

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS

IMPLICACIONES EVOLUTIVAS DE LA PLASTICIDAD DEL RNA

T E S I S
QUE PARA OBTENER EL GRADO DE BIOLOGA
PRESENTA:
Natalia Quiñones Olvera

DIRECTOR DE TESIS:
Dr. Pedro Eduardo Miramontes Vidal



CIUDAD UNIVERSITARIA, CDMX
2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

HOJA DE DATOS DEL JURADO

1. Datos del alumno

Quiñones
Olvera
Natalia
(55) 255 99658
Universidad Nacional Autónoma de México
Facultad de Ciencias
Biología
411017334

2. Datos del tutor

Dr.
Pedro Eduardo
Miramontes
Vidal

3. Datos del sinodal 1

Dr.
Víctor Manuel
Valdés
López

4. Datos del sinodal 2

Dra.
María Elena
Álvarez-Buylla
Roces

5. Datos del sinodal 3

Dr.
Arturo Carlos II
Becerra
Bracho

6. Datos del sinodal 4

M. en C.
Alfonso José
Vilchis
Peluyera

7. Datos del trabajo escrito

Implicaciones evolutivas de la plasticidad del RNA
130 p.
2016

When I heard the learn'd astronomer,
When the proofs, the figures, were ranged in columns before me;
When I was shown the charts and the diagrams, to add, divide, and measure them;
When I, sitting, heard the astronomer, where he lectured with much applause in the lecture-room,
How soon, unaccountable, I became tired and sick;
Till rising and gliding out, I wander'd off by myself,
In the mystical moist night-air, and from time to time,
Look'd up in perfect silence at the stars.

WALT WHITMAN

AGRADECIMIENTOS

Porque creo que la tesis de licenciatura conserva aún cierta inocencia que permite agradecimientos de extensiones ridículas (y afortunadamente más personales que institucionales) me permito desbordarme en las siguientes páginas:

A mi madre, por educarnos con tanta libertad (de actividad, de opinión, de intereses, de culto) mas siempre acentuando la importancia de la disciplina y la responsabilidad. Gracias por siempre ser muestra de que *debe haber otro modo de ser humano y libre*.

A mi hermano Fernando, por ser siempre mi cómplice en las ideas, en la crítica contra las creencias y en los juegos. Por ser el más escéptico y siempre cuestionar mis ideas y motivos tan seguros. Por ayudarme a bajar de la nube de la pretenciosidad cuando lo necesito.

A los Quiñones, por hacer del debate una actividad familiar, por las guerras del diccionario, por la alegría, por el secreto de amor. Por quererme con todo y lo rara, atea e izquierdosa.

A mis viejos amigos, gracias por todos los momentos vividos: a Carla Vigueras por su cariño, por las aventuras, por las largas pláticas antes de dormir. A Mariana Flores por buscar la libertad, por sus preguntas y reflexiones, por los sueños, por la emoción. A Salma Patarroyo por su apoyo, por rescatarme cuando estaba perdiendo la razón. A Deborah Moloeznik por compartir poesía, ideas, risas y tardes. A Adrián Vázquez por ser maestro y amigo, por mostrarnos la belleza de la ciencia, porque gracias a sus clases acabé en Biología en la UNAM. A los Eleonoros por ser siempre familia y casa, por ser cronopios.

A los infinitos descontados: a Rafa Chávez por los cables subterráneos, por entenderme y escucharme, por estar tan descosido como yo, por la poesía, por los cuentos, por los campos de pájaros negros, por todo su amor y complicidad. A Diana Heredia por su sensibilidad, por quererme con mis manías y necesidad, por acompañarme en todo, por las tardes de té y tesis, por sus historias, por tener las ideas más interesantes y por muchas (muchas) otras cosas.

A los amigos que hicieron de la carrera un experiencia inolvidable: a mi roomie, Andrés Ferriño, por sus preguntas hipotéticas, por las mejores sobremesas sobre ciencia y filosofía, por la vida compartida, por sus ideas, por su familia que me hace sentir en casa. A Pancho García Bullé por entenderme, por los recuentos de daños, por su espíritu libre. A Ana Bermúdez por hablar de todo tema, por su sentido del humor, por las excelentes historias, por su amistad genuina. A Gaby Santos por llevarme con Pedro, por llevarnos a Chile y por ser compañera en este curioso camino de la biología teórica. A más personas que conocí en esta hermosísima ciudad hecha de locos maravillosos, en especial a: Santiago Gracia, Kevin Muñoz, Natalia Delgado, Eric Novelo, Mauricio Betancourt, Adriana Miranda, Begoña Iñárritu, Alonso Sandoval, Claudia Guillot, Cecilia González. A Eric M. Tomasini-Bassols por los días cortazarianos.

A todos los docentes que han sido parte de mi formación: nunca tengo suficientes palabras para agradecerles ni para expresar lo mucho que admiro su trabajo. En especial los profesores de la UNAM: a la Dra. Martha Juana Martínez, a la M. en C. Rosa Zugazagoitia, al Dr. Fabrizio Guerrero Mc Manus y su clase de Naturaleza y Sociedad que cambió mi manera de pensar la ciencia. Al Dr. Álvaro Chaos y su clase en donde me presentó a la complejidad. Al Dr. Carlos Martorell, por la mejor clase de la carrera y por permitirme trabajar con él en el servicio social. Al Dr. David P. Sanders por su asesoría en la programación.

Al taller de biología sintética y cía: Viviana Escobar, Claudia Segal y Pablo Padilla, por darnos la oportunidad de trabajar con ellos, por dejarnos inventar proyectos, por la gran experiencia de aprendizaje. A Ángeles Cancino por su respeto a la vida, por gran apoyo y su valiosa amistad.

A los sinodales y sus valiosas revisiones: Dra. Elena Álvarez-Buylla, Dr. Víctor Valdés, Dr. Arturo Becerra y M. en C. Alfonso Vilchis.

Al Dr. Pedro Miramontes, por ser un excelente tutor y persona. Por darme la libertad que necesitaba en el trabajo, por el gusto de conversar con él, por sus libros, idiomas y anécdotas. Por sus ideales de justicia, libertad e integridad. Por su vocación docente.

A la Facultad de Ciencias.

A la UNAM.

ABSTRACT

RNA secondary structure is a convenient abstraction of shape, for which models and software exist that enable quantitative analysis of large sets of sequences. This allows the exploration of physical properties of shape such as *plasticity*, which can be defined as the potential for an RNA sequence to fold into distinct shapes available in a free energy interval from the ground state. Measuring plasticity can be useful to characterize non-coding RNA families, predict relationships between structure and function, and moreover to understand evolutionary properties of shape. A comparative analysis was performed for the secondary structures of 19 different families of RNA for which 4 statistics of plasticity were computed: Probability of Minimum Free Energy Structure, Number of Structures, Structure Variance and Fraction of the Partition Function. The results indicate clear and distinct plastic properties between families as well as differences across taxonomic groups. Also, a pattern of reduced plasticity was observed for several families in which structure is closely related to function, particularly in ancient metabolically relevant families (rRNA 5S, tRNA and bacterial small SRP). This work sets a precedent for future theoretical analysis secondary structure properties across RNA families.

RESUMEN

La estructura secundaria del RNA es una abstracción conveniente para estudiar forma, para la cual existen modelos y software que permiten análisis cuantitativos de grandes grupos de secuencias. A partir de esto se pueden explorar propiedades físicas de la forma como la *plasticidad*, que puede ser definida como el potencial de una secuencia de RNA para plegarse en distintas estructuras, accesibles dentro de un intervalo de energía libre. Esto puede ser útil para caracterizar familias de RNA no codificante, para predecir relaciones entre la estructura y la función, y para entender propiedades evolutivas de la forma. En este trabajo se realizó un análisis comparativo de las estructuras secundarias de 19 familias de RNA para las que se calcularon 4 estadísticas de plasticidad: Probabilidad de la estructura de energía mínima, Número de estructuras, Varianza de las estructuras y Fracción de la función de partición. Los resultados indican que las propiedades plásticas difieren entre las familias comparadas, así como entre grupos taxonómicos. También se observó un patrón de reducción de la plasticidad en diversas familias en donde su función está cercanamente relacionada con la estructura, particularmente en familias antiguas y relacionadas con el metabolismo (rRNA 5S, tRNA y SRP pequeño bacteriano). Este trabajo sienta un precedente para futuros análisis teóricos sobre las propiedades de la estructura secundaria a través de las familias de RNA.

TABLA DE CONTENIDOS

AGRADECIMIENTOS	vii
ABSTRACT	ix
RESUMEN	ix
INTRODUCCIÓN	13
Glosario.....	17
CAPÍTULO 1 PLASTICIDAD Y EL RNA COMO MODELO GENOTIPO-FENOTIPO	19
1.1 Plasticidad fenotípica	19
Evolución y plasticidad	20
Mecanismos de plasticidad	22
1.2 Estructura secundaria como fenotipo.....	24
CAPÍTULO 2 RNA: BIOQUÍMICA, ESTRUCTURA, FUNCIONES Y FAMILIAS	25
2.1 Bioquímica de los ácidos nucleicos	25
Azúcar.....	26
Esqueleto de fosfato.....	26
Bases.....	27
Apilamiento de las bases	28
Apareamiento de bases.....	29
2.2 Estructura del RNA	32
Niveles estructurales.....	34
Estructura secundaria: Elementos estructurales.....	36
2.3 Funciones y procesos	38
Funciones del RNA	38
Algunos procesos	39
2.4 Familias de RNA no codificante.....	42
Familias en Rfam.....	43
Distribución de las familias	44
CAPÍTULO 3 PREDICCIÓN DE ESTRUCTURA SECUNDARIA DEL RNA	49
3.1 Definición formal de la estructura secundaria	49

Representaciones	50
Elementos estructurales	51
3.2 Algoritmos de plegamiento	52
Métodos de combinatoria.....	52
Métodos comparativos.....	54
Métodos de cinética de plegamiento.....	54
ViennaRNA Package.....	55
CAPÍTULO 4 MEDIDAS DE PLASTICIDAD PARA EL RNA ...	57
4.1 Probabilidad de Boltzmann.....	58
4.2 Distancia DE Hamming.....	59
4.3 Medidas.....	59
Medida 1: Probabilidad de la estructura de mínima energía.....	59
Medida 2: Número de estructuras en $5kT$	60
Medida 3: Varianza de las estructuras en $5kT$	60
Medida 4: Fracción de la función de partición.....	60
CAPÍTULO 5 ANÁLISIS DE PLASTICIDAD	63
5.1 Familias elegidas.....	63
Grupo LUCA.....	65
Grupo Transferencia horizontal.....	67
Grupo Exclusivos.....	70
5.2 Método general.....	73
Secuencias.....	73
Análisis.....	74
Reporte de datos.....	75
Interpretación.....	76
RESULTADOS Y DISCUSIÓN	77
Generales.....	77
Particulares.....	84
Consideraciones adicionales.....	89
Perspectivas.....	90
APÉNDICES	93
Apéndice A: Lista de abreviaciones de RNA.....	93
Apéndice B: Algunos detalles adicionales del programa.....	98
Funciones y parámetros del paquete ViennaRNA.....	98
Otros.....	101
Apéndice C: Detalle de grupos comparados.....	103
Selección.....	103
Resultados completos.....	107
REFERENCIAS	127

INTRODUCCIÓN

El RNA es una molécula con gran versatilidad funcional. Originalmente se reconocía únicamente relevante en el proceso de síntesis de proteínas, en donde por un lado cumple con una función informacional: los mRNA que funcionan como intermediarios entre la secuencia del DNA y la maquinaria de traducción de proteínas; y por otro con una función estructural: los tRNA que permiten el reclutamiento del aminoácido codificado y el rRNA que estructura el ribosoma, permitiendo la síntesis de las proteínas. Sin embargo, recientemente se ha observado que existe una gran variedad de familias de RNAs que se transcriben de regiones que no codifican para proteínas y que están fuertemente involucrados en la regulación de la expresión genética. Se ha observado también, que conforme la complejidad del organismo aumenta, las regiones no codificantes predominan en los genomas, mientras que el número de proteínas se mantiene casi estático (Amaral & Mattick, 2008). Todo indica que para comprender los patrones complejos de expresión genética es relevante entender tanto los mecanismos de regulación como las propiedades físicas de los llamados RNAs no codificantes.

Los RNAs no codificantes (ncRNA) son todos aquellos RNAs funcionales que se transcriben pero que no se traducen a proteínas. Su estructura espacial y propiedades físicas están estrechamente relacionadas con su función, por lo que se han desarrollado diversas herramientas para predecir su plegamiento. A pesar de que los algoritmos y el poder de cómputo limitan la predicción de la estructura terciaria, los métodos computacionales para predecir estructura secundaria han sido muy exitosos (Hofacker, 2003). A partir de esto, el RNA ha demostrado ser un buen modelo para el estudio de la relación genotipo-fenotipo, así como para definir y medir formalmente conceptos como *robustez*, *canalización* y *plasticidad* (Ancel y Fontana, 2000), en torno a los cuales se construye la teoría evolutiva contemporánea (Pigliucci, 2001).

OBJETIVO

El presente trabajo busca comparar cuantitativamente la plasticidad entre familias de RNA, desde una perspectiva evolutiva.

JUSTIFICACIÓN

Hasta ahora existen pocos estudios comparativos entre las familias de RNA y están limitados a estudios funcionales o de patrones de distribución taxonómica. Actualmente, los programas de predicción de estructura secundaria permiten hacer análisis cuantitativos extensos sobre propiedades estructurales y parámetros físicos que no han sido interpretados anteriormente en un contexto evolutivo. Caracterizar estas diferencias entre las familias de RNA es relevante en términos de comprensión de su función y su evolución.

ANTECEDENTES

Plasticidad de la estructura secundaria del RNA Los únicos trabajos definiendo formalmente plasticidad en estructura secundaria del RNA son Ancel & Fontana (2000) y Fontana (2002). En estos se aborda extensa y ampliamente el potencial del RNA como modelo para el mapeo genotipo-fenotipo y se describen conceptos como evolutividad, canalización y plasticidad en términos de RNA. Algunos de estos conceptos se definen formalmente y se utilizan para experimentos computacionales sobre evolución dirigida y su efecto en la plasticidad y la evolutividad. Los análisis matemáticos y computacionales se apoyan en el programa **ViennaRNA** y es a partir de estos planteamientos que se propone el análisis realizado en este trabajo.

Plasticidad del RNA y función Nakamura, Ishiguro, y Miyakawa (2012) exhiben la plasticidad conformacional del RNA a través de sus estudios con aptámeros. Llamam *plasticidad del RNA* a la propiedad de esta molécula para formar diversas estructuras que le permiten unirse con alta afinidad a proteínas y otras moléculas. En el artículo se refieren específicamente a la plasticidad de la estructura tridimensional y se discute su posible aplicación en terapias y en desarrollo de biosensores. Aunque no es discutido en un contexto evolutivo, es una evidencia de la diversidad funcional ocasionada por la plasticidad.

Familias de RNA Existen diversos artículos que comparan funcionalmente distintos tipos de RNA (Eddy, 2001; Szweykowska-kulińska, 2014), sin embargo, están basados en categorías informales. La base de datos Rfam es un esfuerzo por crear agrupaciones de RNAs no codificantes por afinidad filogenética y funcional, a estos grupos se les llama *familias de RNA*. Tanto Burge, *et al.* (2013) como Hoepfner, Gardner, y Poole (2012) revisan ampliamente la distribución taxonómica de las familias, basándose en las categorías de la base de datos Rfam. El artículo de Hoepfner, Gardner, y Poole (2012) se usó como base para la elección de los grupos en el análisis realizado en este trabajo. Más información sobre ese artículo se describe en la sección 2.4.

Medidas similares No hay trabajos previos que hayan comparado familias en términos de su plasticidad en sí. Sin embargo, dos artículos han medido características de la distribución de Boltzmann para estructuras secundarias de RNA.

Miklós, Meyer, y Nagy (2005) comparan la distribución de Boltzmann de secuencias de RNA biológicas contra la de secuencias generadas al azar. En su trabajo analizan tres tipos de RNA: miRNAs, tRNAs de *E. coli* y 5S rRNA. A una limitada selección de secuencias de estos grupos se les calculan los siguientes cuatro parámetros: energía mínima de la estructura de cada secuencia, logaritmo de la probabilidad de la estructura de energía mínima, varianza de las energías libres de la distribución de Boltzmann y una estadística de la desviación de la energía mínima esperada. Con este análisis encuentran que las secuencias biológicas tienen menores mínimos de energía y que sus estructuras de mínima energía tienen mayor probabilidad de ocurrir que las secuencias aleatorias. Además, la distribución de Boltzmann de los RNAs biológicos tiene menor varianza y menor desviación que la de los generados al azar. A partir de este estudio concluyen que las secuencias de RNA funcional requieren mínimos de energía termodinámicamente estables.

Otro estudio similar es el de Chan y Ding (2008), en el que utilizan muestreos y características de agrupamiento del ensamble de Boltzmann de estructuras secundarias para comparar entre secuencias biológicas y secuencias aleatorias. Miden 9 características de los ensamblajes en 60 RNAs estructurales (incluyendo tRNAs, rRNAs, RNAs RNasa P, SRP RNAs, tmRNAs, etc.), 12 mRNAs y 46 pre-miRNAs. No encontraron diferencias entre los mRNAs y las secuencias aleatorias, encontraron 5/9 características distintas del azar para los estructurales y 7/9 características distintas del azar para los pre-miRNAs. Con lo anterior concluyen que el ensamble de estructuras de Boltzmann difiere entre distintos tipos de RNA funcionales.

ESTRUCTURA DEL TRABAJO

El Capítulo 1, “Plasticidad y el RNA como modelo genotipo-fenotipo” busca establecer las bases del concepto de plasticidad y enmarcarlo en su contexto evolutivo. También propone una distinción sobre mecanismos de plasticidad que es relevante para extender la definición de plasticidad fenotípica en niveles de organización inferiores a los que tradicionalmente se estudian. Finalmente menciona cómo se puede entender la estructura secundaria del RNA como un fenotipo.

El Capítulo 2, “RNA: Bioquímica, estructura, funciones y familias” busca sentar las bases sobre el RNA como biomolécula. La sección 2.1 es una revisión bioquímica

sobre los ácidos nucleicos. La sección 2.2 discute los niveles estructurales que puede adoptar el RNA, enfocándose en algunas características de la estructura secundaria. La sección 2.3 discute la versatilidad funcional del RNA y su importancia biológica. Finalmente, la sección 2.4 presenta el concepto de familia y menciona características de distribución y función de las mismas.

El Capítulo 3, “Predicción de la estructura secundaria del RNA”, describe la definición formal de la estructura secundaria y revisa los distintos tipos de algoritmos que han sido desarrollados para su predicción. También presenta el **ViennaRNA Package**, que es el programa utilizado para el análisis realizado en este trabajo.

El Capítulo 4, “Medidas de plasticidad”, describe cuatro medidas propuestas por Ancel y Fontana (2000) para cuantificar la plasticidad del RNA a partir de su estructura secundaria. Primero aborda un par de conceptos necesarios para las medidas: la probabilidad de Boltzmann y la distancia de Hamming, y posteriormente describe cada una de ellas y cómo se relacionan con el concepto de plasticidad.

El Capítulo 5, “Análisis de plasticidad”, explica el análisis central de este trabajo, que consiste en medir la plasticidad de diversas familias de RNA, con las medidas propuestas en el Capítulo 4. Primero enlista las familias incluidas y describe su papel biológico con el objetivo de ayudar a la interpretación de los resultados obtenidos. Posteriormente describe el método utilizado y cómo se reportan los resultados.

Finalmente, la sección “Resultados y discusión” presenta el análisis realizado y discute algunas tendencias observadas. Se incluyen patrones generales y particulares, así como otras consideraciones y perspectivas que surgen del trabajo.

GLOSARIO

- apareamiento de bases** Interacción por medio de puentes de hidrógeno entre bases de los ácidos nucleicos.
- apareamiento Watson y Crick** Patrón de apareamiento de bases en el que las purinas interaccionan con las pirimidinas a través de puentes de hidrógeno entre grupos amino y grupos carbonilo. (C-G, A-T y A-U)
- apareamiento no canónico** Patrón de apareamiento de bases distinto al apareamiento Watson-Crick. (i.e. apareamiento por balanceo, por extremos Hoogsteen, etc.)
- apilamiento de las bases** Interacción entre bases apiladas por fuerzas de van der Waals y dipolo-dipolo, ocasionada por el efecto hidrofóbico.
- bucle** (en estructura secundaria) Regiones en las que las bases no están apareadas
- canalización** El refinamiento de las rutas de desarrollo para reducir el ruido de las perturbaciones ambientales (Waddington, '57 en Ancel & Fontana, 2000)
- distancia de Hamming** El número de sustituciones requeridas para convertir una cadena de caracteres en otra.
- elementos estructurales** Subdivisiones arbitrarias de estructuras secundarias que se forman en el RNA, agrupados como hélices o bucles
- estructura primaria** (en RNA) Secuencia lineal de nucleótidos en la molécula de RNA
- estructura secundaria** (en RNA) Relaciones de apareamiento de bases, intramoleculares o intermoleculares de la molécula de RNA
- estructura terciaria** (en RNA) Interacciones tridimensionales que forman una estructura organizada y compacta
- familias de RNA** Agrupaciones de RNAs no codificantes generadas por análisis filogenéticos y estructurales que buscan ser una clasificación natural.
- fenotipo** Las propiedades morfológicas, fisiológicas, bioquímicas, de comportamiento, etc. de un organismo, que se manifiestan durante su vida, especialmente aquellas afectadas por un alelo particular u otra porción del genotipo. (Futuyma, 2005)
- genotipo** Repositorio de información heredable con instrucciones para la producción de moléculas cuyas interacciones, en conjunto con el ambiente, generan y mantienen el fenotipo. (Fontana, 2002)
- hélice** 1. (en estructura secundaria) Regiones con bases apareadas 2. (en estructura terciaria) Estructura tridimensional característica de los ácidos nucleicos apareados
- mecánica estadística** Es una rama de la física que estudia el comportamiento de sistemas de partículas usando teoría de la probabilidad. Se usa comúnmente para explicar

la relación entre el comportamiento de los elementos microscópicos de un sistema y sus propiedades termodinámicas macroscópicas. (Stowe, 2007; Swendsen, 2012)

MFE Mínima energía libre, por sus siglas en inglés *Minimum Free Energy*.

motivo estructural 1. (en estructura primaria) Secuencias cortas conservadas en RNAs funcionales. 2. (en estructura secundaria) Estructuras evolutivamente conservadas 3. (en estructura secundaria) Regiones cortas de apareamientos definidos de bases 4. (en estructura terciaria) Patrón tridimensional formado a partir de hélices, apilamiento, puentes de hidrógeno, etc.

norma de reacción La función que relaciona los ambientes a los que un genotipo puede estar expuesto con el fenotipo que dicho genotipo puede producir. (Pigliucci, 2001)

notación de paréntesis y puntos Representación de la estructura secundaria en donde los puntos corresponden a bases no apareadas y los paréntesis a bases apareadas (que se anidan y se cierran con su pareja correspondiente)

plasticidad (en RNA) La capacidad de una secuencia de RNA para adoptar una variedad de formas energéticamente favorables, al equilibrarse entre ellas a temperatura constante. (Ancel & Fontana, 2000)

plasticidad fenotípica 1. La capacidad (genética) de un individuo para desarrollarse con un fenotipo entre un rango de fenotipos posibles (Ancel & Fontana, 2000) 2. La habilidad de un genotipo para producir fenotipos distintos al estar expuesto a diferentes ambientes durante su ontogenia (Pigliucci, 2005)

plasticidad fenotípica adaptativa Plasticidad ventajosa evolutivamente, que permite a un genotipo tener una mayor tolerancia a condiciones ambientales y consecuentemente tener una mayor adecuación a través de los distintos ambientes en los que se pueda encontrar. (Ghalambor, McKay, Carroll, & Reznick, 2007)

probabilidad de Boltzmann Probabilidad de que un sistema se encuentre en cierto estado en función de su temperatura y de la energía del estado.

pseudonudo Un tipo de bucle en donde sólo una parte del bucle se encuentra apareada con otra región de la molécula

ruido en el desarrollo Variaciones aleatorias que surgen por fluctuaciones micro-ambientales (subcelulares, termodinámicas, etc)

transferencia horizontal Transferencia genética de un organismo a otro que no es su descendiente. Herencia diferente a la *herencia vertical*, que es la transferencia de genes a través de la reproducción sexual o asexual.

último antepasado común universal Organismo hipotético ancestral del cual provienen todos los seres vivos.

CAPÍTULO 1

PLASTICIDAD Y EL RNA COMO MODELO GENOTIPO-FENOTIPO

La palabra plasticidad se puede entender como la capacidad de ser modelado y la posibilidad de cambiar de forma y conservarla. (RAE, 2001) Esta palabra ha sido usada en conceptos biológicos, como es el caso de la plasticidad neuronal y la plasticidad fenotípica, que aunque conservan elementos de la definición de uso común, se refieren por sí mismos a conceptos específicos que es preciso definir con claridad. A continuación se discutirá el término plasticidad fenotípica, su relevancia evolutiva y su definición en términos de RNA.

1.1 PLASTICIDAD FENOTÍPICA

La *plasticidad fenotípica* se puede definir como la habilidad de un genotipo para producir fenotipos distintos al estar expuesto a diferentes ambientes durante su ontogenia. (Pigliucci, 2005) Por otro lado, el *fenotipo* se define como las propiedades morfológicas, fisiológicas, bioquímicas, de comportamiento, etc. de un organismo, que se manifiestan durante su vida, especialmente aquellas afectadas por un alelo particular u otra porción del genotipo. (Futuyma, 2005) Esta definición que se expande a través de los distintos niveles de organización podría incluir, por ejemplo, a las moléculas y cómo sus propiedades son alteradas por el ambiente. A pesar de esto, el cuerpo conceptual alrededor del término de plasticidad fenotípica se ha construido a partir de la definición clásica de fenotipo, refiriéndose en general a características macroscópicas observables a nivel organismo. Por esta razón, el uso de los conceptos canónicos puede ser problemático para definir y describir la plasticidad a través de los distintos niveles de organización.

Wooltereck acuñó en 1909 el término *norma de reacción*, que es hoy en día fundamental para estudiar la plasticidad fenotípica. La norma de reacción se refiere es la función que relaciona los ambientes a los que un genotipo puede estar expuesto con el fenotipo que dicho genotipo puede producir. Sobre-simplificándola se puede representar en una gráfica con el parámetro ambiental

(biótico o abiótico) en el eje x y la medida del fenotipo en el eje y . La plasticidad es entonces un atributo de la norma de reacción, que indica que el genotipo (a través de las interacciones con el ambiente) puede generar distintos fenotipos. La plasticidad sólo es igual a cero cuando la norma de reacción es constante. Cuando no lo es, se le puede caracterizar por su forma y grado. En una norma de reacción lineal, el grado de plasticidad está dado por la pendiente. (Pigliucci, 2001, 2005)

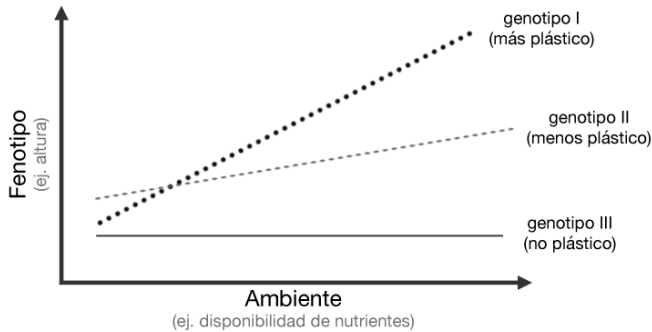


FIGURA 1.1 Normas de reacción Ejemplos de normas de reacción lineales para tres genotipos. [Modificado de (Pigliucci, 2005)]

Evolución y plasticidad

La plasticidad fenotípica se ha convertido en un concepto de gran importancia en teoría evolutiva contemporánea. A pesar de encontrarse en el centro del debate natura vs. nurtura, había sido ignorada durante la primera mitad del siglo XX (Pigliucci, 2001). Actualmente la teoría evolutiva se encuentra en un periodo de discusión y reevaluación en el que se ha extendido el marco teórico para incluir un gran número de conceptos y fenómenos no contemplados en la Síntesis Moderna propuesta en el siglo pasado. La plasticidad es uno de éstos, como se muestra en la figura 1.2 (Pigliucci, 2009) A continuación se discuten algunas relaciones entre evolución y plasticidad:

Valor adaptativo de la plasticidad Se reconoce que la plasticidad puede ser ventajosa evolutivamente cuando permite a un genotipo tener una mayor tolerancia a condiciones ambientales y consecuentemente una mayor adecuación a través de los distintos ambientes en los que se pueda encontrar. A este tipo de plasticidad se le llama *plasticidad fenotípica adaptativa*. Los modelos teóricos predicen que ésta será favorecida cuando las poblaciones están expuestas a ambientes variables, cuando los ambientes producen una señal confiable (para que regule la respuesta plástica), cuando la selección favorece fenotipos distintos en cada

ambiente y cuando ningún fenotipo es adecuado en todos los ambientes. (Ghalambor et al., 2007; Pigliucci, 2001)

Sin embargo, la mayoría de los casos de variación fenotípica inducida por el ambiente parecen tener respuestas plásticas no adaptativas, (Ghalambor et al., 2007) es decir, que la respuesta es en realidad producto de limitaciones ambientales (por ejemplo, el tamaño disminuido de una planta ante la falta de nutrientes en el suelo.) Es importante preguntar cuándo la respuesta plástica es una característica adaptativa y cuándo es un subproducto no adaptativo de las condiciones en distintos ambientes. El análisis de los patrones de variación de un fenotipo en un contexto evolutivo nos permitiría distinguir si efectivamente la plasticidad observada para cierto fenotipo ha evolucionado como una respuesta en sí.

Costos y límites de la plasticidad Al igual que otras características adaptativas, la evolución de la plasticidad está restringida por los límites (genéticos, físicos, etc.) del carácter y por sus costos (desde una perspectiva ecológica). DeWitt, *et al.* (1998) clasificaron los costos y límites de la evolución de la plasticidad de la siguiente manera:

Costos	Límites
Mantenimiento: Costos energéticos de mantener los mecanismos sensibles y regulatorios	Confiabilidad de la información: La información que se puede obtener del ambiente puede no ser confiable o cambiar demasiado rápido.
Producción: Costo de la producción de las características de manera plástica (comparado contra el costo de una respuesta constante)	Tiempo de retraso: La respuesta puede tardar en presentarse después del cambio ambiental.
Adquisición de la información: Gastos de energía al muestrear el ambiente (energía y tiempo no utilizado para otras actividades adaptativas)	Rango del desarrollo: El rango de fenotipos expresados por un genotipo plástico no puede ser infinitamente amplio
Inestabilidad en el desarrollo: Puede dar cierta imprecisión en el desarrollo si la plasticidad implica una reducción de la canalización	Problema del epifenotipo: La evolución del mecanismo plástico tiene que estar anclada en un proceso de desarrollo ya existente por lo que su efectividad no es óptima.
Genética: Mantenimiento de los genes de plasticidad.	

[Modificada de Pigliucci (2001) según DeWitt et al. (1998)]

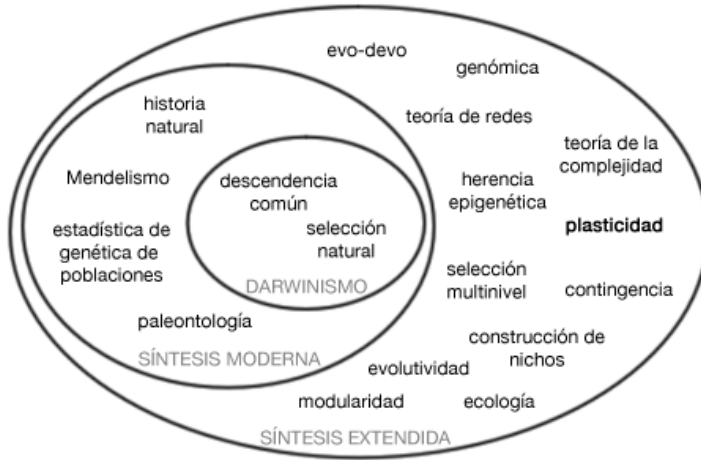


FIGURA 1.2 Expansión de la teoría evolutiva Ideas, fenómenos y campos estudiados en las distintas etapas de la teoría evolutiva [modificado de (Pigliucci, 2009)]

La plasticidad y sus contribuciones a la teoría evolutiva Pigliucci (2001) argumenta que muy pocos componentes de la teoría evolutiva no se ven afectados por los estudios sobre plasticidad fenotípica y enlista algunos: 1) Una perspectiva plástica contribuye a entender cómo es que surgen las adaptaciones. Los organismos deben adaptarse a múltiples condiciones ambientales en diferentes grados y tipos de heterogeneidad ambiental durante toda su vida. Lo anterior limita más estrictamente las posibilidades de la fuerza creadora de la selección natural. 2) Los estudios de plasticidad han demostrado la existencia de variación natural y han ayudado a modelar las circunstancias que llevan a su existencia y mantenimiento. 3) En genética cuantitativa, la plasticidad ha permitido demostrar que la heredabilidad puede ser dependiente del ambiente 4) Se ha estudiado la evolución de la regulación genética y cómo funcionan los mecanismos de respuesta ante distintos ambientes 5) Se han trazado líneas conectando microevolución y macroevolución al estudiar los mecanismos de regulación plástica y su efecto sobre fenotipos macroscópicos.

Mecanismos de plasticidad

Los estudios clásicos de plasticidad fenotípica buscan entender principalmente los mecanismos ontogenéticos que dan origen a la plasticidad, entendiendo la ontogenia como el proceso de desarrollo que sigue un organismo desde que es un cigoto hasta que es un adulto. (Owen & Daintith, 2004) Claramente, este acercamiento resulta inútil para estudiar la plasticidad de fenotipos que carecen de

desarrollo, como las moléculas. Sin embargo, es posible generalizar la idea detrás de esta definición al considerar que para cada fenotipo, en el nivel de organización que sea, existen mecanismos que permiten respuestas plásticas a partir de su interacción con el ambiente aunque la naturaleza de estos mecanismos sea distinta para cada nivel.

Puesto que algunas moléculas como las proteínas o el RNA poseen propiedades físicas inherentes que les permiten plegarse para adoptar conformaciones estructurales específicas y el ambiente cambia la manera en la que esto sucede, podríamos considerar dos tipos de mecanismos de plasticidad:

1. Un *mecanismo de plasticidad intrínseco* de la molécula, es decir, que por sí misma se comporte distinto al cambiar ciertas condiciones físicas, como la temperatura o el pH. Por ejemplo: el plegamiento secundario del RNA, que se da a partir de interacciones de apareamiento entre las bases que conforman a la molécula y cuya energía libre puede oscilar alrededor de un mínimo de energía.
2. Un *mecanismo de plasticidad no-intrínseco*, sería aquel en el que hay una respuesta a un estímulo ambiental que ocurre por la interacción de elementos moleculares coordinados. Por ejemplo: una proteína de choque térmico que guíe el plegamiento de otra proteína en un ambiente estresante o una red de regulación genética que permita la expresión de ciertas moléculas en distintas condiciones ambientales.

Esta distinción es importante para aclarar la diferencia entre estudiar la biología molecular de las respuestas plásticas y estudiar la plasticidad de las moléculas. Aunque la plasticidad de los fenotipos macroscópicos esté mediada de manera última por moléculas, las respuestas plásticas no necesariamente surgen a partir de la plasticidad intrínseca de éstas. Las consideraciones en torno a la evolución de ambos mecanismos son muy distintas: los mecanismos no-intrínsecos evolucionarían como redes de interacción entre elementos genéticos y ambientales que producen respuestas alternativas, mientras que los mecanismos intrínsecos son una propiedad inherente de la molécula que podría ser moldeada a partir de cambios en la secuencia o por cambios de la interacción molecular misma.

Dada la definición de un mecanismo de plasticidad intrínseco: ¿Qué distingue a la plasticidad de las variaciones fenotípicas por *ruido en el desarrollo*? Pigliucci (2001) discute que son distintos fenómenos y que no se encuentran correlacionados. El *ruido en el desarrollo* se entiende generalmente como variaciones aleatorias que surgen por fluctuaciones microambientales (subcelulares, termodinámicas, etc.), mientras que la plasticidad es una respuesta definida a cierta variación

macroambiental. Con esta definición, podría entenderse la plasticidad de las moléculas por mecanismos intrínsecos como un fenómeno de ruido. Sin embargo, considero que no es adecuado hacerlo puesto que la diferencia entre los conceptos de ruido y plasticidad yace en gran parte en cómo se interpreta el resultado de las fluctuaciones microambientales. En niveles de organización superiores podrían ser interpretadas como ruido, pero cuando definimos la molécula como el fenotipo mismo, la respuesta variable posee las características mencionadas en las secciones anteriores y por lo tanto puede ser interpretada como un fenómeno de plasticidad.

1.2 ESTRUCTURA SECUNDARIA COMO FENOTIPO

Los biopolímeros como el RNA se pliegan debido a sus interacciones intramoleculares. La estructura secundaria se refiere al patrón de apareamiento de bases de la molécula de RNA (ver sección 2.2) y es una abstracción interesante que se encuentra en correspondencia con algunas propiedades funcionales de la molécula y da una aproximación geométrica y termodinámica de la estructura terciaria (Ancel & Fontana, 2000). Una ventaja de trabajar con la estructura secundaria es que es un modelo fácilmente computable, que además de tener muchas herramientas ya desarrolladas, ha demostrado ser exitoso. (I. L. Hofacker, 2003) El RNA además tiene la ventaja de poderse estudiar tanto con modelación computacional como con biología molecular experimental. Por estas razones es un buen modelo mínimo de la relación genotipo-fenotipo, siendo el genotipo el gen a partir del que se transcribe la molécula de RNA y el fenotipo la estructura secundaria que adopta la molécula. A pesar de que el plegamiento del RNA no corresponde a un fenómeno de desarrollo, es posible definir operacionalmente conceptos como plasticidad fenotípica, canalización, modularidad, robustez, evolutividad, etc. (Ancel & Fontana, 2000)

Una molécula de RNA se supone plástica porque oscila entre conformaciones debido a pequeñas fluctuaciones de energía. Sin embargo, cuantificar la magnitud de la plasticidad es complicado. El ejemplo ilustrado en la Figura 1.1 supone que los ejes pueden ser fácilmente definidos y medidos, y la plasticidad está dada por la pendiente de la norma de reacción lineal, pero esto no suele ocurrir. Para cuantificar la plasticidad de la estructura secundaria del RNA requerimos medidas que integren distintos tipos de información sobre el comportamiento plástico de la molécula. El acercamiento implementado por Ancel & Fontana (2000) consiste en definir un rango de energía para el cual calcular las distintas respuestas conformacionales y recolectar la siguiente información: el número de estructuras posibles de la molécula, la ubicación de esas estructuras y su variación. Esto se discute con detalle en el capítulo 4.

CAPÍTULO 2

RNA: BIOQUÍMICA, ESTRUCTURA, FUNCIONES Y FAMILIAS

Los ácidos nucleicos DNA (ácido desoxirribonucleico) y RNA (ácido ribonucleico) son repositorios moleculares de información genética. Ambos son polímeros de unidades monoméricas llamadas nucleótidos y es en la secuencia de éstos que yace tanto su información como su diversidad bioquímica y funcional. (Elliot & Ladomery, 2011; Nelson & Cox, 2008) Estas moléculas tienen una relación muy estrecha: el DNA codifica la información para la síntesis de RNA o de proteínas (a través del mRNA). Estos flujos de información fueron originalmente descritos como el dogma central de la biología molecular (Crick, 1970) y han sido posteriormente ampliados y modificados (Shapiro, 2009) generando un panorama cada vez más amplio que esclarece su evolución y funcionamiento.

2.1 BIOQUÍMICA DE LOS ÁCIDOS NUCLEICOS

Los ácidos nucleicos tienen tres componentes que los caracterizan: una base nitrogenada, un azúcar pentosa y un grupo fosfato.

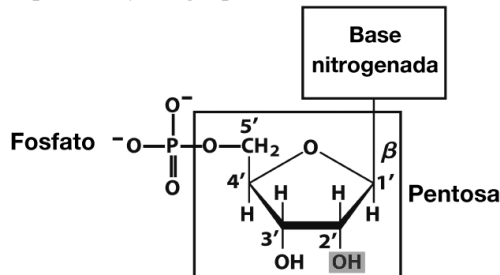


FIGURA 2.1 Estructura general de los nucleótidos. Mostrando la pentosa y los sitios de unión al grupo fosfato y la base nitrogenada. [Modificada de (Nelson & Cox, 2008)]

Azúcar

La pentosa es un monosacárido que tiene forma de ribosa en el caso del RNA y de desoxirribosa en el DNA. Son moléculas cíclicas de cinco carbonos que forman un anillo de cuatro carbonos y un oxígeno, con radicales unidos a cada uno de los carbonos. Los carbonos están numerados del 1' al 5' como una convención de nomenclatura bioquímica para distinguirlos de aquellos carbonos en las bases (Nelson & Cox, 2008). La ribosa tiene un grupo –OH en el carbono 2', mientras que la desoxirribosa tiene un hidrógeno, como se muestra en la figura 2.2. Debido a la alta electronegatividad del oxígeno la ribosa del RNA obtiene un carácter polar y por lo tanto es más reactiva químicamente que la desoxirribosa encontrada en el DNA. En términos de estructura espacial, el anillo de la ribosa tiene una ligera torsión que minimiza las interacciones entre el grupo 2' – OH y el resto de los carbonos, esto tiene consecuencias en la estructura secundaria que pueden formar las ribosomas. (Elliot & Ladomery, 2011)

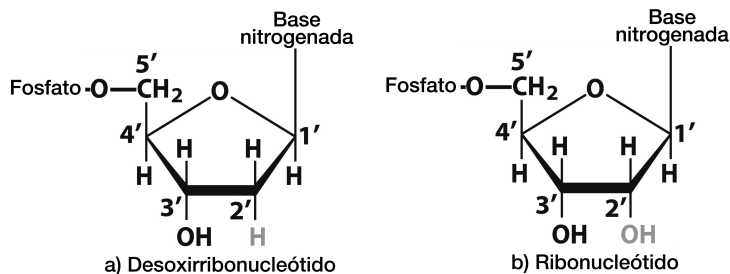


FIGURA 2.2 Pentosas en el DNA y el RNA La diferencia se encuentra en el carbono 2' que tiene a) -H en el caso de los desoxirribonucleótidos del DNA y b) -OH en el caso de los ribonucleótidos del RNA. [Modificada de (Nelson & Cox, 2008)]

Esqueleto de fosfato

El grupo fosfato se encuentra esterificado al carbono 5' como se muestra en la figura 2.1. Es por medio de este grupo que se enlazan los nucleótidos para formar la molécula como polímero tanto en el DNA como en el RNA. Esto sucede por medio de un enlace fosfodiéster del grupo fosfato en 5' en un nucleótido al grupo hidroxilo 3' del siguiente, como se muestra en la figura 2.3. A esta secuencia de pentosas unidas por enlaces fosfodiéster se le conoce como el esqueleto de fosfato mientras que las bases nitrogenadas quedan como residuos libres unidos a la cadena. La región del esqueleto es hidrofílica, pues los grupos hidroxilo del azúcar forman puentes de hidrógeno con el agua. Los grupos de fosfato tienen un pK_a cercano a 0, por lo que tienen carga negativa y se encuentran completamente ionizados en pH 7. Estas cargas negativas suelen estar neutralizadas por medio de

interacciones con cargas positivas de proteínas, iones metálicos y poliaminas. (Nelson & Cox, 2008)

Todos los enlaces fosfodiéster tienen la misma forma que comienza con el grupo fosfato 5' y termina con el grupo hidroxilo 3', esto le da una orientación específica a la molécula y los extremos distintos se denominan simplemente 5' y 3', dependiendo de dónde le falte el nucleótido que seguiría en la cadena.

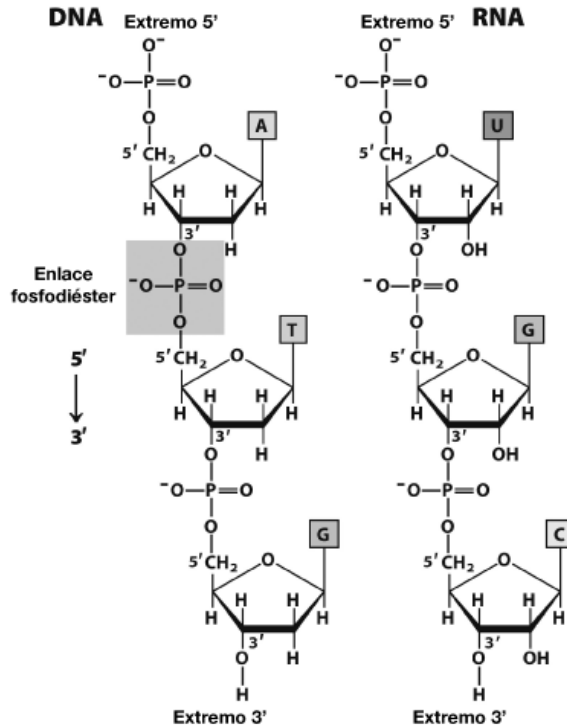


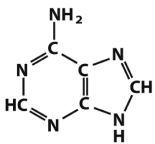
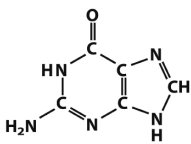
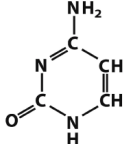
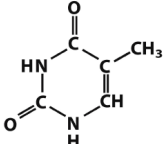
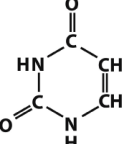
FIGURA 2.3 Polimerización del DNA y el RNA Por medio del enlace fosfodiéster que se muestra sombreado. [Modificado de (Nelson & Cox, 2008)]

Bases

Las bases son compuestos nitrogenados derivados de dos moléculas: pirimidina y purina. Éstas se unen a la pentosa covalentemente por un enlace N-β-glucosídico formado entre el carbono 1' y el nitrógeno 1 para las pirimidinas y el nitrógeno 9 para las purinas. El RNA y el DNA usan principalmente 4 bases: las purinas adenina (A) y guanina (G) y la pirimidina citosina (C) se encuentran tanto en DNA

como en RNA. Como segunda pirimidina el DNA suele usar timina (T) y el RNA usa uracilo (U) (Nelson & Cox, 2008). La tabla 2.1 muestra la estructura molecular de las bases.

TABLA 2.1 Bases nitrogenadas.

Purinas		Pirimidinas		
Adenina (A)	Guanina (G)	Citosina (C)	Timina (T)	Uracilo (U)
DNA y RNA	DNA y RNA	DNA y RNA	DNA	RNA
				

[modificada de Nelson & Cox (2008)]

Se les llama bases porque químicamente son bases débiles. Son moléculas aromáticas, lo cual les da características particulares de distribución de los electrones y por lo tanto de absorción de luz. Son moléculas hidrofóbicas e insolubles en agua en el pH intracelular.

También existen bases menores. Pueden ser bases modificadas por metilaciones, hidroximetilaciones, glucosificaciones, desaminaciones, etc. Estas modificaciones suelen ser relevantes en procesos de regulación o de mantenimiento, así como en ciertos RNAs. La base modificada inosina (I) participa en el apareamiento codón-anticodón de los tRNAs y en procesos de regulación por edición del mRNA, en el que la adenina y citosina son alteradas por desaminación enzimática y cambian a inosina y uracilo, respectivamente. La inosina es después interpretada como guanina durante la traducción. (Nelson & Cox, 2008)

Las interacciones entre las bases son fundamentales en la conformación estructural de los ácidos nucleicos. A continuación se describen dos de ellas: apilamiento y apareamiento. La figura 2.4 muestra estas dos interacciones en un bucle en horquilla de una molécula de RNA.

Apilamiento de las bases

En una cadena polinucleotídica, las bases consecutivas se juntan en pilas, con los planos de sus anillos acomodados de manera paralela. La interacción está dada por fuerzas de van der Waals y dipolo-dipolo. Es ocasionada por el efecto hidrofóbico, por el cual se favorece este acomodo que minimiza el contacto de las bases con el agua. (Nelson & Cox, 2008) Naturalmente, la identidad de las bases apiladas

influye en la energía libre de la molécula plegada, puesto que no todas las bases interactúan de la misma manera.

Bucle en horquilla de RNA

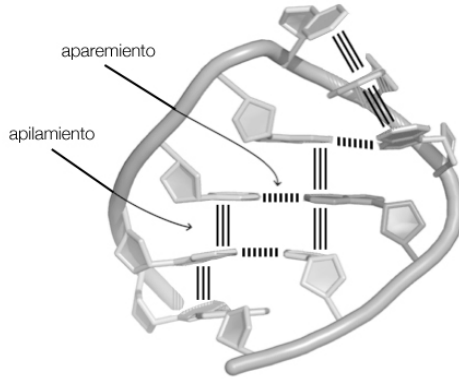


FIGURA 2.4 Interacciones de bases Se muestran con líneas oscuras las interacciones de apareamiento y apilamiento en un bucle en horquilla de RNA. [Modificada de Williams, L. (2012)]

Apareamiento de bases

Watson y Crick describieron un patrón de interacción entre bases por medio de puentes de hidrógeno entre los grupos amino y carbonilos de las bases nitrogenadas. Una purina interactúa con una pirimidina de la siguiente manera: las citosinas se enlazan específicamente con las guaninas, mientras que las adeninas con la timinas o con los uracilos. (Watson & Crick, 1953) Este patrón de interacción forma pares de bases y es conocido como el apareamiento Watson y Crick. La geometría de este tipo de apareamiento se muestra en la figura 2.5.

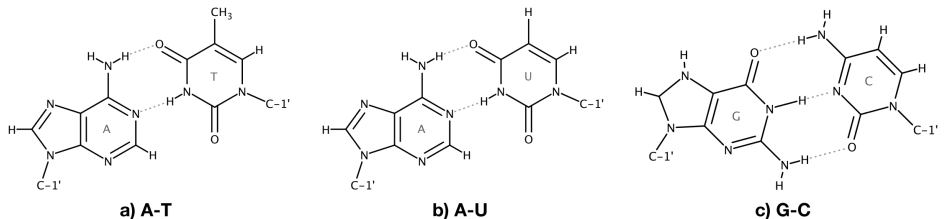


FIGURA 2.5 Apareamiento Watson y Crick a) Adenina y timina apareadas por dos puentes de hidrógeno comúnmente hallado en el DNA b) Adenina y uracilo apareadas, en el caso del RNA, c) Apareamiento de guanina y citosina, unidas por tres puentes de hidrógeno.

Existen otras formas de apareamiento de bases por puentes de hidrógeno, que son particularmente relevantes en los estudios estructurales y funcionales del RNA. El más estudiado y ubicuo en el RNA es el apareamiento por balanceo (o *wobble* en inglés) que propuso originalmente Crick para explicar el apareamiento de codón-anticodón de los tRNAs. (Brown, 2008) La pareja G•U forma un apareamiento por balanceo por medio de dos puentes de hidrógeno como se muestra en la figura 2.6 y ha sido confirmada en casi todas las clases de RNA en los tres dominios filogenéticos. Su estabilidad termodinámica se asemeja a la de los apareamientos Watson y Crick y es mayor que el de otros apareamientos no canónicos. Este apareamiento tiene características bioquímicas que otorgan funciones específicas irremplazables, por lo que su ocurrencia ha sido conservada evolutivamente, por ejemplo: expone muchos grupos donadores de hidrógeno expuestos, que le permiten ser reconocido por otros RNAs, por proteínas u otros ligandos; también crea un gradiente electrostático pronunciado que permite la unión de iones metálicos por el arco mayor, y esto es probablemente importante para la catálisis del RNA. (Varani & McClain, 2000)

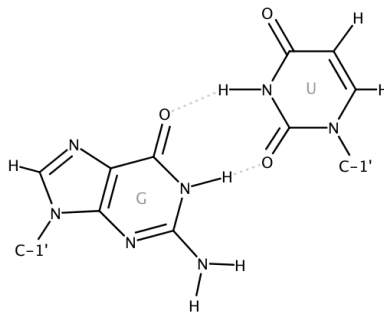
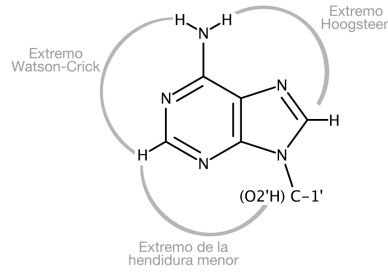
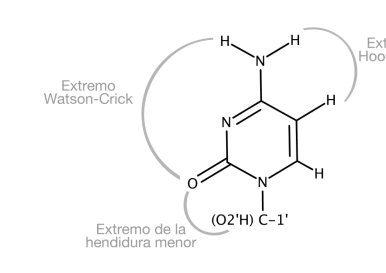
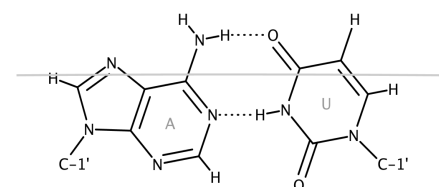
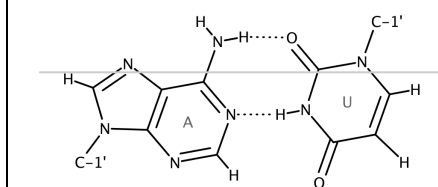


FIGURA 2.6 Guanina y uracilo unidas en apareamiento por balanceo

Además de apareamientos entre bases no tradicionales, existe una gran diversidad de geometrías de apareamiento distintas que son difíciles de clasificar, pero se suelen separar con respecto a los criterios mostrados en la tabla 2.2 (Leontis & Westhof, 2001; Westhof & Fritsch, 2000) que combinadas forman 12 familias de pares de bases (2 posibles orientaciones de las azúcares, 3 sitios posibles de formación de puentes de hidrógeno de la primera base y 3 sitios posibles de formación de puentes de hidrógeno de la segunda base)

TABLA 2.2 Variaciones geométricas en apareamientos no canónicos

Por el sitio de formación de los puentes de hidrógeno		
[1] Extremo Watson-Crick	[2] Extremo Hoogsteen	[3] Extremo de la hendidura menor
 <p style="text-align: center;">Purina</p>		 <p style="text-align: center;">Pirimidina</p>
<p>[1] El extremo Watson-Crick tiene los sitios tradicionales de unión descritos en el modelo de Watson y Crick. [2] Hoogsteen observó en 1963 que podía haber apareamientos con puentes de hidrógeno formados a partir de otro extremos y describió un apareamiento tal entre A y T, posteriormente se han descrito apareamientos de este tipo entre C y G, así como entre extremos Watson-Crick y extremos Hoogsteen [3] El extremo de la hendidura menor es también llamado extremo del azúcar, es común entre adeninas y se han encontrado interactuando con extremos Watson-Crick, Hoogsteen u otros extremos de la hendidura menor.</p>		
Orientación de las azúcares		
[1] cis	[2] trans	
		
<p>Se nombran por la orientación de las azúcares con respecto a la formación de los puentes de hidrógeno, puede ocurrir con apareamientos desde cualquier extremo. [1] es un ejemplo con apareamiento cis-A-U y [2] es un ejemplo con apareamiento trans-A-U</p>		

Además de estas variantes, existen también otras tales como bases unidas por puentes de hidrógeno bifurcados y bases unidas con puentes de hidrógeno con agua insertada. (Leontis & Westhof, 2001; Westhof & Fritsch, 2000) Todos estos apareamientos han sido relacionados con roles funcionales en el RNA (Hermann & Westhof, 1999; Leontis & Westhof, 2001; Westhof & Fritsch, 2000) Sin embargo las interacciones causadas por estos apareamientos no canónicos son particularmente relevantes en estructura terciaria, por lo que los programas que predicen estructura secundaria se han limitado a considerar los apareamientos

Watson-Crick y el G•U por balanceo en sus algoritmos de plegamiento como es el caso del programa ViennaRNA Package 2.0 (Lorenz et al., 2011)

ESPECIFICACIONES SOBRE LA NOTACIÓN

El apareamiento Watson y Crick se suele denotar con un guión sencillo como: A-T, A-U, C-G o alternativamente con guiones representando los puentes de hidrógeno de la siguiente manera: A=T, A=U, C≡G. Mucha literatura muestra los pares por balanceo unidos por un punto obscuro como G•U, mientras que Westhof & Fritsch (2000) prefieren usar un punto blanco como en G°U y usar el punto obscuro para apareamientos distintos a los Watson y Crick con geometrías no canónicas como A•C. En esta tesis se usan los guiones sencillos para referirse a los pares de bases unidos de forma Watson y Crick, y los puntos oscuros para los el apareamiento por balanceo. Los apareamientos con geometrías no canónicas no son considerados por el programa de plegamiento ViennaRNA Package 2.0, por lo que esta notación es innecesaria en este caso.

Apareamiento Watson y Crick	Apareamiento por balanceo
A-T A-U C-G	G•U

2.2 ESTRUCTURA DEL RNA

El RNA se caracteriza por su gran diversidad funcional, producto de su versatilidad de forma. A pesar de ser ácidos nucleicos sutilmente diferentes al DNA, los RNAs tienen estructuras más parecidas a las proteínas que al DNA en términos de complejidad, diversidad e incluso función. (Elliot & Ladomery, 2011)

Las diferencias bioquímicas son relevantes en estas notorias diferencias en su estructura, sin embargo, éstas se deben principalmente a su síntesis, a su procesamiento, a cómo se encuentran en el medio celular y a las interacciones que tienen con otras moléculas; mientras que las hebras de DNA se separan sólo momentáneamente, el RNA se sintetiza a partir de una hebra de DNA y se separa formando un ssRNA (RNA de cadena sencilla, *single-stranded RNA* en inglés) sin cadena complementaria, incapaz de formar una estructura de doble hélice larga. El RNA forma entonces hélices más cortas a partir de apareamiento de bases de la misma molécula o en ocasiones con otras moléculas. La Figura 2.3 muestra lado a lado, como ejemplo, una RNasa P con su tRNA sustrato y una doble hélice de DNA.

Ambas moléculas forman hélices dobles: como se muestra en la Figura 2.4, en el medio celular, las moléculas de RNA forman una hélice A. En la hélice B de DNA, las bases se encuentran perpendiculares al esqueleto de fosfato y hay 10 nucleótidos por vuelta. La hélice A de RNA es más compacta, tiene 11 nucleótidos por vuelta y las bases no son perpendiculares. (Arnott, 2006)

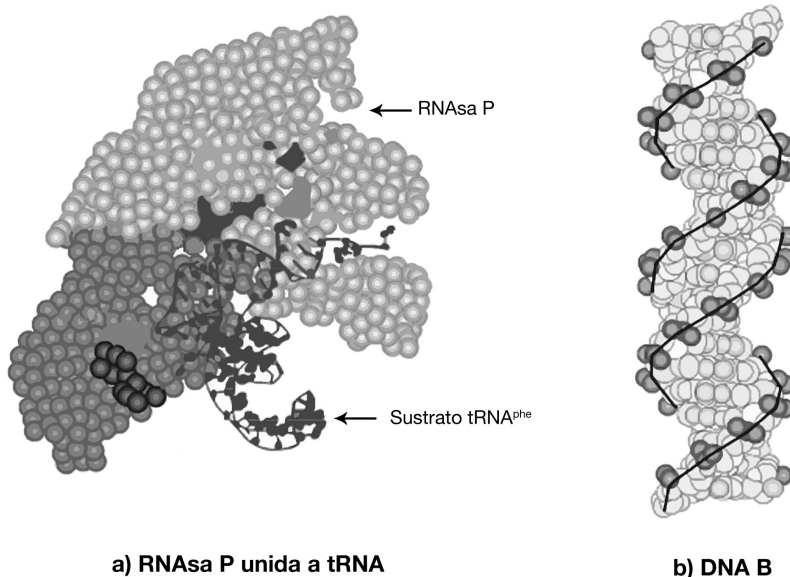


FIGURA 2.3 Estructura tridimensional del RNA y del DNA a) Estructura terciaria de dos moléculas de RNA unidas. La estructura representada con modelo de espacio lleno es el componente de RNA de la RNAsa P (se observa su forma globular). La estructura helicoidal, unida al sitio activo es un tRNA^{phe}. b) Modelo de espacio lleno de la doble hélice del DNA [modificado de Elliot & Ladomery (2011) que a su vez fue reimpresso de Torres-Larios *et al* (2005) y Wang *et al* (1979)]

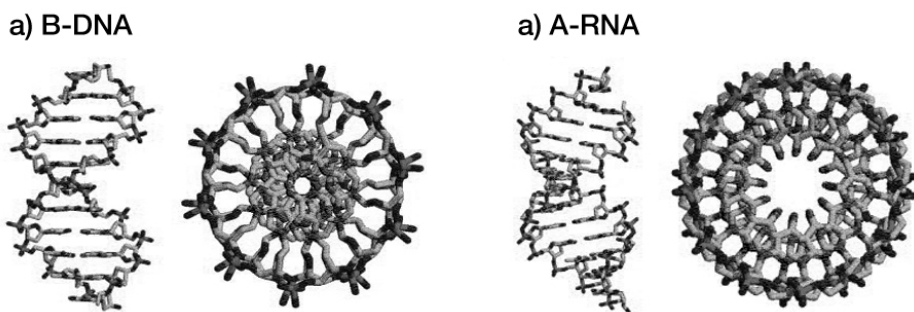


FIGURA 2.4 Hélices de RNA y de DNA Ambas en conformación usual en la célula a) DNA en conformación B, b) RNA en conformación A [Modificada de (Arnott, 2006)]

Niveles estructurales

Las moléculas de RNA se pliegan en estructuras tridimensionales complejas. Para estudiar las relaciones entre los elementos de la molécula que contribuyen a dicha estructura, se pueden identificar tres niveles jerárquicos de organización estructural: estructura primaria, secundaria y terciaria. Antes de definirlos, es importante notar que aunque estos niveles de clasificación son también usados en estudios de estructura de proteínas, no se refieren a lo mismo; no son los mismos tipos de relaciones las que definen al nivel de organización y por lo tanto no son términos equivalentes. De igual manera el concepto de motivo, que en proteínas se refiere a un patrón de plegamiento súpersecundario conservado, en RNA cada nivel estructural tiene un uso distinto del concepto. Esto se discute en los siguientes párrafos y se resume en la tabla 2.3.

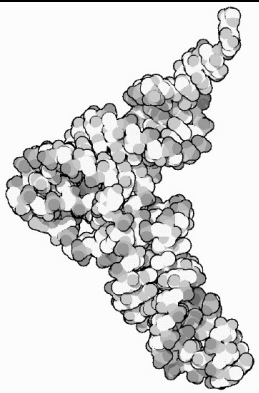
La estructura primaria es la secuencia lineal de nucleótidos en la molécula de RNA. Se representa como una cadena de caracteres en donde cada carácter es una base. Los ‘motivos’ fueron primero definidos en este nivel, y se referían a secuencias cortas, conservadas que ocurrían comúnmente en RNAs funcionales (como tRNAs o rRNAs) usualmente como sitios de reconocimiento. Estos se suelen llamar *motivos de secuencia* y se describen con secuencias consenso, que permiten ver los cambios evolutivos y las partes más conservadas. Un ejemplo es la secuencia Shine–Dalgarno en los mRNAs procariontes, que permite la interacción con el rRNA para iniciar la síntesis de proteínas. (Hendrix, Brenner, & Holbrook, 2005; Leontis, Lescoute, & Westhof, 2006)

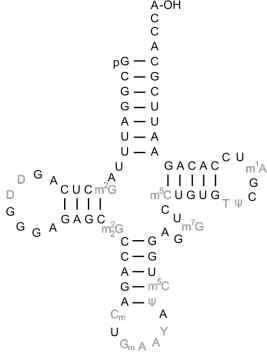
La estructura secundaria se refiere a las relaciones de apareamiento de bases, intramoleculares o intermoleculares. “El término ‘estructura secundaria’ denota cualquier patrón plano de contactos por apareamiento de bases. Es un concepto topológico y no debe ser confundido con algún tipo de estructura bidimensional.” (Fontana, 2002, p. 4) Las estructuras secundarias son típicamente representadas como grafos (ver sección 3.1) En este nivel se suele entender ‘motivo’ de dos maneras distintas: uno es como una estructura secundaria seleccionada y conservada evolutivamente (Hendrix et al., 2005; Leontis et al., 2006) que se puede identificar con métodos comparativos (ver sección 3.2) y otra es como regiones cortas con un patrón específico de apareamiento de bases, separadas en hélices (regiones apareadas) y bucles (regiones no apareadas) que también son llamados más precisamente elementos estructurales (Ancel & Fontana, 2000; Elliot & Ladomery, 2011; Tian, Bevilacqua, Diegelman-Parente, & Mathews, 2004)

La estructura terciaria comprende las interacciones tridimensionales que forman una estructura organizada y compacta. Este último nivel de organización se logra por medio de distintas interacciones, entre ellas: apilamiento coaxial, que son

hélices de RNA que se apilan y estabilizan la molécula; enlaces de hidrógeno distintos a los de apareamiento de bases como los apareamientos triples, como aquellos dados en los extremos Hoogsteen y los zippers de ribosa; iones metálicos como magnesio y sodio que mantienen hélices juntas, entre otras. (Moore, 1999) En este nivel, se entienden los ‘motivos’ como formas tridimensionales recurrentes (muchas como consecuencia directa de la estructura secundaria) pero tomando en cuenta la posición espacial de los átomos de la molécula. Es decir, cuando se habla de hélice en este nivel se refiere a la hélice tipo A con sus parámetros tridimensionales, mientras que una hélice en estructura secundaria se refiere a una región de bases apareadas. Claro que una hélice en estructura terciaria es consecuencia de una hélice en estructura secundaria (si no cuenta con otras interacciones adicionales), sin embargo es importante no confundir los términos y dejar claro a qué nivel se refieren puesto aportan información distinta sobre la conformación de la molécula.

Las definiciones anteriores, tanto las de los niveles estructurales como las definiciones de ‘motivos’ suelen ser ambiguas y usadas sin cuidado en la literatura, sin embargo algunos autores proponen definiciones más claras para los niveles de estructura (Ancel & Fontana, 2000; Fontana, 2002). Por otro lado, Hendrix, et al. (2005) discuten que el concepto de ‘motivo’ debe ser entendido abarcando los tres niveles estructurales como unidades conservadas y sujetas a selección. En todo caso, habría que renombrar adecuadamente los elementos de cada nivel para evitar ambigüedades.

Nivel	Representación ¹	Definición	Definición de motivo
Estructura terciaria		Configuración global tridimensional de la molécula. Incluyendo interacciones espaciales diversas.	Patrón tridimensional formado a partir de hélices, apilamiento, puentes de hidrógeno, etc.

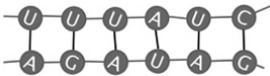
<p>Estructura secundaria</p>		<p>Patrón de contactos por apareamiento de bases, intramoleculares o intermoleculares</p>	<p>Se definen los motivos de dos maneras: -Como estructuras secundarias evolutivamente conservadas. -Como regiones cortas de apareamiento de bases. (También llamados Elementos estructurales o simplemente motivos estructurales.</p>
<p>Estructura primaria</p>	<p>5' – GCGGAUUUA [...] CAGAAUJCGCACCA 3'</p>	<p>Secuencia lineal de nucleótidos en la molécula de RNA</p>	<p>Se definen los motivos como motivos de secuencia: Son secuencias cortas conservadas en RNAs funcionales.</p>

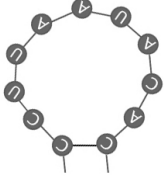
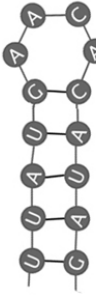
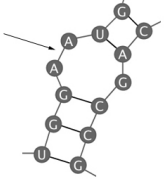
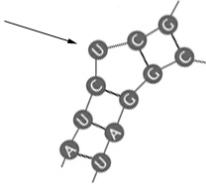
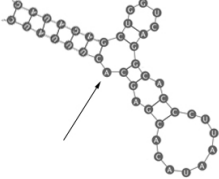

1. La columna muestra un solo tipo de representación por nivel estructural, pero existe un repertorio amplio de representaciones para casa uno. [Figura estructura terciaria de Protein Data Bank, & Goodsell, D. S. (2001); Figura estructura secundaria de Yikrazuul. (2010)]

Estructura secundaria: Elementos estructurales

Como he discutido, las conformaciones estructurales en los distintos niveles tienen nombres ambiguos. En este trabajo, que se enfoca en estructura secundaria del RNA, se usa el término elementos estructurales para referirse a los tipos de estructuras secundarias que se forman en el RNA, agrupados como hélice (para aquellas regiones con bases apareadas), bucle para aquellas regiones en donde las bases no están apareadas y pseudonudo para un tipo especial de bucle en donde sólo una parte del bucle sí está apareada. Aquellas bases que no pertenecen a ninguna de estas estructuras se llaman bases externas y las estructuras o grupos de estructuras delimitados por bases externas se llaman componentes estructurales

La tabla 2.4 muestra los elementos estructurales comunes con sus definiciones, algunos de sus sinónimos en español e inglés y una representación del elemento estructural aislado.

<p>TABLA 2.4 Elementos estructurales comunes</p>		
<p>Hélice</p>	<p>Región con varias bases apareadas. Generalmente son estructuras estabilizadoras de la molécula. también: tallo, stacks</p>	

<p>Bucle</p>	<p>Región incluida en una hélice en donde las bases no están apareadas también: loop, cycle</p>	
<p>Bucle en horquilla</p>	<p>Estructura en donde regiones cercanas de bases complementarias se aparean, separadas por una región no apareada que permite que la secuencia se doble para formar una hélice. Dependiendo de la longitud del bucle recibe el nombre de tetraloop, pentaloop, etc. también: hairpin loop, stem loop</p>	
<p>Bucle interno</p>	<p>Estructura en donde hay regiones no apareadas en ambos lados de la hebra. Puede ser simétrico o asimétrico. también: internal loop</p>	
<p>Protuberancia</p>	<p>Estructura en donde hay una región no apareada en un solo lado de la hebra. también: bulge</p>	
<p>Bucle múltiple</p>	<p>Región en donde se juntan varias hélices. Dependiendo de cuántas hélices involucra recibe los nombres. también: helical junction, multiloop</p>	
<p>Pseudonudo</p>	<p>Es un tipo de bucle especial, en donde sólo una parte del bucle sí está apareada. El pseudonudo más simple consiste de una región libre del RNA apareada con un bucle.</p>	

2.3 FUNCIONES Y PROCESOS

Funciones del RNA

En las primeras secciones del capítulo se describieron características bioquímicas y propiedades estructurales que denotan la versatilidad estructural del RNA. Esto, en conjunto con su capacidad informacional, les permite una gran diversidad funcional. Elliot & Ladomery (2011) presentan una clasificación en clases funcionales que se muestra en la tabla 2.5. Estas clases no pueden pretender ser exhaustivas ni mutuamente excluyentes ya que cualquier clasificación funcional de los RNAs sería en su mayoría producto de agrupaciones artificiales. Sin embargo, conocer el repertorio de funciones es útil para caracterizar clasificaciones naturales, para explorar preguntas evolutivas o incluso para el diseño sintético de RNAs.

Función	Descripción	Ejemplos
Almacenamiento de información genética	El RNA puede tener funciones similares a las del DNA. Algunos virus tienen genomas de RNA. Además, se cree que el RNA fue el principal material genético en la evolución temprana de la vida.	- Retrovirus - Hipótesis del mundo del RNA
Portador de información genética	El mRNA funciona como intermediario de información entre el DNA y la maquinaria de síntesis proteica.	- mRNA
Reconocimiento	Por medio de interacciones de apareamiento específicas, el RNA puede reconocer otras moléculas de RNA o de DNA.	- Interacción codón-anticodón entre mRNA y tRNA - Reconocimiento de la secuencia Shine-Dalgarno por el rRNA 16S - Reconocimiento de mRNAs específicos por miRNAs - Reconocimiento de rRNAs por snoRNAs
Molde	Permite que se inicie síntesis de DNA en regiones específicas.	- RNA telomerasa: Como molde para hacer nuevos extremos de DNA en los cromosomas - Iniciadores
Catálisis	Reacciones químicas catalizadas por RNAs	- Ribozimas
Andamiaje	Permite formar grandes complejos ribonucleoproteicos	- SRP y el ribosoma - Complejo de <i>splicing</i>
Estructural	Las moléculas de RNA suelen tener formas específicas que les permiten interactuar con otros RNAs o proteínas	- Interacción entre las dos subunidades del ribosoma - Interacción entre aptámeros de RNA y sus ligandos

Algunos procesos

Las funciones mencionadas anteriormente están incluidas en redes de procesos complejos entre DNA, RNA y proteínas. En 1958 Crick enunció la primera propuesta sobre los flujos de información relacionando estos tres tipos de moléculas y posteriormente publicó el artículo titulado “Central Dogma of Molecular Biology” (Crick, 1970) en el que por un lado aclaraba su controversial uso de la palabra dogma y por otro hacía una revisión actualizada sobre su propuesta original. (Shapiro, 2009) Es a partir de este Dogma central de la biología molecular que tradicionalmente se han entendido los procesos del RNA, sin embargo, sabemos ahora que no contempla un gran número de procesos asociados al RNA. La figura 2.7 muestra los flujos informacionales revisados en Crick (1970). A continuación se describen los procesos más relevantes en los que participa el RNA:

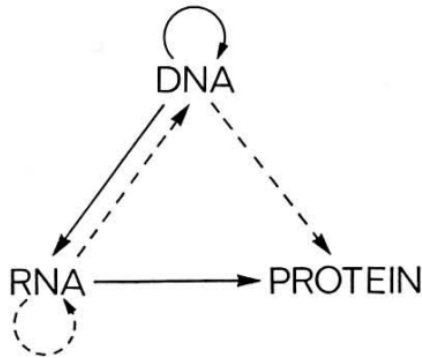


FIGURA 2.7 Flujos informacionales propuestos en el dogma central de la biología molecular. [Reproducida de (Crick, 1970)]

Transcripción El proceso en el que un sistema de enzimas convierte la información de una hebra de DNA en un RNA con una secuencia complementaria. Cuando el RNA transcrito codifica para una secuencia de aminoácidos se le llama mRNA, de lo contrario se le nombra ncRNA. Existen porciones del DNA que nunca son transcritas, entonces al grupo de RNAs transcritos en una célula se le conoce como transcriptoma. El RNA es sintetizado en este proceso por enzimas llamadas RNA polimerasas dependientes de DNA, que van agregando ribonucleótidos en el extremo 3', operando por un mecanismo de apareamiento de bases. Esas enzimas reconocen regiones específicas llamadas promotores a partir de los cuales comienzan la transcripción y una vez que ha finalizado la región a transcribir se detienen en terminadores, que son regiones que

por diversos mecanismos, ocasionan que la polimerasa se detenga. (Nelson & Cox, 2008)

Traducción Es el proceso por el cual la información en un mRNA especifica la secuencia de aminoácidos durante la síntesis proteica. En los mRNAs un triplete de nucleótidos que codifica para un aminoácido específico se llama codón y éstos se leen sin traslaparse a partir de un marco de lectura en el que se identifica el primer codón de la secuencia. El mRNA se une al ribosoma (que contiene rRNA) y se forma un complejo de inicio que permite la subsecuente elongación del péptido. En este proceso los tRNAs se van uniendo por apareamiento codón-anticodón con el mensajero y por medio de complejos formados con el ribosoma, los aminoácidos se van enlazando covalentemente entre sí formando la cadena polipeptídica hasta llegar al codón de terminación. (Nelson & Cox, 2008)

Transcripción reversa La enzima transcriptasa reversa, que es una DNA polimerasa dependiente de RNA, permite el flujo de información de RNA a DNA. La actividad de transcriptasa reversa fue descubierta en virus. Los grupos de virus III, IV, V y VI, según la clasificación de Baltimore, poseen genomas de RNA, y el grupo VI, que incluye a los retrovirus, requiere de la transcriptasa reversa para sintetizar DNA que luego es transcrito a mRNA (Baltimore, 1971). La actividad de transcriptasa reversa se ha encontrado también presente en procariontes y eucariontes. Los análisis genómicos han revelado su importancia en la evolución de los genomas: un tercio del genoma humano consiste en DNA transcrito de RNA. (Shapiro, 2009)

Procesamiento postranscripcional La transcripción no suele generar productos de RNA listos para continuar con su función final. Estos transcritos primarios son modificados por mecanismos complejos y específicos para cada tipo de RNA. Los transcritos primarios suelen ser denotados con el prefijo pre- seguido del tipo de RNA al que pertenecerán, pre-mRNA, por ejemplo. A continuación se describen algunos:

El proceso más ubicuo es el *splicing* (o corte y empalme), que elimina los intrones del transcrito primario. Sucede en muchos tipos de RNAs por medio de distintos mecanismos: (1) El RNA puede ser autocatalítico y cortarse a sí mismo: esto sucede en los intrones del grupo I (que incluye algunos pre- rRNAs, mRNAs y tRNAs) y del grupo II (que generalmente son pre-mRNA de mitocondrias y cloroplastos.) (2) Puede estar mediado por un complejo llamado complejo de *splicing* formado por snRNPs, (small nuclear ribonucleoproteins, en español ribonucleoproteínas nucleares pequeñas) que son complejos de proteínas y snRNA

(small nuclear RNA). Y (3) El *splicing* tipo tRNA que involucra endonucleasas y uso de ATP y sucede sólo en algunos tRNAs. (Nelson & Cox, 2008)

En los pre-mRNAs también pueden ocurrir procesos que modifican los extremos 5' y 3' del transcrito, particularmente en eucariontes. El capping-5' es un residuo de 7-metilguanosina que protege al mRNA de las ribonucleasas y se forma durante las primeras etapas de la traducción. La poliadenilación es un proceso que agrega una cola de poli(A) al extremo 3' que le permite unirse a proteínas específicas para procesos de transporte y para la traducción (Nelson & Cox, 2008)

El procesamiento de RNA ribosomal involucra segmentación de pre-rRNAs largos y modificaciones en algunas bases. En las bacterias los rRNAs 16S, 23S, 5S e incluso algunos tRNAs, se obtienen de un solo precursor que se segmenta y modifica. Las modificaciones en las bases suelen ser ejecutadas por enzimas particulares, mientras que la segmentación se lleva a cabo por ribozimas como RNasa III, RNasa P, RNasa E, etc. En eucariontes ocurre un proceso similar que también puede involucrar *splicing*, pero todo en un proceso acoplado en el nucléolo. Las modificaciones están dadas por complejos llamados snoRNPs que tienen proteínas y RNAs llamados snoRNAs (small nucleolar RNA). Éstos guían la modificación de las bases y algunas reacciones de segmentación. (Nelson & Cox, 2008)

El procesamiento de RNA de transferencia acorta un precursor largo por medio de remoción enzimática de los extremos 5' y 3', llevada a cabo por enzimas como RNasa P y RNasa D, respectivamente. En algunos tRNAs de eucariontes también debe ocurrir *splicing* para eliminar los intrones, así como la adición de los aminoácidos por la enzima nucleotidiltransferasa. (Nelson & Cox, 2008)

Regulación por RNAs La regulación de la expresión genética es controlada por elementos reguladores que interaccionan con una secuencia de DNA o RNA en alguna etapa previa a la síntesis proteica. Así como la proteínas suelen operar con cambios alostéricos, los RNAs usan cambios en su estructura secundaria que resultan en modificaciones de sus interacciones inter o intramoleculares. (Lewin, 2011) Existe un gran número de clases de RNA con funciones regulatorias. Se les suele agrupar como RNAi (RNA de inhibición), pero cada subgrupo recibe además su nombre exclusivo. Tal es el caso de los miRNAs (micro RNAs), siRNAs (small inhibitory , pequeños inhibidores o silenciadores), rasiRNAs (repeat-associated silencing, de silenciamiento asociados a repeticiones), piwiRNAs (asociados a proteínas piwi), asRNA (anti sentido), etc. (Shapiro, 2009) (Una lista larga de abreviaciones utilizadas para RNA se puede encontrar en el Apéndice A. Esta

nomenclatura es problemática, por lo que se discute en la siguiente sección.) A continuación se describen brevemente dos tipos de RNAs reguladores:

Los riboswitches son segmentos de regulación en cis comúnmente encontrados en bacterias y en menor cantidad en algunos eucariontes. Son un tipo de RNA catalítico cuya actividad depende de ligandos pequeños que modifican el plegamiento del RNA y de esta manera regula la transcripción de un mRNA que se encuentra subsecuentemente unido a esta secuencia. Puede suceder por diversos mecanismos: el ligando puede promover la terminación o iniciación de la transcripción, puede controlar el *splicing*, interferir en la transcripción, promover la autocatálisis, etc. (Breaker, 2012; Lewin, 2011)

Los micro RNAs (miRNAs) son silenciadores de genes que al interactuar con los mRNAs, comúnmente en la región 3'UTR, inhiben su traducción u ocasionan su degradación. Se han encontrado en muchos grupos de eucariontes y se ha encontrado que también participan en protección contra RNAs virales y en el control de los transposones. (Lewin, 2011)

Retrotransposición Los transposones son secuencias definidas del genoma que se pueden transportar a otras ubicaciones en el genoma. Los retrotransposones son una clase de estas secuencias que involucra obligatoriamente un intermediario de RNA y por lo tanto requieren la actividad de transcriptasas reversas. Se encuentran principalmente en eucariontes y son una causa importante de variación genética.

2.4 FAMILIAS DE RNA NO CODIFICANTE

Al terminar la década de los 50s, el RNA ya se contemplaba en al menos tres variedades: rRNA, tRNA y mRNA. Sin embargo, como la mayoría de los esfuerzos de la época se centraban en entender el código genético, poco se hizo para estudiar la fracción de RNA que no pertenecía a alguna de estas tres categorías. (Eddy, 2001) El resto de los grupos de RNA eran considerados en el momento como complejos contenedores de moléculas poco abundantes e inestables. A partir de esto se vuelve evidente la razón por la cual los descubrimientos sobre el resto de los RNAs funcionales han sido por lo general aislados y los esfuerzos para encontrarlos y clasificarlos son bastante recientes.

Como se explicó en la sección anterior, hay un gran número de “tipos” de RNAs, pero la mayoría de estos grupos son artificiales en términos filogenéticos o agrupan en niveles completamente distintos entre sí (ver Nota sobre la nomenclatura al final de esta sección). Los RNAs se suelen agrupar de distintas maneras con respecto a su función, pero quizás la manera más amplia de hacerlo es separando en codificantes y no codificantes. Los RNAs codificantes se refieren a los RNAs

mensajeros (mRNA) que tienen un ORF que corresponde a una proteína que es traducida. Los RNAs no codificantes (abreviado ncRNA, también npcRNA, nmRNA o fRNA) son aquellos que a partir de genes de RNA, producen transcritos que por sí mismos tienen funciones estructurales, catalíticas o de regulación.

La tendencia actual busca agrupar a todos aquellos RNAs **no** codificantes para proteínas en familias inferidas a partir de análisis filogenéticos. La base de datos Rfam hace precisamente esto. La versión 12.0 fue la última publicada, incluyendo 2450 familias distintas (E. P. Nawrocki et al., 2014) y actualmente alcanza 2544 familias.

Familias en Rfam

Rfam es una base de datos de familias de RNA no codificantes (ncRNA). Pertenece a un proyecto más grande llamado Xfam, que incluye los proyectos Pfam para familias de proteínas, Treefam para árboles filogenéticos, entre otros. Xfam es actualmente mantenido por el European Bioinformatics Institute que es parte del European Molecular Biology Laboratory. Es de acceso libre y la documentación de las familias se hace vía Wikipedia por medio del WikiProjectRNA. ("WikiProjectRNA", 2016)

Las familias se forman por (1) una alineación múltiple de secuencias representativas de la familia, depurada y revisada manualmente (llamada alineación semilla), (2) la estructura secundaria consenso y (3) un modelo de covarianza construido a partir de la alineación, construido con el programa Infernal (Ver sección 3.2) (Eric P Nawrocki & Eddy, 2013). Con esto se buscan secuencias homólogas putativas en la base European Nucleotide Archive (ENA) también con el programa Infernal, y éstas reciben una puntuación que indica qué tan bien se ajusta la secuencia al modelo de covarianza. Los revisores establecen un umbral para delimitar el puntaje con el que una secuencia puede formar parte de la familia considerando información taxonómica y conocimiento experto. En la base se reporta además el "tipo" de RNA de la familia (integrando así las clasificaciones artificiales antes mencionadas), la literatura referente a cada familia, asociaciones con otras bases de datos y el artículo de Wikipedia. (E. P. Nawrocki et al., 2014) Las familias a su vez están agrupadas en clanes. Las familias en un mismo clan comparten claramente un ancestro común pero divergen demasiado para ser razonablemente alineadas o difieren evidentemente en función. (Gardner et al., 2011) Las familias poseen códigos de acceso con el prefijo *RF*, iniciando en RF00001, en donde los últimos 5 dígitos se asignan consecutiva y crecientemente a las familias agregadas a la base de datos, hasta ahora en la RF02544. Los códigos

de acceso para los clanes inician con *CL* y actualmente se extienden del CL00001 al CL00112. (“Rfam 12.0” EMBL-EBI, 2016)

Distribución de las familias

La hipótesis de el RNA como una molécula primordial en el origen de la vida comenzó a ser propuesta a finales de los 60's por personajes como Woese, Crick y Orgel, pero fue tras el descubrimiento de algunas capacidades catalíticas de esta molécula durante los 80s (Cech, 2011) que se propone la hipótesis del mundo de RNA. (Gilbert, 1986). El mundo del RNA se refiere al estado de evolución temprana de la vida en el que el RNA era tanto la molécula acarreadora de la información genética como el principal catalizador biológico. Actualmente ésta es la hipótesis imperante y se ha alimentado de la gran diversidad funcional de RNAs descubiertos posteriormente. Una idea central para la hipótesis del mundo de RNA es el principio de continuidad, que establece que los sistemas modernos son producto de la evolución gradual de estados anteriores y como consecuencia es posible que algunas familias de RNA sean descendientes directas de moléculas que evolucionaron durante el mundo del RNA. (Hoepfner et al., 2012)

Proponer filogenias de las familias no es tarea sencilla. La corta longitud de los RNAs y el bajo nivel de conservación de la secuencia no permite análisis filogenéticos robustos con los métodos tradicionales y el análisis de distribución no es suficiente para inferirlos, pues la amplia distribución de algunas familias, ocasionada por transferencia horizontal, puede sugerir falsamente antigüedad. Sin embargo, los estudios comparativos entre familias de RNA sí pueden ayuda a formar hipótesis evolutivas sobre el mundo del RNA, el principio de continuidad y revelar patrones generales de las familias. Este es el caso del estudio realizado por Hoepfner, et al. (2012) utilizando precisamente la base Rfam y concluyendo los siguientes puntos:

99% de las familias de RNA están restringidas a un solo dominio Sólo siete familias se encuentran presentes en los tres dominios. La mayoría de las familias son exclusivas de algún dominio y aquellas que comparte más de un grupo podrían deberse tanto a la evolución vertical desde un ancestro común como a la transferencia horizontal entre dominios. La figura 2.8 es un diagrama de Venn reportando la distribución de las familias para Rfam 10.0.

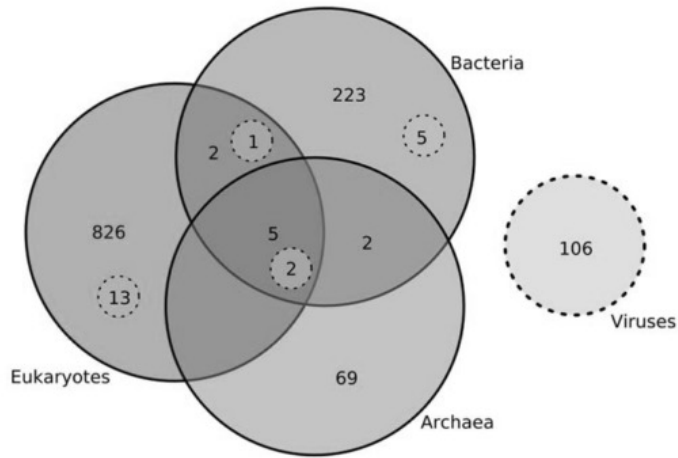


FIGURA 2.8 Diagrama de Venn de la distribución de las familias Número de secuencias a partir de información taxonómica de Rfam [Reproducida de (Hoepfner et al., 2012)]

Las familias presentes en más de un dominio muestran una mezcla de herencia vertical y horizontal De los 12 grupos presentes en más de un dominio, a cinco se les puede atribuir un patrón de herencia vertical, basándose en estudios previos y en su distribución. Todos son RNAs involucrados en la traducción y exportación de proteínas (figura 2.9), lo que reafirma la hipótesis de que son RNAs antiguos, rastreables al *último antepasado común universal* (*LUCA* por sus siglas en inglés, *last universal common ancestor*). Estos grupos son: tRNA (CL00001), rRNA 5S (RF00001), RNA SSU dominio 5' (RF00177), RNasa P (CL00002) y SRP (CL00003). Cuatro grupos muestran evidencia de transferencia horizontal: intrones catalíticos del grupo I (RF00028), Intrones catalíticos del grupo II (RF00029), IsrR (RF01419), y el Pseudonudo del dominio G12 del rRNA 23S (RF01118). Se sospecha la herencia vertical para el riboswitch TPP (RF00059), aunque no está demostrada. Finalmente, aunque la mayoría de familias de CRISPR crRNA se encuentran en un solo dominio, dos clanes CRISPR-1 (CL00014) y CRISPR-2 (CL00015) se encuentran en genomas de Bacteria y Archaea, lo que sugiere transmisión horizontal. Lo anterior es compatible con las hipótesis que proponen que existe transferencia horizontal limitada del sistema CRISPR/Cas.

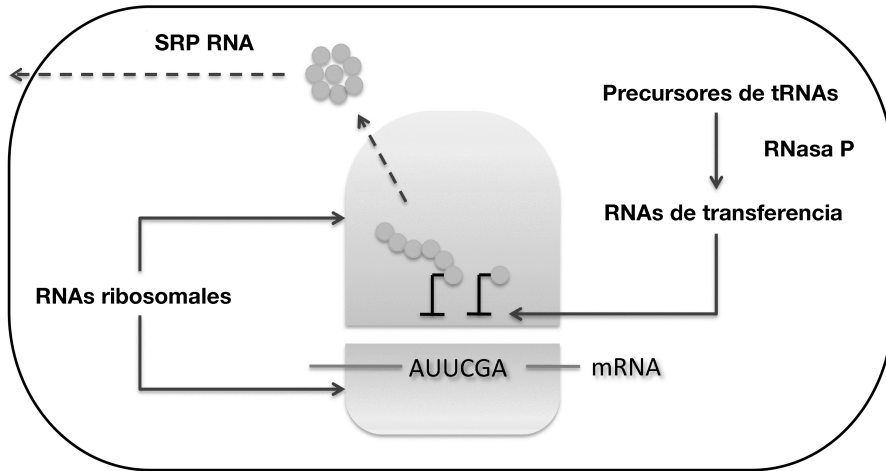


FIGURA 2.9 Procesos de RNA rastreables a LUCA Todas las familias antiguas de RNA que muestran evidencia de herencia vertical están relacionadas con traducción de proteínas (rRNAs, tRNAs, RNasa P) o con su exportación (SRP RNA).

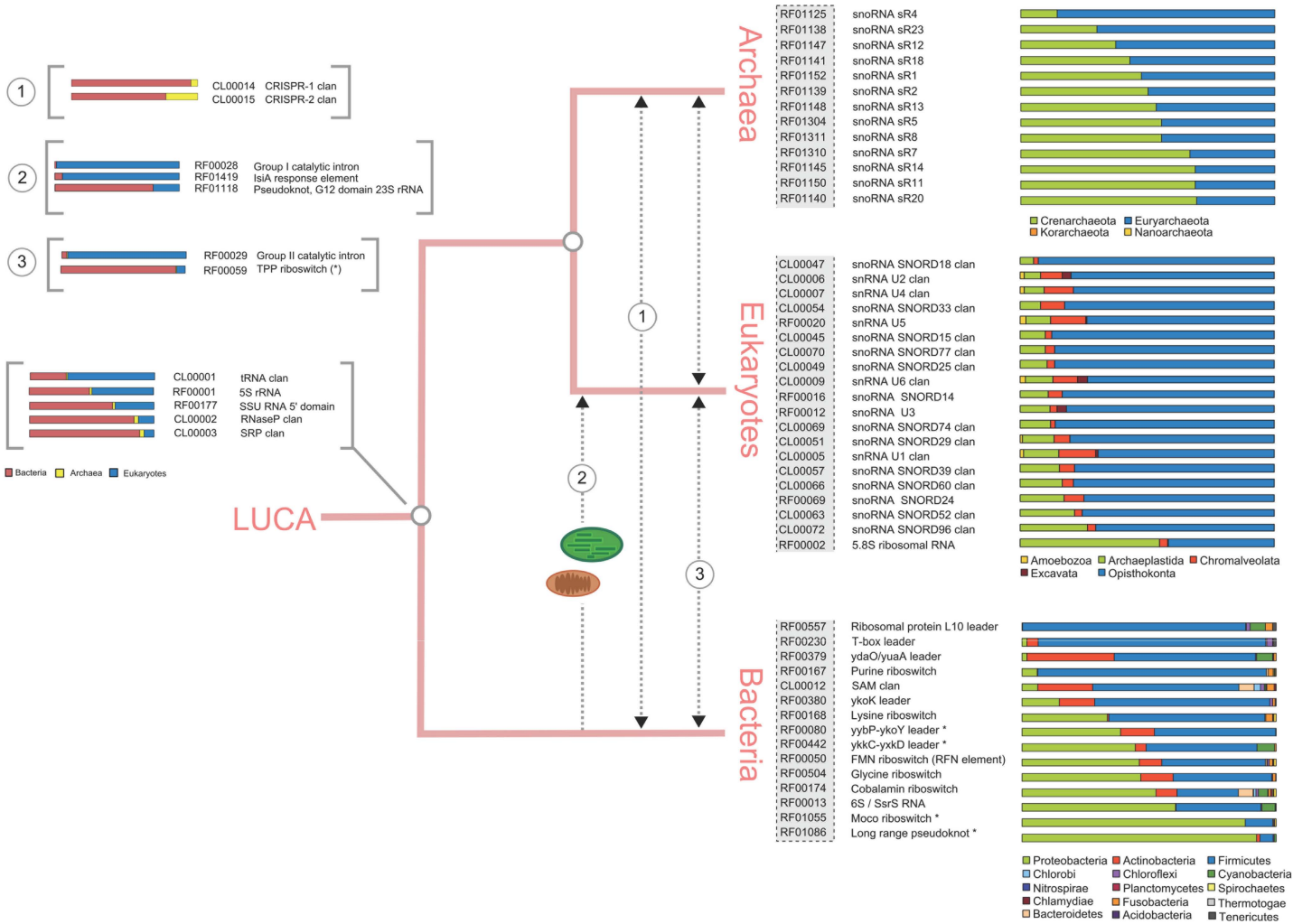
Pocas de las familias exclusivas para un dominio se distribuyen ampliamente El análisis que reveló sólo el 18.8% de los grupos de Archaea, 6.7% de Bacteria y 2.4% de Eukarya están distribuidos ampliamente dentro del dominio y que además éstos suelen ser RNAs regulatorios. También se encontró que los repertorios de Archaea y Eukarya están dominados por snoRNAs.

Los patrones descritos anteriormente podrían deberse a sesgos del muestreo, que es algo común en los datos genómicos. Sin embargo, los datos actuales sugieren que hay una alta proporción de RNAs evolutivamente jóvenes. Aunque sea común discutir la relevancia de las familias de RNA actuales con respecto al mundo del RNA, la evidencia para la hipótesis de continuidad para la mayoría de las familias de RNA es bastante escasa. (Hoepfner et al., 2012) Por otro lado, analizar las características de aquellas familias que sí están ampliamente distribuidas podría revelar patrones evolutivos relevantes.

NOTA SOBRE LA NOMENCLATURA

Hasta el momento no existe nomenclatura estandarizada en las abreviaciones usadas para describir los RNAs. Una lista larga de abreviaciones comúnmente encontradas se recopila en el Apéndice A, que si bien no aclara los problemas producto de esta falta de estandarización, pretende servir como guía al encontrarse con alguna abreviación.

FIGURA 2.10 Reconstrucción de repertorios de RNA distribuidos ampliamente. Las barras representan la abundancia taxonómica por grupo taxonómico en la familia. También se muestran con líneas punteadas las hipótesis de transferencia horizontal. [Reproducida de (Hoepfner et al., 2012)]



CAPÍTULO 3

PREDICCIÓN DE ESTRUCTURA SECUNDARIA DEL RNA

Resolver estructuras secundarias de RNA confiablemente es quizás el objetivo más importante de los estudios estructurales en torno a esta molécula, debido a la íntima relación entre la forma y la función del RNA no codificante. Lograrlo por medio de la computación permitiría elucidar funciones, explicar mecanismos e incluso diseñar RNAs sintéticos, con costos significativamente más bajos y a mayor velocidad que cualquier método experimental. A continuación se revisan algunos de los métodos desarrollados con este objetivo:

3.1 DEFINICIÓN FORMAL DE LA ESTRUCTURA SECUNDARIA

En la sección 2 del segundo capítulo se discutió la definición de estructura secundaria desde una perspectiva biológica y molecular, sin embargo, para desarrollar e interpretar un programa de predicción de estructura secundaria es necesaria una definición formal. La estructura secundaria del RNA puede definirse como un grafo etiquetado en el que los nodos del grafo representan las bases (en las posiciones $i=1, \dots, n$) en la secuencia de la molécula, donde la posición 1 corresponde al extremo 5'; y el conjunto de aristas contiene dos tipos de aristas: las aristas entre nodos consecutivos $(i, i+1)$ que representan el esqueleto de fosfato y las aristas de apareamiento, que se dan entre los pares de bases y representan los puentes de hidrógeno. Estas últimas deben cumplir las siguientes condiciones:

1. Se forman sólo por apareamientos Watson-Crick (A-U y C-G) y apareamiento G•U por balanceo.
2. Un nodo puede tener máximo una arista de apareamiento (la estructura secundaria es, en términos de teoría de grafos, un apareamiento o matching en inglés.)
3. Las aristas de apareamiento deben formarse entre nodos con al menos tres nodos libres entre ellos.

4. Las aristas de apareamiento no pueden cruzarse. Significa que los grafos formados deben ser planares exteriores. En términos biológicos implica que no se forman pseudonudos.

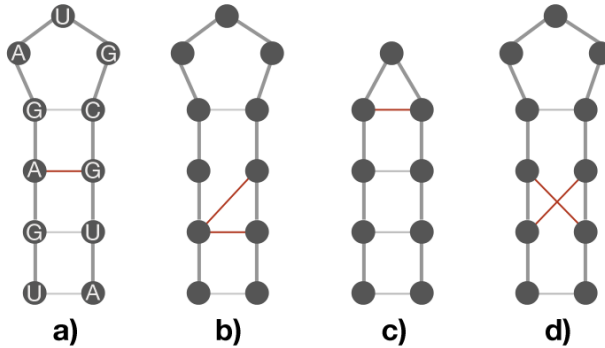


FIGURA 3.1 Condiciones no permitidas para las aristas de apareamiento. Las aristas en rojo son prohibidas por alguna de las cuatro condiciones: a) Ejemplo prohibido por la condición 1, en donde ocurre un apareamiento distinto a los permitidos; b) Ejemplo de arista prohibido por la condición 2, en el que surgen dos aristas de un mismo nodo; c) Ejemplo prohibido por la condición 3, en el que la arista se conecta con nodos a sólo una base de distancia; d) Ejemplo prohibido por la condición 4, en el que los apareamientos se cruzan.

Representaciones

Puesto que esta definición formal de la estructura secundaria se refiere al patrón de contactos entre pares de bases, debe ser entendida como un concepto topológico y no tal cual como una estructura bidimensional de la molécula. (Fontana, 2002) Las representaciones que se muestran en la figura 3.2 son equivalentes y suelen ser usadas en distintos contextos, dependiendo lo que se quiera mostrar con el diagrama. El programa ViennaRNA ha usado consistentemente una *de puntos y paréntesis* (*dot-bracket notation* en inglés, también llamada ViennaRNA dot-bracket notation e identificada por la extensión .dbn) en donde los puntos corresponden a bases no apareadas y los paréntesis a bases apareadas (que se anidan y se cierran con su pareja correspondiente) mostrada en la figura 3.2a. La representación lineal (figura 3.2b) muestra el esqueleto de fosfato en una línea recta y las bases apareadas unidas por arcos. La representación circular (figura 3.2c) coloca el esqueleto de fosfato en una circunferencia y los apareamientos se representan por líneas rectas dentro del círculo y finalmente la representación de grafo plano (figura 3.2d) que usa distancias predeterminadas entre las bases vecinas apareadas y que es útil para visualizar los componentes estructurales de la molécula. (Darty, Denise, & Ponty, 2009)

Además, estas representaciones pueden tener variantes para dibujar las aristas, de tal manera que muestren el tipo de apareamiento con las clasificaciones que se discuten en la sección 2.1. En esta tesis se usa un dibujo de línea simple para todos los apareamientos.

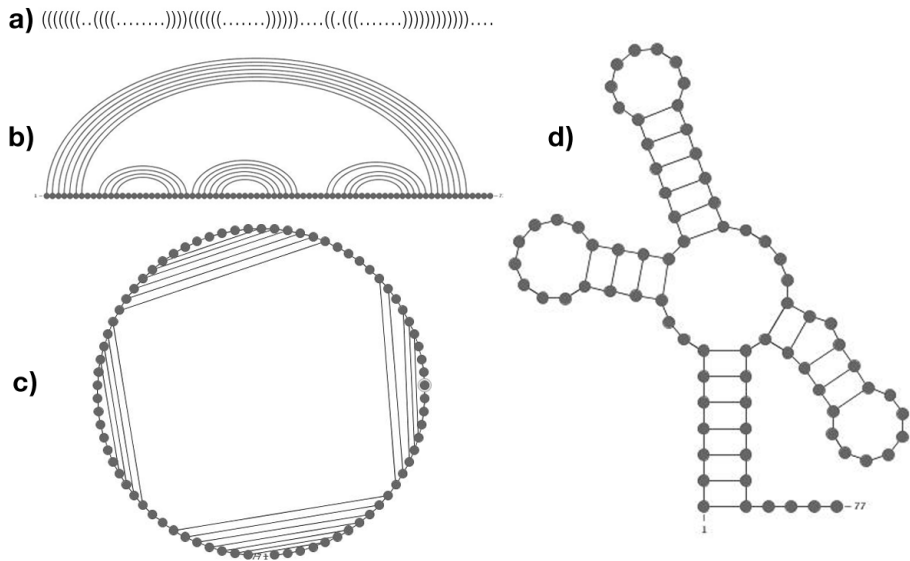


FIGURA 3.2 Representaciones de la estructura secundaria del RNA Todas las figuras muestran la misma molécula de RNA (tRNA^{Ala} de *Homo sapiens* [AC013472] (Cannone et al., 2002)), generadas con VARNA (Darty et al., 2009) a) Notación de puntos y paréntesis, b) Representación lineal c) Representación circular y d) Representación de grafo plano

Elementos estructurales

La figura 3.4 muestra una molécula de RNA con una representación como grafo plano, señalando los elementos estructurales ahí presentes. A pesar de que en la sección 2.2 los elementos estructurales se representaron como grafos, se discutieron únicamente como interacciones inherentes de la molécula y es importante aclarar que están definidos principalmente de manera formal. Todo esto debe ser interpretado con cautela dado que la definición formal y la definición experimental se han retroalimentado mutuamente y ya es imposible separarlas claramente.

Las definiciones formales de los elementos estructurales fueron introducidas como parte de uno de los primeros modelos para predecir el plegamiento de los RNAs, elaborado por Tinoco y sus colaboradores. Este modelo calcula la energía libre de una estructura secundaria al sumar las energías libres de los distintos tipos de

apareamientos y bucles con estimaciones derivadas medidas experimentales. (Tinoco et al, 1971; Tinoco et al., 1973) y sirvió como base para algoritmos posteriores más sofisticados. La figura 3.3 pertenece al primer artículo en el que se publicó este modelo (Tinoco et al., 1971) y muestra la contribución de cada elemento estructural a la estabilidad de la estructura secundaria.

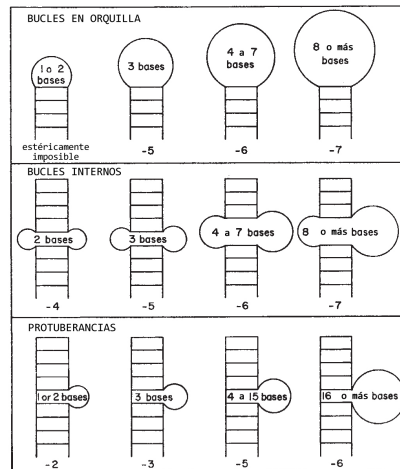


FIGURA 3.3 Contribución de bucles y protuberancias a la estabilidad de la estructura secundaria Un número positivo indica mayor estabilidad. [Modificado de (Tinoco et al., 1971)]

3.2 ALGORITMOS DE PLEGAMIENTO

La predicción de estructuras secundarias es un problema interesante tanto desde una perspectiva biológica, como desde una perspectiva informática. Estos esfuerzos comenzaron con modelos de energía en los 70s (Tinoco et al., 1971) y continúan en la actualidad con el desarrollo de programas que se apoyan en combinaciones de herramientas para llegar a mejores predicciones. Los métodos se pueden separar en tres grandes grupos: métodos de combinatoria, métodos comparativos y métodos de cinética de plegamiento. (Miklós, 2010) Históricamente surgen en ese orden, pero todos tienen versiones revisadas recientes y la tendencia actual apunta hacia la combinación de métodos. A continuación se revisan algunos:

Métodos de combinatoria

En esta categoría se encuentran los métodos de maximizar el número de bases apareadas o aquellos para encontrar la energía mínima. Los métodos de

combinatoria están basados en una función de puntuación (cuya definición depende de la hipótesis de trabajo) y en encontrar, entre las estructuras posibles, aquella que minimice o maximice esta función.

Algoritmo de Nussinov (1980) Es un algoritmo de programación dinámica que maximiza el número de bases apareadas. Es una solución interesante en términos algorítmicos y sirvió como base para otros algoritmos más sofisticados, sin embargo falla para predecir la estructura más estable puesto que no toma en cuenta el apilamiento de bases ni la entropía de distintos bucles, ambas características que contribuyen significativamente a la energía libre de la estructura secundaria. (Miklós, 2010)

Algoritmo Zuker-Sankoff (1984) Este algoritmo se basa en el modelo energético de Tinoco, et al. (1973, 1971) para encontrar la estructura secundaria de mínima energía libre. El modelo energético establece que la energía libre de una estructura secundaria es la suma de las energías libres de los distintos bucles. Se basa en separar los elementos estructurales (de los cuales se han medido energías de manera experimental) para encontrar una estructura óptima. (Miklós, 2010; Wake, 2003)

Algoritmo McCaskill (1990) Este algoritmo responde a la hipótesis biológica de que las secuencias de RNA pueden cambiar su estructura secundaria dinámicamente. Es decir, que no hay un único estado restringido al mínimo de energía sino que hay un conjunto de estructuras secundarias posibles. El algoritmo McCaskill calcula la función de partición de la distribución de Boltzmann (ver sección 4.1) para predecir la probabilidad de ocurrencia de cada una de las estructuras y como extensión, la probabilidad de apareamiento de cada pareja dada. (Miklós, 2010; Wake, 2003)

El problema de los pseudonudos Los algoritmos de combinatoria utilizan programación dinámica, lo que quiere decir que buscan soluciones óptimas usando soluciones de subproblemas. Esto se logra con algoritmos recursivos en donde se calculan primero los óptimos de subsecuencias. (Eddy & Hughes, 2004) Dado que los pseudonudos son apareamientos no anidados entre un bucle y residuos fuera de ese elemento, ocurren cuando la siguiente condición formal no se cumple: (donde las minúsculas i, j, k y l son nodos y P es un apareamiento): si $i < j$ y $k < l$ están en P , entonces $i < k < j$ implica que $i < l < j$ (Ver figura 3.1c). violando la definición recursiva del algoritmo. (Ancel & Fontana, 2000) Para predecir estas estructuras, la recursión no es útil y se requeriría calcular todas las combinaciones posibles de apareamientos de una secuencia, lo cual lo hace un problema np -

complejo. Por esta razón, los pseudonudos no suelen ser contemplados en este tipo de algoritmos. (Miklós, 2010)

Métodos comparativos

Los métodos comparativos se basan en una hipótesis evolutiva: que la estructura se conserva más que la secuencia. Se asume que las secuencias homólogas tienen la misma estructura, pues si ésta está relacionada con la función, las bases coevolucionarían para conservar la estructura secundaria. Estos métodos comparan diversas secuencias homólogas para encontrar una estructura que todas puedan adoptar.

Gramática Knudsen-Hein (1999) Este es un algoritmo basado en una gramática formal, específicamente una gramática probabilística libre de contexto (SCFG por sus siglas en inglés) que es un tipo de gramática transformacional. (Los detalles de lo que esto significa son extensos, pero se pueden encontrar resumidos en Miklós (2010)) El algoritmo genera la estructura secundaria común de un grupo de secuencias de RNA alineadas. De manera muy general, consta de una serie de reglas de reescritura que dependen de parámetros obtenidos de estructuras ya conocidas, de modelos de Markov que describen las sustituciones de los nucleótidos y de árboles de máxima verosimilitud de los RNAs alineados.

Modelos de covarianza (1994) Son una formulación especial de las SCFG para generar modelos estadísticos de alineaciones múltiples de RNA anotadas estructuralmente. Estos capturan la conservación y probabilidad de los residuos en cierta posición. Las posiciones de los apareamientos son dependientes una de la otra lo que permite generar un perfil de covarianza. Los modelos de covarianza son particularmente útiles para identificar patrones homólogos de plegamiento. (Miklós, 2010; E. Nawrocki & Eddy, 2014)

Métodos de cinética de plegamiento

Estos métodos tratan de simular la cinética de plegamiento *in silico* para encontrar la estructura. Debido a que el plegamiento del RNA suele iniciar durante la transcripción, la estructura no necesariamente sería aquella con mínima energía libre. Gulyaev publicó su primer trabajo sobre simulación de vías de plegamiento de RNA en 1991 y fue posteriormente implementado en un programa llamado KineFold en el año 2000. (Miklós, 2010)

ViennaRNA Package

ViennaRNA Package consiste en una biblioteca gratuita de programas escritos en C para la predicción y el análisis de estructuras secundarias de RNA. (Lorenz et al., 2011) Implementa diversos algoritmos para la predicción de estructuras:

Predicción de estructura de energía mínima Es un algoritmo por combinatoria desarrollado por Zuker y colaboradores con el que se obtiene una sola estructura óptima. (Zuker & Stiegler, 1981) En el paquete **ViennaRNA** el programa **RNAfold** lee la secuencia de RNA, calcula la energía mínima (MFE por sus siglas en inglés, *minimum free energy*) y muestra la estructura de mínima energía en notación de paréntesis y puntos.

Algoritmo para la función de partición Calcula la probabilidad de apareamientos de bases en el ensamble termodinámico a partir del algoritmo de McCaskill. (McCaskill, 1990) También es calculado a través del programa **RNAfold**.

Algoritmo de plegamientos subóptimos Genera todas las estructuras subóptimas en un rango de energía dado con el algoritmo de Wuchty, Fontana, Hofacker, & Schuster (1999) Se obtiene a través del programa **RNAsubopt**.

El ViennaRNA Package ha sido utilizado en estudios bioinformáticos del RNA por al menos dos décadas y se encuentra actualmente es su versión 2.3. Cuenta con muchos otros programas y herramientas que no se discuten aquí, sin embargo, la documentación completa se puede encontrar en su sitio web. (Theoretical Biochemistry Group-Universität Wein, "ViennaRNA Package 2", 2016) o en los manuales RNALib (Ivo L. Hofacker, Stadler, Lorenz, *et al.*), 2014, 2016) Además cuenta con una interfaz generada por SWIG (que es un software que conecta programas escritos en C con otros lenguajes de programación de alto nivel) lo que permite usar las funciones del programa directamente en Perl y Python.

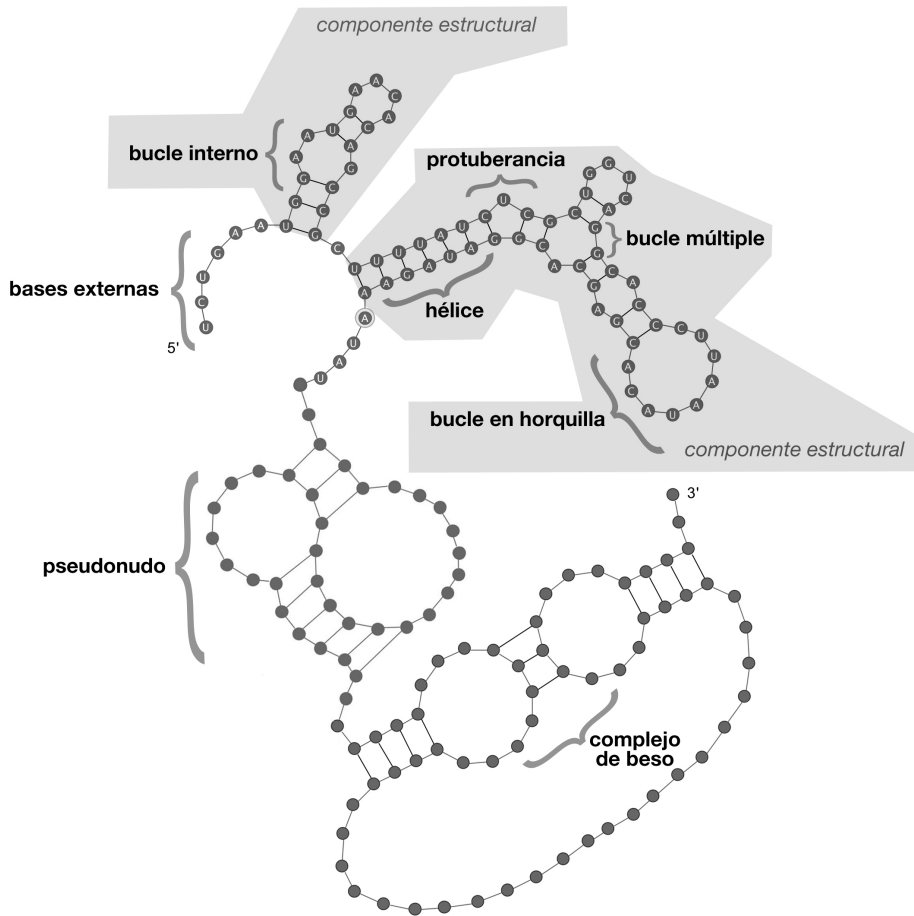


FIGURA 3.4 Grafo de estructura secundaria del RNA. Mostrando algunos los componentes estructurales [Creado con VARNA, Modificado de (Ancel & Fontana, 2000)].

CAPÍTULO 4

MEDIDAS DE PLASTICIDAD PARA EL RNA

Como se discute en el capítulo 1, llegar a una definición operacional (y en este caso formal) de plasticidad no es un asunto sencillo, por lo que un acercamiento conveniente es usar diferentes medidas que interpretadas en conjunto se acerquen a lo que entendemos por plasticidad. Ancel & Fontana (2000) definen algunas medidas cuantitativas para medir la plasticidad de una población de RNAs. Para este trabajo se eligieron cuatro de éstas: 1) Fracción de la función de partición, 2) probabilidad de la estructura de mínima energía, 3) varianza de las estructuras en $5kT$ y 4) número de estructuras en $5kT$. Todas estas involucran parámetros calculados con el ViennaRNA Package, aunque podrían calcularse con otros programas o ser adaptados para modelos equivalentes. A continuación se presentan los planteamientos formales relevantes al cálculo de las medidas de plasticidad según Ancel & Fontana (2000) y Fontana (2002). La figura 4.2 resume, con un ejemplo, los elementos relevantes para entender el proceso por el que se obtienen las cuatro medidas.

Entendemos plasticidad del RNA como la característica intrínseca de una secuencia para plegarse en estructuras alternativas. Asumiendo que una secuencia se equilibra entre distintas estructuras como consecuencia de pequeñas fluctuaciones energéticas, la plasticidad se mide entonces analizando las propiedades del grupo de estructuras accesibles en un rango de energía, Δ , de unos pocos kT (donde T es la temperatura absoluta fijada en 310.15 K y k es la constante de Boltzmann, $k=1.98717 \cdot 10^{-3}$ kcal/K). En este caso se elige $5kT$, o bien $\Delta=3.08$ kcal, lo que corresponde a la pérdida de dos interacciones de apilamiento G-C/C-G.

Para una secuencia S_j , consideramos todas las estructuras σ_i^j cuyas energías libres correspondientes son $E(\sigma_i^j)$. Las estructuras σ_i^j están ordenadas por energía creciente, por lo que el índice 0 se refiere a la estructura de energía mínima. El

grupo de estructuras denotado como $\sigma^j(\Delta)$ incluye las estructuras accesibles en un rango de energía $[E(\sigma_0^j), E(\sigma_0^j) + \Delta]$, y se nombra *repertorio plástico*.

4.1 PROBABILIDAD DE BOLTZMANN

Para estudiar el comportamiento de los grupos de estructuras de RNA se usan herramientas de *mecánica estadística*. La mecánica estadística es una rama de la física que estudia el comportamiento de sistemas de partículas usando teoría de la probabilidad. Se usa comúnmente para explicar la relación entre el comportamiento de los elementos microscópicos de un sistema y sus propiedades termodinámicas macroscópicas. (Stowe, 2007; Swendsen, 2012)

Para calcular la fracción de tiempo que una molécula pasa en una estructura σ , de entre todas las estructuras accesibles, se usa la *probabilidad de Boltzmann*. Ésta se obtiene de la distribución de Boltzmann que es una distribución de probabilidad que indica la probabilidad de que un sistema se encuentre en cierto estado en función de su temperatura y de la energía del estado. En equilibrio termodinámico, la probabilidad de Boltzmann para una estructura σ está dada por:

$$p(\sigma) = \frac{e^{-E(\sigma)/kT}}{Z}$$

en donde $-E(\sigma)$ (también a veces denotado ΔG_σ) es la energía libre de la estructura σ , k es la constante de Boltzmann, T es la temperatura absoluta y Z es la *función de partición* (también a veces denotada Q).

Este sistema específico se considera un *ensamble canónico*, pues es un sistema en contacto térmico con el ambiente, con una temperatura definida y en donde el volumen y el número de partículas del sistema es fijo. (Miklós, 2010) La *función de partición canónica* depende de los grados de libertad del sistema y la Z se calcula por medio del algoritmo de McCaskill. (McCaskill, 1990)

$$Z = \sum_i e^{\frac{-E(\sigma_i)}{kT}}$$

donde i es el índice de los microestados del sistema (las posibles estructuras que la secuencia de RNA puede adquirir) y se suma para todas.

El algoritmo usa programación dinámica parecida a la del algoritmo Zuker-Sankoff donde se van calculando funciones de partición parciales de los plegamientos

posibles de subsecuencias, por lo cual es posible calcular también la probabilidad de observar un apareamiento específico de bases. (Miklós, 2010; Wake, 2003)

4.2 DISTANCIA DE HAMMING

La distancia de Hamming se define como el número de sustituciones requeridas para convertir una cadena de caracteres en otra, o también, el número de posiciones en el que los símbolos correspondientes difieren. (Schuster & Fontana, 1998) Puede ser visualizada también con un cubo n -dimensional en el que n es la longitud de la cadena y la distancia de Hamming es el camino más corto entre los dos vértices.

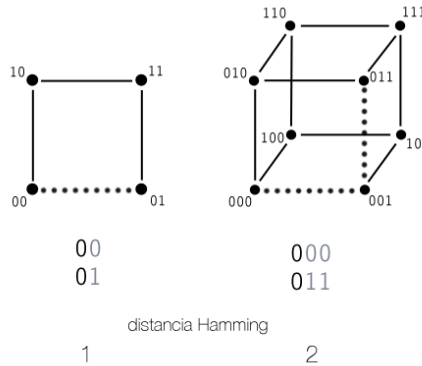


FIGURA 4.1
Representación
en cubo de la
distancia de
Hamming La línea
punteada
representa el
camino más corto
de un vértice a otro.

4.3 MEDIDAS

A partir de una secuencia, se obtiene entonces: la estructura de mínima energía, todas las estructuras subóptimas en un rango de energía de $5kT$, la energía libre correspondiente a cada una de las estructuras y la Z del ensamble. Es con esta información que se calculan las siguientes cuatro medidas de plasticidad:

Medida 1: Probabilidad de la estructura de mínima energía

Esta medida únicamente toma la probabilidad de Boltzmann para la estructura de mínima energía. Una probabilidad alta está relacionada con menor plasticidad por las mismas razones que la medida anterior; un valor alto indica que la secuencia tiende a mantenerse más tiempo en la estructura de mínima energía.

$$p(\sigma_0^j) = \frac{\exp(-E(\sigma_0^j)/kT)}{Z}$$

Medida 2: Número de estructuras en $5kT$

Se refiere al número de estructuras accesibles en $5kT$ y está denotada como $\langle |\sigma^j(\Delta)| \rangle$. Un mayor número de estructuras estaría relacionado con mayor plasticidad, pues indica que existen varios plegamientos alternativos con energías cercanas a la de la estructura de energía mínima.

Medida 3: Varianza de las estructuras en $5kT$

Esta medida es la única que toma en cuenta la forma de las estructuras secundarias predichas. Usando las representaciones de paréntesis y puntos (Ver sección 3.1) se calcula la distancia de Hamming entre cada una de las estructuras subóptimas $\sigma_i^j \in \sigma^j(\Delta)$ y la estructura de mínima energía σ_0^j (denotada como $d(\sigma_i^j, \sigma_0^j)$). Se calcula una varianza común como medida de dispersión de las estructuras de la siguiente manera:

$$\text{var } \sigma^j(\Delta) = \overline{d(\sigma_i^j, \sigma_0^j)^2} - \overline{d(\sigma_i^j, \sigma_0^j)}^2$$

Una mayor varianza está relacionada con mayor plasticidad, puesto que indica que las estructuras en la vecindad del mínimo de energía no son tan parecidas entre sí.

Medida 4: Fracción de la función de partición

Mide la probabilidad total de la ocurrencia de las estructuras accesibles en un rango de $5kT$ con respecto a todas las secuencias accesibles por la secuencia S_j . Está definida como z^j y se calcula de la siguiente manera:

$$z^j = \frac{1}{Z} \sum_{\sigma_i^j \in \sigma^j(\Delta)} \exp\left(\frac{-E(\sigma_i^j)}{kT}\right) = \sum_{\sigma_i^j \in \sigma^j(\Delta)} p(\sigma_i^j)$$

Esta medida indicaría menor plasticidad al acercarse a 1.0, puesto que representa que la secuencia tiende a pasar todo su tiempo en los estados vecinos a la estructura de mínima energía.

La relación general entre la medida propuesta y la definición de plasticidad se muestra en la figura 4.3. Cabe recordar que se trata de una generalización para guiar la interpretación y que justamente el objetivo de usar las cuatro medidas simultáneamente es tener una medida más comprensiva de la plasticidad.

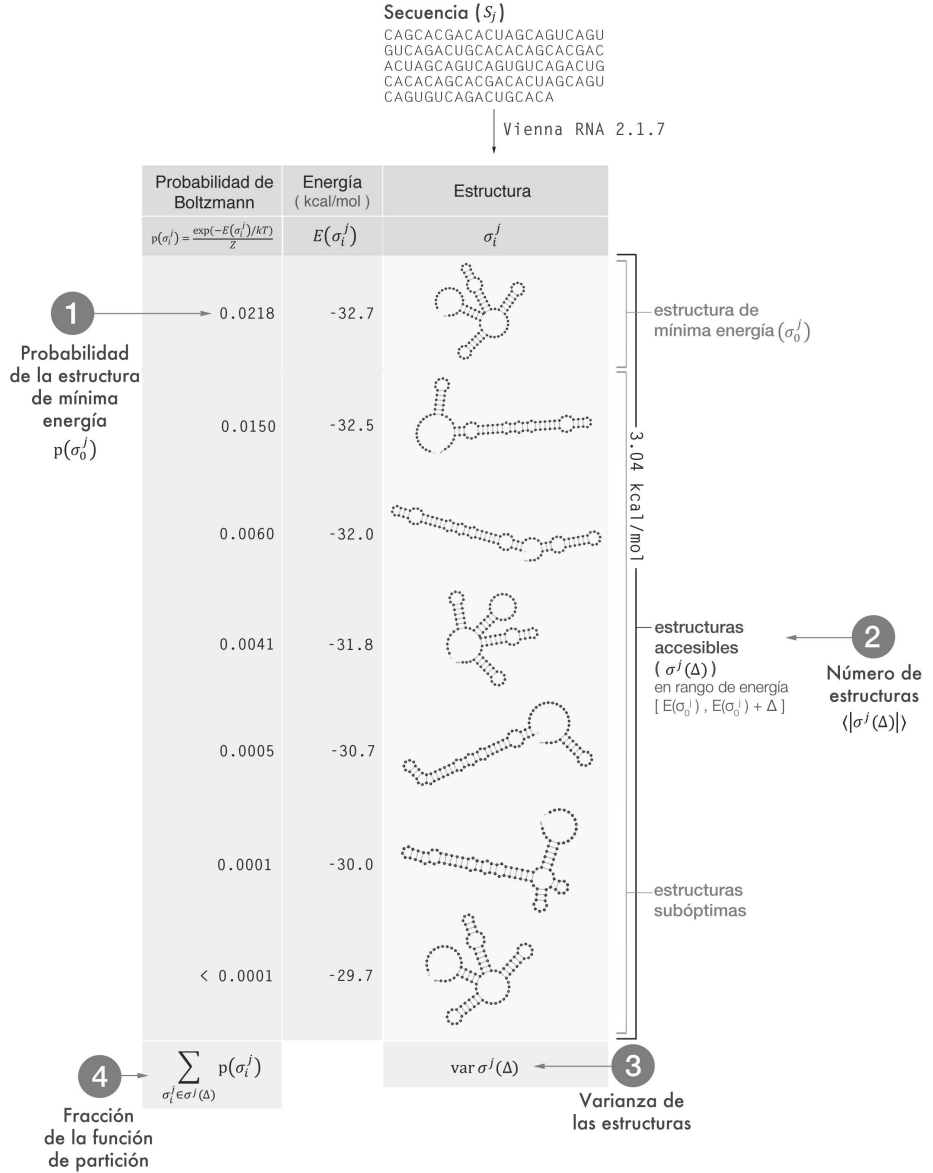


FIGURA 4.2 Medidas de plasticidad Se ilustra un ejemplo (reducido, pero real) del proceso y la información requerida para obtener las cuatro medidas para una sola secuencia.

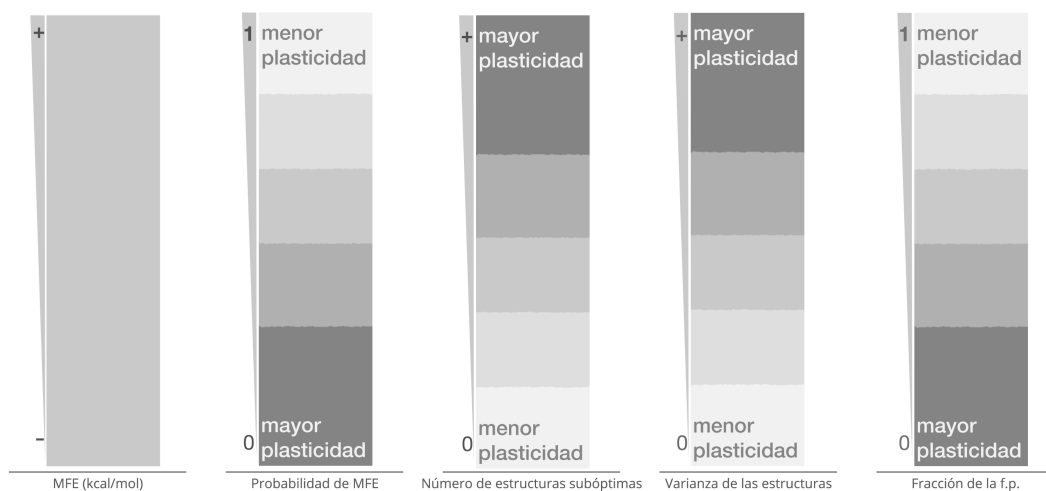


FIGURA 4.3 Relación de las medidas con la plasticidad Se muestran de izquierda a derecha: 1) Energía de la estructura de energía mínima (MFE) en kcal/mol, que no es una de las medidas de plasticidad propuestas, sino una referencia, seguido de las medidas descritas anteriormente: 2) Probabilidad de la MFE 3) Número de estructuras subóptimas, 4) Varianza de las estructuras y 5) Fracción de la función de partición. Todas representan una barra con valor creciente y el gradiente muestra la tendencia esperada de mayor (color oscuro) o menor (color claro) plasticidad. Del lado izquierdo de cada barra se muestran los valores entre los que se encuentra la medida.

CAPÍTULO 5

ANÁLISIS DE PLASTICIDAD

El método descrito en el capítulo anterior permite analizar cualquier secuencia de RNA individualmente. Sin embargo, este trabajo busca detectar tendencias evolutivas de plasticidad en los distintos grupos de RNA, por lo que es necesario elegir (de entre las 2450 familias descritas) aquellos con historias evolutivas interesantes. De manera general, se eligieron familias conservadas a través de los dominios, familias exclusivas para algún dominio, y familias con transferencia horizontal. El siguiente capítulo describe los grupos seleccionados y sus características funcionales (toda la información de WikiProject RNA y Rfam) como marco teórico para la interpretación de los resultados. Posteriormente describe el método y finalmente muestra los resultados globales del análisis.

5.1 FAMILIAS ELEGIDAS

Se eligieron tres grupos de interés para este análisis: 1) **Grupo LUCA**: Aquellos RNAs con un origen evolutivo antiguo, 2) **Grupo Transferencia horizontal**: Aquellos RNAs compartidos entre dominios que probablemente fueron adquiridos mediante eventos de transferencia horizontal, y 3) **Grupo Exclusivos**: Aquellos RNAs que se encuentran en un solo dominio. A partir de estas agrupaciones, descritas y enlistadas en Hoepfner, Gardner, & Poole (2012) (ver sección 2.4) se hizo un arbitraje manual para determinar cuales familias podían ser analizadas basándose en: 1) La longitud promedio de las secuencias pertenecientes a esas familias: el programa falla para calcular todos los plegamientos subóptimos para familias con un secuencias con un promedio de longitud mayor a 200 bases 2) Lo representativo del grupo: varios grupos están considerados a nivel clan, no a nivel familia y la base no cuenta con alineamientos semilla para el clan, por lo que sí era necesario seleccionar por familia, descartando aquellos grupos no representativos del clan o con pocas secuencias. 3) Otros asuntos relacionados con la base de datos: la base **Rfam** crece rápidamente y algunos grupos reportados en el artículo mencionado podían no existir o tener diferencias sustanciales, y 4) Alcance: particularmente en el grupo Exclusivos hay un gran número de familias que podrían ser analizadas, sin embargo se redujo el número de familias hasta

cifras manejables para este trabajo, intentado conservar aquellas con características más interesantes por función o estructura. El detalle de los grupos candidatos y la justificación para incluirlos o no se encuentra detallada en el Apéndice C. La figura 5.1 muestra diagramas equivalentes a los de Hoepfner, Gardner, & Poole (2012) pero mostrando únicamente las familias elegidas así como su distribución en los reinos.

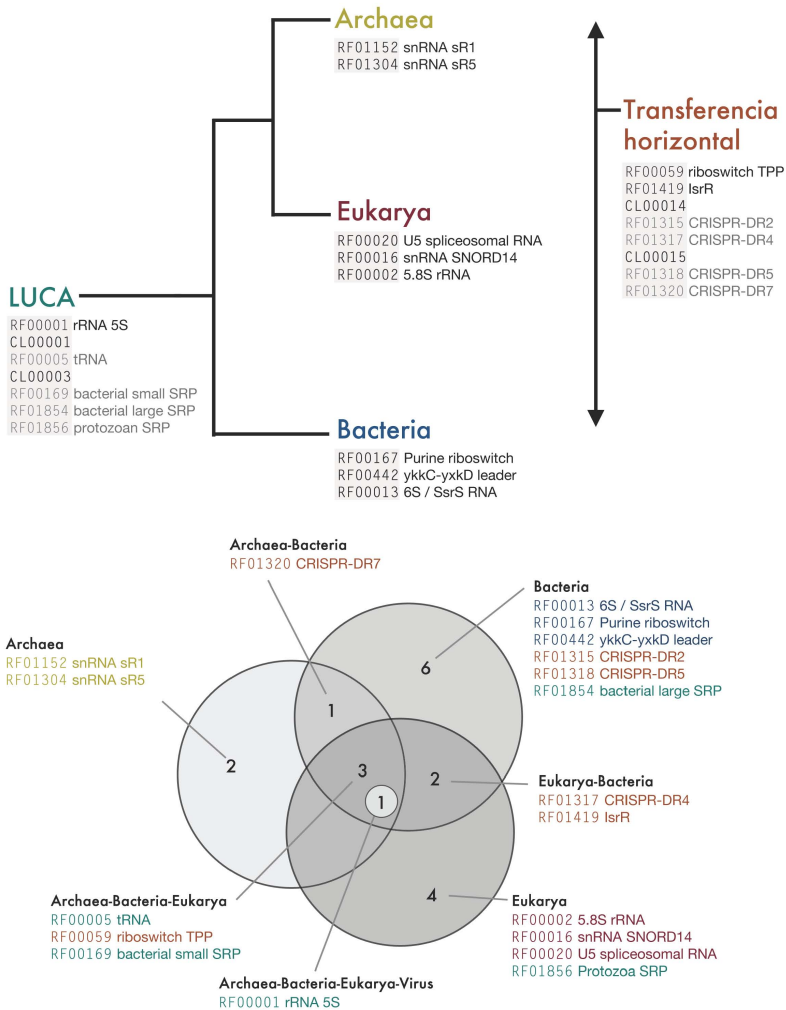


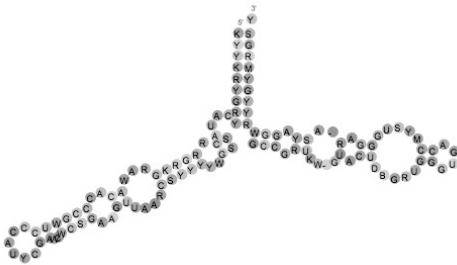
FIGURA 5.1 Distribución de grupos elegidos Las 19 familias analizadas en dos diagramas. (Arriba) La hipótesis evolutiva y (Abajo) la distribución taxonómica.

Grupo LUCA

RF00001 | 5S_rRNA

RNA ribosomal 5S

El RNA ribosomal 5S es un componente estructural y funcional de la unidad grande del ribosoma. Su estructura secundaria forma una Y, sus hélices están ancestralmente conservadas y se cree que uno de sus bucles en horquilla interactúa con el factor de transcripción IIIA. Se desconoce su función exacta, pero su eliminación reduce considerablemente la tasa de síntesis proteica. Estudios sugieren que funge como mediador entre los centros funcionales del ribosoma para la unión del tRNA.



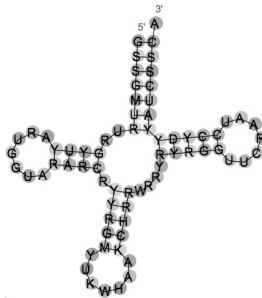
Longitud	73.4 nt
Dominios	A, B, E, V
Tipo	rRNA

CL00001

RF00005 | tRNA

tRNA

El RNA de transferencia es un componente fundamental del proceso de traducción de proteínas, pues es la molécula que lleva el aminoácido hacia el ribosoma, dirigido por un codón en el RNA mensajero. Un extremo de la molécula se aparea con el codón correspondiente por medio de un anticodón, mientras que el otro extremo está unido covalentemente a un aminoácido específico.



Longitud	116.6 nt
Dominios	A, B, E
Tipo	tRNA

CL00003

RF00169 | **Bacteria_small_SRP**

Bacterial small signal recognition particle RNA

Es un componente de la partícula de reconocimiento de señal (SRP), que es un complejo de ribonucleoproteínas que dirige el tráfico de proteínas dentro de la célula. Suelen tener motivos estructurales funcionales y sitios específicos de unión. Esta familia se encuentra principalmente en bacterias, aunque hay algunas secuencias reportadas para Archaea y Eukarya (<0.02%)



Longitud	85.8 nt
Dominios	A, B, E
Tipo	SRP RNA

RF01854 | **Bacteria_large_SRP**

Bacterial large signal recognition particle RNA

(ver RF00169) Principalmente en bacterias, <0.01% en Archaea y Bacteria



Longitud	193.9 nt
Dominios	B
Tipo	SRP RNA

RF01856 | **Protozoa_SRP**

Protozoan signal recognition particle RNA

(ver RF00169) En Eukarya, no restringido a protozoarios (<0.1% en Archaea)



Longitud	181.4 nt
Dominios	E
Tipo	SRP RNA

Grupo Transferencia horizontal

RF00059 | **TPP**

TPP riboswitch

Es un riboswitch que se une directamente al pirofosfato de tiamina para regular la expresión genética por medio de diversos mecanismos. Está presente en Archaea, Bacteria y Eukarya.

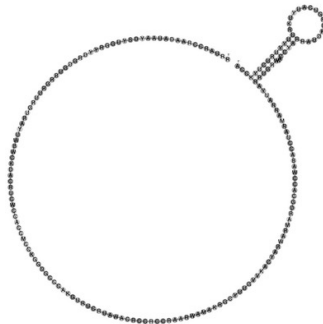


Longitud	107.4 nt
Dominios	A, B, E
Tipo	riboswitch

RF01419 | **IsrR**

RNA antisentido que regula la expresión de isiA

Es un RNA antisentido codificado en cis que regula la expresión de la proteína fotosintética isiA. Su expresión está activada por la proteína reguladora asimiladora de hierro (Fur) en condiciones de estrés por hierro. Tiene un bucle en horquilla conservado y en condiciones suficientes de hierro se une a la región complementaria del mRNA de isiA para que ambos sean degradados.

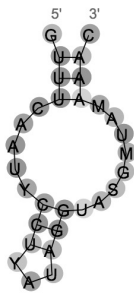


Longitud	185.2 nt
Dominios	B, E
Tipo	asRNA

CL00014 | CRISPR-1 clan

RF01315 | CRISPR-DR2 CRISPR RNA direct repeat element

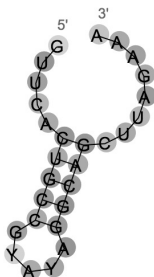
Son RNAs que en el sistema CRISPR/Cas de inmunología procarionte guían a las endonucleasas efectoras al DNA extraño, al que se unen por medio de complementariedad de secuencia. Los mecanismos de su biogénesis son diversos, pero en algunos sistemas las proteínas Cas6 reconocen los bucles en horquilla (que se forman por los elementos de repetición,) los cortan y proceden a madurar por distintos mecanismos. Después se asocian con proteínas Cas para formar los complejos ribonucleotídicos que reconocen ácidos nucleicos extraños.



Longitud	30.0 nt
Dominios	B
Tipo	crRNA

RF01317 | CRISPR-DR4 CRISPR RNA direct repeat element

(ver RF01315)



Longitud	28.0 nt
Dominios	B, E
Tipo	crRNA

CL00015 | **CRISPR-2 clan**

RF01318 | **CRISPR-DR5**
CRISPR RNA direct repeat element

(ver RF01315)



Longitud	37.0 nt
Dominios	B
Tipo	crRNA

RF01320 | **CRISPR-DR7**
CRISPR RNA direct repeat element

(ver RF01315)



Longitud	37.0 nt
Dominios	A, B
Tipo	crRNA

Grupo Exclusivos

Archaea

RF01152 | sR1

Small nucleolar RNA sR1

Los snoRNAs son RNAs de poca longitud que dirigen modificaciones químicas en otros RNAs. Éste pertenece a la clase de snoRNAs de la caja C/D (Gaspin, et al., 2000), que está asociado con la metilación. Este tipo de snoRNA tiene dos motivos de secuencia conservados C y D, así como una región de 10 a 21 nucleótidos complementarios que permite la formación de un dúplex de RNA en el sitio de metilación del RNA objetivo.

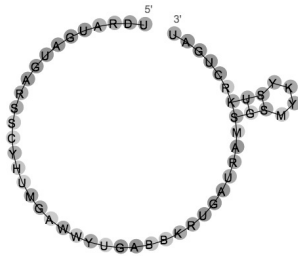


Longitud	55.7 nt
Dominios	A
Tipo	snoRNA

RF01304 | sR5

Small nucleolar RNA sR5

(ver RF01125)



Longitud	53.5 nt
Dominios	A
Tipo	snoRNA

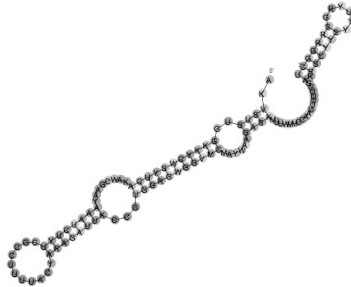
Eukarya

RF00020 | U5

U5 spliceosomal RNA

Es un componente del complejo de *splicing*, se desconocen los detalles de su función, pero se sabe que el bucle del extremo 5' es necesario para la selección del sitio de empalme y la unión de p220, mientras que el bucle en horquilla y el sitio Sm son importantes para la unión de la proteína Sm y la metilación de los *cap*.

Longitud	117.2 nt
Dominios	E
Tipo	snRNA



RF00016 | SNORD14

Small nucleolar RNA SNORD14 (U14)

Se requiere para procesos de corte en los precursores del rRNA eucariontes. En plantas y levaduras existe un bucle en horquilla (conocido como el dominio Y) que es esencial para su función, aunque éste se encuentra ausente en vertebrados.

Longitud	110.6 nt
Dominios	E
Tipo	snRNA



RF00002 | 5_8S_rRNA

5.8S ribosomal RNA

Es un componente de la subunidad grande el ribosoma eucarionte. Se cree que su función está relacionada con la translocación del ribosoma. También se le relaciona con la unión covalente a la proteína supresora de tumores p53.



Longitud	152.2 nt
Dominios	E
Tipo	rRNA

Bacteria

RF00167 | Purine riboswitch

Purine riboswitch

Es una secuencia parte de un RNA mensajero que se une a ligandos de purina y causa un cambio conformacional del mRNA. Esto tiene efectos en la traducción del mensajero ya sea dejando accesible una plataforma de expresión para un gen cercano o formando un bucle con horquilla como terminador de la traducción. Puesto que estos efectos se dan por la abundancia de la purina, las proteínas codificadas en el mensajero suelen estar asociadas con el metabolismo de las purinas.



Longitud	85.8 nt
Dominios	B
Tipo	riboswitch

RF00442 | ykkC-yxkD ykkC-yxkD leader

Es un elemento regulador en cis (región que regula la transcripción de genes cercanos), en dirección 5' de los genes ykkC y yxkD (o genes relacionados). No se conoce bien la función de esta familia, pero se ha sugerido que enciende las bombas de eflujo y sistemas de desintoxicación en respuesta a moléculas peligrosas.



Longitud	108.1 nt
Dominios	B
Tipo	cis-reg

RF00013 | 6S 6S / SsrS RNA

Es un RNA que se asocia con la holoenzima de RNA polimerasa con el factor sigma70 para reprimir la expresión de los promotores dependientes de sigma70.



Longitud	180.0 nt
Dominios	B
Tipo	— (gen)

5.2 MÉTODO GENERAL

Secuencias

Para cada familia una de las 19 familias elegidas, se descargó el alineamiento semilla de la base de datos **Rfam** por medio de su herramienta en línea (un total de 10,103 secuencias). El alineamiento semilla es conveniente pues la lista está depurada y revisada manualmente. Para las familias RF00059 y RF01419 en las que se comparó también entre elementos de la familia pertenecientes a distintos grupos taxonómicos, se solicitaron los archivos directamente a **Rfam** separados por grupo. Se eligió el formato FASTA sin espacios (**FASTA UNGapped**) para todos. Tanto el código de acceso como el número de elementos por familia se puede encontrar en el Apéndice C.

Análisis

Para todo el análisis se usó el **ViennaRNA Package 2.1.7** exportado a **Python 2.7.10** con la interfaz generada por **SWIG** disponible dentro del mismo paquete. El código desarrollado para este trabajo consta de cinco partes principales:

1. **Longitud promedio** Lee el archivo de la familia y calcula la longitud promedio de las secuencias. Es importante asegurarse de que tengan un tamaño adecuado antes de seguir con el análisis ya que el resto del programa falla con RNAs largos. Esto se debe a que las medidas de plasticidad 2, 3, y 4 requieren calcular todas las estructuras subóptimas en un rango de energía y a que el número de éstas aumenta exponencialmente con la longitud de la secuencia. A partir de varias pruebas se determinó que las familias con un promedio de longitud menor a los 200 nucleótidos eran procesables. Ésto se uso como criterio de exclusión inicial.
2. **Análisis de la familia** Lee el archivo de la familia y calcula para cada secuencia: longitud, energía de la estructura de energía mínima (MFE), medida 1, medida 2, medida 3 y medida 4. Genera dos archivos: uno con todos los resultados y otro archivo resumen con los promedios de todo lo calculado.
3. **Repeticiones con secuencias al azar** Realiza un análisis igual al del punto anterior, ahora con secuencias generadas al azar (en lugar de las secuencias biológicas), conservando la longitud de cada secuencia y el número de secuencias. Tiene dos variantes: 1) Secuencias generadas al azar con la misma probabilidad para cada base y 2) secuencias obtenidas al mezclar aleatoriamente los nucleótidos de cada secuencia original para garantizar la misma composición de bases. Finalmente, agrega los promedios de todo lo calculado al archivo resumen generado en el paso anterior.
Se repite este proceso un número n de veces, en este caso 10 repeticiones.
4. **Gráficas** Con el paquete **Plotly**, genera gráficas con la información obtenida en la parte 2 y parte 3. (Los detalles se describen más adelante.)
5. **Estadística** Con el paquete estadístico de **SciPy**, se realiza una prueba t para una muestra, con la que se comparan los promedios de las medidas para la familia contra los promedios obtenidos por los grupos generados al azar. (Se realizaron ensayos previos con la prueba Shapiro-Wilk para asegurar normalidad de los datos. El diagrama de caja y bigotes puede ser útil para discutir este supuesto en casos particulares.)

Todos los resultados de este proceso se pueden encontrar en el Apéndice C.

Reporte de datos

Hay dos tipos de figuras para las que es importante explicar con detalle su construcción:

Gráficas de caja, bigotes y punto Para cada medida se muestra un punto negro, que representa el promedio de la medida calculado para la familia y una gráfica de caja y bigotes que representa la distribución de las medidas para las repeticiones al azar (por alguno de los dos métodos). Esta figura permite ver qué tanto se aleja la plasticidad de la familia de lo esperado para secuencias de la misma longitud y en que dirección.

Ejemplo: RNA biológico (●, Familia rRNA 5S) vs. RNA generado por mezcla aleatoria (■, 10 repeticiones)

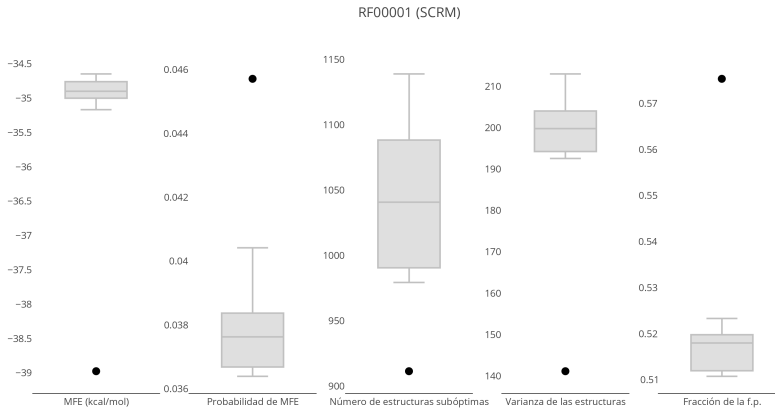


Figura de puntos por medida Debido a que es imposible reportar en un mismo diagrama todas las gráficas de caja y bigotes acompañadas de los resultados del análisis estadístico, se reportan en algunos diagramas de la siguiente manera:



Cada círculo representa una de las medidas de plasticidad. Está marcado como **igual plasticidad** si el punto queda dentro del diagrama de caja y bigotes y/o la

estadística no representa una diferencia significativa en las medias ($p < 0.05$). Está marcado como **menor plasticidad** o **mayor plasticidad** si se encuentra fuera del diagrama de caja y bigotes y el análisis estadístico indica una diferencia significativa en la media. (A partir de lo explicado en la figura 4.3, se indica **menor** si el valor de T es positivo para las medidas 1 y 4 o negativo para las medidas 2 y 3. **Mayor** si el valor de T es negativo para las medidas 1 y 4 o positivo para las medidas 2 y 3.)

Ejemplo: Con la gráfica de caja y bigotes y con el siguiente resultado del análisis estadístico para la familia rRNA 5S

		Aleatorio	RF00001	p-value	T
1	Probabilidad de la MFE	0.0383	0.04567	0.0002	3.7
2	Número de estructuras subóptimas	1031	910	0.0032	-2.9
3	Varianza de las estructuras	194.2	140.8	5.6863 e-17	-8.5
4	Fracción de la función de partición	0.5217	0.5750	5.5104 e-33	12.5

Se obtiene la siguiente figura de puntos por medida, que representa que la plasticidad del rRNA 5S es menor que la esperada en todas las medidas.

RF00001 rRNA 5S ●●●●

Interpretación

Relevancia de la longitud Como es evidente a partir de la figura 6.1, las medidas de plasticidad son dependientes de la longitud de la secuencia. Por esta razón, es más conveniente comparar la medida obtenida contra la esperada al azar para una secuencia de igual longitud. El valor estadístico T sirve como medida estandarizada para saber qué tanto se aleja la media de la familia de la media esperada al azar.

Composición de bases Se reportan las gráficas con las repeticiones con secuencias mezcladas al azar que garantizan la misma composición de bases. Sin embargo, los resultados obtenidos con el método que genera secuencias al azar con composición de bases homogénea muestran también diferencias estadísticamente significativas.

Plasticidad de la estructura secundaria Aunque las características de la estructura secundaria permiten inferir información sobre la estructura terciaria (y de manera última la estructura funcional) del RNA, no es correcto extrapolar los resultados sin cuidado. De manera concisa, lo que se describe al medir la plasticidad de la estructura secundaria del RNA es la tendencia de la secuencia para aparear sus bases de formas alternativas. Es claro que la estructura tridimensional funcional del RNA surge a partir de los distintos patrones de contactos por apareamiento de bases, pero no es la única manera en la que una molécula podría modificar su plasticidad.

RESULTADOS Y DISCUSIÓN

El análisis generó resultados que permiten discutir al menos tres aspectos sobre la plasticidad en el RNA: 1) Características de las medidas: La correlación con otros parámetros, coherencia interna y redundancia. 2) Patrones evolutivos generales, como la plasticidad con respecto a la hipótesis evolutiva de la familia o el clan o bien, con respecto a su distribución taxonómica. 3) Resultados particulares pertinentes a cada grupo, tales como la relación entre la medida obtenida y la función biológica que tienen o los patrones de plasticidad en sus subgrupos taxonómicos. A continuación se muestran y discuten algunos de estos resultados. Se incluyen también algunas consideraciones en torno al trabajo teórico y bioinformático del RNA que surgieron a partir de la investigación para el marco teórico. Finalmente se exponen algunas perspectivas futuras para este trabajo.

GENERALES

Las medidas de plasticidad están correlacionadas con la longitud

La figura 6.1 muestra los valores calculados de todos los RNAs biológicos para cada medida de plasticidad. En todos los casos una mayor longitud (número de nucleótidos) se correlaciona con mayor plasticidad. Sin embargo, también es evidente que para una misma longitud hay un rango muy amplio de valores asignados, es decir, que secuencias de un mismo tamaño pueden tener valores distintos de plasticidad lo cual indica que las medidas sí tienen valor informativo.

Estas correlaciones presentan un problema y una consideración adicional: el problema surge al intentar comparar valores de plasticidad de RNAs de distintos tamaños. Por esta razón se hicieron análisis comparando las medidas obtenidas para las secuencias de RNA biológico contra secuencias de RNA generadas al azar (ver sección 5.2) y sería también adecuado estandarizar las medidas. (ver Perspectivas) Por otro lado, existe una consideración biológica importante: sería necesario preguntar si son las medidas de plasticidad las que están correlacionadas con la longitud o si la plasticidad en sí depende de la longitud. Lo anterior no sería extraño: es claro que una secuencia con mayor longitud aumenta las

combinaciones posibles de apareamientos entre bases y que las secuencias largas tienden a formar estructuras de menor energía, por lo que la plasticidad intrínseca de la molécula estaría dada en parte por la longitud de la secuencia. La implicación biológica de lo anterior es que la evolución podría darse directamente sobre la longitud del RNA o bien, sobre la secuencia para modificar la plasticidad de un RNA de cierto tamaño dado.

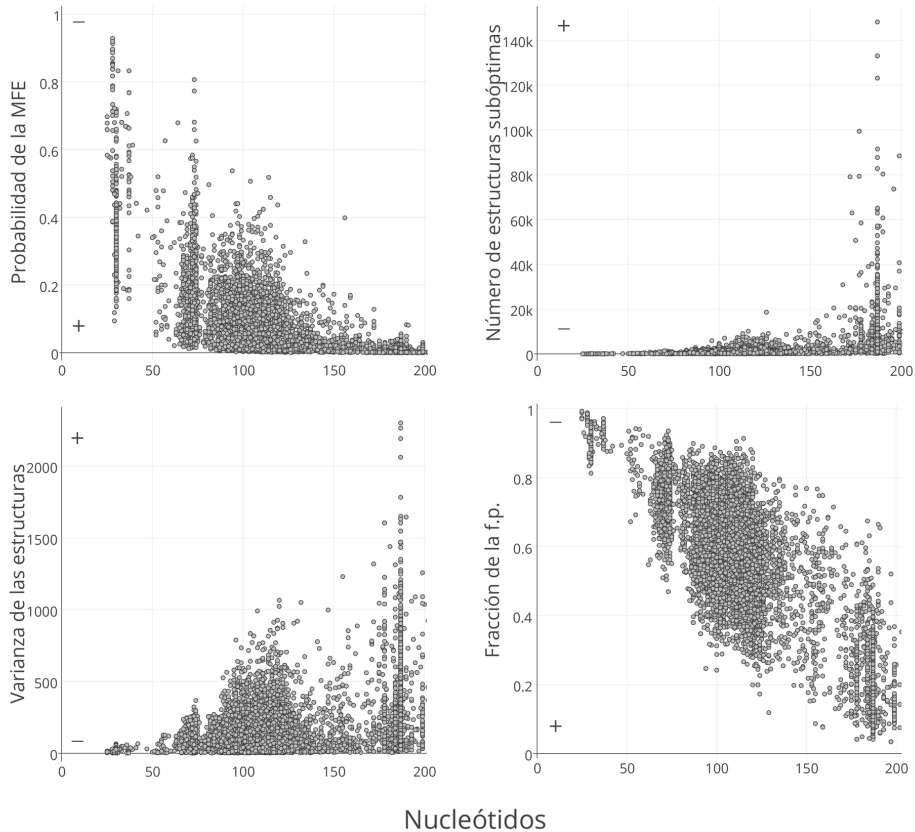


FIGURA 6.1 Medidas de plasticidad y longitud de la secuencia Se muestra el valor obtenido para todas las medidas calculadas de todos los RNAs biológicos procesados ($n = 10,103$) y su respectiva longitud (número de nucleótidos). Como referencia, en el lado derecho de cada eje se muestran los símbolos + y - junto a los valores asociados a mayor y menor plasticidad, respectivamente. Las gráficas están truncadas debido a que los mayores a 200 estaban poco representados por el método de selección de las familias.

Las medidas son coherentes y no redundantes

En la figura 6.2 se puede observar que en las familias analizadas la mayoría de las medidas apuntan hacia un mismo comportamiento de plasticidad y que en ningún

caso se contradicen gravemente. Las medidas tampoco son redundantes puesto que existen algunas que difieren de la tendencia general para la familia y la manera en la que lo hacen no es contante, es decir, ninguna de las medidas repite exactamente el comportamiento de otra. Considerar las cuatro medidas nos da información más completa sobre las características plásticas del grupo y deben ser interpretadas en conjunto.

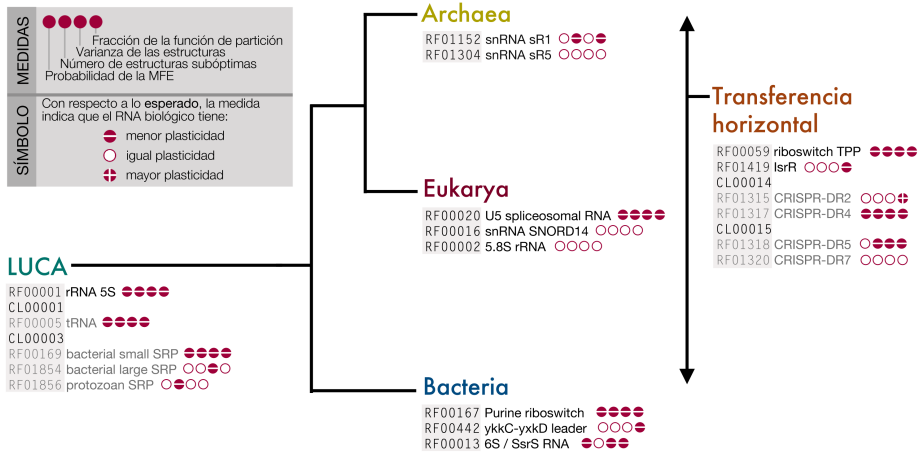


FIGURA 6.2 Plasticidad y relaciones evolutivas de las familias Se muestra el mismo diagrama presentado en la figura 5.1, mostrando los resultados obtenidos de cada medida de plasticidad para todas las familias analizadas.

Todos los grupos distribuidos en los tres reinos son menos plásticos que lo esperado

En la figura 6.3 se muestra la presencia de las familias a través de los reinos y usa el código de color de la figura 6.2. Difieren en que la primera muestra las hipótesis evolutivas de familias y clanes, y debido a que la distribución de un clan está dada por todas las familias que contiene, no necesariamente corresponden ambas distribuciones.

El centro incluye las familias presentes en los tres dominios: Archaea, Bacteria y Eukarya. Se analizaron cuatro familias de este tipo: rRNA 5S, tRNA, riboswitch TPP y SRP pequeño bacteriano y los cuatro grupos muestran menor plasticidad que la esperada en las cuatro medidas. El rRNA 5S, el tRNA y el SRP pequeño bacteriano se consideran familias ancestrales (LUCA) y el riboswitch TPP se cree adquirido por transferencia horizontal.

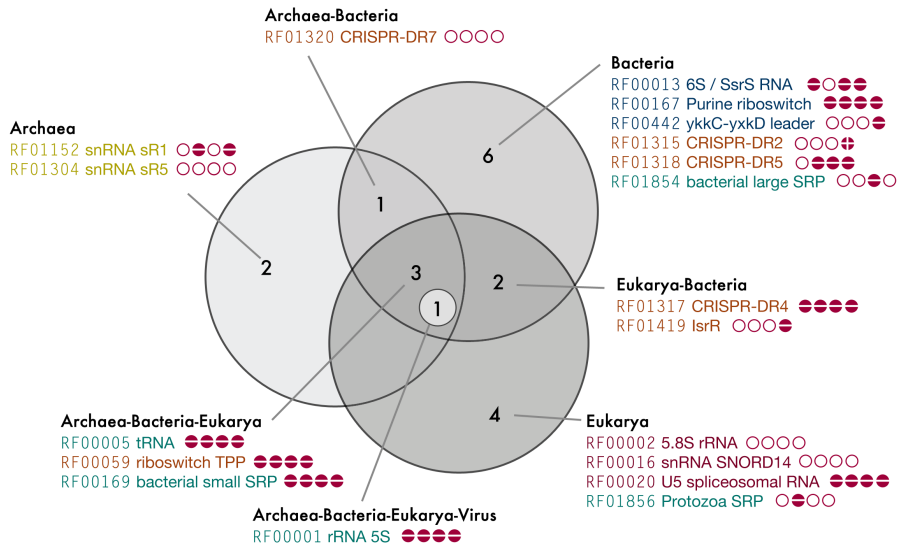


FIGURA 6.3 Plasticidad y distribución de las familias. Se muestra el mismo diagrama presentado en la Figura 5.1, mostrando los resultados obtenidos de cada medida de plasticidad para todas las familias analizadas.

Los tres grupos LUCA ampliamente distribuidos participan en procesos metabólicos fundamentales (ver figura 2.9) y se conoce la relevancia de la estructura del RNA en su función. Es interesante notar también que los grupos LUCA que no están ampliamente distribuidos no presentan las mismas características de plasticidad. Este resultado sugiere una tendencia evolutiva hacia estructuras estables para estas funciones, cuya estabilidad dependa en gran medida de su estructura secundaria.

Tendencia de reducción de plasticidad

La mitad de las familias analizadas son en general menos plásticas que las secuencias generadas *in silico* por alguno de los dos métodos. (Tres o más de las medidas indican menor plasticidad para 9/19 de las familias y 1 tiene dos medidas que indican menor plasticidad y dos que indican igual.) Este patrón apoya la idea de la plasticidad como una característica que puede evolucionar.

Sólo una medida de una familia muestra una tendencia hacia mayor plasticidad (Fracción de la función de partición para CRISPR-DR2). Sería interesante buscar otras familias en las que eso se presente como tendencia general de las medidas para quizás formular hipótesis sobre el tipo de RNAs que requieren esto.



FIGURA 6.4 Medidas de plasticidad comparadas Se muestran los resultados para tres medidas de plasticidad, comparando entre las secuencias biológicas (extraídas de bases de datos) y las secuencias generadas *in silico*.

La composición de bases no tiene un efecto muy claro sobre la plasticidad

Los promedios obtenidos de tres medidas para todas las familias se pueden encontrar en la figura 6.4, acompañados de los promedios obtenidos con los RNAs generados aleatoriamente por ambos métodos. El método para generarlos con la misma composición mezcla aleatoriamente la secuencia original, mientras que el método con composición al azar genera una secuencia de igual longitud en la que las bases son seleccionadas aleatoriamente con $\frac{1}{4}$ de probabilidad cada una. Como se aprecia en la figura, los RNAs biológicos suelen tener valores distintos a los RNAs generados. Sin embargo, no hay una diferencia clara ni constante entre los dos métodos de RNAs generados, en muchos casos son indistintos y en otros hay diferencia que no se mantiene entre familias. Lo anterior indica que probablemente la composición de bases no es muy relevante para la plasticidad de la molécula, a pesar de la estabilidad que brindan ciertos apareamientos de bases y de la composición de bases sesgada en ncRNAs (Schattner, 2002)

La distribución de la familia no está relacionada con el grado de reducción de plasticidad

El valor T de la prueba estadística t de Student nos permite comparar de manera estándar qué tanto se aleja la medida del RNA biológico de la media de los RNAs generados. La figura 6.5 muestran los valores T de todas las familias, agrupadas en aquellas que están presentes en uno, dos o tres dominios. No se observa una diferencia entre dichos grupos. Las medidas muestran de nuevo ser coherentes más no redundantes.

Los resultados son coherentes con los antecedentes

Tanto Miklós, Meyer, y Nagy (2005) como Chan y Ding (2008) parecen estar midiendo aspectos equivalentes a los que se usan para medir la plasticidad en este análisis. Sus resultados indican que hay diferencias físicas, particularmente en propiedades del ensamble de Boltzmann, entre las secuencias biológicas y las secuencias al azar, así como dentro de los distintos tipos de RNA. Este análisis encuentra resultados similares pero con un alcance mucho mayor debido a tres diferencias fundamentales: 1) La agrupación de los tipos de RNA: los grupos que se usan aquí pretenden ser grupos naturales, lo cual ayuda a garantizar que no haya ruido en el resultado 2) La diversidad de grupos y el tamaño de la muestra: mientras que Miklós *et al.* analiza 736 secuencias en tres grupos por tipo (tRNAs, 5S rRNAs y pre-miRNAs) y Chan y Ding (2008) analizan 118 secuencias en tres agrupaciones artificiales (RNAs estructurales, mRNAs y pre-miRNAs) en este análisis se procesaron un total de 10,103 secuencias distintas, distribuidas en 19 familias. 3) La interpretación en un contexto evolutivo: las cuatro medidas

utilizadas buscan interpretar distintos componentes del concepto de plasticidad, además los resultados son interpretados en el contexto evolutivo, funcional y de distribución de las familias.

	Probabilidad de la MFE	Número de estructuras subóptimas	Varianza de las estructuras	Fracción de la función de partición	Distribución por dominio
RF00001 rRNA 5S					A, B, E, V
RF00005 tRNA					A, B, E
RF00059 TPP riboswitch					A, B, E
RF00169 bacterial small SRP					A, B, E
RF01320 CRISPR-DR7					A, B
RF01317 CRISPR-DR4					B, E
RF01419 lsrR					B, E
RF01152 snoRNA sR1					A
RF01304 snoRNA sR5					A
RF00013 6S / SsrS RNA					B
RF00167 purine riboswitch					B
RF00442 ykkC-yxkD leader					B
RF01318 CRISPR-DR5					B
RF01854 bacterial large SRP					B
RF01315 CRISPR-DR2	+			+	B
RF00002 rRNA 5.8S					E
RF00016 snRNA SNORD14					E
RF00020 U5 spliceosomal RNA					E
RF01856 protozoa SRP					E

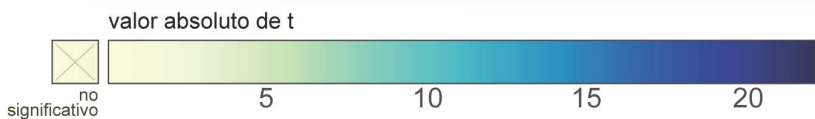


FIGURA 6.5 Valores t para medidas de plasticidad El valor t se obtiene al comparar las medidas de plasticidad de la familia contra sus respectivas mezclas al azar. El color indica a qué grado se aleja la plasticidad de lo esperado al azar. Para todos los casos significativos la plasticidad es menor que la esperada con excepción de los marcados con el símbolo +. Por lo tanto, las bandas horizontales más azules corresponden a los grupos menos plásticos. Se encuentran ordenadas por distribución por dominio, en donde A= Archaea, B= Bacteria, E= Eukarya y V=Virus.

PARTICULARES

La familia tRNA se aleja notablemente de lo esperado

La familia del tRNA (RF00005) es la única que se encuentra constantemente entre las 3 más alejadas de lo esperado hacia menor plasticidad (figura 6.5). Se sabe que la estructura de la molécula es fundamental para la función de los tRNAs y que algunos componentes estructurales están muy bien conservados a pesar de ser una familia tan antigua. El notable alejamiento de lo esperado por azar supone una tendencia evolutiva hacia adquirir plegamientos poco plásticos, probablemente garantizando así el correcto funcionamiento de los tRNAs. Este resultado no es sorprendente, y en conjunto con más información experimental podría ser usado para validar este método para medir plasticidad.

Ambos riboswitches son poco plásticos y parecidos

Tanto el riboswitch TPP (RF00059) como el riboswitch de purina (RF00167) son significativamente menos plásticos que lo esperado. Además, las medidas de plasticidad obtenidas para ambas familias no difieren demasiado y ambas se alejan de lo esperado de manera similar. Lo anterior se puede observar en las figuras 6.6 y 6.7. Esto es interesante debido a que la actividad de los riboswitches consiste en la unión de un ligando que cambia el plegamiento del RNA. Naturalmente, el plegamiento que se predice para este análisis es el que tiene la molécula sin el ligando, por lo que nos revelaría información sobre el aptámero (que es el componente estructural que interacciona con el ligando). Batey (2012) menciona la importancia de la estructura secundaria y terciaria para el funcionamiento de los riboswitches. Ya que al menos la mitad de los riboswitches conocidos controlan la expresión a nivel transcripcional, requieren un plegamiento rápido y certero, antes de que la RNA polimerasa actúe. El evolucionar una estructura secundaria con poca plasticidad y cuya dinámica de plegamiento dependa de principalmente de la estructura secundaria es coherente con la función de estas familias.

La plasticidad de una misma familia cambia entre subgrupos taxonómicos

Se analizaron independientemente algunas familias, separando las secuencias por subgrupos taxonómicos.

El riboswitch TPP (RF00059) se encuentra en Archaea, Eukarya y Bacteria. Todos los subgrupos tienen el mismo patrón que la familia completa: todas las medidas son significativamente diferentes a lo esperado para secuencias aleatorias. Sin embargo, la magnitud de esta diferencia es distinta. Aquellos en el dominio

Archaea son más plásticos, seguido de aquellos en Eukarya y finalmente, Bacteria con los menos plásticos.

Algo similar sucede con el RNA IsrR (RF01419) con los subgrupos para Bacteria y Eukarya. Ambos mantienen el patrón encontrado en la familia, con todas las medidas igual a lo esperado, con excepción de la fracción de la función de partición, que indica menor plasticidad. En este caso, los RNAs IsrR en Eukarya se alejan más de lo esperado que los de Bacteria.

La figura 6.5 indica que no hay una tendencia general de plasticidad con respecto a los dominios en sí, pero estos dos ejemplos muestran que existe una diferencia en la plasticidad en subgrupos taxonómicos dentro de una misma familia. Esto puede ser considerada una evidencia de la evolución de la plasticidad.

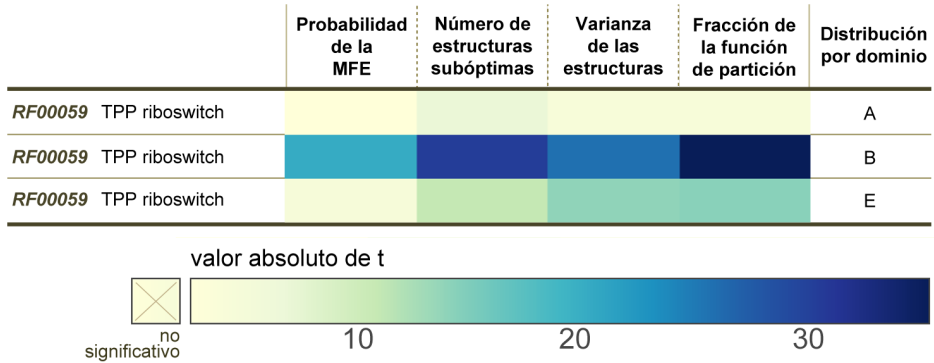


FIGURA 6.6 Valores T para subgrupos de la familia: Riboswitch TPP (*RF00059*)

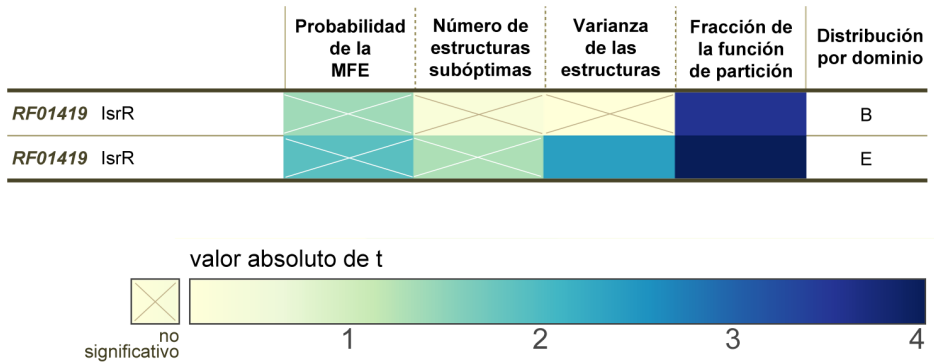


FIGURA 6.7 Valores T para subgrupos de la familia: RNA IsrR (*RF01419*)

Probablemente las familias CRISPR-DR presentan sesgo muestral

Todas están catalogadas como elementos de repeticiones directas, pero se desconoce su función específica. A pesar de que las estructuras secundarias consenso sean bastante similares (ver Apéndice C) las cuatro familias analizadas presentan propiedades de plasticidad muy distintas. CRISPR-DR4 (RF01317) tiene todas sus medidas con significativamente menor plasticidad que la esperada. También aparece constantemente entre los primeros 4 más alejados de lo esperado (figura 6.5) con menor plasticidad, superado globalmente sólo por la familia tRNA. El CRISPR-DR7 (RF01320) presenta igual plasticidad que la esperada para todas sus familias. Por otro lado, CRISPR-DR5 (RF01318) tiene 3 de 4 medidas con significativamente menor plasticidad que la esperada y se aleja menos que CRISPR-DR5 en todas las medidas. Finalmente, CRISPR-DR2 (RF01315) es la única familia que presenta mayor plasticidad de la esperada en una de sus medidas y está constantemente entre las más alejadas hacia mayor plasticidad. (figura 6.5)

Estos resultados probablemente se deban a un efecto causado por la corta longitud de estas familias y al tamaño de la muestra. (CRISPR-DR2: 30 nt, n=19; CRISPR-DR4: 28 nt, n=33; CRISPR-DR5: 37 nt, n=12, y CRISPR-DR7: 37 nt, n=10) Con una muestra más grande que presentara un patrón similar, las medidas de plasticidad podrían servir como indicadores de diversidad funcional de estos elementos, pero es imposible aún con la casi nula información sobre estas familias. Es muy probable que se trate de un sesgo muestral.

El U5 RNA del complejo de *splicing* es poco plástico

Otro grupo que constantemente aparece entre los que se alejan más de lo esperado, hacia menor plasticidad es la familia U5 RNA del complejo de *splicing* (RF00020). Es un componente de ambos tipos de complejo de *splicing* y a pesar de que su función precisa se desconoce, se identifican algunos de sus elementos estructurales como fundamentales para su función como el bucle 5' que selecciona la región de *splicing* y el bucle en horquilla 3' que junto con el sitio Sm son importantes para la unión de la proteína Sm (“U5 RNA”, WikiProject RNA)

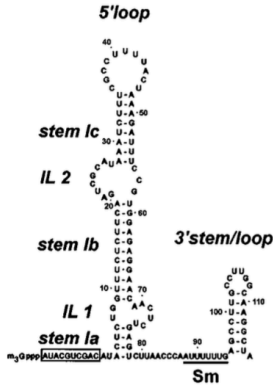


FIGURA 6.8 Modelo de la estructura secundaria del U5 RNA humano. Muestra el sitio de unión Sm, el bucle 5', el bucle en horquilla 3' y otros elementos estructurales (Hinze, Moore, & Bindereif, 1996)

El RNA SRP más pequeño y más ampliamente distribuido es el menos plástico

Se analizaron 3 familias: RNA SRP pequeño bacteriano (RF00169, $\bar{n}\tau=85.8$) RNA SRP protozoario (RF01856, $\bar{n}\tau=181.4$) y RNA SRP grande bacteriano (RF01854, $\bar{n}\tau=193.9$). De éstos, sólo el RNA SRP pequeño bacteriano es significativamente menos plástico que lo esperado en sus cuatro medidas (figuras 6.2 y 6.3) además de ser el que más se aleja (figura 6.5). Los otros dos grupos presentan 3/4 medidas iguales a lo esperado y una menor.

El RNA SRP se ha encontrado en todos los dominios de la vida pero el tamaño y el número de componentes estructurales de la molécula varían considerablemente entre grupos filogenéticos (Rosenblad, Zweib, & Samuelsson, 2009), por lo que no es sorprendente encontrar patrones de plasticidad distintos. La figura 6.9 muestra tres tipos de RNA SRP como referencia a lo diferentes que pueden ser.

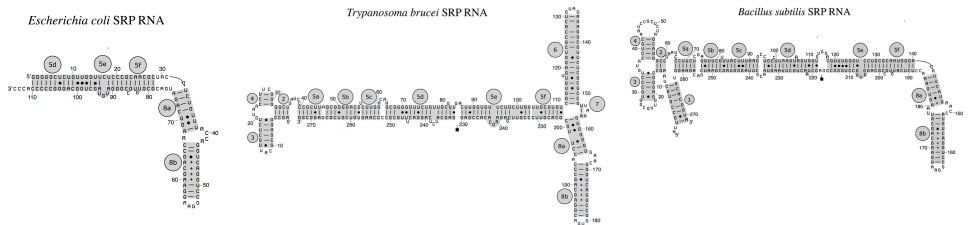


FIGURA 6.9 RNA SRPs. RNA SRP pequeño de *Escherichia coli*, SRP RNA del protozoario *Trypanosoma brucei* y RNA SRP largo de la bacteria *Bacillus subtilis*. (Rosenblad et al., 2009)

Los rRNAs tienen distinta plasticidad

Mientras que el 5S rRNA es menos plástico que lo esperado en todas sus medidas, el 5.8S es igual de plástico en todas sus medidas. Las estructura secundarias consenso del 5.8S rRNA muestra un bucle grande que puede ser parte de la razón para esta diferencia, sin embargo difícilmente se puede formular una hipótesis funcional sin los estudios estructurales específicos.

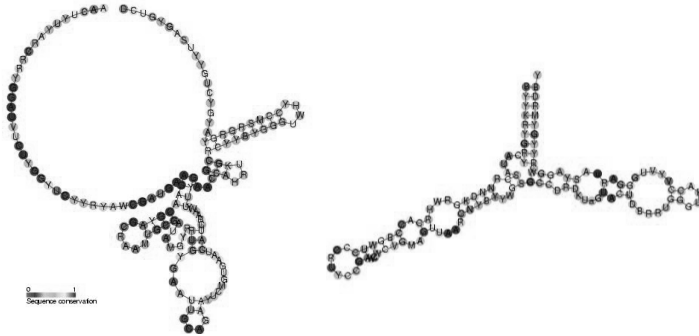


FIGURA 6.10 Estructura secundaria consenso de dos rRNAs 5.8S rRNA (izquierda) y 5S rRNA (derecha)

El líder *ykkC-yxkD* muestra un patrón inesperado

Aunque no se conoce el mecanismo preciso del líder *ykkC-yxkD* (RF00442), se sabe que tiene una estructura regulatoria conservada y se sospecha que podría funcionar como un riboswitch. (“YkkC/YxkD leader”, WikiProject RNA) Sin embargo, presenta plasticidad igual a la esperada para una secuencia aleatoria de la misma longitud.

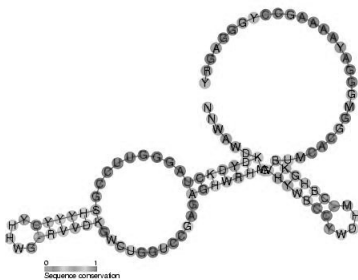


FIGURA 6.11 Estructura secundaria consenso del líder YkkC-TxkD

El RNA *IsrR* muestra un patrón esperado

El RNA IsrR (RF01419) presenta igual plasticidad que la esperada para todas sus medidas, con excepción de la fracción de la función de partición. Es un RNA antisentido (asRNA) que regula la expresión de *isiA*. Los asRNA funcionan por medio de una secuencia complementaria que se aparea con el mRNA (en este caso de *isiA*, que es una proteína parte del proceso fotosintético). La única estructura conservada del RNA IsrR es un tallo en horquilla. La región que no está plegada corresponde a una región complementaria al mRNA que regula por lo que la estructura secundaria de esa región no es relevante para su función.

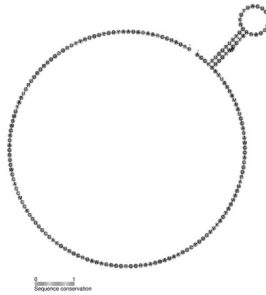


FIGURA 6.12
Estructura secundaria
consenso del RNA
IsrR

CONSIDERACIONES ADICIONALES

Formalización

El campo de estudio requiere urgentemente una estandarización en nomenclatura y una revisión de términos. En la biología existe gran escepticismo hacia la formalización excesiva y se prefiere omitir cuando los conceptos son complejos. No es mayor problema si no afecta la comunicación dentro de la disciplina, pero este no es el caso para la mayoría de los temas tratados en este trabajo. Por mencionar algunos ejemplos: el concepto de *motivo* (ver sección 2.2), las definiciones de los *niveles estructurales* (ver sección 2.2), la nomenclatura de los *apareamientos* (ver sección 2.1), la definición y acotamiento del concepto de *RNA no codificante* (ver sección 2.4) y más notablemente el uso laxo de las abreviaciones para los “tipos” de RNA, que impide la búsqueda eficiente de referencias e interfiere con los esfuerzos para clasificarlos naturalmente. El Apéndice A recopila una tabla con algunas de las abreviaciones encontradas durante la realización de esta tesis que no pretende más que servir como guía para ahorrar tiempo en futuras lecturas.

Validación y vigencia de los resultados

El supuesto más fundamental del trabajo es que el modelo de predicción de la estructura secundaria funciona. Es tentador pensar que obtener los resultados esperados es una manera de validarlo, pero no es así. Se deben interpretar los resultados suponiendo que el programa ViennaRNA predice correctamente los parámetros físicos requeridos, de lo contrario se cae en una lógica circular en donde la hipótesis que se pone a prueba es la misma que se usa para generar los resultados a poner a prueba. Se puede suponer que en algunos casos la estructura secundaria no es suficiente para explicar la función del RNA, pero no se pueden usar los resultados inesperados para invalidar el método de predicción. Validar el modelo debe ser un proceso externo en el que se involucre más evidencia teórica en conjunto con evidencia experimental.

Dicho esto, es seguro que los algoritmos seguirán siendo refinados y que los métodos de predicción cambiarán. Lo mismo sucederá con la clasificación de las familias y con la cantidad de información disponible. Lo anterior es un proceso normal y necesario que podría cambiar en distintos grados los resultados de este análisis. Sin embargo, considero que los resultados obtenidos en este trabajo no son más importantes que la pregunta y el método propuesto.

PERSPECTIVAS

El potencial de los análisis teóricos-computacionales de conceptos evolutivos es inmenso. Realmente sobran las preguntas que se pueden abordar con estos métodos. Con respecto a este trabajo hay muchos aspectos que se pueden mejorar y/o ampliar. El primero y más evidente es la estandarización de medidas con respecto a la longitud del RNA. Compararlos con respecto a la estadística T es complicado y agrega supuestos que podrían mejor omitirse. Otro punto que se podría explorar con más profundidad es el análisis de los subgrupos taxonómicos de una familia, ya que permitiría inferir patrones evolutivos más claros y específicos.

Por otro lado, sería un interesante experimento adicional explorar las propiedades de robustez de la plasticidad con un experimento *in silico* en el que se hagan mutaciones puntuales al azar y se vea el efecto de éstas sobre la plasticidad. Con éste mismo método, combinado con el de mutaciones neutrales estructurales usado para hacer las redes neutrales (Ancel & Fontana, 2000), se podría definir una estructura secundaria y modificar únicamente las propiedades de plasticidad hasta llegar a un valor de plasticidad deseado. Esto se podría comparar contra otras estructuras biológicas para formular hipótesis sobre los mecanismos de la evolución de la plasticidad o bien para diseñar estructuras de RNA estables.

Se podría también evaluar la capacidad de estas medidas para buscar ncRNAs funcionales, ya sea integrado a la búsqueda de elementos por familia o a búsquedas generales en genomas. Tendría que evaluarse la resolución para ver si es posible distinguir entre secuencias individuales y revisar la eficiencia de la búsqueda.

Muchos otros conceptos como canalización y evolutividad pueden también ser explorados a partir de estas herramientas. Los experimentos computacionales son sencillos, de bajo costo, requieren cada vez menos poder de cómputo y todo el tiempo aumenta la información para realizarlos. El alcance de los estudios teóricos y computacionales en biología es inmenso y más aunado a los esfuerzos por mantener la información libre: tanto la base de datos **Rfam** y los programas usados en este análisis, **Python** (y sus paquetes), **ViennaRNA**, y **VARNA** son libres.

APÉNDICES

APÉNDICE A: LISTA DE ABREVIACIONES DE RNA

La siguiente tabla recopila los nombres completos a los que hacen referencia las abreviaciones. Se mantienen los nombres en inglés ya que pretende servir como guía en la búsqueda de referencias.

Abreviación	Nombre
6S RNA	RNA no codificante regulador de la RNA polimerasa Se une a la holoenzima σ^{70} para regular expresión genética.
7SK RNA	snRNA regulador de la transcripción Se encuentra en un complejo de snRNP que controla el factor de elongación de la transcripción, P-TEFb.
7SL RNA	<i>ver SRP RNA</i> (también <i>4.5S SRP RNA</i> , <i>4.5S RNA</i>)
AirRNA	RNA no codificante silenciador de los genes autosómicos <i>Igf2r/Slc22a/Slc22a3</i> en ratón
aRNA	<i>ver asRNA</i>
asmiRNA	<i>anti-sense micro RNA</i>
asRNA	<i>antisense RNA</i> RNA de una sola hebra [<i>ssRNA</i>] que es complementario a un mRNA transcrito en una célula
BC1RNA	<i>small brain-specific non-messenger RNA</i> RNA transportado hacia dendritas de células neuronales donde interviene en la regulación de la traducción cerca de las sinapsis.
cis-NAT	<i>cis-natural antisense transcript</i> RNAs complementarios a otros RNAs que se transcriben del mismo locus genético de su RNA objetivo, pero en la hebra de DNA opuesta y forma apareamientos perfectos.
CRISPR RNA	<i>clustered regularly interspaced short palindromic repeats RNA</i> Los RNAs guías de los sistemas CRISPR/Cas
crRNA	<i>ver CRISPR RNA</i>
DDRNA	<i>DNA damage response RNA</i> ncRNAs pequeños, activados por las endoribonucleasas DICER y DROSHA para la activación de la el sistema de respuesta de daños del DNA (DDR: <i>DNA Damage Response</i>)
diRNA	<i>DNA double-strand break (DSB)-induced small RNAs</i> (<i>ver DDRNA</i>)
DsrARNA	ncRNA con roles en la regulación de la transcripción y traducción

dsRNA	<i>double stranded RNA</i> RNA compuesto de dos hebras
endo-siRNA	<i>endogenous small interfering RNAs</i> Son reguladores endógenos de la expresión genética por medio de interferencia.
exRNA	<i>extracellular RNA</i> RNA que se encuentra fuera de la célula en la que se transcribió
fRNA	<i>functional RNA</i> (ver ncRNA)
gRNA	<i>guide RNA</i> Se puede referir a RNAs que señalan un RNA específico, como aquellos en el sistema CRISPR/Cas o a un grupo específico de RNAs que guían la edición del RNA por inserción o eliminación de residuos de uridina.
hc-siRNA	<i>heterochromatic small interfering RNA</i> , también <i>HC siRNA</i> o <i>hcsiRNA</i>
hnRNA	<i>heterogeneous nuclear RNA</i> RNA inmaduro en el núcleo que puede tener distintos destinos, aunque principalmente se refiere a mRNAs inmaduros
iRNA	<i>interference RNA</i> RNA involucrado en procesos de interferencia, ver RNAi
lincRNA	<i>long intergenic non-coding RNAs</i> lncRNAs transcritos de regiones no codificantes entre genes de proteínas
lncRNA	<i>long non-coding RNA</i> ncRNAs con longitud mayor a los 200 nucleótidos
miRNA	<i>micro RNA</i> ncRNAs pequeños, que participan en el silenciamiento de RNAs y en regulación de expresión genética postranscripcional. Suelen formar estructuras de bucle en horquilla y funcionan a través de apareamiento de bases con los mRNAs objetivo
micRNA	<i>messenger RNA-interfering RNA</i> ver asRNA, iRNA
mRNA	<i>messenger RNA</i> RNAs mensajeros, con información para codificar proteínas. Contraparte de los <i>ncRNAs</i>
mrpRNA	<i>RNase MRP RNA</i> RNA de la ribonucleoproteína RNasa MRP, para el procesamiento de mitocondrias
natsiRNA	<i>natural antisense short interfering RNA</i> , también <i>nat-siRNA</i> Un tipo de <i>siRNA</i> , endógeno que se genera de un transcrito complementario al mRNA
ncRNA	<i>non-coding RNA</i> RNA no codificante, todo aquel cuyo destino último no es codificar una proteína. Contraparte de los <i>mRNAs</i>

nmRNA	<i>non-messenger RNA</i> ver ncRNA
NoRC associated RNA	ver pRNA
npcRNA	<i>non-protein coding RNA</i> ver ncRNA
OxyS RNA	<i>oxidative stress response RNA</i> ncRNA que regula la expresión de un gran número de genes en respuesta al estrés oxidativo
piRNA	<i>piwi-interacting RNA</i> RNAs que interactúan con proteínas piwi (<i>P-element induced wimpy testis</i>) que son proteínas regulatorias en procesos de desarrollo.
pre-mRNA	<i>messenger RNA precursor</i> ver hnRNA, mRNA antes de ser procesado
pRNA	ncRNA que regula la transcripción del rRNA
qiRNA	<i>QDE-2 interacting RNA</i> Un tipo de siRNA en respuesta al daño del DNA
rasiRNA	<i>repeat associated siRNA</i> Tipo de piRNA que mantiene la estructura de la cromatina, controla algunos transcritos y silencia transposones y retrotransposones
RNAi	<i>RNA interference</i> Interferencia de RNA, Proceso en el que el RNA inhibe la expresión genética
RNase MRP	ver mrpRNA
RNase P	<i>ribonuclease P</i> Es una ribozima que procesa precursores de otros RNAs
rRNA	<i>ribosomal RNA</i> Componentes de RNA de los ribosomas
scaRNA	<i>small Cajal body-specific RNA</i> Participan en la maduración de los snRNAs
scnRNA	<i>small-scan RNA</i> Un tipo de siRNA
scRNA	<i>small cytoplasmic RNA o small conditional RNA</i> El small cytoplasmic RNA es un RNA citoplásmico involucrado con la SRP, mientras que el small conditional RNA es un RNA sintético diseñado para cambiar de conformación en respuesta a estímulos específicos
SgrS RNA	<i>sugar transport-related sRNA</i> Un RNA pequeño activado durante la respuesta al estrés de azúcar-fosfato
shRNA	<i>short hairpin RNA</i> RNA sintético para procesos de interferencia por RNA

siRNA	<i>small interfering RNA, short interfering RNA o silencing RNA</i> RNAs de doble hebra involucrados en procesos de interferencia por RNA
SL RNA	<i>spliced Leader RNA</i> RNA de procesamiento, lleva acabo un mecanismo trans de maduración de mRNAs
smRNA	<i>small RNA o small modulatory RNA</i> Agrupación mal definida, es un término que se usa casi exclusivamente para describir RNAs pequeños con funciones no conocidas, o métodos de RNA-seq que los buscan.
SmY RNA	<i>mRNA trans-splicing</i> Familia de snRNAs encontrados en nematodos, involucrados en el <i>splicing trans</i> del mRNA.
snmRNA	<i>small non messenger RNA</i> Es una agrupación mal definida para nombrar RNAs pequeños con funciones desconocidas.
snoRNA	<i>small nucleolar RNA</i> RNAs que guían modificaciones químicas de otros RNAs, en especial de algunos rRNAs, tRNAs y snRNAs.
snRNA	<i>small nuclear RNA</i> Son RNAs del spliceosoma que participan en el procesamiento de los pre-mRNAs en el núcleo. Se encuentran asociados a snRNPs.
snRNP	<i>small nuclear ribonucleic proteins</i> Complejos de proteínas con snRNAs para el procesamiento de pre-mRNAs
sRNA	<i>small RNA,</i> abreviación antiguamente usada para referirse a <i>soluble RNA</i> , ahora conocido como tRNA Es una agrupación mal definida, comúnmente usada para referirse a RNAs de pequeño tamaño con funciones no conocidas.)
SRP RNA	<i>signal recognition particle RNA</i> RNA parte del complejo ribonucleoprotéico SRP (partícula de reconocimiento de señal) que dirige el tráfico y secreción de proteínas en la célula.
ssRNA	<i>single stranded RNA</i> RNA de una sola hebra
SsrS RNA	<i>ver 6S RNA</i>
stRNA	<i>small temporal RNA</i> Agrupación mal definida, usada para miRNAs de regulación en nematodos
tacrRNA	<i>trans-activating crRNA</i> crRNAs con actividad <i>trans</i>
tasiRNA	<i>trans-acting siRNA</i> siRNAs con actividad <i>trans</i>

TERC	<i>telomerase RNA component</i> ncRNA parte de la telomerasa
tmRNA	<i>transfer-messenger RNA</i> RNA bacteriano con propiedades de tRNA y de mRNA. Parte de un complejo ribonucleoprotéico (tmRNP)
tncRNA	<i>tiny non coding RNA</i> Agrupación mal definida, usada para RNAs muy pequeños con funciones no conocidas.
trans-NAT	<i>trans-natural antisense transcript</i> RNAs complementarios a otros RNAs que se transcriben un locus genéticos distinto al de su RNA objetivo, y no suele formar apareamientos perfectos. Incluye a los miRNAs
tRNA	<i>transfer RNA</i> RNA de transferencia
uRNA / U RNA	<i>ver snRNA</i> Usada para referirse a grupos específicos de snRNAs, U1–U8, U11, U12, U14, U18, U48, U49, etc.
vRNA	<i>vault RNA</i> RNAs parte de complejos ribonucleoproteicos en el citoplasma llamados <i>vault</i> , cuya función se desconoce pero se conjetura que está relacionada con la expulsión de xenobióticos de la célula
vtRNA	<i>ver vRNA</i>
XistRNA	<i>X-inactive specific transcript</i> Gen de RNA en el cromosoma x de mamíferos placentados que participa en el proceso de lionización.
Y RNA	RNA componente de la partícula ribonucleoproteica Ro60. Esta partícula es atacada por el sistema inmune en pacientes con lupus eritematoso sistémico.

APÉNDICE B: ALGUNOS DETALLES ADICIONALES DEL PROGRAMA

Funciones y parámetros del paquete ViennaRNA

Los parámetros y funciones del programa pueden ser consultadas con detalle en el manual RNALib-2.1.7 (Ivo L. Hofacker et al., 2014). A continuación se nombran aquellas utilizadas en este trabajo:

RNA.fold(sequence, structure)

Produce la estructura secundaria de mínima energía. A partir de una secuencia de RNA y una estructura dada (opcional, si sólo se busca calcular el valor de la energía).

INPUT

sequence (str)	
	Secuencia de RNA, previamente leída del archivo .fasta por la herramienta SeqIO de Biopython.
structure (str)	= None
	No se uso información previa de estructuras secundarias en ningún caso

OUTPUT

structure (str)	
	Estructura secundaria de mínima energía, en notación de paréntesis y puntos
MFE (float)	
	Energía de la estructura en kcal/mol

RNA.subopt(sequence, structure, delta, file)

La función produce todas las estructuras secundarias en un rango delta ("delta" * 0.01 kcal/mol) del óptimo de energía. A partir de una secuencia de RNA y una estructura dada (opcional).

INPUT

sequence (str)	
	Secuencia de RNA, previamente leída del archivo .fasta por la herramienta SeqIO de Biopython.
structure (str)	= None

	No se uso información previa de estructuras secundarias en ningún caso
<code>delta (integer)</code>	= 308
	El valor delta = 3.08 kcal (ver Capítulo 4)
<code>file</code>	= None

OUTPUT

El programa entrega el resultado en un objeto llamado SOLUTION, del que se puede extraer la siguiente información:

<code>structures (str)</code>	Estructura secundaria de mínima energía, en notación de paréntesis y puntos
<code>MFEs (float)</code>	Energía de cada estructura en kcal/mol

RNA.hamming(structure1, structure2)

Produce la estructura secundaria de mínima energía. A partir de una secuencia de RNA y una estructura dada (opcional, si sólo se busca calcular el valor de la energía).

INPUT

<code>structure1 (str)</code>	Primera secuencia de RNA o estructura de paréntesis y puntos.
<code>structure2 (str)</code>	Segunda secuencia de RNA o estructura de paréntesis y puntos.

OUTPUT

<code>hammingd (int)</code>	La distancia de hamming entre ambas secuencias o estructuras
-----------------------------	--

RNA.pf_fold(sequence)

Calcula la Z de la función de partición para una secuencia de RNA

INPUT

sequence (str)	
----------------	--

Secuencia de RNA

OUTPUT

G (float)	
-----------	--

Regresa la energía libre de Gibbs del ensamble, donde $G = RT \cdot \log(Z)$ en kcal/mol. Para despejar la Z se usan los siguientes parámetros:

$R = 1.98717/1000$ (constante de los gases)
 $T = 273.15 + 37$ (temperatura en Kelvin)

RNA.random_string(length, symbols)

Produce una secuencia aleatoria de una longitud dada

INPUT

length (int)	
--------------	--

Longitud dada por una secuencia de RNA real, para generar una secuencia equivalente aleatoria

symbols (str)	= "AUCG"
---------------	----------

Bases del RNA

OUTPUT

rndsequence (str)	
-------------------	--

Secuencia de RNA generada aleatoriamente

Otros

Para generar las secuencias aleatorias con la misma composición de bases se usó la siguiente función del módulo `random`

```
''.join(random.sample(s,len(s)))
```

Mezcla los elementos de una secuencia dada

INPUT

<code>s (str)</code>	
	Secuencia original

OUTPUT

<code>rndsequence (str)</code>	
	Secuencia de RNA mezclada aleatoriamente

Para las pruebas estadísticas se usó el módulo de funciones estadísticas de SciPy, `scipy.stats`

```
scipy.stats.ttest_1samp(a, popmean, axis=0)
```

Calcula la estadística T para la media de un grupo. Es una prueba de dos colas para la hipótesis nula de que la media esperada de la muestra es igual a una media de la población.

INPUT

<code>a (array_like)</code>	
	Muestra observada para la familia de RNA
<code>popmean (float)</code>	
	Media esperada (obtenida a partir de las poblaciones generadas aleatoriamente)

OUTPUT

<code>statistic (float)</code>	
	Estadística T
<code>pvalue (float)</code>	
	Valor de p para dos colas

APÉNDICE C: DETALLE DE GRUPOS COMPARADOS

Selección

LUCA

Familia ¹	Longitud ²	Secuencias ³	Representación ⁴	Sel. ⁵	Justificación
RF00001	116.6	712	A, B, E	✓	
RF00177	1530.4		A, B, E	×	Excede longitud
CL00001					
RF00005	73.4	954	99.3% del clan A, B, E, V	✓	
RF00023	357.3	477	0.4% del clan B, E, V	×	Excede longitud
RF01849	367.9	111	0.1% del clan B, E	×	Excede longitud
RF01850	310.9	7	<0.0% del clan B	×	Poco representativo
RF01851	284.4	27	<0.0% del clan B, E, V	×	Poco representativo
RF01852	90.8	109	0.2% del clan A, B, E, V	×	Poco representativo
RF02544	78.0	11	<0.0% del clan E	×	Poco representativo
CL00002					
RF00009	313.6	116	10.6% del clan E	×	Excede longitud
RF00010	380.4	458	64.3% del clan B, E	×	Excede longitud
RF00011	367.0	114	13.0% del clan B, E	×	Excede longitud
RF00030	321.4	66	8.3% del clan E	×	Excede longitud
RF00373	312.0	70	3.6% del clan A, B	×	Excede longitud
RF01577	620.5	2	0.1% del clan E	×	Excede longitud
CL00003					
RF00169	85.8	261	44.1% del clan B (A, E)	✓	
RF01502	344.4	50	1.6% del clan E	×	Excede longitud
RF01570		3	<0.0% del clan E	×	Poco representativo, pocas secuencias
RF01854	193.9	92	8.4% del clan B (A, E)*	✓	
RF01855	293.9	92	5.9% del clan E	×	Excede longitud
RF01856	181.4	18	37.4% del clan	✓	

			E (A)		
RF01857	306.3	53	2.6% del clan	×	Excede longitud

TRANSFERENCIA HORIZONTAL

Familia ¹	Longitud ²	Secuencias ³	Representación ⁴	Sel. ⁵	Justificación
RF00059	107.4	109	A, B, E	✓	
RF00029	88.6	92	B, E	×	Interpretación problem.
RF00028	364.8	12	B, E, V	×	Excede longitud
RF01419	185.2	308	B, E	✓	
RF01118	108.0	347	B, E	×	Pseudonudo
CL00014					
RF01315	30.0	19	40.5% del clan B	✓	
RF01317	28.0	33	44.3% del clan B (E)	✓	
RF01327	29.0	5	9.8% del clan B	×	Está incompleta la estructura secundaria
RF01328	—	3	0.2% del clan A	×	Pocas secuencias, errores en la base
RF01338	25.0	5	2.7% del clan A	×	Se eligieron otros rep. del clan
RF01352	29.0	3	1.8% del clan B	×	Se eligieron otros rep. del clan
RF01379	36.8	4	0.7% del clan B	×	Se eligieron otros rep. del clan
CL00015					
RF01318	37.0	12	64.1% del clan B	✓	
RF01320	37.0	10	34.7% del clan A, B	✓	
RF01376	37.0	2	1.1% del clan B, E	×	Se eligieron otros rep. del clan
RF01377	37.0	2	0.1% del clan A	×	Se eligieron otros rep. del clan

Familia ⁶	Secuencias por dominio ⁷	
RF00059	A	16
	B	7195
	E	435
RF00029	B	1195
	E	12650
RF01419	B	138
	E	2063
RF01118	B	9585

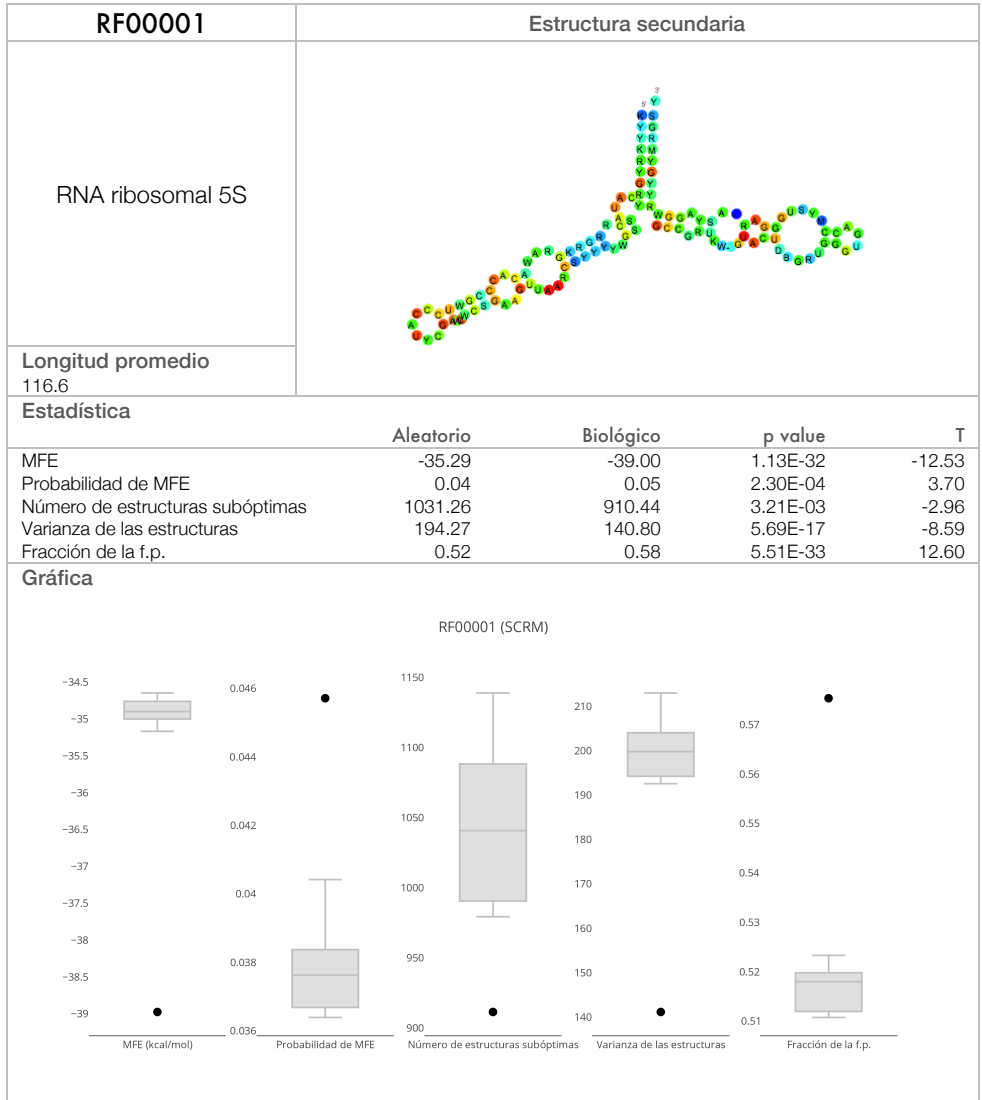
	E	3510
RF01317	B	733
	E	25
RF01320	A	9
	B	28
RF01376	B	2
	E	1

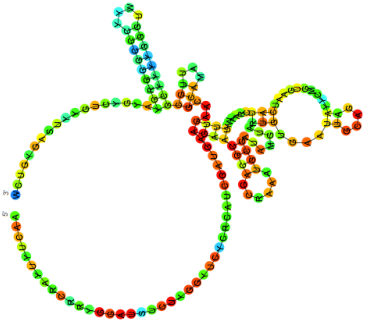
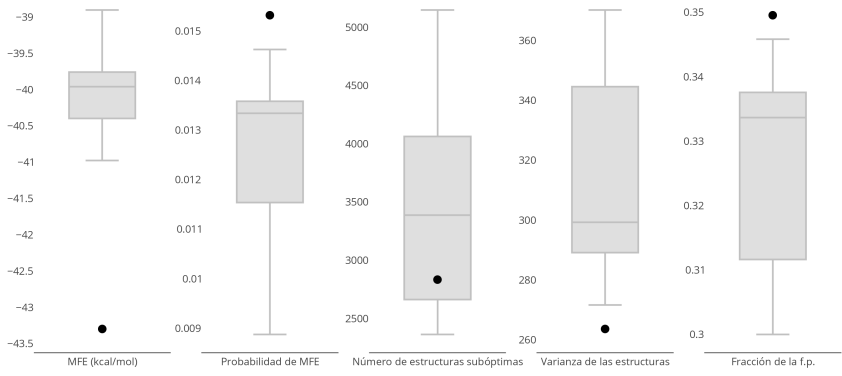
EXCLUSIVOS

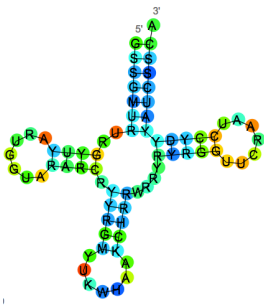





Familia ¹	Longitud ²	Tipo	Secuencias ³	Sel. ⁵	Justificación	
Archaea						
RF01125	—	snoRNA	3	×	No se podían elegir todos por que sería muy difícil analizar todos. Se eligieron algunos que se consideraban representativos de cada grupo.	
RF01138	—	snoRNA	6	×		
RF01147	—	snoRNA	5	×		
RF01141	—	snoRNA	5	×		
RF01152	55.7	snoRNA	12	✓		
RF01139	—	snoRNA	8	×		
RF01304	53.5	snoRNA	11	✓		
RF01310	—	snoRNA	5	×		
RF01145	—	snoRNA	8	×		
RF01150	—	snoRNA	8	×		
RF01140	—	snoRNA	6	×		
Eukarya						
RF00020	117.2	U5 spliceosomal RNA	180	✓		
RF00016	110.6	snoRNA	18	✓		
RF00012	209.0	snoRNA	87	×		
RF00069	77.9	snoRNA	14	×		
RF00002	152.2	5.8S ribosomal RNA	61	✓		
Bacteria						
RF00557	139.4	Ribosomal protein L10 leader	97	×		
RF00230	232.0	T-box leader	48	×		
RF00379	143.7	ydaO/yuaA leader	106	×		
RF00167	85.8	riboswitch	133	✓		
RF00380	168.7	ykoK leader	157	×		
RF00168	183.5	riboswitch	47	×		
RF00080	177.1	yybP-ykoY leader	29	×		
RF00442	108.1	ykkC-yxD leader	97	✓		
RF00050	136.3	riboswitch	144	×		

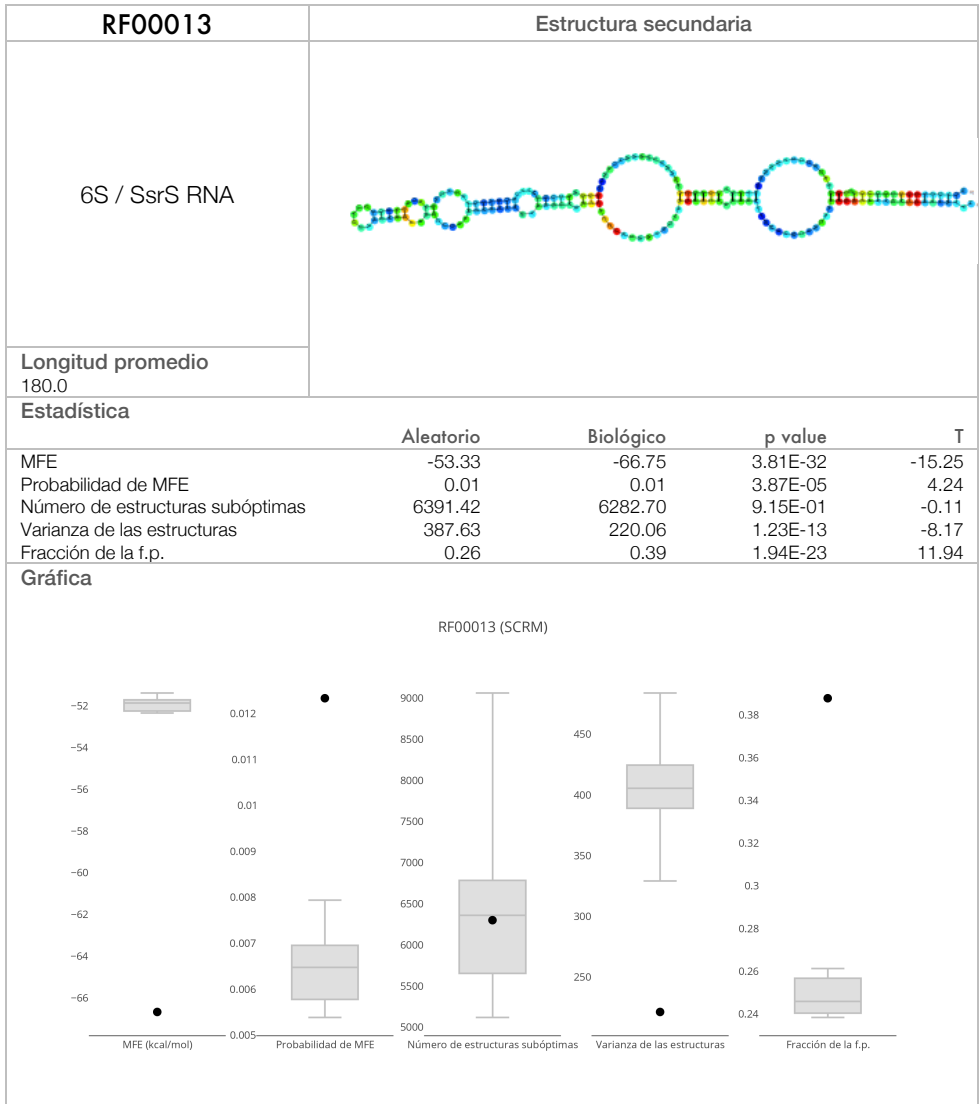
RF00504	100.9	riboswitch	44	×	
RF00174	202.9	riboswitch	430	×	
RF00013	180.0	SsrS RNA	149	✓	
RF01055	142.0	riboswitch	160	×	
Virus					
RF01768	62.0	ribosomal frameshift site	7	×	Forma pseudonudos
RF01833	55.0	ribosomal frameshift site	4	×	Forma pseudonudos
RF01835	44.9	ribosomal frameshift site	19	×	Forma pseudonudos

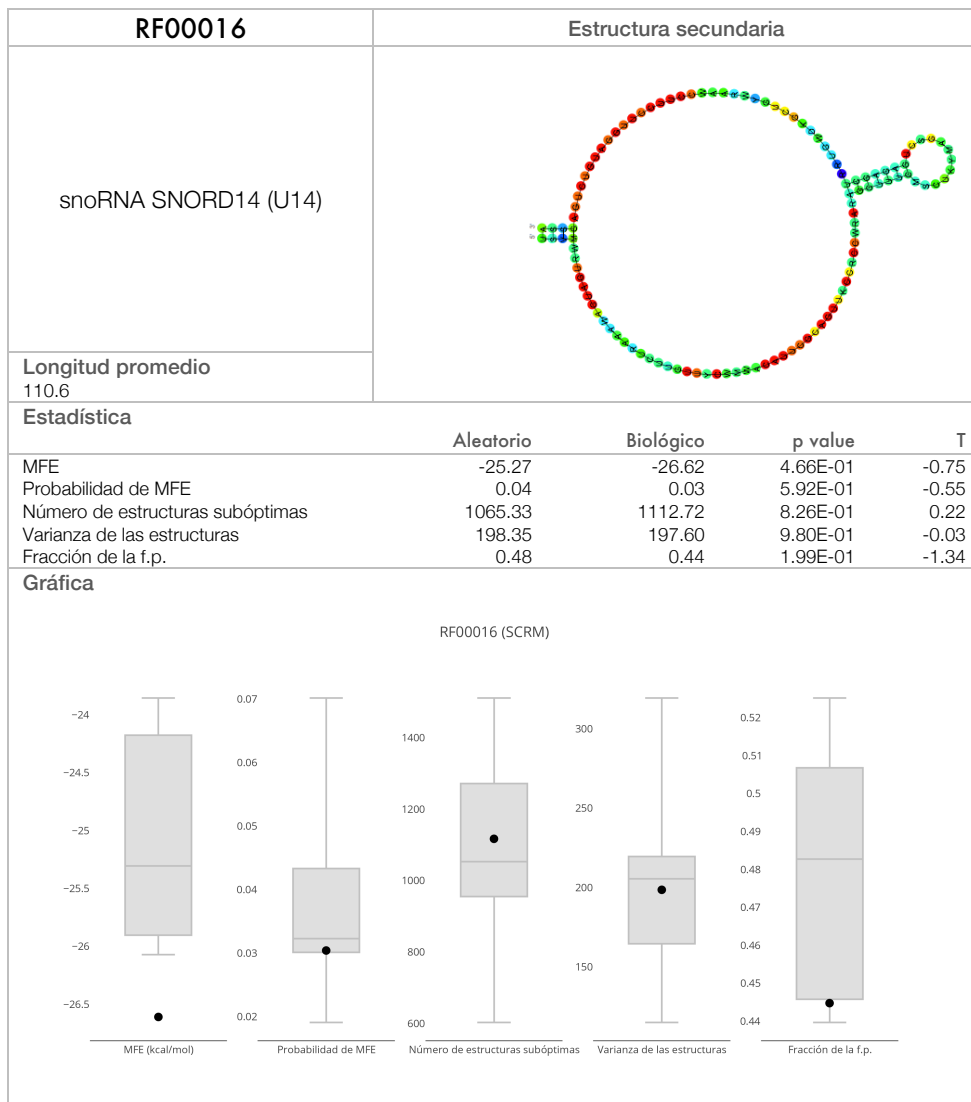
Resultados completos

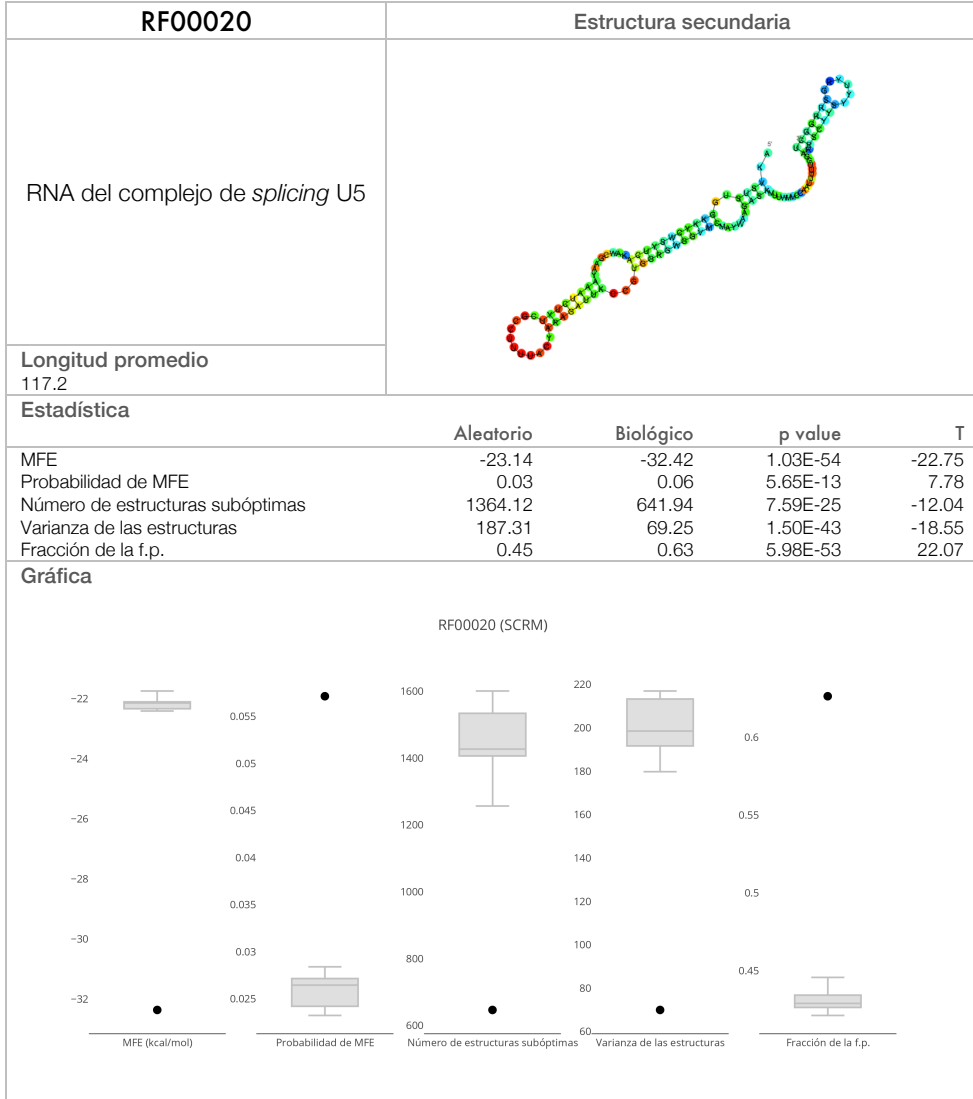


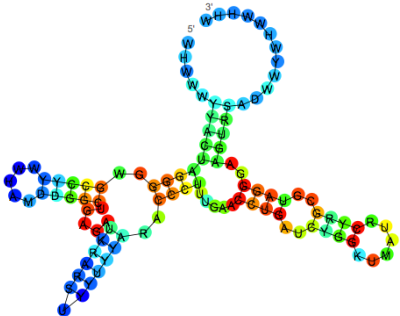
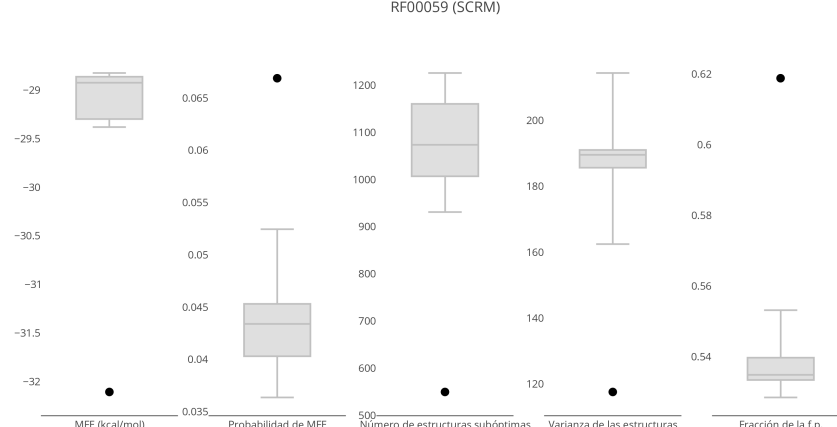
RF00002	Estructura secundaria				
RNA ribosomal 5.8 S					
	Longitud promedio 152.2				
Estadística					
	Aleatorio	Biológico	p value	T	
MFE	-40.33	-43.32	2.47E-03	-3.16	
Probabilidad de MFE	0.01	0.02	4.65E-01	0.74	
Número de estructuras subóptimas	3405.16	2821.36	1.78E-01	-1.36	
Varianza de las estructuras	308.89	263.17	1.13E-01	-1.61	
Fracción de la f.p.	0.33	0.35	2.00E-01	1.29	
Gráfica					
RF00002 (SCRM)					
					

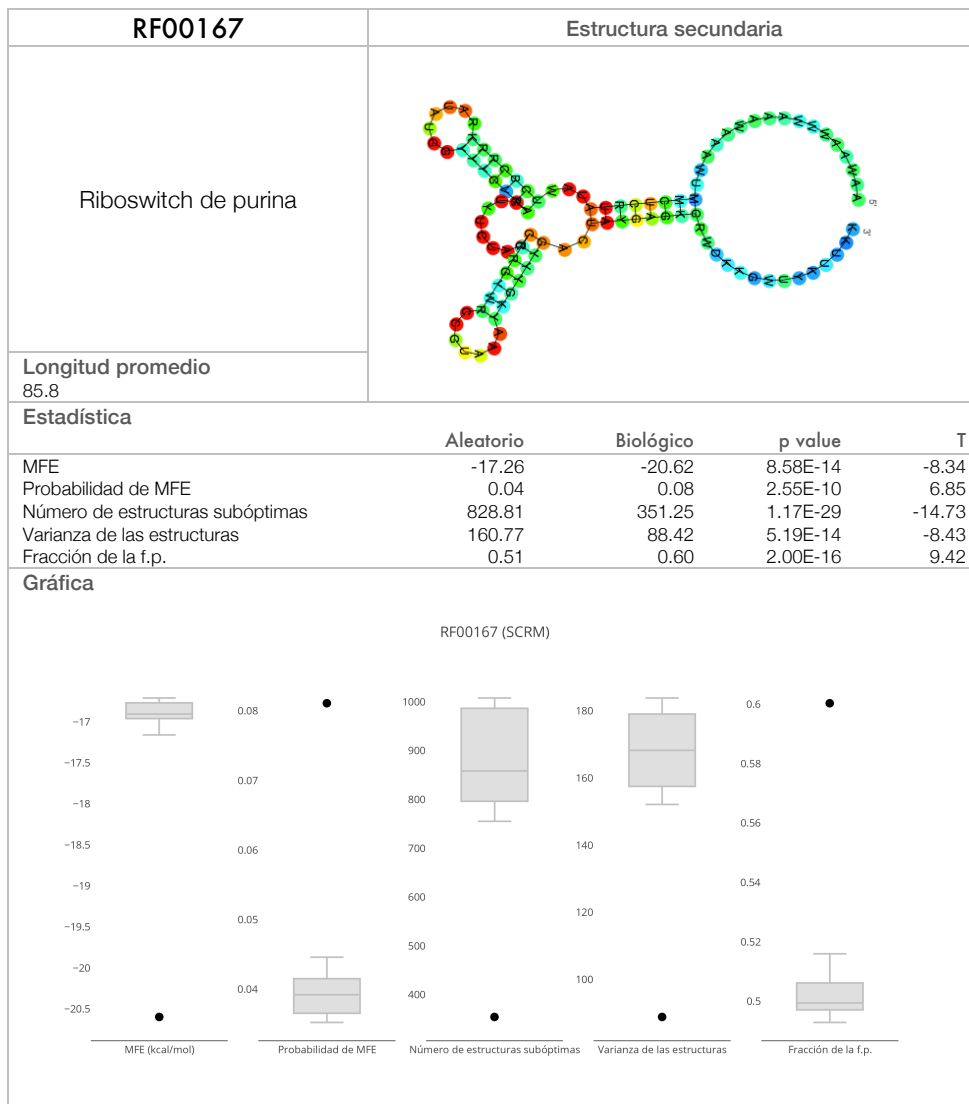
RF00005	Estructura secundaria				
RNA de transferencia					
Longitud promedio 73.4					
Estadística	Aleatorio	Biológico	p value	T	
MFE	-17.06	-22.00	2.03E-72	-19.65	
Probabilidad de MFE	0.11	0.16	5.60E-37	13.27	
Número de estructuras subóptimas	279.58	163.29	9.18E-78	-20.52	
Varianza de las estructuras	102.52	79.13	1.31E-31	-12.13	
Fracción de la f.p.	0.70	0.75	6.19E-61	17.72	
Gráfica					
RF00005 (SCRM)					
					
MFE (kcal/mol)	Probabilidad de MFE	Número de estructuras subóptimas	Varianza de las estructuras	Fracción de la f.p.	

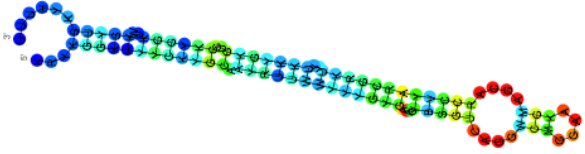
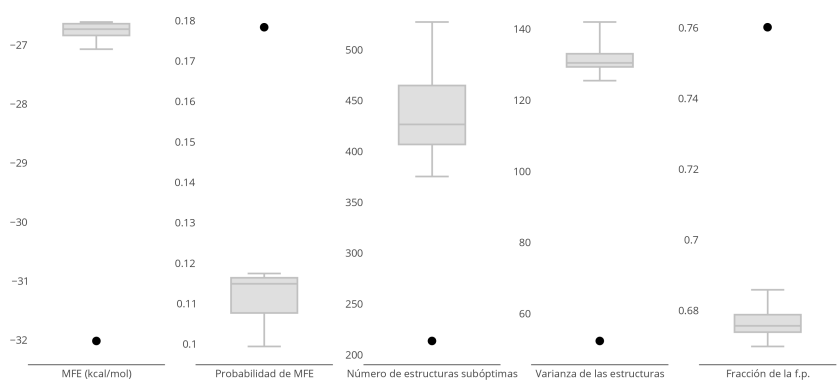


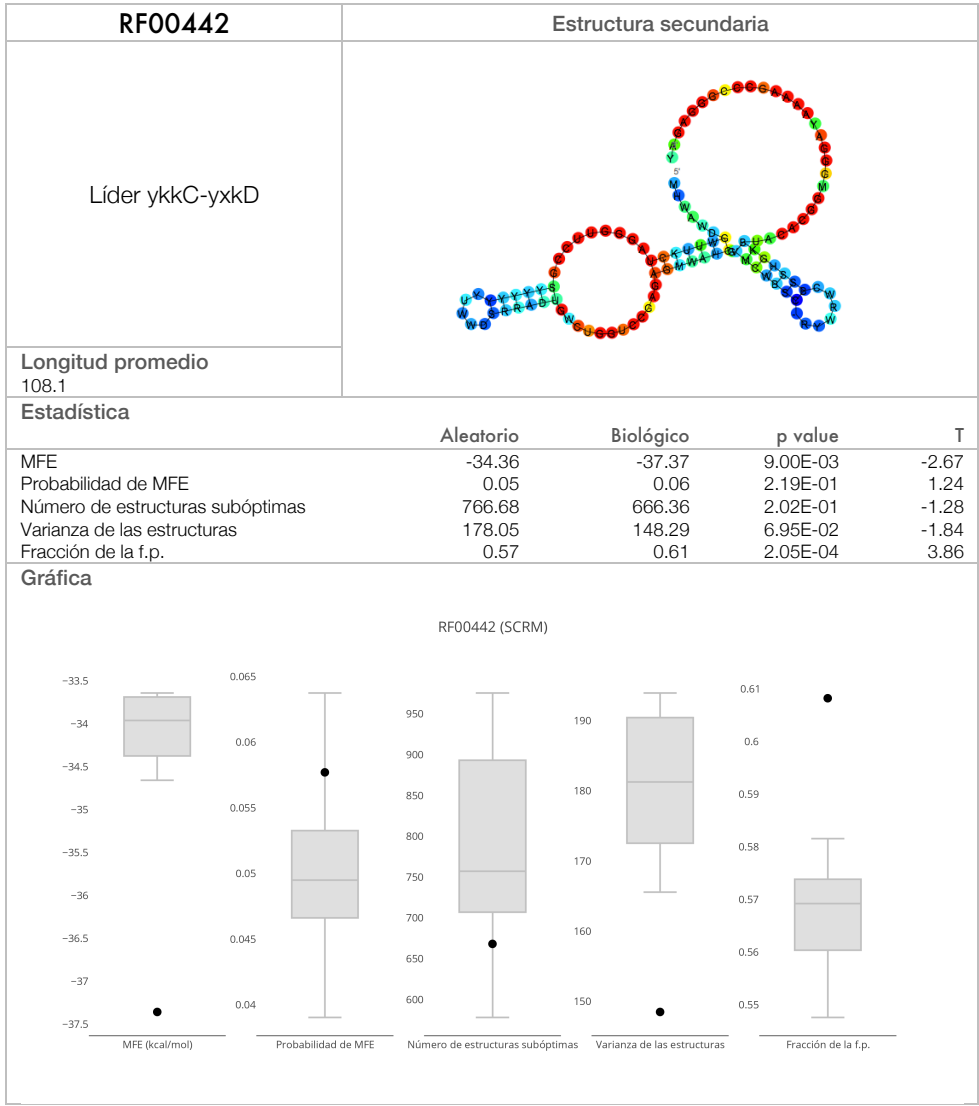


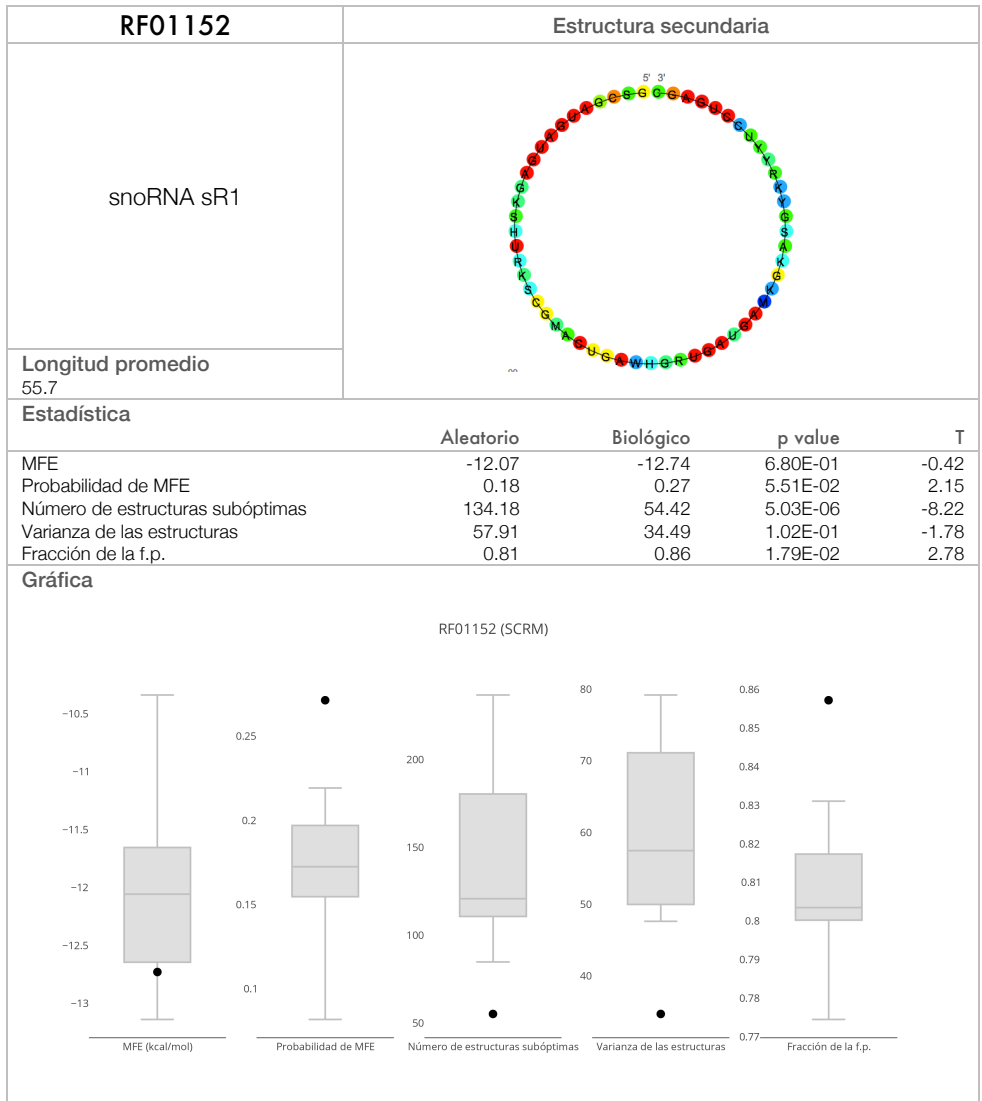


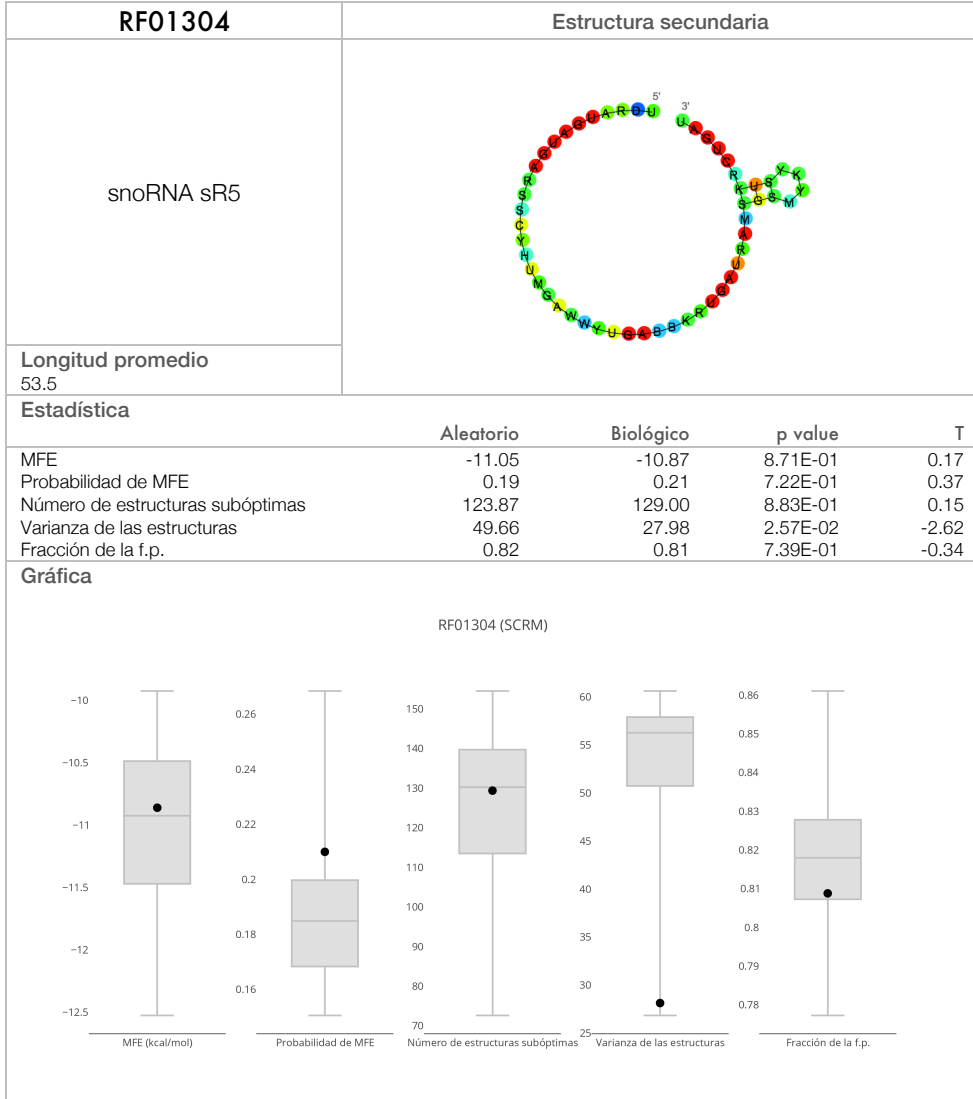
RF00059	Estructura secundaria				
Riboswitch TPP					
Longitud promedio 107.4					
Estadística	Aleatorio	Biológico	p value	T	
MFE	-29.33	-32.12	1.63E-03	-3.23	
Probabilidad de MFE	0.05	0.07	5.29E-04	3.57	
Número de estructuras subóptimas	1027.41	546.39	3.04E-09	-6.46	
Varianza de las estructuras	181.62	117.10	5.19E-09	-6.35	
Fracción de la f.p.	0.54	0.62	8.48E-12	7.66	
Gráfica	RF00059 (SCRM)				
					

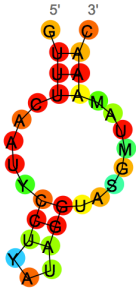
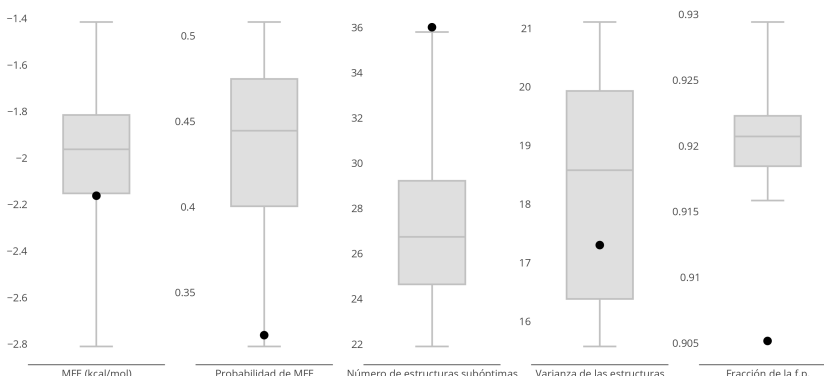


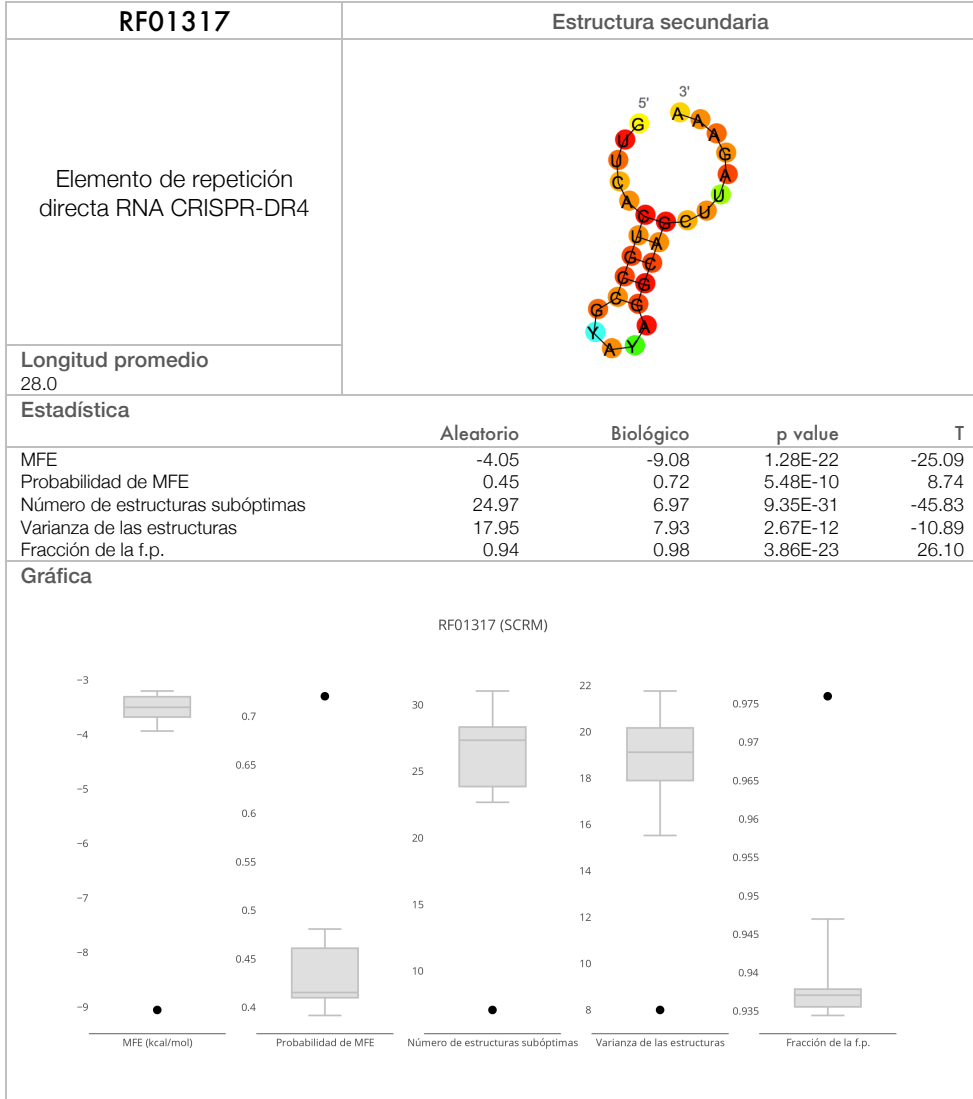
RF00169	Estructura secundaria				
<p>Partícula de reconocimiento de señal (SRP) pequeña bacteriana</p>					
<p>Longitud promedio 85.8</p>					
<p>Estadística</p>	Aleatorio	Biológico	p value	T	
MFE	-27.27	-32.04	2.42E-10	-6.59	
Probabilidad de MFE	0.12	0.18	4.07E-09	6.09	
Número de estructuras subóptimas	416.01	212.22	3.78E-20	-10.01	
Varianza de las estructuras	124.24	51.96	5.33E-41	-16.12	
Fracción de la f.p.	0.68	0.76	3.62E-29	12.73	
<p>Gráfica</p>	<p style="text-align: center;">RF00169 (SCRM)</p> 				

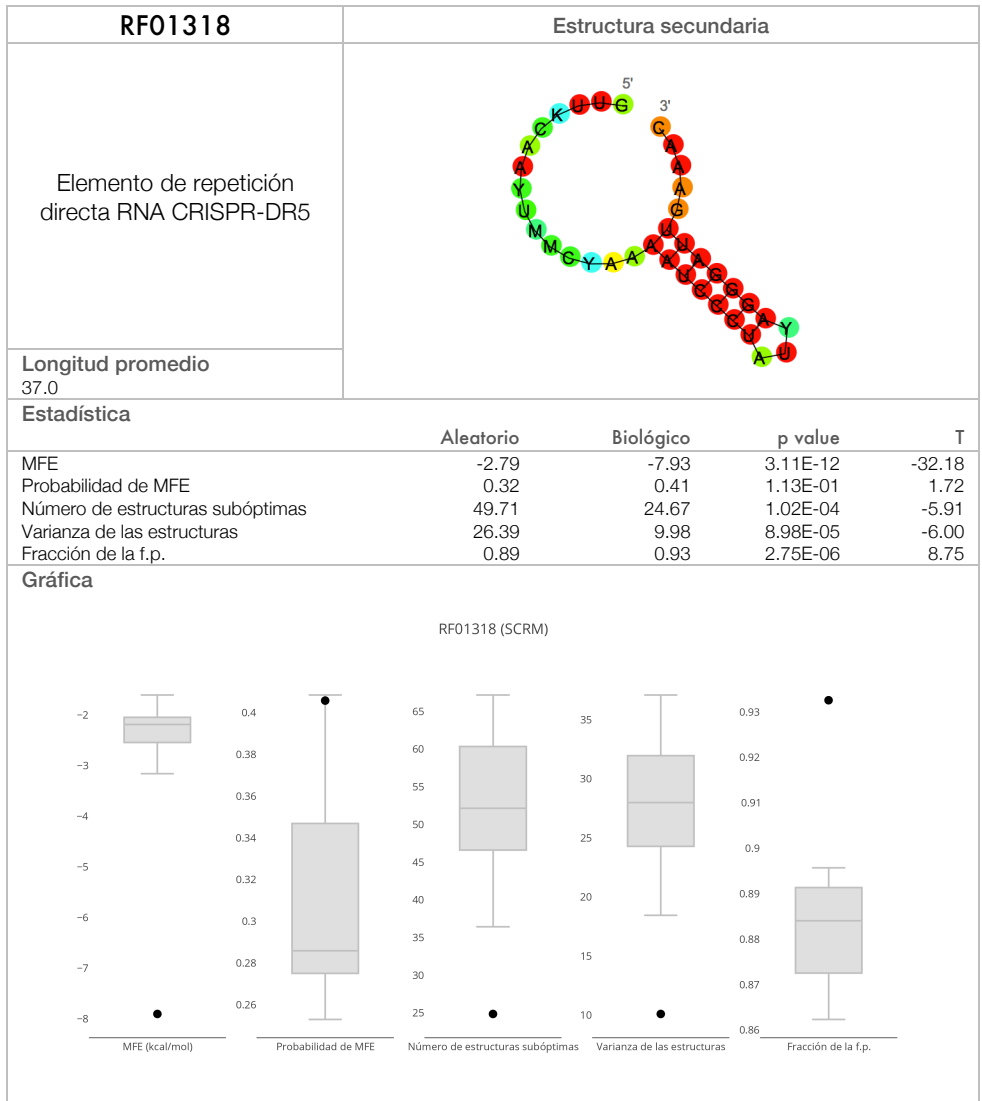


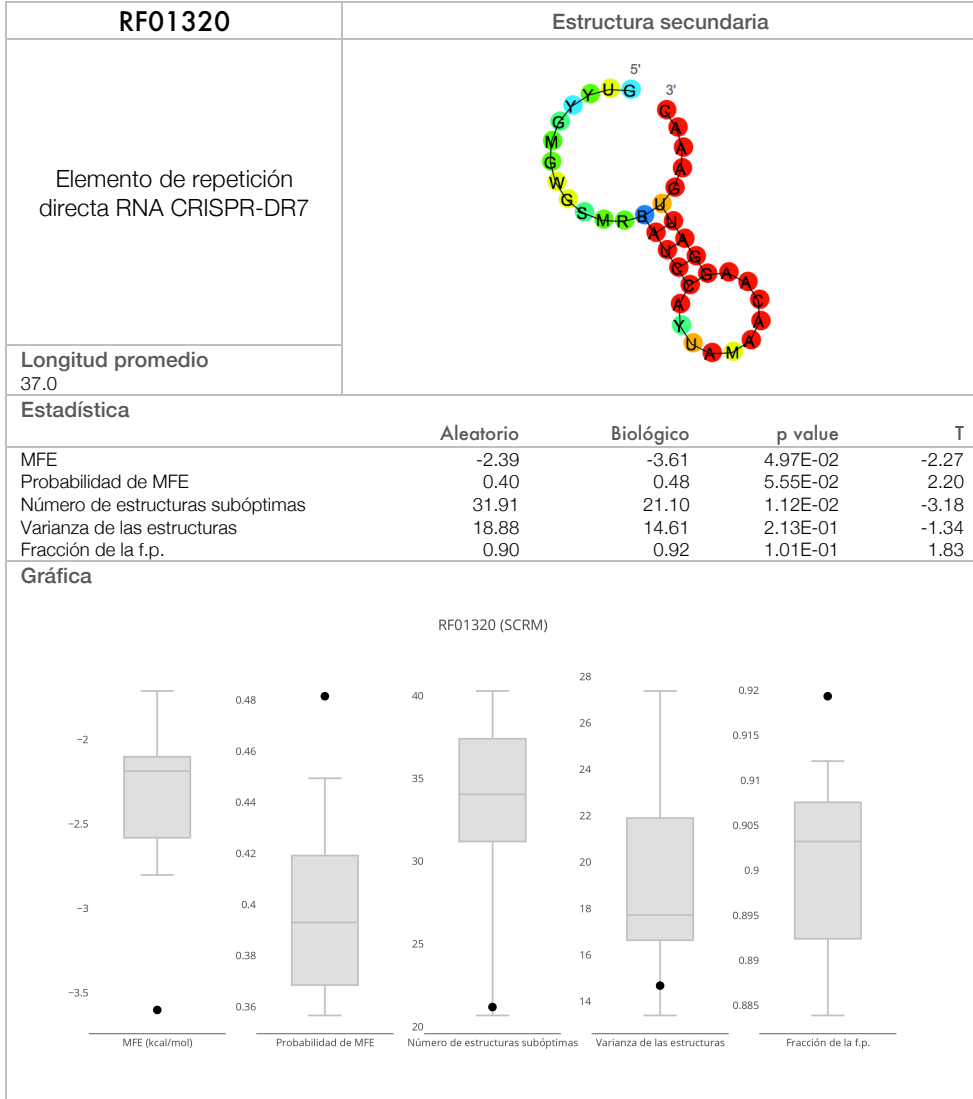


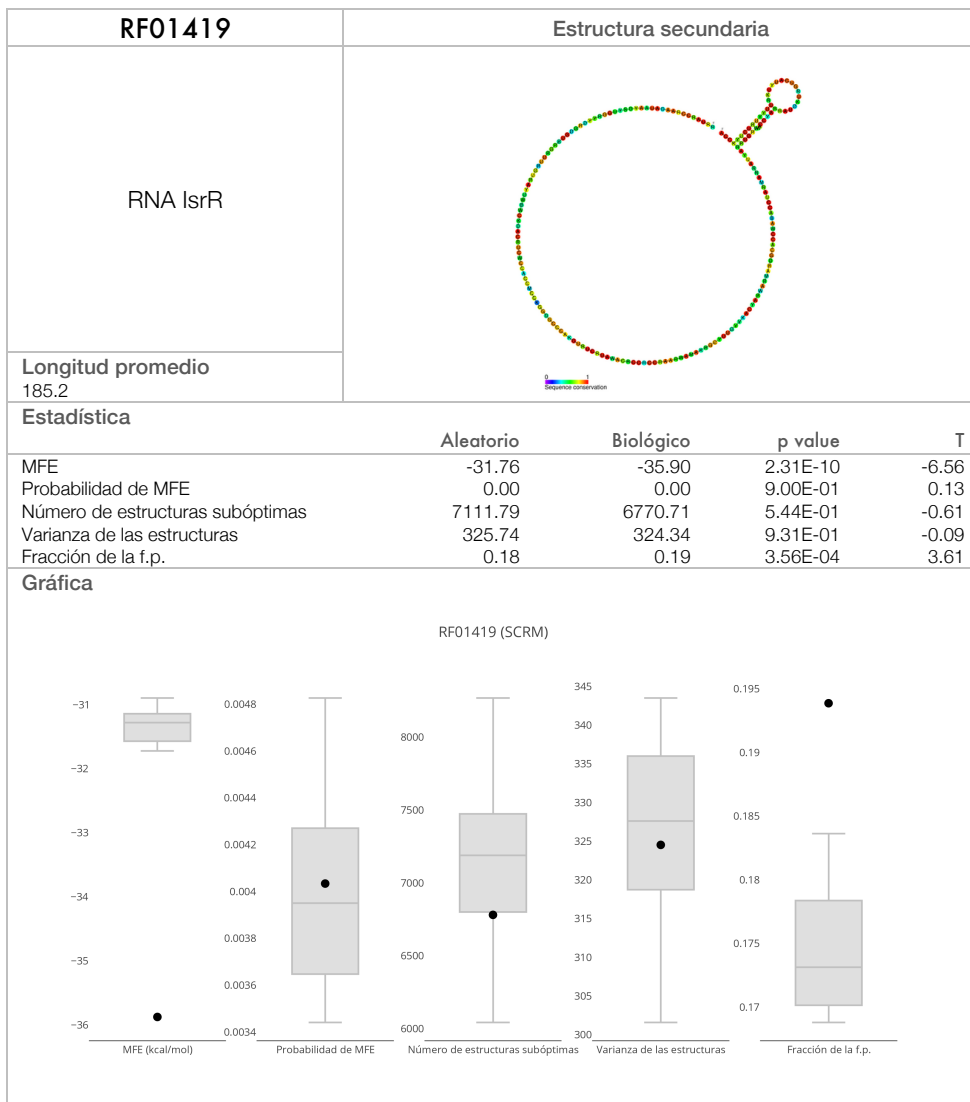


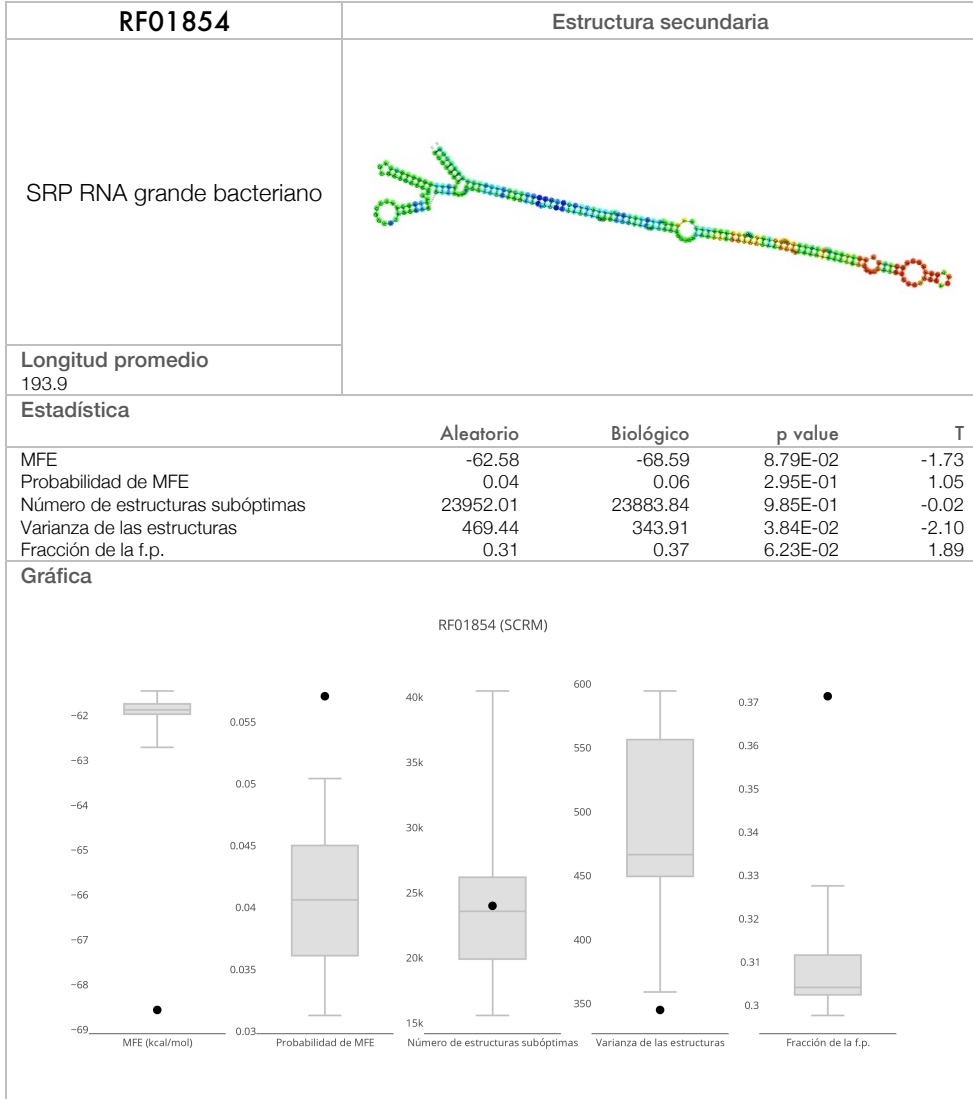
RF01315	Estructura secundaria				
Elemento de repetición directa RNA CRISPR-DR2					
Longitud promedio 30.0					
Estadística	Aleatorio	Biológico	p value	T	
MFE	-2.03	-2.17	6.69E-01	-0.43	
Probabilidad de MFE	0.42	0.32	1.44E-03	-3.76	
Número de estructuras subóptimas	28.12	35.95	6.86E-02	1.94	
Varianza de las estructuras	18.14	17.28	7.08E-01	-0.38	
Fracción de la f.p.	0.92	0.91	3.07E-02	-2.34	
Gráfica					
RF01315 (SCRM)					
					

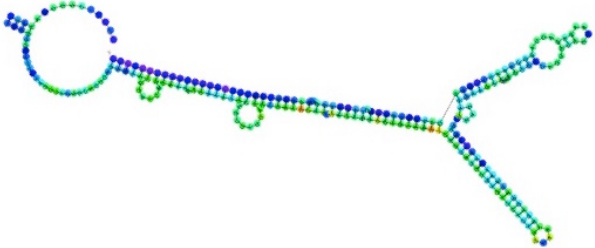
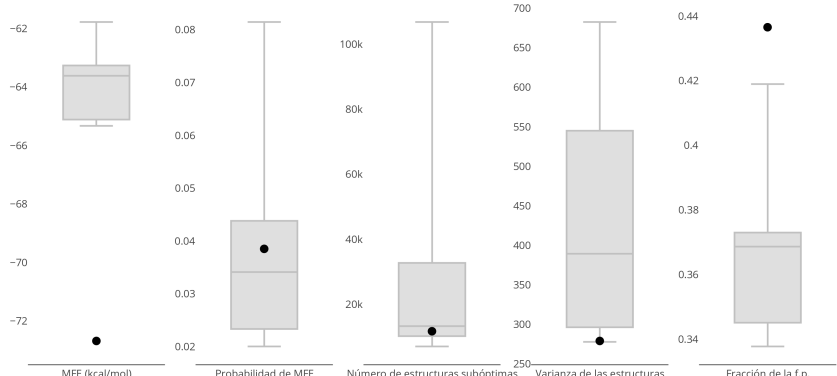










RF01856	Estructura secundaria																						
SRP RNA protozooario																							
Longitud promedio 181.4																							
Estadística	Aleatorio	Biológico	p value	T																			
MFE	-64.67	-72.73	3.79E-01	-0.90																			
Probabilidad de MFE	0.04	0.04	9.63E-01	0.05																			
Número de estructuras subóptimas	23838.12	11291.78	2.38E-02	-2.48																			
Varianza de las estructuras	410.75	277.17	2.32E-01	-1.24																			
Fracción de la f.p.	0.37	0.44	3.81E-01	0.90																			
Gráfica	RF01856 (SCRM)																						
 <p>The figure displays five box plots comparing random (Aleatorio) and biological (Biológico) distributions for various RNA structure metrics. The plots are arranged in two rows. The first row shows MFE (kcal/mol) and Probabilidad de MFE. The second row shows Número de estructuras subóptimas, Varianza de las estructuras, and Fracción de la f.p. Each plot has a y-axis representing the metric value and an x-axis with two categories: Aleatorio and Biológico. Outliers are indicated by black dots.</p> <table border="1"> <caption>Approximate data from the box plots</caption> <thead> <tr> <th>Metric</th> <th>Aleatorio (Random)</th> <th>Biológico (Biological)</th> </tr> </thead> <tbody> <tr> <td>MFE (kcal/mol)</td> <td>Median: ~-65, Range: ~-65 to -62</td> <td>Median: ~-70, Range: ~-72 to -68</td> </tr> <tr> <td>Probabilidad de MFE</td> <td>Median: ~0.035, Range: ~0.02 to 0.045</td> <td>Median: ~0.035, Range: ~0.02 to 0.045</td> </tr> <tr> <td>Número de estructuras subóptimas</td> <td>Median: ~15k, Range: ~10k to 35k</td> <td>Median: ~15k, Range: ~10k to 35k</td> </tr> <tr> <td>Varianza de las estructuras</td> <td>Median: ~380, Range: ~300 to 550</td> <td>Median: ~380, Range: ~300 to 550</td> </tr> <tr> <td>Fracción de la f.p.</td> <td>Median: ~0.36, Range: ~0.34 to 0.42</td> <td>Median: ~0.36, Range: ~0.34 to 0.42</td> </tr> </tbody> </table>						Metric	Aleatorio (Random)	Biológico (Biological)	MFE (kcal/mol)	Median: ~-65, Range: ~-65 to -62	Median: ~-70, Range: ~-72 to -68	Probabilidad de MFE	Median: ~0.035, Range: ~0.02 to 0.045	Median: ~0.035, Range: ~0.02 to 0.045	Número de estructuras subóptimas	Median: ~15k, Range: ~10k to 35k	Median: ~15k, Range: ~10k to 35k	Varianza de las estructuras	Median: ~380, Range: ~300 to 550	Median: ~380, Range: ~300 to 550	Fracción de la f.p.	Median: ~0.36, Range: ~0.34 to 0.42	Median: ~0.36, Range: ~0.34 to 0.42
Metric	Aleatorio (Random)	Biológico (Biological)																					
MFE (kcal/mol)	Median: ~-65, Range: ~-65 to -62	Median: ~-70, Range: ~-72 to -68																					
Probabilidad de MFE	Median: ~0.035, Range: ~0.02 to 0.045	Median: ~0.035, Range: ~0.02 to 0.045																					
Número de estructuras subóptimas	Median: ~15k, Range: ~10k to 35k	Median: ~15k, Range: ~10k to 35k																					
Varianza de las estructuras	Median: ~380, Range: ~300 to 550	Median: ~380, Range: ~300 to 550																					
Fracción de la f.p.	Median: ~0.36, Range: ~0.34 to 0.42	Median: ~0.36, Range: ~0.34 to 0.42																					

REFERENCIAS

- Amaral, P. P., & Mattick, J. S. (2008). Noncoding RNA in development. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 19(7-8), 454–92. <http://doi.org/10.1007/s00335-008-9136-7>
- Ancel, L. W., & Fontana, W. (2000). Plasticity, Evolvability, and Modularity in RNA. *Journal of Experimental Zoology (Molecular and Developmental Evolution)*, 288, 242–283.
- Arnott, S. (2006). Historical article: DNA polymorphism and the early history of the double helix. *Trends in Biochemical Sciences*, 31(6), 349–354. <http://doi.org/10.1016/j.tibs.2006.04.004>
- Baltimore, D. (1971). Expression of Animal Virus Genomes. *Bacteriological Reviews*, 35(3), 235–241.
- Batey, R. T. (2012). Structure and mechanism of purine binding riboswitches. *Quarterly Reviews of Biophysics*, 45(3), 345–381. <http://doi.org/10.1017/S0033583512000078>.Structure
- Breaker, R. R. (2012). Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(2), 1–15.
- Brown, T. (2008). *Genomas*. Ed. Médica Panamericana.
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., ... Bateman, a. (2013). Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research*, 41(D1), D226–D232. <http://doi.org/10.1093/nar/gks1005>
- Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D’Souza, L. M., Du, Y., ... Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3(2).
- Cech, T. R. (2011). The RNA Worlds in Context. In *RNA Worlds: From Life’s Origins to Diversity in Gene Regulation* (pp. 1–5).
- Chan, C. Y., & Ding, Y. (2008). Boltzmann ensemble features of RNA secondary structures: a comparative analysis of biological RNA sequences and random shuffles. *Mathematical Biology*, 56, 93–105. <http://doi.org/10.1007/s00285-007-0129-z>
- Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227, 561–563.
- Eddy, S. R. (2001). Non-Coding RNA Genes and the Modern RNA World. *Nature Reviews Genetics*, 2, 919–929.
- Eddy, S. R., & Hughes, H. (2004). How do RNA folding algorithms work ?, 22(11), 1457–1458.
- Elliot, D., & Ladomery, M. (2011). *Molecular Biology of RNA*. Gosporty, Hampshire: Oxford University Press.
- Fontana, W. (2002). Modelling “evo-devo” with RNA. *BioEssays: News and Reviews in*

- Molecular, Cellular and Developmental Biology*, 24(12), 1164–77.
<http://doi.org/10.1002/bies.10190>
- Futuyma, D. J. (2005). *Evolution*. Sunderland, Massachusetts: Sinauer Associates, Inc.
- Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., ... Bateman, a. (2011). Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Research*, 39(Database), D141–D145. <http://doi.org/10.1093/nar/gkq1129>
- Gaspin, C., Cavallé, J., Erauso, G., & Bachelier, J. (2000). Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: lessons from the Pyrococcus genomes. *Journal Of Molecular Biology*, 297(4), 895–906.
- Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394–407. <http://doi.org/10.1111/j.1365-2435.2007.01283.x>
- Gilbert, W. (1986). The RNA world. *Nature*, 319, 618.
- Hendrix, D. K., Brenner, S. E., & Holbrook, S. R. (2005). RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysics*, 38, 221–243. <http://doi.org/10.1017/S0033583506004215>
- Hermann, T., & Westhof, E. (1999). Non-Watson-Crick base pairs in RNA-protein recognition. *Chemistry & Biology*, 6(12), 335–343.
- Hinz, M., Moore, M. J., & Bindereif, A. (1996). Domain Analysis of Human U5 RNA. *The Journal of Biological Chemistry*, 271(31), 19001–19007.
- Hoepfner, M. P., Gardner, P. P., & Poole, A. M. (2012). Comparative Analysis of RNA Families Reveals Distinct Repertoires for Each Domain of Life. *PLoS Computational Biology*, 8(11), 1–8. <http://doi.org/10.1371/journal.pcbi.1002752>
- Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, 31(13), 3429–3431. <http://doi.org/10.1093/nar/gkg599>
- Hofacker, I. L., Stadler, P. F., Lorenz, R., & Al., (et. (2014). *RNAlib-2.1.7*.
- Hofacker, I. L., Stadler, P. F., Lorenz, R., & Al., (et. (2016). *RNAlib-2.2.3*.
- Leontis, N. B., Lescoute, A., & Westhof, E. (2006). The building blocks and motifs of RNA architecture. *Current Opinion in Structural Biology*, 16, 279–287. <http://doi.org/10.1016/j.sbi.2006.05.009>
- Leontis, N. B., & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *Rna*, 7(4), 499–512. <http://doi.org/10.1017/S1355838201002515>
- Lewin, B. (2011). *Lewin's genes X*. (J. E. Krebs, S. T. Kilpatrick, & E. S. Goldstein, Eds.) (10th ed.). Jones and Bartlett Publishers.
- Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*: *AMB*, 6, 26. <http://doi.org/10.1186/1748-7188-6-26>
- McCaskill, J. S. (1990). The Equilibrium Partition Function and Base Pair Binding Probabilities for RNA Secondary Structure. *Biopolymers*, 29, 1105–1119.
- Miklós, I. (2010). *Computational Biology*. (D. Fenyő, Ed.). Humana Press.
- Miklós, I., Meyer, I. M., & Nagy, B. (2005). Moments of the Boltzmann distribution for RNA secondary structures. *Bulletin of Mathematical Biology*, 67, 1031–1047. <http://doi.org/10.1016/j.bulm.2004.12.003>
- Moore, P. B. (1999). Structural Motifs in RNA. *Annual Review of Biochemistry*, 68(1), 287–300.

- Nakamura, Y., Ishiguro, A., & Miyakawa, S. (2012). RNA plasticity and selectivity applicable to therapeutics and novel biosensor development. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms*, 17(5), 344–64. <http://doi.org/10.1111/j.1365-2443.2012.01596.x>
- Nawrocki, E., & Eddy, S. (2014). *INFERNAL User's Guide*.
- Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., ... Finn, R. D. (2014). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, 43(D1), D130–D137. <http://doi.org/10.1093/nar/gku1063>
- Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinforma*, 29(22), 2933–2935. <http://doi.org/10.1093/bioinformatics/btt509>
- Nelson, D. A., & Cox, M. M. (2008). *Lehninger Principles of Biochemistry* (5th ed.). New York: W.H. Freeman and Company.
- Owen, E., & Daintith, E. (2004). *Dictionary of Evolutionary Biology*. New York: Facts On File, Inc.
- Pigliucci, M. (2001). *Phenotypic Plasticity, Beyond Nature and Nurture*. (S. M. Schenider, Ed.). The Johns Hopkins University Press.
- Pigliucci, M. (2005). Evolution of phenotypic plasticity: where are we going now? *Trends in Ecology & Evolution*, 20(9), 481–6. <http://doi.org/10.1016/j.tree.2005.06.001>
- Pigliucci, M. (2009). An extended synthesis for evolutionary biology. *Annals of the New York Academy of Sciences*, 1168, 218–28. <http://doi.org/10.1111/j.1749-6632.2009.04578.x>
- Rosenblad, M. A., Zweib, C., & Samuelsson, T. (2009). Kinship in the SRP RNA family. *RNA Biology*, 6(5), 508–516. <http://doi.org/10.4161/rna.6.5.9753>
- Schattner, P. (2002). Searching for RNA genes using base-composition statistics. *Nucleic*, 30(9), 2076–2082.
- Schuster, P., & Fontana, W. (1998). Chance and Necessity in Evolution: Lessons from RNA, (February 2008), 37. [http://doi.org/10.1016/S0167-2789\(99\)00076-7](http://doi.org/10.1016/S0167-2789(99)00076-7)
- Shapiro, J. a. (2009). Revisiting the central dogma in the 21st century. *Annals of the New York Academy of Sciences*, 1178, 6–28. <http://doi.org/10.1111/j.1749-6632.2009.04990.x>
- Stowe, K. (2007). *An Introduction to Thermodynamics and Statistical Mechanics* (2nd ed.). Cambridge University Press.
- Swendsen, R. H. (2012). *An Introduction to Statistical Mechanics and Thermodynamics*. Oxford University Press.
- Szweykowska-kulińska, Z. (2014). The Old and New RNA World Abiotic synthesis of RNA components and RNA First RNA-encoded protein syntheses, 83(4), 441–448. <http://doi.org/10.5586/asbp.2014.035>
- Tian, B., Bevilacqua, P. C., Diegelman-Parente, A., & Mathews, M. B. (2004). The Double-Stranded-RNA-Binding Motif: Interference and Much More. *Nature Reviews Molecular Cell Biology*, 5, 1013–1023. <http://doi.org/10.1038/nrm1528>
- Tinoco, I. J., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M., & Gralla, J. (1973). Improved Estimation of Secondary Structure in Ribonucleic-Acids. *Nature-New Biology*, 246(150), 40–41. <http://doi.org/10.1038/newbio246040a0>
- Tinoco, I. J., Uhlenbeck, O. C., & Levine, M. D. (1971). Estimation of Secondary Structure in Ribonucleic Acids. *Nature*, 230, 362–367.
- Varani, G., & McClain, W. H. (2000). The G-U wobble base pair diverse biological systems.

- EMBO Reports*, 1(1), 18–23.
- Wake, T. (2003). *RNA Secondary Structure Prediction Using Applications of the Partition Function*. Boston College.
- Watson, J. D., & Crick, F. H. . (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, (171), 737 – 738. Retrieved from <http://jama.jamanetwork.com/article.aspx?articleid=405512>
- Westhof, E., & Fritsch, V. (2000). RNA folding: beyond Watson–Crick pairs. *Structure*, 8(3), R55–R65. [http://doi.org/10.1016/S0969-2126\(00\)00112-X](http://doi.org/10.1016/S0969-2126(00)00112-X)
- Wuchty, S., Fontana, W., Hofacker, I. L., & Schuster, P. (1999). Complete Suboptimal Folding of RNA and the Stability of Secondary Structures. *Biopolymers*, 49, 145–165.
- Zuker, M., & Sankoff, D. (1984). RNA Secondary Structures and their Prediction. *Bulletin of Mathematical Biology*, 46(4), 591–621.
- Zuker, M., & Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1), 133–148.

Sitios web, bases de datos y software

- Chapman BA and Chang JT. *Biopython: Python tools for computational biology*. ACM SIGBIO Newsletter 2000 Aug; 20, 15-19.
- Darty, K., Denise, A., & Ponty, Y. (2009). *VARNA*: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15), 1974–1975
- EMBL-EBI. (2014, July). *Rfam 12.0*. Retrieved February, 2016, from <http://rfam.xfam.org>
- Lorenz, Ronny and Bernhart, Stephan H. and Höner zu Siederdisen, Christian and Tafer, Hakim and Flamm, Christoph and Stadler, Peter F. and Hofacker, Ivo L. *ViennaRNA Package 2.0 Algorithms for Molecular Biology*, 6:1 26, 201
- Plotly Technologies Inc.*, Collaborative data science, 2015 en <https://plot.ly>
- Protein Data Bank, & Goodsell, D. S. (2001, March). Phenylalanine transfer RNA [PDB: Molecule of the Month]. Retrieved from <http://pdb101.rcsb.org/motm/15>
- Python Software Foundation. *Python Language Reference*, version 2.7.
- Real Academia Española. (2014). plástico. En *Diccionario de la lengua española* (23.a ed.). Recuperado de <http://dle.rae.es/?id=TLksLOY>
- Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. *The NumPy Array: A Structure for Efficient Numerical Computation*, *Computing in Science & Engineering*, 13, 22-30 (2011)
- Wes McKinney. *PANDAS*, Data Structures for Statistical Computing in Python, *Proceedings of the 9th Python in Science Conference*, 51-56 (2010)
- WikiProject RNA* en Wikipedia. Consultado en 2016, de https://en.wikipedia.org/wiki/Wikipedia:WikiProject_RNA
- Williams, L. (2012, December 20). [From "NASA Astrobiology: Yin and Yang: Polypeptide and Polynucleotide"]
- Yikrazuul. (2010, May 19). Estructura secundaria de ARN de transferencia: TRNAPhe de *S. cerevisiae* [Wikimedia Commons under CC BY-SA 3.0]