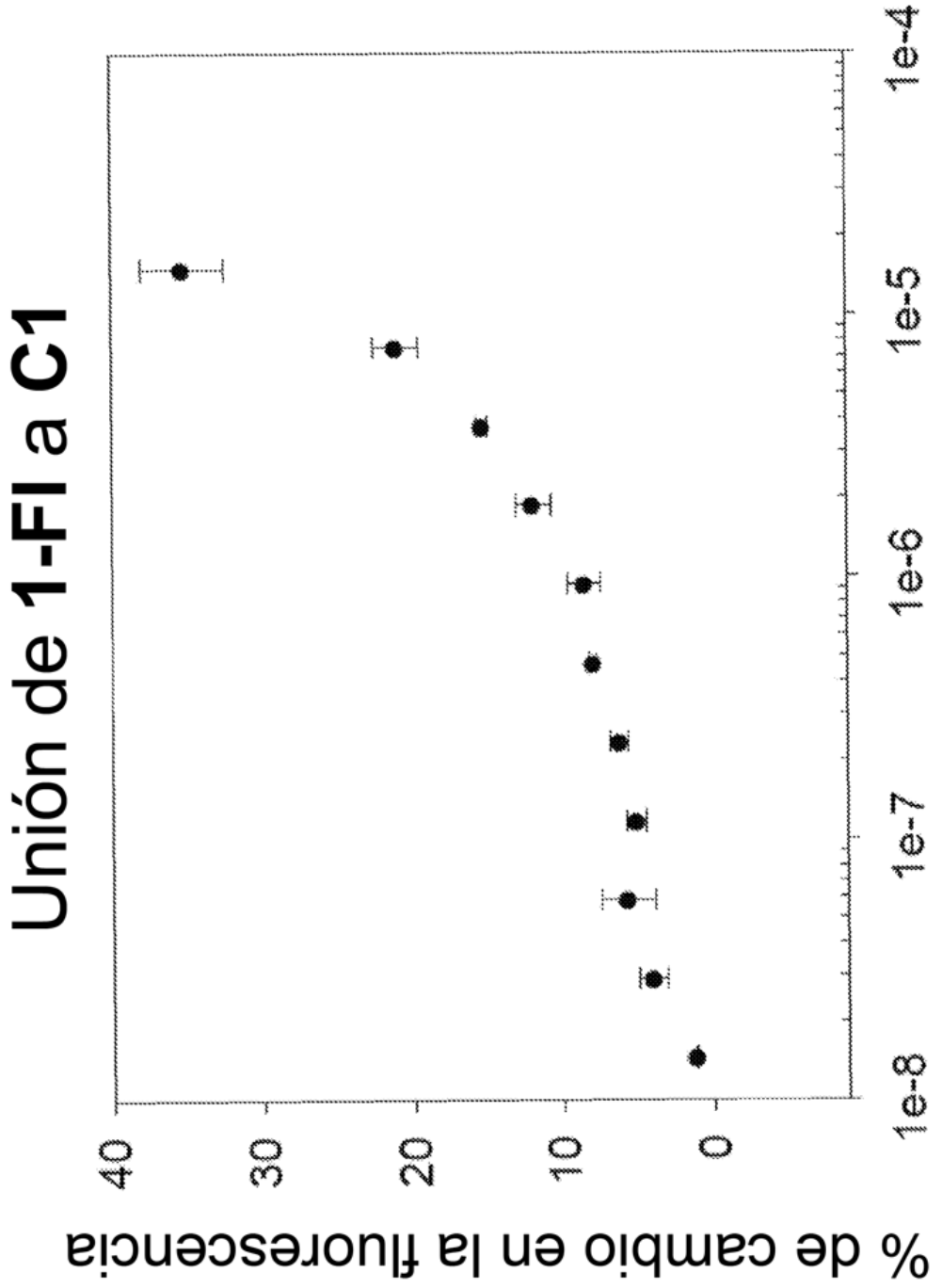

Unión de 1-FI a C1
% de cambio en la fluorescencia
40
30
20
10
0
1e-8
1e-7
1e-6
1e-5
1e-4
[C1] (M)

Fig. 4F

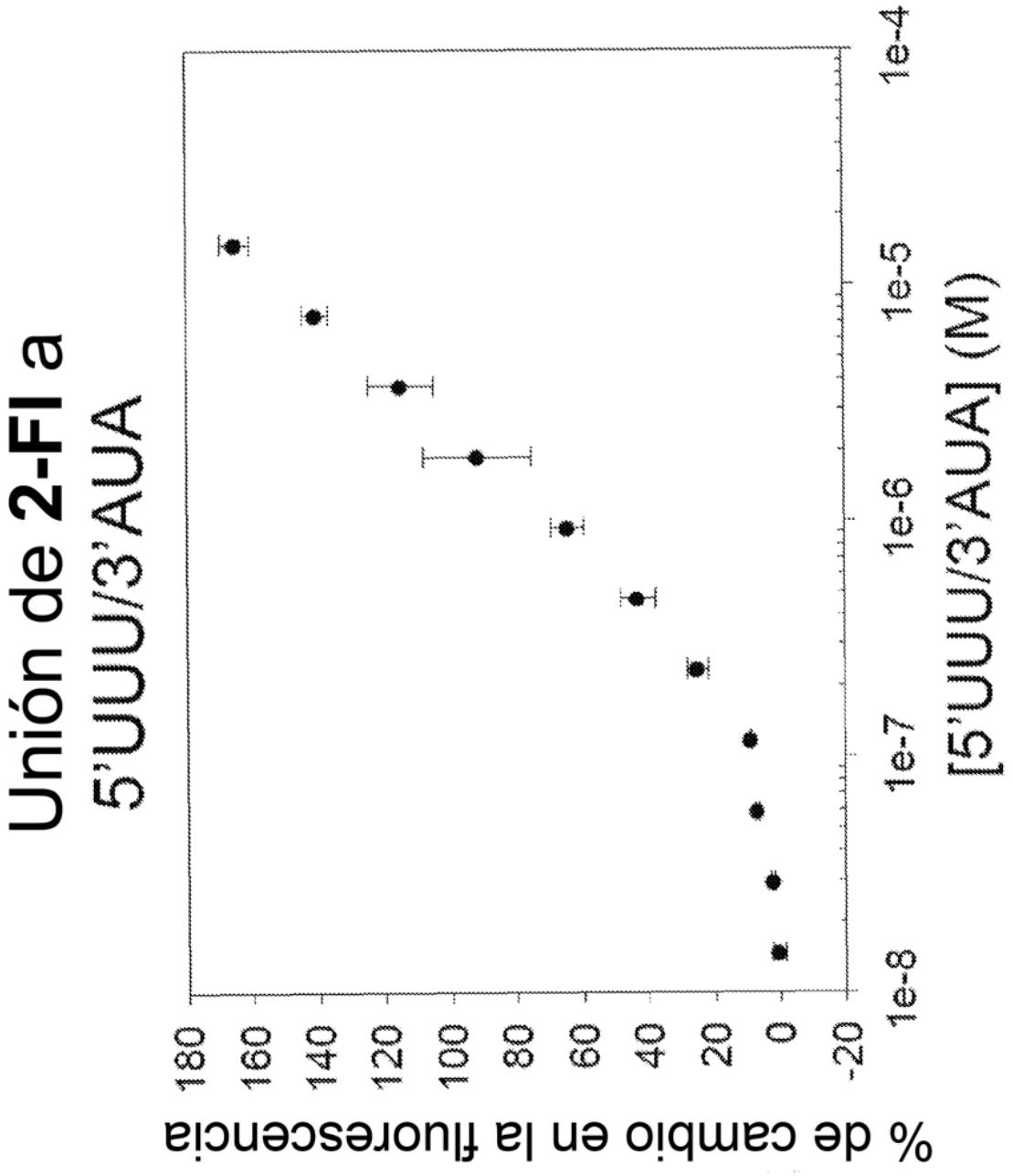

Unión de 2-FI a
5'UUU/3'AUA
% de cambio en la fluorescencia
180
160
140
120
100
80
60
40
20
0
-20
1e-8
1e-7
1e-6
1e-5
1e-4
[5'UUU/3'AUA] (M)

Fig. 4G

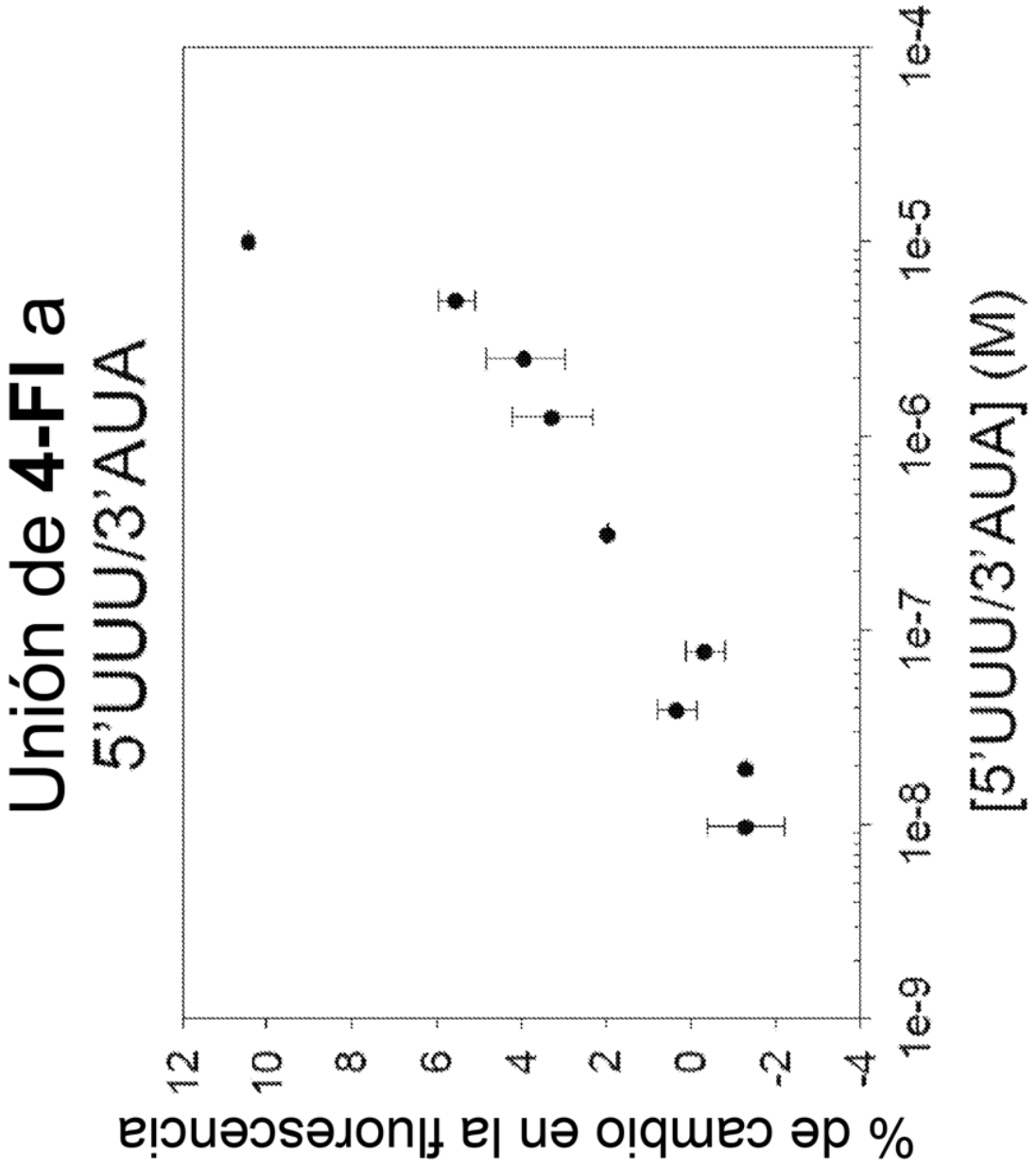

Unión de 4-FI a
5'UUU/3'AUA
% de cambio en la fluorescencia
12
10
8
6
4
2
0
-2
-4
1e-9
1e-8
1e-7
1e-6
1e-5
1e-4
[5'UUU/3'AUA] (M)

Fig. 4H

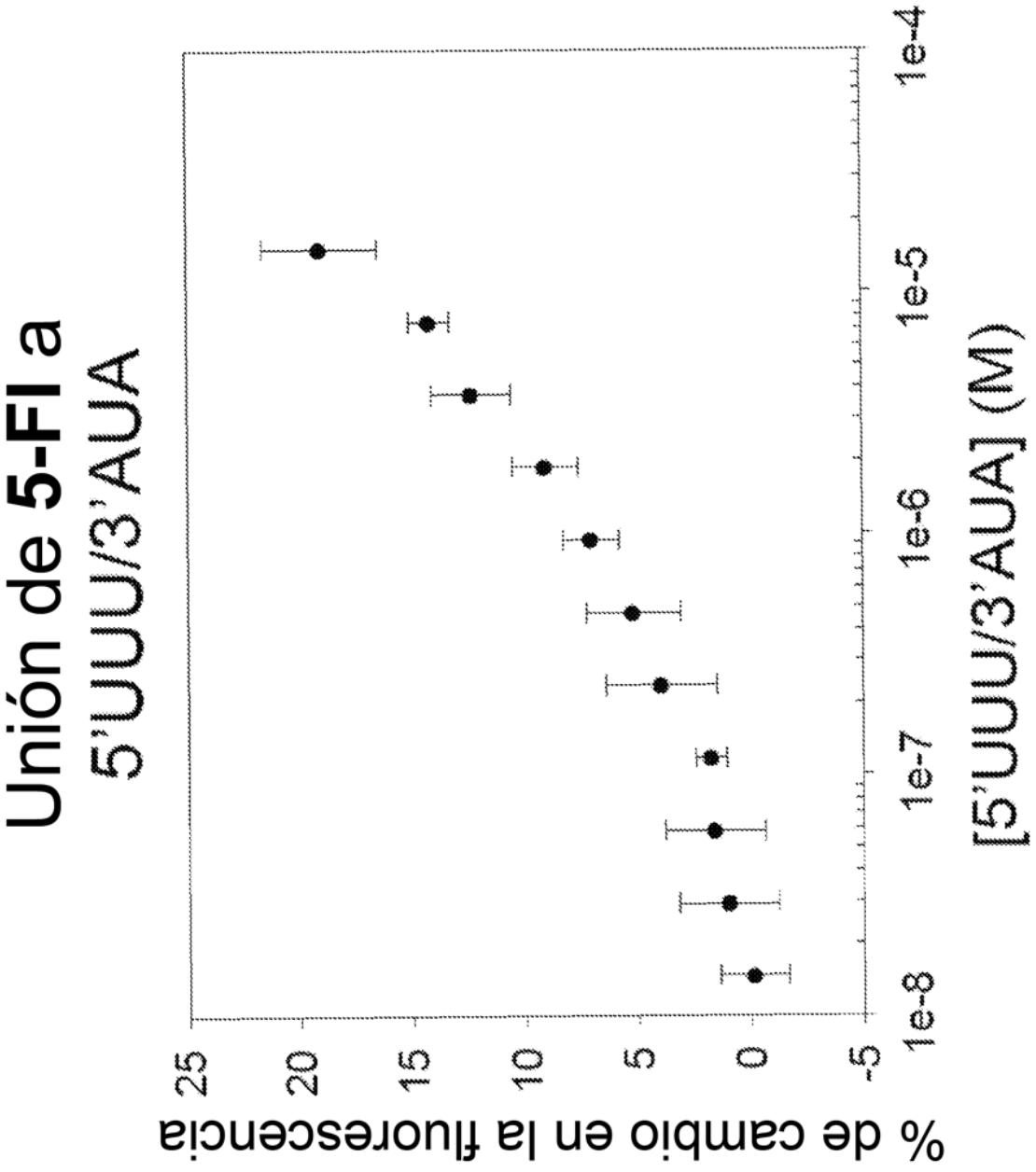

Unión de 5-FI a
5'UUU/3'AUA
% de cambio en la fluorescencia
25
20
15
10
5
0
-5
1e-8
1e-7
1e-6
1e-5
1e-4
[5'UUU/3'AUA] (M)

Fig. 4I

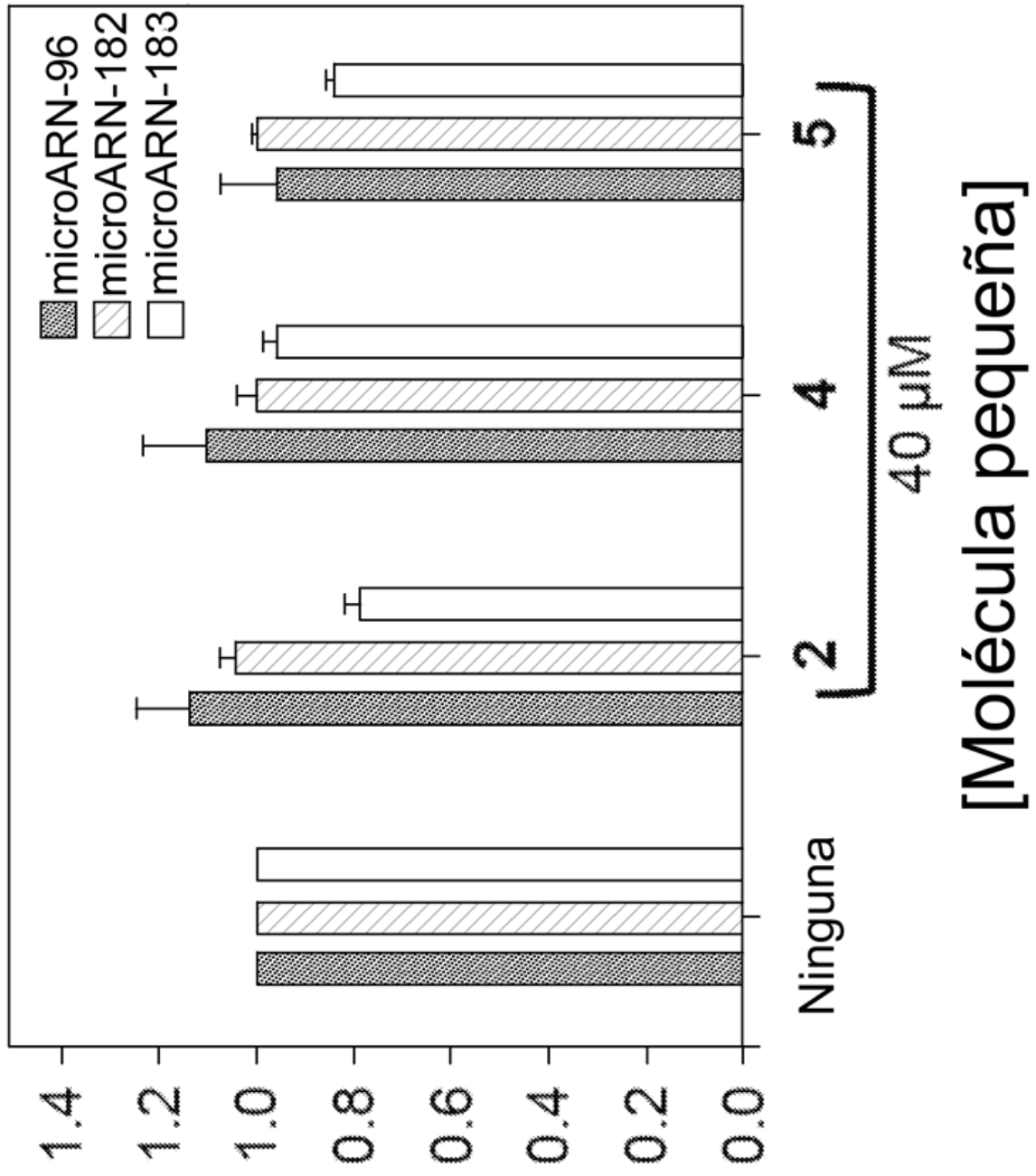

microARN-96
microARN-182
microARN-183
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Ninguna
2
4
5
40 µM
[Molécula pequeña]

Fig. 4J

precursor de miR-96
precursor de miR-182
precursor de miR-183
precursor de miR-210

Fig. 5A

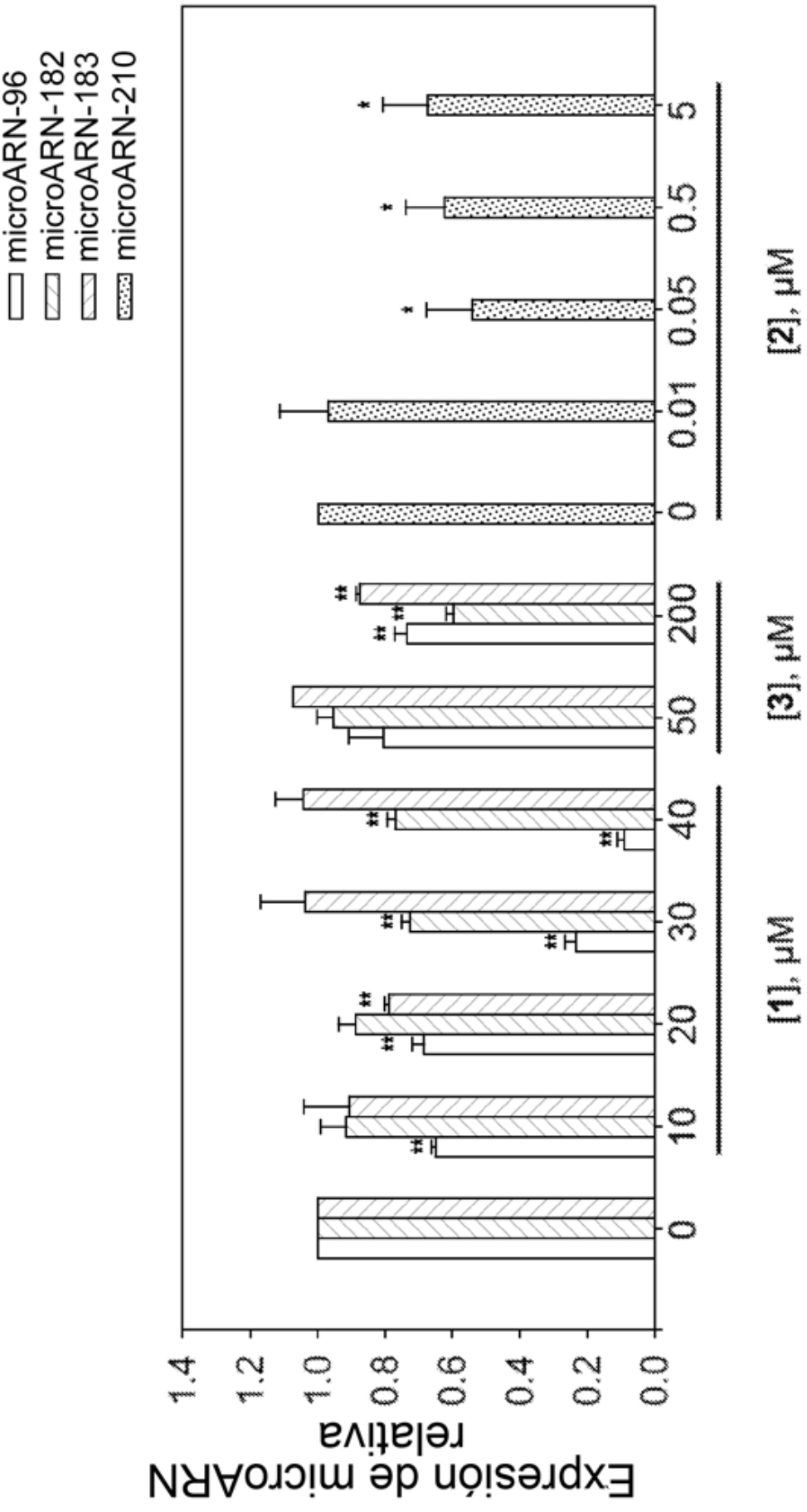

microARN-96
microARN-182
microARN-183
microARN-210
Expresión de microARN relativa
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
200
0
0.01
0.05
0.5
5
[1], µM
[3], µM
[2], µM

Fig. 5B

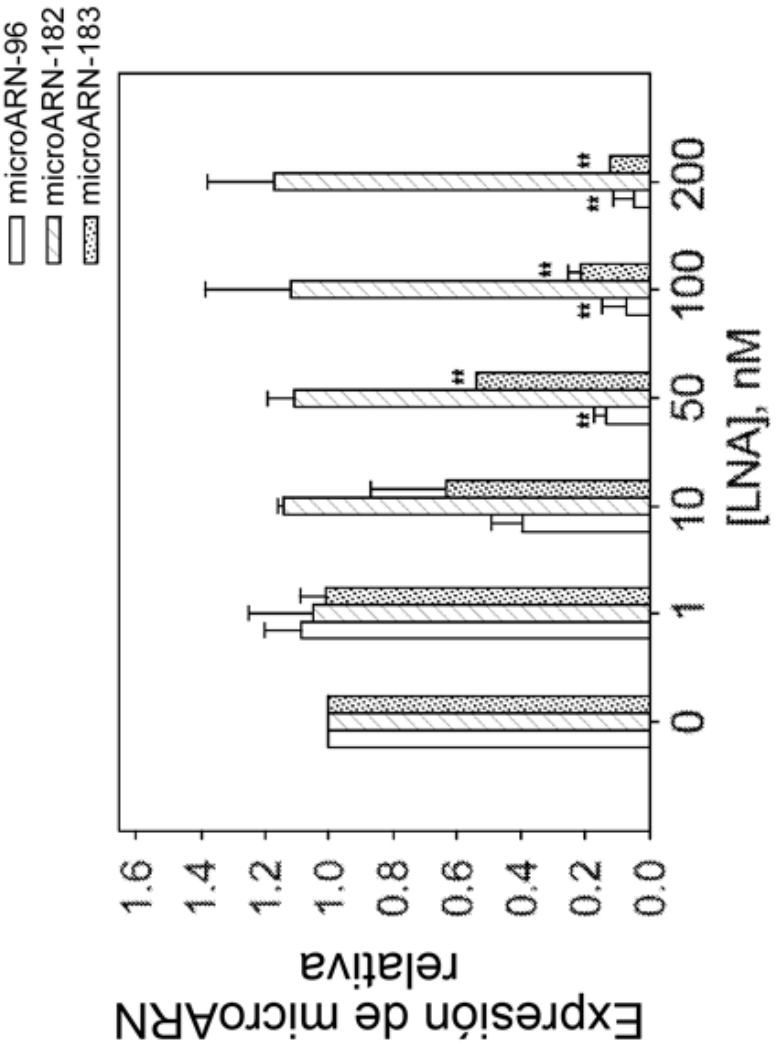


Fig. 5C

miR-96:
5'UUUGGCACUAGCACAUUUUUGCU

miR-182:
5'UUUGGCAAUGGUAGAACUCACACU

miR-183:
5'UAUGGCACUGGUAGAAUUCACU

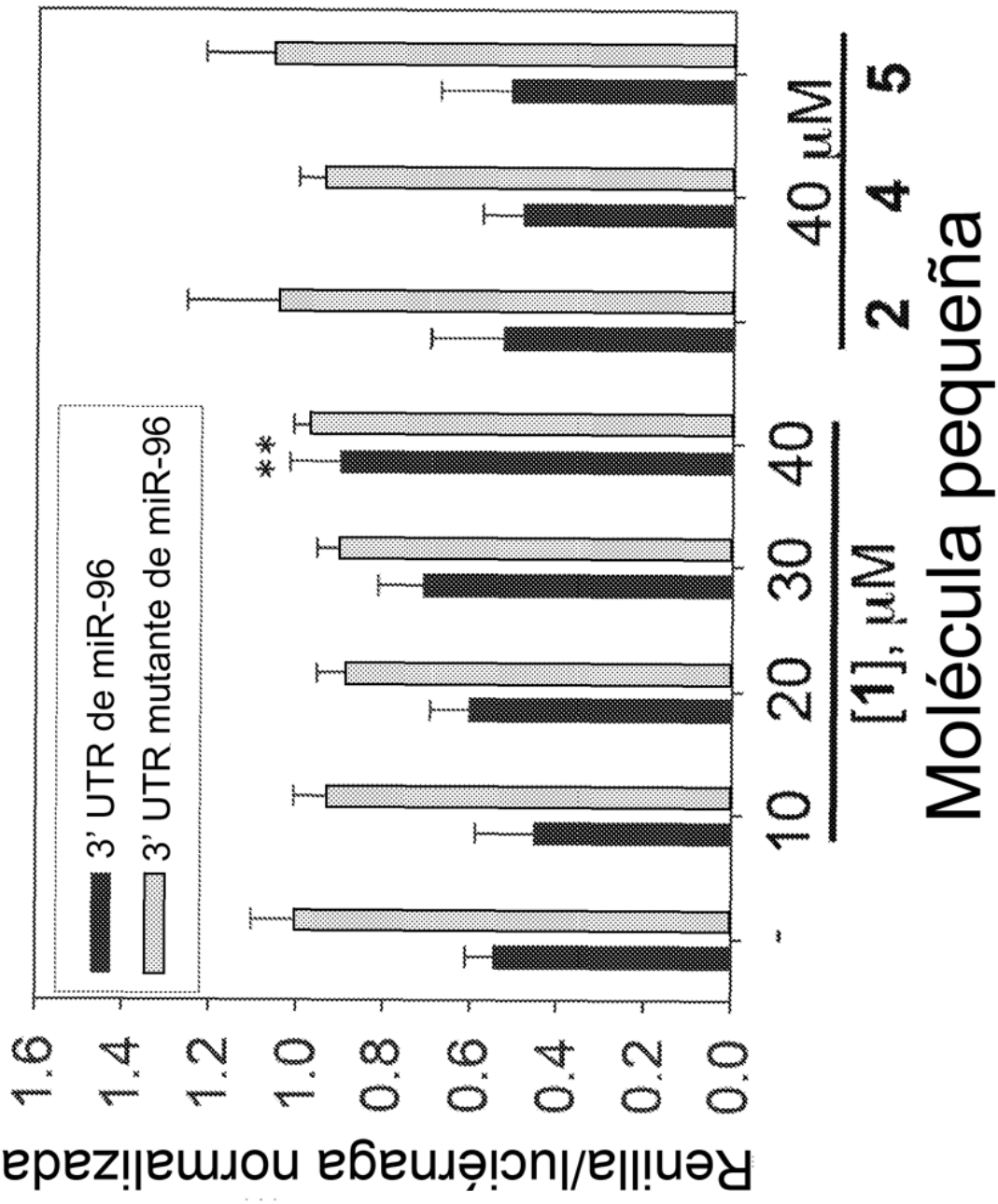

Renilla/luciérnaga normalizada
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
3' UTR de miR-96
3' UTR mutante de miR-96
**
-
10
20
30
40
[1], µM
2
4
5
40 µM
Molécula pequeña

Fig. 6A

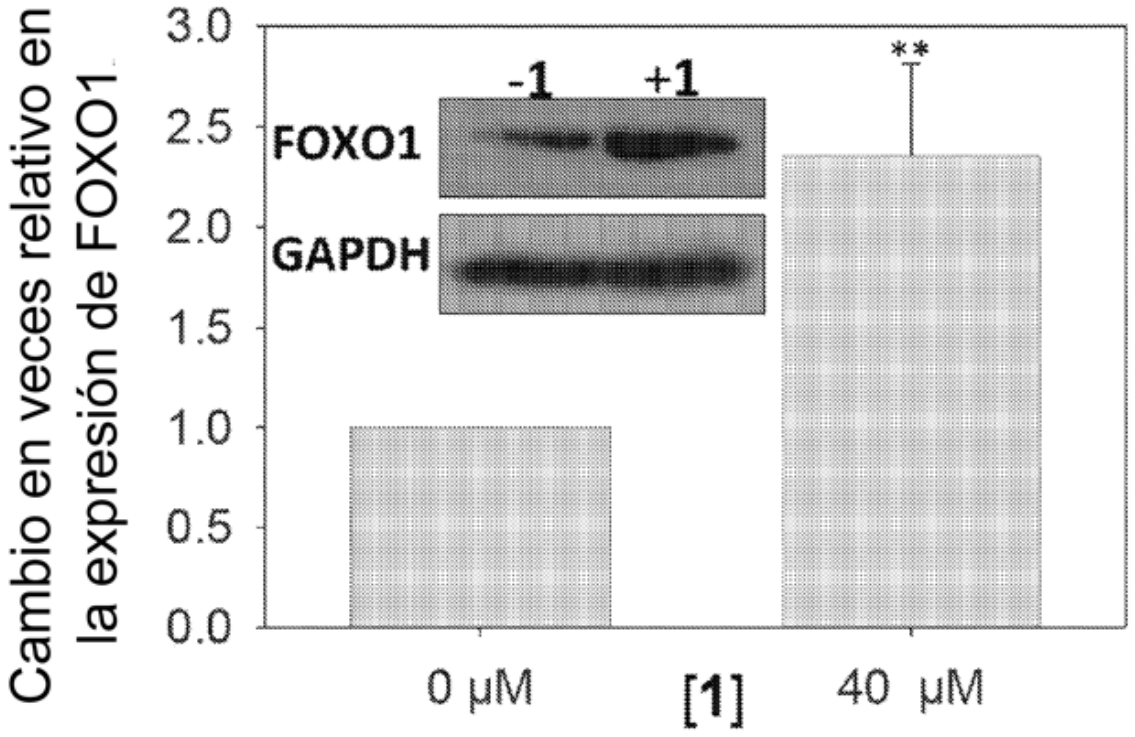

Cambio en veces relativo en la expresión de FOXO1
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-1
+1
FOXO1
GAPDH
**
0 µM
[1]
40 µM

Fig. 6B

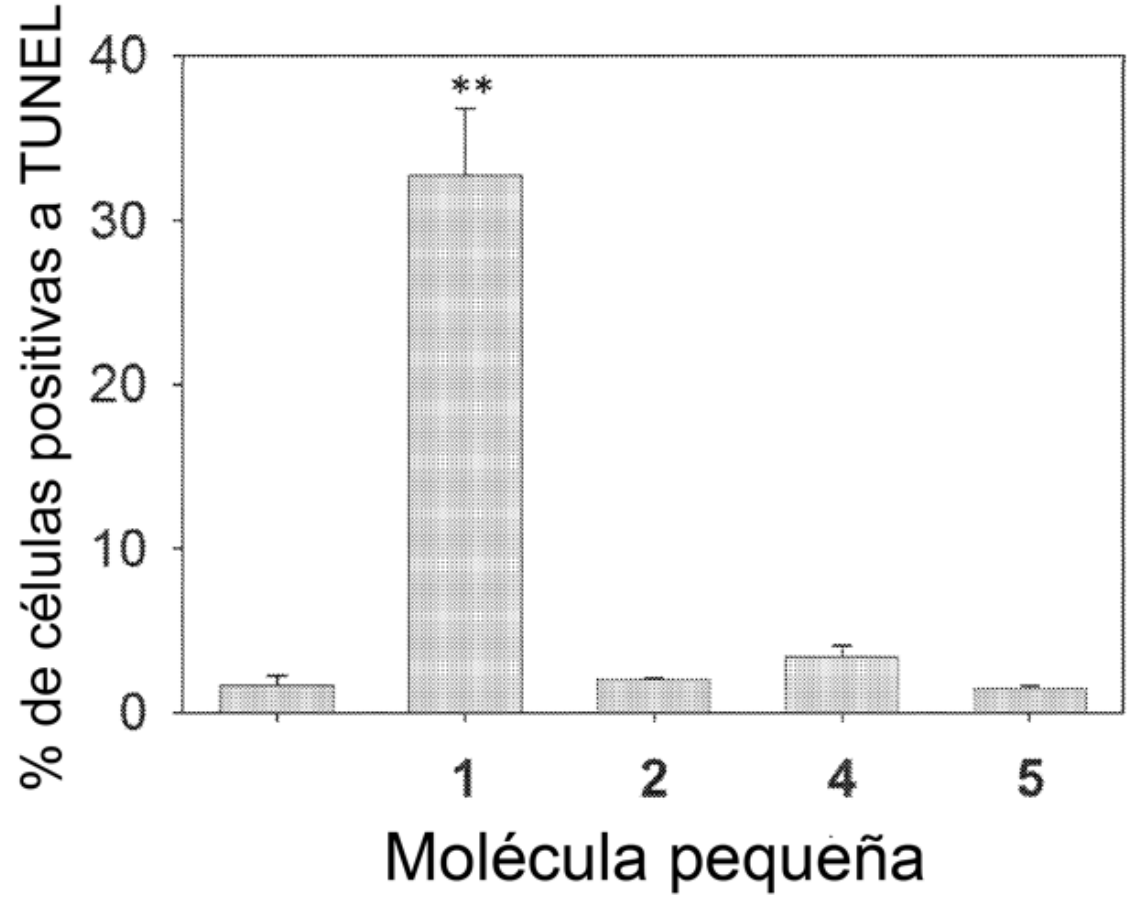

% de células positivas a TUNEL
40
30
20
10
0
**
1
2
4
5
Molécula pequeña

Fig. 6C

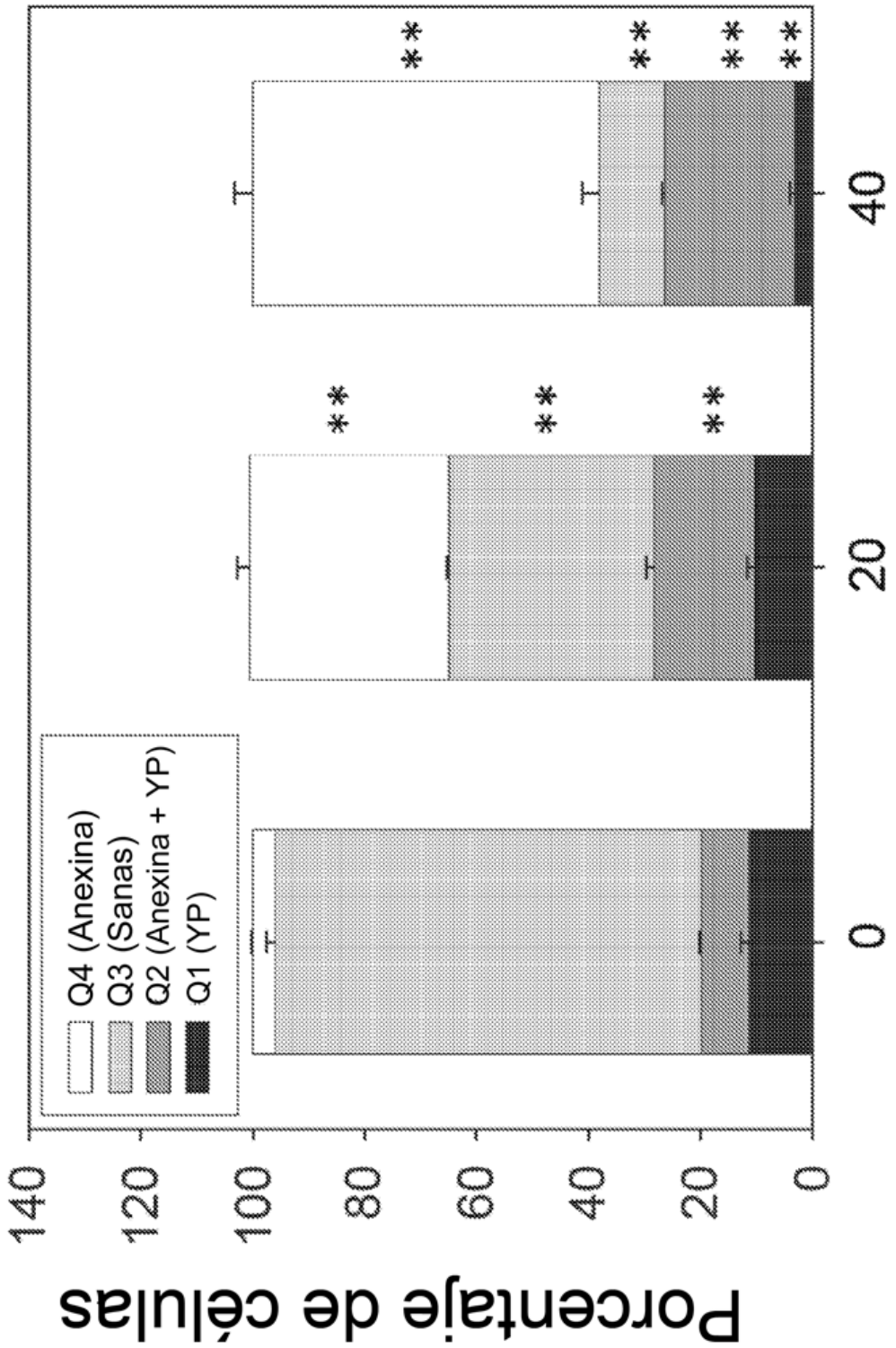

Porcentaje de células
140
120
100
80
60
40
20
0
Q4 (Anexina)
Q3 (Sanas)
Q2 (Anexina + YP)
Q1 (YP)
0
20
40
[1], µM
**

Fig. 6D

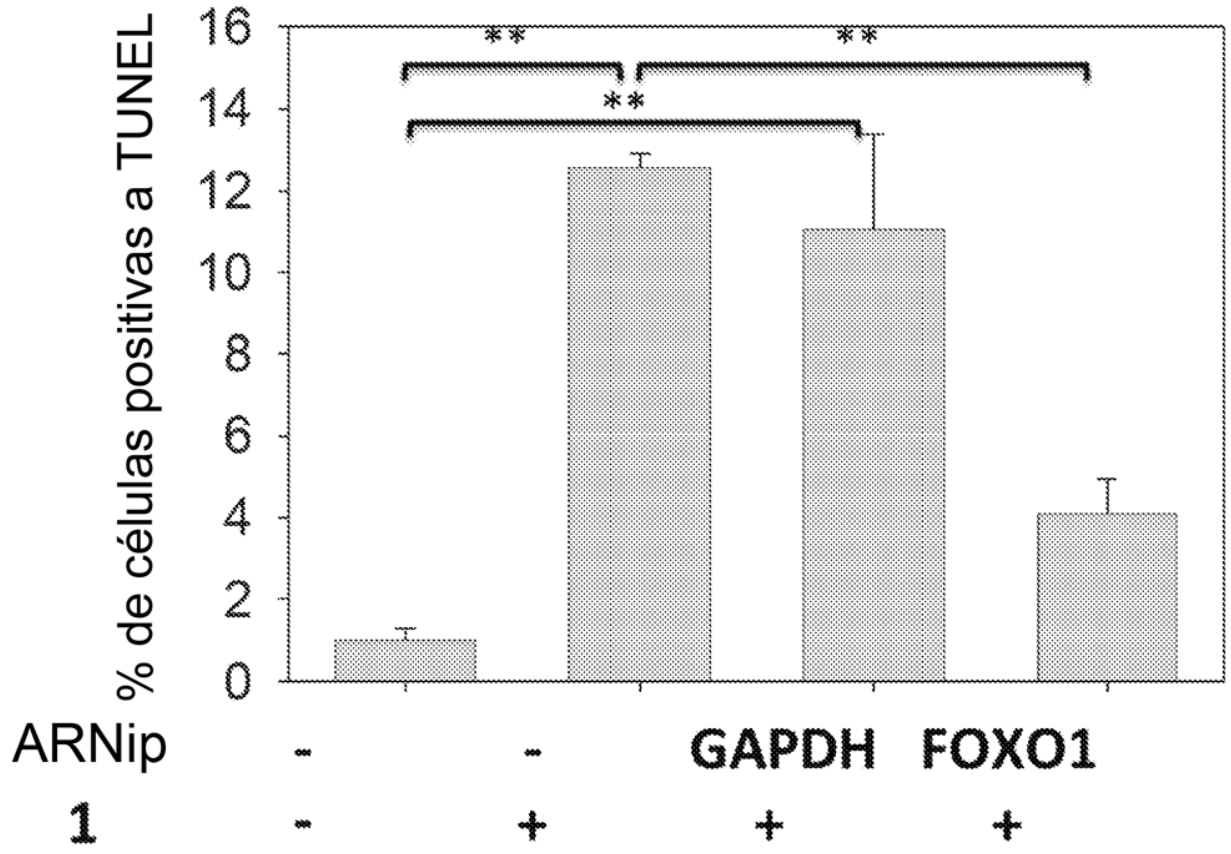

% de células positivas a TUNEL
16
14
12
10
8
6
4
2
0
**
**
**
ARNip
-
-
GAPDH
FOXO1
1
-
+
+
+

Fig. 6E

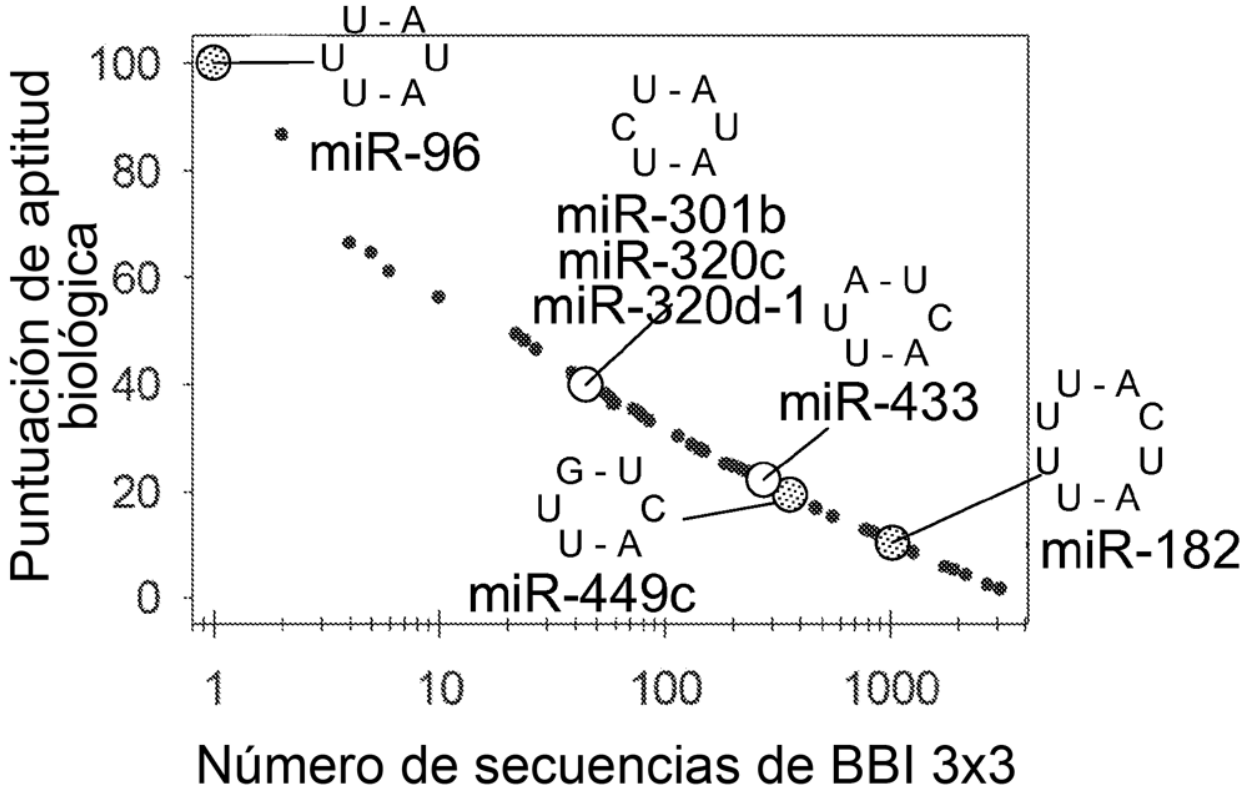

Puntuación de aptitud biológica
100
80
60
40
20
0
1
10
100
1000
Número de secuencias de BBI 3x3
miR-96
miR-301b
miR-320c
miR-320d-1
miR-433
miR-449c
miR-182

Fig. 7A

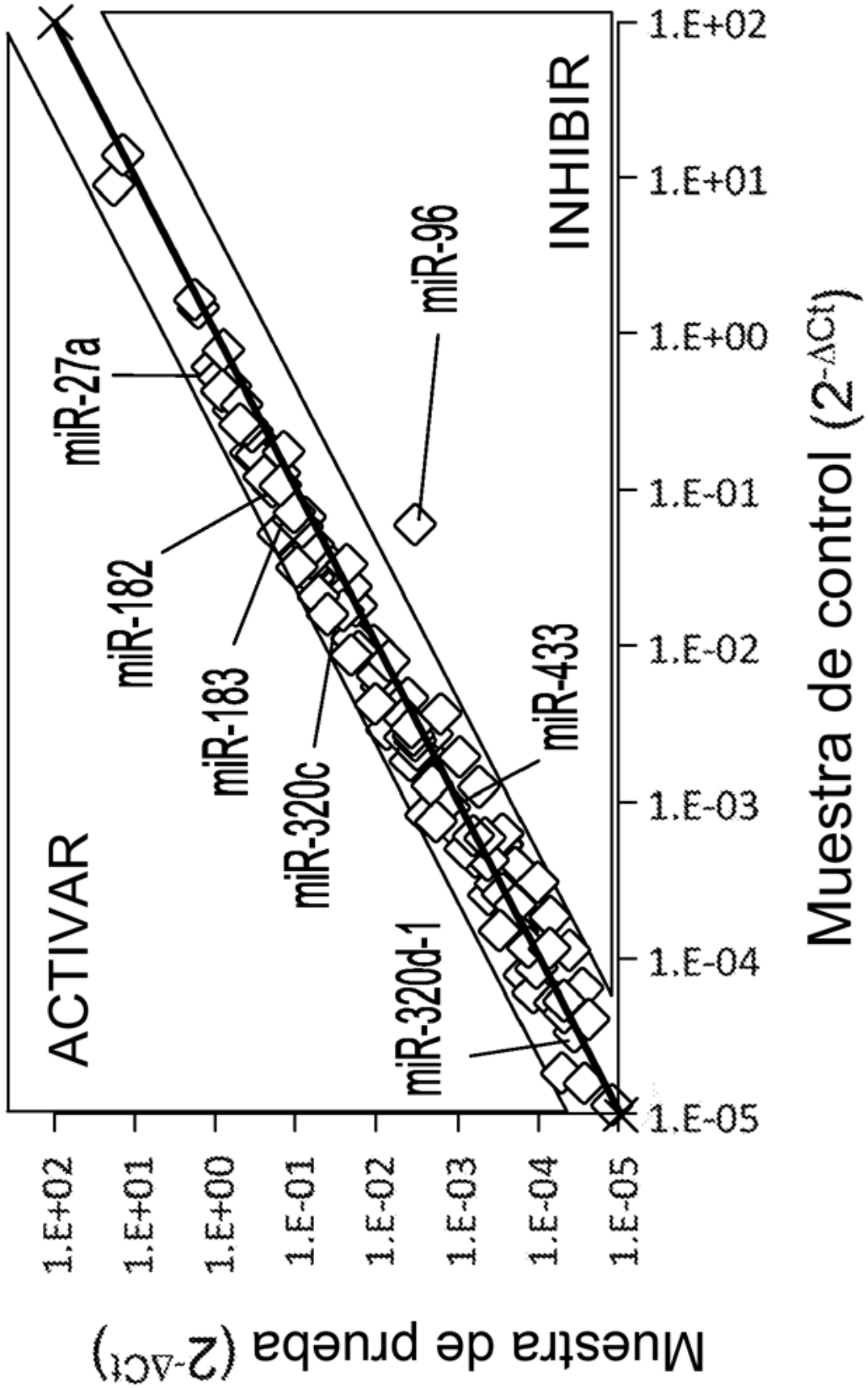

ACTIVAR
INHIBIR
miR-27a
miR-182
miR-183
miR-320c
miR-320d-1
miR-96
miR-433
1.E+02
1.E+01
1.E+00
1.E-01
1.E-02
1.E-03
1.E-04
1.E-05
Muestra de control (2-ΔCt)
Muestra de prueba (2-ΔCt)

Fig. 7B

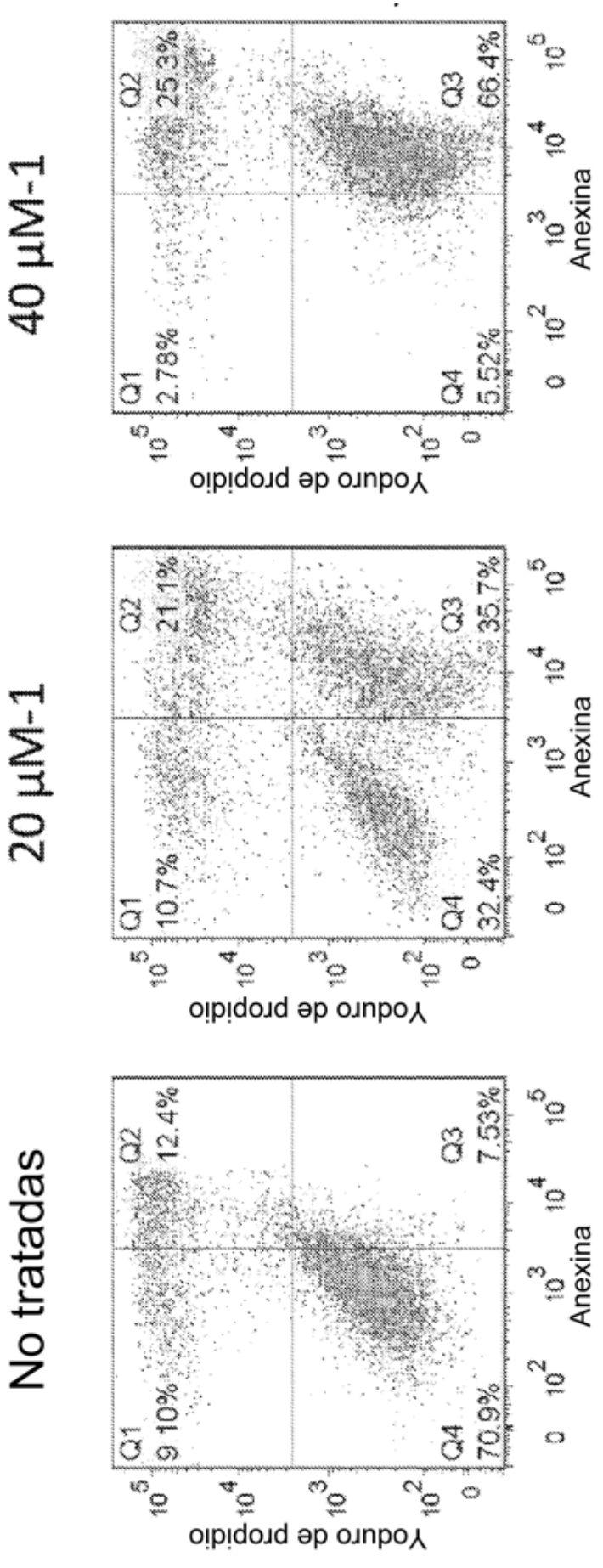

No tratadas
20 µM-1
40 µM-1
Yoduro de propidio
Anexina
Q1
9.10%
Q2
12.4%
Q3
7.53%
Q4
70.9%
Q1
10.7%
Q2
21.1%
Q3
35.7%
Q4
32.4%
Q1
2.78%
Q2
25.3%
Q3
66.4%
Q4
5.52%

Fig. 8A

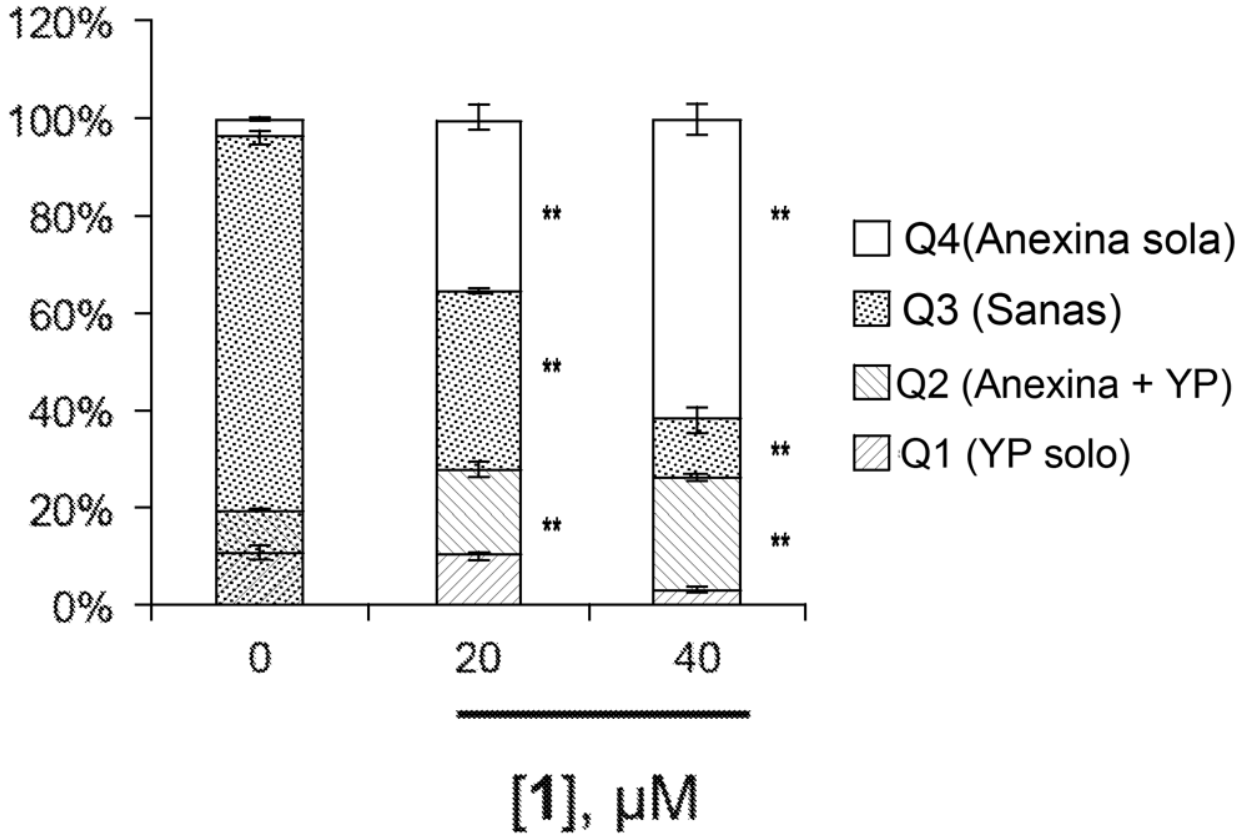

120%
100%
80%
60%
40%
20%
0%
0
20
40
**
**
**
**
**
**
Q4(Anexina sola)
Q3 (Sanas)
Q2 (Anexina + YP)
Q1 (YP solo)
[1], µM

Fig. 8B

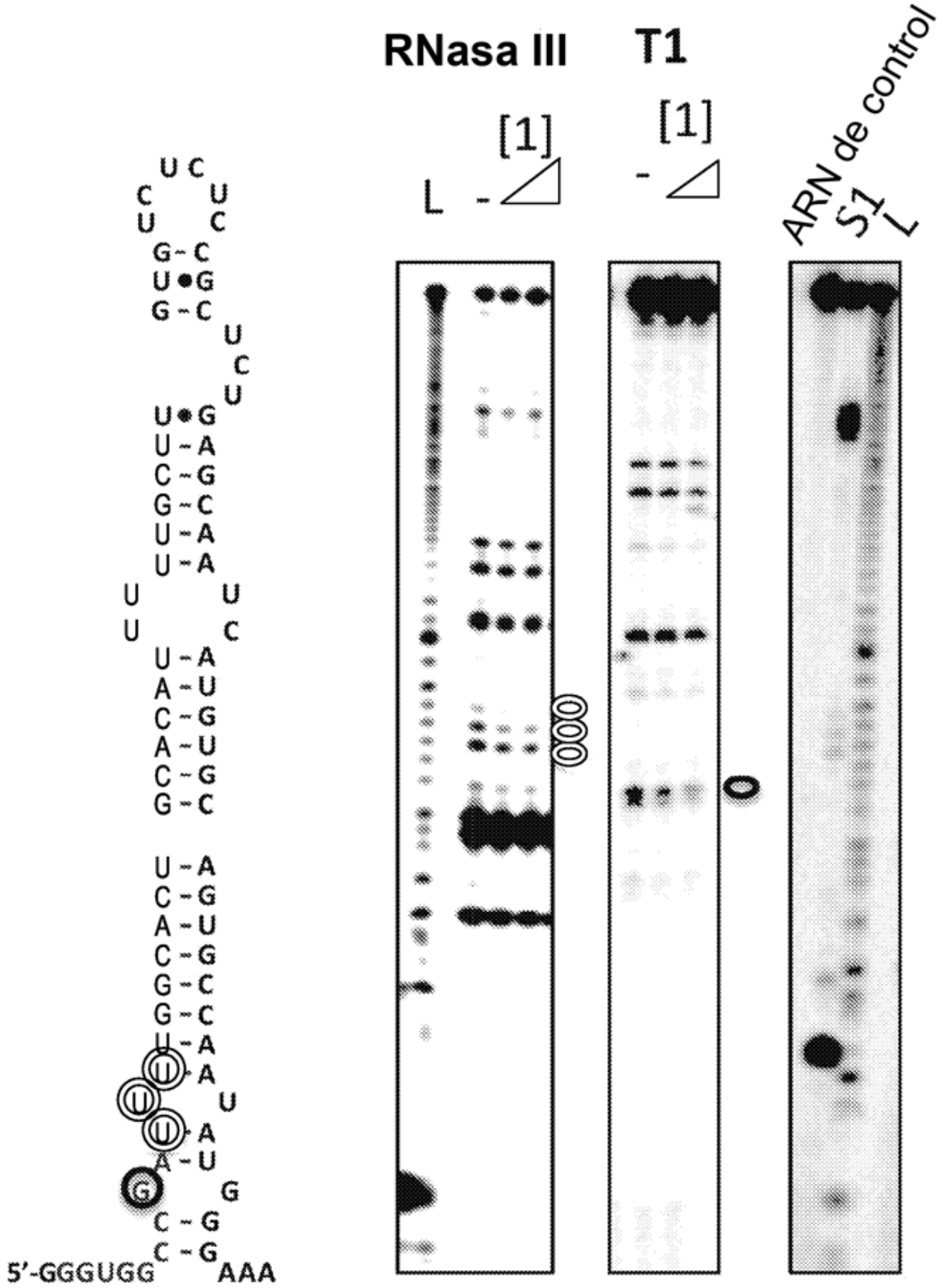


Fig. 9A

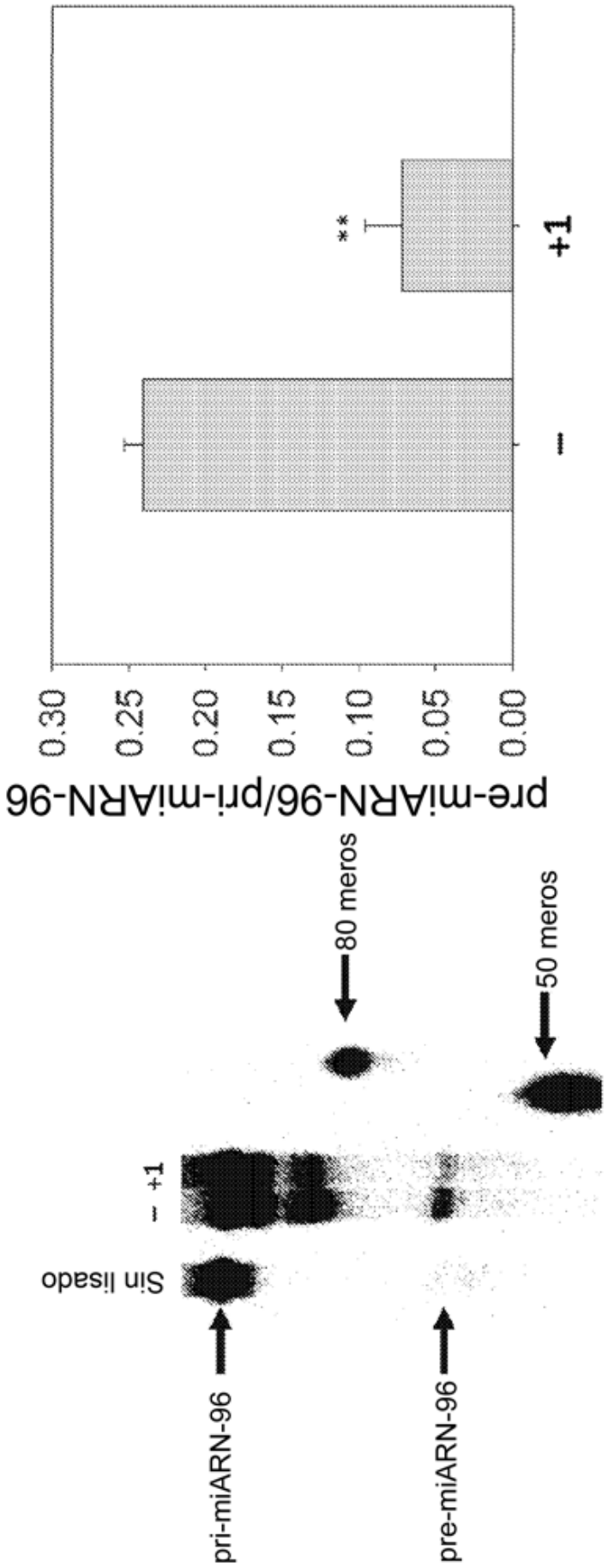

pri-miARN-96
pre-miARN-96
Sin lisado
– +1
80 meros
50 meros
pre-miARN-96/pri-miARN-96
0.30
0.25
0.20
0.15
0.10
0.05
0.00
**
–
+1

Fig. 9B

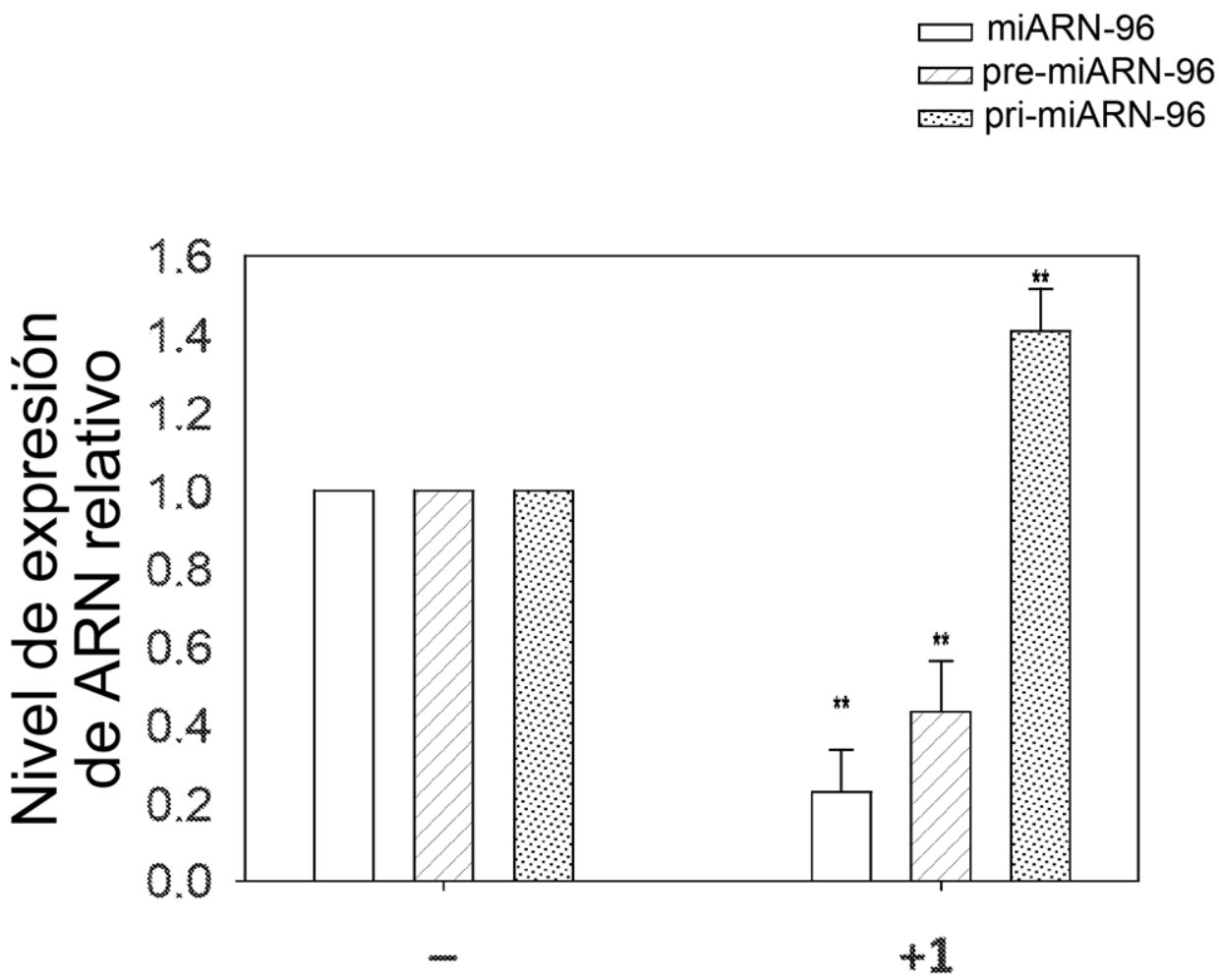

miARN-96
pre-miARN-96
pri-miARN-96
Nivel de expresión de ARN relativo
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
**
**
**
–
+1

Fig. 9C

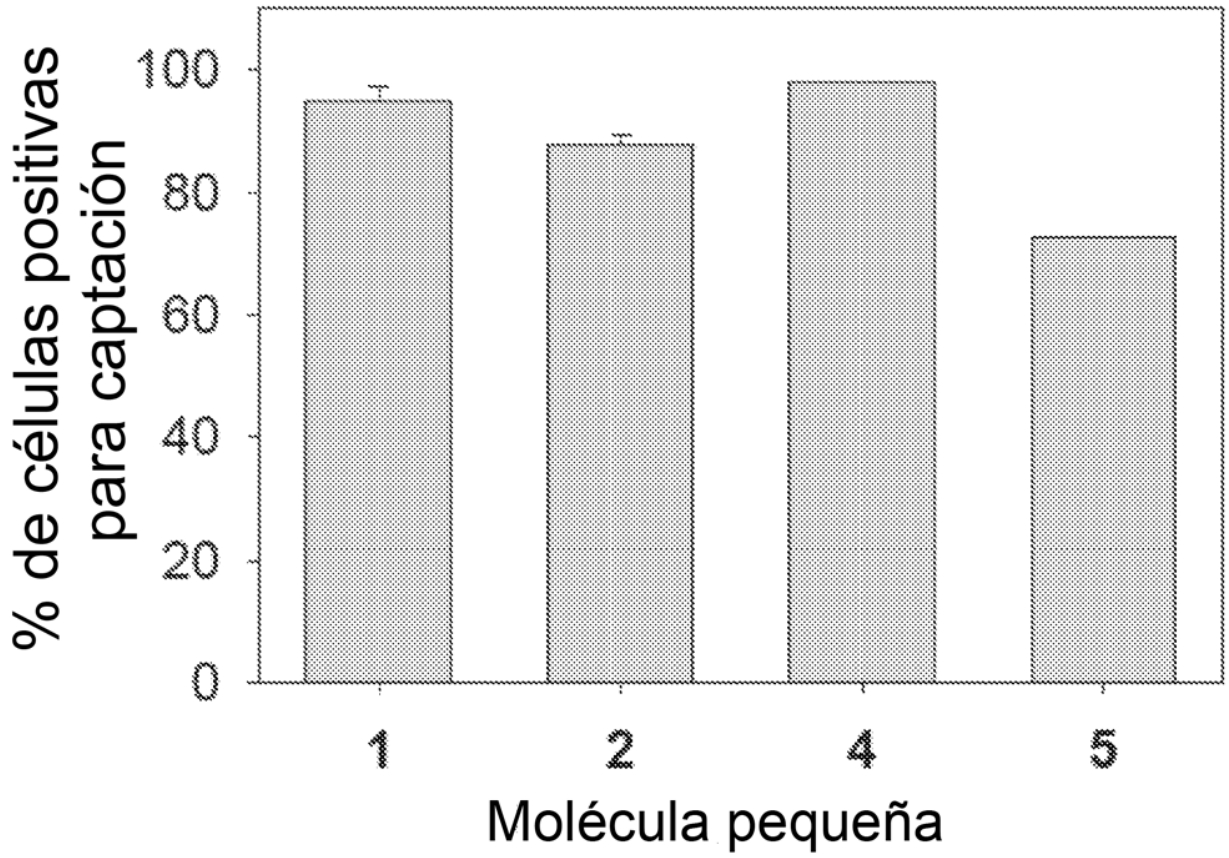
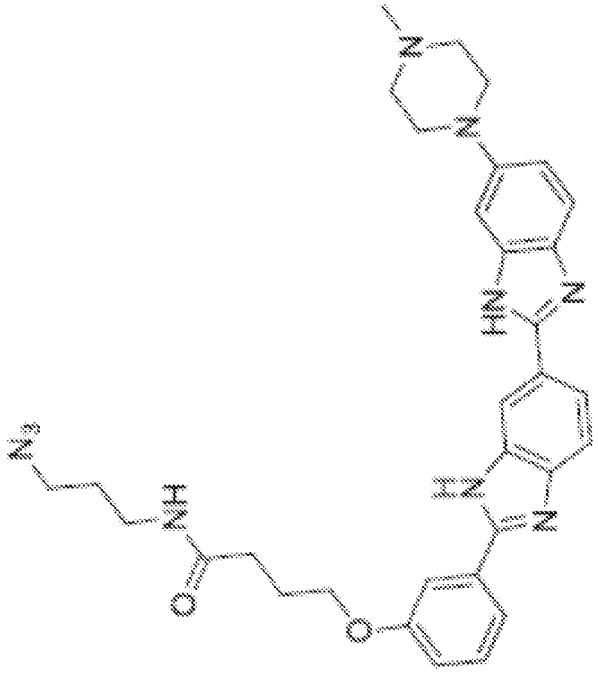
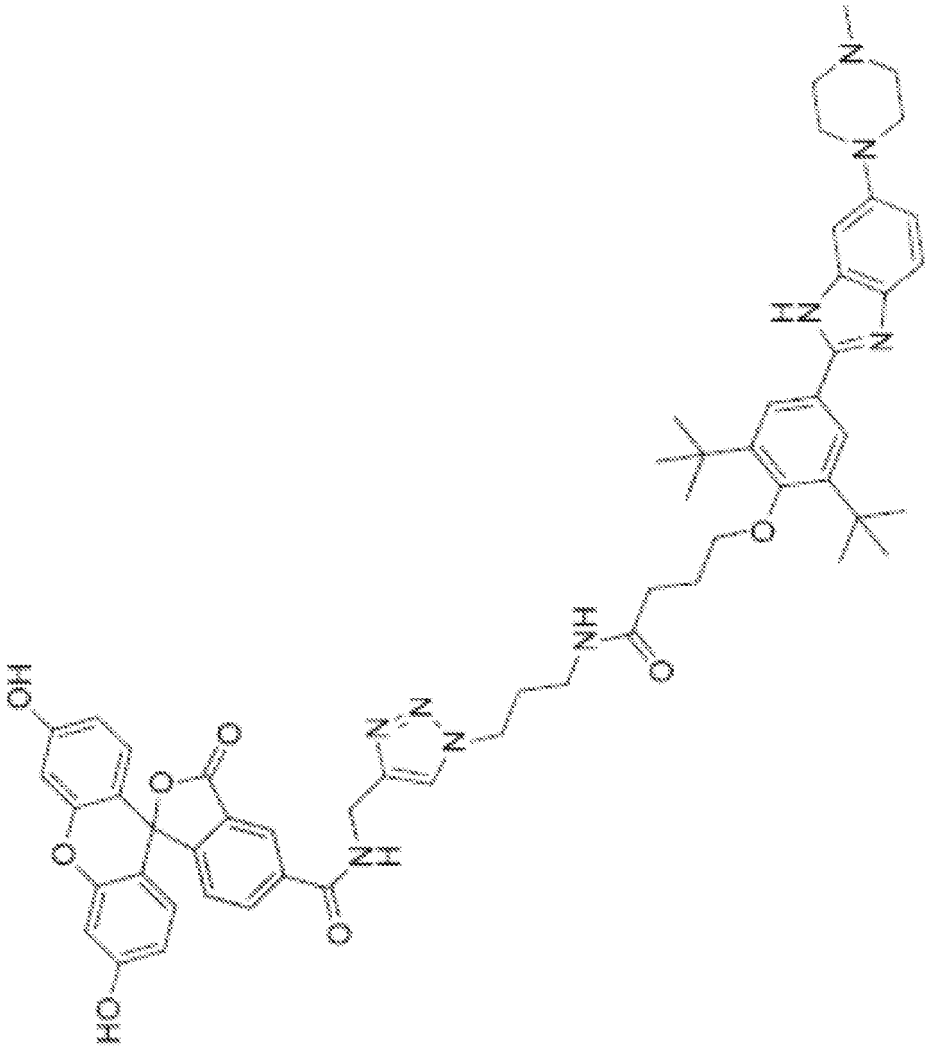

% de células positivas para captación
100
80
60
40
20
0
1
2
4
5
Molécula pequeña

Fig. 10A

Fig. 10B

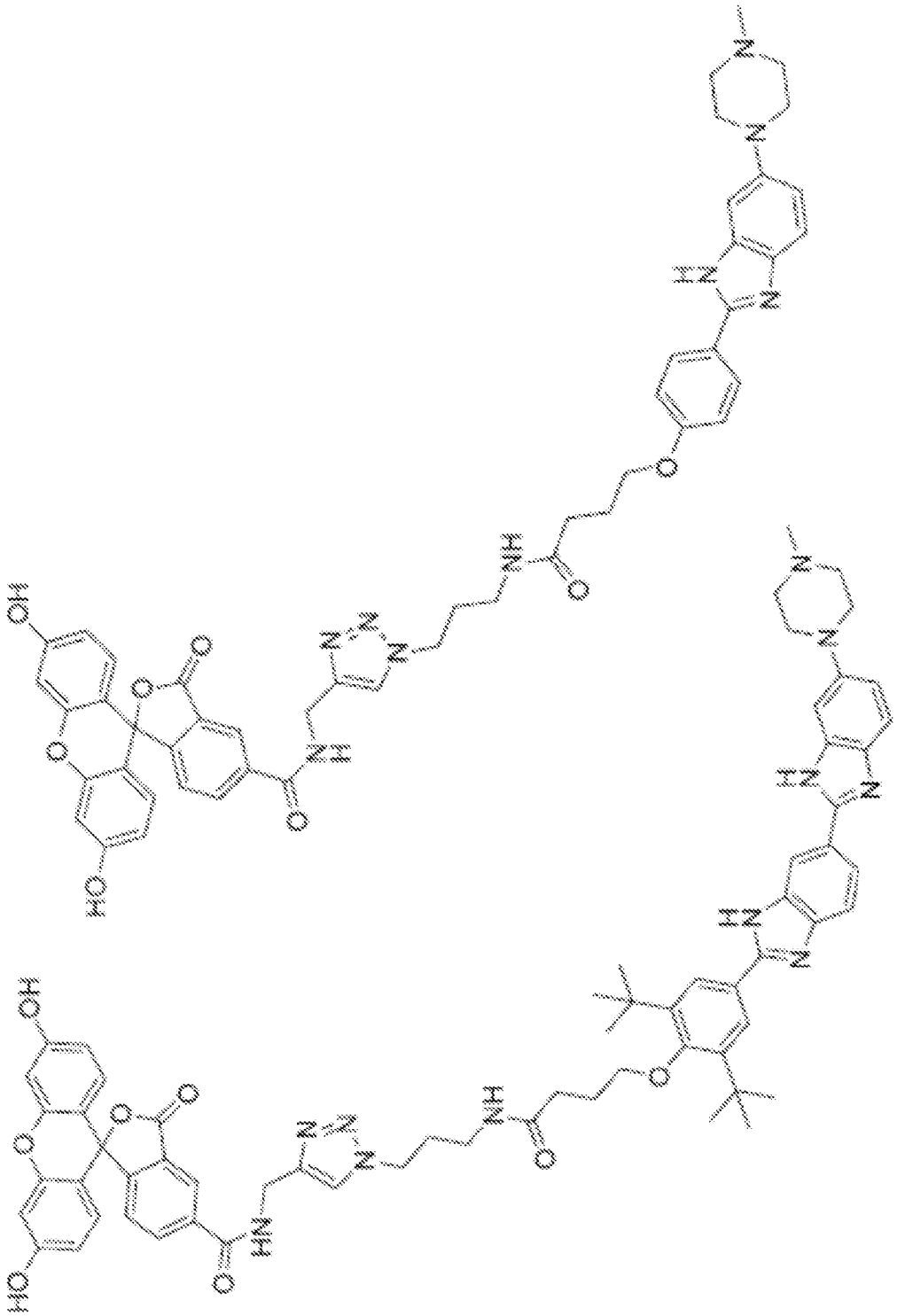


2



1-FI

Fig. 11A–1



4-FI

5-FI

Fig. 11A–2

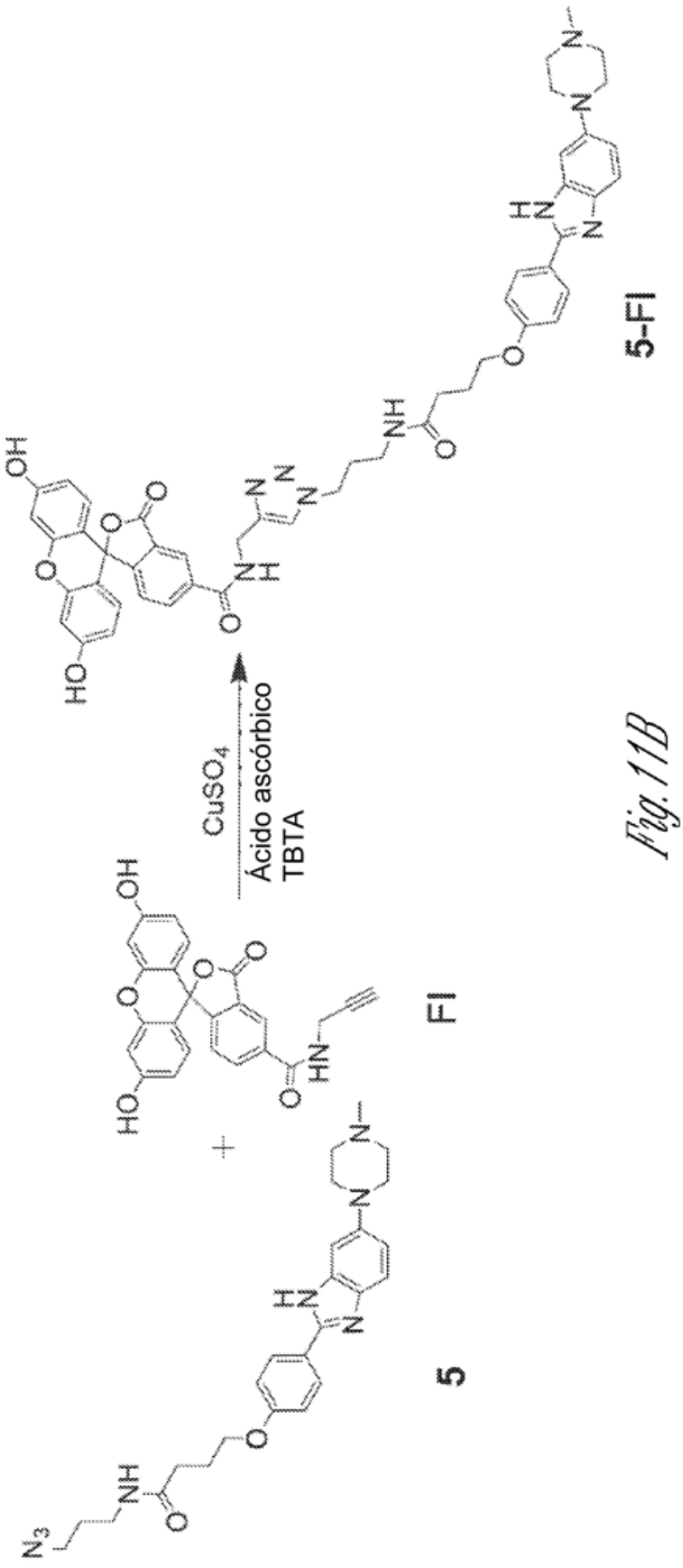


Fig. 11B

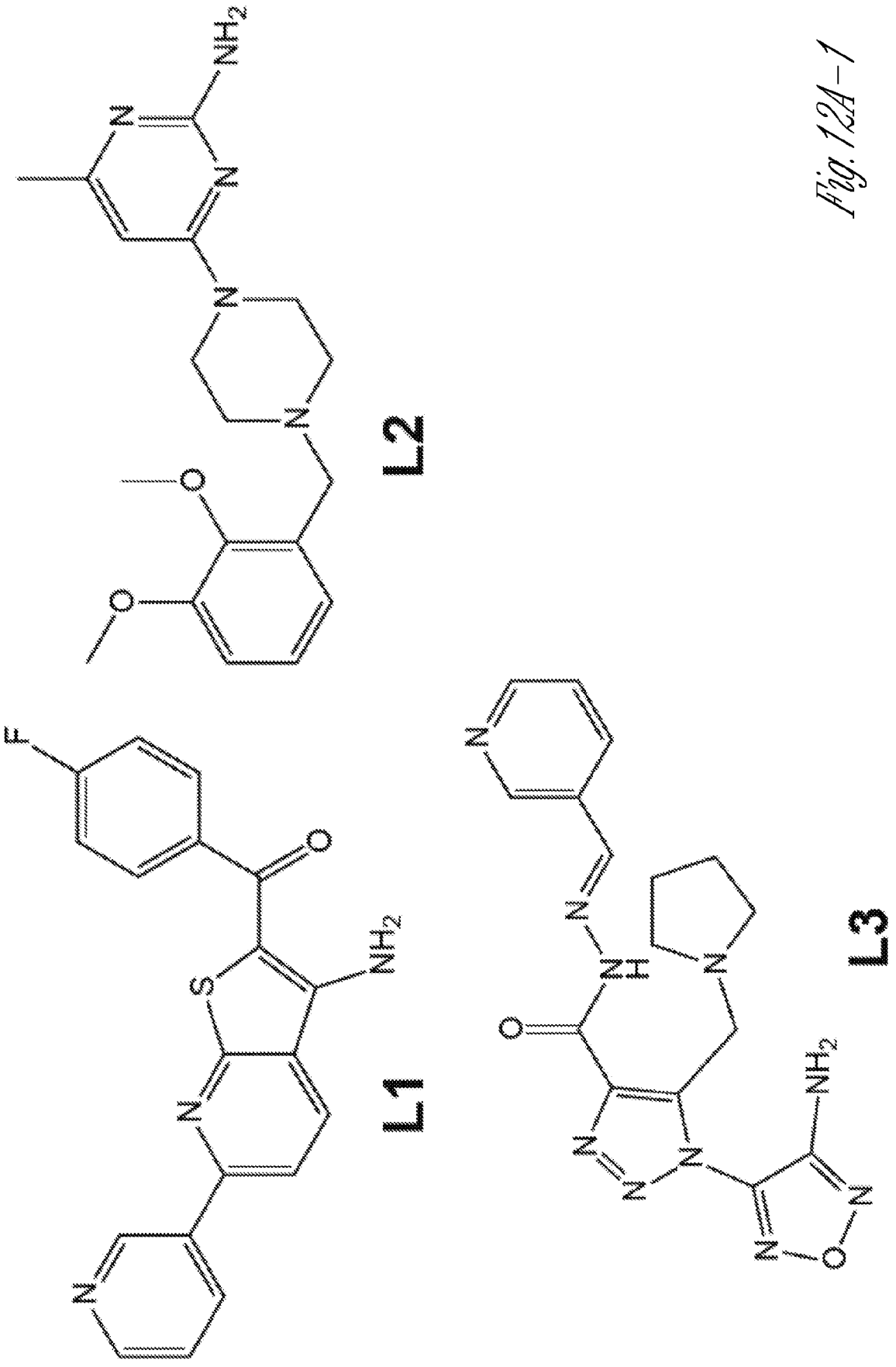


Fig. 12A–1

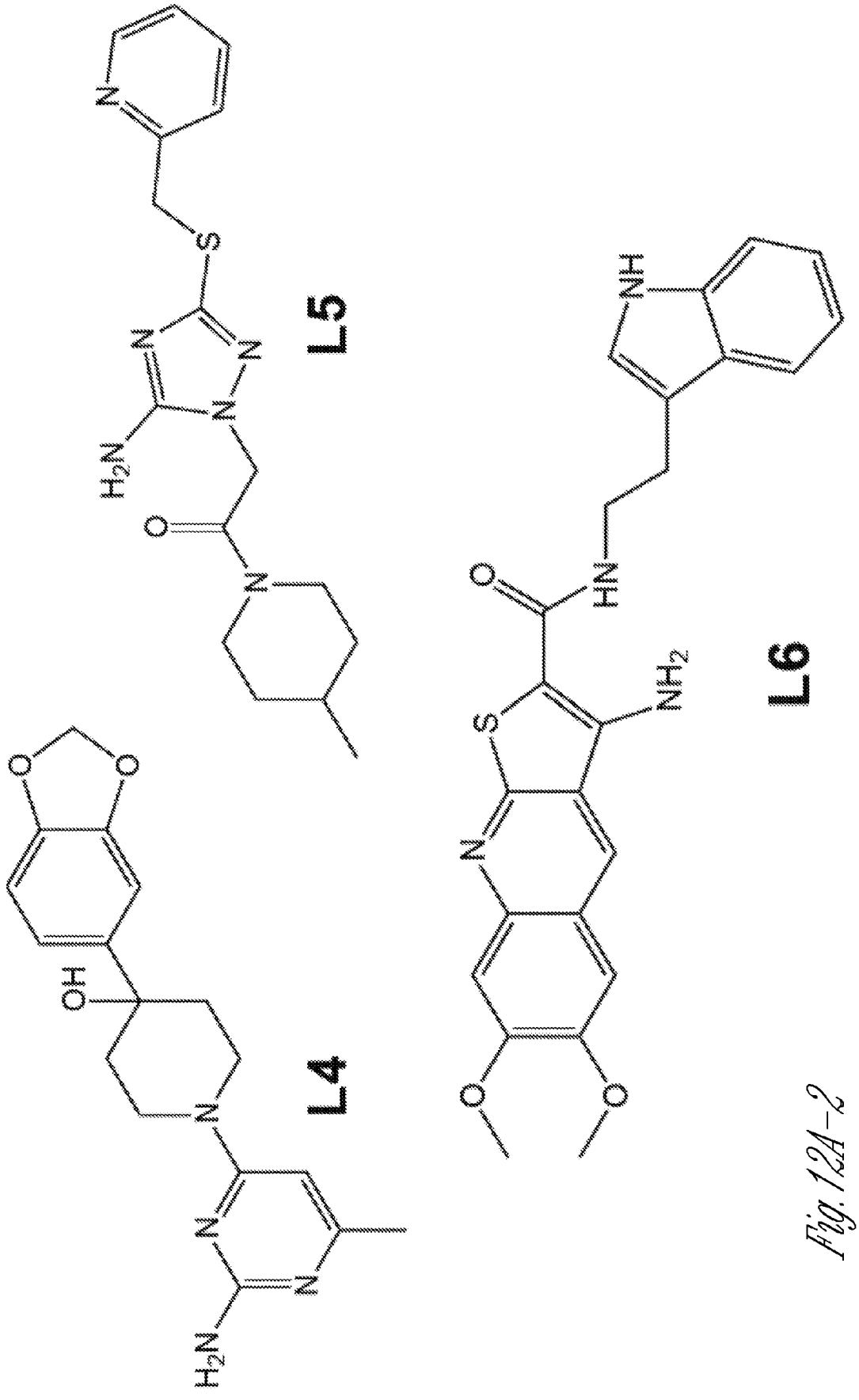


Fig. 12A-2

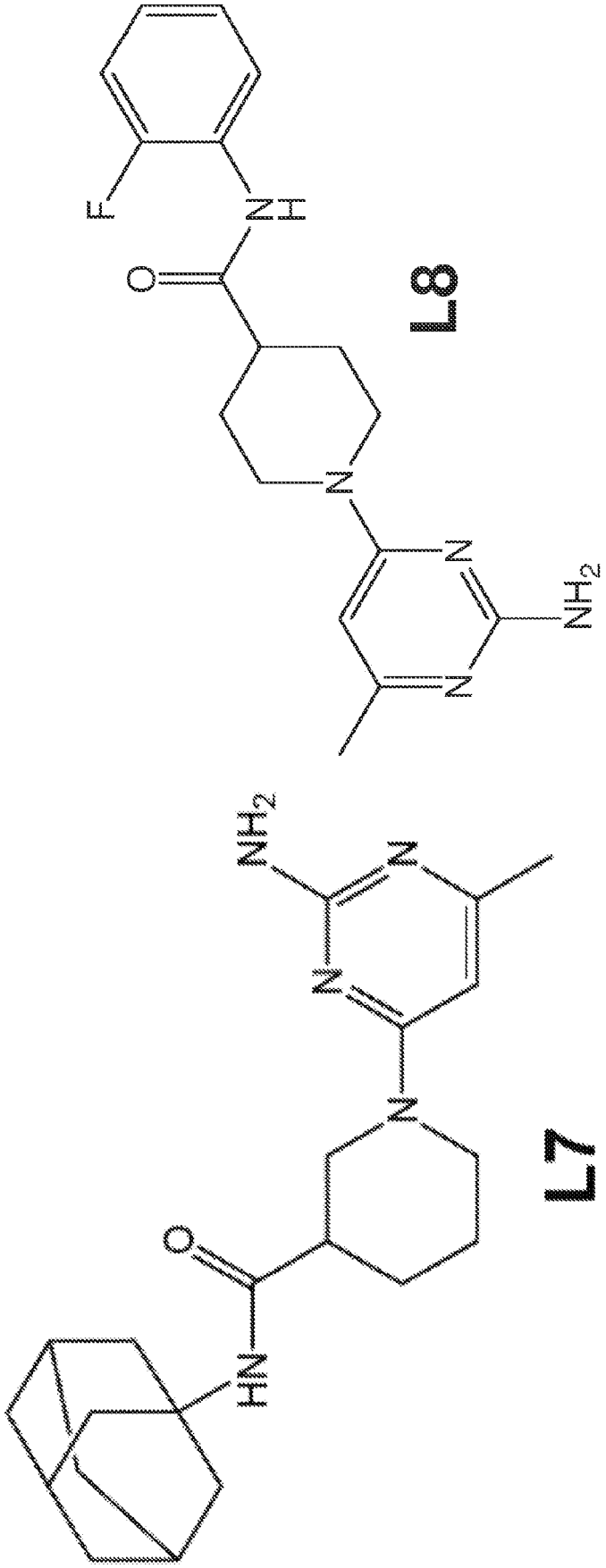


Fig. 12A–3

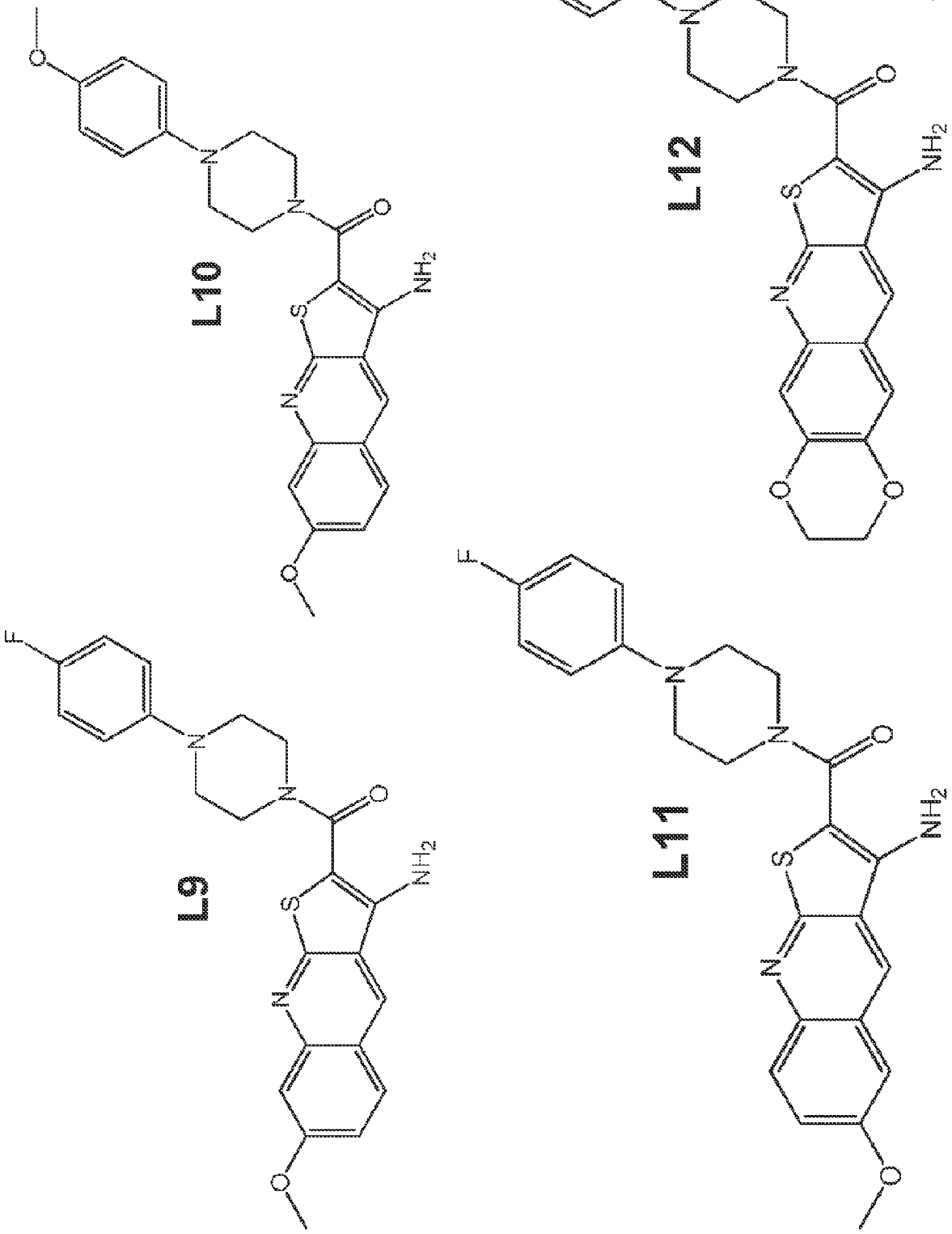


Fig. 12A-4

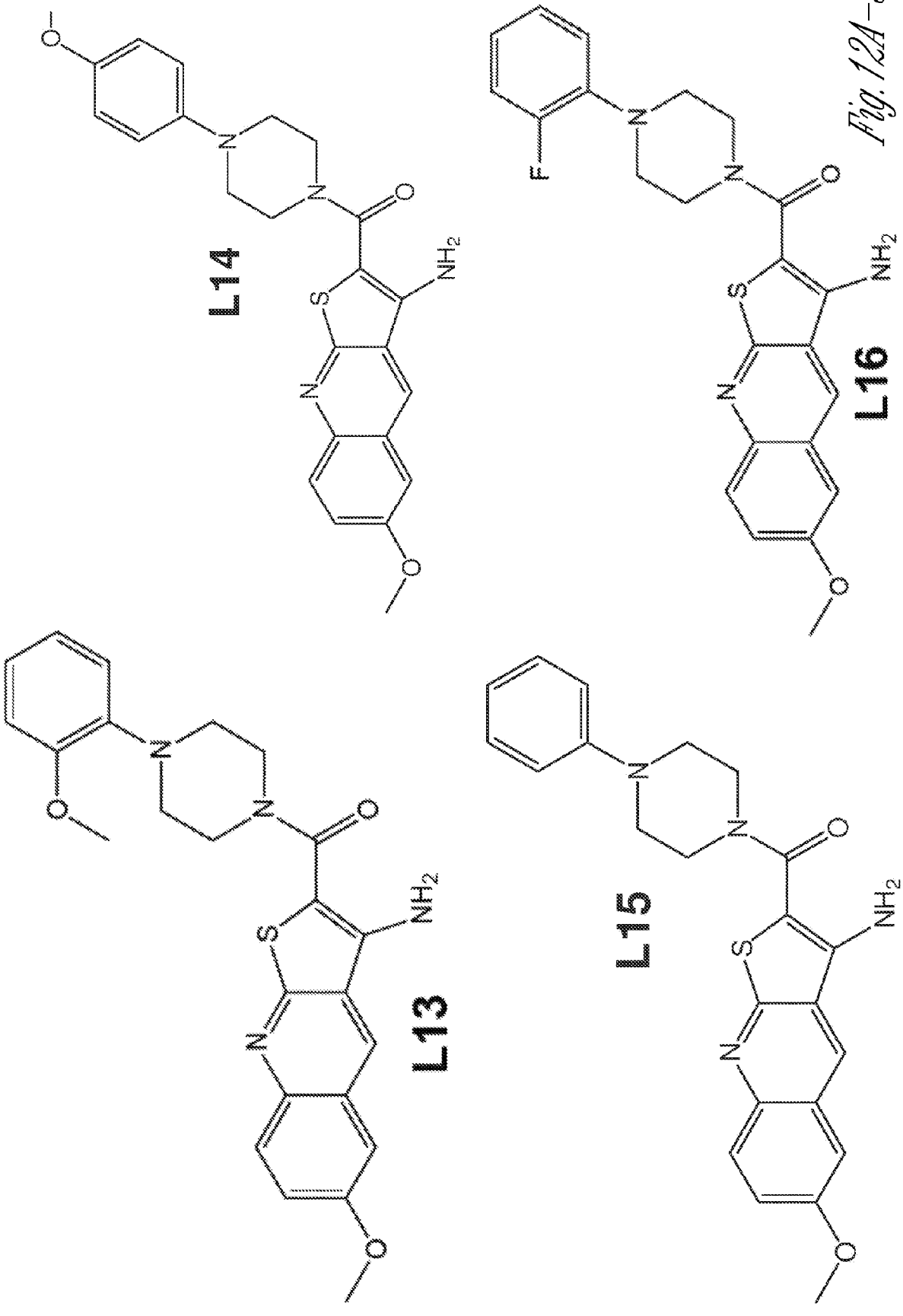


Fig. 12A–5

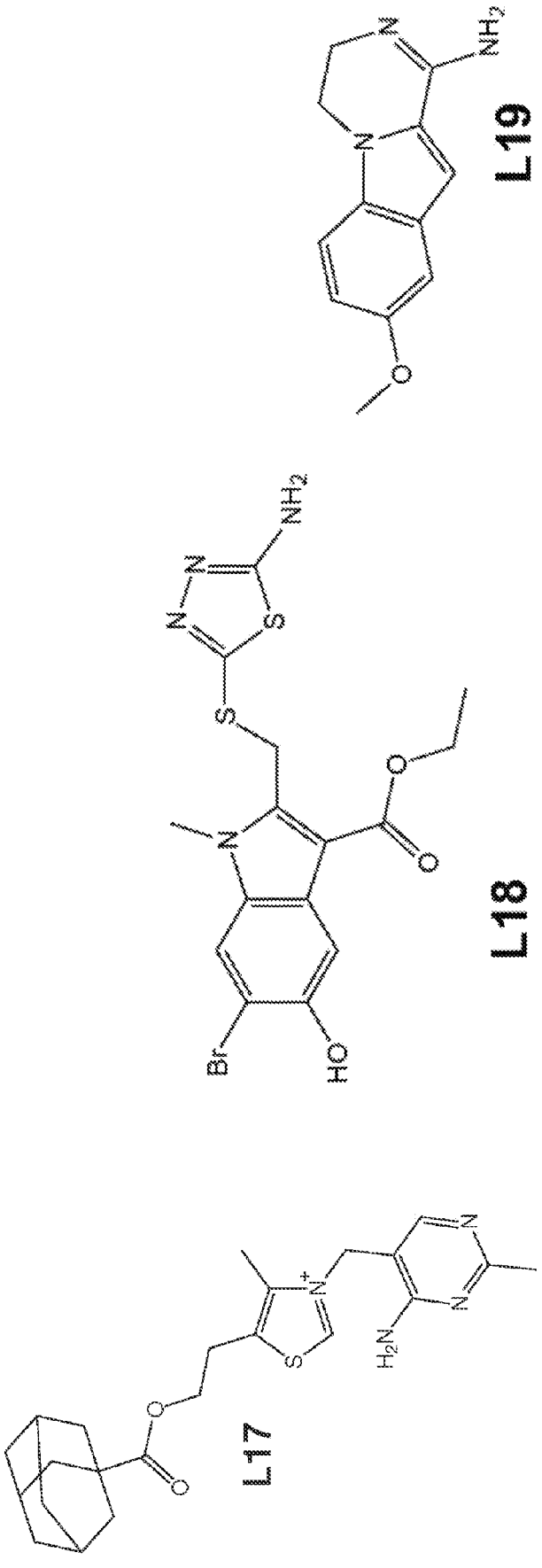


Fig. 12A–6

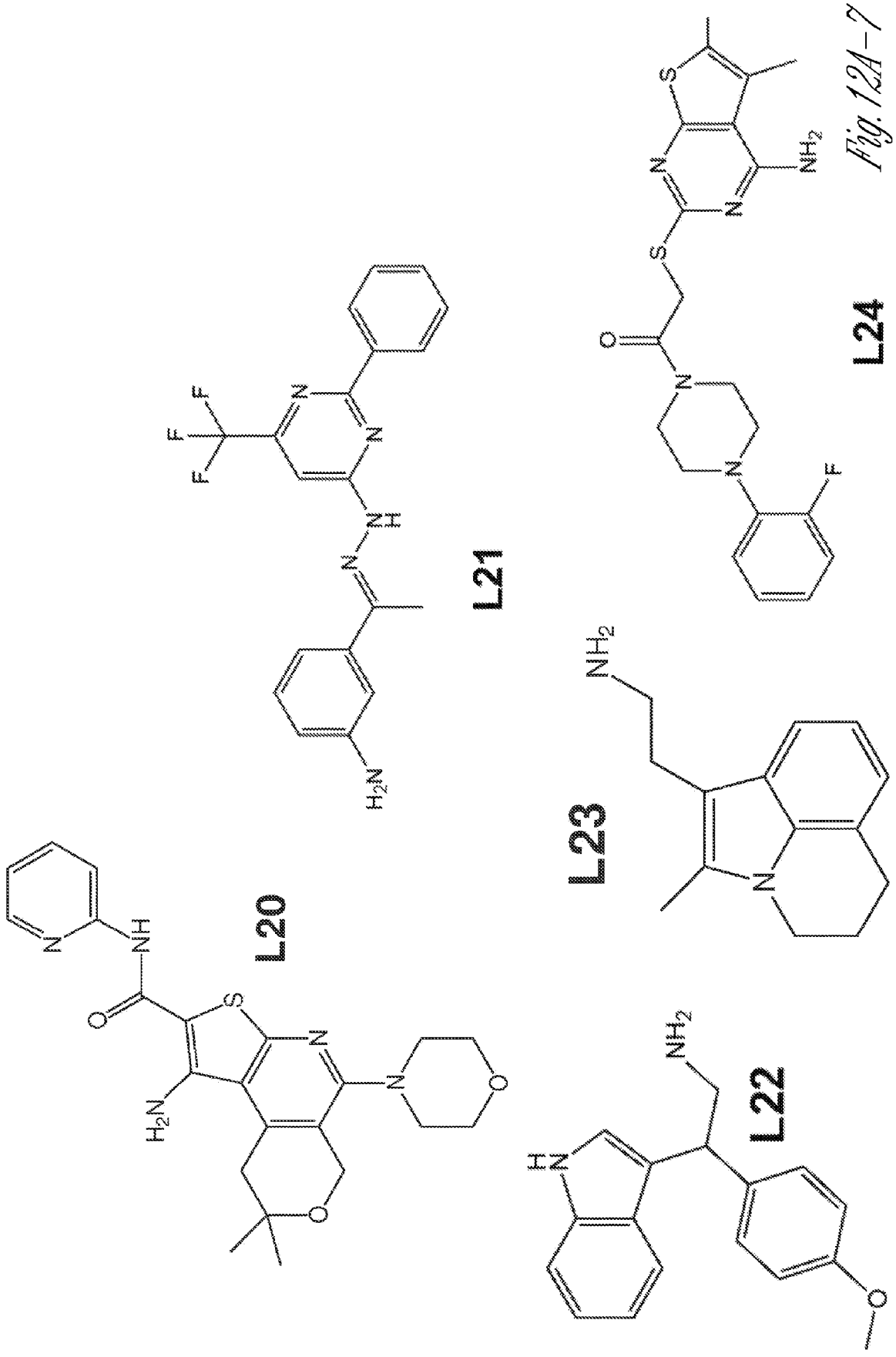


Fig. 12A–7

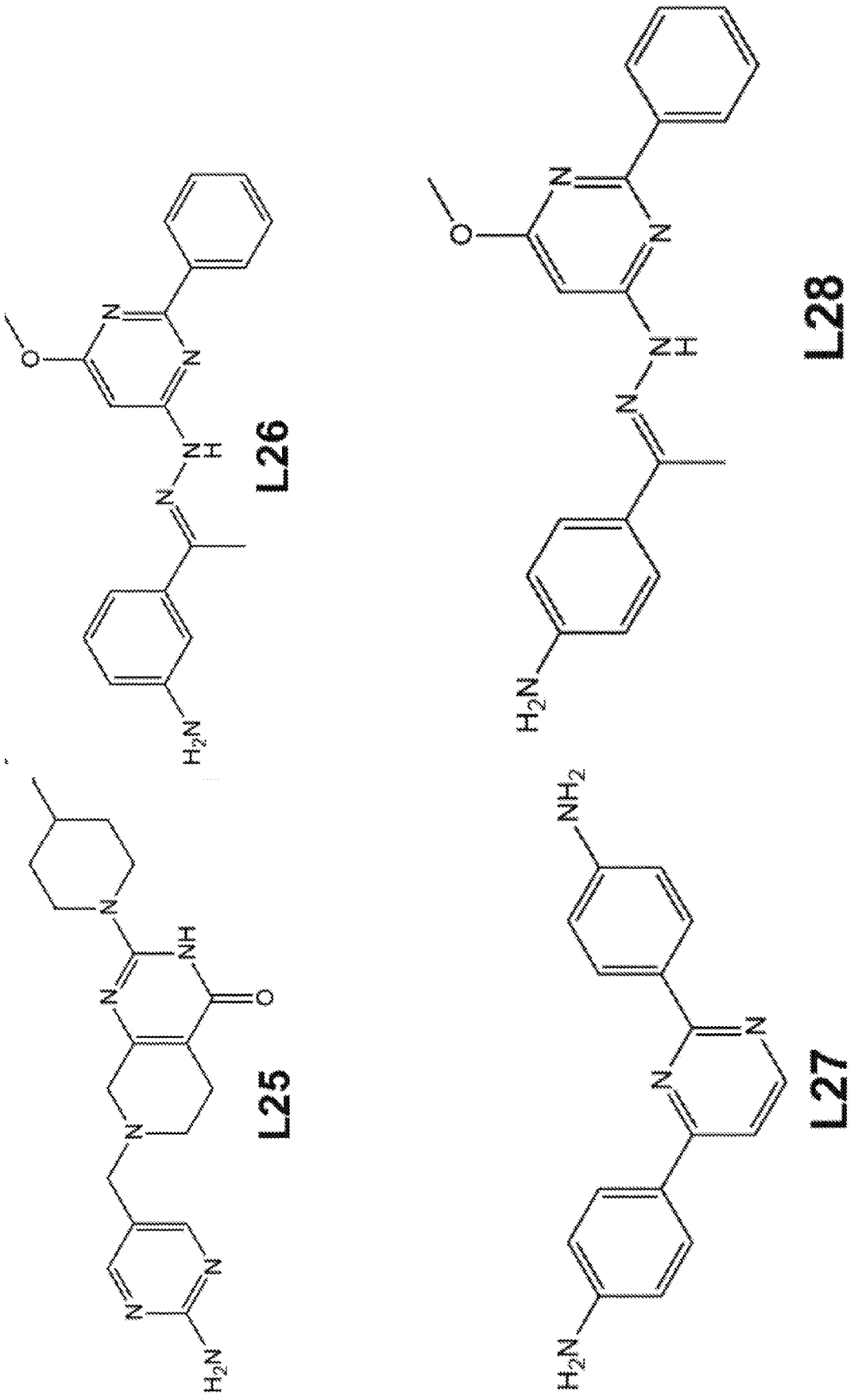


Fig. 12A–8

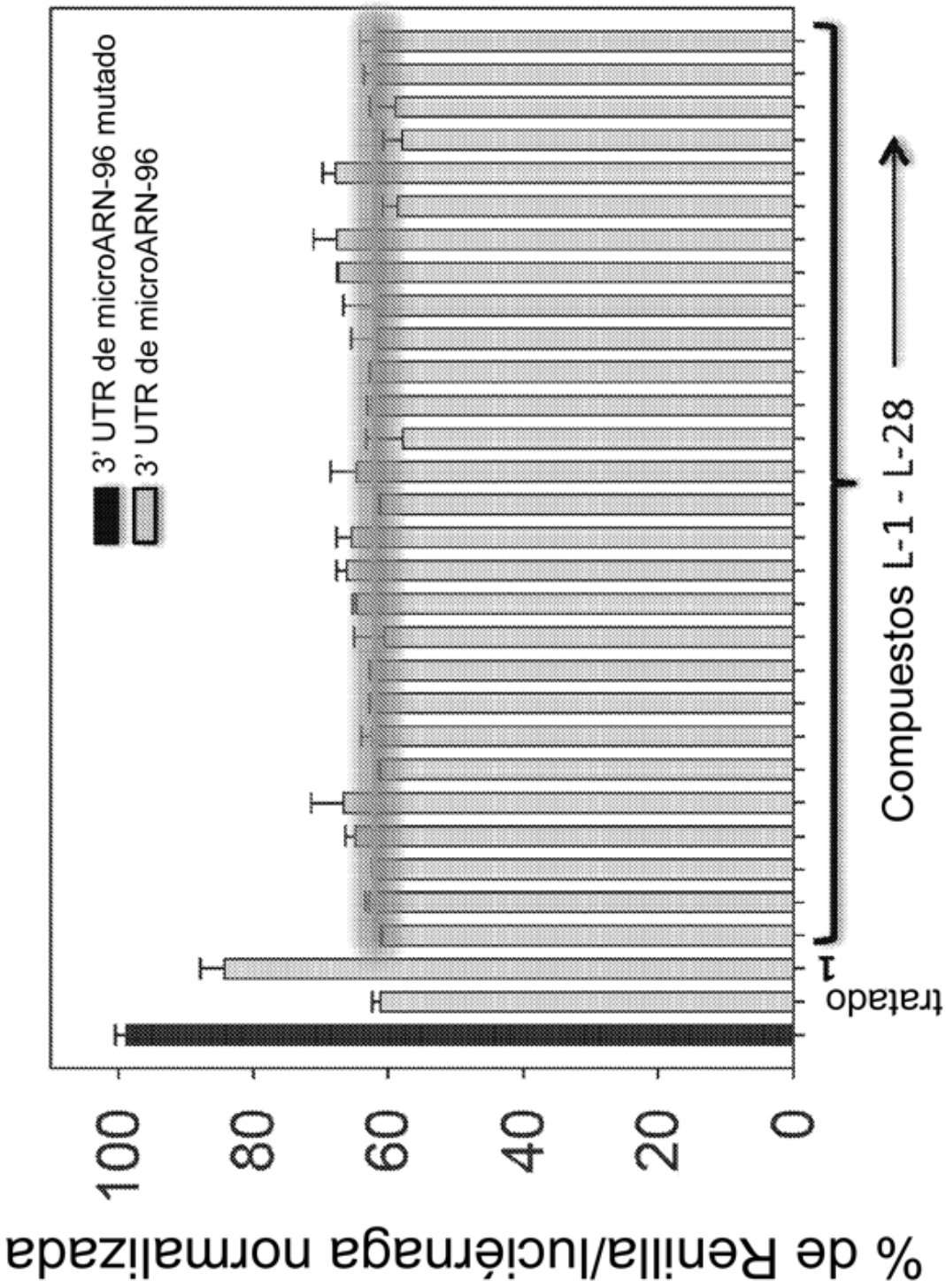

% de Renilla/luciérnaga normalizada
100
80
60
40
20
0
3' UTR de microARN-96 mutado
3' UTR de microARN-96
tratado
1
Compuestos L-1 - L-28

Fig. 12B

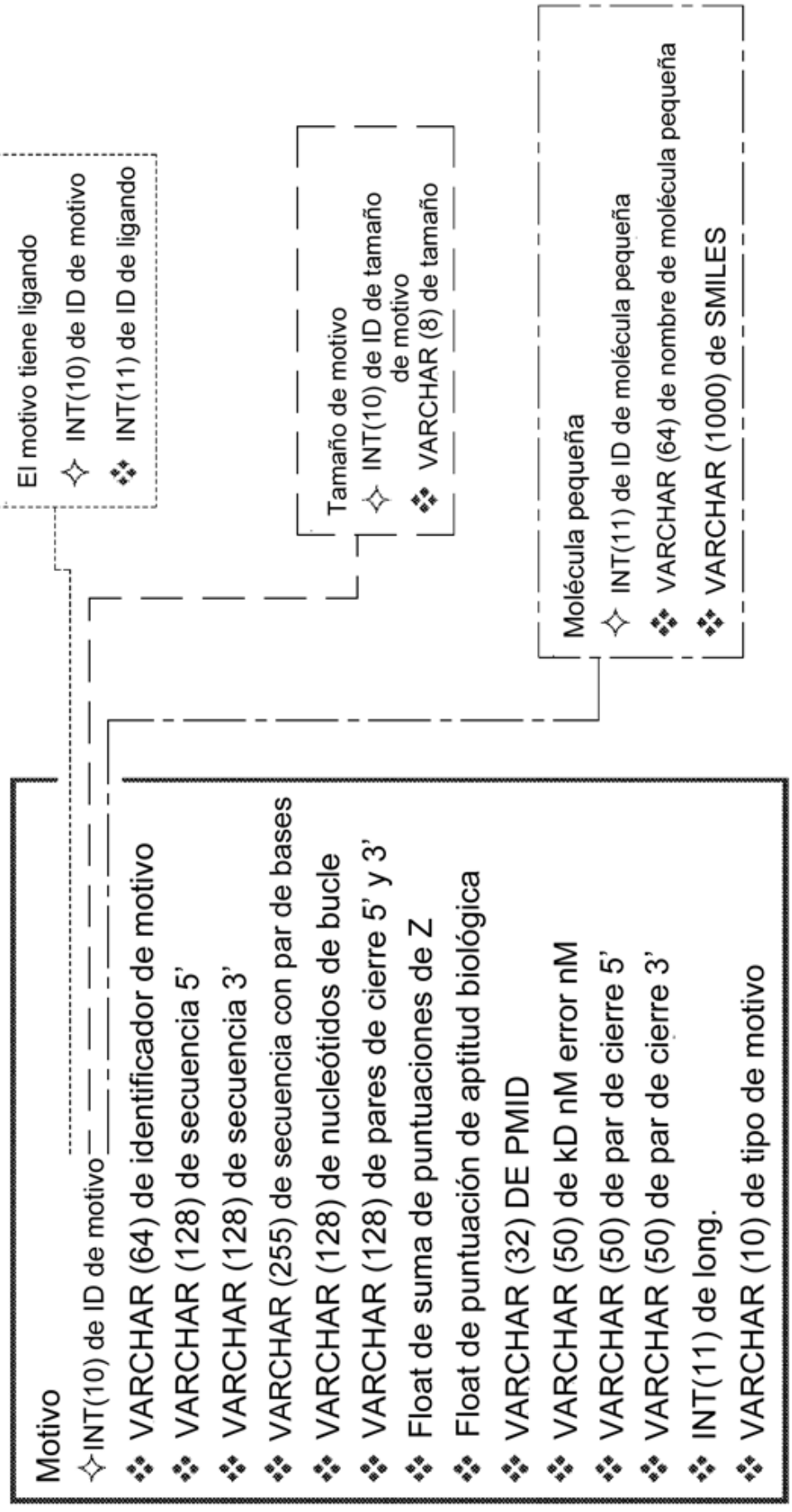

Motivo
INT(10) de ID de motivo
VARCHAR (64) de identificador de motivo
VARCHAR (128) de secuencia 5'
VARCHAR (128) de secuencia 3'
VARCHAR (255) de secuencia con par de bases
VARCHAR (128) de nucleótidos de bucle
VARCHAR (128) de pares de cierre 5' y 3'
Float de suma de puntuaciones de Z
Float de puntuación de aptitud biológica
VARCHAR (32) DE PMID
VARCHAR (50) de kD nM error nM
VARCHAR (50) de par de cierre 5'
VARCHAR (50) de par de cierre 3'
INT(11) de long.
VARCHAR (10) de tipo de motivo
El motivo tiene ligando
INT(10) de ID de motivo
INT(11) de ID de ligando
Tamaño de motivo
INT(10) de ID de tamaño de motivo
VARCHAR (8) de tamaño
Molécula pequeña
INT(11) de ID de molécula pequeña
VARCHAR (64) de nombre de molécula pequeña
VARCHAR (1000) de SMILES

Fig. 13A

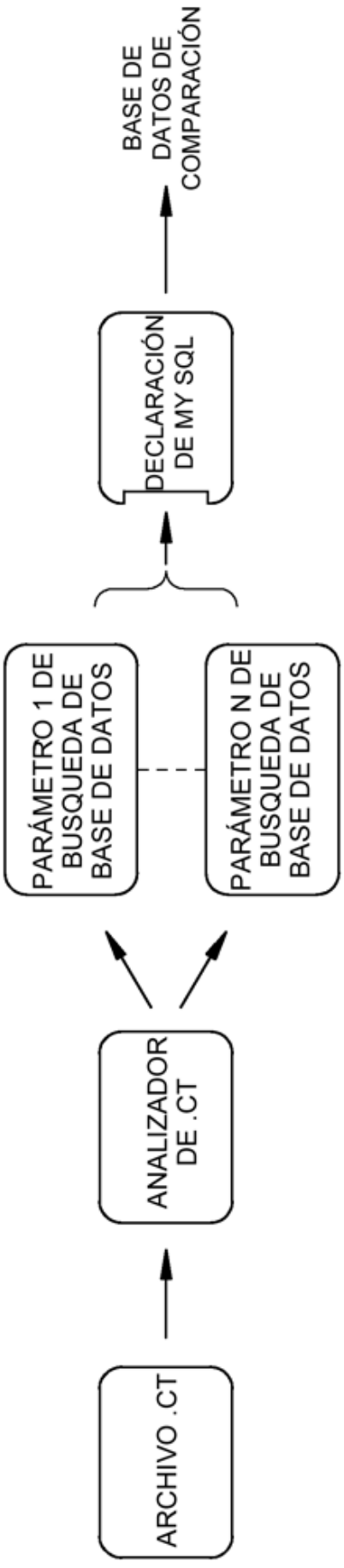

ARCHIVO .CT
ANALIZADOR DE .CT
PARÁMETRO 1 DE BUSQUEDA DE BASE DE DATOS
PARÁMETRO N DE BUSQUEDA DE BASE DE DATOS
DECLARACIÓN DE MY SQL
BASE DE DATOS DE COMPARACIÓN

Fig. 13B

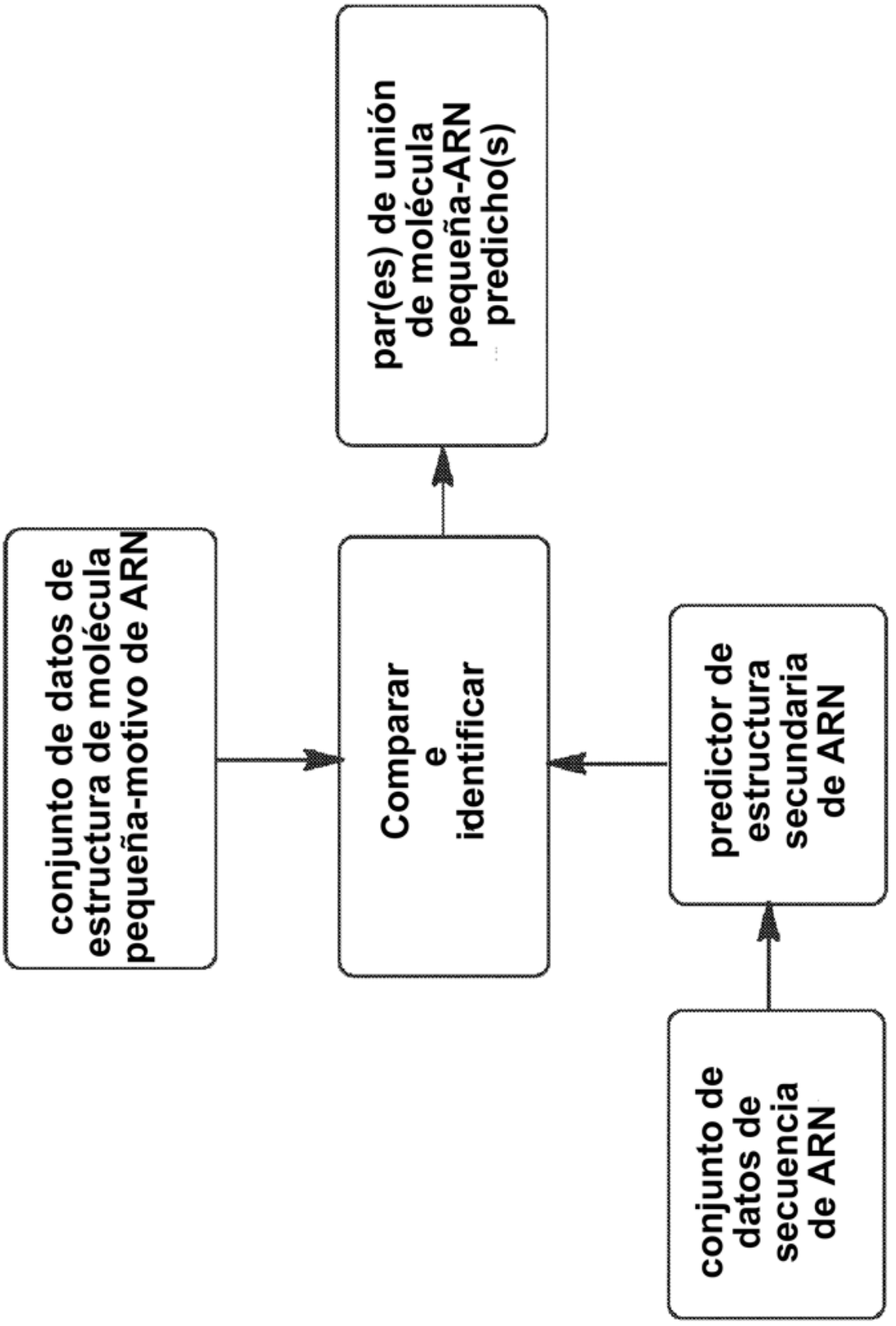

conjunto de datos de estructura de molécula pequeña-motivo de ARN
Comparar e identificar
par(es) de unión de molécula pequeña-ARN predicho(s)
predictor de estructura secundaria de ARN
conjunto de datos de secuencia de ARN

Fig. 14A

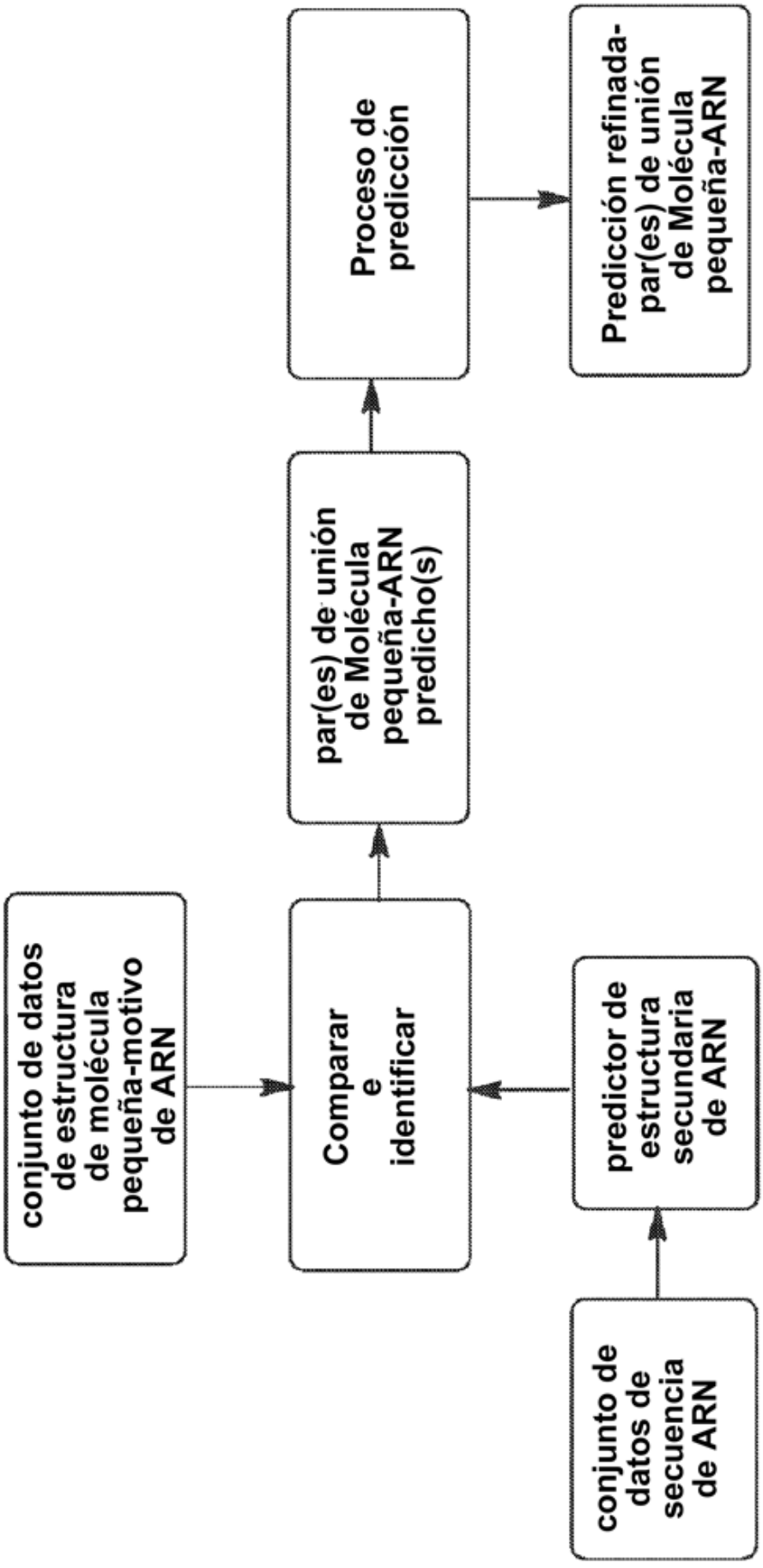

conjunto de datos de estructura de molécula pequeña-motivo de ARN
Comparar e identificar
predictor de estructura secundaria de ARN
conjunto de datos de secuencia de ARN
par(es) de unión de Molécula pequeña-ARN predicho(s)
Proceso de predicción
Predicción refinada-par(es) de unión de Molécula pequeña-ARN

Fig. 14B

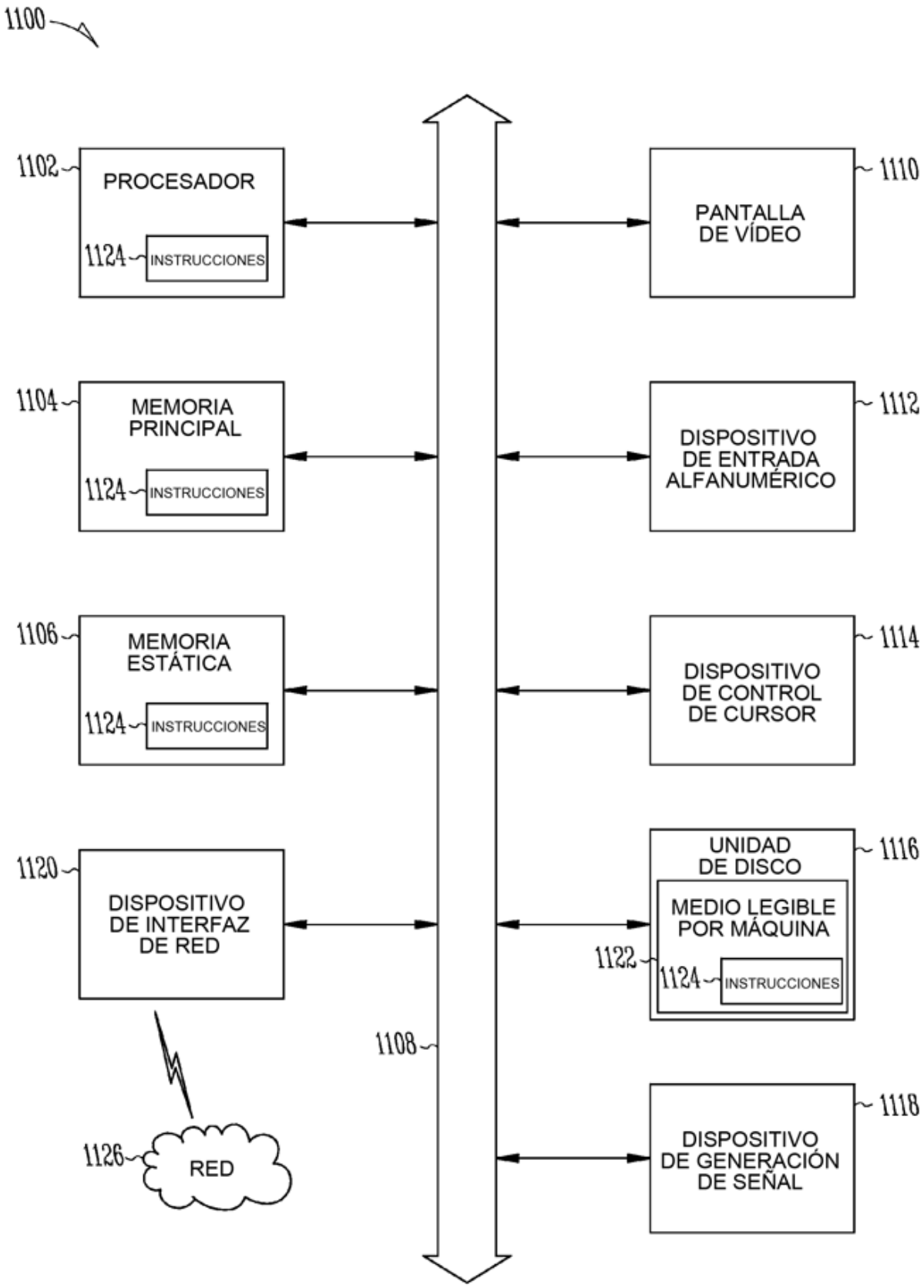

1100
1102
PROCESADOR
1124
INSTRUCCIONES
1104
MEMORIA PRINCIPAL
1124
INSTRUCCIONES
1106
MEMORIA ESTÁTICA
1124
INSTRUCCIONES
1120
DISPOSITIVO DE INTERFAZ DE RED
1126
RED
1108
1110
PANTALLA DE VÍDEO
1112
DISPOSITIVO DE ENTRADA ALFANUMÉRICO
1114
DISPOSITIVO DE CONTROL DE CURSOR
1116
UNIDAD DE DISCO
MEDIO LEGIBLE POR MÁQUINA
1122
1124
INSTRUCCIONES
1118
DISPOSITIVO DE GENERACIÓN DE SEÑAL

Fig. 15

AGENTE DE
UNIÓN A BUCLE
GG 1X1
SÍNTESIS DE DÍMEROS
DE DISEÑADOR
BUSQUEDA DE INFORNA
PARA UNIÓN "IMPERFECTA"
A OTROS SITIOS

Fig. 16A

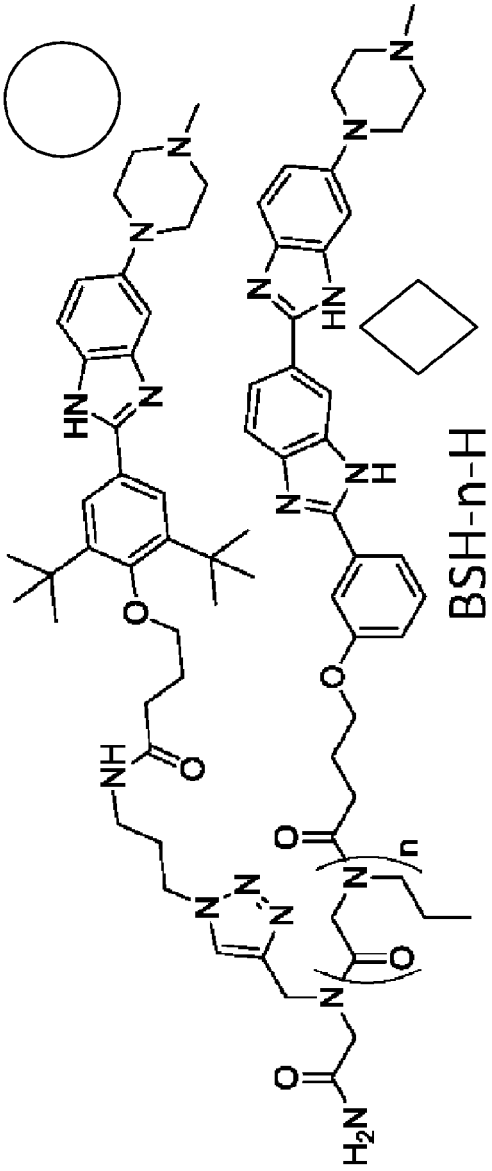


Fig. 16B

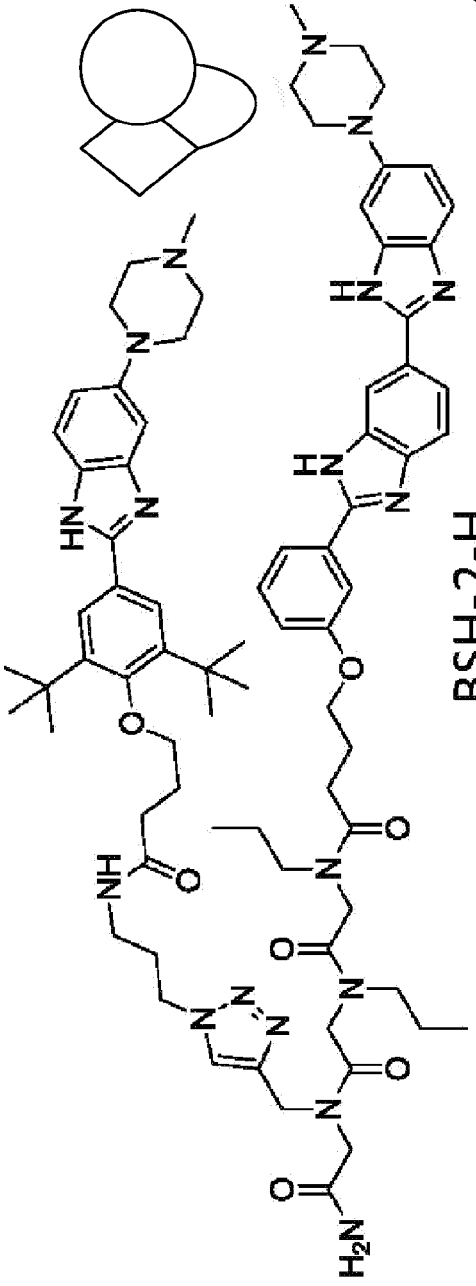


Fig. 16C

ARN de casete
(C1)
5'-GGGAGA
G-C
GCAAGG
5'UUU/3'AUA
5'-GGGAGA
G-C
GCAAGG
5'CGA/3'GGU
5'-GGGAGA
G-C
GCAAGG
5'CGAUU/3'GGUAUA
Horquilla de ADN
(Horquilla H)
5'-C-G

Fig. 17

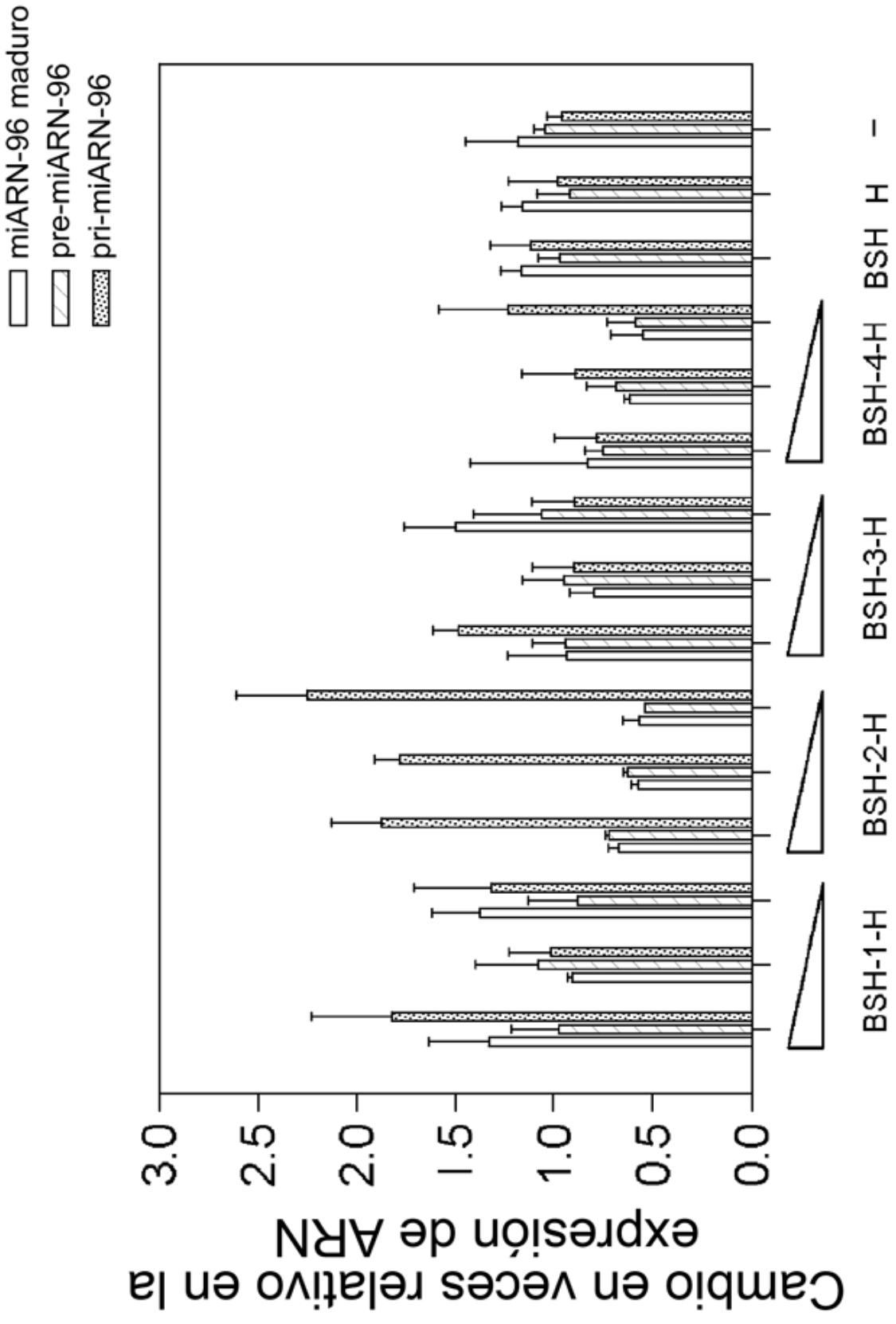

miARN-96 maduro
pre-miARN-96
pri-miARN-96
Cambio en veces relativo en la expresión de ARN
3.0
2.5
2.0
1.5
1.0
0.5
0.0
BSH-1-H
BSH-2-H
BSH-3-H
BSH-4-H
BSH H –

Fig. 18

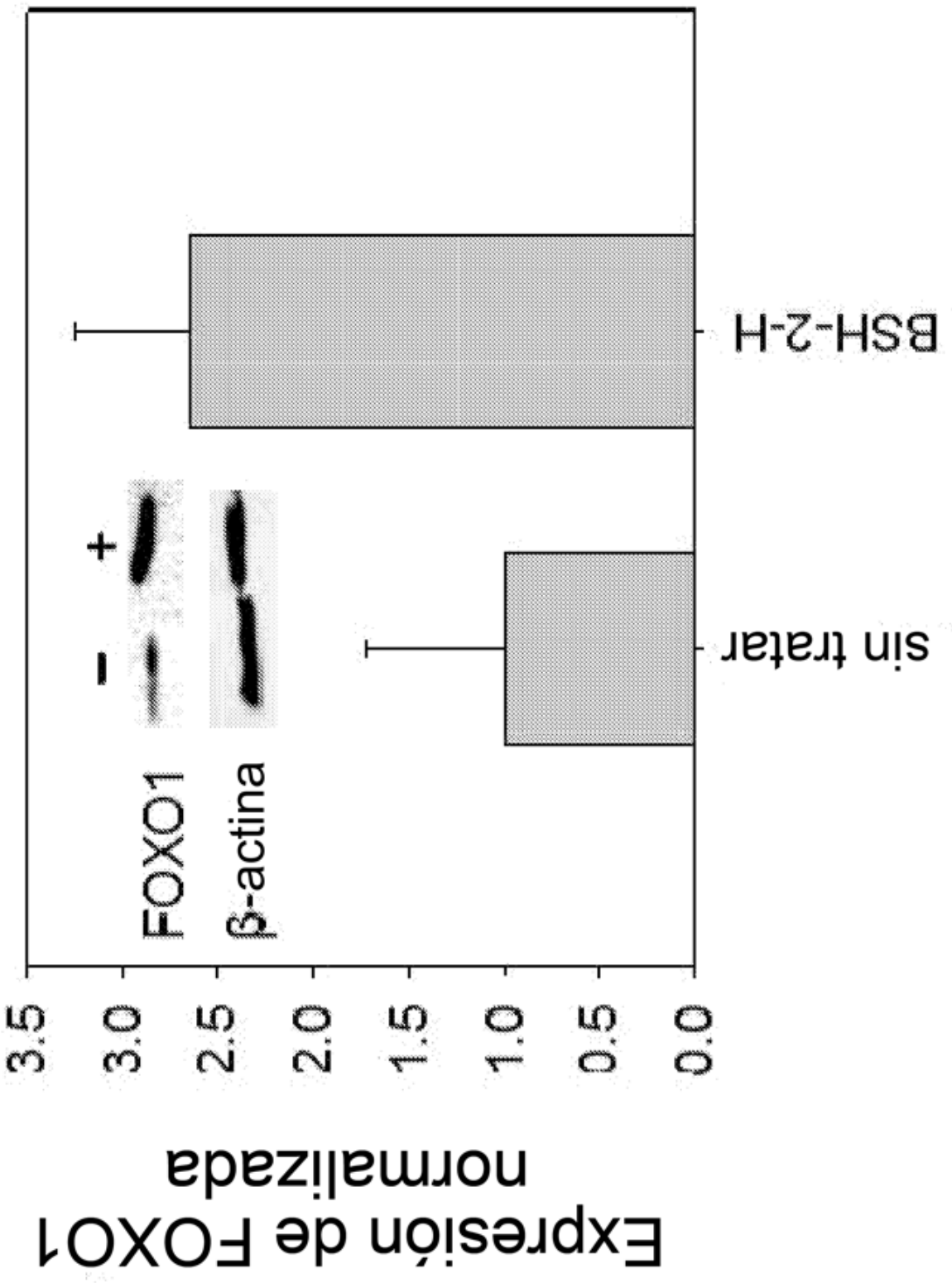

Expresión de FOXO1
normalizada
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
FOXO1
β-actina
+
-
sin tratar
BSH-2-H

Fig. 19

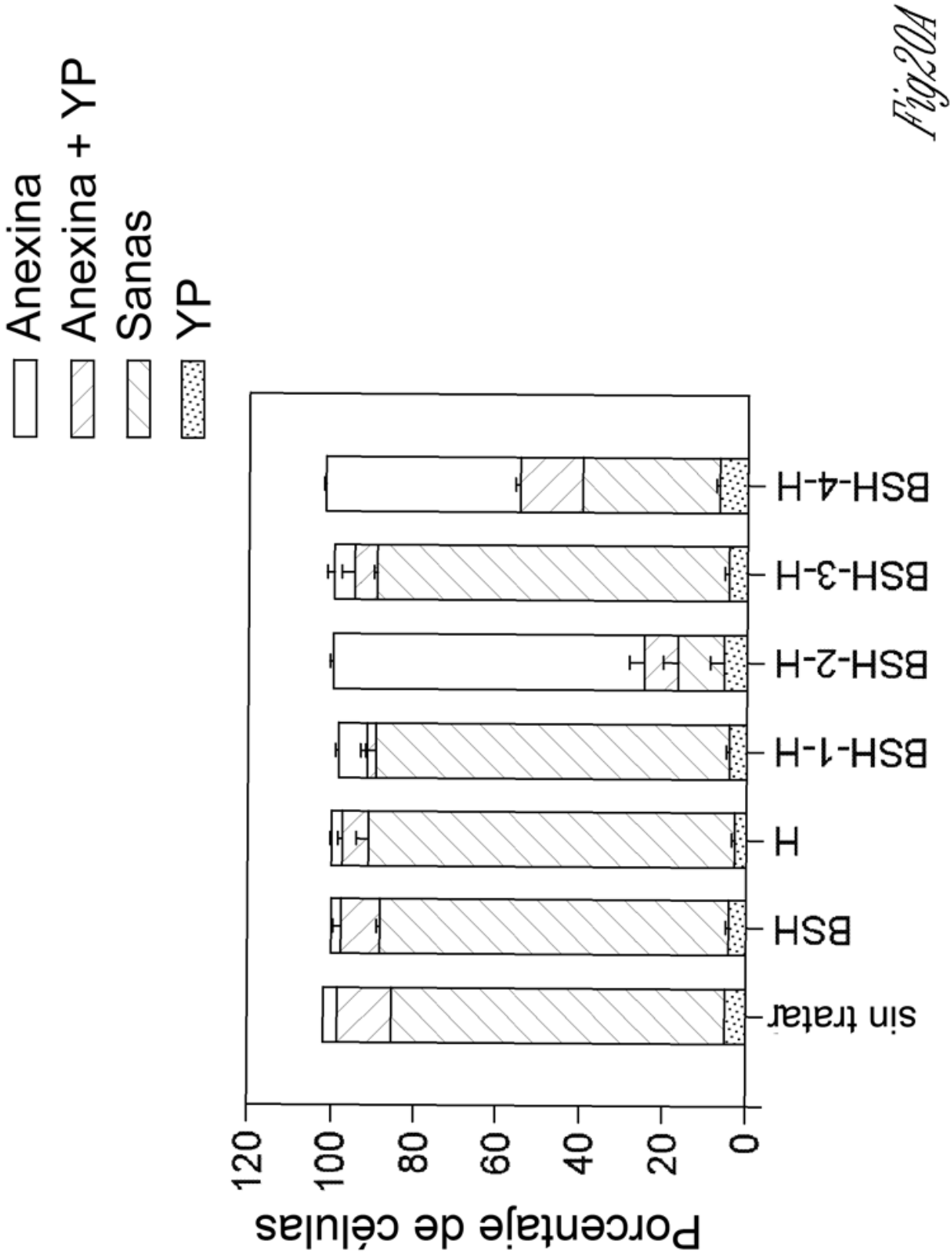

Anexina
Anexina + YP
Sanas
YP
Porcentaje de células
120
100
80
60
40
20
0
sin tratar
BSH
H
BSH-1-H
BSH-2-H
BSH-3-H
BSH-4-H

Fig20A

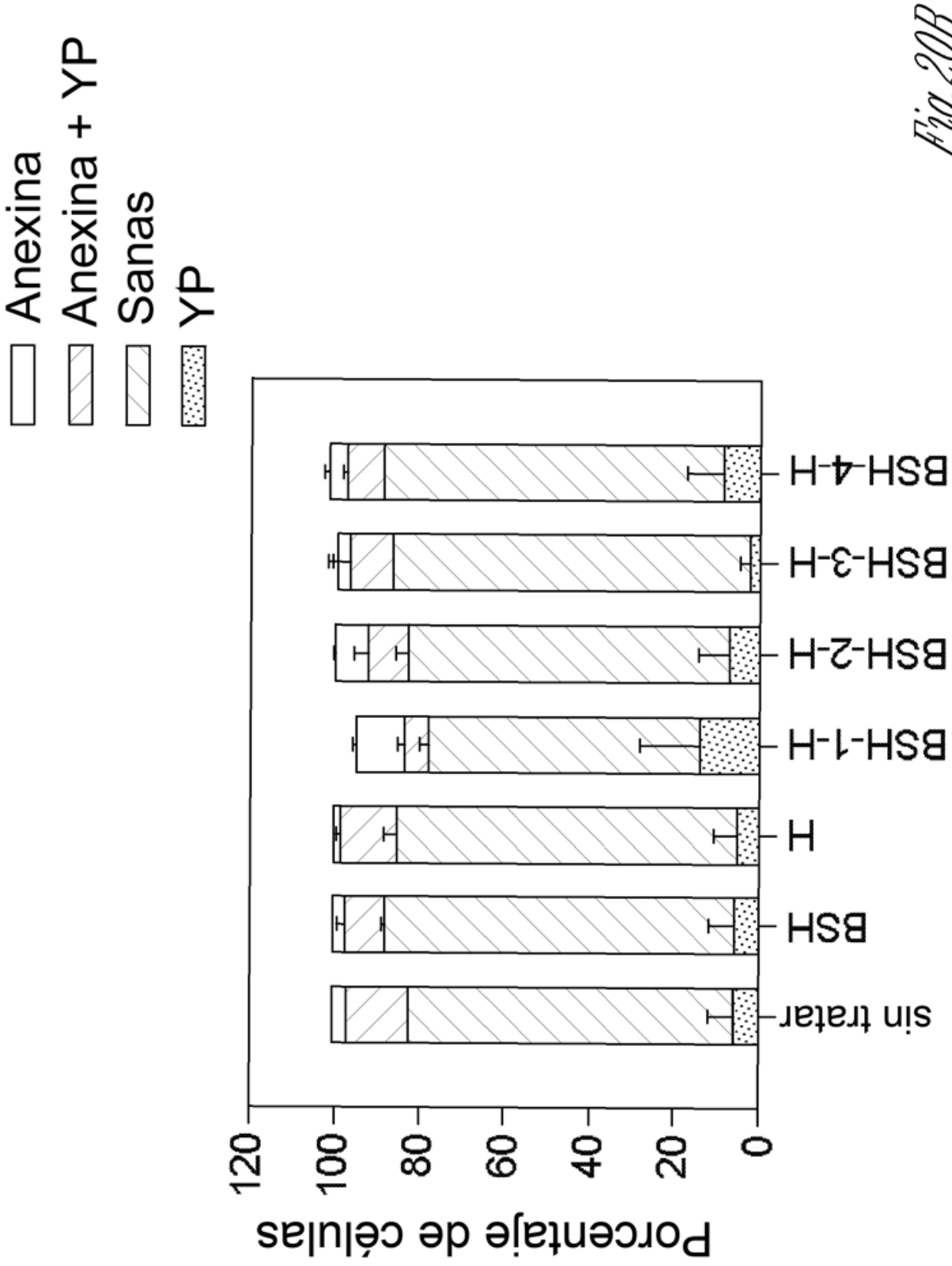

Anexina
Anexina + YP
Sanas
YP
Porcentaje de células
120
100
80
60
40
20
0
sin tratar
BSH
H
BSH-1-H
BSH-2-H
BSH-3-H
BSH-4-H

Fig.20B

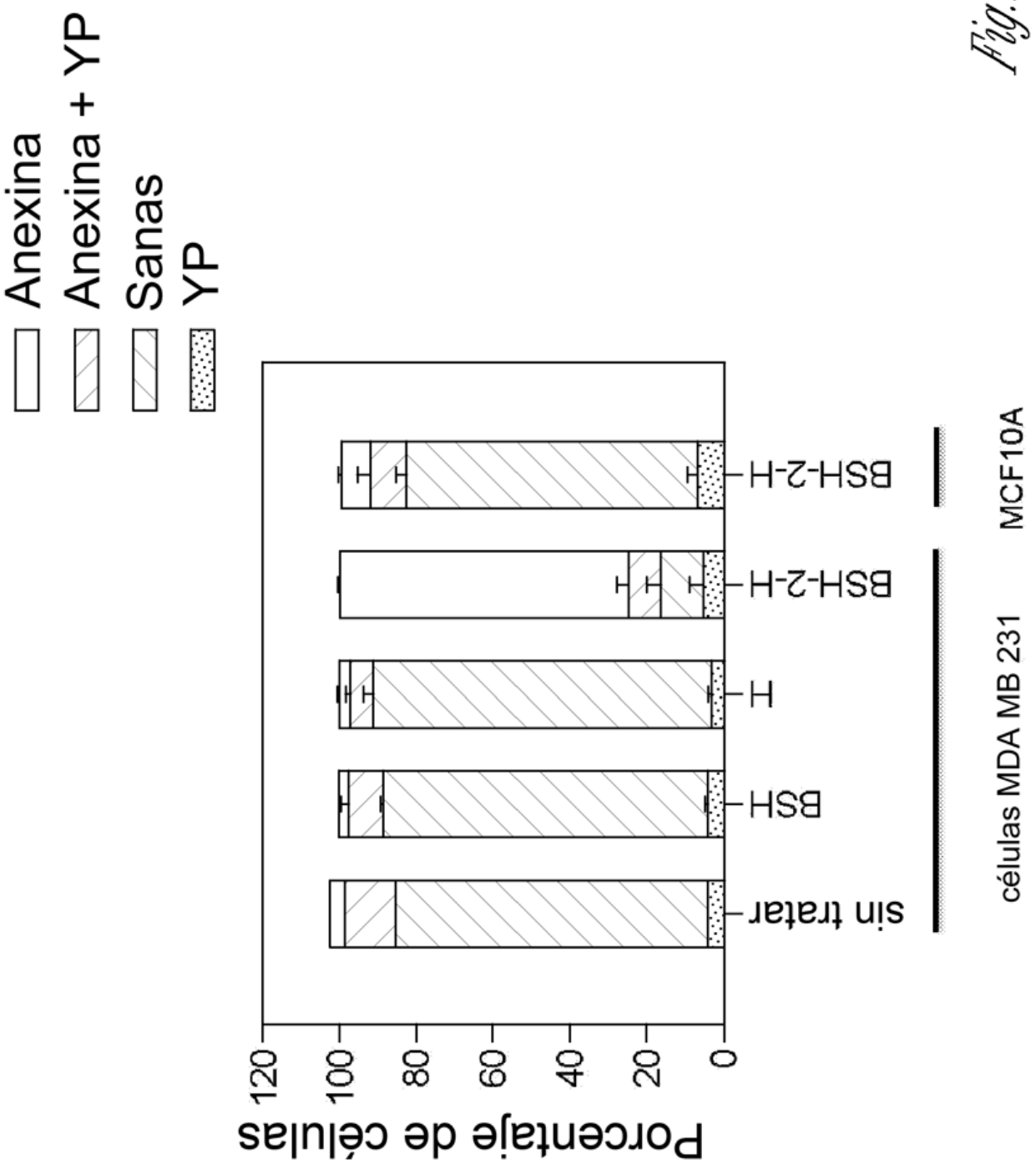

Anexina
Anexina + YP
Sanas
YP
Porcentaje de células
120
100
80
60
40
20
0
sin tratar
BSH
H
BSH-2-H
BSH-2-H
células MDA MB 231
MCF10A

Fig.20C