(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 791 873**

(51) Int. CI.:

C12N 9/12 (2006.01)
C12Q 1/6806 (2008.01)
C12N 15/01 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **07.10.2014 PCT/US2014/059489**

(87) Fecha y número de publicación internacional: **16.04.2015 WO15054247**

(96) Fecha de presentación y número de la solicitud europea: **07.10.2014 E 14851691 (7)**

(97) Fecha y número de publicación de la concesión europea: **08.01.2020 EP 3055413**

(54) Título: **Detección de modificaciones químicas en ácidos nucleicos**

(30) Prioridad:

07.10.2013 US 201361887614 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
06.11.2020

(73) Titular/es:

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL (100.0%)
100 Europa Drive, Suite 430
Chapel Hill, NC 27517, US

(72) Inventor/es:

WEEKS, KEVIN M.;
SIEGFRIED, NATHAN;
HOMAN, PHILIP;
BUSAN, STEVEN y
FAVOROV, OLEG V.

(74) Agente/Representante:

SALVÀ FERRER, Joan

DESCRIPCIÓN

Detección de modificaciones químicas en ácidos nucleicos

SOLICITUDES RELACIONADAS

[**0001**] La materia divulgada en el presente documento se basa y reivindica el beneficio de la Solicitud de Patente Provisional de Estados Unidos Nº de Serie 61/887.614, presentada el 7 de octubre de 2013.

INTERÉS DEL GOBIERNO

[0002] La invención se realizó con apoyo gubernamental bajo las Subvenciones NIH Nos. AI068462 y GM064803. El gobierno tiene ciertos derechos en la invención.

CAMPO TÉCNICO

[0003] La materia divulgada en el presente documento se refiere a la tecnología y los procedimientos para analizar la estructura de moléculas de ácido nucleicos, tales como moléculas de ARN. Más particularmente, la materia divulgada en el presente documento se refiere a procedimientos de detección de modificaciones químicas en ácidos nucleicos, tales como ARN, tales como las modificaciones químicas que estabilizan la estructura terciaria totalmente plegada y funcional del ácido nucleico, tal como ARN.

ANTECEDENTES

[0004] La función biológica de ácidos nucleicos, tales como ARN, está mediada por sus estructuras. Por ejemplo, el ARNm es considerado generalmente como una molécula lineal que contiene la información para dirigir la síntesis de proteínas dentro de la secuencia de ribonucleótidos. Los estudios han revelado una serie de estructuras secundaria y terciaria en el ARNm que son importantes para su función (Tinoco et al (1987) Symp Quant Biol 52: 135). Los elementos estructurales secundarios en el ARN se forman en gran medida mediante interacciones de tipo Watson-Crick entre las diferentes regiones de la misma molécula de ARN. Los elementos estructurales secundarios importantes incluyen regiones intramoleculares de doble cadena, bucles en horquilla, protuberancias en el ARN de doble cadena y bucles internos. Los elementos estructurales terciarios se forman cuando los elementos estructurales secundarios entran en contacto entre sí o con regiones de cadena única para producir una estructura tridimensional más compleja.

[0005] Se sabe muy poco sobre las estructuras tridimensionales precisas de ácidos nucleicos, incluyendo, en particular, el ARN. Sin embargo, ha habido una serie de esfuerzos de investigación que han demostrado que las estructuras de ARN, incluyendo las estructuras de cadena única, secundarias y terciarias, tienen importantes funciones biológicas más allá de simplemente codificar la información para producir proteínas en secuencias lineales (Resnekov et al. (1989) J Biol. Chem. 264: 9953; Tinoco et al (1987) Symp Quant Biol 52: 135; Tuerk et al (1988) PNAS USA. 85: 1364; y Larson et al (1987) Mol Cell. Biochem. 74: 5). Wilkinson, K. et al. describen el análisis cuantitativo de la estructura del ARN a través de análisis de la 2'-hidroxil acilación selectiva mediante extensión con cebadores ("Selective 2'-hydroxyl acylation analyzed by primer extensión (SHAPE): quantitative RNA structure analysis at single nucleotide resolution", NATURE PROTOCOLS, 2006, vol. 1, nº 3, páginas 1610-1616). Lucks, JB et al. decriben la caracterización de estructura multiplexada a través de análisis de 2'-hidroxil acilación selectiva mediante extensión con cebadores ("Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extensión sequencing (SHAPE-Seq)", P.N.A.S., 2011, vol. 108, nº 27, páginas 11063 -. 8). El documento US 2010/035761 A1 se refiere a procedimientos para analizar la estructura de moléculas de ARN, en particular, a procedimientos de, composiciones para, y productos de programas informáticos para el análisis de la estructura del ARN a través de la acilación de 2'-hidroxilo selectivo de alcóxido mediante extensión con cebadores. Por lo tanto, el desarrollo de enfoques para la evaluación de la estructura de las moléculas de ácidos nucleicos, tales como ARN, representa una necesidad actual y en curso en la técnica.

CARACTERÍSTICAS DE LA INVENCIÓN

[0006] La invención se define por las reivindicaciones adjuntas. Se dan a conocer procedimientos para detectar una o más modificaciones químicas en un ácido nucleico. Se da a conocer que el procedimiento comprende proporcionar un ácido nucleico que se sospecha que tiene una modificación química; sintetizar un ácido nucleico utilizando una polimerasa y el ácido nucleico proporcionado como una plantilla, en el que la síntesis se produce en condiciones en las que la polimerasa lee a través de una modificación química en el ácido nucleico proporcionado para producir de este modo un nucleótido incorrecto en el ácido nucleico resultante en el sitio de la modificación química; y detectar el nucleótido incorrecto.

[0007] Se dan a conocer procedimientos para detectar datos estructurales en un ácido nucleico. Se da a conocer que el procedimiento comprende proporcionar un ácido nucleico que se sospecha que tiene una modificación química; sintetizar un ácido nucleico utilizando una polimerasa y el ácido nucleico proporcionado como una plantilla, en el que la síntesis se produce en condiciones en las que la polimerasa lee a través de una modificación química en el ácido nucleico proporcionado para producir de este modo un nucleótido incorrecto en el ácido nucleico resultante en el

sitio de la modificación química; detectar el nucleótido incorrecto; y producir archivos de salida que comprenden datos estructurales para el ácido nucleico proporcionado.

[0008] Se da a conocer que el ácido nucleico proporcionado puede ser ARN. Se da a conocer que los procedimientos pueden comprender la detección de dos o más modificaciones químicas. La polimerasa lee a través de múltiples modificaciones químicas para producir múltiples nucleótidos incorrectos y los procedimientos comprenden la detección de cada nucleótido incorrecto.

[0009] Se da a conocer que el ácido nucleico ha sido expuesto a un reactivo que proporciona una modificación química o la modificación química es preexistente en el ácido nucleico. Se da a conocer que la modificación preexistente es un grupo 2'-O-metilo, y/o se crea por una célula de la que se deriva el ácido nucleico, tal como, pero no limitada a, una modificación epigenética y la modificación es 1-metil adenosina y 3-metil citosina.

[0010] Se da a conocer que el reactivo comprende un electrófilo. Se da a conocer que el electrófilo modifica selectivamente nucleótidos no restringidos en el ARN para formar un aducto 2'-O de ribosa covalente. Se da a conocer que el reactivo puede ser de 1 M7, 1 M6, NMIA, DMS, o combinaciones de los mismos. Se da a conocer que el ácido nucleico puede estar presente en o derivar de una muestra biológica.

[0011] La polimerasa es una transcriptasa inversa. Se da a conocer que la polimerasa puede ser una polimerasa nativa o una polimerasa mutante. Se da a conocer que el ácido nucleico sintetizado puede ser un ADNc.

[0012] Se da a conocer que la detección del nucleótido incorrecto comprende secuenciar el ácido nucleico. Se da a conocer que la información de la secuencia se alinea con la secuencia del ácido nucleico proporcionado. Se da a conocer que la detección del nucleótido incorrecto puede comprender el empleo de secuenciación masiva en paralelo en el ácido nucleico. En algunos, el procedimiento comprende amplificar el ácido nucleico. Se da a conocer que el procedimiento puede comprender amplificar el ácido nucleico usando un enfoque dirigido al sitio usando cebadores específicos, todo el genoma utilizando cebadores aleatorios, todo el transcriptoma utilizando cebadores aleatorios o combinaciones de los mismos.

[0013] Se da a conocer que el procedimiento puede comprender facilitar el diagnóstico o proporcionar un pronóstico de una enfermedad o trastorno en un sujeto a través de la detección de la una o más modificaciones químicas. Se da a conocer que el procedimiento puede comprender normalizar, comparar y/o unir diferentes conjuntos de datos que contienen información estructural del ácido nucleico, tal como, pero no limitada a, la información estructural de ARN. Se da a conocer que la estructura puede comprender un sitio de unión a cebador, un sitio de unión a proteína, un sitio de unión a molécula pequeña, o una combinación de los mismos. Se da a conocer que el procedimiento puede comprender el análisis de la estructura del ácido nucleico en presencia y ausencia de un cebador, una proteína, una molécula pequeña o una combinación de los mismos para identificar un sitio de unión a cebador, un sitio de unión a proteína, un sitio de unión a molécula pequeña, o una combinación de los mismos.

[0014] Se dan a conocer productos de programas informáticos que comprenden instrucciones ejecutables por ordenador incorporadas en un medio legible por ordenador en la realización de etapas que comprenden cualquier etapa del procedimiento de la materia divulgada en el presente documento. Se dan a conocer bibliotecas de ácidos nucleicos producidas mediante cualquier procedimiento de la materia divulgada en el presente documento.

[0015] Se dan a conocer kits que comprenden al menos un reactivo para llevar a cabo cualquier realización de la materia ada a conocer en la actualidad; y, opcionalmente, un recipiente para el mismo. Se da a conocer que el kit puede comprender instrucciones para llevar a cabo una realización de la materia divulgada en el presente documento.

[0016] Se dan a conocer cualquiera y todos los procedimientos, dispositivos, sistemas, kits, aparatos, composiciones y/o usos que se muestran y/o describen expresamente o por implicación en la información proporcionada con el presente documento, incluyendo, pero no limitado a, características que pueden ser evidentes y/o entendidas por los expertos en la técnica tras una revisión de la presente descripción.

[0017] Un objetivo de la materia divulgada en el presente dcumento es proporcionar procedimientos para el análisis de la estructura de ácido nucleico, incluyendo análisis de la estructura del ARN.

[0018] Un objetivo de la materia divulgada en el presente documento habiéndose expuesto anteriormente, y que se logra en su totalidad o en parte por la materia divulgada en el presente documento, otros objetivos resultarán evidentes a medida que avance la descripción cuando se tomen en conexión con los dibujos que se acompañan como mejor se describe en este documento a continuación.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

[0019] Figura 1. Modelado de estructura secundaria representativa para el ARNr 5S (SEQ ID NO: 1) sin y con datos SHAPE. Las predicciones de los pares de bases se ilustran con líneas (que indican pares de bases correctas, incorrectas y ausentes, respectivamente) en las representaciones convencionales de estructura secundaria (parte

superior) y representaciones de círculos (parte inferior). Los nucleótidos están coloreados en función de su reactividad SHAPE en una escala de negro, gris claro, gris oscuro, para baja, media y fuerte reactividad. Los nucleótidos que muestran una fuerte reactividad preferencial con NMIA (> 0,3 unidades) se indican con un símbolo delta.

Figura 2A y 2B. Análisis de la estructura de ARN de una molécula mediante secuenciación masiva en paralelo. (**Figura 2A**) Las moléculas de ARN experimentan variaciones estructurales locales y "respiración" ("breathing") en la que las regiones de un ARN se vuelven reactivas a una sonda química de una manera correlacionada. Los nucleótidos que interactúan muestran una reactividad correlacionada. El análisis de asociación estadística se utiliza para detectar y cuantificar las fuerzas de estas interdependencias. (**Figura 2B**) Las moléculas de ARN pueden adoptar múltiples conformaciones en solución. El análisis de agrupamiento espectral basado en la similitud de los patrones de reactividad de nucleótidos se puede utilizar para separar datos en cadenas de ARN individuales en diferentes conformaciones.

Figuras 3A-3C. Análisis RING de la estructura del ARN. (**A)** Número de mutaciones por transcripción o transcrito detectadas mediante transcripción inversa con (gris) y sin (negro) modificación con DMS. (**B**) Frecuencias de mutaciones inducidas por modificación con DMS en función de la posición de nucleótido. Los datos a partir de muestras tratadas con DMS se muestran en gris y los controles sin reactivos son de color negro. (**C**) RING para los ARN de riboswitch de TPP, dominio P546, y ARNasa P (SEQ ID NOs: 5-7) que muestran correlaciones fuertes y moderadas. Las correlaciones se producen entre las posiciones que son reactivas en la estructura nativa (círculos rellenos) o se vuelven reactivos durante los movimientos de "respiración"' (círculos abiertos), lo que refleja el componente de respiración estructural de interdependencias de reactividad. Los coeficientes de correlación de 0,025 y 0,035 corresponden a la mediana de los incrementos de 2,5 y 2,8 veces, respectivamente, en la probabilidad de mutación en un nucleótido debido a la mutación de un segundo nucleótido. Las estructuras secundarias se dibujan en orientaciones helicoidales relativas aproximados en el espacio tridimensional en base a estructuras conocidas.

Figuras 4A-4D. RING y análisis de la agrupación de riboswitch de TPP (SEQ ID NO: 5) en presencia y ausencia de ligando TPP. El análisis RING en presencia de (Figura 4A) ligando saturante y (Figura 4B) ausencia de ligando. Se muestran las asociaciones internucleotídicos fuertes y moderadas. Los nucleótidos que son menos o más estructurados en el grupo menor, menos poblado, se enfatizan con esferas abiertas y cerradas, respectivamente. El análisis de agrupamiento espectral en (Figura 4C) presencia de ligando saturante y (Figura 4D) ausencia de ligando. Hay dos grupos en cada estado. En presencia de ligando saturante, el grupo principal corresponde al riboswitch totalmente plegado. En ausencia de ligando, el grupo principal refleja un estado no estructurado con pocas interacciones internucleotídicas. El grupo menor importancia en la muestra con ligando saturarte está menos estructurado que el grupo principal y es similar a la estructura no-ligando. El grupo menor en la muestra sin ligando está más altamente estructurado que el grupo mayor específicamente en la región del bolsillo de unión a tiamina (círculos cerrados).

Figuras 5A y 5B. Análisis de agrupamiento espectral del riboswitch de TPP (SEQ ID NO: 5) a la concentración de ligando subsaturante de 200 nM de ligando. (A) Análisis RING de las interacciones de asociación internucleotídica. Las interacciones son menos en número y más débiles que las del ARN bajo condiciones de saturación de ligando (comparar con la Fig. 4A). (B) Tres grupos fueron identificados con fracciones de población de 32, 31, y 37%. Cada uno de estos grupos corresponde a un estado identificado en la concentración de ligando de saturación o en ausencia de ligando con nucleótidos correspondientes a estructuras unidas a ligando o estructuras sin ligando.

Figura 6: Visión general de SHAPE-MaP. Se trata ARN con un reactivo SHAPE que reacciona en nucleótidos conformacionalmente dinámicos. Durante la transcripción inversa, la polimerasa lee a través de aductos químicos en el ARN e incorpora un nucleótido no complementario a la secuencia original en el ADNc. El ADNc resultante se secuencia utilizando cualquier enfoque masivamente en paralelo para crear perfiles mutacionales (MaP). Las lecturas de secuenciación se alinean a una secuencia de referencia, y se calculan las tasas de mutaciones con resolución de nucleótidos, se corrige la base y se normalizan, produciendo un perfil de reactividad SHAPE estándar. Las reactividades SHAPE se pueden usar entonces para modelar estructuras secundarias, visualizar estructuras que compiten y son alternativas, o cuantificar cualquier proceso o función que module la dinámica del ARN de nucleótidos locales.

Figuras 7A y 7B: Precisión del modelado de la estructura secundaria dirigida por SHAPE-MaP. **(Figura 7A)** Precisiones de modelado de estructura secundaria descritas como una función de la sensibilidad (sens) y valor predictivo positivo (vpp) para cálculos realizados sin restricciones experimentales (sin datos), con electroforesis capilar convencional (CE) de datos, y con datos de SHAPE-MaP obtenidos con el reactivo 1M7 o con datos diferenciales de tres reactivos (Dif). Los resultados están sombreados en una escala para reflejar la menor a mayor precisión del modelado. **(Figura 7B)** Relación entre la profundidad de lectura de secuenciación, el nivel de éxito y la precisión del modelado de estructura de ARN. La precisión del modelo (eje vertical) se muestra como la media geométrica de sens y vpp de las estructuras predichas con respecto al modelo aceptado. Las representaciones en cajas resumen el modelado de la estructura secundaria de ARN ribosomal 16S como una función de profundidad de lectura SHAPE-Map simulada. En cada profundidad, se tomaron muestras de 100 trayectorias plegables. La línea en el centro de la caja indica el valor de la mediana y las cajas indican el intervalo intercuartil. Los "bigotes" contienen puntos de datos que están dentro de 1,5 veces el rango intercuartil y los valores atípicos se indican con marcas (+). El nivel de éxito es la señal total por encima de la base normalizada por nucleótido de transcripción.

Figuras 8A-8C: Análisis SHAPE-MaP del genoma NL4-3 de VIH-1. (Figura 8A) Reactividades SHAPE, entropía de Shannon y probabilidad de emparejamiento para el ARN genómico de VIH-1 NL4-3. Las reactividades se muestran como la ventana de 55 nt de mediana centrada, con respecto a la mediana global; las regiones por encima o por debajo de la línea son más flexibles o limitadas que la mediana, respectivamente. Los arcos que representan pares de bases están etiquetados por sus respectivas probabilidades de emparejamiento. Las áreas con muchos arcos superpuestos tienen múltiples estructuras posibles. Los pseudonudos (PK) se indican mediante arcos negros. Los datos mostrados corresponden a un solo experimento representativo; las regiones individuales, incluyendo pseudonudos propuestos, fueron confirmadas por réplicas independientes. **(Figura 8B)** Las regiones de ARN identificadas pore tener funciones

biológicas. Los paréntesis encierran regiones bien determinadas y se dibujan para enfatizar las ubicaciones de estas regiones con respecto a las características conocidas del ARN en el contexto del genoma viral. Las regiones corresponden a los dominios de entropía de Shannon baja-SHAPE baja y se amplían para incluir todas las hélices de intersección de la estructura secundaria de energía libre predicho más baja. Se muestran UTR 5' y 3'; aceptores y donantes de corte y empalme, respectivamente; tractos de polipurina; dominios variables; y los dominios del desplazamiento del marco de lectura y RRE. Estos elementos se encuentran dentro de las regiones con baja SHAPE y baja entropía de Shannon con mucha más frecuencia de lo esperado por casualidad (p = 0,002). (**Figura 8C**) Modelos de estructura secundaria para regiones, identificadas *de novo*, con reactividades SHAPE bajas y entropías de Shannon bajas. Los nucleótidos se muestran por la reactividad SHAPE y las estructuras con pseudonudos están etiquetadas.

Figura 9. Estrategias para el experimento SHAPE-MaP usando cualquiera de los cebadores específicos del gen o la fragmentación al azar para el análisis de muestras y preparación de la biblioteca de secuenciación. Se puede realizar SHAPE-MaP utilizando cebadores específicos del gen (para ARN pequeños o áreas específicas en ARN grandes y para el análisis de los ARN escasos y de concentración baja) o cebadores aleatorios (para el análisis exhaustivo de ARN grandes o transcriptomas completos) para crear el conjunto inicial de ADNc. Para ambos enfoques, el ARN se trata con un reactivo SHAPE o con disolvente bajo condiciones de interés, y una muestra de ARN se modifica bajo condiciones de desnaturalización. Para las muestras específicas de genes, la transcripción inversa y los cebadores de PCR se diseñan basándose en la secuencia diana conocida. Los ARN grandes se fragmentan al azar en una solución de Mg^{2+} tamponada. Se sintetizó ADNc de cadena única usando la transcripción inversa propensa a mutaciones; los sucesos de incorporación incorrecta en el ADNc naciente marcan la ubicación de aductos SHAPE en el ARN en cuestión. Se crearon ADNc de doble cadena mediante PCR (enfoque específico de gen) o síntesis de la segunda cadena (muestras fragmentados al azar). Se añadieron secuencias específicas de la plataforma de secuencias (incluyendo códigos de barras multiplexantes) a las bibliotecas de ADN de doble cadena, ya sea directamente a través de una segunda PCR (enfoque específico de gen) o mediante una ligadura de ADN-ADN de secuencias adaptadoras (muestras fragmentados al azar). A continuación, las bibliotecas preparadas por cualquiera de los procedimientos se secuenciaron, produciendo datos que se procesaron en perfiles de reactividad SHAPE utilizados en aplicaciones de modelado de estructura. SHAPE-MaP es totalmente independiente de la plataforma de secuenciación y el esquema de generación de la biblioteca (una vez que el ADNc inicial se ha sintetizado). Por lo tanto, se puede utilizar cualquier plataforma y cualquier esquema de generación de bibliotecas.

Figura 10. **Diseño de cebador para SHAPE-MaP**. Secuencias con un contenido GC bajo o desigualmente distribuido se benefician de los cebadores basados en LNA recién diseñados utilizados para analizar las secuencias de VIH-1 en este trabajo.

DESCRIPCIÓN DETALLADA

[0020] El modelado de la estructura de ácido nucleico, tal como el modelado de la estructura secundaria del ARN, es un problema difícil, y los éxitos recientes han elevado el nivel de precisión, consistencia y controlabilidad. Se han conseguido grandes aumentos en la precisión mediante la inclusión de datos sobre reactividad hacia sondas químicas: la incorporación de los datos de reactividad SHAPE 1M7 en un algoritmo de la clase mfold da lugar a una mediana en las precisiones para la predicción de par de bases que excede el 90%. Sin embargo, muchas estructuras de ácidos nucleicos, incluyendo estructuras de ARN, se modelan con una precisión significativamente menor. Se dan a conocer en el presente documento, en algunas realizaciones, que enfoques que incorporan reactividades diferenciales de los reactivos NMIA y 1M6, que detectan interacciones no canónicas y terciarias, en algoritmos de predicción, dan lugar a modelos de estructura secundaria de alta precisión para los ácidos nucleicos, tales como ARN, que previamente se demostró que eran díficiles de modelar. Por ejemplo, para los ARN, el 93% de pares de bases canónicas aceptadas se recuperaron en modelos dirigidos por SHAPE. Las discrepancias entre las estructuras aceptadas y modeladas eran pequeños y parecían reflejar diferencias estructurales genuinas. El modelado dirigido por SHAPE de tres reactivos se escala de forma concisa a ácidos nucleicos estructuralmente complejos (incluyendo ARN) para resolver el problema del análisis de la estructura secundaria en solución para muchas clases de ácidos nucleicos, incluyendo el ARN.

[0021] Las estructuras de ácidos nucleicos complejas, tales como estructuras de ARN de orden superior complejas, pueden jugar un papel importante en todas las facetas de la expresión génica; sin embargo, las redes de interacción a través del espacio que definen las estructuras terciarias y gobiernan el muestreo de múltiples conformaciones son poco conocidas. Se da a conocer en el presente documento, en algunas realizaciones, que los enfoques de análisis de la estructura de ácido nucleico de una sola molécula, tales como los enfoques de análisis de la estructura de ARN, en los que múltiples sitios de modificación química se identifican en cadenas de ácidos nucleicos individuales mediante secuenciación masiva en paralelo y, a continuación, se analizan para las interacciones correlacionadas y agrupadas. Por lo tanto, en algunas realizaciones, la estrategia identifica grupos de interacción mediante perfiles mutacionales (RING-MaP) y hace posible múltiples aplicaciones expansivas. Por ejemplo, en el presente documento se da a conocer, en algunas realizaciones, la identificación y la creación, a través de interacciones espaciales, de modelos 3D para los ácidos nucleicos, tales como ARN, que abarcan 80-265 nucleótidos, y la caracterización de amplias clases de interacciones intramoleculares que estabilizan ácidos nucleicos, tales como ARN. Adicionalmente, en el presente documento se dan a conocer, en algunas realizaciones, enfoques que distinguen distintas conformaciones en conjuntos de soluciones y revelan estados ocultos previamente no detectados y reconfiguraciones estructurales a gran escala que se producen en los ácidos nucleicos no plegados, tales ARN, en relación con estados nativos. El examen de la estructura de ácido nucleico de molécula individual RING-MaP permite un un análisis conciso y simple de las arquitecturas globales y múltiples conformaciones que gobiernan la función de los ácidos nucleicos, tales como ARN.

[0022] También se da a conocer en el presente documento, en algunas realizaciones, que la acilación de 2'-hidroxilo selectiva analizada mediante extensión de cebadores y perfiles mutacionales (SHAPE-MAP) hace posible la identificación *de novo* y a gran escala de motivos funcionales en los ácidos nucleicos, tales como ARN. En algunas realizaciones, los sitios de acilación de 2'-hidroxilo por SHAPE se codifican como nucleótidos no complementarios durante la síntesis de ADNc, tal como se mide mediante secuenciación masivamente en paralelo. Por ejemplo, el modelado guiado por SHAPE-MaP identificó más de un 90% de pares de bases aceptadas en ARN complejos de estructura conocida, y se utilizó para definir un nuevo modelo para el genoma de ARN del VIH-1. El modelo de VIH-1 contiene todos los motivos estructurados conocidos y elementos anteriormente desconocidos, incluyendo pseudonudos validados experimentalmente. SHAPE-MaP produce modelos de estructura secundaria precisos y de alta resolución, permite el análisis de ácidos nucleicos de baja abundancia (incluyendo ARN de baja abundancia), desenreda polimorfismos de secuencias en experimentos individuales y, por último, democratizará el análisis de la estructura de ácidos nucleicos, incluyendo el análisis de la estructura del ARN.

[0023] Los detalles de una o más realizaciones de la materia divulgada en el presente documento se exponen en la descripción adjunta a continuación. Otras características, objetos y ventajas de la materia divulgada en el presente documento serán evidentes a partir de la descripción detallada y las reivindicaciones. Todas las publicaciones, solicitudes de patente, patentes, y otras referencias mencionadas en el presente documento se incorporan por referencia en su totalidad. Algunas de las secuencias de polinucleótidos y polipéptidos descritas en este documento son una referencia cruzada a los números de acceso GENBANK®. Las secuencias de referencia cruzada en la base de datos GENBANK® se incorporan expresamente por referencia, ya que son secuencias equivalentes y relacionadas presentes en GENBANK® u otras bases de datos públicas. También se incorporan expresamente en el presente documento por referencia todas las anotaciones presentes en la base de datos GENBANK® asociadas con las secuencias descritas en el presente documento. En caso de conflicto, la presente memoria descriptiva, incluyendo definiciones, prevalecerá.

[0024] A menos que se defina lo contrario, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que se entiende comúnmente para un experto en la técnica a la que pertenece la materia divulgada en el presente documento. Aunque cualquier procedimiento, dispositivo y material similar o equivalente a los descritos en el presente documento puede usarse en la práctica o ensayo de la materia divulgada en el presente documento, se describen a continuación procedimientos, dispositivos y materiales representativos.

[0025] Siguiendo la convención de la ley de patentes, los términos "un", "una" y "el/la" se refieren a "uno o más" cuando se usa en esta solicitud, incluyendo las reivindicaciones. Así, por ejemplo, la referencia a "una célula" incluye una pluralidad de tales células y así sucesivamente.

[0026] A menos que se indique lo contrario, todos los números que expresan cantidades de ingredientes, condiciones de reacción, y así sucesivamente usados en la memoria descriptiva y reivindicaciones han de entenderse como modificados en todos los casos por el término "aproximadamente". En consecuencia, a menos que se indique lo contrario, los parámetros numéricos expuestos en esta memoria descriptiva y reivindicaciones adjuntas son aproximaciones que pueden variar dependiendo de las propiedades deseadas que se buscan obtener por la materia divulgada en el presente documento.

[0027] Tal como se utiliza en el presente documento, el término "aproximadamente", cuando se refiere a un valor o a una cantidad de masa, peso, tiempo, volumen, concentración o porcentaje pretende comprender variaciones en algunas realizaciones ± 20%, en algunas realizaciones ± 10%, en algunas realizaciones ± 5%, en algunas realizaciones ± 1%, en algunas realizaciones ± 0,5%, y en algunas realizaciones ± 0,1% de la cantidad especificada, ya que tales variaciones son apropiadas para llevar a cabo el procedimiento descrito.

[0028] El término "que comprende", que es sinónimo de "que incluye" "que contiene" o "caracterizado por" es inclusivo o abierto y no excluye elementos o etapas del procedimiento adicionales no citados. "Que comprende" es un término de la técnica usado en el lenguaje de las reivindicaciones que significa que los elementos nombrados son esenciales, pero que otros elementos se pueden añadir y todavía forman una construcción dentro del alcance de la reivindicación.

[0029] Tal como se usa en este documento, la frase "que consiste en" excluye cualquier elemento, etapa, o ingrediente no especificado en la reivindicación. Cuando la frase "consiste en" aparece en el cuerpo de una reivindicación, en lugar de inmediatamente después del preámbulo, se limita sólo el elemento que se expone en la cláusula; otros elementos no están excluidos de la reivindicación en su conjunto.

[0030] Tal como se usa en este documento, la frase "que consiste esencialmente en" limita el alcance de una reivindicación a los materiales o etapas especificados, además de aquellos que no afectan materialmente a la característica o características básicas y nuevas de la materia reivindicada.

[0031] Con respecto a los términos "que comprende", "que consiste en" y "que consiste esencialmente en", donde uno de estos tres términos se utiliza en el presente documento, la materia divulgada y reivindicada en el presente documento puede incluir el uso de cualquiera de los otros dos términos.

[0032] Tal como se utiliza en el presente documento, el término "y/o" cuando se utiliza en el contexto de una lista de entidades, se refiere a las entidades que están presentes individualmente o en combinación. Así, por ejemplo, la frase "A, B, C, y/o D" incluye A, B, C, y D individualmente, pero también incluye cualquier y todas las combinaciones y subcombinaciones de A, B, C, y D.

I. PROCEDIMIENTOS, SISTEMAS Y KITS

[0033] Se da a conocer en el presente documento, en algunas realizaciones, procedimientos que emplean la secuenciación de próxima generación (next-gen) con el análisis de ácido nucleico de alta precisión, tal como análisis de la estructura del ARN. En algunas realizaciones, se proporcionan procedimientos SHAPE-MaP. Un ejemplo de dicho procedimiento incluye un procedimiento para detectar modificaciones químicas en el ARN mediante la lectura a través de nucleótidos que contienen un aducto y que tiene una enzima transcriptasa inversa o cualquier otra polimerasa que incorporan un nucleótido incorrecto (no complementario) en el sitio de la modificación química. Otro ejemplo de dicho procedimiento incluye un procedimiento para detectar modificaciones químicas en los ácidos nucleicos, tales como ARN, mediante la lectura a través de los nucleótidos que contienen un aducto y que tienen una enzima transcriptasa inversa mutante que incorpora un nucleótido incorrecto (no complementario) en el sitio de la modificación química. Otro ejemplo de dicho procedimiento incluye un procedimiento para detectar modificaciones químicas a los ácidos nucleicos, tales como RNA, utilizando secuenciación masiva en paralelo. Otro ejemplo de dicho procedimiento es un procedimiento para detectar cualquier modificación química arbitraria a los ácidos nucleicos, tales como RNA, utilizando secuenciación masiva en paralelo y leída como cambios de secuencia en el ácido nucleico complementario sintetizado, tal como ARN.

[0034] También se describe en el presente documento el uso de SHAPE-MaP por los reactivos 1M7, 1M6 y NMIA (ejemplos de realización expuestos en los Ejemplos). De hecho, se proporciona de acuerdo con la materia divulgada en el presente documento el uso de SHAPE-MaP por cualquier agente químico.

[0035] Se da a conocer en el presente documento, en algunas realizaciones, procedimientos RING-MaP. Un ejemplo de dicho procedimiento incluye un procedimiento para detectar múltiples modificaciones químicas en los ácidos nucleicos, tales como ARN, mediante la lectura a través de múltiples sitios utilizando la transcriptasa inversa u otra polimerasa para incorporar múltiples nucleótidos no complementarios, específicamente en cada sitio de mutación. Un ejemplo de dicho procedimiento incluye un procedimiento para inferir la estructura del ácido nucleico, tal como ARN, a partir de la detección de múltiples modificaciones químicas en los ácidos nucleicos, tales como ARN.

[0036] También se describe en el presente documento el uso de RING-MaP por DMS. De hecho, se proporciona de acuerdo con la materia divulgada en el presente documento el uso de RING-MaP por cualquier agente químico.

[**0037]** Tal como se usa en el presente documento y en las reivindicaciones, el término "incorrecto", con respecto a la incorporación de un nucleótido, se refiere a la incorporación de un nucleótido que es no complementario (por las reglas de Watson-Crick; A-U, A-T, G-C) al nucleótido presente en la secuencia original. También incluye pequeñas deleciones en la secuencia.

[0038] La materia descrita en el presente documento para el análisis de modificación química de ARN y/o el análisis de la estructura de ARN incluye kits para llevar a cabo los procedimientos y análisis. La materia descrita en el presente documento para el análisis de modificación química de ARN y/o análisis de la estructura de ARN incluye enfoques de diagnóstico y procedimientos que emplean el análisis.

[0039] En algunas realizaciones, la modificación química es preexistente en el ácido nucleico, tal como ARN, tal como un grupo 2'-O-metilo. Por lo tanto, la modificación química se puede crear mediante cualquier reactivo químico, o puede ser creado por la célula (en el caso de una modificación epigenética), y en algunas realizaciones, utilizando polimerasas nativas y/o mutantes. La detección de una modificación epigenética se puede emplear en un enfoque de diagnóstico, por ejemplo. En algunas realizaciones, la modificación es 1-metil adenosina y/o 3-metil citosina (ambas pueden detectarse de la modificación DMS). Otras modificaciones epigenéticas son 6-metil adenosina, 3-metil uridina, 2-metil guanosina, y otros como será evidente para un experto en la técnica tras una revisión de la presente descripción.

[0040] La materia descrita en el presente documento para el análisis de modificación química de ácido nucleico, tal como ARN y/o el análisis de la estructura de ácido nucleico, tal como ARN, se puede implementar utilizando un producto de programa informático que comprende instrucciones ejecutables por ordenador incorporadas en un medio legible por ordenador. Los medios legibles por ordenador adecuado de ejemplo adecuados para la implementación de la materia descrita en el presente documento incluyen dispositivos de chip de memoria, dispositivos de memoria de disco, dispositivos lógicos programables, y circuitos integrados específicos de aplicación. Además, un producto de programa informático que implementa la materia descrita en el presente documento puede estar situado en un solo dispositivo o plataforma de computación o puede ser distribuido a través de múltiples dispositivos o plataformas de computación. Por lo tanto, la materia descrita en el presente documento puede incluir un conjunto de instrucciones de ordenador, que cuando se ejecutan por un ordenador, realiza una función específica para el análisis de la estructura de ácido nucleico, tal como ARN.

[0041] El ácido nucleico, tal como ARN, puede estar presente en una muestra biológica. La solución de reactivo-disolvente se añade a una solución biológica compleja que contiene ácidos nucleicos, tales como ARN. La solución puede contener diferentes concentraciones y cantidades de proteínas, células, virus, lípidos, monosacáridos y polisacáridos, aminoácidos, nucleótidos, ADN, y diferentes sales y metabolitos. La concentración del reactivo se puede ajustar para conseguir el grado deseado de modificación en los ácidos nucleicos, tales como ARN.

[0042] El término "disolvente aprótico" se refiere a una molécula de disolvente que no puede aceptar ni donar un protón. Los disolventes apróticos típicos incluyen, pero no se limitan a, acetona, acetonitrilo, benceno, butanona, butironitrilo, tetracloruro de carbono, clorobenceno, cloroformo, 1,2-dicloroetano, diclorometano, éter dietílico, dimetilacetamida, N, N-dimetilformamida (DMF), dimetilsulfóxido (DMSO), 1,4-dioxano, acetato de etilo, etilenglicol dimetil éter, hexano, N-metilpirrolidona, piridina, tetrahidrofurano (THF), y tolueno. Ciertos disolventes apróticos son disolventes polares. Ejemplos de disolventes apróticos polares incluyen, pero no se limitan a, acetona, acetonitrilo, butanona, N, N-dimetilformamida, y dimetilsulfóxido. Ciertos disolventes apróticos son disolventes no polares. Ejemplos de disolventes apróticos no polares incluyen, pero no se limitan a, éter dietílico, hidrocarburos alifáticos, tales como hexano, hidrocarburos aromáticos, tales como benceno y tolueno, e hidrocarburos halogenados simétricos, tales como tetracloruro de carbono.

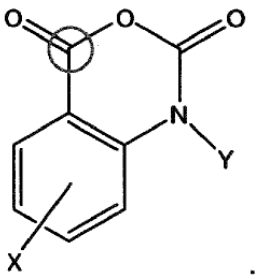
[0043] Continuando, los ácidos nucleicos, tales como ARN, se pueden modificar en presencia de proteína u otros ligandos biológicos pequeños y grandes. Los componentes de la solución que reaccionan directamente con el electrófilo, así como codisolventes orgánicos, incluyendo, por ejemplo, formamida y DMSO, pueden ser bien tolerados, pero pueden requerir ajustar las concentraciones de reactivos. Debido a que la reactividad del electrófilo puede ser fuertemente dependiente del pH, el pH se puede mantener en cualquier intervalo adecuado, tal como, pero no limitado a, pH 7,5 a 8,0. El intervalo dinámico que diferencia los nucleótidos más reactivos (flexibles) y menos reactivos (restringidos) típicamente abarca un factor de 20-50.

[0044] Continuando, el ácido nucleico puede también ser ADN. Las aplicaciones podrían ser detectar las modificaciones epigenéticas en el ADN mediante el uso de una polimerasa o transcriptasa inversa que incorpora un nucleótido incorrecto o no complementario cuando se sintetiza ADN past, una modificación química en el ADN diana. También, se incluyen procedimientos en los que el ADN es una molécula sintética, por ejemplo, un aptámero de ADN. En esta realización, la aplicación del procedimiento SHAPE-MaP o RING-MaP a ADN podría ser utilizado como un diagnóstico para detectar una molécula o proteína de unión al ADN.

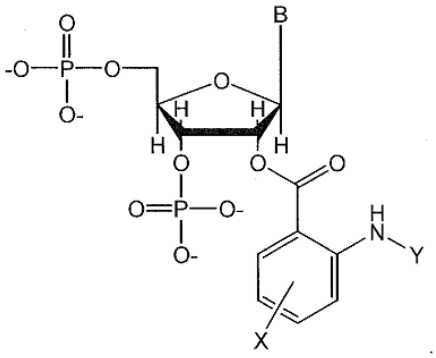
II. ELECTRÓFILOS DE SHAPE

[0045] Tal como se describe en el presente documento de acuerdo con algunas realizaciones de la materia divulgada en el presente documento, la química SHAPE aprovecha el descubrimiento de que la reactividad nucleófila de un grupo ribosa 2'-hidroxilo grupo está canalizada por la flexibilidad de nucleótidos locales. En los nucleótidos limitados por emparejamiento de bases o interacciones terciarias, el anión 3'-fosfodiéster y otras interacciones reducen la reactividad del 2'-hidroxilo. En cambio, las posiciones flexibles adoptan preferentemente conformaciones que reaccionan con un electrófilo, incluyendo, pero no limitado a, NMIA, para formar un 2'-O-aducto. A modo de ejemplo, NMIA reacciona genéricamente con los cuatro nucleótidos y el reactivo experimenta una reacción de hidrólisis paralela auto-inactivante. De hecho, la materia divulgada en el presente documento describe que cualquier molécula que puede reaccionar con un ácido nucleico como se ha descrito inmediatamente arriba se puede emplear de acuerdo con algunas realizaciones de la materia divulgada en el presente documento.

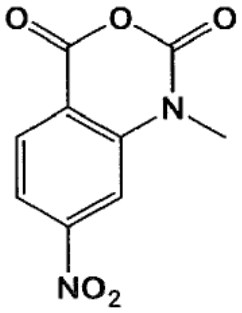
[0046] Se han desarrollado reactivos de SHAPE adicionales. Los reactivos de SHAPE incluyen, pero no se limitan a, derivados de anhídrido isatoico. En algunas realizaciones, los derivados de anhídrido isatoico adecuados para su uso con la metodología SHAPE se representan a continuación, en las que X e Y pueden ser cualquier grupo funcional, y el centro de carbono reactivo es un círculo:



[0047] Un aducto formado entre un derivado de anhídrido isatoico y un nucleótido de ARN puede tener la estructura:



[0048] En algunas realizaciones, el derivado de anhídrido isatoico puede ser anhídrido de 1-metil-7-nitroisatoico (1M7):



[0049] En algunas realizaciones, el sustituyente X del anhídrido isatoico puede ser un grupo funcional que incluye, pero no limitado a, alquilo, alquilo sustituido, cicloalquilo, arilo, arilo sustituido, heteroarilo, alcoxilo, ariloxilo, aralquilo, aralcoxilo, dialquilamino, nitro, carboxilo, halo, acilo, hidroxialquilo, aminoalquilo. En algunas realizaciones, Y puede ser un grupo funcional que incluye, pero no limitado a, alquilo, alquilo sustituido, cicloalquilo, arilo, arilo sustituido, heteroarilo, hidroxialquilo, y aminoalquilo.

[0050] Un grupo funcional llamado "X", "Y", o en algunos casos "R" tendrá generalmente la estructura que es reconocida en la técnica como correspondiente a un grupo que tiene ese nombre, a menos que se especifique lo contrario en el presente documento. A efectos de ilustración, ciertos grupos funcionales representativos llamados "X", "Y", o en algunos casos "R" se definen a continuación. Estas definiciones están destinadas a complementar e ilustrar, no excluir, las definiciones que serían evidentes para un experto ordinario en la técnica tras la revisión de la presente descripción.

[**0051]** Además de las moléculas mencionadas anteriormente, el DMS también se pueden emplear como reactivo de SHAPE. De hecho, la materia divulgada en el presente documento describe que cualquier molécula que puede reaccionar con ARN o ADN para dejar una modificación química permanente se puede emplear, de acuerdo con algunas realizaciones de la materia divulgada en el presente documento.

[0052] Tal como se utiliza en el presente documento, el término "alquilo" se refiere a cadenas de hidrocarburo C_{1-20} inclusives, lineales (es decir, "de cadena lineal"), ramificadas o cíclicas, saturadas o al menos parcialmente y en algunos casos completamente insaturadas (es decir, alquenilo y alquinilo), incluyendo por ejemplo, grupos metilo, etilo, propilo, isopropilo, butilo, isobutilo, terc-butilo, pentilo, hexilo, octilo, etenilo, propenilo, butenilo, pentenilo, hexenilo, octenilo, butadienilo, propinilo, butinilo, pentinilo, hexinilo, heptinilo y alenilo. "Ramificado" se refiere a un grupo alquilo en el que un grupo alquilo inferior, tal como metilo, etilo o propilo, está unido a una cadena de alquilo lineal. "Alquilo inferior" se refiere a un grupo alquilo que tiene de 1 a aproximadamente 8 átomos de carbono (es decir, un alquilo C_{1-8}), por ejemplo, 1, 2, 3, 4, 5, 6, 7 u 8 átomos de carbono. "Alquilo superior" se refiere a un grupo alquilo que tiene de aproximadamente 10 a aproximadamente 20 átomos de carbono, por ejemplo, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 átomos de carbono. En ciertas realizaciones, "alquilo" se refiere, en particular, a alquilos C_{1-8} de cadena lineal. En otras realizaciones, "alquilo" se refiere, en particular, a alquilos C_{1-8} de cadena ramificada.

[0053] Los grupos alquilo pueden estar opcionalmente sustituidos (un "alquilo sustituido") con uno o más sustituyentes de grupos alquilo, que pueden ser el mismo o diferente. El término "sustituyente de grupo alquilo" incluye, pero no se

limita a, alquilo, alquilo sustituido, halo, arilamino, acilo, hidroxilo, ariloxilo, alcoxilo, alquiltio, ariltio, aralquiloxilo, aralquiltio, carboxilo, alcoxicarbonilo, oxo, y cicloalquilo. Se puede insertar opcionalmente a lo largo de la cadena de alquilo uno o más de oxígeno, azufre o átomos de nitrógeno sustituidos o no sustituidos, donde el sustituyente de nitrógeno es hidrógeno, alquilo inferior (también denominado en este documento como "alquilaminoalquilo"), o arilo.

[0054] Por lo tanto, tal como se usa en el presente documento, el término "alquilo sustituido" incluye grupos alquilo, como se definen en el presente documento, en el que uno o más átomos o grupos funcionales del grupo alquilo se reemplazan por otro átomo o grupo funcional, incluyendo, por ejemplo, alquilo, alquilo sustituido, halógeno, arilo, arilo sustituido, alcoxilo, hidroxilo, nitro, amino, alquilamino, dialquilamino, sulfato y mercapto.

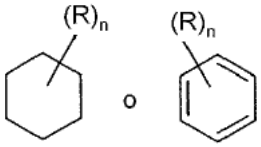
[0055] El término "arilo", se utiliza en el presente documento para referirse a un sustituyente aromático que puede ser un solo anillo aromático o múltiples anillos aromáticos que están condensados, unidos covalentemente, o unidos a un grupo común, tal como, pero no limitado a, un resto metileno o etileno. El grupo de unión común puede ser también un carbonilo, como en la benzofenona, u oxígeno, como en difeniléter, o nitrógeno, como en la difenilamina. El término "arilo" abarca específicamente compuestos aromáticos heterocíclicos. El anillo o anillos aromáticos pueden comprender fenilo, naftilo, bifenilo, difeniléter, difenilamina y benzofenona, entre otros. En realizaciones particulares, el término "arilo" significa un aromático cíclico que comprende de aproximadamente 5 a aproximadamente 10 átomos de carbono, por ejemplo, 5, 6, 7, 8, 9, o 10 átomos de carbono, e incluyendo anillos aromáticos de hidrocarburos y heterocíclicos de 5 y 6 miembros.

[**0056]** El grupo arilo puede estar opcionalmente sustituido (un "arilo sustituido") con uno o más sustituyentes del grupo arilo, que puede ser el mismo o diferente, en el que "sustituyente del grupo arilo" incluye alquilo, alquilo sustituido, arilo, arilo sustituido, aralquilo, hidroxilo, alcoxilo, ariloxilo, aralquiloxilo, carboxilo, acilo, halo, nitro, alcoxicarbonilo, ariloxicarbonilo, aralcoxicarbonilo, aciloxilo, acilamino, aroilamino, carbamoílo, alquilcarbamoílo, dialquilcarbamoílo, ariltio, alquiltio, alquileno, y -NR'R", en donde R' y R" pueden ser cada uno independientemente hidrógeno, alquilo, alquilo sustituido, arilo, arilo sustituido, y aralquilo.

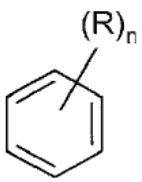
[0057] Por lo tanto, tal como se usa en el presente documento, el término "arilo sustituido" incluye grupos arilo, tal como se define en el presente documento, en el que uno o más átomos o grupos funcionales del grupo arilo se sustituyen por otro átomo o grupo funcional, incluyendo, por ejemplo, alquilo, alquilo sustituido, halógeno, arilo, arilo sustituido, alcoxilo, hidroxilo, nitro, amino, alquilamino, dialquilamino, sulfato y mercapto.

[0058] Los ejemplos específicos de grupos arilo incluyen, pero no se limitan a, ciclopentadienilo, fenilo, furano, tiofeno, pirrol, pirano, piridina, imidazol, bencimidazol, isotiazol, isoxazol, pirazol, pirazina, triazina, pirimidina, quinolina, isoquinolina, indol, carbazol, y similares.

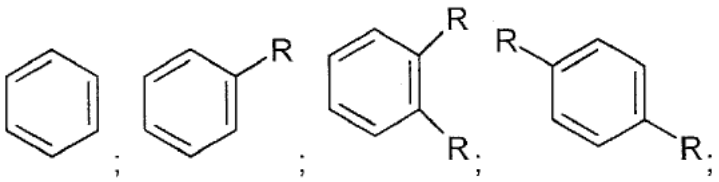
[0059] Una estructura representada generalmente por una fórmula, tal como:



tal como se utiliza en el presente documento se refiere a una estructura de anillo, por ejemplo, pero no limitado a, un compuesto cíclico, alifático y/o aromático de 3 carbonos, 4 carbonos, 5 carbonos, 6 carbonos, y similares, que comprende un grupo sustituyente R, en el que el grupo R puede estar presente o ausente, y cuando está presente, uno o más grupos R pueden estar sustituidos cada uno en uno o más átomos disponibles de carbono de la estructura de anillo. La presencia o ausencia del grupo R y el número de grupos R se determina por el valor del entero n. Cada grupo R, si hay más de uno, está sustituido en un carbono disponible de la estructura del anillo en vez de sobre otro grupo R. Por ejemplo, la estructura:



en la que n es un número entero de 0 a 2, comprende grupos de compuestos, que incluyen, pero no limitado a:



y similares.

[0060] En algunas realizaciones, los compuestos descritos por la materia divulgada en el presente documento contienen un grupo de unión. Tal como se utiliza en el presente documento, el término "grupo de unión" comprende un resto químico, tal como un radical furanilo, fenileno, tienilo y pirrolilo, que está unido a dos o más de otros restos químicos, en particular, grupos arilo, para formar una estructura estable.

[0061] Cuando un átomo mencionado de un anillo aromático o un anillo aromático heterocíclico se define como "ausente", el átomo mencionado se sustituye por un enlace directo. Cuando el grupo de unión o espaciador se define como ausente, el grupo de unión o el grupo se sustituyen por un enlace directo.

[0062] "Alquileno" se refiere a un grupo hidrocarburo alifático bivalente lineal o ramificado que tiene de 1 a aproximadamente 20 átomos de carbono, por ejemplo, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 , 13, 14, 15, 16, 17, 18, 19, o 20 átomos de carbono. El grupo alquileno puede ser lineal, ramificado o cíclico. El grupo alquileno puede estar también opcionalmente insaturado y/o sustituido con uno o más "sustituyentes del grupo alquilo". Se puede insertar opcionalmente a lo largo del grupo alquileno uno o más de oxígeno, azufre o átomos de nitrógeno sustituidos o no sustituidos (también denominado en este documento como "alquilaminoalquilo"), en el que el sustituyente de nitrógeno es alquilo como se describió previamente. Los grupos alquileno de ejemplo incluyen metileno (-CH_2-); etileno (-CH_2-CH_2-); propileno (-$(CH_2)_3$-); ciclohexileno (-$C_6H_{,O}$-)-, -CH=CH-CH=CH-; -CH=CH-CH_2-; -$(CH_2)_q$-N(R$HCH_2)_r$-, en el que cada uno de q y r es independientemente un número entero de 0 a aproximadamente 20, por ejemplo, 0, 1, 2, 3, 4, 5, 6, 7, 8 , 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20, y R es hidrógeno o alquilo inferior; metilendioxilo (-O-CH_2-O-); y etilendioxilo (-O-$(CH_2)_2$-O-). Un grupo alquileno puede tener de aproximadamente 2 a aproximadamente 3 átomos de carbono y puede tener además de 6-20 carbonos.

[0063] Tal como se utiliza aquí, el término "acilo" se refiere a un grupo ácido carboxílico orgánico, en el que el -OH del grupo carboxilo ha sido sustituido por otro sustituyente (es decir, tal como se representa por RCO-, en el que R es un grupo alquilo o un arilo, tal como se define en este documento). Por tanto, el término "acilo" incluye específicamente grupos arilacilo, tales como un grupo acetilfurano y fenacilo. Los ejemplos específicos de grupos acilo incluyen acetilo y benzoílo.

[0064] "Cíclico" y "cicloalquilo" se refieren a un sistema de anillos monocíclico o multicíclico no aromáticos de aproximadamente 3 a aproximadamente 10 átomos de carbono, por ejemplo, 3, 4, 5, 6, 7, 8, 9, o 10 átomos de carbono. El grupo cicloalquilo puede ser opcionalmente parcialmente insaturado. El grupo cicloalquilo también puede estar opcionalmente sustituido con un sustituyente de grupo alquilo, tal como se define en el presente documento, oxo, y/o de alquileno. Se puede insertar opcionalmente a lo largo de la cadena de alquilo cíclico, uno o más átomos de oxígeno, azufre o nitrógeno sustituidos o no sustituidos, donde el sustituyente de nitrógeno es hidrógeno, alquilo, alquilo sustituido, arilo o arilo sustituido, proporcionando de este modo un grupo heterocíclico. Los anillos cicloalquilo monocíclicos representativos incluyen ciclopentilo, ciclohexilo, y cicloheptilo. Los anillos cicloalquilo multicíclicos incluyen adamantilo, octahidronaftilo, decalina, alcanfor, canfano, y noradamantilo.

[0065] "Alcoxilo" se refiere a un grupo alquil-O-, en el que el alquilo es tal como se ha descrito anteriormente. El término, "alcoxilo", tal como se usa en el presente documento, puede referirse a, por ejemplo, metoxilo, etoxilo, propoxilo, isopropoxilo, butoxilo, t-butoxilo, y pentoxilo. El término "oxialquilo" se puede usar indistintamente con "alcoxilo".

[0066] "Ariloxilo" se refiere a un grupo aril-O-, en el que el grupo arilo es tal como se ha descrito anteriormente, incluyendo un arilo sustituido. El término "ariloxilo", tal como se usa en el presente documento, puede referirse a feniloxilo o hexiloxilo, y feniloxilo o hexiloxilo sustituidos con alquilo, alquilo sustituido, halo o alcoxilo.

[0067] "Aralquilo" se refiere a un grupo aril-alquilo en el que arilo y alquilo son como se describió previamente, y se incluyen arilo sustituido y alquilo sustituido. Grupos aralquilo de ejemplo incluyen bencilo, feniletilo y naftilmetilo.

[0068] "Aralquiloxilo" se refiere a un grupo aralquilo-O- en el que el grupo aralquilo es, tal como se ha descrito previamente. Un grupo aralquiloxilo de ejemplo es benciloxi.

[0069] "Dialquilamino" se refiere a un grupo -NRR', en el que cada uno de R y R' es independientemente un grupo alquilo y/o un grupo alquilo sustituido tal como se ha descrito anteriormente. Los grupos alquilamino de ejemplo incluyen etilmetilamino, dimetilamino y dietilamino.

[0070] "Alcoxicarbonilo" se refiere a un grupo alquil-O-CO-. Los grupos alcoxicarbonilo de ejemplo incluyen metoxicarbonilo, etoxicarbonilo, butiloxicarbonilo y t-butiloxicarbonilo.

[0071] "Ariloxicarbonilo" se refiere a un grupo aril-O-CO-. Los grupos ariloxicarbonilo de ejemplo incluyen fenoxi-carbonilo y naftoxi-carbonilo.

[0072] "Aralcoxicarbonilo" se refiere a un grupo aralquil-O-CO-. Un grupo aralcoxicarbonilo de ejemplo es benciloxicarbonilo.

[0073] "Carbamoílo" se refiere a un grupo H_2N-CO-.

[0074] "Alquilcarbamoílo" se refiere a un grupo R'RN-CO- en el que uno de R y R' es hidrógeno y el otro de R y R' es alquilo y/o alquilo sustituido, tal como se describe anteriormente.

[0075] "Dialquilcarbamoílo" se refiere a un grupo R'RN-CO-, en el que cada uno de R y R' es independientemente alquilo y/o alquilo sustituido, tal como se describe anteriormente.

[0076] "Aciloxilo" se refiere a un grupo acilo-O- en el que acilo es tal como se describe anteriormente.

[0077] "Acilamino" se refiere a un grupo acil-NH-, en el que acilo es tal como se describe anteriormente.

[0078] El término "amino" se refiere al grupo -NH_2.

[0079] El término "carbonilo" se refiere al grupo -(C=O)-.

[0080] El término "carboxilo" se refiere al grupo -COOH.

[0081] Los términos "halo", "haluro", o "halógeno", tal como se usan en el present documento, se refieren a fluoro, cloro, bromo, y yodo.

[0082] El término "hidroxilo" se refiere al grupo -OH.

[0083] El término "hidroxialquilo" se refiere a un grupo alquilo sustituido con un grupo -OH.

[0084] El término "aminoalquilo" se refiere a un grupo alquilo sustituido con un grupo -NH_2. Por lo tanto, un grupo "aminoalquilo" puede ser un grupo $NH_2(CH_2)_n$, en el que n es un número entero de 1 a 6 (es decir, 1, 2, 3, 4, 5, o 6).

[0085] El término "mercapto" se refiere al grupo -SH.

[0086] El término "oxo" se refiere a un compuesto descrito previamente en este documento en el que un átomo de carbono se sustituye por un átomo de oxígeno.

[0087] El término "nitro" se refiere al grupo -NO_2.

[0088] El término "tio" se refiere a un compuesto descrito previamente en este documento en el que un átomo de carbono o de oxígeno es reemplazado por un átomo de azufre.

[0089] El término "sulfato" se refiere al grupo -SO_4.

[0090] Cuando se utiliza el término "seleccionado independientemente", los sustituyentes a los que se hace referencia (por ejemplo, grupos R, tales como grupos R_1 y R_2 , o los grupos X e Y), pueden ser idénticos o diferentes. Por ejemplo, tanto X e Y pueden ser sustituidos alquilos, o X puede ser hidrógeno e Y puede ser un alquilo sustituido, o viceversa, y similares.

III. PLEGAMIENTO O PLEGADO DE ARN

[0091] La materia divulgada en el presente documento se puede realizar con ARN generado mediante procedimientos que incluyen, pero no limitado a, la transcripción *in vitro* y el ARN generado en las células y los virus. En algunas realizaciones, los ARN se pueden purificar mediante electroforesis en gel desnaturalizante y se renaturalizan para lograr una conformación biológicamente relevante. Además, se puede sustituir cualquier procedimiento que pliega el ARN a una conformación deseada a un pH deseado (por ejemplo, aproximadamente pH 8). El ARN se puede calentar primero y enfriar inmediatamente en un tampón de baja fuerza iónica para eliminar formas multiméricas. A continuación, se puede añadir una solución de plegado para permitir que el ARN logre una conformación apropiada y para prepararlo para una sonda sensible a la estructura con un electrófilo. En algunas realizaciones, el ARN se puede plegar en una única reacción y posteriormente se separa en reacciones de electrófilo (+) y (-). En algunas realizaciones, el ARN no se

pliega de forma nativa antes de la modificación. La modificación puede tener lugar mientras el ARN se desnaturaliza mediante condiciones de calor y/o bajas en sal.

IV. AGENTE PARA LA POLIMERIZACIÓN

[0092] El agente para la polimerización puede ser cualquier compuesto o sistema que funcione para llevar a cabo la síntesis de un ácido nucleico, incluyendo, por ejemplo, enzimas. La polimerasa puede ser una polimerasa nativa y/o una polimerasa mutante. Las enzimas adecuadas para este propósito incluyen, pero no se limitan a, ADN polimerasa I de E. coli, fragmento Klenow de la ADN polimerasa de E. coli, muteínas de polimerasa, transcriptasa inversa, otras enzimas, incluyendo enzimas térmicamente estables, tales como enzimas de la transcriptasa inversa murinas o aviar.

[0093] La cadena recién sintetizada y su cadena de ácido nucleico complementaria puede formar una molécula de doble cadena bajo condiciones de hibridación descritas en este documento y este híbrido se usa en las etapas posteriores del procedimiento.

EJEMPLOS

[0094] Los siguientes ejemplos se han incluido para proporcionar orientación a una persona con experiencia ordinaria en la técnica para practicar las realizaciones representativas de la materia divulgada en el presente documento. A la luz de la presente descripción y el nivel general de capacidad en la técnica, los expertos pueden entender que los siguientes Ejemplos pretenden ser solamente ejemplos y que numerosos cambios, modificaciones y alteraciones se pueden emplear sin apartarse del alcance de la materia divulgada en el presente documento.

EJEMPLO 1

Protocolo SHAPE-MP

Visión general:

[0095]
1.
A. Modificación SHAPE
B. Precipitación con EtOH/G25 o 50/RNEasy

2.
A. Transcripción inversa (3 horas, sin Mg^{2+}, Mn^{2+} 6 mM)
(de acuerdo para dejar durante la noche)
B. Columna de centrifugación G25/50

3.
A. Primera etapa de PCR utilizando cebadores SGP - sólo unos pocos ciclos (3)
B. Limpieza de PCR (limpieza micro PCR PureLink™)
(de acuerdo para congelar después, en la placa de PCR sellada)

4.
A. Segunda etapa de PCR utilizando cebadores específicos de índice premezclados - más ciclos que en el paso anterior. (27)
B. Limpieza PCR (limpieza micro PCR PureLink™)
(de acuerdo para congelar después, en la placa de PCR sellada)

5. Selección del tamaño de perlas Ampure™

6. Bioanalizador/Análisis Qubit

7. Agrupación de muestras, secuencia

Requerimientos de materiales:

[0096]
1.
A. ARN (100 **ng - 1 ug ***) (100 ul rxn)
B. Precipitación con EtOH/columna G25 o 50/kit RNEasy™

2.
A. Transcripción inversa (3 horas, sin Mg^{2+}) (**1-500 ng**) (-20 ul)
B. Columna de centrifugación G25/50 (50 ul de salida)

1. Modificación SHAPE:

[0097] La siguiente tabla muestra el proceso de modificación, en comparación con modificación SHAPE tradicional, si un volumen de 100 ul después de la reacción es conveniente para un plan de purificación. Sólo añadir agua, etanol y sal si se planea la precipitación con etanol (no recomendado). Si el ARN es demasiado pequeño para utilizar el kit RNeasy ™, en una columna de centrifugación (~ 200 pb) es mucho más eficiente y proporciona un mayor rendimiento de ARN:

Reactivo	**SHAPE normal**	**SHAPE M, +SHAPE**	**SHAPE MP, DMSO**	**SHAPE MP, DC Control**
ARN	1 ul	1 ul	1 ul	1 ul
Tampón de plegado	3 ul	x ul	x ul	89 ul de HEPES 50 mM. EDTA 4 mM, formamida al 50%
H_2O	5 ul	X ul	X ul	X ul
Total:	**9 ul**	**90 ul**	**90 ul**	**90 ul**

Etapa	**SHAPE normal**	**SHAPE M, +SHAPE**	**SHAPE MP, DMSO**	**SHAPE MP, DC Control**
1. Plegado/incubación	37 grados, etc.	37 grados, etc.	37 grados, etc.	95 grados durante 1 min
2. Modificación SHAPE	1 ul 10 mM	10 ul 100 mM	10 ul DMSO	10 ul 100 mM
3. Incubación	5 semividas	5 semividas	5 semividas	5 semividas
4. H_2O	90 ul	--	--	--
5. NaCl 4 M		--	--	--
6. EtOH	400 ul	--	--	--

Si el ARN es > 200 pb, usar la columna RNeasy ™ para eliminar el reactivo SHAPE y pequeños fragmentos, si es más pequeño dividir en 2 alícuotas de 50 ul y pasar a través de una columna de centrifugación G50. Si es necesario, combinar y precipitar con etanol. Volver a suspender en un volumen apropiado de agua o de Tris (pH 8,0).

2. Transcripción inversa, GSP (3 horas, sin Mg^{2+}):

[0098] Esta reacción RT se lleva a cabo durante 3 horas a 42 ° C en un tampón que no contiene $MgCl_2$, sino que utiliza $MnCl_2$ como el divalente de activación. Por ejemplo, hacer 2x 421 tampón sin Mg^{2+}, se indica como 421^- , 2x.

[0099] En primer lugar, hacer 10 x primer tampón de cadena:

5 ml	1 M Tris, pH 8	(500 mM final)
3,75 ml	2 M KCl	(750 mM final)
1,25 Ml	H_2O	
---	---	---
10 ml	Total	

[0100] DTT: Diluir una solución madre 1 M de DTT a 0,2 M

[0101] dNTPs: Hacer una mezcla que es 20 mM en cada uno

421$^-$, 2x:

[0102]
800 ul FSB^-, 10x
400 ul DTT 0,2 M
200 ul de mezcla de dNTP 20 mM

RT:

[0103]
1-500 ng de ARN; por ejemplo 150 - 500 ng
1 ul cebador GSP = 2 pmol (cebador de RT)
Agua a 11 ul
→ 65 °C durante 5 minutos, a continuación hielo
3,5 ul 421^-, 2x
0,24 ul 500 mM Mn^{2+}
5,26 ul H_2O

→ 42 °C durante 2 minutos
1 ul SSII
Incubar a 42 °C durante 180 minutos
Incubar a 70 °C durante 15 minutos
→ Mantener a 4 °C
Ajustar a 50 ul, pasar a través de una columna G50.
Elución es ~ 50 ul.

[0104] **Nota: Estos volúmenes pretenden ser flexibles, si el ARN ocupa un volumen mayor reducir el H_2O antes o después de la etapa de adición 421^- , 2x. Una mezcla maestra se realiza generalmente para consistencia, evitando el pipeteo de 0,24 ul de $MnCl_2$.

3. 1ª etapa de PCR usando cebadores GSP:

[0105] La reacción se lleva a cabo de acuerdo con las instrucciones para la polimerasa Phusion™ o Q5™. La polimerasa Q5™ es más capaz de sobrevivir con un alto contenido de GC. Ambas son enzimas de alta fidelidad y degradarán cebadores de cadena sencilla si se le da suficiente tiempo. Como resultado, se sugiere que la polimerasa se añada en último lugar, a continuación, se colocan las reacciones en un termociclador pre-calentado. **Utilizar una temperatura de hibridación calculada usando la calculadora de cebador NEB:**

[0106] http://www.neb.com/nebecomm/tech_referencia/tmcalc/#. UFYYxRgyHgs Dado que la muestra final es el producto de dos amplificaciones por PCR el número total de ciclos de manera deseable no debe exceder de 30 entre los dos. Se han utilizado 15 ciclos para cada una con éxito. Se han utilizado 2 - 3 ul de ADNc de la etapa anterior, pero esto se puede variar. Es importante comenzar utilizando puntas de ART en este punto para evitar la contaminación de la PCR. Aquí está una reacción típica de 50 ul:

Conc. Final:

[0107]
200 μM de cada dNTP
0,5 μM de cada cebador
50-500 ng de cada plantilla genómica
1X Q5 tampón de reacción
(1X Potenciador de GC elevado - opcional)
desnaturalización a 98 °C
Extensión 20 s/kb a 72 °C
1 unidad ADN polimerasa Q5 de alta fidelidad

Reactivo	Volumen a añadir (ul)	MM (_____)X
Plantilla	15	N/D
5X tampón	10	
dNTP 10 mM	1	
Cebador A 10 uM	2,5	N/D
Cebador B 10 uM	2,5	N/D
Potenciador de GC (opcional)	10	
H_2O	8,5	
Polimerasa Q5	0,5	
MM para cada Rxn:	N/D	

[0108] Las reacciones deben establecerse en el hielo. Después de la PCR, purificar con un kit de limpieza PureLink Micro PCR. La elución es ~9-10 ul.

4. 2ª etapa de PCR utilizando cebadores específicos de índice premezclados:

[0109] Esta reacción añade las secuencias específicas de Illumina, incluyendo la secuencia de índice, sobre los extremos del producto de ADN. Utilizar los cebadores de la 2ª etapa de la PCR premezclados. Estos cebadores se almacenan a una concentración de 10 uM y se marcan por un número de índice. Llevar a cabo la reacción exactamente como anteriormente, utilizando una temperatura de hibridación de 60 °.

Reactivo	Volumen a añadir (ul)	MM (_____)X
Plantilla	10	N/D
5X tampón	10	
dNTP 10 mM	1	
Cebador A 10 uM	2,5	N/D
Cebador B 10 uM	2,5	N/D

Potenciador de GC (opcional)	10	
H_2O	13,5	
Polimerasa Q5	0,5	
MM para cada Rxn:	N/D	35

[0110] Purificar con el kit de limpieza de PCR PureLink™ (no micro). La elución es de -50 ul.

5. Limpieza de perlas Ampure™:

[0111] Dejar calentar perlas Ampure ™-XP hasta temperatura ambiente. Alícuotas de 1,8 ml tardan alrededor de 15 minutos en llegar a este estado. Asegurarse de que los granos estén bien mezclados – mezlarlos con vórtice. Transferir las muestras a una placa de 96 pocillos.

[0112] Preparar etanol nuevo al 80% para esta purificación.
1. Unión a perlas: Añadir 50 ul de perlas mezcladas con vórtice a 50 ul de muestra, pipetear arriba y abajo 10 veces para mezclar.
2. Incubar en la mesa de trabajo durante 15 minutos a temperatura ambiente.
3. Colocar la placa en un soporte magnético 96 pocillos. Dejar durante 5 minutos.
4. Retirar 95 ul de sobrenadante depurado de cada muestra, teniendo cuidado de cambiar las puntas. Dejar la placa en el soporte magnético.
5. Lavar: Añadir 200 ul de etanol al 80% sin alterar las perlas - pipeta hacia el otro lado del pocillo.
6. Dejar incubar durante 30 segundos, a continuación, retirar y desechar todo el sobrenadante.
7. Repetir los pasos 5 y 6, de modo que las perlas se han lavado dos veces.
8. Dejar la placa en el soporte magnético, dejar que las perlas se sequen durante 15 minutos. Colocar una gran tapa de plástico sobre la placa para evitar que el polvo y las corrientes de aire alteren las muestras.
9. Elución de material: Retirar la placa del soporte magnético con cuidado, a continuación, volver a suspender las perlas en 32,5 ul de Tris 10 mM (pH 7,5 - 8,0).
10. Incubar a TA durante 2 minutos. Colocar en el soporte magnético.
11. Después de 5 minutos, retirar 30 ul de sobrenadante depurado y colcoar en tubos Eppendorf individuales.

EJEMPLO 2

Protocolo SHAPE-MP

Visión general:

[0113]
1.
A. Modificación SHAPE
B. EtOH ppt/G25 o 50/RNEasy™

2.
A. Fragmento con 3X FSB
B. Columna centrifugadora G25/50

3.
A. Transcripción inversa (3 horas, sin Mg^{2+}, Mn^{2+} 6 mM)
(de acuerdo para dejar durante la noche)
B. Columna centrifugadora G25/50

4.
A. Síntesis de la segunda cadena
B. Limpieza de PCR (limpieza PureLink™ PCR Micro)
(de acuerdo para congelar después, en la placa de PCR sellada)

5.
A. Reparación final
B. Limpieza de perlas Ampure™ XP
(de acuerdo para congelar después, en la placa de PCR sellada)

6.
A Adición de saliente A
B. La ligación de adaptadores en horquilla
C. Limpieza de perlas Ampure™ XP
D. Limpieza de perlas Ampure™ XP

7.
A. PCR en emulsión
B. Extracciones con éter
C. Limpieza de PCR (limpieza PureLink™ PCR)

8. Selección del tamaño de perla Ampure™
9. Análisis con Bioanalyzer™/Qubit™
10. Agrupación de muestras, secuencia

Requerimientos materiales:

[0114]
1.
A. Modificación SHAPE **(100 ng-1 ug)** (100 ul rxn)
B. Columna G25 o 50/kit RNEasy™

2.
A. Fragmento con 3X FSB (50 ul rxn)
B. Columna centrifugadora G25/50 (50 ul de salida)

3.
A. Transcripción inversa (3 horas, sin Mg^{2+}) **(1-500 ng)** (~20 ul)
B. Columna centrifugadora G25/50 (50 ul de salida)

4.
A. Síntesis de la segunda cadena **(10-100 ng)**
B. Limpieza de PCR (limpieza PureLink™ PCR Micro) (10 ul de salida)

5.
A. Reparación final **(Salida del anterior)**
B. Limpieza de perlas Ampure™ XP

6.
A Adición de saliente A **(Salida del anterior)**
B. La ligación de adaptadores en horquilla
C. Limpieza de perlas Ampure™ XP
D. Limpieza de perlas Ampure™ XP **(30 ul de salida)**

7.
A. PCR en emulsión **(varía, 7-30 ul del anterior)**
B. Extracciones con éter
C. Limpieza de PCR (limpieza PureLink™ PCR)

8. Selección del tamaño de perla Ampure™
9. Análisis con Bioanalyzer™/Qubit™
10. Agrupación de muestras, secuencia **(total, 10 ul de ADN 20 mM)**

[0115] * Esta cantidad no ha sido cuidadosamente investigada a límites aún más bajos, y se espera que se puedan emplear límites inferiores.

1. Modificación SHAPE:

[0116] El ARN de interés se modifica una vez usando una solución madre de reactivo SHAPE 100 mM, reactivo 10 mM en condiciones de reacción finales. Por conveniencia, se usan volúmenes de reacción de 100 ul (10X SHAPE normal). Realizar los pasos 4, 5 y 6 sólo si utiliza etanol para precipitar; si se utilizan los kits G25/50 o RNeasy™, ajustar el volumen apropiadamente y usar las instrucciones del fabricante.

[0117] El control desnaturalizado se lleva a cabo en formamida al 50% FINAL. Hacer un tampón de control 3X DC que contiene HEPES 150 mM y EDTA 12 mM, sin formamida.

[0118] La siguiente tabla muestra el proceso de modificación, en comparación con la modificación SHAPE tradicional:

Reactivo	**SHAPE normal**	**SHAPE MaP, +SHAPE**	**SHAPE MaP, DMSO**	**SHAPE MaP, DC Control**
ARN	1 ul	x ul	x ul	1 ul
Tampón de plegado	3 ul	x ul	x ul	HEPES 50 mM,

				EDTA 4 mM, formamida al 50%
H_2O	5 ul	X ul	X ul	X ul
Total:	**9 ul**	**90 ul**	**90 ul**	**90 ul**

Etapa	**SHAPE normal**	**SHAPE MP, +SHAPE**	**SHAPE MP, DMSO**	**SHAPE MP, DC Control**
1. Plegado/incubación	37 grados, etc.	37 grados, etc.	37 grados, etc.	95 grados durante 1 min
2. Modificación SHAPE	1 ul 100 mM	10 ul de reactivo SHAPE 100 mM	10 ul de reactivo SHAPE 100 mM	10 ul de reactivo SHAPE 100 mM
3. Incubación	5 semividas	5 semividas	5 semividas	5 semividas
4. H_2O	90 ul	--	--	--
5. NaCl 4 M		--	--	--
6. EtOH	400 ul	--	--	--

[0119] Si el ARN es > 200 pb, usar la columna RNeasy ™ para eliminar el reactivo SHAPE y pequeños fragmentos. Si es más pequeño, reducir el volumen de rxn a 50 ul total y pasar a través de una columna G25. Se puede utilizar etanol para precipitar pero tarda más y da lugar a una mayor pérdida de muestra.

2. Fragmento con 3X FSB:

[0120] La fragmentación se lleva a cabo con Mg^{2+} 9 mM como el divalente funcional; esto es equivalente a 3X FSB suministrado con las enzimas Superscript. Por ejemplo, utilizar una incubación de 4 minutos a 94 °C. Esta digestión se lleva a cabo en una placa de PCR de modo que todas las muestras se pueden calentar duerante exactamente la misma cantidad de tiempo.

Protocolo:

[0121] Crear un programa PCR con una tapa calentada a incubado durante el tiempo seleccionado, colocar primero una etapa de mantenimiento a 94 °C, terminando con un mantenimiento a 4ºC. Intenta estar allí cuando el tiempo de digestión ha acabado para colocar la placa en hielo inmediatamente.

x ul RNA
Y ul 5X FSB (FSB + Mg^{2+}) - Se suministra con enzimas SS
Si es necesario, ajustar el volumen a un volumen total de 50 ul con agua, pasar a través de una columna de centrifugación G25/50. La elución es de ~ 50 ul.

3. Transcripción inversa, GSP (3 horas, sin Mg^{2+}):

[0122] Esta reacción RT se lleva a cabo durante 3 horas a 42 °C en un tampón que no contiene $MgCl_2$, sino que utiliza $MnCl_2$ como el divalente de activación. Por ejemplo, hacer 2x 421 tampón sin Mg^{2+}, opcionalmente se indica como 421^-, 2x.

[0123] En primer lugar, hacer 10 x primer tampón de cadena:

5 ml	1 M Tris, pH 8	(500 mM final)
3,75 ml	2 M KCl	(750 mM final)
1,25 Ml	H_2O	
---	---	---
10 ml	Total	

[0124] DTT: Diluir una solución madre 1 M de DTT a 0,2 M

[0125] dNTPs: Hacer una mezcla que es 20 mM en cada uno

421^-, 2x:

[0126]
800 ul FSB^-, 10x
400 ul DTT 0,2 M
200 ul de mezcla de dNTP 20 mM

RT:

[0127]
1-500 ng de ARN; se han utilizado 150 - 500 ng
1 ul de nonámeros aleatorios (50-250 ng) – se han utilizado 200 ng
Agua a 11 ul
→ 65 °C durante 5 minutos, a continuación hielo
3,5 ul 421^-, 2x
0,24 ul 500 mM Mn^{2+}
5,26 ul H_2O
→ 25 °C durante 2 minutos
1 ul SSII
-> 25 °C durante 10 minutos
Incubar a 42 °C durante 180 minutos
Incubar a 70 °C durante 15 minutos
→ Mantener a 4 °C
Ajustar a 50 ul, pasar a través de una columna G50.
Elución es ~ 50 ul.

[0128] **Nota: Estos volúmenes pretenden ser flexibles, si el ARN ocupa un volumen mayor reducir el H_2O antes o después de la etapa de adición 421^- , 2x. Una mezcla maestra se realiza generalmente para consistencia, evitando el pipeteo de 0,24 ul de $MnCl_2$.

4. Síntesis de la segunda cadena:

[**0129]** Usando el híbrido RNA/DNA de la etapa anterior, se crea una segunda cadena de ADN y se reparan las discontinuidades ("nick") en el ADN, por ejemplo, utilizando un kit de NEB - # E6111S/L. El ADN de entrada es 10 - 100 ng de la primera cadena de ADN en un volumen de 20 ul. El siguiente paso del protocolo añade 48 ul de H_2O, se cree que para la dilución del tampón restante de la reacción de RT. Dado que no hay ninguno, la entrada es de 10-100 ng en 68 ul H_2O.

Añadir:

[0130] 68 ul de ADN/ARN
8 ul 10x tampón de síntesis de segunda cadena
4 ul de mezcla de enzimas
80 ul de volumen total

[0131] Mezclar e incubar a 16 °C durante 2,5 horas. Purificar con un kit de limpieza Micro PCR PureLink™. La elución es de ~9-10 ul.

5. Reparación de extremos:

[0132] A partir de este punto en adelante, se utiliza una placa de PCR de 96 pocillos, por ejemplo, una fila a la vez; se llevan a cabo reacciones en una, purificar las perlas, a continuación, pasar a la siguiente fila. Sellar los pocillos utilizados anteriormente para evitar la contaminación cruzada. Las placas son caras, pero permiten el tratamiento de todas las muestras de manera más similar y son las más susceptibles de una separación rápida de las perlas. Esta etapa elimina los bordes salientes, fija fosfatos cíclicos y prepara el ADN de doble cadena para los pasos de ligación posteriores. El ADN de doble cadena resultante tiene extremos romos.

[0133] El módulo de reparación de extremos NEBNext (E6050 S o L) sugiere el uso de una reacción de 100 ul (vol. total) para 1 a 5 ug de ADN fragmentado. Por ejemplo, utilizar una reacción de 50 ul para de 0,5 a 2,5 ug de ADN fragmentado de entrada teórica. Sacar las perlas Ampure™ XP y permitir que lleguen a temperatura ambiente durante 30 minutos. Mezclar los siguientes:

X ul de ADN de doble cadena
5 ul de tampón de reaparación de extremos NEBNext™ (10x)
2,5 ul de mezcla de enzimas NEBNext™
x ul H_2O
50 ul de volumen total

[0134] Incubar a 20 °C durante 30 minutos. Ajustar a 100 ul mediante la adición de 50 ul de Tris (pH 8,0).
1. Añadir 160 ul de perlas Ampure XP, ajustar la pipeta a 200 ul, mezclar hacia arriba y hacia abajo 10 veces. Incubar a temperatura ambiente durante 15 minutos.
2. Colocar la placa en soporte magnético, esperar 5 minutos y, a continuación, retirar y desechar **127,5 ul** de sobrenadante aclarado.
3. Repetir el paso 2 una vez, dejar la placa sobre soporte magnético.
4. Añadir 200 ul de EtOH al 80% recién preparado a cada pocillo, no lavar sobre perlas.

5. Esperar 30 segundos, retirar todo el sobrenadante, repetir una vez para 2 lavados.
6. Dejar la placa sobre soporte magnético durante 15 minutos para secar las perlas.
7. Retirar la placa de soporte magnético, añadir 17,5 ul de Tris 10 mM, pH 8,0 y volver a suspender las perlas. Incubar las perlas resuspendidas durante 2 minutos a temperatura ambiente.
8. Colocar la placa sobre soporte magnético, esperar 5 minutos, retirar **15 ul** y colocar en nuevos pocillos.

6. Colas en los extremos:

[0135] Módulo de cola dA NEBNext E6053 S/L, 1-5 ug de ADN romo de extremos reparados (100-1000 pb). Mezcla:
15 ul de ADN
2,0 ul 10X de tampón de cola dA
1,2 ul de fragmento Klenow
1,8 ul de agua
20,0 ul de volumen total

[0136] Incubar durante 30 minutos a 37 °C en un termociclador. Proceder con la ligadura O llevar a cabo la purificación de perlas (añadir 1,8 perlas de volumen). Eluir en 23 ul de agua o Tris 10 mM (pH 8,0). Proceder con 1 ul en un chip bioanalizador Agilent™ (ADN de alta sensibilidad) para determinar la concentración relativa de adaptador para uso en la ligación. Es probable que una cantidad mucho menor de adaptador pueda ser utilizada en el siguiente paso en comparación con el sugerido por el fabricante Illumina.

7. Ligación con adaptadores:

Mezclar los siguientes:

[0137]
20,0 ul de ADN de la etapa anterior
7,5 ul de tampón Rxn de ligación NEBNext Quick (5X) (NO 2X)
2,5 ul de adaptador de ADN
3,75 ul de ADN ligasa de T4 Quick
3,75 ul de H_2O
37,5 ul de volumen total

[0138] Incubar durante 15 minutos a 20 °C. Añadir 5 ul de H_2O para llevar el volumen a 42,5 ul totales. Proceder con la limpieza de perlas Ampure™ XP.

Primera limpieza:

[0139]
1. Añadir 42 ul de perlas a cada muestra
2. Ajustar la pipeta a 85 ul y mezclar cada muestra 10 veces
3. Incubar durante 15 minutos a temperatura ambiente
4. Colocar la placa sobre soporte magnético y dejar reposar durante 5 minutos
5. Retirar **79,5 ul** de sobrenadante y desechar
6. Añadir 200 ul de etanol al 80% recién preparado a cada pocillo
7. Dejar incubar las perlas durante 30 segundos
8. Retirar todo el sobrenadante y desechar
9. Repetir los pasos 6-8 una vez más (dos lavados)
10. Dejar que las perlas se sequen durante 15 minutos sobre el soporte magnético
11. Retirar del soporte, resuspender en **52,5 ul** de agua o Tris 10 mM
12. Incubar durante 2 minutos a temperatura ambiente
13. Colocar sobre el soporte magnético durante 5 minutos
14. Retirar 50 ul de sobrenadante de cada muestra y colocar en un nuevo pocillo en la placa de 96 pocillos.

Segundo limpieza:

[0140]
1. Añadir **50 ul** de perlas a cada muestra
2. Ajustar la pipeta a 100 ul y mezclar cada muestra 10 veces
3. Incubar durante 15 minutos a temperatura ambiente
4. Colocar la placa sobre soporte magnético y dejar reposar durante 5 minutos
5. Retirar **95 ul** de sobrenadante y desechar
6. Añadir 200 ul de etanol al 80% recién preparado a cada pocillo
7. Dejar incubar las perlas durante 30 segundos
8. Retirar todo el sobrenadante y desechar
9. Repetir los pasos 6-8 una vez más (dos lavados)

10. Dejar que las perlas se sequen durante 15 minutos sobre el soporte magnético
11. Retirar del soporte, resuspender en **22,5 ul** de agua o Tris 10 mM
12. Incubar durante 2 minutos a temperatura ambiente
13. Colocar sobre el soporte magnético durante 5 minutos
14. Retirar 20 ul de sobrenadante de cada muestra y colocar en tubos eppendorf pequeños.

8. PCR en emulsión:

[0141] Esta etapa amplifica las bibliotecas de ADNc de una manera que impide en gran medida los eventos de recombinación. La creación de una emulsión de reacción de PCR acuosa en una mezcla de aceite-detergente crea millones de pequeñas burbujas que actúan como cámaras de reacción pequeñas. Como resultado, se producen pocas recombinaciones. El protocolo se lleva a cabo según Williams et al (Nature Methods vol. 3 no.7, 2006 pág 545), preparar la mezcla de aceite, tal como se ha descrito.

[0142] Los cambios en este protocolo:
1. Usar la polimerasa Q5™, no Pfu. Por consiguiente, el usar el cebador Q5™ y las concentraciones de dNTP, así como las condiciones de ciclado de Q5™.
2. A menos que utilice Q5™ "hot start", añadir la enzima inmediatamente antes del goteo de la fase acuosa en su mezcla de emulsión. Esto evitará la degradación mediada por polimerasa de sus cebadores.
3. Usar un paso de desnaturalización a 95 °C durante sus ciclos.
4. Antes de utilizar los tubos de placa en speed vac, abrir en un bloque de calor a ~ 65 °C durante ~ 5 minutos para eliminar el éter en exceso.

[0143] Una 1X mezcla usando Q5 (en este caso no se utilizó ningún potenciador de GC):

10 ul	Plantilla de ADN
52 ul	Tampón Q5
26 ul	BSA
13 ul	Cebador 1
13 ul	Cebador 2
5,2 ul	dNTPs (10 mM)
2,6 ul	Polimerasa Q5
138,2 ul	H2O

[0144] Debe fabricarse una mezcla madre excluyendo la plantilla de ADN y la enzima. La purificación de las emulsiones rotas sometidas a speed-vac se realiza utilizando un kit de limpieza de PCR PureLink™ (no micro). Se lleva a cabo una limpieza final de las perlas utilizando un volumen 1 a 1 de las perlas de muestra: los granos (protocolo de Illumina) o el protocolo de Iowa de mayor volumen.

Limpieza Ampure - Amplicones - Protocolo de Iowa

[0145]
1. Llevar las perlas Ampure™ XP a temperatura ambiente.
2. Preparar EtOH al 70% recién preparado (se necesitan ~ 3 ml por muestra).
3. Llevar un volumen de ADN a 45 ul en TE/EB (suficiente con ~200 ng) en un tubo de baja unión de 1,5 ml.
4. Mezclar bien las perlas Ampure™ XP y añadir 72 ul de perlas a los 45 ul de ADN (relación 1,6:1,0).
5. Girar en vórtice, centrifugar e incubar a temperatura ambiente durante 5 minutos.
6. Colocar sobre el soporte magnético (Sugerido: Invitrogen Dynal™, Red-Silver).
7. Esperar que se forme el gránulo y retirar el sobrenadante.
8. Añadir 500 ul de EtOH al 70 % al lado opuesto del imán (no molestar a los gránulos).
9. Esperar 30 segundos y retirar el EtOH.
10. Repetir los pasos 8-9 una vez más.
11. Permitir que las perlas se sequen por completo (se pueden retirar del imán y colocar en un bloque de calor a 37 °C durante 3-4 minutos). A veces habrá grietas visibles en el gránulo cuando está seco.
12. Para eluir añadir 45 ul de TE (Qiagen o EB), girar en vórtice, centrifugar y colocar sobre soporte magnético.
13. Después de que se forme el gránulo, transferir 45 ul con el ADN eluido en un nuevo tubo.
14. Repetir los pasos 4-13 una o dos veces más.
15. Aplicar 1 ul en el bioanalizador Agilent™ (chip de ADN de alta sensibilidad).

Ejemplo 3

Modelado de la estructura secundaria de ARN a una precisión elevada consistente utilizando SHAPE diferencial

[0146] El ARN es un portador de información central en biología (Sharp 2009). La información se codifica en el ARN a dos niveles diferentes: en su secuencia primaria y en su capacidad de plegarse en estructuras de orden superior (Leontis et al 2006; Dethoff et al 2012). El nivel más fundamental de la estructura de orden superior es el patrón de

apareamiento de bases o estructura secundaria. La definición de la estructura secundaria de un ARN es también un primer paso importante en el modelado de la estructura terciaria (Hajdin et al 2010; Weeks 2010; Bailor et al 2011.). Las estructuras de las moléculas de ARN modulan las numerosas funciones de ARN y las interacciones de ARN con proteínas, moléculas pequeñas, y otros ARN en el empalme, traducción, y otros mecanismos reguladores (Mauger et al. 2013). El modelador *de novo* preciso de la estructura secundaria del ARN es un reto: en ausencia de restricciones experimentales, los algoritmos actuales predicen los patrones de apareamiento de bases que contienen, en promedio, 50%-70% de los pares canónicoss (G-C, A-U y G-U) en estructuras secundarias establecidas a través de análisis filogenético o procedimientos experimentales de alta resolución (Mathews et al 2004; Hajdin et al 2013.). El reto del modelado resulta del hecho de que sólo hay cuatro nucleótidos de ARN; y estos nucleótidos tienen el potencial para disponerse en muchas estructuras secundarias de ARN, a menudo energéticamente similares, aunque muchos ARN adoptan unas pocas o solamente estructuras individuales (Tinoco y Bustamante 1999). Las características que son difíciles de extraer únicamente a partir de la secuencia, tales como mecanismos cinéticos, facilitadores de proteínas y unión a ligando, también influyen en el plegado de ARN.

[0147] La identificación de la estructura secundaria de ARN correcta también resulta mucho más difícil a medidad que aumenta la longitud del ARN. Se pueden utilizar reactivos de acilación de 2'-hidroxilo selectivo analizado mediante extensión de cebadores (Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) para investigar la flexibilidad de casi todos los nucleótidos en un ARN (Merino et al 2005; McGinnis et al 2012). La reactividad en el 2'-hidroxilo hacia el reactivo anhídrido de 1-metil-7-nitroisatoico (1M7) mide la flexibilidad local de nucleótidos. Debido a que los nucleótidos apareados con bases también están limitados estructuralmente, la reactividad SHAPE es aproximadamente inversamente proporcional a la probabilidad de que un nucleótido esté apareado. La incorporación de información de reactividad SHAPE en algoritmos de plegado de ARN da lugar a precisiones > 90% para la mayoría de los ARN, incluyendo aquellos con pseudonudos individuales (Deigan et al 2009; Hajdin et al 2013.). SHAPE se ha utilizado para crear modelos con resolución de nucleótidos para los genomas virales de VIH-1 (Watts et al. 2009) y STMV (Archer et al. 2013) y para analizar los cambios conformacionales en el VIH-1 (Wilkinson et al. 2008) y el virus de la leucemia murina de Moloney (Grohman et al. 2013). Aunque el plegamiento dirigido por SHAPE produce modelos casi perfectos para muchos ARN, quedan unos ARN cuyas estructuras son difíciles de recuperar usando un experimento de sondeo de una sola estructura (Cordero et al 2012; Leonard et al 2013.). Estos ARN "duros" se modelan con sensibilidades en el intervalo de 75%-85%.

[0148] La utilidad de los modelos de estructura secundaria a diferentes precisiones se puede resumir de la siguiente manera. Los modelos con sensibilidades de predicción < 60% contienen grandes errores en la estructura global y no son generalmente útiles para la generación de hipótesis biológicas. Los algoritmos sólo computacionales logran precisiones de predicción de mediana del ~ 70%. Un modelo individual que recupera el 70% de los pares de bases aceptados tendrá algunas hélices correctas y también errores críticos. Aunque los enfoques que recuperan el 70% de los pares de bases aceptados incluyen pares correctos e incorrectos, generalmente es difícil determinar qué hélices son correctas y cuáles no. Utilizando un modelado dirigido por SHAPE, las estructuras predichas para la mayoría de los ARN desafiantes contienen 80%-85% de pares de bases aceptados. En algunos casos, los pares de bases incorrectamente pronosticados están dispersos por todo el ARN de tal manera que el modelo general es bastante bueno. En otros casos, los errores se encuentran en elementos estructurales que se sabe que son funcionalmente importantes.

[0149] En promedio, el modelado dirigido por SHAPE actualmente recupera ~93% de los pares de bases aceptados en conjuntos desafiantes de moléculas de ARN. Este nivel de sensibilidad es suficiente para la generación de hipótesis biológicas robustas y para el modelado de la estructura tridimensional. Muchos de los modelos generados en este nivel de precisión se diferencian de los modelos aceptados por unos pocos pares de bases y se deben considerar casi perfectos. La mejora de la precisión a un nivel mayor del 90% para todos los ARN es el desafío actual en el modelado de la estructura secundaria dirigida de forma experimental. La inclusión de información bioquímica completa y rica en información adicional podría informar mejor y potencialmente resolver el problema del modelado de la estructura secundaria de ARN.

[0150] Un enfoque denominado en el presente documento como SHAPE "diferencial" revela interacciones no canónicas locales y de estructura terciaria basadas en experimentos de sondeo bioquímico sencillo. En esta estrategia, se comparan las reactividades específicas de posición de dos reactivos, anhídrido N-metilisatoico (NMIA) y anhídrido 1-metil-6-nitroisatoico (1M6). El primer reactivo, NMIA, tiene una semivida relativamente larga en solución y reacciona preferentemente con nucleótidos que exprimentan una dinámica lenta. A menudo, estos nucleótidos están en la conformación rara de ribosa C2'-endo y han sido implicados como temporizadores moleculares capaces de gobernar el plegamiento en ARN grandes (Gherghe et al 2008; Mortimer y Weeks 2009). Para el segundo reactivo, el grupo nitro de 1M6 hace que el sistema de dos anillos sea pobre de electrones, y este reactivo es capaz de apilarse con nucleobases de ARN que no están protegidas por las interacciones con otros nucleótidos en una estructura de ARN (Steen et al. 2012). Esta conformación es inusual ya que la mayoría de nucleobases se apilan con otras bases en ambas caras (Leontis et al. 2006). Tomando la diferencia en los perfiles de reactividad para estos dos reactivos selectivos de 2'-hidroxilo, se pueden identificar los nucleótidos implicados en las interacciones estructuralmente distintivas dentro de una estructura de ARN. Debido a que el análisis SHAPE diferencial es específicamente sensible a interacciones no canónicas y terciarias en el ARN (Steen et al. 2012), este enfoque puede ayudar a identificar los nucleótidos que están restringidos (y por lo tanto son no reactivos a 1 M7-SHAPE), pero que no participan en el emparejamiento de bases

canónicas.

[0151] En este documento se proporciona de acuerdo con aspectos de la materia divulgada en el presente documento un término de pseudo-energía libre que incluye información de las reactividades SHAPE diferencial lenta y en apilamiento para producir modelos de estructura secundaria casi perfectos en un experimento concisa que escala a ARN de cualquier tamaño.

RESULTADOS

[0152] Selección de un conjunto de pruebas desafiantes. Para evaluar la utilidad de la incorporación de los datos de SHAPE diferencial en un algoritmo de modelado, se eligió un conjunto de diversos ARN con estructuras secundarias bien establecidas para los que la predicción de estructura secundaria dirigida por SHAPE con un solo reactivo siegue siendo compleja (Tabla 1). Incluían seis dominios de aptámero riboswitch que requieren la unión a ligando para plegarse en sus estructuras aceptadas (los riboswitch TPP, adenina, glicina, cíclico-di-GMP, M-Box y de lisina); cuatro ARN más largos de 300 nucleótidos (nt), incluyendo varios dominios de los ARN ribosomales 16S y 23S de Escherichia coli y; cuatro ARN que contienen pseudonudo; y todos los ARN de los que somos conscientes que contienen hasta un pseudonudo para los que la exactitud de modelado con un solo reactivo 1M7 es < 90% (Cordero et al 2012;. Hajdin et al 2013; Leonard et al 2013; Tabla 1).

[0153] Incorporación de SHAPE diferencial en el modelado de la estructura secundaria. Se realizaron experimentos SHAPE con 1M7, NMIA, y 1M6 en ARN preincubados en presencia de ligando afín, si es apropiado ,pero sin proteína. Basado en el trabajo piloto en tres ARN cortos, las señales de reactividad SHAPE de NMIA y 1M6 se correlacionan fuertemente en la mayoría de posiciones (Steen et al. 2012). Un algoritmo para escalar de ventana se empleó para normalizar localmente entre sí los perfiles SHAPE de NMIA y 1M6 (ver Materiales y Procedimientos, en el presente documento a continuación) y luego se restaron los perfiles normalizados para generar las repesentaciones de la reactividad SHAPE diferencial.

[0154] Se utilizó un enfoque potencial estadístico (Rohl et al 2004; Cordero et al 2012) para evaluar las señales de SHAPE diferencial. Este enfoque infiere una energía libre de la diferencia en las distribuciones de nucleótidos apareados y no apareados. La función de energía fue lineal y demostró ser robusta cuando se sometió a un análisis de jackknife de dejar uno fuera. Durante el ajuste, se evaluaron señales diferenciales negativas de NMIA y 1M6. La señal de amplitud negativa de 1M6 no fue tan altamente correlacionada con carácter monocatenario en los sitios de reactividad diferencial como fue la señal de amplitud positiva. Se evaluaron reactividades SHAPE normalizadas de las reacciones con NMIA y 1M6 y se calcularon las reactividades SHAPE diferenciales (Steen et al. 2012) mediante primero el escalado las reactividades de 1M6 a NMIA sobre una ventana en movimiento y a continuación restando las reactividades de 1M6 de NMIA. Los fuertes mejoras de la reactividad diferencial (> |0,3| unidades SHAPE) se observaron para NMIA y para 1M6. Estos sitios corresponden a los nucleótidos con dinámica lenta y aquellos con una cara disponible para el apilamiento, respectivamente. Las posiciones de nucleótidos que muestran reactividades diferenciales de amplitud positiva fuerte (favoreciendo NMIA) incluían las posiciones 53, 58, 62, 66, 69 y 108.

[0155] El término cambio de pseudoenergía libre de reactividad diferencial para cada nucleótido se tomó como:

$$\Delta GDiff = d \times (\text{señal diferencial de amplitud positiva}), \quad (1)$$

donde d es 2,11 kcal/mol. Esta penalización de energía se añadió a la pseudoenergía libre basada en 1M7 estándar, tal como se aplica en *ShapeKnots* (Low and Weeks 2010; Hajdin et al 2013.); la inclusión de esta penalización mejoró las predicciones para muchos ARN. Para cada modelo de ARN, la precisión de una predicción de estructura secundaria en términos de su sensibilidad (sens; fracción de pares de bases en la estructura aceptada predicha correctamente) y el valor predictivo positivo (ppv; la fracción de pares predicha que se producen en la estructura aceptada) se describen en la Tabla 1.

[0156] Impacto de ΔG_{Dif} sobre el modelado de estructura. En ausencia de restricciones experimentales, el algoritmo mfold predice sólo 10 de las 35 pares de bases (29%) en la estructura aceptada del ARNr 5S de E. coli (SEQ ID NO: 1; Fig 1, izquierda.). La adición de limitaciones 1M7-SHAPE produjo una mejora sustancial: 86% de los pares de bases aceptados estaban presentes en el modelo dirigido por SHAPE. Como es común para las predicciones en este nivel de precisión, la mayor parte de la estructura se modela correctamente. Las excepciones son pares de bases en un elemento, una hélice en un cruce de tres vías (Fig. 1, la estructura central, las posiciones 102-107). Cuando se añadieron datos de SHAPE diferencial como limitaciones, se obtuvo un modelo estructural sustancialmente mejorado (Fig. 1, derecha). Los errores en el modelo basado en SHAPE diferencial son menores e implican la adición de unos pocas pares de bases en la segunda hélice de la estructura cerca del nucleótido 30. Estas pares de bases pueden formarse de hecho en las condiciones de investigación, ya que este ARN se investigó en ausencia de subunidades ribosómicas y proteínas.

[0157] La adición de información SHAPE diferenciañ también mejoró la exactitud de la predicción de la estructura de riboswitch de glicina (SEQ ID NO: 2). Con los datos de 1M7 solamente, el modelo previsto para el riboswitch de glicina tenía 55% sens. y 49% PPV. El principal error en el modelo es la predicción de un pseudonudo falso que propaga

entonces otros errores. La inclusión de la penalización SHAPE diferencial dio lugar a sens. y ppv de 95%. En este caso, el uso de la penalización de reactividad diferencial corrigió errores importantes (por ejemplo, las reactividades diferenciales en las posiciones 12-13 y 112) y eliminó el falso pseudonudo positivo. Además, las reactividades diferenciales de menor magnitud desplazaron el panorama de plegado de los nucleótidos 39-49 para dar lugar a un acuerdo de las estructuras predichas y aceptados.

[0158] La estructura predicha del riboswitch M-Box, a 83% de sensibilidad (Tabla 1), era formalmente el modelo de calidad más bajo en el conjunto de pruebas. Las restricciones de la reactividad diferencial mejoraron la predicción en un solo par de bases con respecto a la estructura predicha usando los datos de 1M7 únicamente (SEQ ID NO: 3). La topología global del ARN M-Box es en gran medida correcta independientemente de la inclusión de información SHAPE diferencial: La unión de tres hélices y todas las hélices principales se predicen correctamente. La mayor diferencia entre las estructuras modeladas y aceptadas se produce en la hélice P1 que conecta los extremos 5' y 3' del ARN. Los nucleótidos en esta hélice son moderadamente reactivos frente a los reactivos SHAPE, lo que sugiere que la hélice P1 no es especialmente estable en las condiciones utilizadas para la investigación de la estructura. En la estructura cristalina que es la base para el modelo aceptado, la hélice P1 se estabiliza mediante tres pares de bases GC (Dann et al. 2007) que no estaban presentes en el transcrito que se analizó por SHAPE. Los datos de SHAPE sugieren que la hélice P1 de secuencia nativa es conformacionalmente dinámica. Para la secuencia de ARN investigada en este trabajo, se infiere que la estructura limitada por SHAPE es esencialmente correcta.

[0159] ARN sensible y no sensible. Para los ARN en el conjunto de pruebas, las predicciones mejoraron significativamente con la adición de datos de SHAPE diferencial o se vieron afectados sólo modestamente. La mejora estructural se define como significativa si la sensibilidad o ppv o ambos aumentaban en al menos el 3%. Siete ARN en el conjunto de datos mostraron una mejora significativa por este criterio (Tabla 1, la parte superior, ARN sensibles). Las estructuras predichas para estos ARN aumentaron en la sensibilidad desde un promedio de 84,5% a un promedio de 93,4%. La mejora de valor predictivo positivo (ppv) fue incluso más sustancial: de 78,1% a 91,2%. De los ARN en la categoría de menos sensible, cuatro de los ocho mostraron pequeñas mejoras en la sensibilidad o ppv (Tabla 1, parte central), y los cambios en la estructura de energía libre más baja implicaron ajustes relativamente menores en el apareamiento de bases con respecto a las estructuras predichas usando sólo los datos de 1M7. Cabe destacar que, a pesar de que las predicciones para múltiples ARN mejoraron mediante la adición de restricciones de SHAPE diferencial, ninguna de las predicciones resultó sustancialmente peor con la excepción del intrón del grupo I de *Tetrahymena* (Tabla 1).

[0160] La estructura modelada para el intrón del grupo I de Tetrahymena resultó menos parecida que la estructura aceptada tras la inclusión de la información de reactividad diferencial: La sensibilidad disminuyó del 93% al 85% (SEQ ID NO: 4; Tabla 1). La hélice P7 comprende un pseudonudo en la estructura del ARN aceptada. Una cadena de la hélice P7 es reactiva por SHAPE y no está presente en el modelo dirigido por SHAPE. Estos datos sugieren que la hélice P7 es conformacionalmente dinámica bajo las condiciones en solución de sondeo utilizadas en este trabajo.

DISCUSIÓN

[0161] El desarrollo de modelos de estructura secundaria precisos para ARN largos es un requisito previo deseable para entender el papel de la estructura del ARN y las interacciones ARN-ligando en la mayoría de las fases de la regulación de genes (Mauger et al. 2013). Además, un modelo de estructura secundaria preciso puede facilitar el modelado de la estructura terciaria (Hajdin et al 2010; Bailor et al 2011). Un enfoque deseable para el modelado de la estructura de ARN debe equilibrar una alta precisión con experimentación concisa y escalable. El modelo termodinámico vecino más próximo desarrollado por Turner y colegas (Mathews y Turner 2006) proporciona una base para el modelado de la estructura secundaria. Sin embargo, hay características de plegamiento del ARN que son difíciles de extraer de la secuencia, incluyendo los efectos de unión de ligando y proteínas, interacciones de estructura terciaria no canónicas y de largo alcance, y la historia cinética de la reacción de plegado del ARN. La inclusión de estructura experimental de un solo reactivo que sondea datos proporciona una mejora sustancial en el modelado de precisión para muchos ARN (Deigan et al 2009; Hajdin et al 2013., pero esta mejora no fue suficiente para producir modelos de estructura secundaria precisos para todos los ARN en nuestro conjunto de pruebas (véase, por ejemplo, Figura 1). Se demuestra en este documento que la inclusión de información de un experimento SHAPE diferencial aumenta sustancialmente la sensibilidad y el valor predictivo positivo de los modelos de estructura secundaria para un conjunto de pruebas de ARN diseñados para ser tan difíciles como sea posible (Tabla 1).

[0162] El contenido de información de un modelado de estructura de ARN dirigida por SHAPE con tres reactivos parece ser superior a la de los enfoques con sondeo químico descritos anteriormente. La adición de sulfato de dimetilo (DMS) y la información de reactividad CMCT en el contexto de un conjunto de datos de seis ARN pequeños produjo una mejora de aproximadamente tres pares de bases en un ARN (Tabla 2; Kladwang et al 2011b; Cordero et al 2012.). En cambio, el experimento con SHAPE diferencial produjo grandes mejoras, estructuralmente significativas, en siete ARN (Tabla 1, parte superior) y menos mejoras drásticas en otros cuatro ARN (Tabla 1, parte central) además de un modelado dirigido por 1M7 con un solo reactivo. Se observó una gran mejora para el ARNr 5S, que no mejoró con la adición de DMS y datos CMCT (Cordero et al. 2012). Además, los modelos desarrollados usando un sondeo SHAPE con tres reactivos tiene precisiones de predicción que igualan o superan las de los enfoques que implican sondear conjuntos amplios de mutantes (Kladwang et al. 2011a). Por lo tanto, los datos de SHAPE diferencial tienen un alto contenido de información

que se obtiene en un experimento conciso que se escala fácilmente para ARN grandes.

[0163] El uso de SHAPE diferencial para la predicción de la estructura secundaria de ARN representa un avance significativo en el modelado de la estructura de ARN. Con la información de SHAPE diferencial, las estructuras de algunas de las moléculas de ARN que se vieron anteriormente como las más difíciles, incluyendo el ARNr 5S, el riboswitch de glicina, y algunos dominios ribosomales, se modelaron en acuerdo casi perfecto con las estructuras aceptadas (Tabla 1). Una tendencia intrigante era que los ARN que eran más sensibles a la penalización de reactividad diferencial fueron aquellos con estructuras predichas más deficientemente en ausencia de información de SHAPE diferencial. Los ARN en esta clase probablemente tienen interacciones no canónicas que se describen de manera incompleta por el algoritmo del vecino más próximo o los datos de un solo reactivo. En varios casos en los que los modelos dirigidos por SHAPE no están de acuerdo con las estructuras aceptadas, las “errores” de los riboswitches de MBox y lisina y el Intrón del grupo I de *Tetrahymena* I, parecen reflejar diferencias entre conformaciones en el cristal y en solución para estos ARN.

[0164] Actualmente, sólo una pequeña base de datos de ARN con estructuras aceptadas bien definidas (Rivas et al 2012; Leonard et al 2013) está disponible. Actualmente hay muy pocos ARN grandes con estructuras complejas cuyas estructuras están bien verificadas. También, los enfoques para el modelado de pseudonudos han avanzado significativamente (Hajdin et al. 2013), pero el modelado preciso de más de un solo pseudonudo en un ARN complejo sigue siendo un desafío, tanto debido a las limitaciones en los modelos actuales de energía como debido a los requisitos computacionales para muchos algoritmos. El presente trabajo se puede utilizar como parte de los esfuerzos para abordar estas cuestiones.

[0165] Aunque el presente trabajo se centró en pares de bases canónicas y no modeló explícitamente pares no canónicos, en muchos casos estas se puede deducir de su falta de reactividad hacia 1M7. Además, los algoritmos de plegado dirigidos por SHAPE en la actualidad incluyen socios de emparejamiento de bases dentro de 600 nt. En general, esta es una buena suposición y, por ejemplo, permite modelar ARN ribosómicos de longitud completa con alta precisión (Deigan et al. 2009). Sin embargo, hay importantes interacciones ARN-ARN que se producen sobre distancias de 1.000 nt o más (Álvarez et al 2005; Jin et al 2011.). Finalmente, las reactividades SHAPE reflejan el conjunto estructural presente en solución en el momento del sondeo. Si un ARN está mal plegado parcialmente o muestra múltiples conformaciones, el perfil SHAPE resultante reflejará estas contribuciones.

[0166] El modelado de estructura secundaria del ARN de alta precisión descrito aquí implica experimentos sencillos con tres reactivos 1M7, 1M6, y NMIA. Este trabajo examinó las estructuras de ARN complejas, incluyendo > 3800 nt, y se centró específicamente en los ARN que se cree que comprende los desafíos de modelado conocidos más difíciles. El sondeo de la estructura SHAPE de tres reactivos es experimentalmente conciso, produce modelos estructurales de ARN consistentemente precisos, y se puede aplicar a ARN de cualquier complejidad y tamaño, incluyendo genomas virales completos y los constituyentes de transcriptomas enteros.

MATERIALES Y PROCEDIMIENTOS

[**0167]** Se ha descrito anteriormente una investigación química mediante datos de SHAPE diferencial para los dominios de aptámero de riboswitch de tiamina pirofosfato (TPP) de *E. coli*, riboswitch de adenina de *Vibrio vulnificus* y riboswitch de lisina de *Thermotoga maritime* (Steen et al. 2012). Se codificaron plantillas de ADN (IDT) para ARNr 5S de *E. coli* y el $ARNt^{Phe}$, riboswitch de glicina de *Fusobacterium nucleatum*, riboswitch M Box de *Bacillus subtilis*, intrón de grupo I de *Tetrahymena thermophila* y los ARN del intrón del grupo II de *Oceanobacillus iheyensis* en el contexto de cassettes de estructura 5' y 3' flanqueantes (Wilkinson et al. 2006), se amplificaron por PCR y se transcribieron en ARN utilizando ARN polimerasa de T7. Los ARN se purificaron usando electroforesis en gel de poliacrilamida desnaturalizante, se escindieron del gel, y pasivamente eluyeron durante la noche a 4 °C. Los ARN ribosomales 16S y 23S se aislaron de células DH5a durante la fase semilogarítmica usando condiciones no desnaturalizantes (Deigan et al. 2009). Los ARN se volvieron a plegar en HEPES 100 mM, pH 8,0, NaCl 100 mM, y $MgCl_2$ 10 mM (Steen et al. 2012). El ARN aptámero de glicina se incubó con glicina 5 μM final durante el plegdo. Después de plegarse, todos los ARN fueron modificados en presencia de reactivo SHAPE 8 mM y se incubaron durante 3 min (1M6 y 1M7) o 22 min (NMIA) a 37 °C. Los controles sin reactivos, que contienen DMSO puro en lugar de reactivo SHAPE, se realizaron en paralelo.

[0168] Después de la modificación y la precipitación con etanol, los ARN reactivo y de control se sometieron a transcripción inversa con kit SUPERSCRIPT III™ (Invitrogen) utilizando cebadores marcados con fluorescencia (colorante VIC, Invitrogen) que reconoció el cassette de estructura 3’ (Wilkinson et al. 2006). Se usó un segundo cebador interno para el intrón del grupo II para leer a través del extremo del ARN. También se realizó una reacción de secuenciación de transcripción inversa usando ddC y un cebador marcado con NED para permitir la alineación de secuencias. Las reacciones con reactivo o de control sin reactivos se combinaron con reacciones de secuenciación y se analizaron usando un instrumento de electroforesis capilar ABI 3500. Los datos resultantes se procesaron utilizando *QuShape* (Karabiber et al. 2013).

[0169] Los ARN ribosomales se analizaron mediante un nuevo enfoque, SHAPEMaP, que se describe en el presente documento a continuación. Para todos los ARNs, se normalizaron las reactividades SHAPE de 1M7 utilizando el enfoque de boxplot (Hajdin et al. 2013). En este enfoque, las reactividades se ordenaron primero, y las reactividades por

encima de 1,5 X intervalo del intercuartil o el percentil 90, el que fuera mayor, fueron excluidos como valores atípicos. A continuación, se calculó un factor de normalización promediando el siguiente 10% de reactividades SHAPE. A continuación, el conjunto de datos original se dividió por el factor de normalización recién calculado para producir los datos procesados finales.

[0170] Análisis de datos SHAPE diferencial. Las reactividades SHAPE con NMIA y 1M6 se normalizaron mediante la exclusión de la parte superior del 2% de reactividades y dividiendo por el promedio del siguiente 8% de reactividades. Las reactividades de 1M6, a continuación, se escalaron con más precisión a reactividades de NMIA minimizando la diferencia de reactividad sobre una ventana corredera de 51 nt. Las reactividades de 1M6 escaladas se restaron de reactividades de NMIA para producir un perfil de SHAPE diferencial. Este algoritmo, implementado en un programa de pitón, se describe en otra parte en este documento.

[0171] Penalización del pesudocambio de energía libre de SHAPE diferencial. Los ARN con estructuras secundarias derivadas de procedimientos de alta resolución (cristalografía o RMN) se utilizaron para clasificar la conformación de nucleótidos como apareados (G-C, A-U o G-U) o no apareados. A continuación, se creó un histograma de reactividades diferenciales (reactividad de NMIA menos reactividad de 1M6) para cada categoría usando una anchura de intervalo de 0,2 unidades de SHAPE. Las reactividades positivas y negativas de SHAPE diferencial fueron tratadas por separado. A continuación, se ajustó un potencial de energía estadística ΔG_{Dif} usando un enfoque análogo a los utilizados ampliamente para el modelado de proteínas (Rohl et al. 2004) y recientemente para el modelado de ARN (Cordero et al. 2012). Los histogramas de nucleótidos diferenciales apareados y no apareados de todos los ARN se agruparon y se ajustaron a una distribución γ. Una energía libre a una temperatura (T) de 310 K se calculó utilizando la relación Gibbs:

$$\Delta GDif = -kbTlnPx \text{ apareado } Px \text{ no apareado}$$

.

P (x) apareado y P (x) no apareado son las probabilidades de que un nucleótido esté apareado o no apareado en reactividad SHAPE x, respectivamente; k_b es la constante de Boltzmann; y ΔG_{Dif} es la penalización del cambio de energía libre resultante que se debe aplicar a una reactividad SHAPE diferencial particular, x. La función resultante fue lineal con una intersección cerca de cero. Para simplificar el cálculo y para hacer la función de energía continua para todas las reactividades diferenciales, ΔG_{Dif} se ajustó a una ecuación lineal con una intersección de cero. Se estimó una medida de error estándar del ajuste mediante un enfoque de jackknife de dejar uno fuera; el ajuste resultante fue una línea con una pendiente de 2,11 kcal/mol y una intersección de cero.

[0172] Exploración de potenciales de energía SHAPE diferencial más simple. Exploramos la posibilidad de omitir el experimento con 1M6 y de cálcular reactividades de SHAPE diferencial basadas únicamente en experimentos con 1M7 y NMIA. Se calcularon las diferencias de reactividad entre NMIA y 1M7 para cada nucleótido usando el algoritmo de sustracción de la diferencia descrito anteriormente. La relación fue lineal con una pendiente de 2,91 kcal/mol. Los errores estándar que resultan de un análisis jackknife de dejar uno fuera fueron de magnitud similar a los de la relación entre las reactividades de NMIA y 1M6. Esta versión de dos reactivos del experimento SHAPE diferencial produjo mejoras significativas al moldeado de estructura secundaria de ARN (Tabla 3); sin embargo, el análisis de tres reactivos en última instancia produjo modelos de estructura más precisos (véase Tabla 1 y Tabla 3). Debido al contenido con mayor información del análisis diferencial de NMIA-1 M6, se sugiere el uso de tres reactivos (1M7, 1 M6, y NMIA) para conseguir deseablemente altas precisiones en el modelado de la estructura secundaria. Durante el transcurso del ajuste de nuevos datos de SHAPE diferencial, el potencial de energía libre de 1M7 también se reajustó usando un potencial estadístico y conjunto de datos de ARN previamente publicado (Hajdin et al. 2013). Las distribuciones de nucleótidos apareados y no apareados se ajustaron a una mezcla de dos distribuciones gamma, y se calculó un término de cambio de energía libre usando la relación de Gibbs. La función de cambio de energía libre resultante fue comparable en magnitud y la intersección con x de la función logarítmica optimizada de búsqueda de rejilla anterior. Por lo tanto, se eligió para su uso la función logarítmica original para incorporar los datos de 1M7 en el modelado de estructura dirigida por SHAPE-dirigida.

[0173] Implementación en RNAstructure Fold y ShapeKnots. Se creó un archivo de energía SHAPE modificada para su uso en *RNAstructure Fold* (Reuter y Mathews 2010) y *ShapeKnots* (Hajdin et al. 2013) para incorporar la información de SHAPE diferencial. Los valores diferenciales de cambio de pseudo energía libre (ΔG_{Diff}) para cada nucleótido se calcularon a partir de las reactividades diferenciales de amplitud positiva (D):

$$\Delta G(d)_{Dif} = \begin{cases} 2.11d & \text{si } d > 0 \\ 0 & \text{si } d \leq 0 \end{cases}$$

Los cambios de pseudoenergía libre SHAPE se calcularon a partir de reactividades de 1M7 usando la ecuación de SHAPE en forma logarítmica (Hajdin et al 2013.):

$$\Delta G_{SHAPE} = 1.8\, ln(\text{SHAPE} + 1) - 0.6$$

Estas dos energías libres se sumaron, y se calculó un archivo de reactividad SHAPE modificada para su uso en *Fold* o *ShapeKnots*, de manera que, cuando se utilizó con pendiente de 1,0 y una intersección de -1,0, el algoritmo de plegado aplica el término de cambio de pseudoenergía libre apropiado:

$$\text{SHAPE} = e^{(\Delta G_{\text{SHAPE}} + \Delta G_{\text{Dif}} + 1)} - 1$$

Las futuras versiones de *ShapeKnots* y *Fold* simplificarán este procedimiento y permitirán que las magnitudes de 1M7 y SHAPE diferencial se puedan introducir directamente desde un archivo de datos. Para *ShapeKnots*, se utilizaron los parámetros de pseudonudo optimizados (P1 = 3,5, P2 = 6,5) (Hajdin et al. 2013). La opción *maxtracebacks* se fijó en 100 y la opción de *ventana* se fijó a 0 para maximizar el número de estructuras identificadas potenciales.

[0174] El cálculo para plegar ARN utilizando 1M7 en lugar de 1M6 como reactivo diferencial se realizó de la misma manera, excepto que la pendiente diferencial fue de 2,91. Los pliegues resultantes se resumen en la Tabla 3. En general, el uso de *ShapeKnots* se sugiere para el modelado de la estructura secundaria de ARN debido a su capacidad para predecir pseudonudos (Hajdin et al 2013.); a nivel práctico, este programa se limita a ARN por debajo de ~700 nt de longitud.

[0175] Representaciones y figuras. Las representaciones de la estructura secundaria se construyeron utilizando *VARNA* (Darty et al. 2009), y se realizaron representaciones de círculos usando *CircleCompare*, una parte de *RNAstructure* (Reuter y Mathews 2010). El modelo sens se calculó como el número de pares de bases correctas dividido por el número total de pares de bases de la estructura aceptada; ppv se calculó como el número de pares de bases correctos dividido por el número total de pares de bases predichas. Los valores sens y ppv para dominios ribosomales se calcularon después de omitir regiones (Deigan et al. 2009) en las que las reactividades SHAPE eran claramente no consistentes con el patrón de apareamiento de bases en el modelo de estructura secundaria aceptado.

		SHAPE	−	+	+	
		diferencial	−	−	*NMIA-1M6*	
	longitud (nts)		sens	ppv	sens	ppv	sens	ppv	
riboswitch TPP, *E. coli*	79		77.3	85.0	96.5	91.3	95.5	100.0	sensible
riboswitch di-GMP cíclico, *V. cholerae*	97		75.0	77.8	89.2	86.2	96.4	93.1	sensible
ARNr 5S, *E. coli*	120		28.6	25.0	85.7	76.9	94.3	91.7	sensible
Riboswitch de glicina, *F. nuleatum*	158		70.0	60.9	55.0	48.9	95.0	95.0	sensible
Dominio III de ARNr 23S, *E. coli*	372		46.9	43.1	82.7	74.3	90.8	83.2	sensible
Intrón grupo I, *T. thermophila*	425		83.3	75.0	93.2	91.2	84.9	89.7	sensible
Dominio 3' de ARNr 16S, *E. coli*	478		26.7	21.2	89.5	77.6	97.1	86.1	sensible
Promedio			**58.3**	**55.4**	**84.5**	**78.1**	**93.4**	**91.2**	
Riboswitch de adenina, *V. vulnificus*	71		100.0	100.0	100.0	100.0	100.0	100.0	no sensible
ARNt phe, *E. coli*	76		100.0	91.3	100.0	75.0	100.0	77.8	no sensible
Riboswitch M-Box, *B. subtilis*	154		87.5	91.3	83.3	90.9	83.3	93.0	no sensible
Riboswitch de lisina, *T. maritime*	174		75.8	84.8	84.9	90.3	84.9	90.3	no sensible
Intrón de grupo II, *O. iheyensis*	412		88.0	97.5	93.2	96.9	92.5	98.4	no sensible
Dominio 5' de ARNr 16S, *E. coli*	530		61.3	57.9	97.8	91.8	97.8	91.8	no sensible
Dominio II de ARNr 23S, *E. coli*	685		87.6	78.6	97.8	87.4	96.8	88.2	no sensible
Promedio			**85.7**	**85.9**	**93.9**	**90.3**	**93.6**	**91.4**	
Promedio global			**72.0**	**70.7**	**89.2**	**84.2**	**93.5**	**91.3**	

Tabla 1. Precisiones del modelado de la estructura secundaria de ARN con la información SHAPE con 1M7 y diferencial. En esta tabla se incluyen todos los ARN bien plegados que contienen hasta un pseudonudo, del cual somos conscientes, para el que una predicción de la estructura secundaria limitada a un solo reactivo 1M7 da lugar a menos del 90% de sensibilidad. Los ARN se indican en base a si el modelado es sensible o no a la información de reactividad diferencial: (parte superior) predicciones que mejoran y (parte inferior) predicciones que muestran pequeños cambios o ningún cambio. Se evaluó si los ARN eran sensibles a los datos de SHAPE diferencial si sens. o ppv cambiaban en menos de un 3%. Se calcularon los promedios por separado para cada clase para todos los ARN juntos.

Fuente		Este trabajo*		Cordero *et al.* 2012		Cordero *et al.* 2012		Cordero *et al.* 2012		Hajdin *et al.* 2013*
Procedimiento		Tres reactivos		NMIA		DMS		NMIA+DMS		1M7
	longitud (nts)	sens	ppv	sens	ppv	sens	ppv	sens	ppv	sens
Riboswitch de adenina ., *V. vulnificus*	71	100.0	100.0	100.0	91.3	100.0	91.3	100.0	91.3	100.0
ARNT phe , *E. coli*	76	100.0	77.8	100.0	95.5	100.0	95.5	100.0	95.5	100.0
Riboswitch TPP , *E. coli*	79	95.5	100.0	--	--	--	--	--	--	95.5
riboswitch di-GMP cíclico , *V. cholerae*	97	96.4	93.1	96.2	92.5	80.7	91.3	96.2	92.5	89.3
ARNr 5S , *E. coli*	120	94.3	91.7	85.3	76.3	85.3	76.3	85.3	76.3	85.3
Riboswitch M-BOx , *B. subtilis*	154	83.3	93.0	--	--	--	--	--	--	87.5
Riboswitch de glicina *F. nuleatum*	158	95.0	95.0	94.9	86.0	97.5	92.8	97.5	92.8	--
Riboswitch de lisina , *T. maritime*	174	84.9	90.3	--	--	--	--	--	--	87.3
Dominio III de ARNr 23S , *E. coli*	372	90.8	83.2	--	--	--	--	--	--	--
Intrón del grupo II , *O. iheyensis*	412	92.5	98.4	--	--	--	--	--	--	93.2
Intrón del grupo I , *T. thermophila*	425	84.9	89.7	--	--	--	--	--	--	93.9
Dominio 3' de ARNr 16S , *E. coli*	478	97.1	86.1	--	--	--	--	--	--	--
Dominio 5' de ARNr 16S , *E. coli*	530	97.8	91.8	--	--	--	--	--	--	--
Dominio II de ARNr 23S , *E. coli*	685	96.8	83.2	--	--	--	--	--	--	--

Tabla 2. Precisiones del modelado de estructura secundaria de ARN comparando SHAPE diferencial con tres reactivos con trabajos recientes relacionados. Los enfoques que permiten pseudonudos se indican con un asterisco. Los procediimentos que utilizaron parámetros optimizados utilizando grupos de datos pequeños se indican con una cruz.

		SHAPE	–	+	+	
		diferencia	–	–	*NMIA-1M7*	
	longitud	(nts)	sens ppv	sens ppv	sens ppv	
RIboswitch TPP , *E. coli*	79		77.3 85.0	96.5 91.3	95.5 100.0	sensible
ARNr 5S , *E. coli*	120		28.6 25.0	85.7 76.9	94.3 91.7	
Riboswitch de glicina , *F. nuleatum*	158		70.0 60.9	55.0 48.9	97.5 95.1	
Intrón del grupo I , *T. thermophila*	425		83.3 75.0	93.2 91.2	82.6 88.7	
Dominio 3' de ARNr 16S , *E. coli*	478		26.7 21.2	89.5 77.6	97.1 86.8	
Dominio 5' de ARNr 16S , *E. coli*	530		61.3 57.9	97.8 91.8	88.3 83.4	
Promedio			**57.9 54.2**	**86.3 79.6**	**92.5 91.0**	
Riboswitch de adenina , *V. vulnificus*	71		100.0 100.0	100.0 100.0	100.0 100.0	no sensible
ARNt phe , *E. coli*	76		100.0 91.3	100.0 75.0	100.0 77.8	
Riboswitch de di-GMP cíclico , *V. cholerae*	97		75.0 77.8	89.2 86.2	89.2 86.2	
Riboswitch M-Box , *B. subtilis*	154		87.5 91.3	83.3 90.9	83.3 93.0	
Riboswitch de lisina , *T. maritime*	174		75.8 84.8	84.9 90.3	84.9 90.3	
Dominio III de ARNr 23S , *E. coli*	372		46.9 43.1	82.7 74.3	82.7 74.3	
Intrón del grupo II , *O. iheyensis*	412		88.0 97.5	93.2 96.9	94.0 98.4	
Dominio II de ARNr 23S , *E. coli*	685		87.6 78.6	97.8 87.4	97.8 87.9	
Promedio			**82.6 83.0**	**91.4 87.6**	**91.5 88.5**	
Promedio global			**72.0 70.7**	**89.2 84.2**	**91.9 89.5**	

Tabla 3. Precisiones del modelado de estructura secundaria del ARN para un experimento SHAPE diferencial con dos reactivos utilizando 1M7 y NMIA. Los ARN se indican según si la predicción de la estructura era sensible o no a las reactividades diferenciales NMIA-1M7. El experimento con dos reactivos produjo mejoras significativas en el modelado en relación con los datos de 1M7 solo, pero las mejoras no fueron tan amplias como las del experimento con tres reactivos (tabla 1).

Ejemplo 4

Sondeo químico de ARN correlacionado con una molécula individual

[0176] Las moléculas de ARN funcionan como el conducto central de la transferencia de información en biología. Para hacer esto, que codifican información tanto en sus secuencias como en sus estructuras de orden superior. La comprensión de la estructura de orden superior del ARN sigue siendo un reto. En este Ejemplo se proporciona un enfoque simple, experimentalmente conciso y exacto para el examen de la estructura del ARN de orden superior mediante la conversión de secuenciación masiva en paralelo ampliamente utilizada en un experimento con una sola molécula de fácil aplicación para la detección de interacciones a través del espacio y múltiples conformaciones. Este enfoque se emplea entonces para analizar la estructura del ARN de orden superior, detectar estados ocultos biológicamente importantes, y refinar los modelos de estructura tridimensionales precisos.

[0177] Las funciones del ARN están mediadas por niveles estratificados de información: el más sencillo es la secuencia primaria, y la más compleja es la estructura de orden superior que rige las interacciones con ligandos, proteínas y otros ARN (Sharp PA (2009) Cell 136.: 577-5802; Leontis NB, et al (2006) Curr Opin Struct Biol 16: 279-287). Muchos ARN pueden formar más de una estructura estable, y estas conformaciones distintas a menudo tienen diferentes actividades biológicas (Montange RK, Batey RT (2008) Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys. 37: 117-133; Dethoff EA, Chugh J, Mustoe AM, Al-Hashimi HM (2012) Functional complexity and regulation through RNA dynamics. Nature 482: 322-330). Actualmente, la velocidad de describir nuevas secuencias de ARN supera con mucho la capacidad de examinar sus estructuras.

[0178] En este ejemplo se describe la caracterización de interacciones a través del espacio y múltiples conformaciones en ARNs individuales mediante la fusión del sondeo químico y la secuenciación masivamente en paralelo. Debido a que la secuenciación masivamente en paralelo informa de las secuencias de plantillas individuales, cada lectura es fundamentalmente una observación de una sola molécula (Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol. 26: 1135-1145). Nosotros modificamos primero el ARN con un reactivo que es sensible a la estructura del ARN subyacente y, a continuación, se detectaron múltiples aductos en cadenas de ARN individuales (Figs. 2A y 2B). Los aductos químicos se detectaron como mutaciones de secuencia en base a su capacidad para inducir una mala lectura eficiente del nucleótido plantilla por una enzima transcriptasa inversa, un enfoque denominado perfilado o perfil mutacional, o MaP (descrito en otra parte en este documento). Los datos de sondeo de una sola molécula se utilizaron de dos formas distintas: para detectar modificaciones de ARN correlacionadas que reflejan interacciones a través del espacio de orden superior (Fig 2A.) y para examinar múltiples conformaciones en conjuntos individuales en solución (Fig 2B.).

RESULTADOS Y DISCUSIÓN

[0179] Reactividad del sulfato de dimetilo multisitio con ARN. Se utilizó sulfato de dimetilo (DMS) para investigar las estructuras de tres ARN: riboswitch de pirofosfato de tiamina (TPP) de *Escherichia coli* (79 nt) (Serganov A, et al (2006) Nature 441: 1167-1171), el dominio P546 de intrón del grupo I de *Tetrahymena (*160 nt) (Cate JH et al (1996) Science 273: 1678-1685), y el dominio catalítico de ARNasa P de *Bacillus stearothermophilus* (265 nt) (Kazantsev AV, et al (2009) RNA. 15: 266-276). Los ARN fueron seleccionados para ilustrar las distintas características de plegado del ARN y para enfatizar retos de análisis cada vez más difíciles.

[0180] El riboswitch de TPP se une al ligando TPP para actuar en la regulación génica. El dominio P546 tiene una estructura en forma de U estabilizada por una interacción terciaria de largo alcance que abarca aproximadamente 45 pares de bases. El dominio de la ARNasa P se encuentra entre los ARN más grandes para los que se ha perseguido el modelado automatizado. El DMS forma aductos en la posición N1 de la adenosina y en la posición N3 de la citosina. Hemos optimizado las condiciones para producir múltiples modificaciones en una cadena de ARN sin interrumpir el plegado del ARN nativo. Las muestras de ARN se trataron con DMS 170 mM en 10 mM de Mg2+ y 300 mM de cacodilato en tampón de pH 7 durante 6 min. Las reacciones se inactivaron mediante la adición de un exceso de 2-mercaptoetanol. Los nucleótidos de citosina y adenosina se metilaron con igualdad de eficiencia, y un ~12% de los nucleótidos se modificaron en estas condiciones.

[0181] Hemos detectado sitios de metilación con DMS directamente como mutaciones inducidas por aducto en los ADNc de longitud completa generados durante la transcripción inversa (Figs. 2A y 2B). Visualizamos los patrones generales de reactividad para cada ARN en los perfiles de frecuencia de mutaciones en 2D (Fig. 3B). La comparación con las estructuras de alta resolución para estos ARN (Serganov A, et al, (2006) Nature 441 (7097): 1167-1171; Cate JH, et al, (1996) Science 273 (5282): 1678-1685; Kazantsev AV, Krivenko AA, Pace NR (2009) RNA 15 (2): 266-276) mostró que los nucleótidos modificados a niveles altos eran aquellos que no participan en el apareamiento de bases o interacciones terciarias (figura 3C). Como era de esperar, la modificación de nucleótidos era dependiente de la concentración de reactivos, tiempo de reacción y la estabilidad de la estructura del ARN. Las altas frecuencias de mutaciones observadas sobre la base permitió analizar las mutaciones inducidas por DMS sin el requisito de corrección de la base (fondo) (Fig. 3B).

[0182] Para el riboswitch de TPP, el dominio P546, y el dominio de la ARNasa P, se detectaron promedios de dos, cinco y siete aductos en cada lectura de secuenciación, respectivamente (Fig. 3A). Aproximadamente el 15% de los nucleótidos A y C de una cadena en cada ARN se modificaron mediante DMS, comparable con el nivel de modificación de los nucleótidos libres en estas condiciones. Debido a que se detectaron múltiples eventos de modificación química en las lecturas de secuenciación de cadenas de ARN individuales, se pudieron cuantificar interdependencias correlacionadas en la reactividad. Los nucleótidos modificados de una forma correlacionada comprenden grupos de interacción de ARN, o RING. Las reactividades correlacionadas se midieron mediante perfiles mutacionales (MaP) (descrito en otra parte en el presente documento), produciendo un experimento RING-MaP.

[0183] Interacciones de ARN a través del espacio detectadas mediante análisis estadístico de asociación. Los nucleótidos implicados en las interacciones a través del espacio mostrarán reactividades químicas correlacionadas, que refleja un mecanismo de "respiración" en el que un nucleótido de ARN se vuelve transitoriamente accesible para su modificación. Este mecanismo de respiración sugiere que el sondeo correlacionado será selectivo para interacciones transitorias, dinámicas, en lugar de diferencias estructurales estáticas (Fig. 2A).

[0184] Se utilizó una estrategia de dos partes para identificar pares de nucleótidos reactivos con correlaciones estadísticamente significativas y cuantificar las fuerzas de estas correlaciones. Las interdependencias para reactividades con DMS para cualquiera de dos posiciones en una sola cadena de ARN se evaluaron primero mediante una prueba de $\chi 2$. A continuación, la fuerza de la interacción entre cada par de nucleótidos correlacionados se cuantificó usando la medida de phi de Pearson.

[0185] Los RING para el riboswitch TPP, dominio de intrón P546 y ARN de ARNasa P (Fig. 3C) incluyen nucleótidos que se sabe que interactúan sobre la base de estructuras de alta resolución (Serganov A, et al, (2006) Nature 441 (7097): 1167-1171; Cate JH, et al, (1996) Science 273 (5282): 1678-1685; Kazantsev AV, Krivenko AA, Pace NR (2009) RNA 15 (2): 266-276). Por ejemplo, las posiciones correlacionadas en el riboswitch de TPP corresponden a los nucleótidos implicados en una interacción de acoplamiento entre el bucle L5 y la hélice P3 y en la formación del bolsillo de unión a ligando. En el dominio de P546, se observaron modificaciones correlacionadas en los nucleótidos en la interacción de acoplamiento del bucle L5b y la hélice P6a, dentro de la región bisagra J5, y a lo largo de las longitudes de las hélices P5a y P5B. En el ARN de la ARNasa P, RING describe interacciones terciarias entre los bucles L5 y L15.1 y en el núcleo estructural. También se observó un segundo conjunto de interacciones en el elemento de la ARNasa P P19.

[0186] RING describe interacciones de ARN de orden superior y terciario. Se evaluaron las diferencias en las interacciones terciarias, según lo descrito por RING, bajo diferentes condiciones de solución o en ARN mutantes. En presencia de Mg^{2+}, el dominio P546 forma una estructura en forma de U en la que se forma una interacción tetrabucle-receptor entre L5b y P6a y la región J5 actúa como una bisagra (Murphy FL, Cech TR (1994) J Mol Biol 236 (1): 49-63; Szewczak AA, Cech TR (1997) RNA 3 (8): 838-849). Estas interacciones se describieron correctamente mediante múltiples modificaciones químicas correlacionadas en estos elementos estructurales. La alteración de esta estructura terciaria plegando el ARN en ausencia de Mg^{2+} eliminó la mayoría de las interacciones observadas. La estructura terciaria del dominio P546 también puede verse perturbada por mutaciones en la hélice P6a y en la bisagra J5. La mutación del par de bases C223-G250 en la hélice P6a a A-U altera la interacción L5-P6a (Murphy FL, Cech TR (1994) J Mol Biol 236 (1): 49-63). El análisis RING de este mutante mostró que la correlación entre L5 y P6b se perdió y que otras partes del ARN también experimentaron una reconfiguración significativa. Las interacciones que implicaban la bisagra parecían verse reforzadas y se observaron correlaciones más y más fuertes dentro de los dominios helicoidales de P5a y P5b.

[0187] Las mutaciones que dan como resultado el apareamiento de bases de nucleótidos en la bisagra J5 probablemente producen una conformación lineal para el dominio P546 (Szewczak AA, Cech TR (1997) RNA 3 (8): 838-849). El análisis RING del mutante J5 mostró la pérdida esperada en los nucleótidos correlacionadas dentro de la región J5 y por consiguiente la pérdida de la interacción L5b-P6a. Las correlaciones entre los nucleótidos en la hélice P5b se reforzaron en este mutante con respecto a las observadas en el ARN de tipo salvaje, pero no se observaron cambios en las correlaciones entre los nucleótidos en la hélice P5a. Este análisis del dominio P546 demostró que los RING reflejan con precisión las interacciones estructurales en una gran molécula de ARN a la resolución de nucleótidos.

[0188] Detección de múltiples y ocultas conformaciones de ARN. En el enfoque RINGMaP, cada cadena de ARN se secuencia independientemente mediante secuenciación masiva en paralelo. Las cadenas de ARN de diferentes conformaciones tenderán a exhibir grupos distintos de nucleótidos coreactivos (Fig. 2B). Tales grupos pueden ser detectados por agrupación espectral y serán un reflejo estructuras individuales relativamente estables distintas en solución. El agrupamiento espectral produce una estimación objetiva del número de grupos y por lo tanto el número de conformaciones adoptadas por un ARN particular en solución.

[0189] El análisis del agrupamiento espectral de los datos de modificación obtenidos en el riboswitch TPP y los ARN de ARNasa P indicó que cada ARN formó múltiples conformaciones distintas en las condiciones utilizadas en nuestros experimentos de sondeo. RING identificados para el riboswitch TPP con ligando saturante revelaron interacciones en la estructura acoplada L5-P3 atracado estructura y en el bolsillo de unión a ligando (Fig. 4A). Hubo significativamente menos interacciones terciarias internucleotídicas en ausencia de ligando TPP que en su presencia; sin embargo, aún se

observaron las interacciones específicas en J2-4 (Fig. 4B). El agrupamiento espectral reveló que tanto el ARN con ligando saturante como los ARN sin ligando son estados compuestos con conformaciones mayores y menores constituyentes (Figs. 4C y 4D). El agrupamiento menor en el ARN con ligando saturante se caracteriza por un aumento de reactividades de DMS precisamente en las posiciones que resultaron reactivas cuando no se unía ligando (Figs. 4A y 4C, círculos abiertos). Por lo tanto, incluso en condiciones de ligando saturante, el ARN de riboswitch TPP muestra conformaciones características de ambos estados unido a ligando y no unido. En ausencia de ligando, el agrupamiento mayor tiene un patrón de reactividad de DMS similar al estado menos estructurado en presencia de ligando. En cambio, el agrupamiento menor detectado en ausencia de ligando tiene reactividades de DMS reducidas precisamente en el bolsillo de unión a tiamina, sugestivo de una conformación que es más altamente estructurada que la del agrupamiento mayor (Figs. 4B y 4D). Deducimos que, en ausencia de ligando, el riboswitch de bolsilla de unión a tiamina muestra una estructura preplegada "oculta" similar a la formada tras la unión a ligando. A continuación, se probó el ARN del riboswitch TPP en presencia de concentraciones subsaturantes de ligando [200 nM de TTP; Kd ~50-200 nM (Kulshina N, Edwards TE, Ferré-D'Amaré AR (2010) RNA 16 (1): 186-196)]. El análisis de la agregación de los datos de reactividad química produce tres agrupamientos bien definidos en la relación de 1:1:1,2 que corresponde a (i) el estado unido a ligando totalmente plegado, (ii) el estado en que el bolsillo de unión a ligando está estructurado, pero el resto del ARN muestra interacciones internucleotídicos débiles, y (iii) el estado relativamente no estructurado con sólo unos pocos nucleótidos que interactúan (Figs. 5A y 5B). Cada uno de estos agrupamientos corresponde a los estados identificados anteriormente, ya sea en los ARN con ligando aturante o sin ligando. Por lo tanto, el análisis de agrupamiento espectral identificó múltiples conformaciones distintas de un conjunto de ARN individual en solución, incluyendo un estado no caracterizado previamente en el que el bolsillo de unión a ligando esá preplegado. Este estado parcialmente plegado es probablemente importante para el reconocimiento del ligando TPP. Por último, se analizó la estructura del ARN de ARNasa P en función de Mg^{2+}. Las redes de interacciones fueron sorprendentemente diferentes en presencia y ausencia de Mg^{2+}. Las fuertes redes de interacciones entre L5 y L15.1 y en el núcleo estructural desaparecieron en ausencia de Mg^{2+} y fueron reemplazadas por interacciones entre P5.1 y P2 y dentro de P7. El agrupamiento espectral identificó dos grupos en el estado más Mg^{2+}. El agrupamiento menor en la muestra más Mg^{2+} es distinto de la estructura de ARN totalmente plegada y la estructura sin Mg^{2+}. Los nucleótidos reactivos en el agrupamiento menor comprenden la L5-L15.1 e interacciones del núcleo estructurales, lo que indica que estas interacciones se debilitan en este estado. Críticamente, el análisis del agrupamiento espectral de una sola molécula muestra que el riboswitch TPP y los ARN de ARNasa P adoptan de forma nativa múltiples estados únicos, incluso en condiciones generalmente asumidas para promover la formación de una sola estructura.

[0190] Principios del plegado de ARN. La formación de la estructura de ARN puede ser útilmente aproximada suponiendo que hélices estables, formadas localmente y estabilizadas por emparejamiento de Watson-Crick, se organizaron posteriormente en una estructura 3D mediante interacciones terciarias de mayor alcance. El análisis de RING de los tres ARN estudiados aquí es consistente con esta jerarquía estructura bien establecida. Por ejemplo, observamos RING que reflejan pares de bases no canónicas e interacciones terciarias bucle-hélice y bucle-bucle que se han observado ampliamente en estudios estructurales anteriores (Butcher SE, Pyle AM (2011) Acc Chem Res 44 (12): 1302-11; Brion P, Westhof E (1997) Annu Rev Biophys Biomol Struct 26: 113-137).

[0191] El análisis RING también identificó interacciones cuya prevalencia fue previamente no completamente apreciada. Aproximadamente un tercio de todas las interacciones correlacionadas implican nucleótidos de cadena sencilla o de bucle en los extremos opuestos de las hélices individuales. Estas interacciones a través de la hélice significan que la comunicación estructural en el ARN se puede extender a través de largas distancias. En algunos casos, el acoplamiento estructural a través de la hélice se extiende a través de múltiples hélices apiladas. Además, el análisis RING indica que las interacciones terciarias no son independientes, sino que son fuertemente dependientes de otros elementos estructurales. Hemos observado interacciones acopladas entre motivos de estructura terciaria individuales bien definidos tanto en el ARN del riboswitch TPP como en el ARN del dominio P546. La alteración de cualquier interacción terciaria, por exclusión de ligando o por mutación, dio como resultado la pérdida del propio motivo terciario y también alteró otras interacciones. Los datos de RING también apoyan la importancia del empaquetamiento helicoidal próximo. Estas interacciones son especialmente obvias en el riboswitch TPP y en el núcleo estructural del ARN de la ARNasa P.

[0192] Los análisis de los riboswitch TPP, el dominio P546, y el dominio de la ARNasa P indican que las mutaciones o ausencia de ligando (Figs 4 y 5) o iones divalentes no simplemente "sustraen" una interacción de la estructura, sino que causan un reorganización a gran escala del plegado de ARN. Ninguno de los estados no plegados o menos plegados caracterizados era simplemente una versión menos estructurada de un estado completamente plegado. En su lugar, encontramos que los estados menos estructurados son estabilizados por conjuntos únicos de interacciones interdependientes que, en general, no se han detectado en los estudios de conjunto o de moléculas individuales anteriores.

[0193] Refinamiento de la estructura tridimensional de ARN. Debido a que el análisis ANILLO identifica matrices densas de interdependencias de nucleótidos que reflejan la estructura terciaria del ARN, hemos explorado si estas interacciones se podrían utilizar como restricciones para modelar los plegamientos de ARN 3D. Un pequeño número de limitaciones, que reflejan la estructura del ARN a través del espacio, son a menudo suficientes para producir modelos de estructuras de alta calidad (Gherghe CM, et al, (2009) J Am Chem Soc 131 (7): 2541-2546; Lavender CA., et al, (2010) Biochemistry 49 (24): 4931-4933). Se utilizó un potencial de interacción de dos etapas para introducir bonos de energía

libre cuando los nucleótidos constituyentes se aproximan durante una simulación de dinámica molecular discreta (Gherghe CM, et al, (2009) J Am Chem Soc 131 (7): 2541-2546; Lavender. CA, et al, (2010) Biochemistry 49 (24): 4.931-4.933; Ding F, et al (2008) RNA 14 (6): 1164-1173). La introducción de restricciones RING causó que cada ARN mostrara preferentemente estados colapsados durante la simulación. Después de filtrar por radio de giro, se selecionaron las estructuras representativas mediante agrupación jerárquica. Para cada ARN caracterizado, obtuvimos modelos de alta calidad y estadísticamente significativos (Hajdin CE, et al, (2010) RNA 16 (7): 1340-1349) que correctamente recapitularon la arquitectura del ARN definidas por estructuras de alta resolución (Serganov A, et al, (2006) Nature 441 (7097): 1167-1171; Cate JH, et al (1996) Science 273 (5282): 1678-1685; Kazantsev AV, AA Krivenko, Pace NR (2009) RNA 15 (2): 266-276). Para el ARN de ARNasa P, se observó RING solapantes que abarcaban dos/tercios de la molécula y un segundo conjunto no solapante en P19 (Fig. 3C); esto sugiere que el elemento P3-P2-P19 no está estructuralmente unido con el resto de la arquitectura de ARN. La precisión del modelo 3D para el ARN de la ARNasa P es especialmente alta en el núcleo estructural (excluyendo el elemento P3-P2-P19) con un rmsd de 14,4 A ($P < 10-6$) en comparación con la estructura cristalina. Para los tres ARN estudiados aquí, la significación estadística relativa del modelado de estructuras realmente aumentó con el tamaño, probablemente debido a un número mayor de modificaciones químicas, y por lo tanto más interacciones de nucleótidos se detectan con ARN de mayor tamaño (Fig. 3A). De este modo, las interacciones de la red de RING hacen posible tanto la identificación de novo de elementos estructurales como el modelado de dominios plegdos para ARN grandes.

[0194] Perspectiva. El análisis de la estructura de una sola molécula mediante sonde de química correlacionada, tal como se detecta por secuenciación, representa un enfoque muy simple y genérico para el análisis de las arquitecturas globales de ARN o ADN funcionalmente importantes. La idea fundamental de que un experimento de una sola molécula puede crearse mediante simplemente el registro de múltiples eventos en la misma cadena de ARN o ADN es completamente general y debe inspirar el desarrollo de numerosas nuevas clases de experimentos y descubrimientos biológicos. RING-MaP es único en su sencillez y concisión experimental y puede ser aplicado a prácticamente cualquier ARN biológico, sin el requisito para introducir mutaciones, optimizar el análisis biofísico, o introducir sondas estructurales artificiales. El enfoque de obtención del perfil mutacional de una sola molécula (MaP) que se describe aquí utilizando DMS se puede extender fácilmente a otros agentes de modificación de ARN, a experimentos que cuestionan los cuatro nucleótidos simultáneamente, a la reticulación ARN-proteína y al análisis de bibliotecas complejas de mutantes. Los experimentos MaP de una sola molécula que exploran interacciones proteína-ARN, ARN-ARN y mediadas por ADN simultáneamente son claramente factible.

[0195] La estructura de orden superior está estrechamente vinculada a la función biológica. Por lo tanto, de forma análoga al nuevo descubrimiento basado en la caracterización de la estructura secundaria (Ejemplo 5 en este documento), la identificación de grupos a través RING espaciales en grandes ARN y transcriptomas permitirá el descubrimiento generalizado de motivos funcionales biológicos.

Procedimientos

[0196] Las descripciones detalladas del enfoque RING-MaP, el análisis de asociación estadística, la agrupación espectral, y el modelado de la estructura se proporcionan en el presente documento. Los datos procesados y el software están libremente disponibles en el sitio web del autor correspondiente (chem.unc.edu/rna). Están disponibles los datos de secuenciación en el National Center for Biotechnology Information Sequence Read Archive

[0197] Procedimientos SI de caracterización de la reacción entre sulfato de dimetilo y nucleobases de ARN. La formación de aducto entre sulfato de dimetilo (DMS) (Sigma-Aldrich) y ATP, CTP y UTP marcado con [γ-32P] se realizó mediante la adición de DMS al 10% (vol/vol) (1 ul; 1,7 M en etanol absoluto) a [γ-32P] NTP en 1 x tampón de reacción [9 µl, MgCl2 10 mM y cacodilato de sodio 300 mM (pH 7,0)] a 37 °C. Las reacciones se inactivaron con un volumen igual de 2-mercaptoetanol (2ME) puro después de 10, 30, 60, 120, 180, 360, y 900 s. Para reacciones de control de preinactivación, se añadió primero una solución 1,3 M de DMS [1:2:5 (vol/vol) DMS:etanol:H2O] a un volumen igual de 2-mercaptoetanol puro. Esta mezcla (2,8 µl, DMS 625 mM) se añadió inmediatamente a [γ-32P]NTP en 7,2 µl de 1,4x tampón de reacción [7,2 µl; MgCl2 14 mM y cacodilato de sodio 417 mM (pH 7,0)], y la reacción se incubó a 37 °C durante 15 min. Las reacciones desactivadas se separaron por electroforesis en gel (poliacrilamida al 30%; 29:1 acrilamida:bisacrilamida; gel 0,4 mm x 28,5 cm x 23 cm; 30 W, 45 min) y se cuantificó mediante “fosforoimágenes”. Los datos fueron consistentes con un mecanismo en el que DMS forma aductos en la posición N1 de la adenosina y la posición N3 de la citosina y no reacciona con uridina. El cambio en el pH durante la formación de aducto de DMS fue seguido en reacciones sin NTP a 37 ° C utilizando un medidor de pH Accument 25. Las mediciones directas de la formación de aducto De3 MS con citosina y adenosina sugieren reactividades más o menos iguales con estos dos nucleótidos. Esta observación difiere de la opinión generalizada de que la adenosina reacciona más rápidamente que la citosina con DMS. Atribuimos este concepto erróneo a la capacidad relativamente ineficiente de la N3-metil citosina para inhibir las enzimas de la transcriptasa inversa.

[0198] Construcciones de ARN. Plantillas de ADN para el riboswitch de tiamina pirofosfato (TPP) de *Escherichia coli*, dominio P546 de intrón I del grupo de *Tetrahymena* y ARN de dominio catalítico de ARNasa P de *Bacillus stearothermophilus*, cada uno dentro de secuencias flanqueantes de casete de estructura 5' y 3' se generaron mediante PCR (Wilkinson, KA, et al, (2006) Nat Protoc 1 (3): 1610-1616). Los ARN se transcribieron in vitro [1 ml; Tris 40 mM (pH 8,0), $MgCl_2$ 10 mM, DTT 10 mM, espermidina 2 mM, Triton X-100 al 0,01% (vol/vol), poli(etilen)glicol 8000 al 4%

(peso/volumen), 2 mM de cada NTP, 50 µl de plantilla generada por PCR, 0,1 mg/ml de ARN polimerasa de T7; 37 °C; 4 h] y se purificó mediante electroforesis en gel de poliacrilamida desnaturalizante (poliacrilamida al 8%, urea 7 M, 29:1 acrilamida:bisacrilamida, gel 0,4-mm x 28,5 cm x 23 cm; 32 W, 1,5 h). Los ARN se escindieron del gel, se recuperaron mediante elución pasiva durante la noche a 4 °C, y se precipitaron con etanol. Los ARN purificados se resuspendieron en 50 µl de Tris 10 mM (pH 7,5), EDTA 1 mM (TE) y se almacenaron a -20 °C.

[0199] Plegado o plegamiento de ARN y modificación con DMS. Se realizaron experimentos de investigación de la estructura del ARN en MgCl2 10 mM y cacodilato 300 mM a pH 7,0. Los ARN [5 pmol en 5 ul de Tris 5 mM (pH 7,5), EDTA 0,5 mM (1/2x TE)] fueron desnaturalizados a 95 °C durante 2 min, se enfriaron en hielo, se trataron con 4 µl de 2,5x tampón de plegado [cacodilato 750 mM (pH 7,0) y $MgCl_2$ 25 mM], y se incubaron a 37 °C durante 30 min. Después de plegarse, el dominio P546 y los ARN del dominio catalítico de la ARNasa P se trataron con DMS (1 ul; 1,7 M en etanol absoluto) y se dejaron reaccionar a 37 °C durante 6 min. Las reacciones de control sin reactivo se realizaron con 1 ul de etanol absoluto. Las reacciones se inactivaron mediante la adición de un volumen igual de 2ME puro e inmediatamente se colocaron en hielo. Los experimentos sin Mg2+ se realizaron de forma idéntica, excepto que el 2,5 x tampón de plegamiento omitió el Mg2+. El ARN de riboswitch TPP se incubó en tampón de plegado a 37 °C durante 10 min, después de lo cual el ligando TPP se añadió a la concentración deseada y las muestras se incubaron a 37 °C durante 20 min. (Nota: DMS es un conocido carcinógeno y 2ME puro tiene un olor muy fuerte Las manipulaciones que implican DMS y 2ME se debe realizar en una campana química de humos. Las soluciones que contienen DMS deben ser neutralizadas con NaOH 5 N. Las soluciones que contienen DMS o 2ME deben desecharse como residuos químicos.)

[0200] Transcripción inversa y detección de aductos. La estrategia de la transcripción inversa y la detección de aductos fue adaptada del enfoque de obtención de perfiles mutacionales (MaP) SHAPE (Ejemplo 5 en este documento). Después del tratamiento con DMS, los ARN se purificaron usando columnas de centrifugación G-50 (GE Healthcare). Las reacciones de transcripción inversa se realizaron utilizando transcriptasa inversa SuperScript II (Invitrogen) durante 3 horas a 42 °C [dNTP mezclados previamente 0,5 mM, Tris HCl 50 mM (pH 8,0), KCl 75 mM, MnCl2 6 mM y DTT 10 mM]. Las reacciones se desalaron con columnas de centrifugación G-50 (GE Healthcare). En estas condiciones (tiempo largo de incubación, en presencia de Mn2+), la transcriptasa inversa lee aductos de metilo en las posiciones N1 y N3 de la adenosina y citosina, respectivamente, produciendo una mutación en el sitio del aducto. Las bibliotecas de ADN de doble cadena con los adaptadores y los índices compatibles con la secuenciación basada en Illumina se generaron mediante PCR. Las bibliotecas resultantes se agruparon y se secuenciaron con un instrumento Illumina MiSeq™ (kit de 500 ciclos) de modo que la lectura de la primera secuenciación en modo de extremos apareados cubría la secuencia de ARN de interés. Los archivos de datos FASTQ resultantes se alinearon con las secuencias de referencia y se calcularon las tasas de mutación por nucleótido utilizando un canal interno (Ejemplo 5 en este documento). Se requería que las puntuaciones Phred para contar las mutaciones fueran ≥ 20.

[0201] Medición de interacciones internucleotídicas mediante análisis de asociación estadística. Para detectar interdependencias de reactividad de nucleótidos, todos los posibles pares de nucleótidos se sometieron a la versión corregida de Yates de la prueba $\chi 2$ de Pearson de independencia frente a asociación (Yates F (1934) Supl J Roy Stat Soc 1: 217-235). La estadística de chi cuadrado de Yates corregido se calculó como:

$$\chi^2_{Yates} = \frac{N\left(|ad-bc|-0.5N\right)^2}{(a+b)(c+d)(a+c)(b+d)} \qquad (1)$$

donde $N = (a + b + c + d)$ es el número total de cadenas en el conjunto de datos, y a, b, c, y d se definen por la siguiente tabla de contingencia 2 × 2 de los números observados de cuatro posibles coposibilidades:

Nucleótido *i* / Nucleótido *j*	No mutado 0	Mutado 1
No mutado 0	*a*	*b*
Mutado 1	*c*	*d*

Se tomó que un par de nucleótidos tenían una asociación estadísticamente significativa si $\chi_{Yate}^2 > 20$ ($p < 0{,}00001$). Con este alto umbral de aceptación para un par de nucleótidos individuales esperamos realizar no más de una determinación de falsos positivos para los ARN de al menos hasta 500 nt de longitud.

[0202] Para pares de nucleótidos que pasaron la prueba de significación χ2a, el signo y la fuerza de la asociación estadística se determinó mediante el cálculo de coeficiente de correlación de Pearson, p. En el caso de dos variables binarias, p es igual a la medida de asociación de Pearson, el coeficiente phi. El coeficiente de correlación y la estadística de chi-cuadrado están relacionadas:

$$\rho^2 = \chi^2 / N \qquad (2)$$

Aunque los coeficientes de correlación eran típicamente inferiores a 0,05, los coeficientes fueron altamente significativos. Sobre la base de la estadística de χ^2, la probabilidad de que los nucleótidos correlacionados identificados fueran independientes fue de menos de 0,00001.

[0203] Las siguientes directrices se impusieron también para el análisis de asociación de nucleótidos y la agrupación. Se requirió que el número promedio de modificaciones detectadas por lectura fuera del ~ 15% del número estimado de nucleótidos de cadena simple, dando un promedio igual a o mayor que dos mutaciones por lectura. Los nucleótidos con una tasa de mutaciones mayor que 0,05 en el control sin modificación se excluyeron del cálculo de χ2. Los pares de nucleótidos correlacionados con una SD de su coeficiente de correlación (estimado mediante bootstrapping) mayor que el 20% no se utilizaron como grupo de restricciones del grupo de interacción de ARN (RING). Las iteraciones por bootstrapping eran suficientemente grandes, de manera que la diferencia absoluta entre los coeficientes de correlación con bootstrapping y calculados fue de menos del 1%.

[0204] Agrupamiento espectral de conformaciones múltiples en un único conjunto de ARN individuales. Los nucleótidos que forman un ARN definen las dimensiones de un espacio abstracto de muchas dimensiones, en el que cualquier lectura individual de una cadena de ARN está representada por un punto cuyas coordenadas están definidas por las que los nucleótidos, si lo hay, reaccionan con el reactivo químico (cada coordinada se ajusta a 1 o 0, dependiendo de si el nucleótido que la coordenada representa era reactivo o no). En este espacio, las cadenas con conformaciones estructurales similares tenderán a agruparse, reflejando las diferencias en la frecuencia de perfiles de modificación para cada conformación. Se utilizó el agrupamiento espectral (Shi J, Malik J (2000) IEEE Trans Pattern Anal Mach Intell 22: 888-905; Ng AY, Jordan MI, Weiss Y (2002) Advances in Neural Information Processing Systems 14 eds Dietterich TG, Becker S , Ghahramani Z (MIT Press, Cambridge, MA), pág. 849-856; Luxburg von U (2007) Stat Comput 17: 395-416), que es particularmente eficaz en la búsqueda de grupos con forma allrbitraria para definir grupos estructurales de ARN, sin hacer ninguna hipótesis sobre la forma de los grupos de datos.

[0205] Para detectar la presencia de múltiples conformaciones estructurales en un conjunto de ARN usando el agrupamiento espectral, se resumieron los lugares en la secuencia primaria (N nucleótidos de longitud) de los nucleótidos que se modificaron por el reactivo químico en las cadenas M de ARN, en una matriz "hit", HM × N. Este conjunto de datos se trata como un gráfico ponderado, simple, completo, no dirigido, en el que cada nucleótido está representado por un vértice con todos los N vértices unidos por los bordes. A cada borde se le asignó un peso correspondiente a la similitud de los patrones de reactividad de los dos nucleótidos, medida como el número de lecturas en el conjunto de datos en el que se modificaron ambos nucleótidos:

$$S = H^T H \qquad (3)$$

A continuación, esta matriz de similitud S se utilizó para construir una matriz Laplaciana del gráfico normalizado:

$$L_{NCut} = D^{-1/2} \cdot (D - S) \cdot D^{-1/2} \qquad (4)$$

donde D es una matriz diagonal, en la que $D_{ii} = \Sigma_j S_{ij}$. Los vectores propios de la matriz L_{Ncut} se utilizaron para realizar una partición de corte normalizada del conjunto de datos en grupos mediante el corte de los bordes entre los vértices de una manera que minimizaba la suma de los pesos de los bordes cortados y maximizaba la suma de pesos de los bordes conservados (Shi J, Malik J (2000) Pattern IEEE Trans Anal Mach Intell 22: 888-905; Ng AY, Jordan MI, Weiss Y (2002) Advances in Neural Information Processing Systems 14, eds Dietterich TG, Becker S, Ghahramani Z (MIT Press, Cambridge, MA), pág. 849-856; Luxburg von U (2007) Comput Stat 17: 395-416). En nuestra aplicación a los datos de mutación de ARN, los valores propios y los vectores propios de la matriz Lapaciana del gráfico normalizado L_{Ncut} se

utilizaron para (i) determinar la cantidad de conformaciones estructurales presentes en en el conjunto de ARN estudiados, (ii) estimar las fracciones relativas de las diferentes conformaciones en la muestra, y (iii) reconstruir los perfiles de frecuencia de mutaciones para las conformaciones individuales. Estos procedimientos se describen en detalle en los párrafos siguientes. El agrupamiento espectral se aplicó a los nucleótidos de adenosina y citosina que se modificaron con frecuencias superiores a 0,01. Las cadenas sin modificaciones no llevan ninguna información y se excluyeron del agrupamiento espectral.

[0206] Los valores propios se clasificaron en orden ascendente, desde el valor propio más pequeño, λ1, al más grande, λN. El primer valor propio, λ1, es siempre cero. Los valores propios expresan la eficacia de cada partición de corte normalizada de los vértices. Cuanto más eficaz sea un corte particular en los bordes de corte entre vértices diferentes (tales como los que pertenecen a diferentes grupos) conservando al mismo tiempo los bordes entre vértices similares (tales como los que pertenecen al mismo grupo), más pequeño es su valor propio. Por lo tanto, si un conjunto de datos tiene K grupos distintos, los primeros valores propios K serán claramente más pequeños que el valor propio K + 1 y el resto de los valores propios. Por lo tanto, para estimar el número de grupos, K, en un conjunto de datos, se optó por K, de manera que todos los valores propios λ2, λ3 ... λK eran relativamente pequeñas, y λK + 1 era relativamente grande. Para hacer que los saltos entre valores propios consecutivos sean más claros, se evaluaron "espacios propios" (definidos como una diferencia Δλi = λi + 1 - λi, con el primer espacio propio, Δλ1, dijado a cero) en lugar de valores propios (Luxburg von U (2007) Stat Comput 17: 395-416). En general, si un conjunto de datos tiene K grupos, la representación de espacios propios tendrá un espacio propio destacado en la posición K (ΔλK) y probablemente también a la izquierda de la misma, pero no a la derecha de la misma.

[0207] Estimación de fracciones relativas de diferentes conformaciones en muestras de ARN. Si un conjunto de datos tiene grupos K, se pueden utilizar vectores propios ~x2... ~xK para asignar cadenas de ARN individuales a grupos. Específicamente, si se reconoce que un conjunto de datos de cadenas M tiene grupos K, las puntuaciones de las cadenas se calculan como:

$$Y_{M\times(K-1)} = H \cdot \left[\vec{x}_2, \vec{x}_3, \cdots \vec{x}_K\right], \qquad (5)$$

Donde $\vec{x}_2, \vec{x}_3, \cdots \vec{x}_K$ son los vectores propios segundo a k-ésimo. Las puntuaciones tienen K-1 dimensiones, y las cadenas se dividen en grupos K mediante la realización de la agrupación de *K-means* de puntos de datos *M* en el espacio puntuación dimensional *K-1*.

[0208] Reconstrucción de perfiles de frecuencia de modificación de conformaciones individuales en una muestra de ARN. Una vez las cadenas están asignadas a grupos que reflejan distintas conformaciones, pueden calcularse los perfiles de frecuencia de modificación específicamente para cada conformación. La exactitud de dicha reconstrucción del perfil mejoró mediante el cálculo de las frecuencias de modificación de cada nucleótido por separado usando el siguiente procedimiento. Para calcular las frecuencias de modificación del nucleótido i, este nucleótido se eliminó primero de la matriz “hit” H y, a continuación, se realizó el agrupamiento espectral y la separación de cadenas mediante el agrupamiento con K-means sobre esta matriz reducida (sin la contribución de este nucleótido). A continuación, se calculó la frecuencia de modificación del nucleótido i por separado para cada grupo separado de cadenas, dando estimaciones para diferentes conformaciones.

[0209] Modelado de la estructura de ARN tridimensional. La reconstrucción de los pliegues del ARN tridimensional se realizó usando un enfoque de dinámica molecular limitada en el que se incorporaron bonos de energía libre basado en correlaciones de nucleótidos por pares (Gherghe CM, et al, (2009) J Am Chem Soc 131 (7): 2541-2546; Lavender CA, et al, (2010) Biochemistry 49 (24): 4.931-4.933). Cada nucleótido se modeló como tres pseudoátomos correspondientes a los grupos fosfato, azúcar y base. Las interacciones por pares que incluyen el apareamiento de bases, el apilamiento de bases, las interacciones del empaquetamiento y la repulsión electrostática se aproximan usando los potenciales de pozo cuadrado (Ding F, et al (2008) RNA 14 (6): 1164-1173.). Las disposiciones de apareamiento de bases aceptadas (Serganov A, et al, (2006) Nature 441 (7097): 1167-1171; Cate JH, et al (1996) Science 273 (5282): 1678-1685; Kazantsev AV, AA Krivenko, NR Pace (2009) RNA 15 (2): 266-276) se utilizaron para limitar el modelado.

[0210] Para incorporar la información de análisis RING, se aplicaron potenciales de energía libre durante simulaciones de dinámica molecular discreta (DMD) entre pares de nucleótidos que se encontró que interactuaban. Se incluyó un bono de energía libre para interactuar pares de nucleótidos si el coeficiente de correlación absoluto estaba por encima de 0,025. Los potenciales de energía libre no se incluyeron si dos nucleótidos estaban implicados al estar en contacto por proximidad en la secuencia primaria o por la participación en un elemento de estructura secundaria común. Para vecinos de secuencia primaria, no se incluyó un potencial de energía libre entre los nucleótidos dentro de las 11 posiciones en la secuencia. Para los vecinos de estructura secundaria, los potenciales no se añadieron para los nucleótidos en el mismo elemento estructural. Los elementos estructurales se definieron como los nucleótidos en las posiciones n_i y n_j en un par RING y los nucleótidos en las posiciones m_i y m_j en cualquier par de bases dada, si $|n_i - m_j|$ +

$|n_j - m_j|$ es menor que o igual a 11 nucleótidos. El umbral de 11-nt fue seleccionado basándose en el número de pares de bases en un solo giro de una hélice de ARN en forma de A.

[0211] Se impusieron potenciales de energía libre entre pares de nucleótidos correlacionados con RING basados en distancias a través del espacio entre nucleótidos constituyentes durante simulaciones. Para distancias entre los nucleótidos correlacionados dentro de 36 Å y 23 Å, los bonos aplicados fueron -0,3 y -0,6 kcal/mol, respectivamente. El bono máximo de -0,6 kcal/mol es equivalente a la estabilización proporcionada por la interacción de apilamiento de un solo ARN en el campo de fuerzas de dinámica molecular.

[0212] Se realizaron simulaciones de dinámica molecular usando el motor de DMD con intercambio de réplicas (Ding F, et al, (2012) Methods Nat 9 (6): 603-608). Ocho réplicas se desarrollaron en paralelo para 1 millón de unidades de tiempo, cada uno con valores de factores de temperatura de réplica de 0,1000, 0,1375, 0,1750, 0,2125, 0,2500, 0,2875, 0,3250, y 0,3625. De cada réplica, se tomaron modelos a cada 100 unidades de tiempo. A continuación, se filtró esta lista de modelos basado en el radio de giro. Los radios de giro para estos modelos se compararon frente a una simulación de control donde no se incorporaron potenciales a base de RING. Para ambos modelos dependientes de RING y de control, se construyeron histogramos de radios de giro. El histograma de control se escaló para minimizar su diferencia del histograma experimental, y entonces la frecuencia del histograma de control se restó del experimental. Para este histograma de diferencia, se encontró una distribución logarítmica normal mediante ajuste de mínimos cuadrados que describe la distribución de radios de giro para estructuras colapsadas de dependientes de limitaciones. Para mayor consideración, los modelos dependientes de RING tenían que estar dentro de una SD geométrica de la media geométrica descrita por esta distribución de ajuste.

[0213] Después de la filtración por el radio de giro, los 250 modelos con las energías más bajas se analizaron a continuación mediante agrupación jerárquica (Lavender CA, et al, (2010) Biochemistry 49 (24): 4931-4933). La agrupación se realizó teniendo en cuenta los valores de rmsd entre los modelos analizados. La agrupación se limitó de tal manera que el máximo de rmsd entre dos miembros constituyentes de un grupo fue menor que la suma del promedio y SD de la distribución rmsd predicha para la estructura analizada. El medoide del grupo más poblado se tomó como la estructura predicha.

EJEMPLO 5

Descubrimiento de motivos de ARN mediante SHAPE y perfil mutacional (SHAPE-Map). RESULTADOS

[0214] Estrategia MaP. Los experimentos SHAPE utilizan reactivos selectivos de 2'-hidroxilo que reaccionan para formar 2'-O-aductos covalentes en nucleótidos de ARN conformacionalmente flexibles, tanto bajo condiciones de solución simplificadas (Wilkinson, KA et al. PLoS Biol. 6, E96 (2008), Merino, EJ, et al., J. Am. Chem. Soc. 127, 4.223-4.231 (2005) y en células (Tyrrell, J., et al., Biochemistry 52, 8.777-8.785 (2013); McGinnis, JL & Weeks, KM Biochemistry 53, 3237-3247 (2014); Spitale, RC et al Nat Chem Biol 9, 18-20 (2013)). Los datos de SHAPE se pueden emplear como restricciones en algoritmos para la predicción de la estructura del ARN para proporcionar modelos muy precisos de la estructura secundaria para ARN estructuralmente complejos (Ejemplo 3 en el presente documento anteriormente; Hajdin, CE et al Proc Natl Acad Sci USA. 110, 5.498-5.503 (2013). En este Ejemplo, las modificaciones químicas de SHAPE (Mortimer, SA & Weeks, KMJ Am Chem Soc 129, 4144-4145 (2007); Ejemplo 3 en el presente documento anteriormente; Merino, EJ, et al J. Am Chem Soc 127, 4223 -4231 (2005); Steen, K.-A., Rice, GM & Weeks, KMJ Am. Chem. Soc. 134, 13160-13163 (2012)) en el ARN se cuantifican en una sola etapa directa mediante secuenciación masiva en paralelo (Fig. 6). El enfoque explota las condiciones que causan que la transcriptasa inversa lea erróneamente los nucleótidos modificados por SHAPE e incorporen un nucleótido no complementario a la secuencia original en el ADNc recién sintetizado.

[0215] Las posiciones y las frecuencias relativas de aductos de SHAPE de este modo se registran inmediata, directa y permanentemente como mutaciones en la secuencia primaria de ADNc, creando de este modo un SHAPE-MAP. En un experimento SHAPE-MaP, el ARN se trata con un reactivo de SHAPE o se trata sólo con disolvente, y el ARN se modifica bajo condiciones de desnaturalización para controlar los sesgos específicos de secuencia en la detección de mutaciones inducidas por aducto. El ARN de cada condición experimental se transcribe de forma inversa y los ADNc resultantes se preparan, a continuación, para la secuenciación masivamente en paralelo. Las posiciones reactivas se identifican restando de datos para la muestra tratada de los datos obtenidos para la muestra no tratada y mediante la normalización a los datos para el control desnaturalizado (Figs. 6 y 9).

[0216] Modelado de la estructura: validación. La estructura del dominio de aptámero del riboswitch de pirofosfato de tiamina (TPP) de *Escherichia coli* se examinó inicialmente en presencia y ausencia de concentraciones saturantes del ligando TPP. SHAPE-MaP recapituló el patrón de reactividad conocido para el ARN unido a ligando plegado y describió con precisión las diferencias de reactividad con resolución de nucleótidos que se producen tras la unión al ligando. Estos resultados, y un análisis del ARNr 16S de 1542 nt de *E. coli* demuestran la capacidad de SHAPE-MaP para capturar detalles finos estructurales para conformaciones de ARN distintas a una resolución de nucleótidos, con precisión, de forma reproducible y con independencia del tipo nucleótido. Dado que los perfiles SHAPE se reconstruyen a partir de las frecuencias de mutaciones derivadas de todas las lecturas de secuenciación, las incertidumbres en reactividades SHAPE se puede estimar a partir de la distribución de Poisson de sucesos o eventos de mutaciones.

[0217] El uso de datos de SHAPE como los términos de cambio de pseudoenergía libre para restringir el modelado de la estructura secundaria ha sido ampliamente comparado utilizando conjuntos de pruebas de ARN elegidos específicamente como desafíos para el modelado de estructura secundaria convencional (Ejemplo 3 en el presente documento anteriormente; Hajdin, CE et al . Proc. Natl. Acad. Sci. USA. 110, 5.498-5.503 (2013)). Para evaluar la exactitud de SHAPE-MaP, se sondeó un subconjunto de estos ARN, que variaban en tamaño de 78 nucleótidos (nt) a 2904 nt, con el reactivo 1M7 bien validado. El experimento SHAPE "diferencial" que utiliza dos reactivos adicionales, 1M6 y NMIA, también se evaluó para detectar interacciones no canónicas y terciarias y para producir modelos estructurales de ARN con una alta precisión consistente, incluso para ARN especialmente difíciles (Ejemplo 3 en el presente documento anteriormente; Steen, K.-A., Rice, GM & Weeks, KM Fingerprinting noncanonical and tertiary RNA structures by differential SHAPE reactivity. J. Am. Chem. Soc. 134, 13160-13163 (2012). La precisión global del modelado dirigido por SHAPE-MAP de la estructura de ARN usando reactividades diferenciales, medida en términos de sensibilidad y el valor predictivo positivo, fue similar a y, a menudo superior, a la de reactividades SHAPE convencionales basadas en la terminación mediada por aducto de la extensión del cebador detectada por electroforesis capilar. La precisión para la recuperación de pares de bases canónicas aceptadas superó el 90% (Fig. 7A).

[0218] Las reactividades SHAPE obtenidas utilizando la estrategia MaP se miden como muchos eventos individuales mediante secuenciación masiva en paralelo. La fiabilidad depende de la medición adecuada de las tasas de mutación. Se logró un modelado preciso de la estructura del ARNr 16S utilizando una profundidad de lectura por nucleótido de 2.000-5.000. Esto corresponde a 6-15 modificaciones por encima de la base por nucleótido ribosomal en promedio (Fig. 7B). Aunque se han realizado varios estudios anteriores (Kertesz, M. et al Nature 467, 103-107 (2010); Ding, Y. et al Nature 505, 696-700 (2014); Rouskin, S., et al Nature 505, 701-705 (2014)), en los que todos los ARN en un transcriptoma determinado estaban físicamente presentes durante la fase de sondeo del experimento, este análisis a nivel de éxito indica que sólo unos pocos miles de nucleótidos en cada caso fueron muestreados a una profundidad que permitiría la recuperación completa de la información estructural subyacente. El modelado dirigido por SHAPE-MAP preciso se consiguió usando los mismos parámetros definidos originalmente para experimentos basados en electroforesis capilar y se obtuvieron precisiones altas comparables utilizando experimentos específicos de ARN y cebados aleatoriamente. Los datos eran altamente reproducibles entre réplicas biológicas completas realizadas con meses de separación por diferentes individuos, lo cual enfatizó la robustez de SHAPE-MaP.

[0219] Un modelo de alta resolución actualizado para un genoma de ARN de VIH-1. Se obtuvo información estructural con resolución de un solo nucleótido para un ARN genómico de VIH-1 auténtico completo (cepa NL4-3, ~9200 nt) en experimentos y se realizó el análisis de datos durante aproximadamente 2 semanas. Los datos de SHAPE-MAP de 1M7 y diferencial se procesaron para producir perfiles de reactividad SHAPE y modelos de estructura secundaria utilizando algoritmos eficientes y totalmente automatizados (Figs. 8 y 9). El enfoque MAP, implementado en este Ejemplo, produce datos de reactividad con resolución de nucleótidos para grandes ARN que son iguales o superiores a los datos de electroforesis capilar estándar de oro anteriores (Fig. 7A). Por lo tanto, la estructura del genoma del VIH-1 que aquí se presenta constituye un nuevo modelo de mayor resolución para los elementos bien definidos en este ARN.

[0220] Identificación *de novo* de estructuras bien determinadas. Casi cualquier secuencia de ARN larga formará algunas estructuras secundarias (Doty, P., et al. Proc. Natl. Acad. Sci. USA. 45, 482-499 (1959), pero no todas estas estructuras son biológicamente importante o están bien definidas. Por lo tanto, se utilizó el modelado dirigido por SHAPE, cuya función de energía subyacente produce modelos muy precisos para ARN con estructuras secundarias bien definidas (Figs. 7A y 7B), para calcular una probabilidad de cada par de bases en todas las estructuras posibles en el conjunto de Boltzmann de estructuras predichas para el ARN del VIH-1. Estas probabilidades se utilizaron para calcular entropías de Shannon (Huynen, M., Gutell, R. & Konings, DJ Mol. Biol. 267, 1104 -1112 (1997); Mathews, DH RNA 10, 1178-1190 (2004)) (Fig 8). Las regiones con mayores entropías de Shannon probablemente forman estructuras alternativas, y aquellas con bajas entropías de Shannon corresponden a regiones con estructuras de ARN bien definidas o de una sola cadena persistente, tal como se determina por la reactividad SHAPE. La representación de la probabilidad de emparejamiento a lo largo de todo el genoma de VIH-1 revela estructuras de ARN bien determinadas y variables en el ARN genómico de VIH-1 (Fig. 8A). Las regiones estructuradas anteriormente caracterizadas, tales como la región no traducida 5' (UTR), el elemento de respuesta Rev (RRE), el elemento de desplazamiento de marco y el tracto de polipurina están bien determinados en el modelo. En cambio, también hay grandes regiones (tales como nucleótidos 3.200-4.500 y 6.100-6.800) que tienen reactividades SHAPE elevadas y una alta entropía de Shannon, y, por lo tanto, es probable que muestren muchas conformaciones. Este enfoque de visualización destaca las regiones con estructuras únicas, probablemente estables, y aquellas regiones en las que múltiples estructuras son susceptibles de estar en equilibrio.

[0221] Un análisis de entropías Shannon y reactividades SHAPE proporciona un enfoque para el descubrimiento *de novo* de regiones con estructura bien definida en los ARN largos. Quince regiones en el ARN genómico de VIH-1 tenían valores de reactividad SHAPE bajos (indicando un alto grado de estructura de ARN) y bajas entropías Shannon (que proporciona confianza en una única estructura secundaria predominante) (Figs. 8A y 8B). Se crearon modelos de estructura con resolución de nucleótido para cada una de estas regiones (Fig. 8C). Los modelos de estructuras de regulación funcionalmente importantes conocidas (RRE, elemento de respuesta que actúa en trans 5' (TAR), sitio de unión a cebador, elemento de empaquetamiento (Psi), sitio de iniciación de la dimerización, elemento del marco de

lectura ribosomal y 3' TAR) concordaban estrechamente con los modelos propuestos anteriormente para estas regiones. Además, la hélice continua más larga, las horquillas que flanquean el tracto de polipurina y otras características permanecen consistente entre el anterior modelo (Watts, JM et al., Nature 460, 711-716 (2009)) y el modelo actual (Tabla 4).

[0222] A continuación, se obtuvo una lista de todos los elementos reguladores que probablemente funcionan a través de un motivo de ARN (Fig. 8B y Tabla 5). A continuación, la ubicación de estos elementos estructurales de ARN se compararon con las regiones altamente estructuradas y de baja entropía identificadas *de novo* por SHAPE-MaP. Los elementos de ARN funcionales se producen mayoritariamente en regiones de bajo SHAPE y baja entropía de Shannon ($P = 0,002$; Fig. 8), lo que indica que la mayoría de las funciones mediadas por ARN operan en el contexto de una estructura de ARN subyacente. Varias regiones de bajo SHAPE y baja entropía de Shannon en el genoma del VIH-1 se producen en las regiones no asociadas previamente con elementos funcionales de ARN conocidos: estas regiones son dianas de alto valor para el descubrimiento de nuevos motivos de ARN.

[0223] Descubrimiento de motivos y deconvolución de polimorfismo estructural. Los pseudonudos parecen ser raros en grandes ARN y son difíciles de identificar, pero estos motivos parecen estar sobrerrepresentados en regiones funcionalmente importantes de muchos ARN (Staple, DW y Butcher, SE PLoS Biol 3, E213 (2005); Brierley, I., Pennell, S. y Gilbert, RJC Nat. Rev. Microbiol. 5, 598-610 (2007)). Como una prueba rigurosa de los avances acumulativos actuales en el modelado de la estructura dirigida por SHAPE y de los propios datos de SHAPE-MaP de alto rendimiento, se realizó una búsqueda (Hajdin, CE et al. Proc. Natl. Acad. Sci. USA. 110, 5.498-5503 (2013)) para nuevos pseudonudos en el ARN del genoma de VIH-1. En el modelo de este ejemplo, hay cuatro pseudonudos en regiones de baja reactividad SHAPE y baja entropía de Shannon (Fig. 8C). El pseudonudo adyacente a la señal de poliadenilación 5' en el ARN de VIH-1 (5'$_{PK}$) se ha validado previamente (Wilkinson, KA et al PLoS Biol 6, E96 (2008); Paillart, J.-C., et al. J. Biol. Chem. 277, 5995-6004 (2002)). Los tres nuevos pseudonudos adicionales se prevé que se formen en la región de codificación de la transcriptasa inversa (RT_{PK}), al inicio de env (ENV_{PK}) y en la región U3 adyacente a la señal de poliadenilación 3' ($U3_{PK}$). Como control negativo, se analizó un pseudonudo adicional predicho por el algoritmo de ShapeKnots que se encuentra en una región de alta reactividad SHAPE y de alta entropía de Shannon (CA_{PK}, los nucleótidos 961-1.014).

[0224] Las mutaciones silenciosas diseñadas para alterar cada pseudonudo se introdujeron en el genoma de VIH-1 de longitud completa. Las características especiales de la región $U3_{PK}$ ilustran el poder del enfoque de MaP. Las secuencias U3 aparecen en ambos extremos 5' y 3' del genoma viral en ADN de VIH-1 proviral, pero sólo en el extremo 3' en el ARN viral. Durante la transfección del plásmido que codifica el provirus, estas secuencias pueden someterse a recombinación. Cuando se introdujeron mutaciones en la secuencia U3 sola (en el extremo 3') con la secuencia nativa U3 en el extremo 5' del ADN proviral, los experimentos SHAPE-MaP revelaron que ambas secuencias nativas y mutantes estaban presentes en los extremos 3' de los ARN genómicos individuales en la muestra que contenía $U3_{PK}$ mutante. Debido a que los nucleótidos se analizan en el contexto de regiones de ARN no fragmentadas en el enfoque de MaP, ambos alelos pueden ser controlados de forma independiente en el mismo experimento, separarse computacionalmente y se puede construir perfiles SHAPE individuales para los ARN nativos y mutantes. Se observaron diferencias notables en la reactividad SHAPE entre U3 nativo y mutante, producidas por virus en competencia directa entre sí y consistente con una alteración precisa de la estructura de $U3_{PK}$. Las mutaciones introducidas en el lado 5' de la hélice con pseudonudo $U3_{PK}$ indujeron cambios en la pareja de emparejamiento 3' predicha, que se encontraba a más de 100 nucleótidos de distancia. SHAPE-MaP es por lo tanto particularmente útil para el análisis estructural y el descubrimiento de motivo en los sistemas que contienen mezclas complejas de ARN, y para detectar y desconvolucionar consecuencias estructurales de polimorfismos de un solo nucleótido y otros polimorfismos alélicos.

[0225] Todas las construcciones mutantes se analizaron utilizando SHAPE-MaP y en ensayos basados en células para la capacidad viral. Las mutaciones en $U3_{PK}$ redujeron la propagación viral en las células Jurkat en aproximadamente diez veces en relación con NL4-3 y redujeron la capacidad viral en competencia directa con NL4-3, con una diferencia de capacidad relativa media de -0,32 en relación con NL4-3 31. Este gran efecto sobre la capacidad viral por mutaciones en el U3PK es consistente con la importancia general de UTR 3' en la regulación de la estabilidad y la traducción del ARNm (Matoulkova, E., et al. RNA Biol. 9, 563-576 (2012)) y, más particularmente, con un papel para la organización espacial específica de orden superior de la señal de poli(A) y elementos de secuencia en dirección 5' en el montaje del mecanismo de poliadenilación (Gilmartin, GM, et al EMBO J. 11, 4.419-4.428 (1992); Klasens, BI, et al. Nucleic Acids Res. 27, 446-454 (1999)). Los cambios SHAPE en el mutante de RT_{PK} también se localizaron directamente en o inmediatamente adyacente a la hélice con pseudonudos. Las mutaciones en RT_{PK} mostraron una disminución más pequeña, pero reproducible, de la propagación viral y la capacidad viral, con una capcidad relativa media de -0,14, en comparación con NL4-3. Los cambios en reactividades SHAPE también se observaron en las secuencias 5' y 3' para el mutante de ENV_{PK} de “larga distancia”, incluidos los cambios que se extienden 5 desde la hélice con pseudonudos, lo cual sugiere un replegamiento local causado por la alteración de este pseudonudo. La propagación viral y la capacidad viral no se redujeron por el mutante de ENV_{PK}, lo que podría reflejar el desafío de detectar algunas características de la replicación de VIH-1 en cultivo celular. Las mutaciones en CA_{PK}, que se analizó como control negativo, no apoyaron la existencia de una estructura con pseudonudos en este lugar mediante el análisis de SHAPE-MaP, en concordancia con el perfil de alta entropía de Shannon.

DISCUSIÓN

[**0226]** Con perfiles mutacional, la información estructural de ácido nucleico se registra directamente y de forma concisa en la secuencia del ADNc complementario y resultó insensible a sesgos en la preparación de la biblioteca y la secuenciación masivamente en paralelo. MaP de este modo convierte la transcripción inversa o la síntesis de ADN en un motor directo para el descubrimiento de la estructura del ácido nucleico. MaP es totalmente independiente de la estrategia de secuenciación y por lo tanto se puede utilizar en cualquier enfoque de secuenciación con una tasa de error de base suficientemente baja para cuantificar las modificaciones químicas en cualquier ARN detectable de baja abundancia mediante transcripción inversa. La detección de aductos químicos en ambos ARN y ADN a través de lectura directa puede acoplarse con las estrategias para la selección de la polimerasa (Ghadessy, FJ y Holliger, P. et al Methods Mol Biol 352, 237-248 (2007); Chen, T. y Romesberg, FE FEBS Lett. 588, 219-229 (2014) para registrar, como perfiles mutacionales o MaP, una amplia variedad de modificaciones post-transcripcionales y epigenéticas.

[0227] Los datos de SHAPE-MaP contienen estimaciones de error y se integran fácilmente en algoritmos revisados totalmente automatizados para el modelado de la estructura y el descubrimiento de motivos de toda la transcripción. A gran escala y en estudios a escala de genoma de la estructura del ARN, se pueden identificar verdaderos elementos funcionales en la base del conjunto complejo de estructuras que se forman en cualquier ARN grande. La combinación de análisis de SHAPE-MaP con el análisis de las probabilidades de emparejamiento, calculada a lo largo de grandes regiones de ARN, identificó casi todos conocidos los elementos funcionales conocidos a gran escala en el genoma del VIH-1, con la excepción del tracto de polipurina central, que parece tener una estructura conservada (Pollom,. E. et al. PLoS Pathog 9, e1003294 (2013)). Por lo tanto, la sensibilidad de detección del elemento funcional por SHAPE-MaP es muy alta. Además, a pesar del hecho de que el genoma del VIH-1 es uno de los ARN más intensamente estudiados en la historia científica, el SHAPE-MaP cuantitativo y de alta resolución permitió, sin embargo, un rápido descubrimiento *de novo* y la validación directa de nuevos motivos funcionales, específicamente tres pseudonudos, un motivo que tradicionalmente ha sido un reto de predecir. El valor predictivo positivo de los enfoques desarrollados aquí es por lo tanto también correspondientemente alto. SHAPE-MaP proporciona simplicidad experimental y precisión estructural y se puede escalar a sistemas de ARN de cualquier tamaño y complejidad.

PROCEDIMIENTOS

[0228] Visión general experimental de SHAPE-MaP. Los experimentos SHAPE-MaP usan condiciones para la transcripción inversa que promueven la incorporación de nucleótidos no complementarios al ARN en el ADNc naciente en los lugares de aductos de SHAPE. Los sitios de aductos de ARN de este modo corresponden a mutaciones internas o deleciones en el ADNc, con relación a la comparación con los ADNc transcritos de ARN no tratado con reactivo SHAPE. La transcripción inversa se puede llevar a cabo utilizando cebadores específicos de gen o aleatorios (Figura 9); ambos enfoques se describen a continuación. Una vez que la síntesis de ADNc se ha completado, la información estructural del ARN esencialmente se registra de forma permanente en la secuencia y es por lo tanto independiente de sesgos introducidos durante cualquier esquema de construcción de bibliotecas con múltiples etapas. La preparación de bibliotecas es similar a la de un experimento de secuenciación de ARN (RNA-seq), se puede adaptar fácilmente a cualquier plataforma de secuenciación, y permite multiplexar utilizando códigos de barras de secuencia. Las roturas de cadena simple y la degradación de base no interfieren intrínsecamente con los experimentos de SHAPE-MaP (en contraste con SHAPE convencional y otros ensayos dependientes de parada de la transcriptasa inversa), ya que no se detectan durante la lectura a lo largo de la secuenciación. Tampoco hay degradación o decrecimiento de señal en el enfoque de MaP, que de otro modo requiere una corrección compleja, parcialmente heurística.

[0229] Desarrollo y eficacia de SHAPE-MaP. Las enzimas de transcriptasa inversa pueden, en algunos casos, leer a través de enlaces 2'-O y aductos inusuales, después de la pausa de la enzima (Lorsch, JR, Bartel, DP y Szostak, Nucleic Acids Res 23, 2811-2814 (1995); Patterson, JT, Nickens, DG & Burke, DH ARN Biol. 3, 163 (2006)). La hipótesis es que la lectura causa, o resulta de, la distorsión estructural en el sitio activo de la transcriptasa inversa, lo que resulta en una mayor tasa de incorporación errónea de nucleótidos en la ubicación de un aducto de SHAPE inductor de la pausa. Múltiples enzimas de transcriptasa inversa fueron seleccionadas para su uso en SHAPE-MapP en función de la concentración de nucleótidos, tiempo de reacción, condiciones del tampón e identidad de iones de metal divalente. Las condiciones de la enzima se buscaron para las paradas mínimas producidas de la transcripción inversa inducidas por aducto y productos de ADNc de longitud completa máxima. De los iones metálicos divalentes analizados (incluyendo magnesio, manganeso, cobre, cobalto, níquel y plomo), Mn^{2+} promovió más eficazmente la lectura de la enzima en los sitios de 2'-O-aductos voluminosos, en particular, utilizando trancriptasa inversa del virus de la leucemia murina de Moloney (SUPERSCRIPT™ II, Invitrogen). Esta observación es consistente con la alta actividad de la transcriptasa inversa de Moloney en Mn^{2+} (Roth, MJ, Tanese, N. & Goff, SPJ Biol. Chem. 260, 9326 a 9335 (1985)) y la capacidad de este ion para promover el comportamiento mutagénico en ADN polimerasas (Beckman, RA, Mildvan, AS & Loeb, LA Biochemistry 24, 5810 a 5817 (1985)). Las clases precisas de los sucesos de incorporación incorrecta inducidos por aducto se determinaron mediante la comparación de las tasas de sustitución y deleción en las posiciones de nucleótidos no apareados y apareados en el ARNr 16S.

[0230] Las tendencias de incorporación errónea fueron similares entre los tres reactivos SHAPE (1M7 (Mortimer, S.A. & Weeks, K.M. J Am. Chem. Soc. 129, 4144-4145 (2007))) y los reactivos "diferenciales" NMIA y 1M6 (Ejemplo 3 en el presente documento más arriba). Generalmente, la presencia de un aducto SHAPE hace que los nucleótidos no sean bien leídos como A o T, o como eventos de deleción, aunque hay un contenido de información sustancial en otros

eventos de incorporación errónea. Los nucleótidos flexibles en un sustrato modelo de dinucleótidos con una única posición reactiva (AddC) (Mortimer, S.A. & Weeks, K.M. J Am. Chem. Soc. 129, 4144-4145 (2007)) se modifican con una eficiencia de ~ 2% en NMIA o 1M7 en condiciones similares a las utilizadas en el presente documento. Las tasas de mutaciones por encima de la base en las posiciones flexibles en el ARNn 16S son ≥ 0,5 %, con muchas de las posiciones más reactivas por encima del 2%. Teniendo en cuenta estos valores límite, se estimó que la estrategia MaP detecta aductos SHAPE con una eficiencia de ≥ 50%.

[0231] Plegamiento de ARN y sondeo SHAPE de los ARN modelo. Se sintetizaron plantillas de ADN (IDT) para los ARN de $ARNt^{Phe}$, riboswitch de TPP, 5S de *E. coli*, dominio IRES del virus de la hepatitis C, intrón del grupo I de *T. thermophila* o intrón del grupo II de *O. iheyensisen* el contexto de cassettes de estructura 5' y 3' flanqueantes. Las plantillas se amplificaron por PCR y se transcribieron en ARN utilizando ARN polimerasa de T7 (Wilkinson, KA, et al. Nat. Protoc 1, 1610-1616 (2006)). Los ARN se purificaron mediante PAGE desnaturalizante, se escindieron las regiones apropiadas y los ARN se eluyeron pasivamente del gel durante la noche a 4 °C. Se aislaron ARN 16S y 23S de células DH5α durante la fase semilogarítmica usando condiciones no desnaturalizantes (Deigan, K.E., et al. Proc. Natl. Acad. Sci. USA. 106, 97-102 (2009)). Para cada muestra, se replegaron 5 pmol de ARN en HEPES 100 mM, pH 8,0, NaCl 100 mM y $MgCl_2$ 10 mM en un volumen final de 10 µl. Después de plegarse, los ARN fueron modificados en presencia de reactivo SHAPE 10 mM y se incubaron a 37 °C durante 3 min (1M6 y 1M7) o 22 min (NMIA). Los controles no reactivos, que contienen DMSO puro en lugar de reactivo SHAPE, se realizaron en paralelo. Para justificar los sesgos específicos de secuencia en la detección de aductos, los ARN fueron modificados utilizando NMIA, 1M7 o 1M6 bajo condiciones fuertemente desnaturalizantes en HEPES 50 mM (pH 8,0), EDTA 4 mM y formamida al 50 % a 95 ºC. Después de la modificación, los ARN se aislaron usando columnas de afinidad de ARN (RNeasy™ MinElute™; Qiagen) o columnas de centrifugación G-50 (GE Healthcare).

[0232] Plegamiento de ARN y sondeo SHAPE del ARN genómico de VIH-1. Para el SHAPE-MaP de genoma completo de VIH-1 (cepa NL4-3; grupo M, subtipo B), el virus se produjo y se purificó como se ha descrito (Watts, JM et al., Nature 460, 711-716 (2009).). El ARN viral se extrajo suavemente y se purificó de la proteína, a continuación, se precipitó con etanol a partir de una solución que contenía NaCl 300 mM. Aproximadamente el 30% del ARN genómico es de longitud completa cuando se preparó de esta manera (Watts, JM y otros, Nature 460, 711-716 (2009)); la naturaleza fragmentada de las muestras del genoma de VIH-1 nativo dio lugar a una disminución de la recuperación de la muestra durante las purificaciones en columna (RNeasy™ MinElute™, Qiagen). Por lo tanto, se utilizó ~ 1 g de ARN de VIH-1 por muestra, más ARN que los 250 ng requeridos para experimentos SHAPE-MaP de más ARN intactos. Los virus mutantes se produjeron mediante transfección de células 293T= utilizando FuGene6™ (Promega) o XtremeGene™ HP (Roche). Los sobrenadantes virales se concentraron usando concentradores centrífugos (Vivaspin™ 20, Sartorius), seguido de precipitación (concentrador Lenti-X™, Clontech) para concentrar los viriones. Los viriones sedimentados se resuspendieron en tampón de lisis viral (HEPES 50 mM (pH 8,0), NaCl 200 mM y $MgCl_2$ 3 mM) (Watts, JM et al., Nature 460, 711-716 (2009)), y se lisaron con SDS al 1% (p/v) y 100 µg/ml de proteinasa K (25 °C, 30 min). Se extrajo el ARN con fenol:cloroformo:alcohol isoamílico al menos tres veces, seguido por dos extracciones con cloroformo y precipitación con etanol.

[0233] Aproximadamente 1 µg de ARN genómico de VIH-1 se suspendió en tampón de modificación (HEPES 50 mM (pH 8,0), acetato de potasio 200 mM (pH 8,0) y $MgCl_2$ 3 mM) y se incubaron a 37 °C durante 15 min (paramuestras modificadas con SHAPE y sin tratar) o en tampón desnaturalizante (HEPES 50 mM (pH 8,0), EDTA 4 mM y formamida al 50%) y se incubaron a 95 °C durante 2 min. A continuación, las muestras se trataron a continuación con el reactivo SHAPE (10 mM final) o disolvente puro.

[0234] SHAPE-MaP usando muestras fragmentadas. Después de la modificación SHAPE y la purificación, las muestras de VIH-1, intrón del grupo II, IRES de VHC y ARNr se fragmentaron (produciendo longitudes de ~250-350 nt) mediante una incubación de 4 min a 94 °C en un tampón que contiene $MgCl_2$ 9 mM, KCl 225 mM y Tris HCl 150 mM (pH 8,3). Los fragmentos de ARN se desalaron usando columnas de centrifugación G-50. Las muestras fragmentadas (250-500 ng de masa total) se sometieron a transcripción inversa durante 3 horas a 42 °C (usando superscript™ II, Invitrogen). Las reacciones se cebaron utilizando 200 ng de cebadores nonámeros aleatorios (NEB) para el ARN del ribosoma, intrón del grupo II e IRES de VHC o con cebadores LNA personalizados (Figura 10) para genomas de ARN de VIH-1. El tampón de transcriptasa inversa contenía 0,7 mM de dNTPs premezclados, Tris = HCl 50 mM (pH 8,0), KCl 75 mM, $MnCl_2$ 6 mMy DTT 14 mM. Después de la transcripción inversa, las reacciones se desalaron usando columnas de centrifugación G-50 (GE Healthcare). Bajo estas condiciones (largos tiempos de incubación y utilizando Mn^{2+} 6 mM como el único ion divalente) la transcriptasa inversa lee a través de sitios de modificación en 2'-O mediante un reactivo SHAPE, incorporando un nucleótido no complementario en el sitio del aducto.

[0235] Las bibliotecas de ADN de cadena doble para la secuenciación masiva en paralelo se generaron utilizando módulos de preparación de muestras NEBNext™ para Illumina. La síntesis de la segunda cadena (NEB E6111) de la biblioteca de ADNc se realizó usando 100 ng de ADN de entrada, y la biblioteca se purificó usando un kit de limpieza Micro PCR PureLink™ (Invitrogen = K310250). La reparación de extremos de las bibliotecas de ADN de doble cadena se realizó usando el módulo de reparación de extremos NEBNext™ (NEB E6050). Los volúmenes de reacción se ajustaron a 100 µl, se sometieron a una etapa de limpieza (perlas Agencourt AMPure™ XP A63880, relación 1,6:1 de perlas con respecto a muestra), se añadió una cola d(A) (NEB E6053) y se ligaron con adaptadores bifurcados

compatibles Illumina (TruSeq™) con un módulo de ligación rápida (NEB M2200). Se realizó una emulsión PCR44 (30 ciclos) usando polimerasa de alta fidelidad Q5 “hot-start” (NEB M0493) para mantener la diversidad de muestras de la biblioteca. Las bibliotecas resultantes se cuantificaron (fluorímetro Qubit™; Life Technologies), se verificaron usando un Bioanalyzer™ (Agilent), se agruparon y se sometieron a secuenciación usando la plataforma Illumina MiSeq™ o HiSeq™. Se obtuvo una sola réplica de los ARN del grupo II e IRES de VHC y las reactividades SHAPE-MaP concordaban bien con las reactividades previamente generadas derivadas de SHAPE-CE. Para el genoma de ARN de VIH-1, se obtuvieron dos réplicas biológicas completas a diferentes concentraciones de reactivos. El análisis que aquí se presenta se basa principalmente en una de estas réplicas (las características de secuenciación se resumen en la Tabla 5). Se sondaron segmentos individuales seleccionados, incluyendo todas las regiones con pseudonudos, mediante estrategias fragmentadas y dirigidas (específicas de gen) y mostraron una concordancia excelente.

[0236] SHAPE-mAp usando cebadores específicos de gen dirigidos. Los ARN de ARNTPhe, riboswitch de TPP, ARNr 5S, intrón del grupo I y construcción mutante de VIH-1 se sometieron a transcripción inversa utilizando un cebador específico de ADN al cassette de estructura 3’ (5'-GAA CCG GAC CGA AGC CCG-3') (SEQ ID NO: 8) para los ARN pequeños o a las secuencias VIH-1 específicas que flanquean un pseudonudo usando condiciones de tampón y de reacción descritas anteriormente. Se generaron bibliotecas de secuenciación utilizando un enfoque de PCR modular de dos etapas dirigido que hace posible la generación de forma económica y eficiente de datos para muchas dianas de ARN diferentes. Las reacciones de PCR se realizaron utilizando ADN polimerasa de alta fidelidad Q5 “hot-start”. El cebador de PCR directo (5'-GAC TGG AGT TCA GAC GTG TGC TCT TCC GATC NNNNN-cebador específico de gen-3') (SEQ ID NO: 9) incluye una región específica de Illumina en el extremo 5', seguido de cinco nucleótidos aleatorios para optimizar la identificación de grupos en el instrumento MiSeq™ y termina con una secuencia complementaria al extremo 5' del ARN diana. El cebador inverso (5 'CCC TAC ACG ACG CTC TTC CGA TCT NNNNN-cebador específico de gen-3') (SEQ ID NO: 10) incluye una región específica de Illumina, seguido de cinco nucleótidos aleatorios y una secuencia que es el complemento inverso del extremo 3' del ARN diana. La biblioteca de ADNc se “etiquetó” mediante PCR de 5 ciclos limitada para amplicones o una reacción de PCR 25 ciclos más larga cuando se utilizaron concentraciones muy bajas de ARN. El cebador en exceso, que no se utiliza en los primeros ciclos, se extrajo (kit de limpieza de Micro PCR PureLink™; Invitrogen). La segunda ronda de PCR añadió las secuencias restantes específicas de Illumina necesarias para la amplificación celular en el flujo y se añadieron códigos de barras en las muestras para la multiplexación. El cebador directo (CAA GCA GAA GAC GGC ATA CGA GAT (código de barras) GT GAC TGG AGT TCA GAC) (SEQ ID NO: 11) contiene un código de barras y reconoce una secuencia en el cebador directo de PCR 1. El cebador inverso (AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT T CCC TAC AC GAC GCT CTT CCG) (SEQ ID NO: 12) contiene una secuencia específica de Illumina y reconoce el cebador inverso de PCR 1. La PCR 2 se realizó durante 25 o 5 ciclos para generar la biblioteca final para secuenciación (no superior a 30 ciclos totales). Los experimentos SHAPE-MaP típicos de virus mutantes utilizaron de ~ 150 ng a 200 ng de ARN por condición experimental. Sin embargo, cuando el material es limitante, solo 50 ng de ARN de entrada es suficiente.

[0237] Proyecto de análisis de datos SHAPE-MaP. Se creó un proyecto de análisis de datos, llamado ShapeMapper, que puede ser ejecutado en la mayoría de plataformas basadas en Unix y acepta como lecturas de secuenciación de entrada archivos en formato FASTQ, secuencias de referencia en formato FASTA y un archivo de configuración editado por el usuario. Sin intervención adicional del usuario, el software crea un perfil de reactividad SHAPE y estimaciones de error estándar para cada secuencia de referencia. Se proporcionan otros productos de salida útiles, incluyendo recuentos de mutación, profundidades de secuenciación y estructuras secundarias predichas. El software de análisis incorpora varios programas de terceros. Se requiere Python 2.7 (python.org/); se utiliza Software Bowtie™ 2 para la alineación de lecturas (Langmead, B. & Salzberg, SL Nat. Methods 9, 357-359 (2012)); los perfiles de reactividad se generan utilizando el matplotlib de bibliotecas de Python (Hunter, JD Comput Sci Eng 9, 90-95 (2007)); la predicción de la estructura secundaria utilizó el software RNAstructure™ (Reuter, JS & Mathews, DH BMC Bioinformatics 11, 129 (2010)); y el dibujo de la estructura secundaria utiliza el servicio web Pseudoviewer™ (Byun, Y. y Han, K. Nucleic Acids Res. 34, W416-W422 (2006)).

[0238] Configuración. Se utiliza un archivo de configuración para especificar las secuencias de referencia presentes en cada muestra y cuyas muestras se deben combinar para crear perfiles de reactividad. El formato es flexible, permitiendo la alineación de cada muestra a múltiples dianas de secuencia, así como el tratamiento de múltiples muestras en los análisis unificados. Los parámetros para cada etapa del análisis también se pueden personalizar.

[0239] Recorte de calidad. Se separaron las lecturas de entrada en los archivos mediante la secuenciación del código de barras (este paso está integrado en la mayoría de plataformas de secuenciación). La primera etapa de análisis recorta lecturas mediante calidad de llamada de bases. Cada lectura se recortó en dirección 3’ de la primera llamada de bases con una puntuación de calidad phred por debajo de 10, que corresponde a la precisión esperada del 90%. Las lecturas con 25 o más nucleótidos restantes se copiaron en los archivos nuevos FASTQ para la alineación.

[0240] Alineación de lecturas. Las lecturas fueron alineadas a nivel local a las secuencias de referencia utilizando Bowtie™ 2 (Langmead, B. & Salzberg, SL Nat Procedimientos 9, 357-359 (2012)); se eligieron parámetros para proporcionar alta sensibilidad, para detectar desajustes de nucleótidos únicos, y para permitir deleciones de hasta aproximadamente 200 nucleótidos. La longitud de semillas (-L) fue de 15 nucleótidos. Se permitió una desajuste por semilla (-N). El número máximo de intentos de semillas (-D) se fijó en 20. El máximo de intentos 're-semilla' (-R) se estableció en 3. El “padding” de programación dinámica (-dpad) se fijó en 100 nucleótidos. El bono de coincidencia (-

ma) fue 2. Las penalizaciones por desajuste mínimas y máximas (-mp) fueron 6 y 2, respectivamente. Los huecos abiertos y parámetros de extensión (-rdg, -rfg) fueron 5 y 1, respectivamente. Se utilizó la función de puntuación de alineación mínima por defecto. Se conectó el "soft-clipping". Se utilizó la alineación de extremos apareados por defectos. Los productos de salida Bowtie™ 2 alinearon las lecturas como archivos SAM.

[0241] Análisis de alineación, eliminación de la alineación ambigua y recuento de mutaciones. Se combinaron las lecturas de los extremos apareados en archivos SAM y se seleccionaron llamadas de bases de mayor calidad, donde las parejas de lectura no concordaban. Los desajustes y deleciones contribuyen al recuento de mutaciones; las inserciones se ignoraron. Como la transcripción inversa propensa a errores genera la mayor parte de las mutaciones en cada lectura, se trató un cambio de secuencia que abarcaba múltiples nucleótidos adyacentes como un evento de mutación única localizada en el nucleótido mayoritariamente en 3'. Si se utilizaron cebadores aleatorios, se excluyó una región un nucleótido más largo que la longitud del cebador desde el extremo 3' de cada lectura. Las lecturas con cualidades de mapeo de < 30 se excluyeron. Las deleciones son una parte importante de la señal de mutación, pero las deleciones que se alinean ambiguamente pueden confundir esta señal, evitando la resolución de un solo nucleótido. Para resolver este problema, se llevó a cabo una simple realineación local para identificar y eliminar las deleciones alineadas de forma ambigua. Se almacenó la secuencia de referencia que rodeaba una deleción. A continuación, la deleción se deslizó en dirección 3' y 5' de un nucleótido a la vez hasta una compensación máxima igual a la longitud de deleción. En cada compensación, la secuencia de referencia circundante se comparó con la secuencia almacenada. Si cualquier secuencia de compensación coincidía, esto indicaba una posible alineación alternativa y se excluyó la eliminación. Este algoritmo identificó correctamente las deleciones ambiguas en las regiones homopoliméricas, así como secuencias repetidas.

[0242] Creación de perfiles de reactividad. La tasa de mutación (mutr) en un nucleótido determinado es simplemente el recuento de mutaciones (desajustes y deleciones alineadas sin ambigüedades) dividido por el recuento de lecturas en ese lugar. Las reactividades primas se generaron para cada nucleótido utilizando la siguiente expresión, donde S corresponde a una muestra modificada con SHAPE, U a la muestra no tratada y D a la reacción bajo condiciones de desnaturalización:

$$R = \frac{mutr_S - mutr_U}{mutr_D} \qquad (1)$$

El error estándar (stderr) asociado con la tasa de mutación en un nucleótido determinado en las muestras *S*, *U*, o *D* se calculó como:

$$stderr = \frac{\sqrt{mutr}}{\sqrt{\text{lecturas}}} \qquad (2)$$

El error estándar final de la reactividad en un nucleótido determinado es:

$$= \sqrt{\left(\frac{stderr_S}{mutr_D}\right)^2 + \left(\frac{stderr_U}{mutr_D}\right)^2 + \left(stderr_D \times \frac{(mutr_S - mutr_U)}{mutr_D{}^2}\right)^2} \qquad (3)$$

Las reactividades se normalizaron a una escala estándar que abarcaba de cero (sin reactividad) a ~ 2 (alta reactividad SHAPE), tal como se ha descrito (Hajdin, CE et al. Proc. Natl. Acad. Sci. USA. 110, 5498-5503 (2013)). Los nucleótidos con mayores tasas de mutación que 5% en muestras de control no reactivas se excluyeron del análisis, así como los nucleótidos con profundidades de secuenciación de menos de 10 en cualquier muestra. Es necesaria una profundidad mayor para datos de alta calidad y el modelado de la estructura (ver Figs. 7A y 7B).

[0243] Salida de datos finales. Los perfiles de reactividad SHAPE (.shape) se extrajeron como archivos de texto delimitados por tabuladores con la primera columna indicando el número de nucleótido y la segunda columna indicando la reactividad. Un archivo de reactividad SHAPE-MaP fue también de salida (.map). Este archivo está en el formato de archivo SHAPE con la adición de dos columnas: error estándar y secuencia de nucleótidos. Otro archivo (.csv) que contiene el recuento de mutaciones, profundidades de lectura, tasas de mutaciones, reactividades primas, reactividades normalizadas y los errores estándar para las muestra modificadas por SHAPE, no tratadas y desnaturalizadas. Se

pueden generar archivos que contienen figuras que muestran los histogramas de tasas de mutaciones, profundidades de secuenciación y perfiles de reactividad (.PDF). Estos son útiles en el diagnóstico de problemas experimentales potenciales (incluyendo la profundidad de secuenciación insuficiente o baja eficiencia de mutagénesis).

[0244] Plegado de ARN y dibujo de la estructura automática del proyecto SHAPEMaP. Para las secuencias más cortas que ~4.000 nucleótidos y con suficiente profundidad de lectura, el proyecto automatizado permite estructuras secundarias para ser modeladas automáticamente usando software RNAstructure™, aunque esta capacidad no se utilizó para los ARN en este trabajo. Los archivos de secuencias FASTA se convierten en archivos SEQ requeridos por software RNAstructure™. Las reactividades SHAPE se incorporan en RNAstructure™ como las pseudoenergías libres utilizando parámetros estándar para el reactivo 1M7 (Hajdin, CE et al. Proc. Natl. Acad. Sci. USA. 110, 5.498-5.503 (2013)) (pendiente (-SM) 1,8, intersección (-si) -0,6). Las reactivos SHAPE diferencial están soportados por el software RNAstructure™ y se incorporan en una versión de la proyecto automatizado. Las estructuras predichas se escriben en los archivos .ct. Las estructuras secundarias predichas de menor energía pueden dibujarse y anotarse mediante la reactividad SHAPE. Esta etapa se consulta en el servicio web Pseudoviewer™ (Byun, Y. y Han, K. Nucleic Acids Res. 34, W416-W422 (2006)) a través de una conexión a Internet activa. Un cliente personalizado (pvclient.py) envía peticiones al servidor y obtiene respuestas. Este cliente también se ocupa de la coloración de los nucleótidos por la reactividad. Los dibujos de la estructura coloreada son archivos vectoriales .eps vectoriales. Las estructuras coloreadas también se convierten automáticamente en archivos .xrna (rna.ucsc.edu/rnacenter/xrna/) para la edición manual opcional.

[0245] Filtrado por el factor Z para los datos de SHAPE diferencial. SHAPE-MAP permite que los errores en las mediciones de reactividad SHAPE se estimen a partir de una distribución de Poisson que describe las tasas de mutaciiones medidas en cada nucleótido. El error de reactividad SHAPE estimado por Poisson se puede utilizar para evaluar la significación estadística al comparar dos señales de SHAPE. Se identificaron diferencias significativas entre la reactividad de NMIA y 1M6 usando una prueba del factor Z (Zhang, J., Chung, T. & Oldenburg, KJ Biomol. Screen. 4, 67-73 (1999)). Esta prueba con resolución de nucleótidos compara la diferencia absoluta de las medias con el error de medición asociado:

$$Z_{factor} = 1 - \frac{3(\sigma_{NMIA} + \sigma_{1M6})}{|\mu_{NMIA} - \mu_{1M6}|} \qquad (4)$$

[0246] Cada nucleótido en un experimento de SHAPE-MaP tiene una reactividad calculada μ y un error estándar asociado σ. El umbral de significación para los factores de Z se fijó en $Z > 0$, lo que equivale a una diferencia de reactividad SHAPE para 1M6 y NMIA de al menos 3 desviaciones estándar (s.d.). Las reactividades diferenciales de nucleótido que no cumplieron con este criterio de significación se fijaron a 0.

[0247] Modelado de la estructura. El modelado de estructura secundaria para los ARN de menos de 700 nt de longitud se realizó como se describe (Ejemplo 3 anteriormente en el presente documento; Hajdin, C.E. et al Proc Natl Acad Sci USA. 110, 5498-5503 (2013)); se incorporaron datos de SHAPE diferencial después de la filtración por el factor Z. Para el genoma del ARN de VIH-1, hemos desarrollado un enfoque de modelado automatizado de ventana, implementado en un proyecto de plegado SHAPE-MaP, en el que cálculos de la estructura se dividieron en etapas diseñadas para aumentar la eficiencia computacional, generar estructuras de ARN realistas y reducir efectos finales causados por la selección de un falso extremo 3' o 5' de un pliegue interno en una ventana. Este enfoque facilitó el descubrimiento de pseudonudos, la identificación de pares de bases probables y la generación de estructuras de energía libre mínima. Los cálculos representativos para el plegado de las subunidades ribosómicas, realizados usando un plegado de ua etapa y en ventana mostraron grados altos comparables de precisión y reducciones sustanciales en el 'tiempo real' del cálculo utilizando una estación de trabajo de sobremesa típico del enfoque de plegado en ventana. Para los ARN más cortos, tales como el ARNr 16S, hay una penalización de rendimiento modesto por romper el ARN en ventanas más pequeñas. Sin embargo, para los ARN más largos que ~2,000 nucleótidos, el tiempo de cálculo se escala aproximadamente de forma lineal con la longitud.

[0248] Casi todas las estructuras de ARN funcionales conocidas y bien validadas se modelan de forma idéntica en el estudio actual y la investigación previa (Watts, JM et al., Nature 460, 711-716 (2009) y en la Tabla 4). Mejoras sustanciales en la adquisición de datos digitales (MaP), la mejora de funciones de energía basada en SHAPE del Ejemplo 3 anterior en el presente documento; Hajdin, CE et al. Proc. Natl. Acad. Sci. USA. 110, 5498-5503 (2013) y el análisis automatizado de datos (Fig. 9) favorecen los modelos actuales de estrcutura de VIH-1 sobre los modelos anteriores en las regiones de desacuerdo. Este Ejemplo también refleja otras innovaciones y análisis, en particular que no todas las regiones de un ARN probablemente forman una estructura individual bien definida. Como resultado, un aspecto de este ejemplo es la identificación de regiones en el genoma de ARN del VIH-1 que no forman estructuras individuales bien definidas.

[0249] Predicción de pseudonudo. Durante la primera etapa, el genoma de ARN de longitud completa del VIH-1 se

plegó en ventanas correderas de 600 nt mueve en incrementos de 100 nt utilizando ShapeKnots con pendiente, intersección, parámetros P1 y P2 establecidos a los valores previamente definidos (1,8, -0,6, 0,35, 0,65), utilizando los datos de SHAPE de 1M7 (Ejemplo 3 anteriormente; Hajdin, C.E. et al Proc Natl Acad Sci USA 110, 5498-5503 (2013)). Se calcularon pliegues adicionales en los extremos del genoma para aumentar el número de ventanas que cubren las secuencias terminales. Los pseudonudos predichos se retenían si la estructura aparecía en la mayoría de las ventanas y tenía una baja reactividad SHAPÑE en ambos lados de la hélice con pseudonudo. Se utilizó esta lista de pseudonudos para todas las etapas posteriores de modelado.

[0250] Modelado de la función de partición. La función de partición se calculó utilizando el software Partition™ (Mathews, D.H. RNA 10, 1178-1190 (2004); Reuter, J.S. & Mathews, DH BMC Bioinformatics 11, 129 (2010)) e incluía los datos SHAPE de 1M7 y diferencial y en la penalización de energía libre. La distancia de emparejamiento máxima se fijó a 500 nt. La partición se ejecutó en ventanas de 1600 nt con un tamaño de paso de 375 nt. Se ejecutaron dos ventanas adicionales (longitudes de 1550 nt y 1500 nt) en las secuencias de extremo 5' y 3' para incrementar la toma de muestras en los extremos verdaderos y para reducir el efecto de la selección del sitio de corte no óptimo. Se restringieron seis secuencias (el sitio de unión al cebador, secuencia de dimerización y cuatro pseudonudos conocidos por estar implicados en interacciones inusuales o especiales) como una sola cadena durante los cálculos de la función de partición. A partir de los archivos de la función de partición individuales, se calculó la entropía de Shannon de apareamiento de bases como:

$$H_i = -\sum_{j=1}^{J} p_{i,j} \log_{10} p_{i,j} \qquad (5)$$

donde $p_{i,j}$ es la probabilidad de emparejamiento para los nucleótidos *i* y *j* sobre todos las parejas *J* potenciales (Huynen, M., et al. J. Mol. Biol. 267, 1104-1112 (1997)). Después de este cálculo, se recortaron 300 nt de los extremos 5' y 3' de cada ventana que no flanqueaban los extremos 5' y 3' verdaderos del ARN. Este cálculo retuvo valores internos más consistentes y descartó valores sesgados por efectos finales. Las ventanas de la entropía de Shannon se combinaron obteniendo el promedio, creando un único archivo de entropía.

[0251] A continuación, los pares probables individuales de cada ventana se recortaron utilizando el mismo enfoque descrito para la entropía de Shannon. Los pares de bases que se formaban con una probabilidad de menos de 10^{-4} fueron retirados para disminuir el tiempo de cálculo. Se combinaron las ventanas y todos los pares restantes fueron promediados sobre la totalidad de las ventanas en las que podrían haber aparecido. Se desarrolló una escala de colores heurística a partir del archivo de partición combinado para indicar la probabilidad relativa de un par que aparece en la estructura final. Los pares resultantes se representaron como arcos (Fig. 8). Los pares de bases con una mayor probabilidad que 0,99 se usaron como limitaciones de doble cadena en la siguiente etapa.

[0252] Modelado de energía libre mínima. Se generó una estructura de energía libre mínima usando Fold™ (Reuter, JS & Mathews, DH BMC Bioinformatics 11, 129 (2010)), datos de SHAPE de 1M7 y datos de SHAPE diferencial. Se utilizó un tamaño de ventana de 3000 nt con un tamaño de paso de 300 nt para generar estructuras potenciales sobre cada ventana. También se generaron cuatro pliegues (3100 nt, 3050 nt, 2950 nt y 2900 nt desde los extremos) para aumentar el número de los modelos de estructura en los extremos. Estos pliegues de ventanas superpuestas se combinaron a continuación en una estructura completa mediante la comparación de pares de bases comunes a cada ventana y que requieren que los pares en la estructura final aparezcan en la mayoría de las ventanas potenciales. Como etapa final, se incorporaron hélices con pseudonudos.

[0253] Análisis de error y determinación de un número mínimo de lecturas necesario para la modelación exacta de la estructura del ARN. Las tasas de mutación para cada una de las señales contributivas (modificado con SHAPE, no tratado, desnaturalizado) se modelaron utilizando una distribución de Poisson porque eventos discretos de lecturas individuales contribuyen a la señal global. La varianza de una distribución de Poisson es igual al número de observaciones; por lo tanto, el error estándar (SE) de una tasa "verdadero" puede ser modelado como:

$$SE_{\text{tasa}} = \frac{\sqrt{\lambda}}{\text{lecturas}} = \frac{\sqrt{\text{tasa}}}{\sqrt{\text{lecturas}}} \qquad (6)$$

donde λ es el número de eventos (mutaciones observadas), *lecturas* es la profundidad de lectura en el nucleótido modelado (mutaciones y no mutaciones), y la *tasa* es el número de eventos por lectura. Como era de esperar, el "bootstrapping" del error estándar de la reactividad SHAPE mostró una relación de potencia $x^{-1/2}$ en función de la profundidad de lectura.

[0254] Usando un ARN profundamente secuenciado (mayor que 50.000 lecturas para cada nucleótido), se conoce el número de eventos de mutación esperados a profundidades de lectura mucho más inferiores con gran precisión. Los eventos de mutaciones pueden ser muestreados a partir de una distribución de Poisson en todo el ARN para crear perfiles de datos SHAPE plausibles. Para determinar un umbral mínimo para el número de lecturas necesario para un modelado dirigido por SHAPE preciso de estructura secundaria, se examinó el ARNr 16S porque se modela mal en ausencia de datos experimentales (~ 50% de sensibilidad). Para cada profundidad de lectura simulada, se crearon 100 trayectorias SHAPE basadas en la varianza de Poisson esperada a la profundidad de lectura simulada y se modelaron usando software RNAstructure Fold ™ (Fig. 7B). Como era de esperar, la precisión del modelado mejoró a medida que aumentaba la profundidad de lectura. Para un modelado preciso de estructuras con resolución de nucleótidos, se sugieren al menos 5.000 lecturas, pero incluso a 500 lecturas, la medición es útil para el modelado de la estructura (Fig. 7B).

[0255] Cálculo del nivel de éxito y comparación con otros informes. El análisis de la estructura por SHAPE-MaP como se lee mediante secuenciación masivamente en paralelo presenta una herramienta valiosa para la interrogación estructural del ARN en un solo nivel de nucleótidos. Se han desarrollado otros enfoques con objetivos similares. Para comparar el requisito de profundidad de lectura de SHAPE-MaP (y su lectura de perfiles mutaciones) con otros enfoques, se calculó un "nivel de éxito". La medida del nivel de éxito cuantifica la señal total restada de la base por nucleótido de la transcripción:

$$\text{nivel de éxito} = \frac{\text{sucesos totales}_S - \frac{\text{prof. lectura}_S}{\text{prof. lectura}_B} \times \text{sucesos totales}_B}{\text{longitud de transcrito}} \qquad (7)$$

donde los subíndices *S* y *B* indican la muestra experimental y control de base, respectivamente; *sucesos* son paradas o mutaciones de secuencias detectadas por ligadura, dependiendo del procedimiento de lectura, y *profundidad de lectura* corresponde a la mediana del número de lecturas que solapan cada nucleótido en la transcripción. Se obtuvo un nivel de éxito de 160 para el ARNr 16S. Como los recuentos de mutaciones en SHAPE MaP son proporcionales a las profundidades de lectura, la relación entre la profundidad de lectura de secuenciación lee profundidad y el nivel de éxito se calculó dividiendo nuestro nivel de éxito observado por la profundidad de lectura media en una condición experimental. Se requiere un nivel de éxito de ~ 15 para recuperar completamente la información de la estructura de ARN como se investiga mediante SHAPE, aunque se obtuvieron modelos de estructura muy útiles consistentemente a niveles de éxito tan bajos como 5 (Fig. 7B). La investigación y el modelado de la estructura del ARN de alta resolución requieren que la mayor parte o la totalidad de un ARN sea investigado a un alto nivel de éxito. Regiones individuales sondeadas a bajos niveles de éxito, incluso si el nivel de éxito promedio general es ≥ 5, son susceptibles de contener errores notables. En un análisis paralelo de experimentos para la estructura de ARN (PARS), se requería un umbral mínimo de 1 parada de lectura promedio por nucleótido de transcripción (Kertesz, M. et al., Nature 467, 103-107 (2010); Wan, Y. et al. Nature 505, 706-709 (2014)); correspondiente al nivel de éxito de 1, suponiendo una base cero para los datos de escisión enzimática. Del mismo modo, un informe que describe una secuenciación de estructura con un sondeo químico con DMS, utilizó un umbral similar de ≥ 1 parada promedio por nucleótido A o C 10; esto corresponde a un nivel de éxito estimado (como se define aquí) de 0.2, asumiendo una relación señal:base de 1,7 (estimada a partir de Ding, Y. et al., Nature 505, 696-700 (2014); datos ampliados de la figura 1d) y que la mitad de todos los nucleótidos de transcripción son A o C. Se requirió un mínimo de 15 lecturas por A o C en promedio por los creadores de DMS-seq11. Esto corresponde a un nivel de éxito de 3,3, asumiendo una relación señal:base de 1,8 (estimada a partir de Rouskin, S., et al Nature 505, 701-705 (2014), la figura 1c). El análisis de evaluación comparativa ("benchmarking") y "bootstrapping" para modelar la exactitud descrita aquí (Fig. 7B) no se ha aplicado en los análisis de estructura de ARN basados en la secuenciación masivamente en paralela anteriores (Kertesz, M. et al., Nature 467, 103-107 (2010), Underwood, JG et al. Nat. Methods 7, 995-1001 (2010), Lucks, JB et al. Proc. Natl. Acad. Sci. USA. 108, 11063-11068 (2011), Ding, Y. et al., Nature 505, 696-700 (2014), Rouskin, S., et al., Nature 505, 701-705 (2014), Wan, Y. et al., Nature 505, 706-709 (2014)). Este análisis del nivel de éxito hace hincapié en que, aunque se han realizado varios estudios previos en los que el complemento completo de los ARN en un transcriptoma determinado estaba presente durante la fase de sondeo del experimento, sólo unos pocos miles de nucleótidos en cada caso fueron muestreados a una profundidad consistente con la recuperación de la información de la estructura subyacente obtenible usando DMS o sondas de enzimas. (Underwood, JG et al. Nat. Methods 7, 995-1001 (2010)).

[0256] Descubrimiento algorítmico de regiones de VIH-1 con una baja entropía de Shannon y baja reactividad SHAPE. La superposición de regiones con baja reactividad SHAPE y baja entropía de Shannon se utilizó para identificar las regiones propensas a tener una única estructura bien determinada. En primer lugar, se calcularon la mediana local de la reactividad SHAPE y la entropía de Shannon sobre ventanas correderas centradas de 55 nt. A continuación, se seleccionaron las regiones en las que la mediana local cayó por debajo de la media global para más de 40 nt, tanto en la entropía de Shannon como en la reactividad SHAPE. Las regiones se combinaron si estaban separadas por menos de 10 nt. Por último, las regiones se ampliaron para incluir estructuras secundarias anidadas a partir del modelo de energía libre predicha mínima.

[0257] Para excluir la posibilidad de que las regiones estructuradas algorítmicamente descubiertas solaparan elementos de ARN conocidos simplemente por casualidad, se generó un conjunto aleatorizado de segmentos y se calculó la distribución esperada de nucleótidos de superposición (Tabla 4). Se mantuvieron el mismo número y la longitud de los segmentos, pero sus localizaciones se pusieron al azar en el genoma de 9173 nt. De 105 ensayos, solamente 219 mostraron un solapamiento más grande que el observado, lo que corresponde a un valor de P de 0,002.

[0258] Mutagénesis de VIH-1. Se introdujeron mutaciones en los subclones de VIH-1 pNL4-3 que abarcaban la región de interés mediante mutagénesis dirigida al sitio (QuikChange ™ XL, Agilent) y se verificaron por secuenciación. El fragmento del subclón mutado se reintrodujo en el plásmido pNL4-3 de longitud completa (Adachi, A. et al. J. Virol. 59, 284-291 (1986)). La secuencia del genoma mutante de longitud completa se verificó mediante secuenciación automatizada convencional utilizando 16 o más cebadores superpuestos. Los virus de los plásmidos NL4-3 mutantes y de tipo salvaje se produjeron por transfección, tal como se describe anteriormente (para dar ~ 12 ng de ARN viral por ml de sobrenadante viral). La producción de viriones se midió mediante un ensayo p24 (kit p24 del VIH AlphaLISA™, PerkinElmer AL207C). Los virus se midieron para determinar la infectividad en células 53 indicadoras de TZM-bl (usando tampón Glo Lysis™ y el sistema de ensayo de luciferasa; Promega).

[0259] Se diseñaron mutaciones para interrumpir la secuencia primaria de pseudonudos, pero manteniendo la identidad de los aminoácidos en las secuencias de codificación. La hélice primaria de pseudonudos en $U3_{PK}$ se solapa parcialmente con un sitio de unión para el factor de transcripción SP1. Existe un total de tres sitios de unión SP1 consecutivos en el VIH-1. Dos mutaciones puntuales introducidas en la construcción de $U3_{PK}$ se solapan con el tercer sitio de unión SP1. El trabajo previo ha demostrado que sólo se requiere un único sitio de unión para una función viral completa (Harrich, D. et al. J. Virol. 63, 2585-2591 (1989). La proteína SP1 tolera la variación en varios lugares en la secuencia de consenso de unión y las mutaciones introducidas aquí mantuvieron un sitio de unión SP1 canónico. Para analizar los efectos sobre la producción de virus que resultan de la mutación de SP1, las mismas mutaciones de U3 se introdujeron en la región U3 5' del clon U3 pNL4-3 (con y sin mutaciones en 3' concomitantes). Los virus resultantes no tenían fenotipo (las mutaciones U3 en 5' solas) o el mismo fenotipo que el mutante de $U3_{PK}$ original (los que contienen mutaciones de U3 en 5' y 3'), lo que sugiere que la alteración del sitio de unión SP1 no interrumpió la producción de ARN viral. La especie de doble mutante se utilizó para los ensayos de propagación y de competición viral, tal como se describe a continuación.

[0260] Separación de datos SHAPE-MaP de $U3_{PK}$ mutante y NL4-3 de tipo salvaje. Los datos de SHAPE-MaP de cebadores específicos de genes para $U3_{PK}$ (para la construcción en la que solo se mutó el U3 en 3') revelaron que los tres nucleótidos dirigidos para la mutagénesis mostraron inusualmente altas tasas de mutación cuando se alinearon con la secuencia mutante, lo que sugiere la presencia de múltiples poblaciones de secuencia. La cuantificación de la abundancia relativa de cada secuencia variante mostró que el 61,8% de las lecturas contenía la secuencia nativa, el 36,0% contenía la secuencia mutante diseñada y una pequeña fracción (2,2%) contenía otras secuencias. Estas relaciones indican que la recombinación entre la secuencia nativa y regiones U3 mutantes se produjo durante la transfección, produciendo virus VIH-1 en mejores condiciones que creció más rápidamente que el virus mutante durante el cultivo de virus. Este virus mutante se cultivó en células H9 (ATCC, Manassas, Virginia, Estados Unidos de América) durante 3 semanas antes de la extracción del ARN y las pruebas con SHAPEMaP. Los perfiles de reactividad para la secuencia mutante diseñada y la de tipo salvaje fueron creados computacionalmente separando las lecturas después de la alineación. Además, los datos de reactividad de SHAPE-MaP para los tres nucleótidos mutados se obtuvo seleccionando las lecturas que reconocen dos nucleótidos mutados a la vez para asignar de tipo salvaje o mutante, lo que permite la variación en el tercer nucleótido para determinar la tasa de mutación. Este enfoque es ampliamente aplicable para sondear químicamente poblaciones de ARN en las que cada fracción de secuencia es mayor que la tasa esperada de mutación por nucleótido inducida por transcripción inversa (~ 1% en este trabajo).

[0261] Ensayos de replicación del VIH. Los virus se ensayaron para propagación de célula a célula en líneas de células Jurkat (ATCC) y líneas de células T H9. Las células se confirmaron que estaban libres de contaminantes biológicos. Se normalizaron inóculos de virus mediante la infectividad TZM-bl a una multiplicidad de infección baja (menos de 0,01) antes de la infección y se utilizaron para infectar 5 x 105 células en 1 ml de medio RPMI-1640 en placas de 12 pocillos; las infecciones se llevaron a cabo por duplicado. Se realizó un cambio de medio completo 3 d después de la infección (d.p.i.), y el medio de cada pocillo se recogió y se sustituyó a 4 dpi, 5 dpi y 6 dpi. Las concentraciones virales fueron cuantificados mediante un ensayo p24 (AlphaLISA™ HIV; PerkinElmer).

[0262] Ensayos de competición de VIH. Virus de secuencia mutante y nativa se mezclaron en una proporción de 10:1 y se utilizó para infectar 5×10^5 células Jurkat en 1 ml de volumen total en placas de 12 pocillos. Las infecciones se llevaron a cabo utilizando como mucho la mitad mutante y 20 veces menos de virus de tipo salvaje con respecto a los ensayos de replicación viral. Los experimentos de competición se llevaron a cabo por duplicado. El medio se recogió inicialmente a 2 d.p.i. para representar el inóculo inicial. El medio se recogió a 3 d.p.i., 4 d.p.i., 5 d.p.i. y 6 d.p.i., y se cuantificó p24 (proteína de la cápside) en un medio (kit p24 de VIH AlphaLISA™). Se requirió que los niveles de p24 aumentaran exponencialmente hasta el día 6 para asegurar que las células no infectadas estuvieran en exceso a lo largo de las infecciones. El ARN viral se purificó del medio (mini kit de ARN viral QIAamp™, Qiagen) y se llevó a cabo la transcripción inversa usando superscript™ III (Life Technologies) usando los cebadores Primer ID (Jabara, CB, et al. Proc. Natl. Acad. Sci. USA. 108, 20.166-20.171 (2011) para colocar un código de barras a cada ADNc producido y

eliminar sesgos de población introducidos durante la PCR. La preparación de la muestra posterior se realizó como se ha descrito anteriormente para SHAPE-MaP usando cebadores específicos de genes dirigidos. Después de la secuenciación, las lecturas de los extremos apareados se fusionaron en lecturas sintéticas más largas usando un ajuste de longitud rápida de lecturas cortas (FLASH) (magoc, T. & Salzberg, SL Bioinformatics 27, 2957-2963 (2011)). A continuación, las lecturas sintéticas se alinearon con la secuencia de NL4-3 esperada para las regiones reconocidas usando software Bowtie™ 2 (Langmead, B. & Salzberg, Nat. Methods 9, 357-359 (2012)) (utilizando parámetros por defecto). Se construyó una lectura de consenso para cada PrimerID basada en una medición de votación de puntuación Phred. Se requirió que los identificadores (IDs) que se emparejaban con cualquiera de las secuencias nativa o mutante tuvieran las mutaciones puntuales previstas en todas las ubicaciones con el fin de ser consideradas. La fracción de ID mutantes se expresó como el número de ID mutantes de la suma de los ID mutantes y nativos. La capacidad relativa de virus mutantes se determinó a partir de la tasa de cambio de la relación de mutante a NL4-3 medida con el tiempo (Resch, W., et al. J. Virol. 76, 8.659-8.666 (2002)).

[0263] Cálculo de diferencias en las reactividades SHAPE en mutantes de pseudonudo. Las mediciones de error estándar de reactividades SHAPE, estimadas a partir de la distribución de Poisson, dependen del número de lecturas obtenido para cada muestra. La observación de que el error estándar disminuye con la inversa del cuadrado de la profundidad de lectura se utilizó para derivar una ecuación de escalado que normaliza a una profundidad común de 8000 lecturas para justificar las diferencias en la profundidad de secuenciación entre las muestras. El factor de escala de error estándar, f_0, se calculó para cada muestra basándose en la profundidad de lectura promedio, r_{ave}, del componente secuenciado más bajo (condiciones modificadas or SHAPE, no tratadas y desnaturalizantes) que contribuye al perfil de reactividad SHAPE:

$$f_0 = \frac{(r_{ave})^{-\frac{1}{2}}}{(8000)^{-\frac{1}{2}}} \qquad (8)$$

[0264] Después de escalar los errores estándar a una profundidad de lectura común, la significación para cada punto se calculó usando un ensayo de factor z modificado (Zhang, J., et al. J. Biomol. Screen. 4, 67-73 (1999)) que requiere diferencias mayores de 1,96 veces la suma de los errores estándar. Las puntuaciones del factor Z mayores que cero se consideraron significativos:

$$Z_{factor} = 1 - \frac{1.96(\sigma_{PK} + \sigma_{WT})}{|\mu_{PK} - \mu_{WT}|} \qquad (9)$$

[0265] Los cambios de reactividad aislados pueden ser vistos como ruido en el contexto de un desplazamiento de la estructura global resultante de la alteración de un pseudonudo. Por lo tanto, además de la prueba del factor z, se requirió que las diferencias fueran consecutivas.

Tabla 4

Estructuras en el genoma de ARN de VIH-1 que son similares entre los modelos actuales y del 2009 (Watts, JM et al., Nature 460, 711-716 (2009))

Nucleótidos participantes	Elemento o elementos estructurales
1-56	5’ TAR
58-104, 399-484	5’ pseudonudo y hélices circundantes
105-344	elemento PBS, DIS y Psi (ampliamente consistente)
581-657	horquillas ramificadas
1177-1198, 1294-1312	hélice
1214-1247	horquilla
1418-1451	tallo-bucle
1531-1541	horquilla muy corta
1568-1707	desplazamiento del marco (ampliamente consistente)
2015-2171	hélice continua más larga y horquillas cercanas
2629-2654	tallo-bucle
2714-2745	2 horquillas cortas
2781-2802	horquilla
3140-3149	horquilla muy corta
4754-4770	horquilla corta

5267-5343	tallos-bucles cortos
5600-5639	tallo-bucle
5793-5878, 5985-6013	2 hélices y 2 horquillas
5943-5964	horquilla
6869-6926, 7037-7065	hélices y bucles internos
7244-7600	RRE
7991-8015	horquilla
8518-8641	PPT y tallos-bucles flanqueantes (ampliamente consistente)
8867-8906	horquilla
9042-9067	2 horquillas cortas en dirección 5’ de 3’ TAR
9070-9138	3’TAR

Tabla 5

Elementos de ARN con estructuras funcionales verificadas utilizados para examinar la significación de motivos algorítmicamente descubiertos en el genoma del VIH-1. La numeración de secuencia se basa en la referencia NL4-3 con el nucléotido de ARN más 5’ definido como +1

Elemento de ARN	**Posición de nucleótido**
5’ LTR R	1-97
5’ LTR U5	98-180
Elemento PSI	227-353
Desplazamiento de marco ribosomal	1559-1719
Donadores de corte y empalme	286-297, 5587-5598
Tracto de polipurina central (CPPT)	4331-4345
Aceptores de corte y empalme	5303-5323, 5465-5522, 8700-8713
Elemento sensible a Rev (RRE)	7294-7550
V1/V2	6097-6174
V3	6644-6756
V4	6921-7020
V5	7149-7179
Tracto de polipurina (PPT)	8605-8619
5’ U5/R pseudonudo (5’$_{PK}$)	79-86, 442-449
ENV$_{PK}$	5666-5676, 5971-5981
RT$_{PK}$	2682-2689, 2828-2835
U3$_{PK}$	9027-9033
3’ LTR R	8622-9076
3’ LTR U3	9077-9173

Tabla 6

Recuento de lecturas de secuenciación representativas para análisis SHAPE-MaP de VIH-1

Muestra	**Tipo de ejecución**	**Lecturas apareadas**	**Lecturas no apareadas**	**Lecturas totales**
1M7 modificado	Hi-Seq 2x100	4.141.728	188.353	4.330.081
1M6 modificado	Hi-Seq 2x100	3.851.060	148.671	3.999.731
NMIA modificado	Hi-Seq 2x100	4.406.294	207.074	4.613.368
1M7 desnaturalizado	Mi-Seq 2x150	1.017.303	16.836	1.034.139
1M6 desnaturalizado	Mi-Seq 2x150	1.231.002	19.637	1.250.639
NMIA desnaturalizado	Mi-Seq 2x150	1.245.518	20.934	1.266.452
No tratada (DMSO)	Hi-Seq 2x100	6.687.508	281.251	6.968.759

REFERENCIAS

[0266]

Adachi, A. et al. 1986. J. Virol. 59, 284-291.
Alvarez DE, et al. 2005. J Virol 79: 6631-6643.
Archer EJ, et al. 2013. Biochemistry 52: 3182-3190.
Bailor MH, et al. 2011. Curr Opin Struct Biol 21: 296-305.
Beckman, R. A., et al. Biochemistry 24, 5810-5817 (1985).
Brierley, I., et al. 2007. Nat Rev Micro 5, 598-610.

Bustamante C. 1999. J Mol Biol 293: 271-281.
Byun, Y. & Han, K. Nucleic Acids Res. 34, W416-22 (2006).
Chen, T. & Romesberg. 2014. FEBS Lett. 588, 219-229.
Cordero P, et al. 2012. Biochemistry 51: 7037-7039.
Dann CE, et al. 2007. Bioinformatics 25: 1974-1975.
Deigan KE, et al. 2009. Proc Natl Acad Sci 106: 97-102.
Derdeyn, C. A. et al. 2000. J. Virol. 74, 8358-8367.
Dethoff EA, et al. 2012. Nature 482:322-330.
Ding Y et al. 2014. Nature 505:696-700.
Doty P et al. 1959. Proc. Natl. Acad. Sci. 45: 482-499
Ghadessy, FJ & Holliger. 2007. Methods Mol. Biol. 352: 237-248.
Gherghe CM, et al. 2008. J Am Chem Soc 130: 8884-8885.
Gherghe C et al. Proc. Natl. Acad. Sci. 107, 19248-19253 (2010).
Gilmartin, GM, et al. 1992. EMBO J. 11: 4419-4428.
Grohman JK, et al. 2013. Science 340: 190-195.
Hajdin CE, et al. 2010. RNA 16: 1340-1349.
Hajdin CE, et al. 2013. Proc Natl Acad Sci 110:5498-5503.
Harrich, D. et al. J. Virol. 1989. 63, 2585-2591.
Hunter, J. D. Matplotlib: Comput. Sci. Eng. 90-95 (2007).
Huynen M, et al. 1997. J. Mol. Biol. 267: 1104-1112.
Jabara, C. B., etal. 2011. Proc. Natl. Acad. Sci. 108, 20166-20171.
Jin Y, Yang Y, Zhang P. 2011. RNA Biol 8: 450-457.
Karabiber F, et al. 2013. RNA 19: 63-73.
Kertesz, M. et al. 2010. Nature 467: 103-107.
Kladwang W, et al. 2011a. Nat Chem 3: 954-962.
Kladwang W, et al. 2011b. Biochemistry 50: 8049-8056.
Klasens, BI, et al. 1999. Nucleic Acids Res. 27: 446-454.
Leonard CW, et al. 2013. Biochemistry 52: 588-595.
Leontis NB, Lescoute A, Westhof E. 2006. Curr Opin Struct Biol 16: 279-287.
Low JT, Weeks KM. 2010. Methods 52: 150-158.
Lorsch, J.R, et al. 1995. Nucleic Acids Res. 23: 2811-2814.
Lucks, J. B. et al. 2011. Proc. Natl. Acad. Sci. 108: 11063-11068.
Magoč, T. & Salzberg. 2011. Bioinformatics 27, 2957-2963.
Matathias, A., Fox, D. & Crouse, J. 1999. SuperScript II RNase H- reverse transcriptase. 18064-3, (Focus On, Life Technologies).
Mathews DH, Turner DH. 2006. Curr Opin Struct Biol 16: 270-278.
Mathews DH, et al. 2004. Proc Natl Acad Sci 101: 7287-7292.
Mathews DH. 2004. RNA 10: 1178-1190.
Matoulkova, E, et al. 2012. RNA Biol 9: 563-576 Mauger DM, Siegfried NA, Weeks KM. 2013. FEBS Lett 587: 1180-1188.
Mauger, DM & Weeks, KM. 2010. Nat. Biotechnol 28: 1178-1179. McGinnis JL, et al. 2012. J Am Chem Soc 134: 6617-6624.
McGinnis JL & Weeks, K. M. 2014. Biochemistry. 53: 3237-3247.
Merino EJ, et al. 2005. J Am Chem Soc 127:4223-4231.
Mortimer SA, Weeks KM. 2007. J Am Chem Soc 129: 4144-4145.
Mortimer SA, Weeks KM. 2009. Proc Natl Acad Sci 106: 15622-15627.
Munroe R. 2012. Star Ratings. http://xkcd.com/1098/.
Paillart JC., et al. 2002. J. Biol. Chem. 277: 5995-6004.
Patterson, JT, et al. 2006. RNA Biol 3: 163.
Pollom, E. et al. 2013. PLoS Pathog. 9: e1003294.
Resch, W, et al. 2002. J. Virol. 76, 8659-8666.
Reuter JS, Mathews DH. 2010. BMC Bioinformatics 11: 129.
Rice GM, et al. 2014. RNA 20: 846-854.
Rivas E et al. 2012. RNA 18: 193-212.
Rohl CA, et al. 2004. Methods Enzymol/Numerical Computer Methods 383: 66-93.
Rouskin S, et al. 2014. Nature 505: 701-705.
Sharp PA. 2009. Cell 136: 577-580.
Spitale, RC et al. 2013. Nat. Chem. Biol. 9: 18-20.
Steen KA, Rice GM, Weeks KM. 2012. J Am Chem Soc 134: 13160-13163.
Staple DW & Butcher SE. 20015. PLoS Biol. 3: e213.
Staple, DW. et al. 2012. Nat. Meth. 9, 357-359 (2012).
Talkish, J, et al. RNA 20, 713-720 (2014).
Tyrrell, J, et al. 2013. Biochemistry 52, 8777-8785.
Underwood, J. G. et al. 2010.Nat. Meth. 7: 995-1001.
Wan, Y. et al. Nature. 2014. 505, 706-709.
Watts JM, et al. 2009. Nature 460: 711-716.

Weeks KM. 2010. Curr Opin Struct Biol 20: 295-304.
Weeks, KM, 2011. Proc. Natl. Acad. Sci. 108: 10933-10934.
Weeks, KM & Mauger, DM. 2011. Acc. Chem. Res. 44: 1280-1291.
Wilkinson KA, et al. 2006. Nat Protoc 1:1610-1616.
Wilkinson KA, et al. 2008. PLoS Biol 6: e96
Williams, R. et al. Nat. Meth. 3, 545-550 (2006).
Zhang, J., et al. 1999. J. Biomol. Screen. 4, 67-73.

LISTADO DE SECUENCIAS

[0267]

<110> The University of North Carolina at Chapel Hill
Week, Kevin M
Siegfried, Nathan
Homan, Philip
Busan, Steven
Favorov, Oleg

<120> DETECCIÓN DE MODIFICACIONES QUÍMICAS EN EL ARN

<130> 421/338 PCT

<150> US 61/887,614
<151> 2013-10-07

<160> 12

<170> PatentIn version 3.5

<210> 1
<211> 120
<212> ARN
<213> Escherichia coli

<400> 1
ugccuggcgg ccguagcgcg gugguccccac cugaccccau gccgaacuca gaagugaaac 60

gccguagcgc cgaugguagu gugggucuc cccaugcgag aguagggaac ugccaggcau 120

<210> 2
<211> 158
<212> ARN
<213> Fusobacterium nucleatum

<400> 2
gauaugagga gagauuucau uuuaaugaaa caccgaagaa guaaaucuuu cagguaaaaa 60

ggacucauau uggacgaacc ucuggagagc uuaucuaaga gauaacaccg aaggagcaaa 120

gcuaauuuua gccuaaacuc ucagguaaaa ggacggag 158

<210> 3
<211> 154
<212> ARN
<213> Bacillus subtilis

<400> 3
cuucguuagg ugaggcuccu guauggagau acgcugcugc ccaaaaaugu ccaaagacgc 60

caauggguca acagaaauca ucgacauaag gugauuuuua augcagcugg augcuugucc 120

uaugccauac agugcuaaag cucuacgauu gaag 154

<210> 4
<211> 425

<212> ARN
<213> Tetrahymena thermophila

<400> 4
cucucuaaau agcaauauuu accuuuggag ggaaaaguua ucaggcaugc accugguagc 60

uagucuuuaa accaauagau ugcaucgguu uaaaaggcaa gaccgucaaa uugcgggaaa 120

ggggucaaca gccguucagu accaagucuc aggggaaacu uugagauggc cuugcaaagg 180

guaugguaau aagcugacgg acaugguccu aaccacgcag ccaaguccua agucaacaga 240

ucuucuguug auauggaugc aguuccagac uaaaugucgg ucggggaaga uguauucuuc 300

ucauaagaua uagucggacc ucuccuuaau gggagcuagc ggaugaagug augcaacacu 360

ggagccgcug ggaacuaauu uguaugcgaa aguauauuga uuaguuuugg aguacucgua 420

aggua 425

<210> 5
<211> 80
<212> ADN
<213> Escherichia coli

<400> 5
ggactcgggg tgcccttctg cgtgaaggct gagaaatacc cgtatcacct gatctggata 60

atgccagcgt agggaagttc 80

<210> 6
<211> 158
<212> ADN
<213> Tetrahymena thermophila

<400> 6
gaattgcggg aaaggggtca acagccgttc agtaccaagt ctcaggggaa actttgagat 60

ggccttgcaa agggtatggt aataagctga cggacatggt cctaaccacg cagccaagtc 120

ctaagtcaac agatcttctg ttgatatgga tgcagttc 158

<210> 7
<211> 268
<212> ADN
<213> Bacillus stearothermophilus

<400> 7
gttaatcatg ctcgggtaat cgctgcggcc ggtttcggcc gtagaggaaa gtccatgctc 60

gcacggtgct gagatgcccg tagtgttcgt ggaaacacga gcgagaaacc caaatgatgg 120

taggggcacc ttcccgaagg aaatgaacgg agggaaggac aggcggcgca tgcagcctgt 180

agatagatga ttaccgccgg agtacgaggc gcaaagccgc ttgcagtacg aaggtacaga 240

acatggctta tagagcatga ttaacgtc 268

<210> 8
<211> 18
<212> ADN
<213> Secuencia artificial

<220>

<223> Cebador de oligonucleótidos sintetizado artificialmente

<400> 8
gaaccggacc gaagcccg 18

<210> 9
<211> 36
<212> DNA
<213> Secuencia artificial

<220>
<223> Cebador de oligonucleótidos sintetizado artificialmente

<220>
<221> misc_feature
<222> (1)..(36)
<223> El cebador de oligonucleótidos puede tener opcionalmente un cebador específico del gen unido a su extremo 3'

<400> 9
gactggagtt cagacgtgtg ctcttccgat cnnnnn 36

<210> 10
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador de oligonucleótidos sintetizado artificialmente

<220>
<221> misc_feature
<222> (1)..(29)
<223> El cebador de oligonucleótidos puede tener opcionalmente un cebador específico del gen unido a su extremo 3'

<400> 10
ccctacacga cgctcttccg atctnnnnn 29

<210> 11
<211> 41
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador de oligonucleótidos sintetizado artificialmente

<220>
<221> misc_feature
<222> (24)..(25)
<223> El cebador de oligonucleótidos puede incluir opcionalmente un código de barras unido al 24º y/o 25º nucleótido

<400> 11
caagcagaag acggcatacg agatgtgact ggagttcaga c 41

<210> 12
<211> 54
<212> ADN
<213> Secuencia artificial

<220>
<223> Cebador de oligonucleótidos sintetizado artificialmente

<400> 12
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccg 54

REIVINDICACIONES

1. Procedimiento para la detección de datos estructurales en un ácido nucleico, comprendiendo el procedimiento:
a. proporcionar un ARN que ha sido expuesto a un reactivo que proporciona una modificación química, en el que el reactivo comprende un electrófilo que modifica selectivamente nucleótidos sin restricciones en el ARN para formar un aducto 2'-O- de ribosa covalente o en el que el reactivo es sulfato de dimetilo, DMS y forma aductos en la posición N1 de la adenosina y la posición N3 de la citosina;
b. sintetizar un ácido nucleico usando una transcriptasa inversa y el ARN proporcionado en la etapa (a) como una plantilla, en el que la síntesis tiene lugar en presencia de Mn^{2+}, en el que la transcriptasa inversa lee a través de la modificación química en el ARN proporcionado en la etapa (a) para producir de este modo un nucleótido no complementario o una deleción en el ácido nucleico resultante en el sitio de la modificación química; y
c. detectar el nucleótido no complementario al nucleótido plantilla o la deleción mediante la secuenciación del ácido nucleico, en el que la información de la secuencia se alinea con la secuencia de ARN proporcionada en la etapa (a).

2. Procedimiento, según la reivindicación 1, que comprende detectar dos o más modificaciones químicas.

3. Procedimiento, según la reivindicación 2, en el que la transcriptasa inversa lee a través de múltiples modificaciones químicas para producir múltiples nucleótidos no complementarios o deleciones y la detección comprende detectar cada nucleótido no complementario o deleción.

4. Procedimiento, según la reivindicación 1, en el que el reactivo es 1M7, 1M6, NMIA, DMS, o combinaciones de los mismos.

5. Procedimiento, según cualquiera de las reivindicaciones 1-4, en el que el ARN proporcionado en la etapa (a) está presente en o deriva de una muestra biológica.

6. Procedimiento, según cualquiera de las reivindicaciones 1-5, en el que la transcriptasa inversa es una transcriptasa inversa nativa o una transcriptasa inversa mutante.

7. Procedimiento, según cualquiera de las reivindicaciones 1-6, en el que la detección del nucleótido no complementario al nucleótido plantilla o la deleción comprende emplear secuenciación masiva en paralelo en el ácido nucleico.

8. Procedimiento, según cualquiera de las reivindicaciones 1-7, que comprende amplificar el ácido nucleico.

9. Procedimiento, según cualquiera de las reivindicaciones anteriores, que comprende además:
d. producir archivos de salida que comprenden datos estructurales para el ARN proporcionado en la etapa (a).

10. Procedimiento, según la reivindicación 9, que comprende, además, normalizar, comparar y/o unir diferentes conjuntos de datos que contienen información estructural del ácido nucleico, tal como, pero no limitado a, la información estructural del ARN.

11. Procedimiento, según cualquiera de las reivindicaciones 1-5 o 7-10, en el que la estructura comprende un sitio de unión a cebador, un sitio de unión a proteína, un sitio de unión a molécula pequeña, o una combinación de los mismos.

12. Procedimiento, según cualquiera de las reivindicaciones 1-5 o 7-11, que comprende analizar la estructura del ácido nucleico en presencia y ausencia de un cebador, una proteína, una molécula pequeña o una combinación de los mismos, para identificar un sitio de unión a cebador, un sitio de unión a proteína, un sitio de unión a molécula pequeña, o una combinación de los mismos.

13. Procedimiento, según cualquiera de las reivindicaciones 9 a 11, en el que cualquier etapa del procedimiento, según cualquiera de las reivindicaciones 9 a 11, se realiza mediante instrucciones ejecutables por ordenador incorporadas en un medio legible por ordenador.

14. Biblioteca de ácidos nucleicos producida mediante el procedimiento, según cualquiera de las reivindicaciones 1-5 o 7-12.

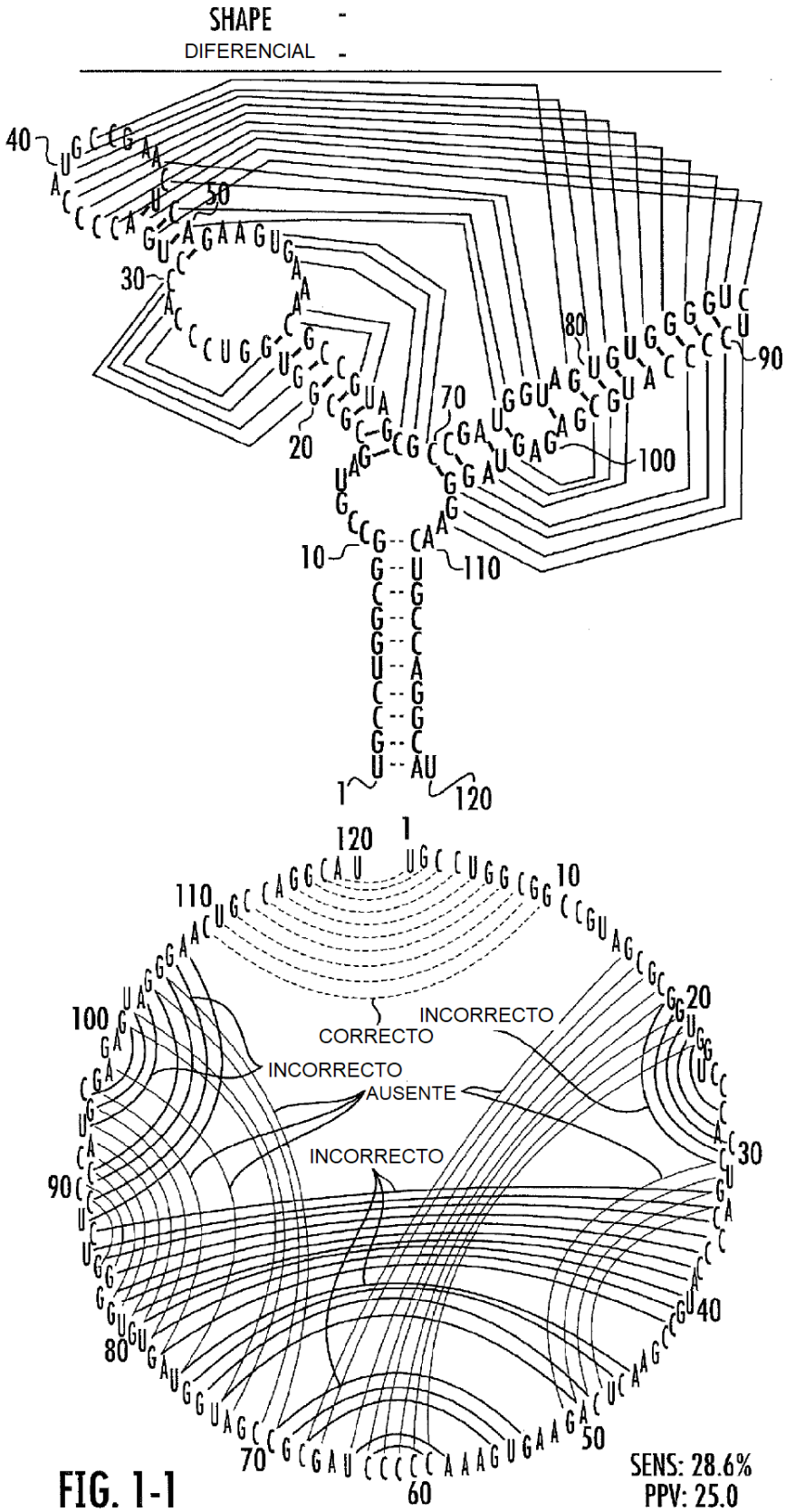

SHAPE
DIFERENCIAL
CORRECTO
INCORRECTO
AUSENTE
INCORRECTO
INCORRECTO
SENS: 28.6%
PPV: 25.0

FIG. 1-1

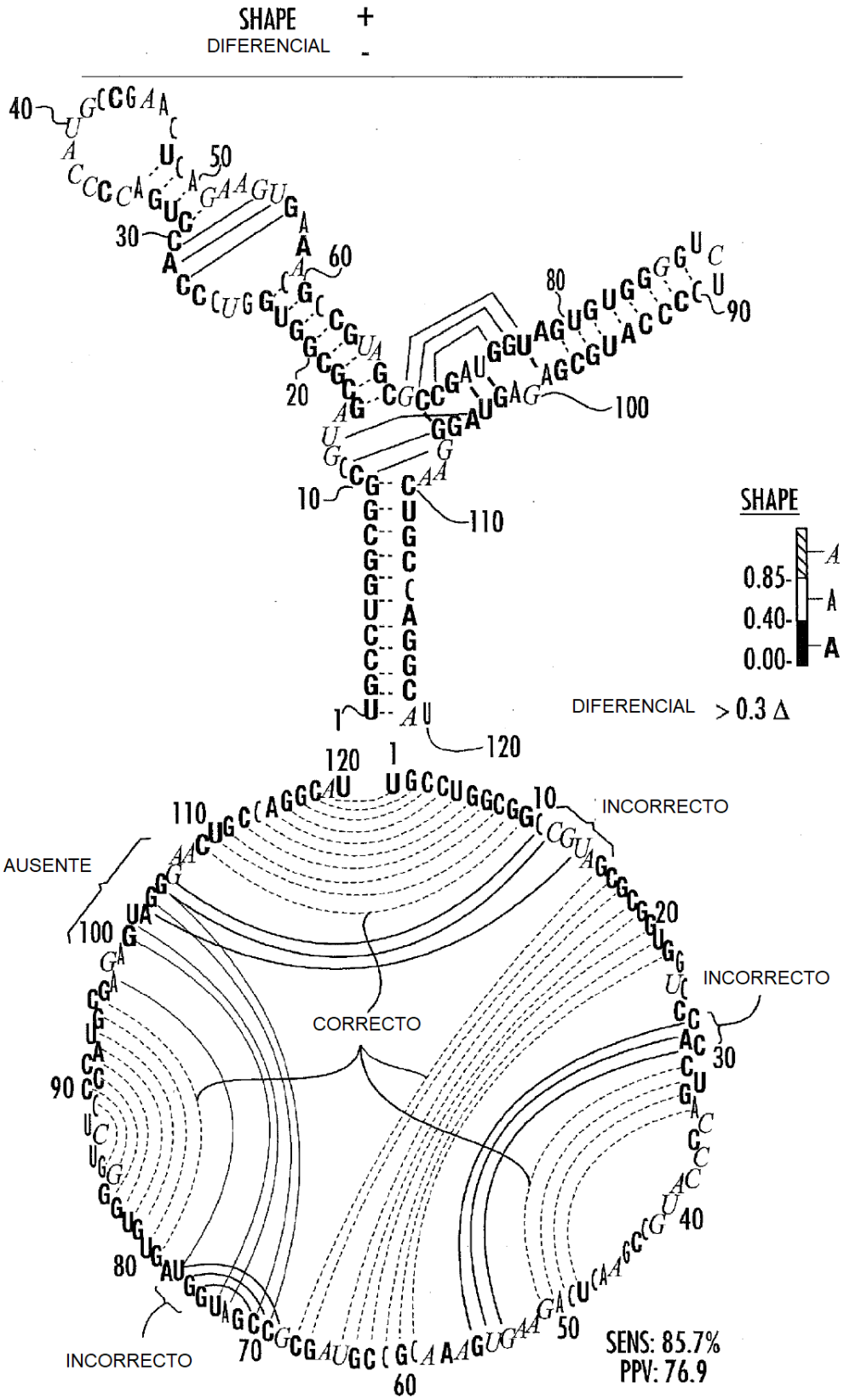

SHAPE
DIFERENCIAL
+
-
SHAPE
0.85-
0.40-
0.00-
A
A
A
DIFERENCIAL > 0.3 Δ
INCORRECTO
AUSENTE
CORRECTO
INCORRECTO
INCORRECTO
SENS: 85.7%
PPV: 76.9

FIG. 1-2

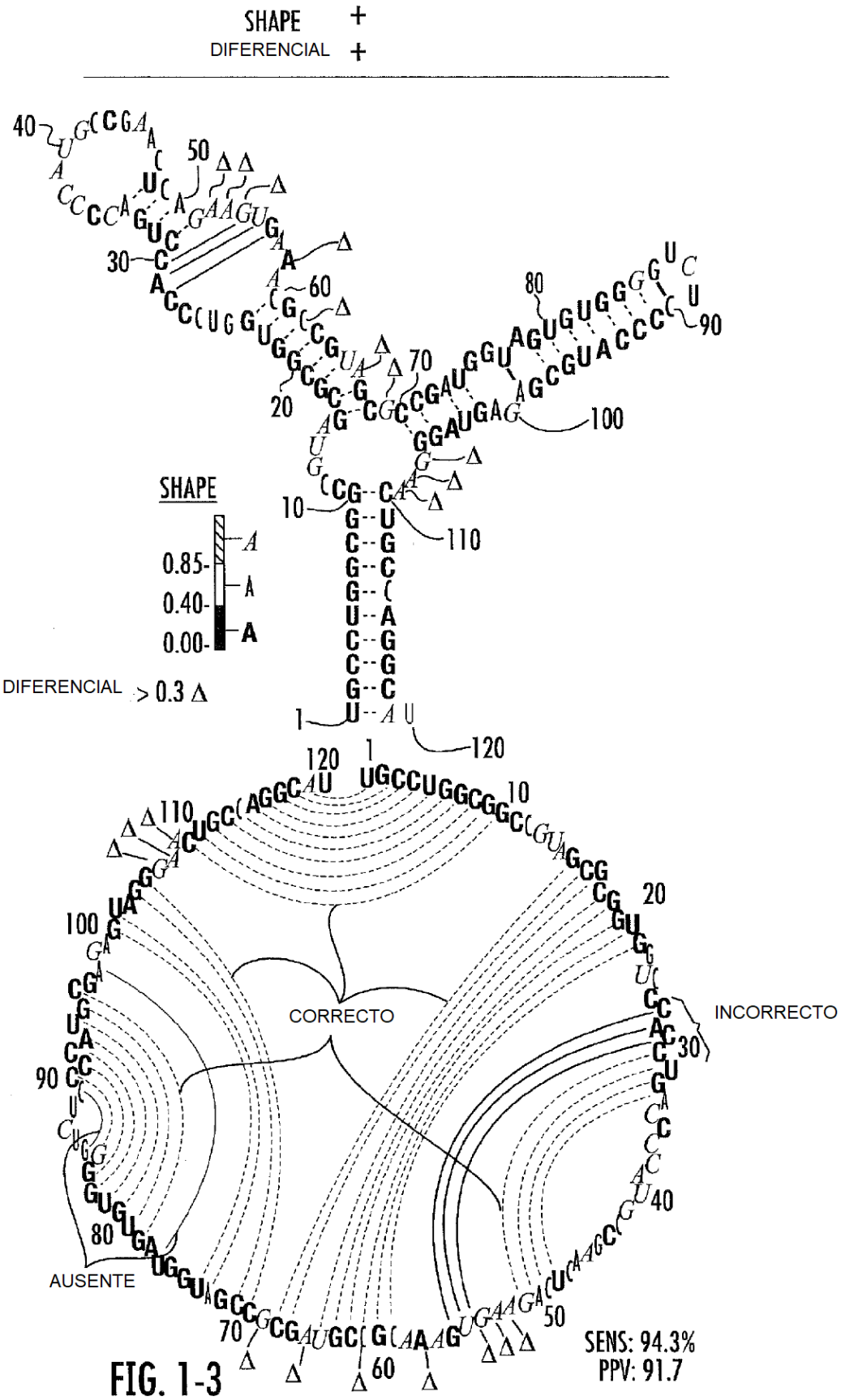

SHAPE +
DIFERENCIAL +
SHAPE
0.85-
0.40-
0.00-
A
A
A
DIFERENCIAL > 0.3 Δ
CORRECTO
INCORRECTO
AUSENTE
SENS: 94.3%
PPV: 91.7

FIG. 1-3

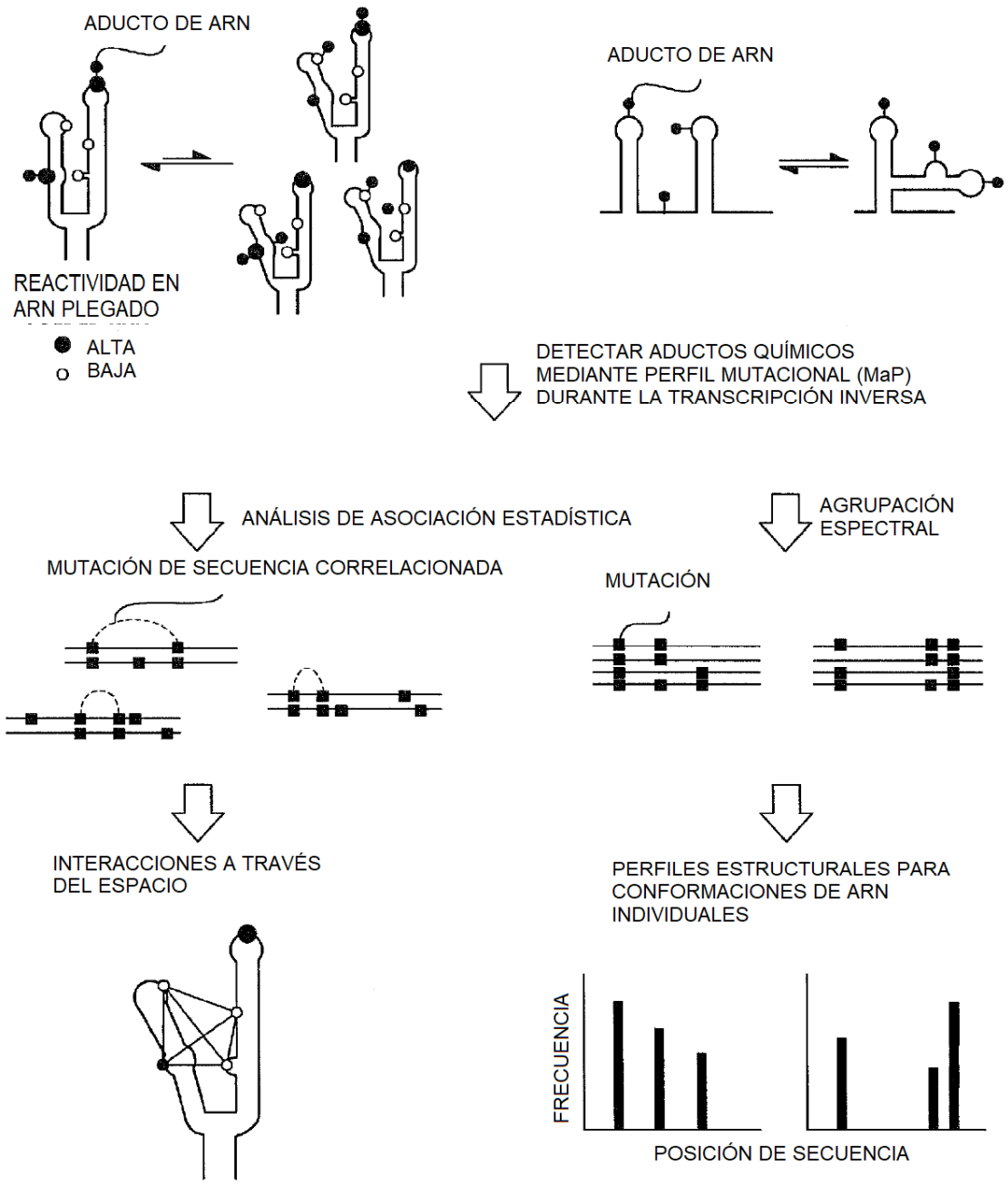

ADUCTO DE ARN
REACTIVIDAD EN ARN PLEGADO
ALTA
BAJA
ADUCTO DE ARN
DETECTAR ADUCTOS QUÍMICOS MEDIANTE PERFIL MUTACIONAL (MaP) DURANTE LA TRANSCRIPCIÓN INVERSA
ANÁLISIS DE ASOCIACIÓN ESTADÍSTICA
MUTACIÓN DE SECUENCIA CORRELACIONADA
INTERACCIONES A TRAVÉS DEL ESPACIO
AGRUPACIÓN ESPECTRAL
MUTACIÓN
PERFILES ESTRUCTURALES PARA CONFORMACIONES DE ARN INDIVIDUALES
FRECUENCIA
POSICIÓN DE SECUENCIA

FIG. 2A

FIG. 2B

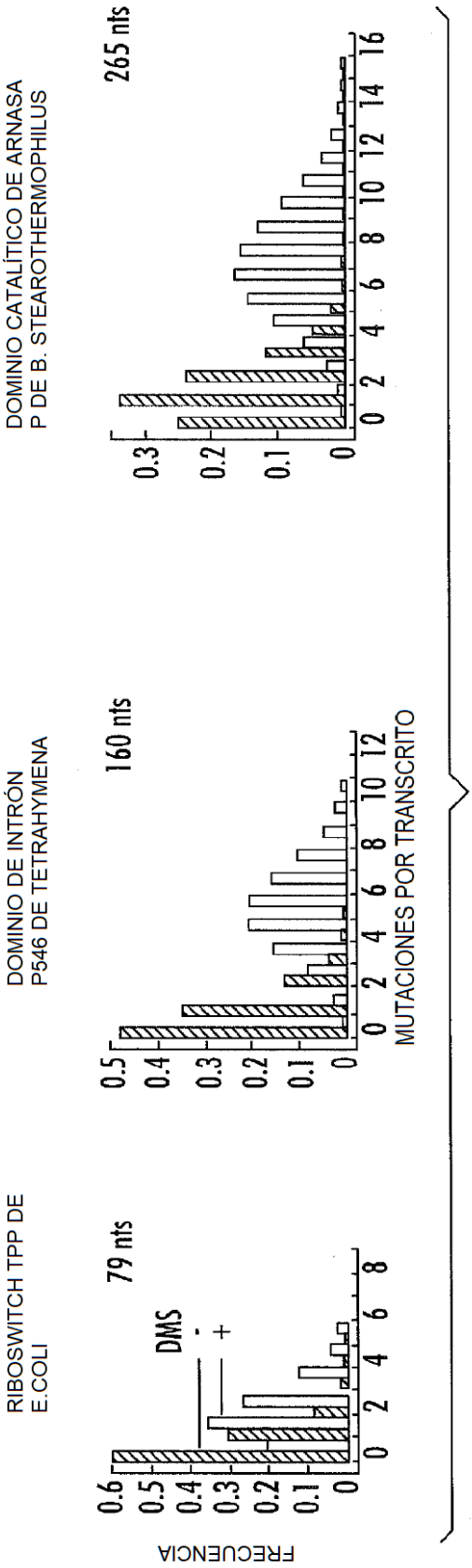

RIBOSWITCH TPP DE E.COLI
79 nts
DMS
-
+
FRECUENCIA
DOMINIO DE INTRÓN P546 DE TETRAHYMENA
160 nts
MUTACIONES POR TRANSCRITO
DOMINIO CATALÍTICO DE ARNASA P DE B. STEAROTHERMOPHILUS
265 nts

FIG. 3A

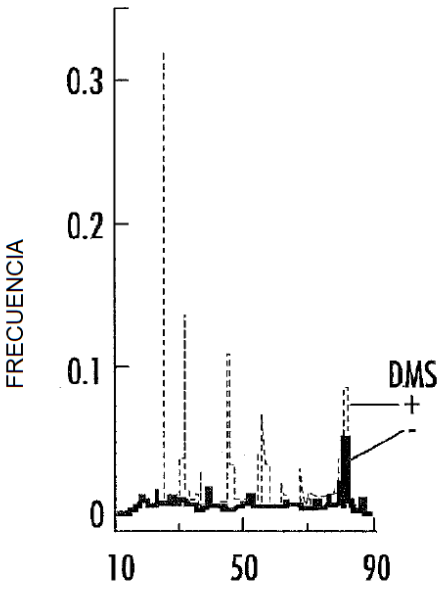


FIG. 3B-1

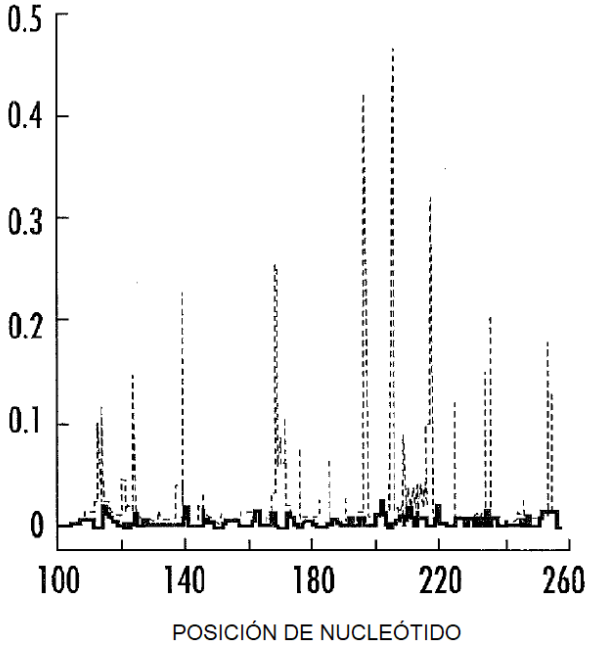


FIG. 3B-2

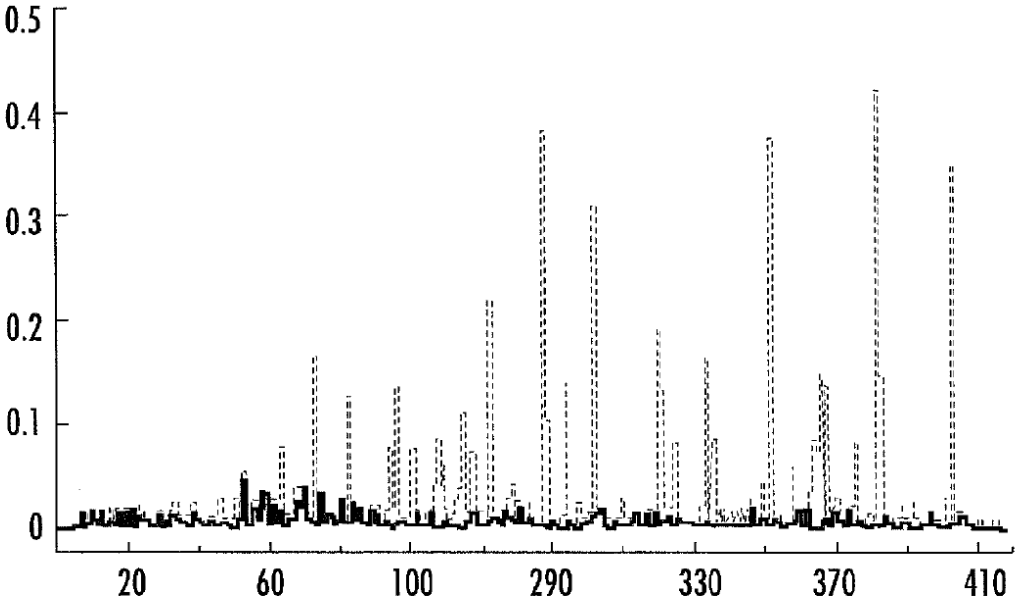


FIG. 3B-3

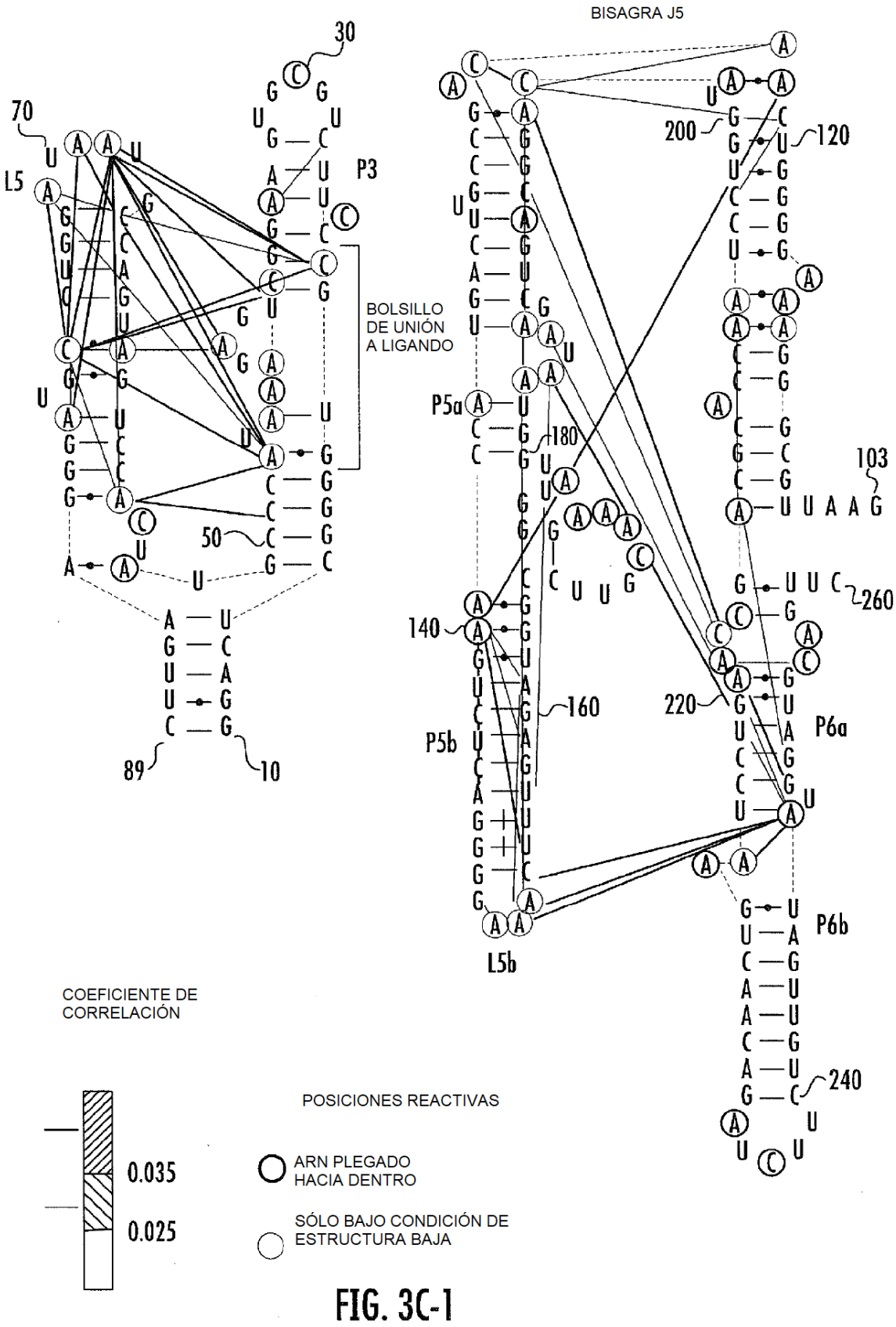

BISAGRA J5
30
70
L5
P3
BOLSILLO
DE UNIÓN
A LIGANDO
50
89
10
P5a
180
140
160
P5b
L5b
200
120
103
260
220
P6a
P6b
240
COEFICIENTE DE
CORRELACIÓN
0.035
0.025
POSICIONES REACTIVAS
ARN PLEGADO
HACIA DENTRO
SÓLO BAJO CONDICIÓN DE
ESTRUCTURA BAJA

FIG. 3C-1

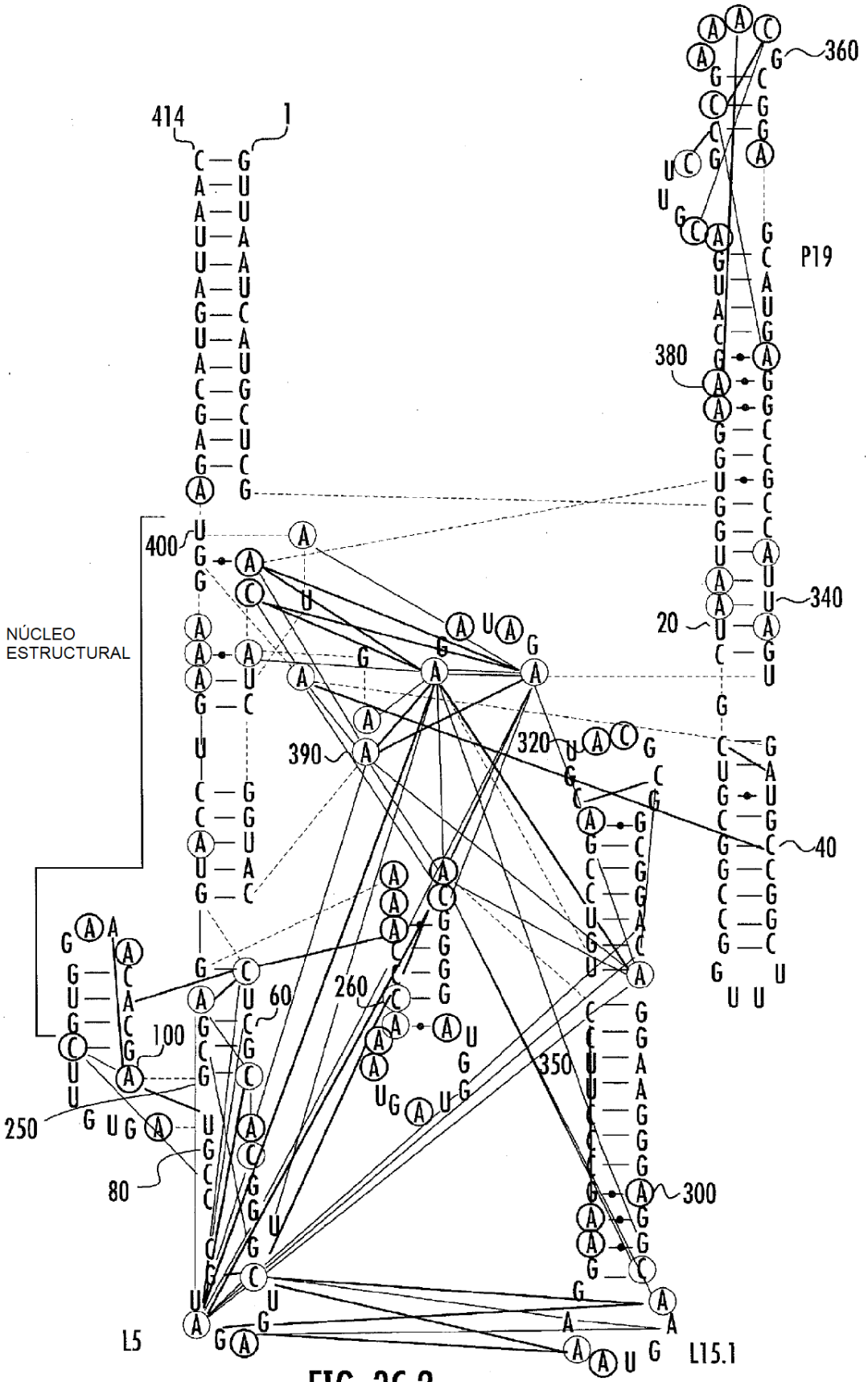
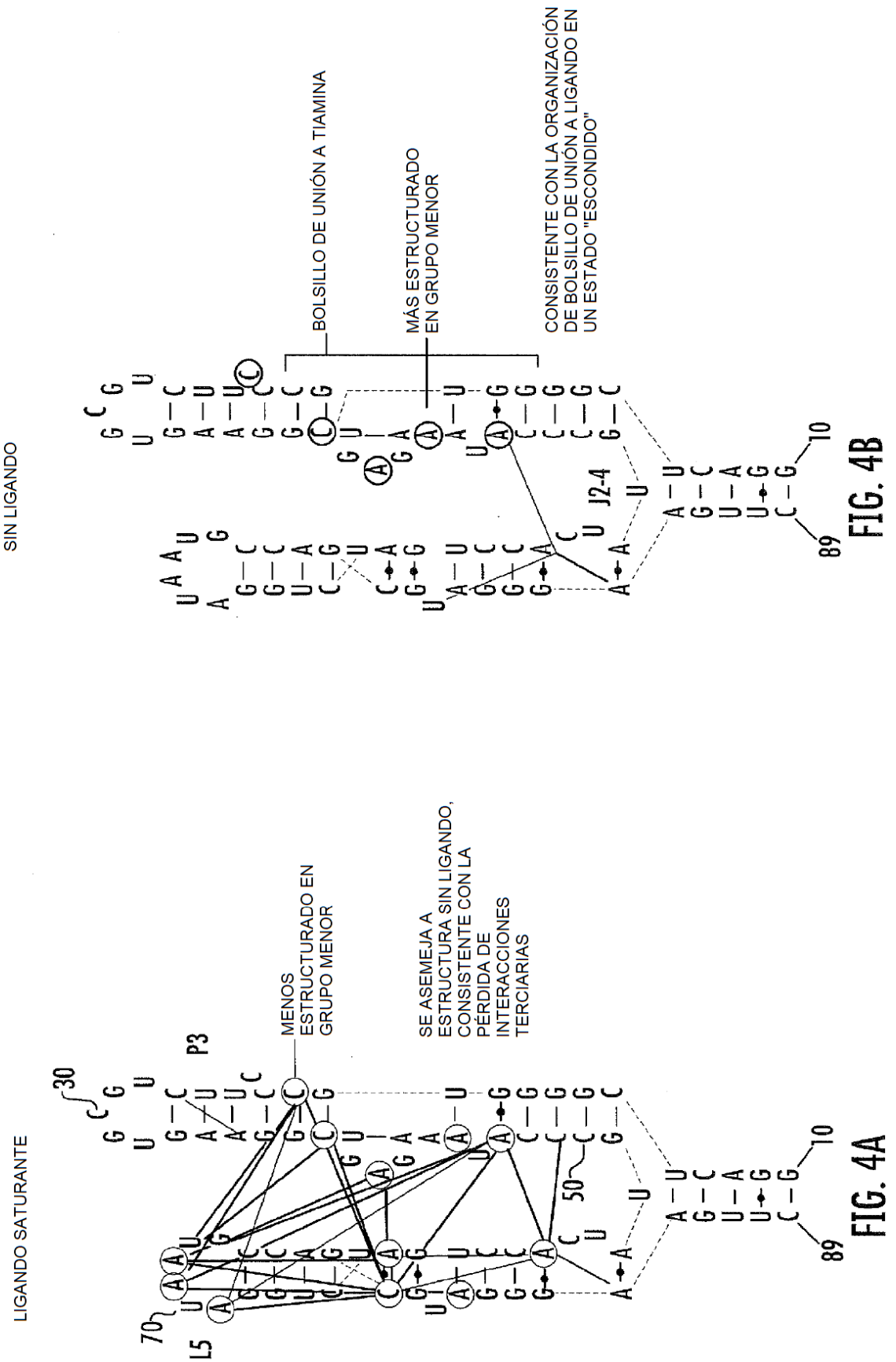

414
360
P19
380
400
NÚCLEO
ESTRUCTURAL
340
20
390
320
40
60
260
100
350
250
80
300
L5
L15.1

FIG. 3C-2


SIN LIGANDO
BOLSILLO DE UNIÓN A TIAMINA
MÁS ESTRUCTURADO EN GRUPO MENOR
CONSISTENTE CON LA ORGANIZACIÓN DE BOLSILLO DE UNIÓN A LIGANDO EN UN ESTADO "ESCONDIDO"
J2-4
89
10
FIG. 4B
LIGANDO SATURANTE
30
P3
MENOS ESTRUCTURADO EN GRUPO MENOR
SE ASEMEJA A ESTRUCTURA SIN LIGANDO, CONSISTENTE CON LA PÉRDIDA DE INTERACCIONES TERCIARIAS
70
L5
50
89
10
FIG. 4A

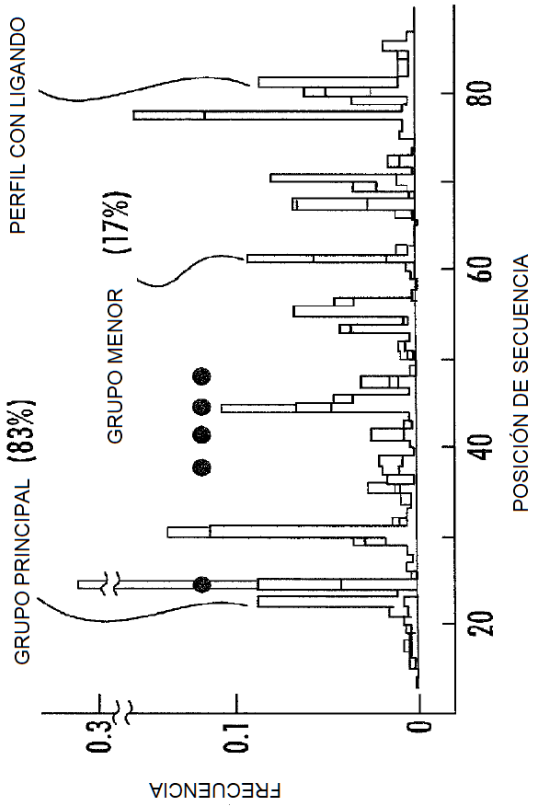

PERFIL CON LIGANDO
GRUPO PRINCIPAL (83%)
GRUPO MENOR (17%)
FRECUENCIA
0.3
0.1
0
20
40
60
80
POSICIÓN DE SECUENCIA

FIG. 4D

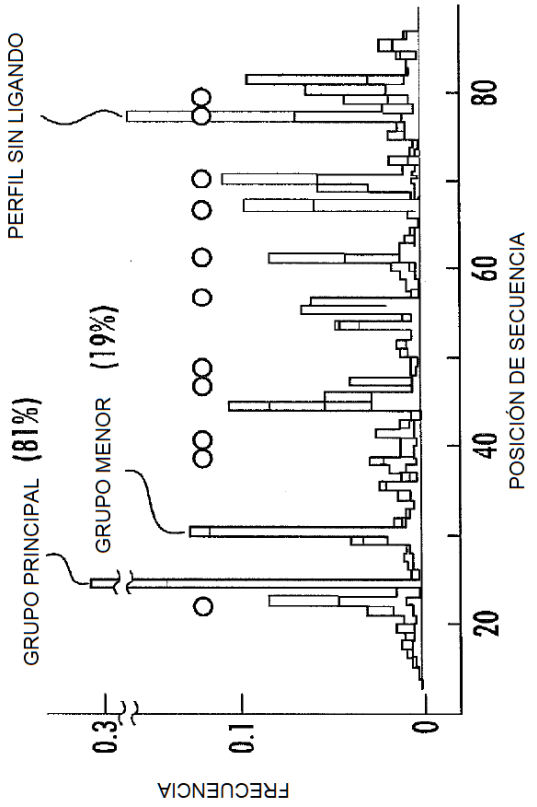

PERFIL SIN LIGANDO
GRUPO PRINCIPAL (81%)
GRUPO MENOR (19%)
FRECUENCIA
0.3
0.1
0
20
40
60
80
POSICIÓN DE SECUENCIA

FIG. 4C

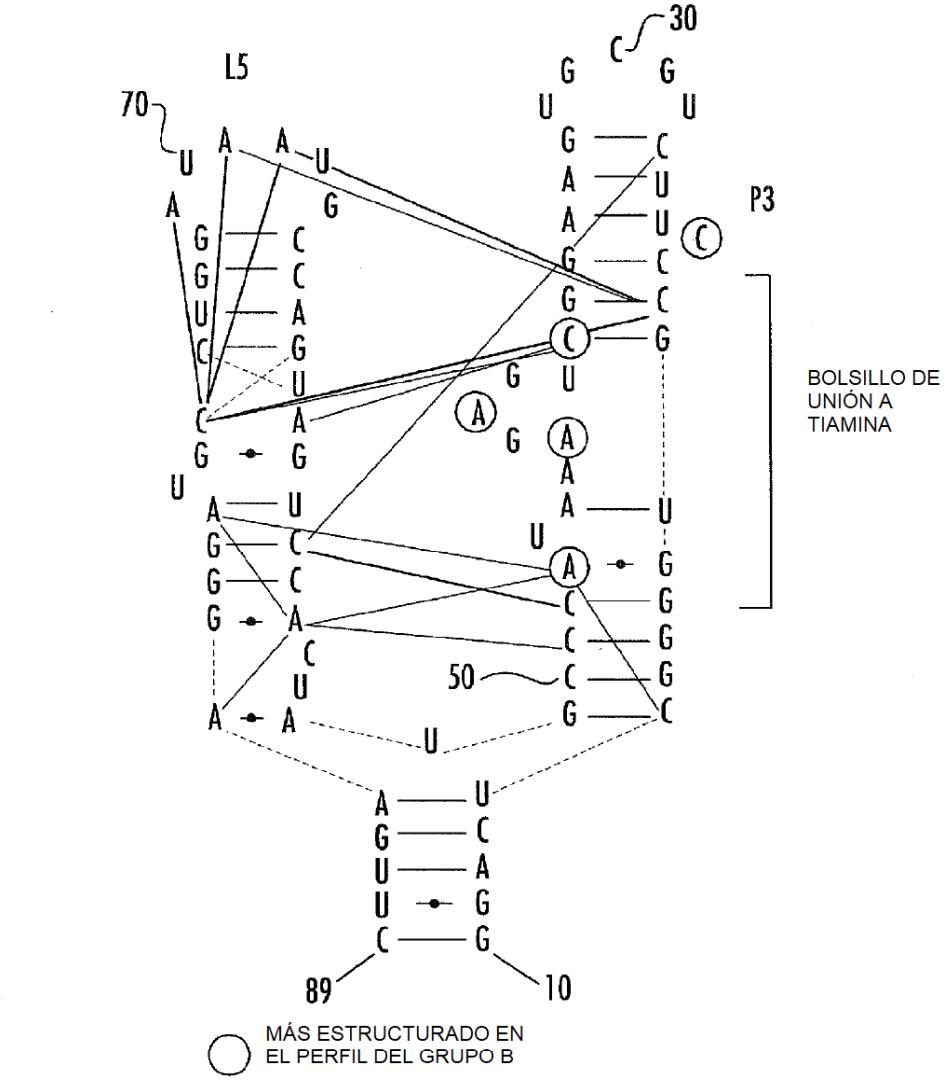
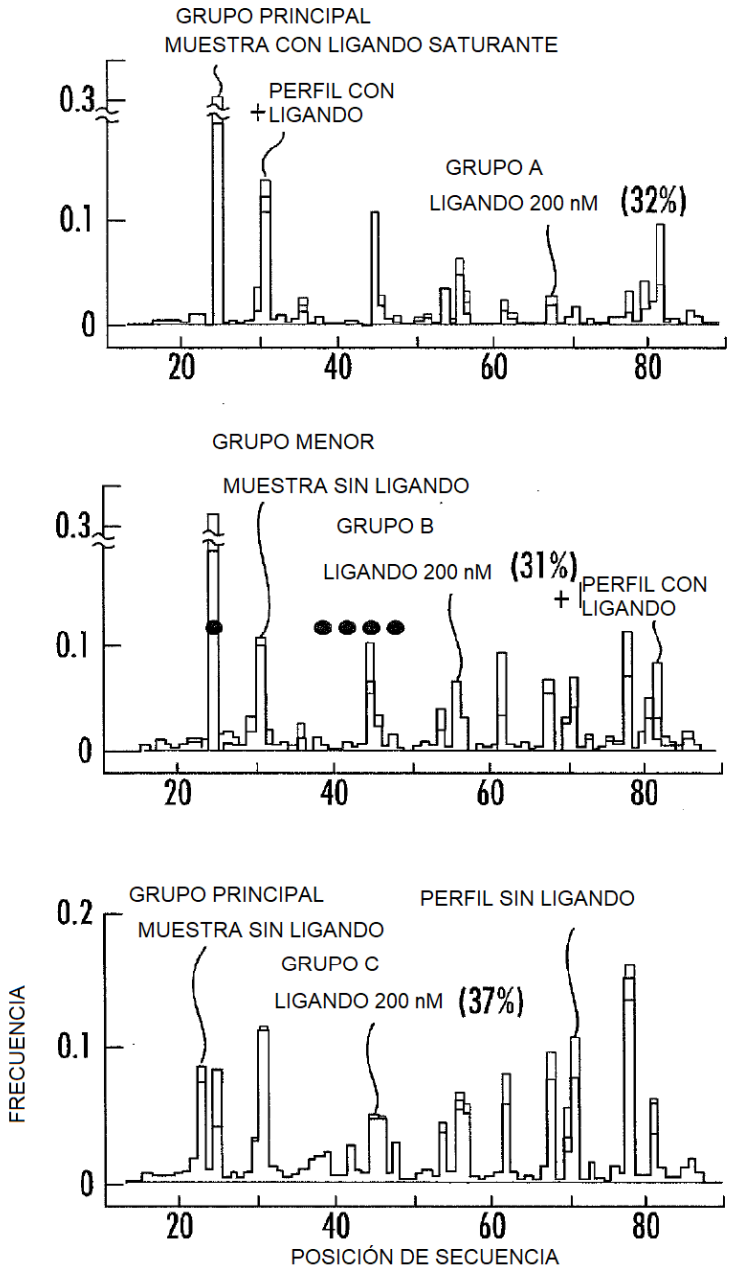

L5
70
30
P3
BOLSILLO DE UNIÓN A TIAMINA
50
89
10
MÁS ESTRUCTURADO EN EL PERFIL DEL GRUPO B
COEFICIENTE DE CORRELACIÓN
0.035
0.025

FIG. 5A



GRUPO A

CORRESPONDE A GRUPO PRINCIPAL EN MUESTRA DE LIGANDO SATURANTE

REFLEJA ARN UNIDO A LIGANDO TOTALMENTE PLEGADO

GRUPO B

CORRESPONDE A GRUPO MENOR EN MUESTRA SIN LIGANDO

REFLEJA ARN PLEGADO PARCIALMENTE Y BOLSILLO DE UNIÓN A LIGANDO ORGANIZADO

GRUPO C

CORRESPONDE A GRUPO PRINCIPAL EN MUESTRA SIN LIGANDO

REFLEJA CONFORMACIÓN NO ESTRUCTURADA CON ALGUNAS ASOCIACIONES ENTRE NUCLEÓTIDOS

FIG. 5B

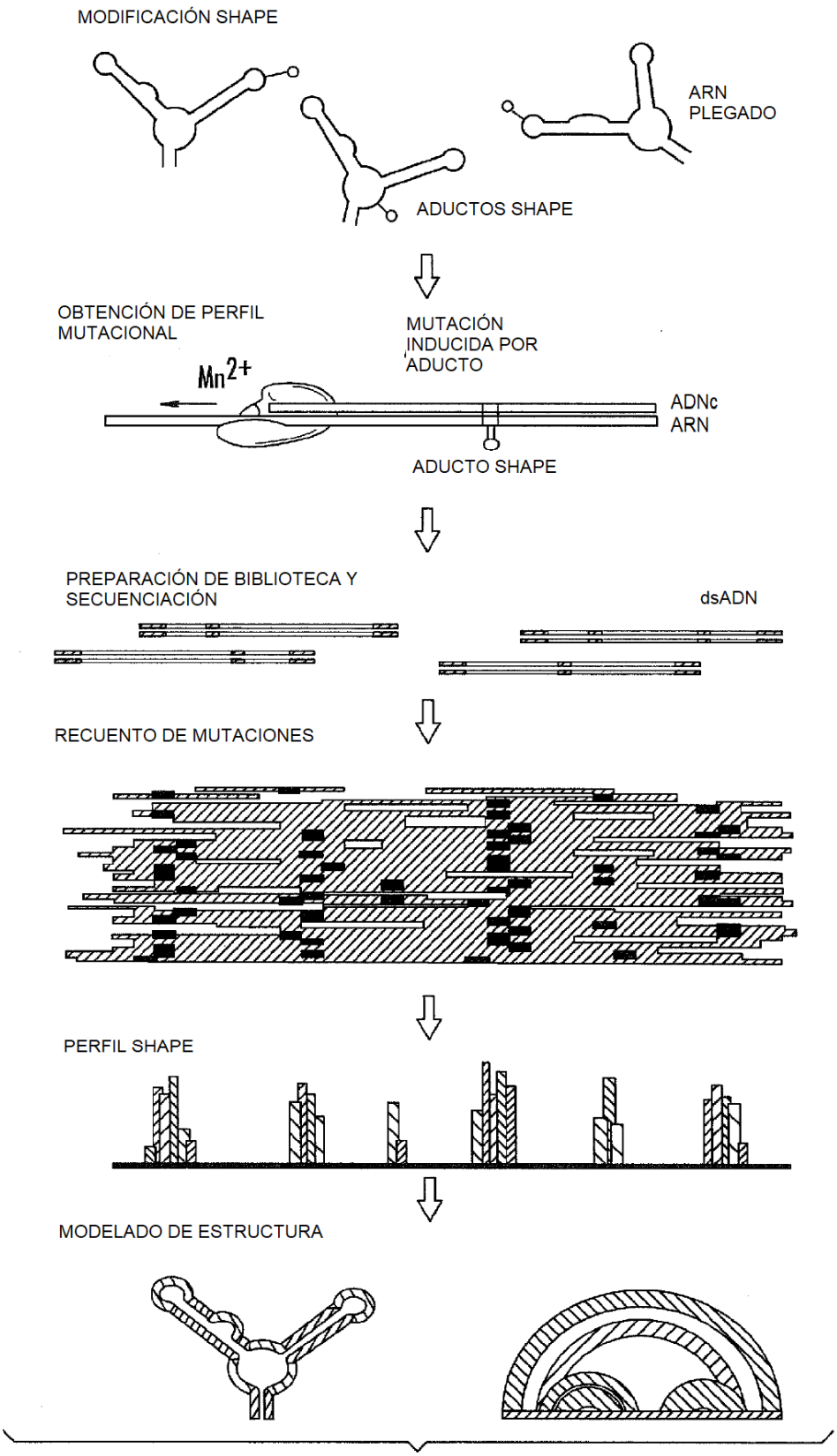

MODIFICACIÓN SHAPE
ARN
PLEGADO
ADUCTOS SHAPE
OBTENCIÓN DE PERFIL
MUTACIONAL
MUTACIÓN
INDUCIDA POR
ADUCTO
Mn2+
ADNc
ARN
ADUCTO SHAPE
PREPARACIÓN DE BIBLIOTECA Y
SECUENCIACIÓN
dsADN
RECUENTO DE MUTACIONES
PERFIL SHAPE
MODELADO DE ESTRUCTURA

FIG. 6

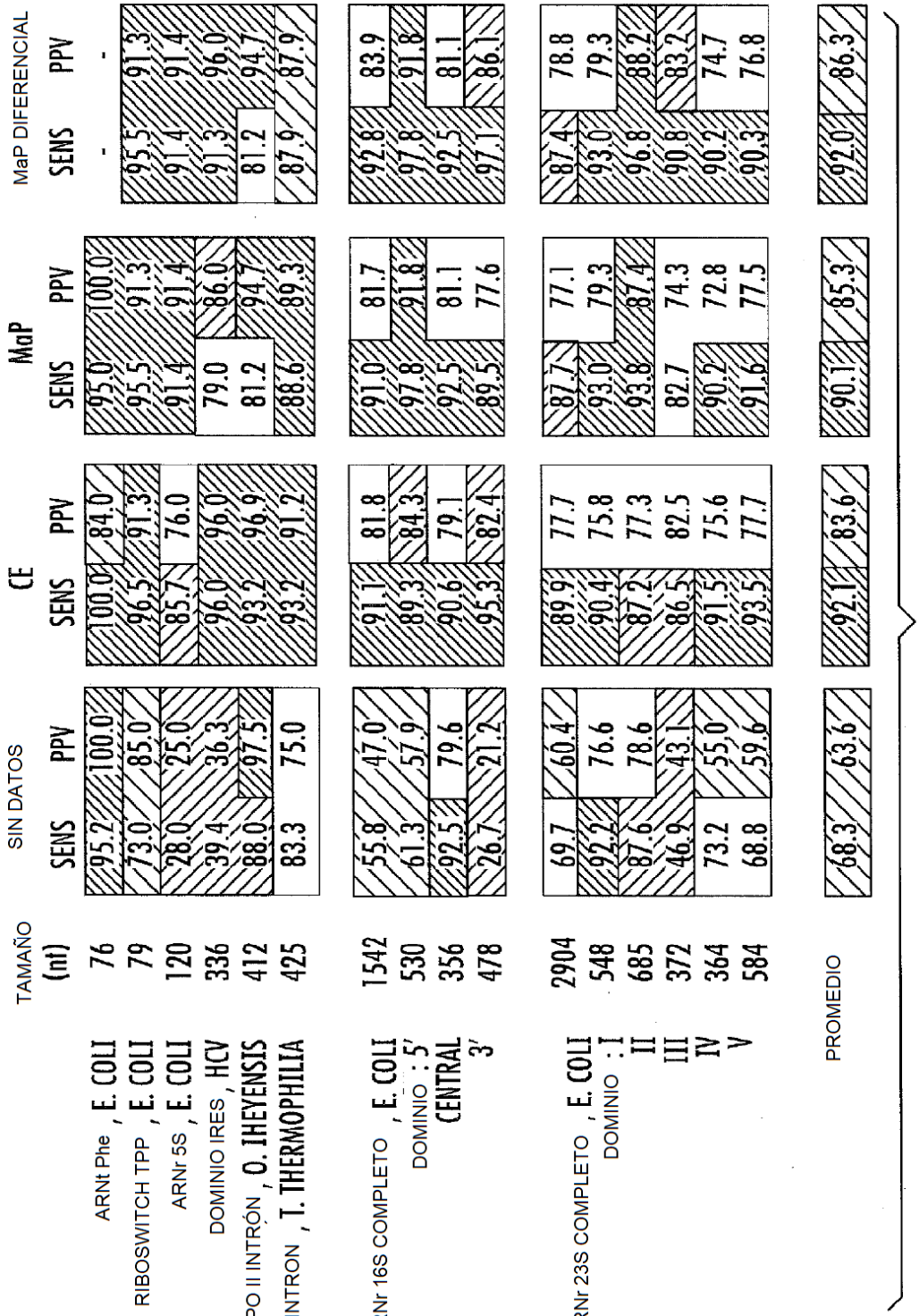

TAMAÑO (nt)
SIN DATOS
CE
MaP
MaP DIFERENCIAL
SENS
PPV
ARNt Phe , E. COLI 76 95.2 100.0 100.0 84.0 95.0 100.0 - -
RIBOSWITCH TPP , E. COLI 79 73.0 85.0 96.5 91.3 95.5 91.3 95.5 91.3
ARNr 5S , E. COLI 120 28.0 25.0 85.7 76.0 91.4 91.4 91.4 91.4
DOMINIO IRES , HCV 336 39.4 36.3 96.0 96.0 79.0 86.0 91.3 96.0
I GRUPO II INTRÓN , O. IHEYENSIS 412 88.0 97.5 93.2 96.9 81.2 94.7 81.2 94.7
GRUPO I INTRON , T. THERMOPHILIA 425 83.3 75.0 93.2 91.2 88.6 89.3 87.9 87.9
ARNr 16S COMPLETO , E. COLI 1542 55.8 47.0 91.1 81.8 91.0 81.7 92.8 83.9
DOMINIO : 5' 530 61.3 57.9 89.3 84.3 97.8 91.8 97.8 91.8
CENTRAL 356 92.5 79.6 90.6 79.1 92.5 81.1 92.5 81.1
3' 478 26.7 21.2 95.3 82.4 89.5 77.6 97.1 86.1
ARNr 23S COMPLETO , E. COLI 2904 69.7 60.4 89.9 77.7 87.7 77.1 87.4 78.8
DOMINIO : I 548 92.2 76.6 90.4 75.8 93.0 79.3 93.0 79.3
II 685 87.6 78.6 87.2 77.3 93.8 87.4 96.8 88.2
III 372 46.9 43.1 86.5 82.5 82.7 74.3 90.8 83.2
IV 364 73.2 55.0 91.5 75.6 90.2 72.8 90.2 74.7
V 584 68.8 59.6 93.5 77.7 91.6 77.5 90.3 76.8
PROMEDIO 68.3 63.6 92.1 83.6 90.1 85.3 92.0 86.3

FIG. 7A

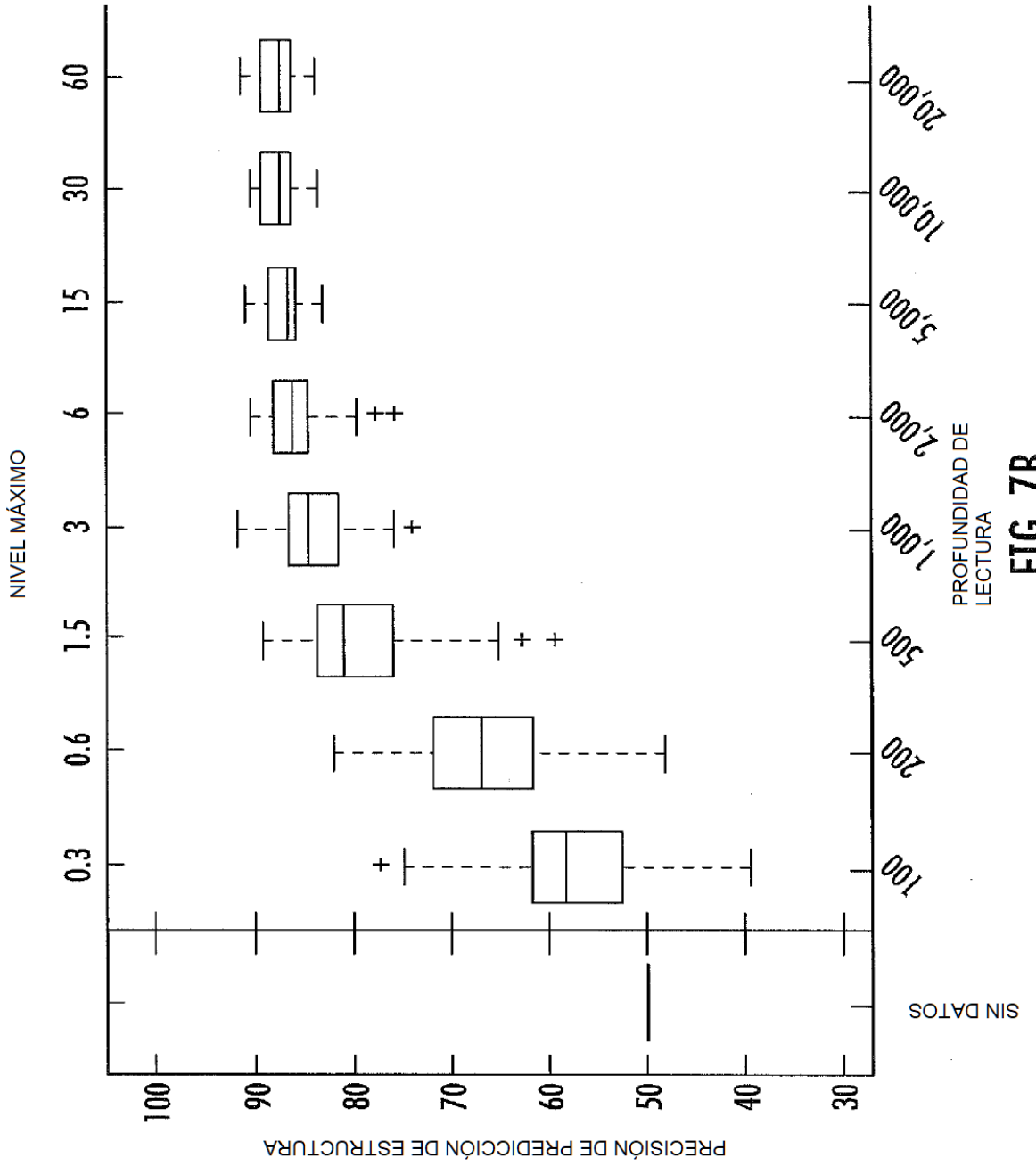

NIVEL MÁXIMO
0.3
0.6
1.5
3
6
15
30
60
PROFUNDIDAD DE LECTURA
100
200
500
1,000
2,000
5,000
10,000
20,000
SIN DATOS
PRECISIÓN DE PREDICCIÓN DE ESTRUCTURA
100
90
80
70
60
50
40
30

FIG. 7B

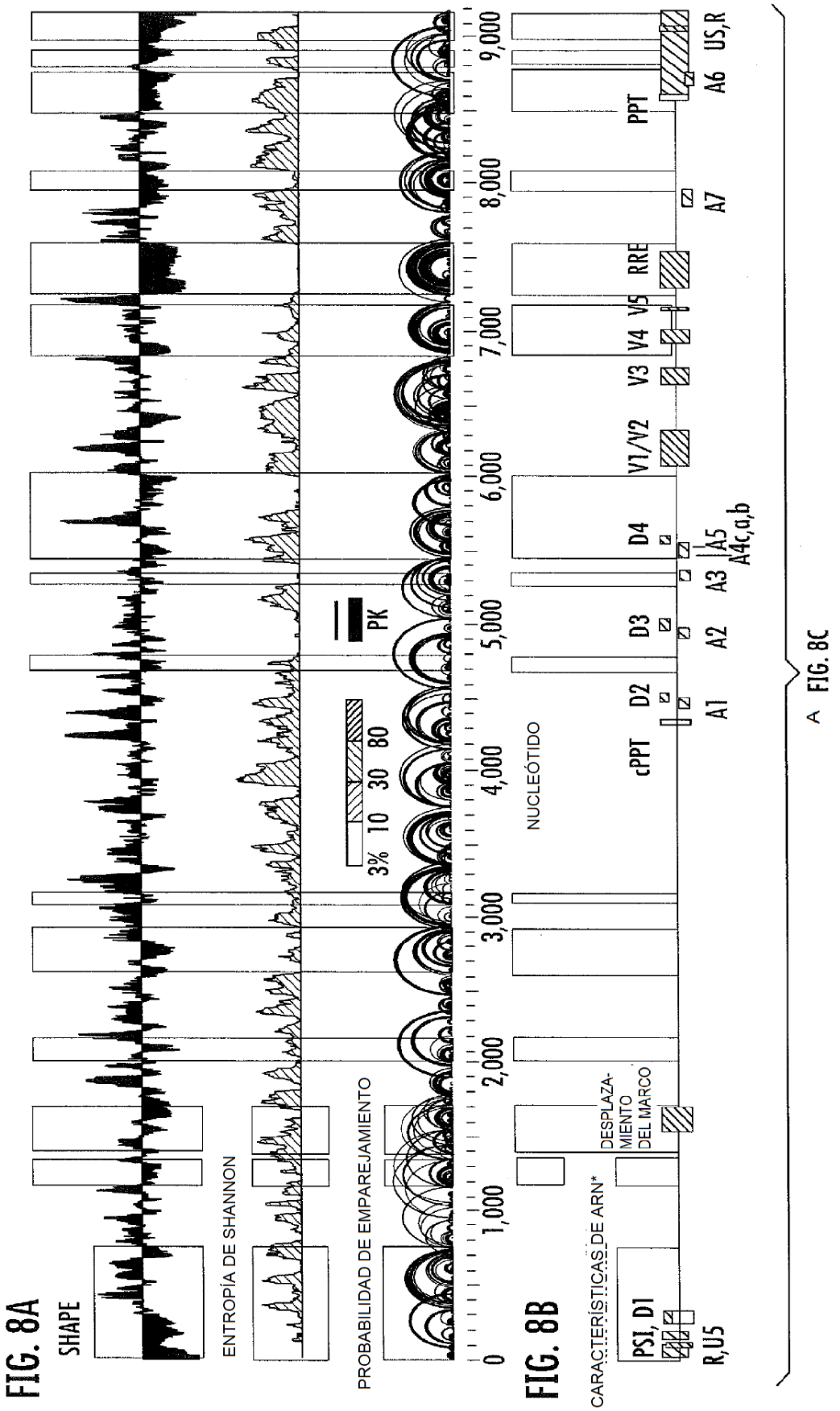


FIG. 8A
SHAPE
ENTROPÍA DE SHANNON
PROBABILIDAD DE EMPAREJAMIENTO
3% 10 30 80
PK
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
NUCLEÓTIDO
FIG. 8B
CARACTERÍSTICAS DE ARN*
PSI, D1
R,U5
DESPLAZA-
MIENTO
DEL MARCO
cPPT
D2
A1
D3
A2
A3
D4
A4c,a,b
A5
V1/V2
V3
V4
V5
RRE
A7
PPT
A6
US,R
A FIG. 8C

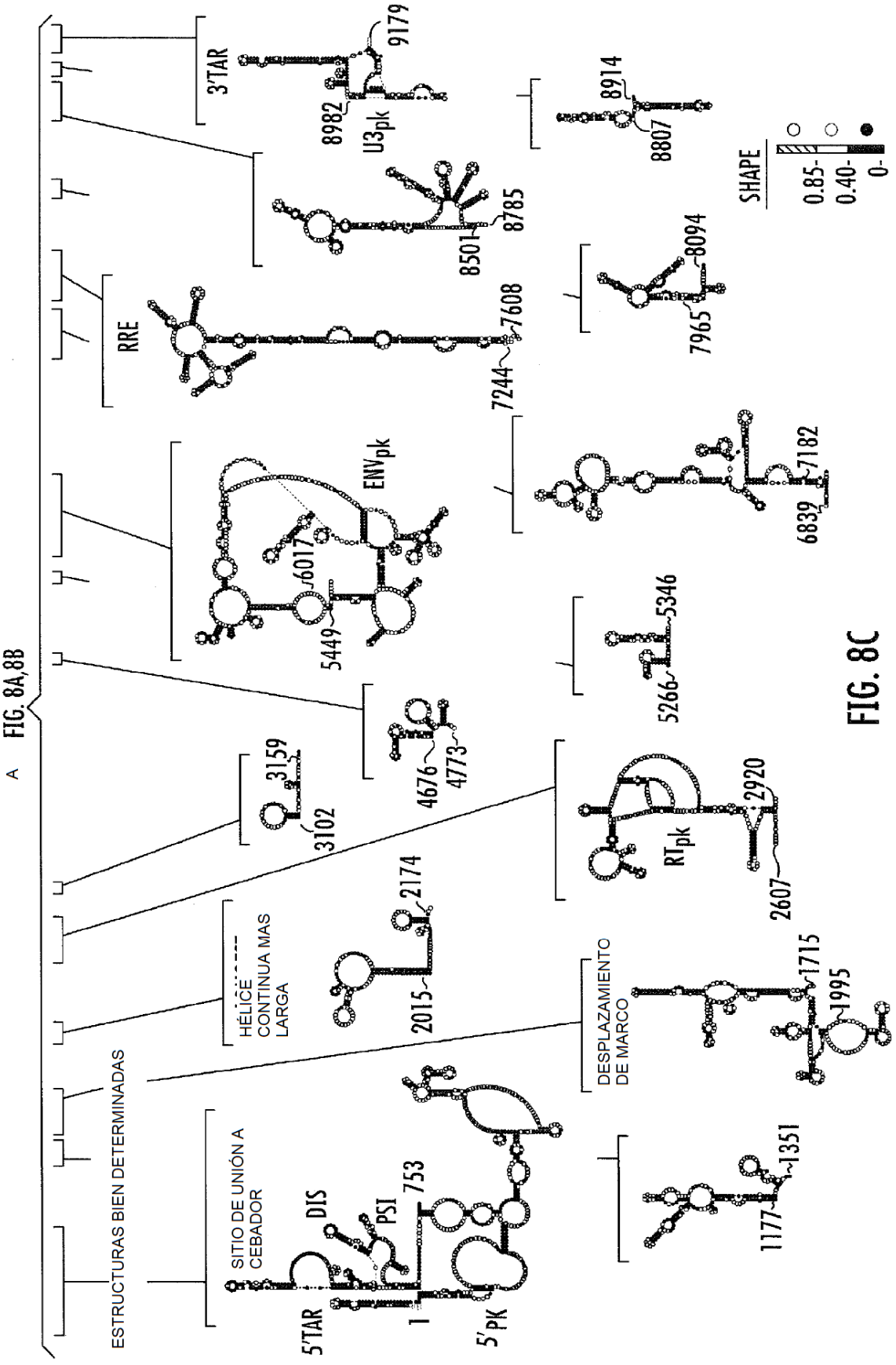

A FIG. 8A,8B
ESTRUCTURAS BIEN DETERMINADAS
SITIO DE UNIÓN A CEBADOR
HÉLICE CONTINUA MAS LARGA
DESPLAZAMIENTO DE MARCO
5'TAR
DIS
PSI
753
1
5'PK
2015
2174
3102
3159
4676
4773
5449
6017
ENVpk
RRE
7244
7608
8501
8785
8982
U3pk
9179
3'TAR
1177
1351
1715
1995
2607
RTpk
2920
5266
5346
6839
7182
7965
8094
8807
8914
SHAPE
0.85-
0.40-
0-

FIG. 8C

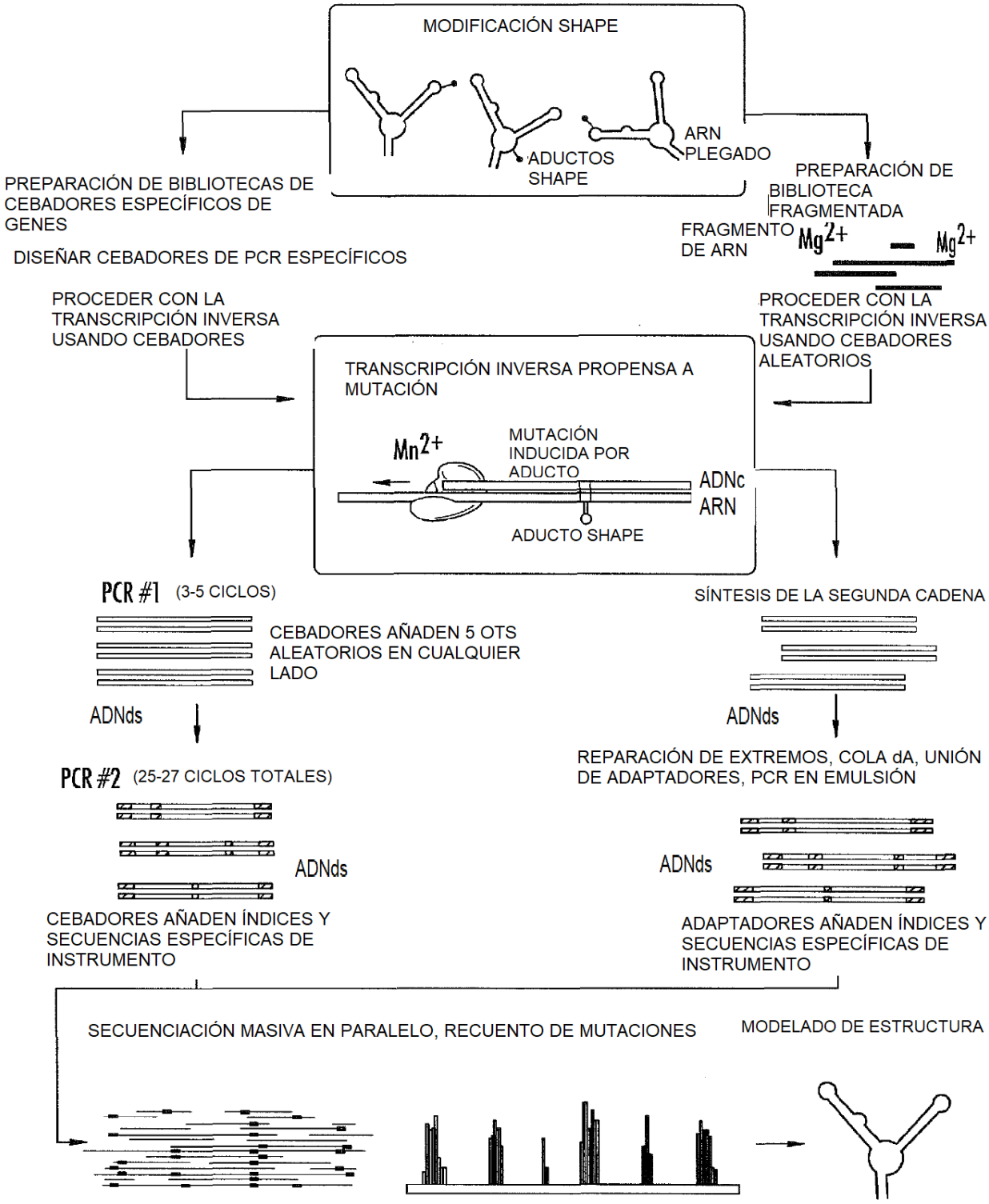

MODIFICACIÓN SHAPE
ARN PLEGADO
ADUCTOS SHAPE
PREPARACIÓN DE BIBLIOTECAS DE CEBADORES ESPECÍFICOS DE GENES
DISEÑAR CEBADORES DE PCR ESPECÍFICOS
PREPARACIÓN DE BIBLIOTECA FRAGMENTADA
FRAGMENTO DE ARN
Mg2+
Mg2+
PROCEDER CON LA TRANSCRIPCIÓN INVERSA USANDO CEBADORES
PROCEDER CON LA TRANSCRIPCIÓN INVERSA USANDO CEBADORES ALEATORIOS
TRANSCRIPCIÓN INVERSA PROPENSA A MUTACIÓN
Mn2+
MUTACIÓN INDUCIDA POR ADUCTO
ADNc
ARN
ADUCTO SHAPE
PCR #1 (3-5 CICLOS)
CEBADORES AÑADEN 5 OTS ALEATORIOS EN CUALQUIER LADO
ADNds
SÍNTESIS DE LA SEGUNDA CADENA
ADNds
PCR #2 (25-27 CICLOS TOTALES)
ADNds
REPARACIÓN DE EXTREMOS, COLA dA, UNIÓN DE ADAPTADORES, PCR EN EMULSIÓN
ADNds
CEBADORES AÑADEN ÍNDICES Y SECUENCIAS ESPECÍFICAS DE INSTRUMENTO
ADAPTADORES AÑADEN ÍNDICES Y SECUENCIAS ESPECÍFICAS DE INSTRUMENTO
SECUENCIACIÓN MASIVA EN PARALELO, RECUENTO DE MUTACIONES
MODELADO DE ESTRUCTURA

FIG. 9

DISEÑO DE CEBADORES PARA SHAPE-MaP CEBADO AL AZAR

9 UNIDADES

$5'\text{-}N_1N_2N_3N_4N_5N_6N_7N_8N_9\text{-}3'$

LNA+

$5'\text{-}L_1X_2L_3X_4L_5X_6X_7X_8X_9X_{10}\text{-}3'$

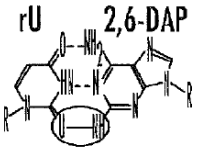
L_1, L_3, L_5 = A, G O T (LNA) BLOQUEADO

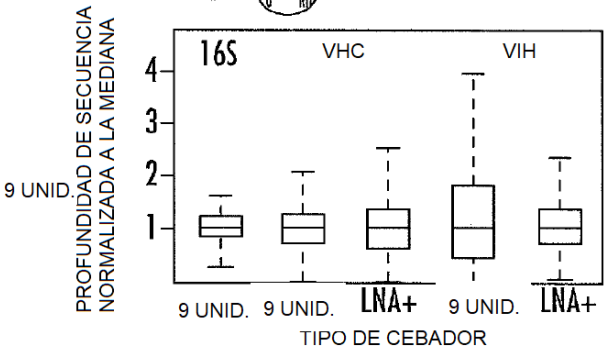
$X_2, X_4, X_{6\text{-}10}$ = 2,6-DIAMINOPURINA, dG O dT

CEBADORES DE 9 UNIDADES SON NONÁMEROS ALEATORIOS ESTÁNDAR

LOS CEBADORES LNA+ SE DISEÑAN PARA REDUCIR LA PROPAGACIÓN DE LA PROFUNDIDAD DE SECUENCIACIÓN PARA ARN CON REGIONES DE BAJO CONTENIDO DE GC

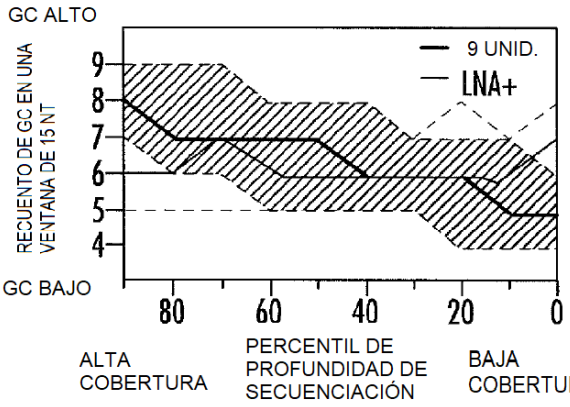
- LA CITOSINA SE OMITE PARA DESFAVORECER LA UNIÓN A GUANOSINA
- SE INCLUYE 2,6-DAP PARA FAVORECER LA UNIÓN A URIDINA
- LNA DE A, G Y T FAVORECEN LA UNIÓN A URIDINA, GUANOSINA Y ADENINA



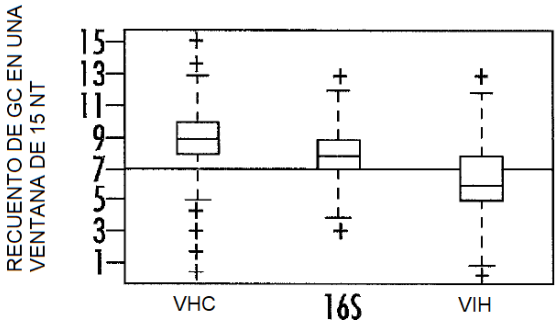


LA PROPAGACIÓN DE LA PROFUNDIDAD DE SECUENCIACIÓN ES:

- BAJA SOBRE ARN RIBOOMAL 16S USANDO CEBADORES DE 9 UNIDADES
- BAJA SOBRE ARN GENÓMICO DE VHC USANDO CEBADORES DE 9 UNIDADES Y NO SE MEJORA USANDO CEBADORES LNA+
- ALTA SOBRE ARN GENÓMICO DE VIH-1 USANDO CEBADORES DE 9 UNIDADES, PERO SE MEJORA NOTABLEMENTE USANDO CEBADORES LNA+



LOS CEBADORES LAN+ REDUJERON SIGNIFICATIVAMENTE EL SESGO DE LA PROFUNDIDAD DE SECUENCIACIÓN FRENTE A REGIONES BAJAS EN GC EN ARN DE VIH-1. SE MUESTRA LA MEDIANA DEL RECUENTO DE GC COMO LÍNEAS CONTINUAS Y LOS INTERVALOS INTERCUARTILES COMO LÍNEAS DISCONTINUAS



REGLA: UTILIZAR CEBADORES LNA+ PARA ARN CON RECUENTOS DE GC DE MEDIANA EN UNA VENTANA DE 15 NT POR DEBAJO DE 7

FIG. 10