



UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ
POSGRADO EN CIENCIAS INTERDISCIPLINARIAS



Plasticidad y evolución estructural de moléculas de RNA *in silico*

TESIS

REALIZADA Y PRESENTADA POR
Alexis Gabriel Cruz Moreno

PARA OBTENER EL GRADO DE
Maestro en Ciencias Interdisciplinarias

CON DIRECCIÓN DE
Dr. Carlos A. Espinosa Soto

COMITÉ EVALUADOR
Dra. Vanesa Olivares Illana
Dr. Enrique González Tovar
Dr. Edgardo Ugalde Saldaña

SAN LUIS POTOSÍ, SLP.

JUNIO 2026

Plasticidad y evolución estructural de moléculas de RNA *in silico*

© 2026 by Alexis Gabriel Cruz Moreno.
Licensed under CC BY-NC-SA 4.0.

To view a copy of this license, visit:
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Resumen

Biomoléculas como el RNA requieren adoptar determinada conformación para desempeñar cierta actividad. Esto refuerza la idea de que ‘estructura es igual a función’. Sin embargo, estas moléculas en realidad son capaces de adoptar un repertorio de diferentes estructuras. Esta capacidad de adoptar diversos fenotipos, sin que el cambio se deba a una modificación genética, se conoce como plasticidad estructural. Este estudio es motivado por la intuición de que este repertorio de posibles estructuras favorece la capacidad de las moléculas para evolucionar estructuras innovadoras.

Estudios previos intentaron modelar este comportamiento, resultando en un estancamiento evolutivo provocado por la aparición de estructuras sumamente robustas, tanto a perturbaciones genéticas como ambientales. Este anómalo resultado se denominó ‘confinamiento neutral’. En este proyecto, me propongo demostrar que moléculas de RNA, simuladas *in silico*, son capaces de acceder a una estructura específica τ , mediante evolución dirigida, más fácilmente cuando se considera la capacidad de plasticidad estructural. Para lo anterior, utilizo un novedoso modelo basado en la unión RNA-ligando para determinar la adecuación de cada molécula.

Abstract

In biology we often say that ‘structure equals function’, referring to the fact that biomolecules such as RNA need to adopt a certain conformation to perform a specific activity. However, these molecules can actually adopt a repertoire of different structures. This ability to adopt diverse phenotypes, without any genetic modification, is known as structural plasticity. This study is motivated by the intuition that this repertoire of possible structures promotes the molecules’ capacity to evolve innovative structures.

Previous studies attempted to model this behavior, resulting in an evolutionary stagnation or a dead end caused by the emergence of extremely robust structures, both to genetic and environmental perturbations. This anomalous result was called ‘neutral confinement’. In this project, we propose to demonstrate that RNA molecules, simulated *in silico*, can access a specific structure τ through directed evolution while considering their structural plasticity. This is achieved by employing a novel model based on RNA-ligand binding to determine the fitness of each molecule.

Distribución del contenido

1. Introducción	6
1.1. Plasticidad fenotípica en la Naturaleza	6
1.1.1. Camuflaje perfecto	7
1.1.2. Fotosensibilidad	7
1.1.3. Termosensibilidad	8
1.2. Plasticidad fenotípica en moléculas de RNA	9
1.2.1. La hipótesis del mundo de RNA	9
1.2.2. Niveles estructurales del RNA	10
1.2.3. Plasticidad estructural en el RNA	11
2. Planteamiento del Problema	13
2.1. Algoritmos evolutivos	13
2.1.1. Cálculo de adecuación	13
2.1.2. Selección	14
2.1.3. Mutación	14
2.1.4. Reemplazo poblacional	14
2.2. Modelos anteriores de plasticidad estructural	15
2.3. Confinamiento neutral	15
2.4. Modelo de unión a Ligando	16
2.5. Propósito de este proyecto	17
2.5.1. Objetivo General	17
2.5.2. Objetivos Específicos	18
3. Marco Metodológico	19
3.1. Predicción de estructuras secundarias	19
3.1.1. ¿Por qué usamos energía libre de Gibbs?	19
3.1.2. Probabilidad de plegamiento en una estructura	21
3.2. Representación de estructuras	22
3.2.1. Cálculo de distancia estructural	24
3.3. Criterios de adecuación	24
3.3.1. Modelo de Ancel y Fontana	24
3.3.2. Modelo de Unión a Ligando	25
3.4. Muestreo de red neutral de tRNA _{phe}	29

3.5. Exploración de estructuras accesibles	30
3.6. Simulaciones evolutivas	31
3.7. Análisis y pruebas estadísticas	32
3.8. Acceso al código desarrollado	33
4. Resultados	34
5. Discusión	44
5. Referencias	46
6. Apéndice	49

Introducción

Los seres vivos cambiamos de manera constante. Algunos de estos cambios son imperceptibles en el corto plazo, pero con el paso del tiempo se acumulan y dejan un rastro visible. Muchos de estos cambios en las características de un organismo se deben a la modificación de una especie de ‘manual interno’ presente en cada ser vivo, conocido como ‘genoma’.

El genoma está formado por el total de diversos bloques funcionales, llamados individualmente ‘gen’ y en agrupaciones ‘genes’. Dicho esto, se conoce como ‘genotipo’ al conjunto de genes cuya función está relacionada con el control de algún ‘fenotipo’ (característica, forma o comportamiento) del ser vivo. Debido a que el genotipo puede verse como un subconjunto del genoma, y este último es el mecanismo de herencia, muchos de los fenotipos de un organismo suelen heredarse a sus descendientes. Los cambios en las características de una población conforme transcurren las generaciones ocurren debido a modificaciones de los genotipos de cada individuo que conforma la población. Los fenotipos más benéficos para la población son los que tienden a aumentar con el tiempo de acuerdo con la teoría de la evolución por selección natural [1].

Considere que relación entre genotipo y fenotipo no es necesariamente de uno a uno. A partir de un genotipo es posible producir una diversidad de fenotipos sin la necesidad de alteraciones al gen o genes que lo regulan. Esta capacidad es conocida como ‘plasticidad fenotípica’. No obstante, si los cambios en el genotipo no son responsables de la variación en el fenotipo, ¿qué es lo que la produce? La respuesta está en diferentes factores ambientales o externos, razón por la que en diferentes casos, la variedad de fenotipos accesibles es fundamental para permitir a los seres vivos adaptarse en respuesta a su entorno. Para ilustrar lo anterior, conviene revisar algunos ejemplos de plasticidad fenotípica en diferentes seres vivos.

1.1. Plasticidad fenotípica en la Naturaleza

En la naturaleza podemos encontrar múltiples ejemplos de plasticidad fenotípica, tantos que se vuelve complicado clasificarlos. Considere que la manera en la que presento los siguientes ejemplos no sigue ninguna clasificación rigurosa. Expongo los diferentes casos de la forma que me parece más ilustrativa.

1.1.1. Camuflaje perfecto

Al hablar de adaptación al entorno pocas habilidades son tan buenas como la capacidad de diferentes cefalópodos de camuflarse con su entorno. La mayoría de esta clase de invertebrados son conocidos por modificar el color, patrón e incluso textura aparente de su piel. Esta capacidad permite mimetizarse con el ambiente a su alrededor.

En el caso de los pulpos por ejemplo, el camuflaje que adoptan depende principalmente de la capacidad del cefalópodo para identificar visualmente la textura, forma y profundidad de su entorno. El fenotipo que finalmente adopta es una combinación de diferentes estímulos provenientes de su entorno [2].

Figura 1. ¡Encuentra al Pulpo!

Los pulpos pueden cambiar la apariencia de su piel para camuflarse con el entorno.

Autor: Jérôme Paillet, Ifremer.

Licencia: CC BY 4.0.

Fuente: <https://image.ifremer.fr/data/00618/73015/hd/28887.jpg>



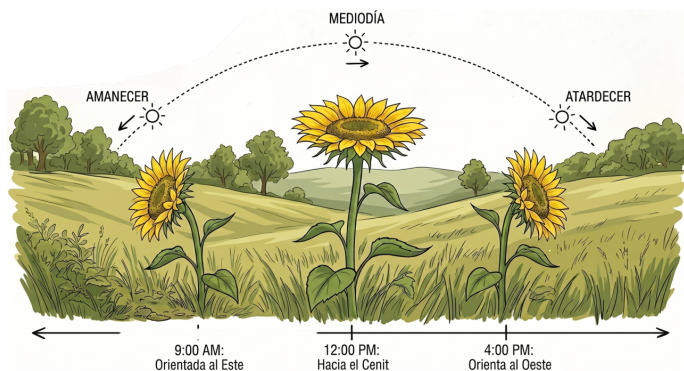
Parte del éxito de los pulpos para camuflarse está en que poseen células especializadas conocidas como cromatóforos. Estas células exhiben algún fenotipo dependiendo de una o más características de la luz, aunque no abordaré ningún caso particular. Este tipo celular no es exclusivo de pulpos, pero es un claro ejemplo de plasticidad fenotípica sensible a la luz o fotosensible [3].

1.1.2. Fotosensibilidad

Un ejemplo cotidiano de este tipo de plasticidad es el heliotropismo. El cambio de orientación que realizan algunas flores persiguiendo la posición del sol. En los girasoles (*Helianthus annuus*) esta respuesta fenotípica se ha asociado con aumento en la captación de luz UV, aumento en la temperatura de la superficie de la flor y mayor captación de organismos polinizadores como consecuencia [4].

Figura 2. Heliotropismo.

Los girasoles modifican su orientación conforme el sol se desplaza a lo largo del día. Imagen generada por Google Gemini bajo instrucción y edición del autor.



Otro ejemplo interesante es el de *Arabidopsis halleri*. En este caso, además de la luz del entorno, la humedad y temperatura del ambiente son factores que afectan la posición que las flores adoptan para resguardar su polen y/o algún polinizador durante la lluvia [5].

Figura 3. Protección temporal. Ilustra la posición que adopta *Arabidopsis halleri* bajo ciertas condiciones del ambiente. Imagen generada por Google Gemini bajo instrucción y edición del autor.

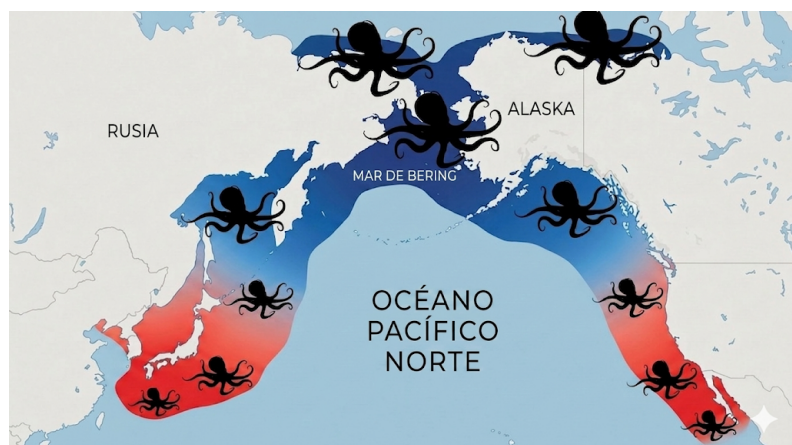


1.1.3. Termosensibilidad

La temperatura del ambiente es otro factor que puede influenciar el fenotipo observado. Tal es el caso de *Enteroctopus dofleini*, conocido como el pulpo gigante del Pacífico. Es la especie de pulpo con mayor tamaño reportado y su hábitat se extiende por gran parte del Océano Pacífico Norte.

En el océano las aguas cercanas al ecuador tienden a ser más cálidas que las próximas a los polos. Por esta razón *Enteroctopus dofleini* se encuentra expuesta a diferentes temperaturas dependiendo de la región en la que habite. Los especímenes que crecen en aguas frías, con temperaturas entre 7-9.5 °C expresan mayor crecimiento corporal frente los que habitan en aguas más cálidas [6]. Esto es apoyado por estudios en los que miden atributos de especímenes capturados en diferentes regiones. Siendo más grandes y pesados los pulpos capturados en las zonas más frías como Alaska y el Mar de Bering [7, 8, 9].

Figura 4. Temperatura y tamaño. Hábitat de *Enteroctopus dofleini*. Azul intenso representa aguas con menor temperatura, rojo las más cálidas. El tamaño de los especímenes se reduce conforme la temperatura incrementa. Imagen generada por Google Gemini bajo instrucción y edición del autor.



1.2. Plasticidad fenotípica en moléculas de RNA

En los ejemplos que hemos abordado es fácil intuir que ventaja brinda la plasticidad fenotípica. Para que esta ventaja tenga sentido en un contexto evolutivo, debe encontrarse codificada genéticamente. Es decir que debe existir un genotipo o red de genes específica que controla el fenotipo y modula su sensibilidad ambiental [10].

Si caracterizamos la relación genotipo-fenotipo(s) en un genoma completo, obtendríamos lo que conocemos como un mapa genotipo-fenotipo. Realizar esto es una ardua labor debido a la cantidad de diferentes interacciones y mecanismos que rigen el funcionamiento de los organismos [10]. Una propuesta más adecuada es centrarnos en el estudio de la plasticidad fenotípica en escenarios más simples y fundamentales.

1.2.1. La hipótesis del mundo de RNA

Imaginemos una etapa temprana de nuestro planeta, momento en el que solo coexisten diversas moléculas simples. No supondremos qué son, sólo que son moléculas sencillas flotando en algún ambiente primigenio. De vez en cuando, algunas de estas moléculas reciben un exceso de energía proveniente de alguna fuente de calor cercana o de rayos UV. Este aporte de energía provoca que interactúen entre sí. Algunas de estas interacciones son momentáneas y no producen nada nuevo, pero otras dan lugar a moléculas de mayor complejidad que se mantienen durante más tiempo [11].

Debido a que las conformaciones más estables se mantienen durante mayor tiempo, serán cada vez más frecuentes en el ambiente. Esto establece algo fundamental, que las composiciones inestables se disocian y desaparecen, mientras que las estructuras estables prevalecen en el tiempo. Eventualmente, cuando las configuraciones más complejas sean suficientemente abundantes, competirán en un proceso similar ensamblándose en compuestos cada vez más complejos [11, 12].

El surgimiento de moléculas cada vez más complejas continúa indefinidamente, originando en algún momento una conformación cuya particularidad lo convierte en ‘el primer replicador’. Una conformación capaz de reclutar elementos del entorno y utilizarlos para construir una copia de sí mismo. Sin embargo, este replicador de vez en cuando se equivoca y produce nuevos replicadores ligeramente diferentes del original [11, 12].

Asumiendo que el ambiente tiene recursos limitados, los diferentes replicadores competirán por los elementos necesarios para replicarse. La hipótesis del mundo del RNA sugiere que estos replicadores podría tratarse de moléculas similares al RNA. Puesto que son moléculas construidas a partir del ensamble de otras menos complejas. Además, la función de estas moléculas está asociada con determinada conformación que la molécula debe adoptar para desempeñarla. Una de estas posibles funciones es realizar actividad catalítica sobre otros RNA [13].

Los replicadores, moléculas de RNA primigenias, desarrollarían diversas estrategias que les dieran ventaja frente a otras. Por ejemplo, barreras a su alrededor para protegerse de daños ambientales. Así como complejos andamiajes que les permitan censar el entorno, almacenar recursos o desplazarse mejor por el medio [14]. Suponemos que en algún momento estos

replicadores tuvieron que evolucionar hacia un sistema más estable de almacenamiento de información como el DNA. Aunque el RNA se mantuvo como un sistema de interacción entre estos replicadores y las complejas máquinas que construyeron, conocidas actualmente como células [12, 14].

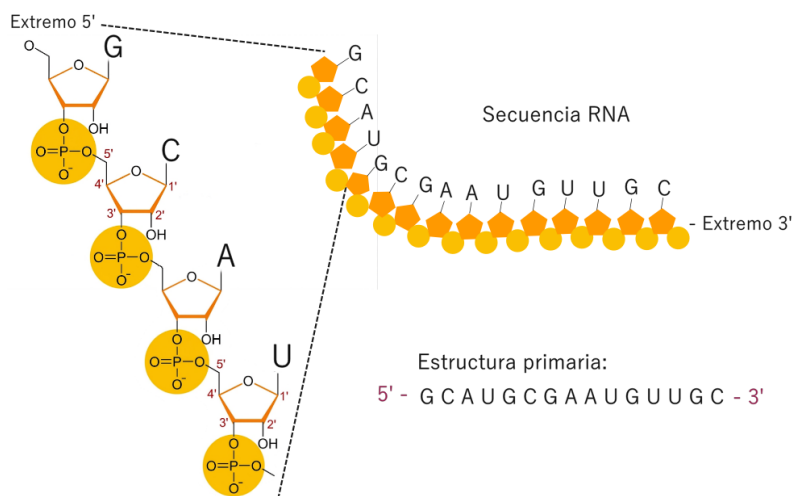
Visto lo anterior, resulta interesante estudiar el RNA como un genoma primigenio. Donde el mapa genotipo-fenotipo se describe según la relación entre la composición de la molécula y la conformación que produce. Además, entender el mecanismo evolutivo de estas moléculas, contribuye a determinar lo que las hizo tan indispensables desde la historia de las primeras células hasta el contexto celular actual.

1.2.2. Niveles estructurales del RNA

Formalmente llamado Ácido Ribonucleico, una molécula de RNA puede entenderse como una cadena formada por cuatro tipos de eslabones unidos entre sí. Cada eslabón está construido por una molécula de azúcar ribosa, y otra molécula (grupo fosfato) que funciona como la unión con el siguiente eslabón. Además, cada azúcar tiene unido lateralmente una de cuatro posibles moléculas llamadas bases nitrogenadas. Dependiendo de la base nitrogenada asociada, el eslabón es nombrado como ribonucleótido de Adenina (A), Guanina (G), Citosina (C) o Uracilo (U) [15].

Cualquier molécula de RNA tendrá una cierta longitud l de ribonucleótidos enlazados. La caracterización ordenada de todos los ribonucleótidos que conforman la cadena, se conoce como ‘estructura primaria’ del RNA [15].

Figura 5. Estructura primaria. Se representa una secuencia de RNA corta. El esqueleto formado por las ribosas y fosfatos se representa en naranja y amarillo. Las bases asociadas a cada ribonucleótido se visualizan mediante su letra representativa. La orientación 5' a 3' se define de acuerdo con la posición de los carbonos que conforman la ribosa.

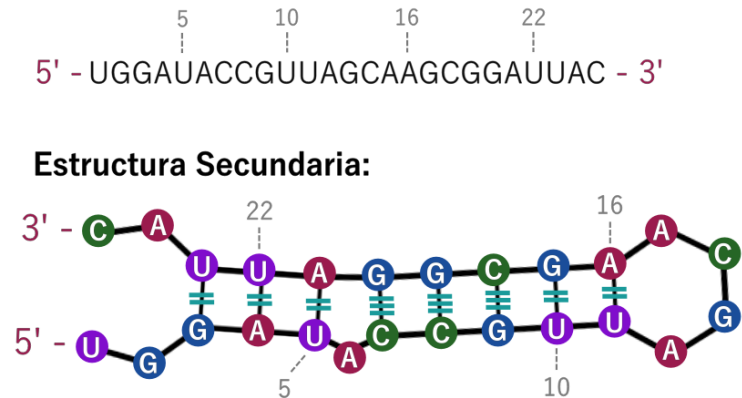


Las cuatro bases nitrogenadas difieren en su estructura molecular de tal manera que pueden establecer interacciones de naturaleza electrostáticas con otras bases nitrogenadas. Estas interacciones electrostáticas se conocen como puentes de hidrógeno y se forman entre parejas canónicas: Guanina con Citosina y Adenina con Uracilo. Así como también son posibles otro tipo de interacciones no canónicas, como Guanina con Uracilo o Citosina con Uracilo [16].

Si conocemos la estructura primaria y las interacciones electrostáticas posibles entre las diferentes bases de la cadena, podemos dibujar una representación bidimensional de la estructura que podría formarse como consecuencia de estas interacciones. Este nivel de representación es lo que conocemos como ‘estructura secundaria’ [15].

Existen múltiples formas de representar la estructura secundaria de una molécula de RNA. Aquí presentamos una representación plana que muestra tanto la identidad de los ribonucleótidos como las interacciones entre ellos y la estructura general de la molécula. Otras representaciones relevantes son profundizadas en el apartado metodológico.

Figura 6. Estructura secundaria. Representación plana de una posible estructura secundaria a partir de la secuencia propuesta. La cantidad de — en las interacciones electrostáticas indican la cantidad de puentes de hidrógeno que se forman en cada interacción.



Aunque en este proyecto solo utilizamos estructura primaria y secundaria, también existe una representación tridimensional o ‘estructura terciaria’ que representa el plegamiento de manera más realista. Sin embargo, las interacciones consideradas en la estructura secundaria son las que más contribuyen al plegamiento global de la molécula y suelen ser suficiente para intuir la función de un específico RNA [17].

Finalmente, debido a que estas moléculas tienen la capacidad de interactuar con otras semejantes o diferentes, pueden adoptar configuraciones tridimensionales que no podrían alcanzar por sí solas. Esta configuración accesible mediante interacciones con alguna otra molécula, se puede considerar como ‘estructura cuaternaria’ del RNA [15].

Hasta este punto solo hemos considerado qué (i) las moléculas de RNA adquieren una cierta estructura debido a las interacciones entre ribonucleótidos que la componen, y (ii) que esta estructura está asociada con la actividad o función que desempeña la molécula. Pero no hemos abordado el concepto de plasticidad fenotípica en estas moléculas. Comencemos estableciendo que el tipo de plasticidad en el que nos centraremos, se denomina plasticidad estructural.

1.2.3. Plasticidad estructural en el RNA

Consideremos una secuencia cualquiera de RNA. Supongamos que elegimos un ribonucleótido al azar que resulta tener asociada la base nitrogenada citosina (C), cuya interacción más común es con una guanina (G). Es de esperarse que a lo largo de la secuencia haya múltiples guaninas candidatas para interactuar. Extendiendo esta lógica para todos los ribonucleótidos que componen la secuencia, es evidente que hay diferentes posibles estructuras

secundarias para la misma molécula de RNA.

En el contexto evolutivo, la secuencia de ribonucleótidos representa el genotipo capaz de producir un conjunto de diferentes fenotipos (estructuras secundarias). Este tipo de plasticidad fenotípica es lo que se conoce como ‘plasticidad estructural’. Mientras que el conjunto de estructuras a las que puede acceder una secuencia se denomina su ‘repertorio plástico’. Habrá estructuras con mayor estabilidad al entorno que otras. Razón por la cual una molécula adoptará algunas de las estructuras con mayor frecuencia que otras [18].

En nuestro caso, al considerar solo estructura secundaria, la estabilidad que ofrece cada estructura se aproxima únicamente mediante las interacciones entre los ribonucleótidos en este nivel estructural. Al calcular la estabilidad de cada estructura dentro del repertorio plástico, podemos aproximar la probabilidad con la que una molécula se plegará en cada estructura en términos de su estabilidad. Estos métodos y algoritmos pueden ser consultados ampliamente en la sección metodológica.

Lo relevante en esta sección es que una molécula de RNA puede adoptar diferentes estructuras, cada una de las cuales ocurre con cierta probabilidad asociada con la estabilidad que ofrece. Aunque es común que solo una de las muchas estructuras esté reportada como funcional [18].

Es posible que, en época del mundo de RNA la plasticidad estructural resultaba fundamental para encontrar estructuras funcionales. La razón es que al no conocer cuáles estructuras son funcionales, cualquier estructura podría esconder alguna función novedosa. En otras palabras, encontrar nuevas estructuras funcionales implica explorar estructuras desconocidas.

Planteamiento del Problema

Este proyecto se origina desde el planteamiento de que la plasticidad fenotípica tiene relevancia en la evolución de los fenotipos. Desde esta perspectiva el modelo clásico, que consiste en seguir el fenotipo predominante a lo largo del tiempo, es una representación incompleta de la historia. Creo que incorporar la plasticidad fenotípica no solo es más realista, sino que es una mejor descripción del mecanismo por el cual evolucionan los fenotipos.

En este sentido, el RNA resulta ideal como elemento de estudio debido a su accesible representación del mapa genotipo-fenotipo. Además de la gran relevancia que desempeña en múltiples procesos celulares, es protagonista de muchas aplicaciones biotecnológicas y conocemos de su importancia aparente en etapas tempranas de la vida [19].

Un buen punto de partida es establecer la forma en que simulamos la evolución *in silico* de moléculas de RNA. Para ello seguimos lo que se conoce como algoritmo evolutivo, adaptando las condiciones necesarias para modelar la evolución de estructuras secundarias de RNA.

2.1. Algoritmos evolutivos

Un proceso evolutivo consiste en la serie de cambios en el tiempo que transforman un ‘algo’ desde una situación A , hasta otra diferente B . Un algoritmo evolutivo se puede definir como la serie de instrucciones que dirigen el proceso evolutivo $A \rightarrow B$. [20]

En nuestro caso dicho ‘algo’ equivale a una población de N células primitivas. Cada una de las cuales se representa mediante su genotipo, que consiste en una secuencia de RNA. Al inicio de las simulaciones todas las células comparten el mismo genotipo. Por tanto, tienen acceso al mismo repertorio de estructuras, donde el fenotipo más estable es una estructura denominada λ .

El objetivo de las simulaciones es que un porcentaje definido de la población ($\geq 80\%$) sean capaces de producir la estructura objetivo τ como fenotipo más estable de su repertorio estructural (tal que $\tau \neq \lambda$). Para este fin aplicamos iterativamente cuatro procesos consecutivos sobre las N células primitivas, hasta que se cumpla la condición deseada o se rebase un límite de iteraciones determinado.

2.1.1. Cálculo de adecuación

El primer proceso consiste en asignar a cada individuo de la población una puntuación de acuerdo con cierto criterio de adecuación. Este criterio consiste en una ecuación matemática

que califica un genotipo en función de una o más características del fenotipo que produce. Por ejemplo, en el caso tradicional, podríamos calificar únicamente que tan cercano a τ es el fenotipo más estable producido por la secuencia. Los genotipos con mejor adecuación serían aquellos cuyo fenotipo se asemeja más a τ .

Nótese que introducir la plasticidad estructural, significa incorporar un criterio en el que se considere cada una de las estructuras accesibles en el cálculo de la adecuación. Las siguientes secciones se centran en los criterios empleados para esto. Además, pueden ser consultados con mayor detalle en el apartado metodológico.

2.1.2. Selección

En nuestras simulaciones el proceso de reproducción se considera asexual. Al tratarse de células, lo que ocurre es la duplicación del genoma para dar origen a una célula hija. Durante cada generación se realizan N sorteos estocásticos independientes que determinan los genotipos de la siguiente generación. La probabilidad de que una célula sea elegida es proporcional a su adecuación calculada en el paso anterior.

Este se denomina ‘selección’, ya que emula el proceso de selección natural, en el que los organismos con fenotipos mejores son los que más oportunidades tienen de transmitir sus genes a la siguiente generación. Las probabilidades de reproducción de cada individuo se calculan como razón entre su adecuación obtenida y la sumatoria de adecuaciones de toda la población.

Una vez establecida la probabilidad de reproducción para cada individuo, se realizan N sorteos estocásticos independientes entre sí. Seleccionando los N genotipos que serán heredados a la siguiente generación.

2.1.3. Mutación

Antes de pasar a la siguiente generación, a cada genotipo seleccionado se le aplica un proceso de mutación que emula la variación natural por mutación. Esto es que cada ribonucleótido de estos genotipos, tiene una probabilidad μ de que su base nitrogenada sea reemplazada por otra diferente.

2.1.4. Reemplazo poblacional

Nuestras simulaciones son extintivas, lo que significa que cada generación la población de células padres desaparece, siendo reemplazada por la nueva población de células hijas. Esto quiere decir que a lo largo de todas las generaciones el tamaño de la población se mantiene constante. Pero que cada nueva generación de genotipos, proviene de los seleccionados y sometidos al proceso de mutación de la generación previa.

2.2. Modelos anteriores de plasticidad estructural

Ancel y Fontana (2000) realizaron simulaciones evolutivas emulando un ensayo de evolución dirigida con moléculas de RNA *in silico*. En su estudio plantearon incorporar la plasticidad estructural de las moléculas mediante el criterio de adecuación descrito en la Eq(2.1). En este criterio la adecuación de un genotipo es calculada como la sumatoria de promedios obtenida al evaluar cada fenotipo accesible m , mediante una razón entre su probabilidad de ocurrencia B_m y una expresión que varía de acuerdo a la diferencia estructural entre m y el fenotipo objetivo τ [21].

$$W_{c/plasticidad} = \sum_m \frac{B_m}{C + d(m, \tau)} \quad (2.1)$$

Donde C representa una constante arbitraria necesaria para evitar valores inválidos cuando la distancia estructural es cero. Cuando consideramos el caso tradicional o 'sin plasticidad', la estructura más estable se considera el único fenotipo m accesible. Sin estructuras adicionales que compitan, asumimos que la probabilidad con la que se producirá este fenotipo es uno. Resultando en el criterio de adecuación descrito mediante la Eq(2.2).

$$W_{s/plasticidad} = \frac{1}{C + d(m, \tau)} \quad (2.2)$$

2.3. Confinamiento neutral

Durante las simulaciones evolutivas, Ance y Fontana reportaron que la población bajo el criterio con plasticidad fenotípica era incapaz de alcanzar el fenotipo objetivo. La evolución de estas poblaciones se estancó debido a la aparición de genotipos altamente resistentes tanto a perturbaciones genéticas como ambientales [21].

La resistencia a perturbaciones genéticas se refiere a que las mutaciones de ribonucleótidos se vuelven menos efectivas para modificar las estructuras que el genotipo produce. En cambio, la resistencia a perturbaciones ambientales se traduce en que las secuencias disminuyen la capacidad de acceder a diversos fenotipos. Este fenómeno de estancamiento evolutivo, en el que las características de la poblacional imposibilitan que la evolución produzca fenotipos cada vez más cercanos al fenotipo objetivo τ , fue nombrado 'confinamiento neutral' [18, 21].

Aunque Ance y Fontana no profundizan en las razones que llevan a la aparición de estos genotipos resistente. Su estudio parece insinuar que la plasticidad fenotípica no favorece la evolución [21]. Dejando al menos hasta entonces este fenómeno como algo anómalo.

Loreto-Velázquez A., Ugalde E., y Espinosa Soto, C. (2025) contribuyeron a entender el fenómeno de confinamiento neutral con más profundidad. En su trabajo realizaron simulaciones evolutivas *in silico* bajo selección estabilizadora, lo cual significa que las secuencias producían desde el inicio la estructura deseada. En tales condiciones demostraron que el modelo de Ance y Fontana inducía el fenómeno de confinamiento neutral. Pero que es posible atenuar dicho fenómeno empleando un criterio de adecuación diferente [22].

2.4. Modelo de unión a Ligando

El modelo de unión a ligando supone que cada célula de la población contiene un genotipo que produce una cierta concentración [RNA] de copias de sí mismo. Además, hay presente una cierta concentración [L] de un ligando capaz de unirse a RNA y formar determinada concentración [X] de complejo (Ver Eq. 2.3) [22].



Las múltiples copias de RNA se mantendrán fluctuando entre los diferentes fenotipos que son capaces de producir. Cuando un ligando se une a cualquiera de las copias, secuestra dicha copia plegada en alguna estructura m . Asumimos que la afinidad entre el ligando y cualquier RNA es siempre la misma, siendo k_+ la constante de asociación. Mientras que la constante de disociación k_- del complejo consiste en una función de sujeción. Una expresión que describe que tan fuerte es la unión entre el ligando y el RNA secuestrado (Eq. 2.4). [22]

$$m + L \xrightleftharpoons[k_-]{k_+} X \quad (2.4)$$

tal que $k_- = a \cdot 2^{b \cdot d(m, \tau)}$

Donde k_- se calcula a partir de la similitud estructural entre el fenotipo expresado por la secuencia secuestrada y una estructura de referencia con afinidad máxima al ligando. Esta distancia se expresa como $d(m, \tau)$, siendo τ el fenotipo objetivo de la simulación y a , b constantes. El procedimiento por el que se obtiene la distancia entre dos estructuras se incluye dentro de la metodología aplicada [22].

A partir de la Eq. 2.4 se puede intuir que la cantidad de complejo formado, entre RNA plegado en la estructura m y moléculas de ligando, se puede determinar solucionando la siguiente ecuación diferencial:

$$\frac{dC_m}{dt} = k_+[m][L] - k_-[C_m] \quad (2.5)$$

Siendo $[m]$ la cantidad de RNA no unido en complejo, pero plegado en la estructura m ; similarmente $[L]$ representa la cantidad de ligando no unido en complejo. Considerando lo anterior, podemos rescribirlo con más detalle:

$$\frac{dC_m}{dt} = k_+B_m(R_T - C_m)(L_T - C_m) - k_-[C_m] \quad (2.6)$$

En esta versión expandida B_m representa la proporción de la población plegada en esa específica estructura; $(R_T - C_m)$ y $(L_T - C_m)$ simbolizan el RNA y el ligando no unido a complejo, respectivamente. Dado que $m \in \mathcal{P}$, donde \mathcal{P} es el repertorio plástico con M estructuras diferentes, lo que tenemos es un sistema de M ecuaciones diferenciales tal que $\{m = 1, \dots, M\}$.

$$\begin{aligned}
\frac{dX_1}{dt} &= k_+ B_1 \left(R_T - \sum_{m=1}^M C_m \right) \left(L_T - \sum_{m=1}^M C_m \right) - k_{-,1} C_1 \\
&\vdots \\
\frac{dC_m}{dt} &= k_+ B_m \left(R_T - \sum_{m=1}^M C_m \right) \left(L_T - \sum_{m=1}^M C_m \right) - k_{-,m} C_m \\
&\vdots \\
\frac{dC_M}{dt} &= k_+ B_M \left(R_T - \sum_{m=1}^M C_m \right) \left(L_T - \sum_{m=1}^M C_m \right) - k_{-,M} C_M
\end{aligned} \tag{2.7}$$

Al considerar el sistema completo la cantidad de RNA y ligando libres se calcula restando el ocupado en todos los diferentes complejos ($\sum_{m=1}^M C_m$). El sistema de ecuaciones además asume la proporción de moléculas que se pliegan en cada estructura es la esperada en el equilibrio. Esta simplificación se adoptó después de considerar el modelo que permite las transiciones entre estructuras y demostrar que la solución al sistema coincide en ambos escenarios. Dicha demostración, así como la solución al sistema se incluye dentro del apartado metodológico [22].

Solucionar este sistema de ecuaciones, implica encontrar la cantidad total de cada tipo de complejo formado. La adecuación de un genotipo se obtiene sumando la cantidad de complejo total que fue capaz de formar a partir de su repertorio de estructuras accesibles [22].

2.5. Propósito de este proyecto

Aunque el modelo de unión a ligando demostró ser capaz de reducir los efectos producidos por confinamiento neutral, no ha sido probado en simulaciones direccionales. Es decir, que no se ha explorado realmente su capacidad de evolucionar con éxito un conjunto de moléculas desde un fenotipo inicial hacia otro diferente.

En este proyecto nos me propongo realizar simulaciones evolutivas bajo selección direccional, con el objetivo de explorar el funcionamiento del modelo de unión a ligando como criterio de adecuación. Sugiero que este modelo será capaz concluir su recorrido evolutivo considerando la plasticidad estructural de las moléculas, sin que se produzcan estancamientos evolutivos.

2.5.1. Objetivo General

Explorar el funcionamiento del modelo de unión a ligando como criterio de adecuación en simulaciones evolutivas *in silico* de moléculas de RNA, bajo selección direccional.

2.5.2. Objetivos Específicos

- Generar código en C++ para realizar modelado, manipulación y simulación evolutiva de moléculas de RNA *in silico*.
- Obtener un muestro de la red neutral de tRNA_{phe} o conjunto de secuencias que comparten la estructura de tRNA_{phe} como estado más estable.
- Explorar el espacio de estructuras accesibles mediante mutaciones en la red neutral muestreada previamente.
- Seleccionar las estructuras secundarias más convenientes como objetivos evolutivos.
- Realizar múltiples simulaciones evolutivas y registrar datos históricos a lo largo de cada simulación.
- Aplicar análisis estadísticos sobre los datos históricos generados y realizar inferencias sobre el comportamiento del modelo.

Marco Metodológico

3.1. Predicción de estructuras secundarias

Resolver una estructura secundaria de RNA con pseudonudos es considerado un problema NP-completo. Esto quiere decir que no hay un algoritmo conocido que resuelva en tiempo polinomial este problema. En cambio, el caso sin pseudonudos es posible resolverlo empleando algoritmos de programación dinámica como el de Zuker y Stiegler (1981) [23, 24].

Cualquier algoritmo de programación dinámica debe cumplir con cuatro aspectos fundamentales. (i) Definir las características que establecen una solución como óptima; (ii) Descomponer el problema en subproblemas siguiendo una dirección de resolución; (iii) Almacenar los valores de los subproblemas resueltos en forma matricial; (iv) Capacidad de rastreo de resultados de la matriz para encontrar la solución óptima. [25]

En el algoritmo de Zuker y Stiegler (1981) la solución óptima es la estructura secundaria que ofrece mayor estabilidad termodinámica en términos de energía libre de Gibbs (ΔG). Divide la estructura completa en las subestructuras que la componen, evalúa la contribución energética que ofrece cada una y las suma para obtener la estabilidad total [24, 25].

Almacenar los datos de esta forma permite rastrear la ruta de interacciones que otorga mayor estabilidad y retornar la estructura producto del conjunto de interacciones, junto con el valor ΔG calculado para la estructura completa. También es posible utilizar algoritmos como el de Wuchty, S. et al. (1999) que siguen la misma metodología descrita, pero rastrean el conjunto de estructuras posibles en un determinado intervalo de energía con respecto al de menor energía libre (mfe) [24, 25, 26].

Tanto el algoritmo de Zuker y Stiegler (1981) como el Wuchty, S. et al. (1999) están implementados dentro de la biblioteca para predicción de estructuras secundarias ‘ViennaRNA’ (Lorenz, R. et al. 2011). Los datos de interacciones entre ribonucleótidos provienen de diversos conjuntos de datos experimentales que pueden ser consultados en la base de datos ‘NNDB’ (Turner, 2004) [27, 28].

3.1.1. ¿Por qué usamos energía libre de Gibbs?

En cualquier sistema termodinámico deben poder identificarse dos tendencias. La primera describe como el sistema opta por estados de menor energía; mientras que la segunda implica la tendencia de la energía a dispersarse, empujando el sistema a un estado de mayor entropía[29].

Consideremos el sistema que conceptualmente me parece el más simple. La máquina de calor utiliza un flujo de energía térmica, desde una temperatura alta T_1 hasta una menor T_2 , con el objetivo de extraer la diferencia de energía y convertirla en trabajo. La energía que podemos extraer para este propósito es denominada ‘energía libre’[29].

La intención es que todo el cambio en la energía interna del sistema (ΔU) se convierta en trabajo. Sin embargo, parte de la energía térmica será cedida para aumentar la T_2 . En un sistema cerrado, cada uso de la máquina implica que las temperaturas T_1, T_2 se acercan más al equilibrio. Esto es la tendencia de nuestro sistema de aumentar su entropía. Cuando $T_1 = T_2$ no hay flujo térmico y la máquina se detiene, llegando a su estado de menor energía[29].

El cambio en la entropía de un sistema se puede definir como $\Delta S = Q/T$, donde Q es la energía cedida y T es una temperatura constante. Entonces es lógico proponer que la energía libre del sistema se puede expresar como $U - Q$, tal que $Q = T \cdot \Delta S$. Lo anterior, resulta en la expresión conocida como energía libre de Helmholtz, utilizada para describir sistemas a temperatura constante[29].

$$F = U - T \cdot S \quad (3.8)$$

A diferencia del ejemplo anterior, los sistemas biológicos no son sistemas cerrados. Reciben energía del exterior que convierten en trabajo para incrementar la organización y reducir la entropía del sistema. Se ha propuesto que esto es posible porque como resultado final devuelven al entorno una forma energía más desorganizada, lo que implica que la entropía del universo aumente[29].

El interior de una célula por ejemplo, es un sistema que tienden a funcionar con temperatura y presión constantes. Lo cual implica que se intercambie energía con el entorno mediante trabajo de expansión (pV) [29, 30, 31]. Así, la energía libre de este sistema es la ecuación de energía libre de Helmholtz 3.8 al adicionarle este trabajo de expansión.

$$G = (U - T \cdot S) + p \cdot V \quad (3.9)$$

La energía que las células reciben proviene de diversas fuentes. La característica de esta energía es que está empaquetada en una forma de baja entropía. Tras ser utilizada debe retornarse al sistema transformada en energía de mayor entropía como el calor. Siendo la entalpía (ΔH) la expresión que describe flujo de calor entre el sistema y el entorno, podemos reescribir nuestra expresión de energía libre como

$$G = H - TS$$

(3.10)

ya que $H = U + pV$

Considerando un cambio finito (Δ) en el sistema, llegamos a la expansión de energía libre de Gibbs. Nombrada así por Josiah Willard Gibbs quien definió esta ecuación considerando las condiciones del interior celular.

$$\Delta G = \Delta H - T\Delta S \quad (3.11)$$

Esta es la medida de energía libre más utilizada en biología y resulta especialmente útil porque describe direccionalidad de reacciones bioquímicas. En nuestro caso particular, me

interesa que se entienda cómo ΔG representa la tendencia de una molécula de RNA para adoptar una estructura secundaria [29].

Para una cierta estructura secundaria, ΔG se determina con referencia al estado lineal de la molécula. Me refiero al estado en el que no hay ninguna interacción electrostática entre los ribonucleótidos que componen la secuencia. Esta comparación significa ir del estado de máxima entropía hacia alguno otro más organizado, implicando una reducción en la entropía del sistema ($\Delta S < 0$). Debido a esto, el término $T\Delta S$ es negativo y empuja hacia un valor positivo de energía libre (3.11).

Para que ΔG sea negativo, ΔH debe entonces ser negativo y suficientemente exotérmico. Es decir que la liberación de energía al entorno tiene que ser más grande que la penalización $T\Delta S$ por ordenarse. Mientras más exotérmico es el plegamiento en una estructura, menor es el valor ΔG y más propenso a ocurrir [31].

3.1.2. Probabilidad de plegamiento en una estructura

Para un conjunto de moléculas de RNA idénticas las diferentes estructuras accesibles competirán entre sí, haciendo que cada molécula fluctúe en los diferentes estados. Debido a que algunos son más propensos a ocurrir que otros, la distribución de moléculas en cada estado no será al azar. La distribución de moléculas en cada estado será cada vez más notable hasta alcanzar un punto de equilibrio. Esto es un punto tal, que el número de moléculas que abandonan una conformación son el mismo número de nuevas moléculas que lo adoptan [32].

La distribución de Boltzmann es una expresión de probabilidad que asigna a cada estado del sistema una probabilidad según su energía termodinámica. Lo que permite aproximar la frecuencia de una estructura/estado en el punto de equilibrio [32].

El factor de Boltzmann se origina de dos ideas básicas. La primera es que considerando sistemas independientes, las energías se suman y las probabilidades se multiplican. Una función que cumple naturalmente con esa propiedad es la exponencial [30, 32].

Siendo \mathcal{P} el repertorio plástico de una secuencia. Para cualquier estructura ($\sigma \in \mathcal{P}$) con energía libre ΔG_σ , su factor de Boltzmann se calcula mediante la expresión

$$\begin{aligned} b(\sigma) &:= e^{-\Delta G_\sigma/k_B T} \\ k_B &= 1,987204259 \times 10^{-3} \text{ kcal mol}^{-1} \text{ K}^{-1} \\ T &= 310,15 \text{ K} \quad (37^\circ\text{C}) \end{aligned} \tag{3.12}$$

La segunda idea es que un estado con mayor energía debe ser menos probable que uno con menor energía, por lo que se incorpora un signo negativo en el exponente. Además, la energía debe compararse con una escala térmica, para lo cual se incorpora $k_B \cdot T$. Finalmente, lo que convierte este factor en una ecuación de probabilidad es normalizarlo respecto la suma de factores de Boltzmann para todos los estados accesibles (1, ..., M) en \mathcal{P} [30, 32].

$$B_\sigma = \frac{e^{-\Delta G_\sigma/k_B T}}{\sum_{m=1}^M b(m)} \tag{3.13}$$

Esta última siendo la expresión que utilizamos para calcular la cantidad de copias de RNA que adoptarán una cierta estructura en el modelo de unión a ligando (2.7). Para reducir tiempo de cómputo, \mathcal{P} contiene solo estructuras accesibles cuyo valor de energía libre (ΔG) tenga una diferencia máxima de 2(kcal/mol) con respecto a la estructura de mfe.

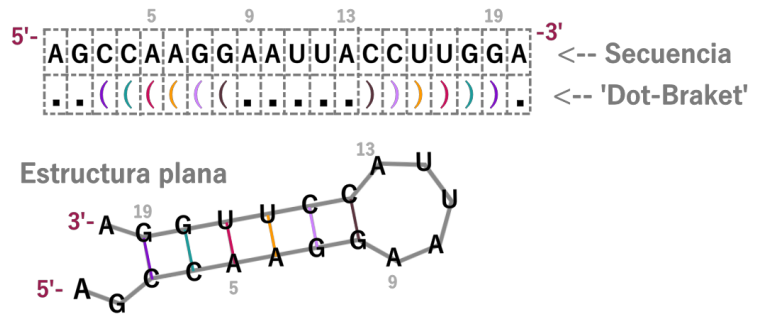
3.2. Representación de estructuras

Aunque la visualización plana de una estructura secundaria de RNA brinda mucha información, otras representaciones permiten trabajar más fácilmente con estas estructuras. Para el desarrollo de este proyecto conviene abordar dos notaciones de estructuras secundarias.

La primera representación se conoce como ‘Dot-Bracket’, traducida al español como notación de paréntesis y puntos. La representación es equivalente en posición de ribonucleótidos a la estructura primaria de la molécula. Con la diferencia que un punto indica que la posición está desapareada, y cada pareja de paréntesis representa un par de ribonucleótidos apareados electrostáticamente entre sí. Es la notación más común para trabajar con moléculas de RNA en software, incluyendo la biblioteca de funciones *ViennaRNA* que utilizamos para este proyecto [27].

Figura 7. Notación ‘Dot-Bracket’.

Se ilustra la notación ‘Dot-Bracket’ o paréntesis y puntos en comparación con la secuencia y representación plana de la estructura secundaria. Se utilizan colores para facilitar la visualización de las interacciones indicadas por los paréntesis.

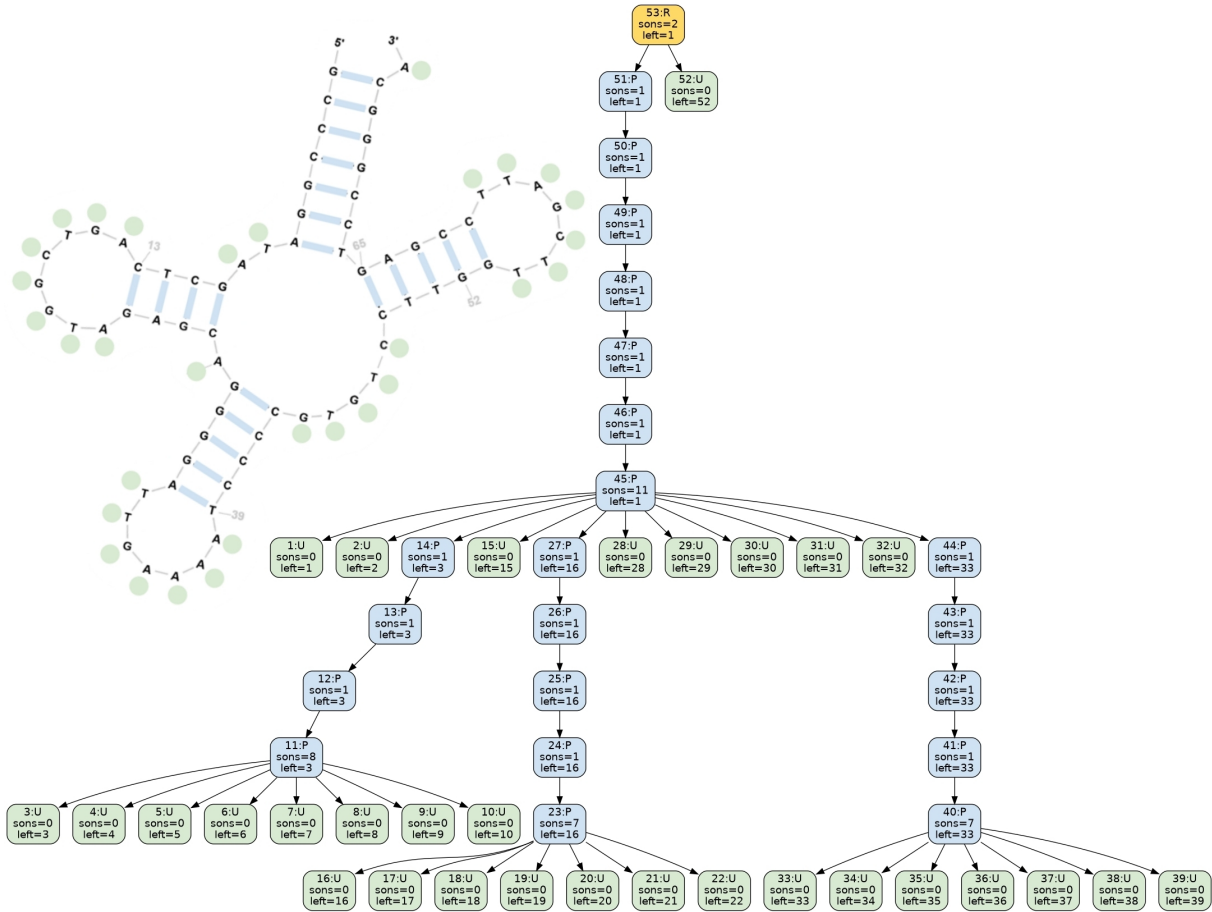


El segundo tipo de representación que es necesario introducir es la representación de árbol. Consiste en un grafo organizado de manera jerárquica, donde se pueden diferenciar regiones apareadas y desapareadas de la estructura. La idea de esta representación es capturar la organización general de la estructura secundaria. No usamos directamente esta representación, pero sí para medir la distancia entre estructuras secundarias [33, 34].

La mejor forma de entender la representación de árbol es visualizándola. Consideremos la estructura que utilizamos como fenotipo inicial en nuestras simulaciones, y correspondiente al RNA de transferencia de fenilalanina o tRNA^{phe} (véase Figura 8).

En la representación de árbol hay nodos (verdes) que representan ribonucleótidos desapareados. Estos nodos no poseen descendientes o ramificaciones. Mientras que los nodos (azules) con descendencia indican pares de ribonucleótidos interactuando. Debemos leer de izquierda a derecha los nodos en el mismo nivel, indicando la dirección 5'→3' de la secuencia.

En la Figura 8, la estructura inicia formando un tallo de siete bases al que continúa una estructura de burbuja central. En esta burbuja se visualizan tres tallos burbuja como tres



Secuencia utilizada para producir las estructuras:
GCCCGAUAGCUCAGUCGGUAGAGCAGGGGAUUGAAAAUCCCGUGUCCUUGGUUCGAUUCGAGUCCGGGCA
 (((((((.....(((.....((((((.....)))))).....)))))))).

Figura 8. Comparación de representaciones de estructura secundaria de tRNA_{phe}. Se muestra la representación plana y de árbol para la estructura secundaria tRNA_{phe}. En la parte inferior se reporta la secuencia utilizada para generar ambas representaciones, así como la estructura en notación Dot-Bracket. La imagen del árbol se realizó siguiendo las especificaciones de cada nodo reportadas por el software *ViennaRNA*. Herramienta utilizada para generar la representación plana y Dot-Bracket. Los colores se añadieron para ilustrar mejor la relación entre ambas representaciones. Imagen diseñada por el autor.

grupos de nodos descendientes que terminan en un conjunto de nodos desapareados. El nodo amarillo al comienzo nos indica que el último ribonucleótido de la secuencia está desapareado.

3.2.1. Cálculo de distancia estructural

La representación de estructuras secundarias mediante diagramas de árbol es útil para describir las subestructuras contenidas dentro de la estructura general. El software de *ViennaRNA* no genera imágenes de estas representaciones. Aunque sí convierte notación dot-Bracket en metadatos como los incluidos en cada nodo de la Figura 8; que pueden ser utilizados para construir dicha representación [21, 22, 33].

Dadas dos estructuras secundarias (σ, τ) en notación dot-Bracket, provenientes de secuencias del mismo tamaño, *ViennaRNA* incluye funciones específicas para obtener los metadatos de sus respectivas formas de árbol. Esto permite determinar cuántos nodos es necesario modificar para transformar una estructura en la otra. La cantidad de operaciones de edición necesarias es lo que empleamos como medida de ‘distancia estructural’ tanto para 2.1, 2.2, como en nuestro modelo de unión a ligando al determinar la constante de disociación para un complejo formado 2.4. [21, 22, 34]

3.3. Criterios de adecuación

El criterio de adecuación es el modelo matemático que nuestras simulaciones utilizan para puntuar o calificar que tan bueno es un genotipo. Ya describimos previamente estos modelos de manera amigable. Por lo que aquí seremos más breves y concretos en las descripciones.

3.3.1. Modelo de Ancel y Fontana

En este modelo un genotipo se califica según

$$W_{AF} = \sum_{\sigma=1}^M (B_{\sigma} \cdot (d(\sigma, \tau) + C)^{-1})$$

(3.14)

donde

$$C = 0,001$$

$$\sigma \in \mathcal{P} : (1, \dots, M)$$

Siendo una sumatoria de promedios calculados entre la probabilidad de Boltzmann (3.13) y la distancia estructural, para cada estructura en el repertorio \mathcal{P} del genotipo.

Nótese que en este modelo hay dos formas de aumentar la adecuación que recibe el genotipo. (i) Reducir la distancia estructural al fenotipo objetivo. (ii) Aumentar la probabilidad de Boltzmann de las estructuras con mejor distancia estructural.

Al principio de las simulaciones de Ancel y Fontana los fenotipos iniciales son muy diferentes al objetivo, lo que hace fácil reducir la distancia estructural. Pero conforme se hace

más difícil encontrar mutaciones que mejoren la estructura, aumentar la probabilidad de Boltzmann de las mejores estructuras del momento se vuelve la forma más simple de subir la adecuación [22].

Lo anterior provoca que los genotipos incrementen indiscriminadamente la probabilidad de Boltzmann de su mejor estructura. Conforme progresan las generaciones los genotipos seleccionados dejan de poder acceder a estructuras novedosas y terminan en un callejón sin salida. Incapaces de producir fenotipos nuevos, son víctimas de su propio éxito [22].

Modelo de Ancel y Fontana, caso sin plasticidad

Cuando no se considera plasticidad, el repertorio plástico \mathcal{P} solo contiene la estructura de mínima energía libre (mfes). Siendo la única estructura evaluada no hay sumatoria de promedios. Consideramos que su probabilidad de ocurrencia es uno y nos queda una versión que depende solo de la distancia estructural entre la mfes y el fenotipo objetivo.

$$W_{AFs/plasticidad} = \frac{1}{C + d(\text{mfes}, \tau)}$$

manteniendo

$$C = 0,001$$
(3.15)

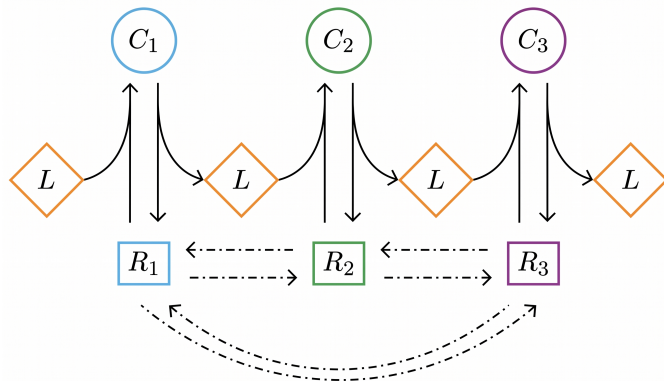
3.3.2. Modelo de Unión a Ligando

Para este modelo asumimos que un genotipo produce una concentración \mathcal{R} de moléculas copia de sí misma, por tanto, comparten el repertorio plástico \mathcal{P} de tamaño M . Estas moléculas compiten por unirse a una concentración \mathcal{L} de cierto ligando. Todo dentro de un ambiente celular aislado e independiente de otros.

Previamente, describimos este modelo bajo supuesto, de que el conjunto de moléculas adopta inmediatamente la distribución esperada en el equilibrio. Lo que permite asumir la proporción de moléculas que adoptan una estructura según su probabilidad de Boltzmann (3.13), y describir el modelo mediante el sistema de ecuaciones 2.7. Este sistema simplificado se ve representado en la Figura 9 sin considerar las flechas puntuadas.

Figura 9. Reacciones del sistema.

El ejemplo considera que \mathcal{P} contiene $M=3$ estructuras diferentes representadas con cuadrados. Los rombos simbolizan moléculas ligando libres, y los círculos los diferentes complejos que se pueden formar. Imagen diseñada a partir de Fig. S1 en Loreto-Velázquez A., Ugalde E., y Espinosa Soto, C. (2025).



El modelo que considera las flechas puntuadas, permite que las moléculas no unidas a complejo cambien su estructura por cualquier otra incluida en su repertorio. Siendo ese el caso, la cantidad de complejo formado para una estructura $m \in \mathcal{P}$ se expresa

$$\frac{dC_m}{dt} = k_+ R_m (L_T - \sum_{\sigma=1}^M C_\sigma) - k_- C_m \quad (3.16)$$

Donde R_m representa la cantidad libre de moléculas plegadas en la estructura m . Cantidad que cambia de acuerdo con la expresión

$$\begin{aligned} \frac{dR_m}{dt} = & \left(k_- C_m - k_+ R_m (L_T - \sum_{\sigma=1}^M C_\sigma) \right) \\ & + \left(\sum_{j=1, j \neq m}^{M-1} z_{m,j} R_j + z_{m,M} (R_T - \sum_{\sigma=1}^M C_\sigma - \sum_{\sigma=1}^{M-1} R_\sigma) - R_m \sum_{j=1, j \neq m}^M z_{j,m} \right) \end{aligned} \quad (3.17)$$

Observese que la primera parte de esta expresión es lo mismo descrito en la 3.16 con signo contrario. Mientras que la segunda parte engloba las transiciones posibles entre las diferentes estructuras del repertorio \mathcal{P} . Siendo $z_{j,m}$ por ejemplo, la tasa de transición de la estructura j hacia la estructura m .

Ya que hay M estructuras en el repertorio \mathcal{P} , el modelo estará formado por dos sistemas de M ecuaciones. El primero describe la formación de complejos; el segundo la cantidad de moléculas plegadas en cada estructura. Sin embargo, solucionar el sistema aún significa encontrar la cantidad de complejos formados cuando se llega al equilibrio. Por definición, en el equilibrio el cambio en la cantidad cualquier complejo debe ser cero.

$$\frac{dC_m}{dt} = k_+ R_m (L_T - \sum_{\sigma=1}^M C_\sigma) - k_- C_m = 0 \quad (3.18)$$

Por lo tanto,

$$\frac{dR_m}{dt} = (0) + \left(\sum_{j=1, j \neq m}^{M-1} z_{m,j} R_j + z_{m,M} (R_T - \sum_{\sigma=1}^M C_\sigma - \sum_{\sigma=1}^{M-1} R_\sigma) - R_m \sum_{j=1, j \neq m}^M z_{j,m} \right)$$

Lo anterior implica que la segunda parte de la ecuación, que describe las transiciones entre las diferentes estructuras, tiene que ser cero para satisfacer las condiciones del equilibrio.

$$\left(\sum_{j=2}^{M-1} z_{1,j} R_j + z_{1,M} (R_T - \sum_{\sigma=1}^M C_\sigma - \sum_{\sigma=1}^{M-1} R_\sigma) - R_1 \sum_{j=2}^M z_{j,1} \right) = 0 \quad (3.19)$$

para cualquier estructura $\in \mathcal{P}$

Resumiendo, las condiciones necesarias para resolver el sistema de ecuaciones que considera las transiciones entre estados, se cumplen cuando cada $R_\sigma \in \mathcal{P}$ tiene la concentración esperada en el equilibrio. Estas concentraciones son las que dicta la probabilidad de Boltzmann para cada estructura 3.13. Asumir que dichas concentraciones se adoptan inmediatamente, permite describir el sistema como una versión simplificada 2.7.

Solución al modelo de unión a ligando

Partimos del sistema de ecuaciones 2.7 que describe la versión simplificada del modelo de unión a ligando. sólo como recordatorio, en este sistema B_m es la probabilidad de Boltzmann 3.13 de una cierta estructura m del repertorio de tamaño M . Las concentraciones de RNA y Ligando del sistema se representan como R_T, L_T respectivamente. La constante de asociación k_+ es la misma para todas las estructuras, mientras que la de disociación (2.4) se determina en función de la distancia estructural al fenotipo objetivo.

$$\begin{aligned} \frac{dX_1}{dt} &= k_+ B_1 \left(R_T - \sum_{\sigma=1}^M X_\sigma \right) \left(L_T - \sum_{\sigma=1}^M X_\sigma \right) - k_{-,1} X_1 \\ &\vdots \\ \frac{dX_m}{dt} &= k_+ B_m \left(R_T - \sum_{\sigma=1}^M X_\sigma \right) \left(L_T - \sum_{\sigma=1}^M X_\sigma \right) - k_{-,m} X_m \\ &\vdots \\ \frac{dX_M}{dt} &= k_+ B_M \left(R_T - \sum_{\sigma=1}^M X_\sigma \right) \left(L_T - \sum_{\sigma=1}^M X_\sigma \right) - k_{-,M} X_M \end{aligned}$$

En el equilibrio el cambio en la cantidad de complejos formados es cero. Esto quiere decir que la suma de todos los complejos formados es una constante que podemos llamar β .

$$\begin{aligned} \frac{dX_m}{dt} &= k_+ B_m (R_T - \beta) (L_T - \beta) - k_{-,m} X_m = 0 \\ \text{siendo } \beta &= \sum_{m=1}^M X_m \end{aligned} \tag{3.20}$$

Además, la cantidad de RNA y ligando total (R_T, L_T) son constantes. Entonces el producto $(R_T - \beta) (L_T - \beta)$ también puede expresarse como una sola constante Z .

$$\begin{aligned} \frac{dX_m}{dt} &= k_+ B_m (Z) - k_{-,m} X_m = 0 \\ \text{siendo } Z &= (R_T - \beta) \cdot (L_T - \beta) \end{aligned} \tag{3.21}$$

Esta expresión nos permite despejar la cantidad de complejo X_m formado en función de valores que podemos conocer. A excepción de la constante Z .

$$X_m = Z \cdot \frac{k_+ B_m}{k_{-,m}}$$

donde

$$k_+ = \text{Cte. de asociación} \quad (3.22)$$

$$k_{-,m} = \text{Cte. de disociación para la estructura } m \text{ 2.4}$$

$$B_m = \text{Prob. de Boltzmann para la estructura } m \text{ 3.13}$$

Lo que a nosotros nos interesa en realidad, es la suma total de los múltiples complejos formados que en 3.20 definimos como β . Este es valor que se emplea como la adecuación. Nótese que si aplicamos la sumatoria para cada estructura dentro del repertorio ($1 \dots M$) en la ecuación 3.22, obtenemos algo equivalente a β .

$$\beta = \sum_{m=1}^M (X_m) = Z \cdot \sum_{m=1}^M \left(\frac{k_+ B_m}{k_{-,m}} \right) \quad (3.23)$$

siendo $Z = (R_T - \beta) \cdot (L_T - \beta)$

Tras sustituir Z , obtenemos una expresión cuya única incógnita es B y puede desarrollarse hasta una ecuación cuadrática.

$$a\beta^2 + b\beta + c = 0$$

donde

$$a = \sum_{m=1}^M \frac{k_+ B_m}{k_{-,m}} \quad (3.24)$$

$$b = - \left(R_T \cdot \sum_{m=1}^M \frac{k_+ B_m}{k_{-,m}} + L_T \cdot \sum_{m=1}^M \frac{k_+ B_m}{k_{-,m}} + 1 \right)$$

$$c = R_T \cdot L_T \cdot \sum_{m=1}^M \frac{k_+ B_m}{k_{-,m}}$$

Esta ecuación puede resolverse mediante la fórmula general

$$\beta = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (3.25)$$

Aunque esta ecuación cuadrática proporciona dos posibles raíces, el sistema posee una única solución con sentido físico. Dado que β representa la suma total de las concentraciones de complejos formados, no puede ser negativo. Además por simple conservación de masas, no puede exceder las concentraciones de RNA y Ligando totales (R_T) y (L_T).

$$0 \leq \beta < L_T \text{ y } 0 \leq \beta < R_T \quad (3.26)$$

Regresando a la expresión 3.24, podemos replantearla como una función $F(\beta)$, en la que cada tipo de complejo se describe como estado $f_m(\beta)$.

$$f_m(\beta) = \frac{k_+ B_m}{k_{-,m}} (R_T - \beta)(L_T - \beta) \quad (3.27)$$

Ya que las constantes k_+ , $k_{-,m}$ y concentraciones R_T , L_T son valores estrictamente positivos, podemos evaluar los extremos de esta función. El límite inferior se da cuando $\beta = 0$, en cuyo caso $f_m(0) > 0$. Mientras que el límite superior se define cuando la cantidad de complejo formado (β) iguala la cantidad de ligando o RNA disponible en el sistema ($\beta = L_T$ o $\beta = R_T$), resultando en ambos casos $f_m(L_T) = 0$ y $f_m(R_T) = 0$ respectivamente.

Por otro lado, la primera y segunda derivada de un componente $f_m(\beta)$ respecto a β son

$$\begin{aligned} \frac{df_m(\beta)}{d\beta} &= \frac{k_+ B_m}{k_{-,m}} (2\beta - R_T - L_T) \\ \frac{d^2 f_m(\beta)}{d\beta^2} &= \frac{2k_+ B_m}{k_{-,m}} \end{aligned} \quad (3.28)$$

Siendo $(2\beta - R_T - L_T)$ un valor negativo en el dominio 3.26, la primera derivada de la función $F(\beta)$ será también negativa. Esto significa que la función $F(\beta)$ decrece conforme aumenta la concentración de complejos β . En cambio, la segunda derivada es siempre positiva en el mismo dominio, ya que se compone únicamente del producto y cociente de constantes positivas. Lo cual describe un comportamiento continuo con forma cóncava hacia arriba.

La conclusión geométrica es que la curva empieza en un valor positivo en el origen, y decrece de forma constante hasta tocar el valor de cero cuando se agota el ligando o el RNA. Debido a que cruzará la línea identidad (donde $F(\beta) = \beta$) exactamente solo una vez. Esa única intersección representa el único punto de equilibrio con validez física del sistema [22].

3.4. Muestreo de red neutral de tRNA_{phe}

La estructura secundaria del RNA de transferencia unido a fenilalanina, tRNA_{phe}, es el fenotipo que elegimos para iniciar las simulaciones evolutivas. Con esto quiero decir que el genotipo inicial, el cual comparten todos los individuos de la población, produce la estructura tRNA_{phe} como su fenotipo más común. Para ver diferentes representaciones estructurales de tRNA_{phe} regrese a la Figura 8.

Múltiples secuencias son capaces de producir una misma estructura secundaria. El conjunto de secuencias capaces de producir esa estructura como estado de mínima energía libre (mfe), se conoce como la red neutral de la estructura.

Se ha reportado que las redes neutrales de estructuras de RNA tienden a estar muy fragmentadas. Esto quiere decir que la red suele estar dividida en regiones. Las secuencias de una región se parecen bastante entre sí en su estructura primaria, pero son muy diferentes a las que hay en otra región de la red. Lo anterior tiene que ser considerado para hacer un muestro de la red neutral de tRNA_{phe}. A partir de una secuencia aleatorizada empleo la función *inversefold* del software de *ViennaRNA*. Función que realiza permutaciones en los

ribonucleótidos de la secuencia, hasta producir una que adopta la estructura de interés como estado mfe[34, 27].

La secuencia que retorna *inversefold* pertenece una región de la red neutral de tRNAp_{he}. Aunque es posible es que este método nos reporte las estructuras de fácil acceso. Por este motivo realizamos una caminata aleatoria partiendo de la secuencia producida con *inversefold*. Cada paso es una mutación de ribonucleótido al azar, que no altera la identidad de la estructura producida como estado de mfe. Al final de esto nos queda secuencia mucho más representativa de la región en la que nos encontramos de la red neutral[35].

Realicé un muestreo de diez mil secuencias de la red neutral de tRNAp_{he}, siguiendo la estrategia descrita. Para cada secuencia se generó una secuencia aleatoria, se le aplicó *inversefold* y una caminata aleatoria de ochenta pasos para la estructura tRNAp_{he} (3.1).

Cuadro 3.1: Muestreo de la red neutral de tRNAp_{he}.

Input: Estructura tRNAp_{he}, muestreo $n = 10000$, pasos de caminata $p = 80$

Output: Conjunto \mathcal{M} de secuencias de la red neutral

```

Inicializar conjunto vacío  $\mathcal{M} \leftarrow \emptyset$ 
for  $i = 1$  to  $n$  do
     $S_{rand} \leftarrow$  GenerarSecuenciaAleatoria()
     $S_{actual} \leftarrow$  inverse_fold( $S_{rand}$ , tRNAphe)
     $pasos \leftarrow 0$ 
    while  $pasos < p$  do
         $S_{candidata} \leftarrow$  CopiarSecuencia( $S_{actual}$ )
         $S_{actual} \leftarrow$  MutarRibonucleotidoAlAzar( $S_{actual}$ )
        if mfe( $S_{actual}$ ) == tRNAphe then
             $pasos \leftarrow$   $pasos + 1$ 
        else % Retroceder mutación
             $S_{actual} \leftarrow S_{candidata}$ 
        end
    end
    Agregar  $S_{actual}$  al conjunto  $\mathcal{M}$ 
end
return  $\mathcal{M}$ 

```

3.5. Exploración de estructuras accesibles

A fin de establecer destinos para las simulaciones evolutivas, interesa explorar a qué red neutral se puede acceder desde la red de tRNAp_{he}. Debido a que este es un estudio exploratorio y no tenemos certeza en cómo se va a comportar el modelo de unión a ligando, conviene elegir destinos evolutivos cercanos. Es decir, estructuras que se pueden producir con una mutación puntual de ribonucleótido.

Partiendo de la muestra de red neutral de tRNA_{phe}. Apliqué una mutación aleatoria a cada una de las diez mil secuencias. Algunas de estas secuencias, modificaron su estructura de mínima energía libre (mfe) como consecuencia. Para este grupo registré la nueva estructura producida y su distancia estructural a tRNA_{phe}.

Una vez anotadas todas las estructuras producidas, realicé el conteo de cuantas veces se registró cada una de las estructuras. Obtuve así un compendio estructuras que contiene su frecuencia de aparición y su distancia estructural respecto a tRNA_{phe}. Mayor frecuencia sugiere más facilidad para acceder a esa red neutral, mientras que la distancia estructural describe que tan diferente es la estructura en comparación con la que producía originalmente la secuencia.

Ya que desconocemos de qué manera afectarán tanto frecuencia como distancia estructural las simulaciones evolutivas, elegí un grupo de estructuras tal que ambos factores sean incluidos (véase 3.2).

Cuadro 3.2: Estructuras seleccionadas τ .

Id.	Estructura Secundaria (mfe)	Frec.	Dist.
15_12	(((((.....(((.....))))).(((.....))))...((((.....)))))))))).	15	12
15_8	(((((.....(((.....))))).((((.....))))).((((.....)))))))))).	13*	8
15_4	(((((.....(((.....))))).(((.....))))).((((.....)))))))))).	15	4
30_4	(((((.....(((.....))))).(((.....))))).((((.....)))))))))).	29*	4
45_4	(((((.....(((.....)))))((((.....))))).((((.....)))))))))).	45	4

Nota: El identificador (Id.) de cada fila se construye a partir de la frecuencia de aparición y la distancia estructural de las configuraciones fenotípicas encontradas (**frecuencia_distancia**).

* Valores que fueron redondeados para la construcción del identificador.

3.6. Simulaciones evolutivas

Algo que quiero dejar muy claro, es que en las simulaciones, cada genotipo representa una célula que contiene un sistema independiente al resto de N de la población. Cada generación estas células mueren/desaparecen y son reemplazados por N células nuevas que son nuevos sistemas independientes entre sí.

Se realizaron doscientas simulaciones evolutivas por cada estructura destino 3.2. Cada simulación se ejecutó una vez bajo cada criterio de adecuación, pero manteniendo la misma generación de números aleatorios en las tres ocasiones (véase 3.3).

Al iniciar las simulaciones se genera una secuencia aleatoria a la que se aplica *inversefold* con la estructura tRNA_{phe}. La secuencia producida representa el genotipo inicial de cada individuo de la población. Cada simulación evolutiva termina cuando supera veinte mil generaciones. Los parámetros utilizados en general y por criterio de adecuación se resumen a continuación (véase 3.4).

Cuadro 3.3: Diseño experimental y número total de simulaciones evolutivas.

Id.	Criterios de Adecuación	Cant. simul.
	Ancel y Fontana (sin plasticidad)	× 200
15_12	Ancel y Fontana (con plasticidad)	× 200
	Modelo de unión a ligando	× 200
15_8	Los mismos tres criterios anteriores	× 600
15_4	Los mismos tres criterios anteriores	× 600
30_4	Los mismos tres criterios anteriores	× 600
45_4	Los mismos tres criterios anteriores	× 600
Total		3000

Nota: Para cada una de las 5 estructuras destino se ejecutan 200 réplicas independientes bajo 3 condiciones de selección distintas, resultando en un total de 600 simulaciones por estructura y 3000 ejecuciones en todo el estudio.

Cuadro 3.4: Parámetros globales y específicos de las simulaciones evolutivas.

Parámetros	Símbolo	Valor
Generales		
Tamaño de población	N	1000
Número de generaciones	t	20 000
Tasa de mutación	μ	0,0005
Modelos de Ancel y Fontana		
Constante C	C	0,01
Modelo de unión a ligando		
RNA total	R_T	20
Ligando total	L_T	80
Constante de asociación	k_+	0,001
Constante de disociación a	a	0,001
Constante de disociación b	b	0,05

Nota: Aunque no se menciona como un parámetro, las estructuras que se incluyen dentro del repertorio plástico son aquellas en el rango de 2 kcal/mol con respecto al valor de ΔG del estado con menor energía libre.

3.7. Análisis y pruebas estadísticas

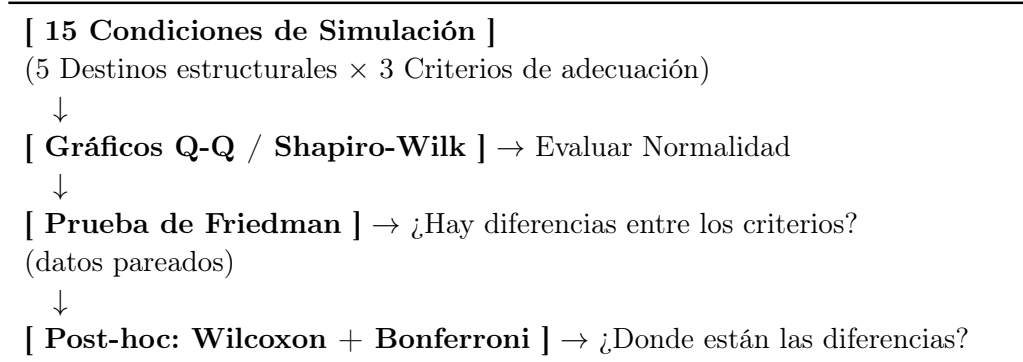
Para visualización de datos utilicé la biblioteca de *matplotlib* disponible para el lenguaje de programación Python. Así como la biblioteca *SciPy* para la aplicación de pruebas estadísticas.

La normalidad de los datos obtenidos se evaluó mediante pruebas de Shapiro-Wilk. También se utilizaron graficos Q-Q como conformación del comportamiento de los datos. Siendo

la mayoría de los datos no normales, decidí utilizar técnicas de análisis no paramétricas.

Para cada grupo de simulaciones asociado a un mismo destino estructural (véase 3.3), las diferencias estadísticas entre los tres criterios de adecuación las evalué mediante una prueba de varianza de dos vías de Friedman (equivalente no paramétrico al ANOVA de medidas repetidas). Posteriormente realicé comparaciones múltiples por pares, empleando la prueba de rangos con signo de Wilcoxon; aplicando la corrección de Bonferroni para el ajuste de los valores p .

Cuadro 3.5: Flujo del análisis estadístico.



3.8. Acceso al código desarrollado

Este proyecto requirió de gran cantidad de tiempo dedicada a desarrollar códigos en el lenguaje de programación C++. Todo el código utilizado se puede descargar siguiendo el QR o link incluido a continuación.



https://drive.google.com/drive/folders/1hpk_S0np1frSRATESd195TnjXa6dAq_k?usp=drive_link

En caso de presentar algún problema para acceder a los archivos, por favor escribir al correo electrónico: alegabriel.moreno@gmail.com.

Resultados

En cada simulación registré la adecuación máxima y el promedio de la población en cada generación. Estos valores se normalizaron respecto al máximo posible en cada criterio de adecuación, transformándolos en valores de adecuación relativa. Para realizar una comparación justa, exploré cuántas de las simulaciones alcanzaron diferentes umbrales de adecuación y en qué generación. Para cada caso elegí un umbral tal que al menos el ochenta por ciento de las simulaciones en cada estructura y destino evolutivo alcanzaran este valor (véase Figura 10). Los umbrales elegidos son 0.82 para adecuación promedio y 0.90 para adecuación máxima.

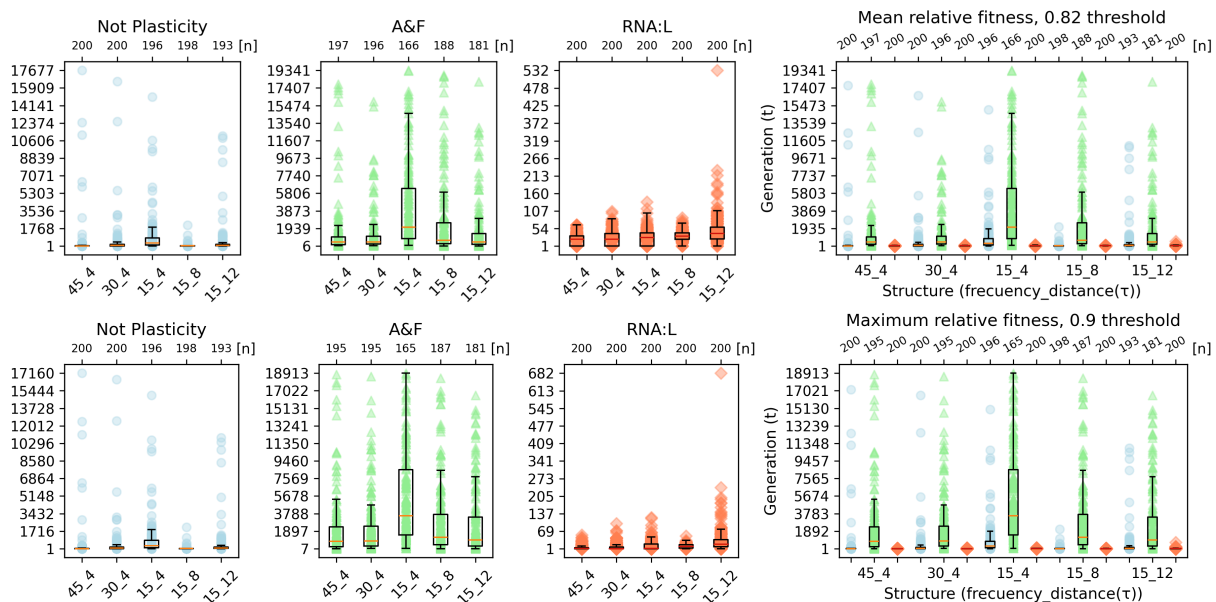


Figura 10. Umbrales elegidos para valores de adecuación.

Las gráficas describen en qué generación las simulaciones alcanzan/superan el umbral en el valor rastreado. Las gráficas superiores se refieren al valor de adecuación relativa promedio, las inferiores a la adecuación relativa máxima. Simulaciones bajo el criterio sin plasticidad estructural se representan con círculos azules, triángulos verdes para simulaciones con plasticidad estructural según el criterio de Ancel y Fontana (A&F). Los rombos naranjas son simulaciones realizadas bajo el criterio de unión a ligando (RNA:L). El valor [n] indica cuántas de las simulaciones están representadas, las no incluidas no alcanzan el umbral de adecuación indicado. Debajo de cada grupo se indica el destino estructural, el primer número indica la accesibilidad de la estructura y el segundo la distancia de árbol respecto la estructura inicial de la población (consulte 3.2). Todos los umbrales explorados se incluyen en el apéndice.

En la Figura 10, las gráficas de caja tienen muchos puntos atípicos. La escala dificulta la interpretación debido a que una gran cantidad de puntos está representada en generaciones tempranas. Esto sugiere que las tendencias de convergencia siguen distribuciones fuertemente sesgadas con una cola muy larga. Considerando este comportamiento de los datos, decidí revisar si encajaban mejor con una distribución log-normal (véase Figura 11).

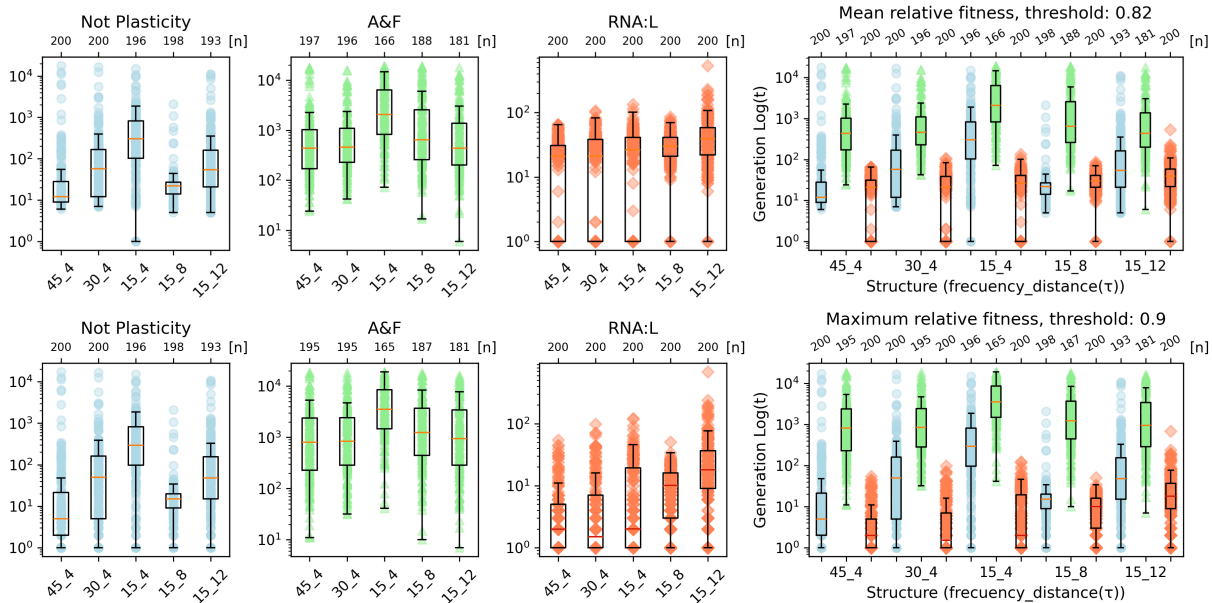


Figura 11. Umbrales de adecuación con generaciones en escala logarítmica.

A diferencia de la Figura 10, estas gráficas presentan el número de generaciones en una escala logarítmica para facilitar la visualización de los datos debido al sesgo y la presencia de valores atípicos. Las gráficas describen en qué generación las simulaciones alcanzan/superan el umbral en el valor rastreado (0.82 para adecuación promedio en las gráficas superiores y 0.90 para adecuación máxima en las inferiores). Se mantiene la misma codificación: círculos azules para simulaciones sin plasticidad, triángulos verdes para plasticidad según AnceL y Fontana (A&F) y rombos naranjas para unión a ligando (RNA:L). El valor [n] indica el número de simulaciones representadas. Debajo de cada grupo se indica el destino estructural, siendo el primer número la accesibilidad de la estructura y el segundo la distancia de árbol respecto la estructura inicial de la población (consulte 3.2).

En este caso, la representación a escala logarítmica ayuda notablemente a visualizar la distribución de nuestros datos. Realicé pruebas de Shapiro-Wilk a cada grupo, evaluando la normalidad bajo esta escala (véase Tabla 4.6). Para la gran mayoría de los grupos se reportó un valor $p < 0,001$ en esta prueba, lo que permite rechazar formalmente la hipótesis de normalidad. Dado que la prueba de Shapiro-Wilk es altamente sensible a pequeñas desviaciones en muestras grandes, decidí corroborar el comportamiento de las distribuciones mediante el análisis visual de gráficos Q-Q (véase Figura 12).

Tanto los gráficos Q-Q como las pruebas de Shapiro-Wilk evidenciaron que la gran mayoría de los grupos siguen distribuciones no normales. Razón por la cual opté por el uso de estadística no paramétrica. Para confirmar si algún criterio de adecuación confirió una ven-

Cuadro 4.6: Prueba de normalidad de Shapiro-Wilk.

Id	Modelo/Criterio	Adecuación Promedio		Adecuación Máxima	
		Estadístico W	Valor p	Estadístico W	Valor p
45_4	Sin plasticidad	0.7009	$1,173 \times 10^{-18}$	0.8575	$1,012 \times 10^{-12}$
45_4	Ancel y Fontana	0.9763	$2,008 \times 10^{-03}$	0.9875	$8,253 \times 10^{-02}$
45_4	De unión a Ligando	0.7404	$1,992 \times 10^{-17}$	0.8096	$6,791 \times 10^{-15}$
30_4	Sin plasticidad	0.9356	$9,623 \times 10^{-08}$	0.9614	$2,778 \times 10^{-05}$
30_4	Ancel y Fontana	0.9762	$2,014 \times 10^{-03}$	0.9865	$6,001 \times 10^{-02}$
30_4	De unión a Ligando	0.7423	$2,305 \times 10^{-17}$	0.7850	$7,329 \times 10^{-16}$
15_4	Sin plasticidad	0.9622	$4,143 \times 10^{-05}$	0.9245	$1,655 \times 10^{-08}$
15_4	Ancel y Fontana	0.9750	$4,273 \times 10^{-03}$	0.9449	$4,917 \times 10^{-06}$
15_4	De unión a Ligando	0.7473	$3,375 \times 10^{-17}$	0.8189	$1,667 \times 10^{-14}$
15_8	Sin plasticidad	0.7990	$3,095 \times 10^{-15}$	0.8640	$2,563 \times 10^{-12}$
15_8	Ancel y Fontana	0.9761	$2,578 \times 10^{-03}$	0.9827	$2,051 \times 10^{-02}$
15_8	De unión a Ligando	0.7928	$1,457 \times 10^{-15}$	0.9028	$3,803 \times 10^{-10}$
15_12	Sin plasticidad	0.9573	$1,442 \times 10^{-05}$	0.9839	$2,571 \times 10^{-02}$
15_12	Ancel y Fontana	0.9763	$3,591 \times 10^{-03}$	0.9812	$1,540 \times 10^{-02}$
15_12	De unión a Ligando	0.9460	$7,954 \times 10^{-07}$	0.9623	$3,544 \times 10^{-05}$

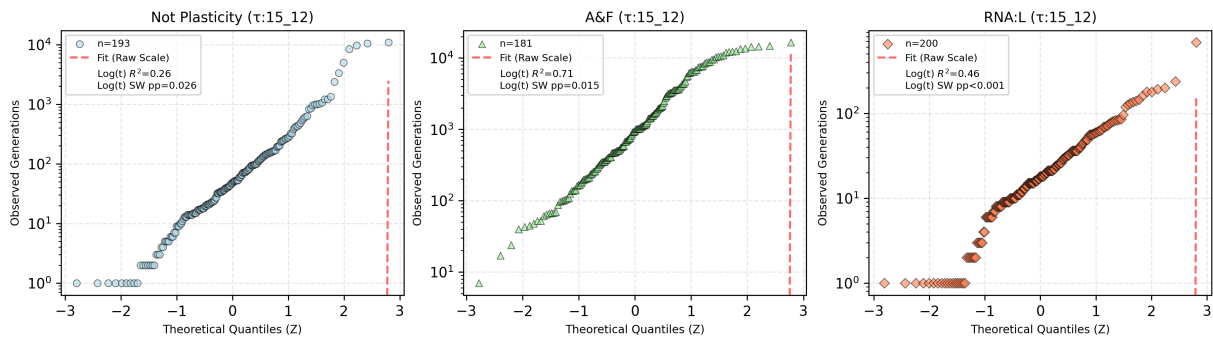


Figura 12. Gráficos de cuantiles-cuantiles (Q-Q), adecuación relativa máxima.

Se muestran los resultados correspondientes a las simulaciones con destino estructural 15_12 (véase 3.2) bajo las tres condiciones: sin plasticidad estructural (círculos azules), con plasticidad según el criterio de Ancel y Fontana (triángulos verdes) y bajo el modelo de unión a ligando (rombos naranjas). Los cuantiles observados de los datos transformados (\log_{10}) se contrastan en el eje horizontal con los cuantiles teóricos esperados de una distribución normal; la línea roja puntuada representa el ajuste esperado para los datos no transformados. Para cada escenario se reporta el valor p de la prueba de Shapiro-Wilk aplicada a los datos transformados y el coeficiente de determinación (R^2) respecto a la distribución teórica.

taja evolutiva significativa frente a los otros, apliqué la prueba de Friedman para muestras pareadas.

Recordemos que cuando hablo de destino estructural, me refiero a una estructura secundaria que se representa con dos números separados por un guion bajo. Mientras más grande es el primer número, más accesible es llegar a ella mediante mutaciones. En cambio, el segundo valor indica su distancia de árbol con respecto a la estructura desde la cual inicia la evolución.

En todos los grupos de simulaciones (45_4, 30_4, 15_4, 15_8, 15_12) se encontraron diferencias globales altamente significativas ($p < 0,001$). Posteriormente, realicé un análisis *post-hoc* mediante la prueba de rangos con signo de Wilcoxon a dos colas, aplicando la corrección de Bonferroni para comparaciones múltiples (véase Tabla 4.7). La gran mayoría de las comparaciones por pares presentaron diferencias estadísticamente significativas. Más allá del valor estadístico, como se observa visualmente en las distribuciones, el modelo RNA:L no solo difiere de los demás, sino que logra alcanzar los umbrales de adecuación en décadas de generaciones frente a miles en los otros modelos.

Si bien el análisis de la adecuación relativa permite observar cómo actúa cada criterio matemático, para comprobar que la simulación realmente cumple su objetivo biológico, registré cuántas de las estructuras de mínima energía (MFE) expresadas por la población tenían una distancia estructural de cero con respecto al fenotipo objetivo. Es decir, en cada generación evalué la estructura de mínima energía que produce cada genotipo, calculé su distancia estructural hacia el fenotipo objetivo de la simulación y anoté cuántos genotipos de la población total producen físicamente el fenotipo objetivo como su MFE.

Al igual que para los valores de adecuación relativa, determiné un umbral de convergencia. Específicamente, busqué el número de individuos que producen la estructura objetivo tal que todos los conjuntos de simulaciones, agrupados por destino estructural y modelo, alcancen este umbral en al menos el ochenta por ciento de los casos (véase Figura 13).

Cuadro 4.7: Prueba de Friedman y comparaciones *post-hoc* de Wilcoxon.

Id	<i>n</i>	Friedman (<i>p</i>)	Comparación	Wilcoxon (<i>p_{ajustado}</i>)
Adecuación relativa promedio				
45_4	197	$1,968 \times 10^{-58}$	Sin Plast. vs. A&F	$1,4 \times 10^{-26}$
			Sin Plast. vs. RNA:L	$4,0 \times 10^{-02}$
			A&F vs. RNA:L	$1,4 \times 10^{-33}$
30_4	196	$7,201 \times 10^{-65}$	Sin Plast. vs. A&F	$3,4 \times 10^{-22}$
			Sin Plast. vs. RNA:L	$1,1 \times 10^{-21}$
			A&F vs. RNA:L	$1,9 \times 10^{-33}$
15_4	164	$6,756 \times 10^{-60}$	Sin Plast. vs. A&F	$5,7 \times 10^{-22}$
			Sin Plast. vs. RNA:L	$1,5 \times 10^{-26}$
			A&F vs. RNA:L	$3,5 \times 10^{-28}$
15_8	187	$8,983 \times 10^{-66}$	Sin Plast. vs. A&F	$4,8 \times 10^{-31}$
			Sin Plast. vs. RNA:L	$1,2 \times 10^{-14}$
			A&F vs. RNA:L	$6,0 \times 10^{-32}$
15_12	176	$1,417 \times 10^{-48}$	Sin Plast. vs. A&F	$3,2 \times 10^{-21}$
			Sin Plast. vs. RNA:L	$3,5 \times 10^{-09}$
			A&F vs. RNA:L	$4,4 \times 10^{-30}$
Adecuación relativa máxima				
45_4	195	$2,763 \times 10^{-72}$	Sin Plast. vs. A&F	$2,0 \times 10^{-27}$
			Sin Plast. vs. RNA:L	$3,8 \times 10^{-16}$
			A&F vs. RNA:L	$2,8 \times 10^{-33}$
30_4	195	$1,705 \times 10^{-71}$	Sin Plast. vs. A&F	$4,6 \times 10^{-26}$
			Sin Plast. vs. RNA:L	$2,7 \times 10^{-26}$
			A&F vs. RNA:L	$2,8 \times 10^{-33}$
15_4	163	$1,415 \times 10^{-64}$	Sin Plast. vs. A&F	$3,0 \times 10^{-24}$
			Sin Plast. vs. RNA:L	$7,4 \times 10^{-27}$
			A&F vs. RNA:L	$5,1 \times 10^{-28}$
15_8	186	$6,321 \times 10^{-66}$	Sin Plast. vs. A&F	$4,5 \times 10^{-31}$
			Sin Plast. vs. RNA:L	$6,2 \times 10^{-09}$
			A&F vs. RNA:L	$8,5 \times 10^{-32}$
15_12	176	$1,788 \times 10^{-61}$	Sin Plast. vs. A&F	$5,2 \times 10^{-23}$
			Sin Plast. vs. RNA:L	$1,6 \times 10^{-20}$
			A&F vs. RNA:L	$3,9 \times 10^{-30}$

Nota: Criterios de adecuación: **Sin Plast.** (Sin plasticidad estructural), **A&F** (Plasticidad según Ancel y Fontana) y **RNA:L** (Modelo de unión a ligando).

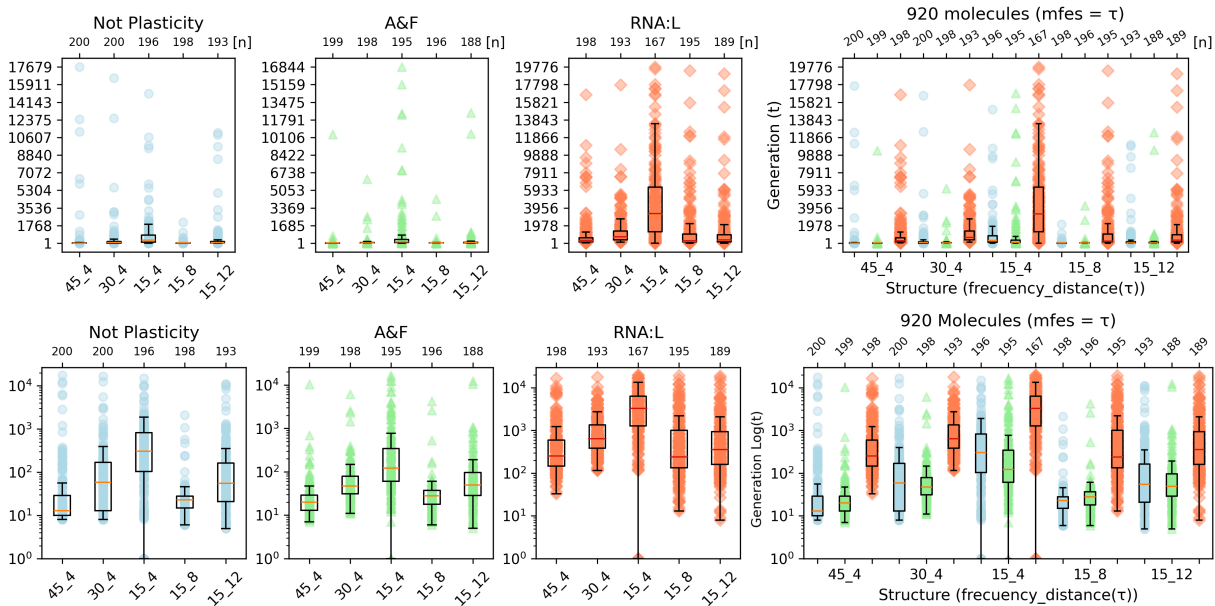


Figura 13. Umbral de estructuras mínima energía = τ .

Las gráficas describen en cuántas generaciones las simulaciones alcanzan/superan el umbral de 920 individuos cuya estructura mfe sea τ . Círculos azules para simulaciones sin plasticidad, triángulos verdes para plasticidad según Ancel y Fontana (A&F) y rombos naranjas para unión a ligando (RNA:L). El valor [n] indica el número de simulaciones representadas. Debajo de cada grupo se indica el destino estructural.

Para el análisis estadístico de esta métrica estructural apliqué las mismas pruebas previamente descritas: Shapiro-Wilk y gráficos Q-Q para evaluar la normalidad (Tabla 4.8), seguidas de la prueba de Friedman para detectar diferencias globales y la prueba *post-hoc* de Wilcoxon con corrección de Bonferroni para las comparaciones por pares (Tabla 4.9). Los gráficos Q-Q y las figuras correspondientes a todos los umbrales evaluados se incluyen en el apéndice.

En las pruebas *post-hoc*, solo en un caso no se encontró diferencia estadística significativa: la comparación entre los criterios sin plasticidad y con plasticidad de Ancel y Fontana con destino estructural 15_12, puesto que su valor p superó el umbral de significancia ($p_{ajustado} = 1,0$).

Para las tres variables rastreadas: adecuación relativa promedio, adecuación relativa máxima y la cantidad de individuos que expresan el destino evolutivo como su estructura MFE, representé su dinámica evolutiva a lo largo de las generaciones en las Figuras 14, 15 y 16, respectivamente.

Cuadro 4.8: Prueba Shapiro-Wilk para cantidad de estructuras $MFE = \tau$.

Id	Modelo/Criterio	Estadístico W	Valor p
45_4	Sin plasticidad	0.6860	$4,346 \times 10^{-19}$
	Ancel y Fontana	0.7982	$2,616 \times 10^{-15}$
	Unión a ligando	0.9385	$1,920 \times 10^{-07}$
30_4	Sin plasticidad	0.9299	$3,331 \times 10^{-08}$
	Ancel y Fontana	0.8935	$1,153 \times 10^{-10}$
	Unión a ligando	0.9692	$3,009 \times 10^{-04}$
15_4	Sin plasticidad	0.9642	$6,927 \times 10^{-05}$
	Ancel y Fontana	0.9637	$6,447 \times 10^{-05}$
	Unión a ligando	0.8251	$7,205 \times 10^{-13}$
15_8	Sin plasticidad	0.7920	$1,646 \times 10^{-15}$
	Ancel y Fontana	0.8132	$1,389 \times 10^{-14}$
	Unión a ligando	0.9642	$7,301 \times 10^{-05}$
15_12	Sin plasticidad	0.9539	$6,573 \times 10^{-06}$
	Ancel y Fontana	0.9346	$1,692 \times 10^{-07}$
	Unión a ligando	0.9849	$4,013 \times 10^{-02}$

Cuadro 4.9: Friedman y post-hoc con Wilcoxon para cantidad de estructuras $MFE = \tau$.

Id	n	Friedman (p)	Comparación	Wilcoxon ($p_{ajustado}$)
45_4	197	$9,557 \times 10^{-60}$	Sin Plast. vs. A&F	$3,2 \times 10^{-02}$
			Sin Plast. vs. RNA:L	$4,5 \times 10^{-27}$
			A&F vs. RNA:L	$8,5 \times 10^{-32}$
30_4	191	$1,337 \times 10^{-56}$	Sin Plast. vs. A&F	$4,3 \times 10^{-03}$
			Sin Plast. vs. RNA:L	$3,0 \times 10^{-27}$
			A&F vs. RNA:L	$1,4 \times 10^{-31}$
15_4	163	$2,181 \times 10^{-43}$	Sin Plast. vs. A&F	$8,2 \times 10^{-05}$
			Sin Plast. vs. RNA:L	$1,1 \times 10^{-23}$
			A&F vs. RNA:L	$9,1 \times 10^{-25}$
15_8	192	$1,942 \times 10^{-67}$	Sin Plast. vs. A&F	$1,3 \times 10^{-10}$
			Sin Plast. vs. RNA:L	$1,6 \times 10^{-31}$
			A&F vs. RNA:L	$1,3 \times 10^{-31}$
15_12	177	$2,253 \times 10^{-43}$	Sin Plast. vs. A&F	$1,0 \times 10^{+00}$
			Sin Plast. vs. RNA:L	$6,0 \times 10^{-21}$
			A&F vs. RNA:L	$2,4 \times 10^{-25}$

Nota: El valor $p_{ajustado}$ corresponde a la corrección de Bonferroni. Abreviaturas: **Sin Plast.** (Sin plasticidad), **A&F** (Ancel y Fontana), **RNA:L** (Unión a ligando).

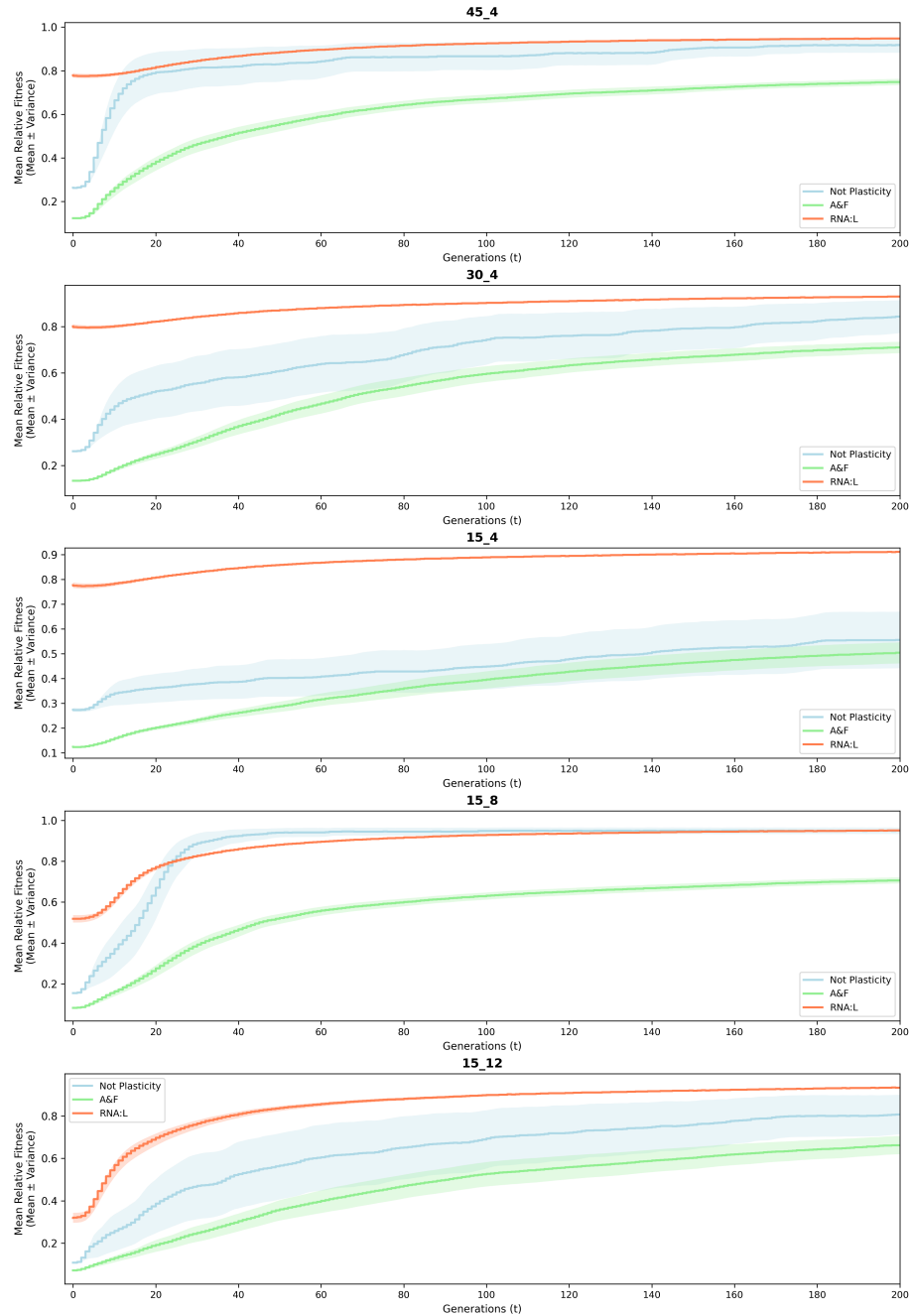


Figura 14. Dinámica evolutiva promedio de las simulaciones respecto a la adecuación relativa promedio.

Para el total de las simulaciones realizadas, promedié los valores por generación e incorporé la varianza como medida de dispersión. Abreviaturas: **Sin Plast.** (Sin plasticidad), **A&F** (Ancel y Fontana) y **RNA:L** (Unión a ligando). La máxima generación mostrada en el eje horizontal se seleccionó de manera independiente para cada grupo con el fin de optimizar la visualización hasta que la dinámica de los tres modelos se estabilizara. En la parte superior de cada panel se indica el destino estructural de las simulaciones.

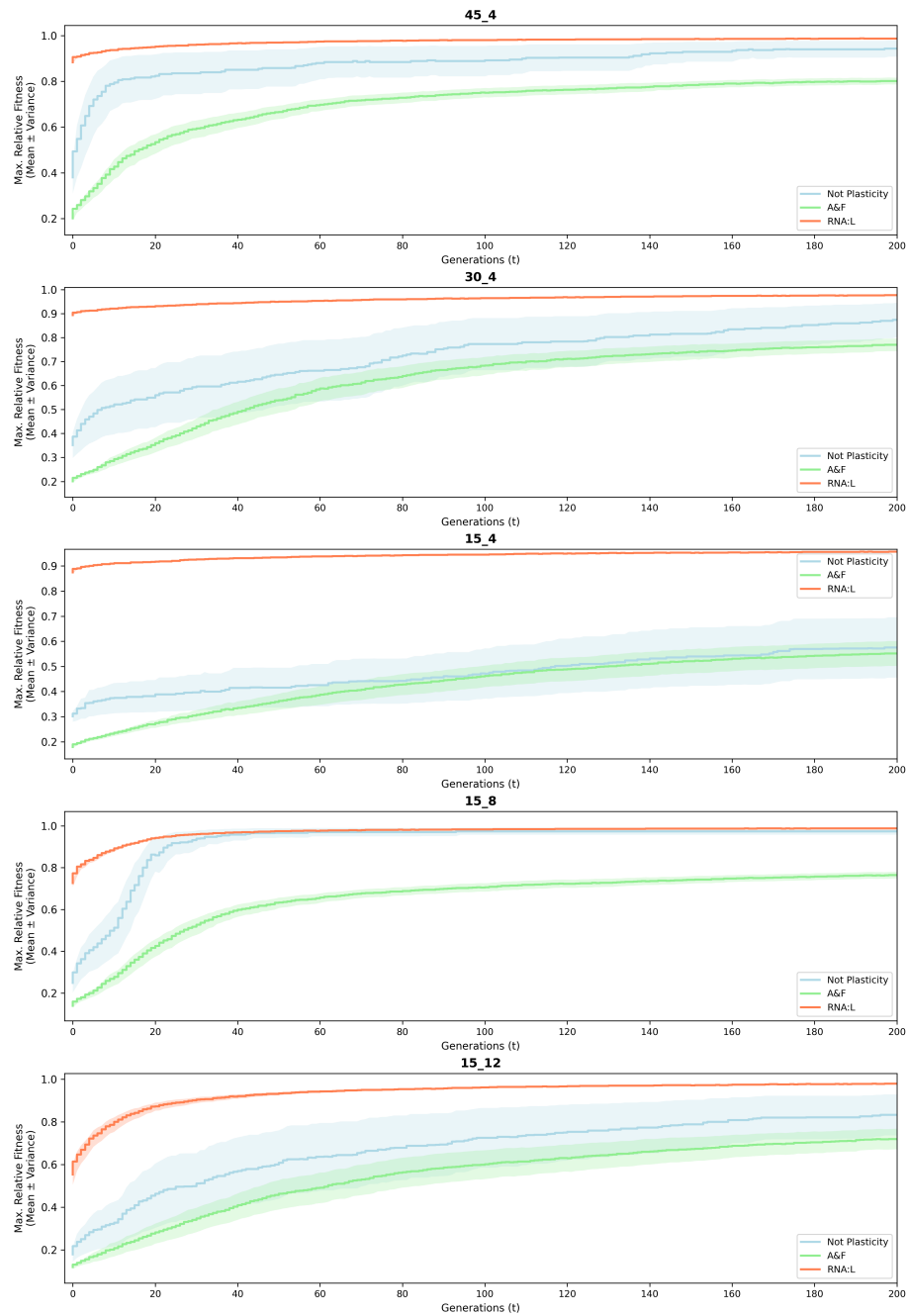


Figura 15. Dinámica evolutiva promedio de las simulaciones respecto a la adecuación relativa máxima.

Se presenta la trayectoria generacional del valor máximo de adecuación alcanzado por las poblaciones. Las curvas representan el promedio de las réplicas por generación para cada modelo, acompañadas de su respectiva varianza como banda de dispersión. Se conserva la codificación de colores y abreviaturas de la figura anterior. El truncamiento del eje temporal se ajustó individualmente por destino estructural para capturar con detalle la ventana de máxima tasa de cambio hasta la saturación del valor adaptativo.

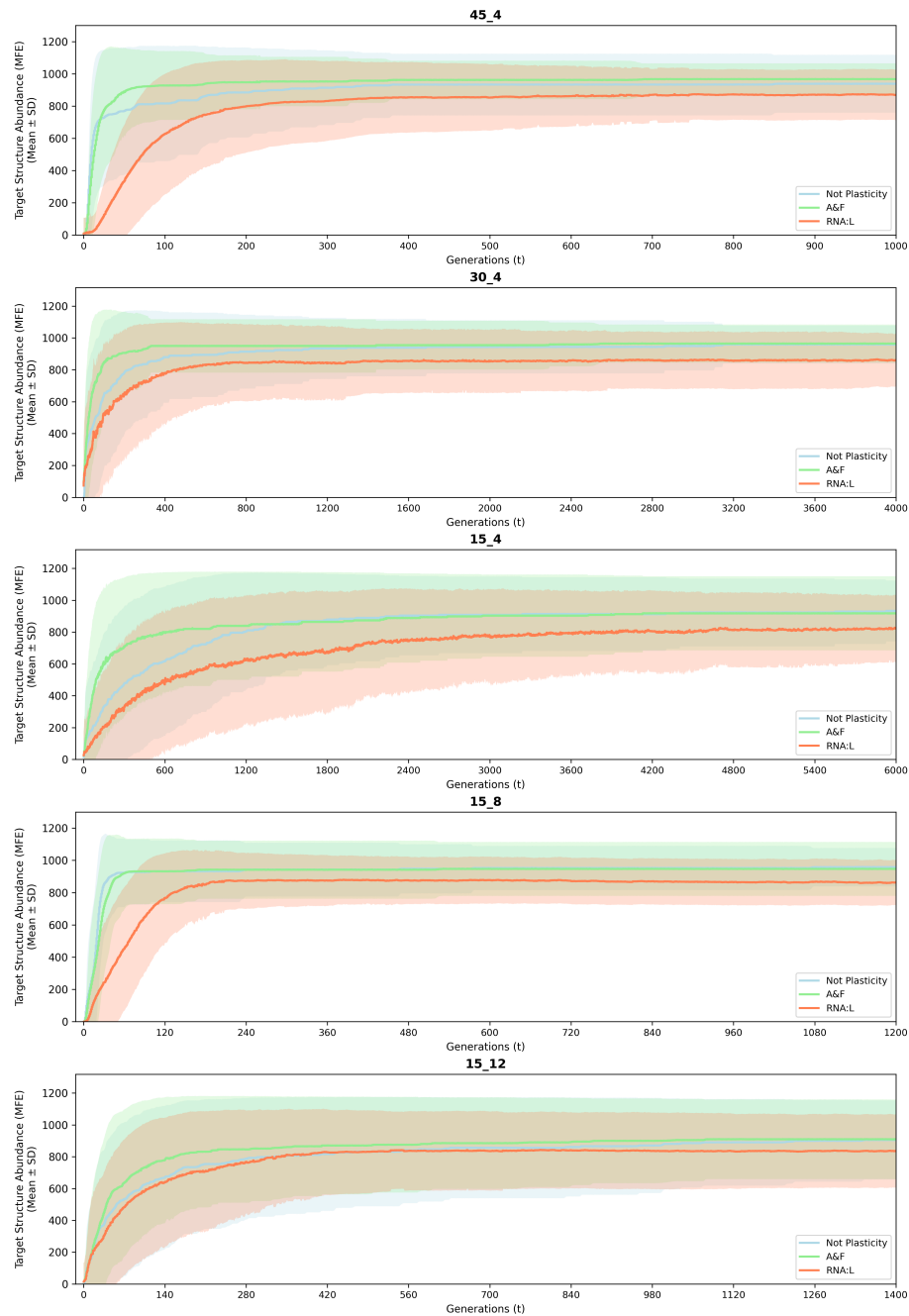


Figura 16. Dinámica evolutiva promedio de la cantidad de genotipos en la población que expresan el fenotipo objetivo como su estructura de mínima energía ($MFE = \tau$).

Las curvas describen el curso temporal del número total de individuos de la población (escala absoluta) que logran consolidar el fenotipo objetivo como su estructura óptima MFE. Se graficó el valor promedio por generación de las simulaciones correspondientes a cada modelo matemático junto con su varianza asociada. Esta representación permite contrastar directamente la velocidad física de fijación estructural en la población.

Discusión

La plasticidad estructural y la formación de complejos RNA-ligando facilitan la evolución de nuevas estructuras secundarias de RNA

El objetivo central de este proyecto fue evaluar si el modelo de unión a ligando (RNA:L) permite a las poblaciones de RNA *in silico* evolucionar estructuralmente de manera eficiente, superando los estancamientos evolutivos reportados en el modelo de Ancel y Fontana. Los resultados confirman que esta premisa se cumple, aunque con ciertas limitantes importantes en el estudio.

En principio, la dinámica observada en la Figura 16 insinúa que el modelo de unión a ligando tarda más tiempo en fijar el fenotipo objetivo como la estructura de mínima energía (MFE) absoluta en la población. Esto ocurre porque, bajo el modelo RNA:L, un genotipo adquiere una ventaja selectiva considerable simplemente conteniendo la estructura como un fenotipo accesible dentro de su repertorio plástico, sin la necesidad estricta de que esta sea inmediatamente su estructura de MFE [22].

En términos de adecuación relativa, el modelo de unión a ligando demostró ser notablemente superior a los otros dos criterios. Esta aceleración adaptativa se debe a que el modelo RNA:L evalúa la capacidad de una molécula externa (ligando) para secuestrar y estabilizar transitoriamente la conformación objetivo. Si la secuencia puede acceder al fenotipo funcional dentro de su repertorio estructural, incluso con una probabilidad de Boltzmann baja, el ligando favorece que esta estructura se establezca. Este favoritismo inducido por el ligando es precisamente lo que retarda su fijación como la estructura de mayor probabilidad para las secuencias. Sin embargo, este mecanismo de secuestro es exactamente lo que le debe permitir a la población escapar del fenómeno de confinamiento neutral al recorrer distancias estructurales más retadoras.

El modelo de Ancel y Fontana tiende a generar las condiciones que provocan el estancamiento evolutivo conforme transcurren más generaciones [21, 22]. Un problema considerable en mis simulaciones fue que convergieron hacia el destino evolutivo en muy pocas generaciones. Ante esta velocidad, el fenómeno de confinamiento neutral simplemente no tuvo el tiempo necesario para ocurrir. Con esto quiero decir que el modelo de unión a ligando demostraría su verdadera valía cuando el destino evolutivo se vuelve más distante y el paisaje adaptativo más abrupto. De hecho, esta tendencia se hace evidente en el caso de la estructura 15_12 (Figura 16), donde al incrementarse la complejidad, la dinámica de fijación del modelo RNA:L finalmente se empareja con la del modelo sin plasticidad.

Otro factor metodológico a destacar radica en la selección de la secuencia inicial. Al inicio del experimento, utilicé un algoritmo de *inversefold* para determinar la secuencia de la que partiría la evolución. En retrospectiva, considero que debí haber realizado una caminata aleatoria (*random walk*) sobre la red neutral. Es altamente probable que las secuencias de inicio generadas por *inversefold* resultaran ser muy ‘fáciles de convencer’ para cambiar su estructura, debido a que este algoritmo corre el riesgo de generar secuencias en zonas de alta conectividad y fácil acceso dentro de su red neutral.

Limitaciones del diseño experimental y perspectivas futuras

A pesar de la claridad en la divergencia de los modelos, el diseño exploratorio de este proyecto presentó una limitación metodológica importante impuesta por la proximidad estructural de los objetivos elegidos y el sesgo de la secuencia inicial. Al partir de una secuencia predefinida mediante *inversefold*, las estructuras seleccionadas como destinos evolutivos se encontraban a distancias estructurales relativamente cortas. Como consecuencia, la convergencia fue tan rápida para todos los criterios evaluados que el diseño experimental limitó severamente la ventana temporal para observar fenómenos evolutivos a largo plazo.

Una perspectiva fundamental que surge de esta investigación es que el ensayo arrojará información mucho más relevante cuando se garantice que la estructura de inicio y el fenotipo objetivo sean estructuralmente más lejanos. La realización de futuras simulaciones partiendo de secuencias obtenidas mediante caminatas en la red neutral, y apuntando a destinos evolutivos distantes, será indispensable para explotar el potencial del modelo RNA:L y observar a detalle cómo la plasticidad mediada por ligando evita el confinamiento neutral.

Finalmente, la relevancia biológica del modelo RNA:L se ve fuertemente respaldada con trabajo complementario realizado recientemente dentro de nuestro grupo de investigación. En este proyecto paralelo, se analizaron complejos RNA-proteína depositados en el *Protein Data Bank* (PDB). Al extraer las secuencias de RNA y acceder a su repertorio plástico *in silico*, se encontró que la conformación unida a la proteína aparece frecuentemente dentro del repertorio de la molécula con alta probabilidad de Boltzmann. Esto apoya que la interacción RNA-ligando opera mediante un mecanismo de selección conformacional y refuerza que el modelo de unión a ligando evaluado en esta tesis representa una aproximación biológica realista (Cruz-Moreno, Toro-Cruz, Olivares-Illana y Espinosa-Soto, en preparación).

Referencias

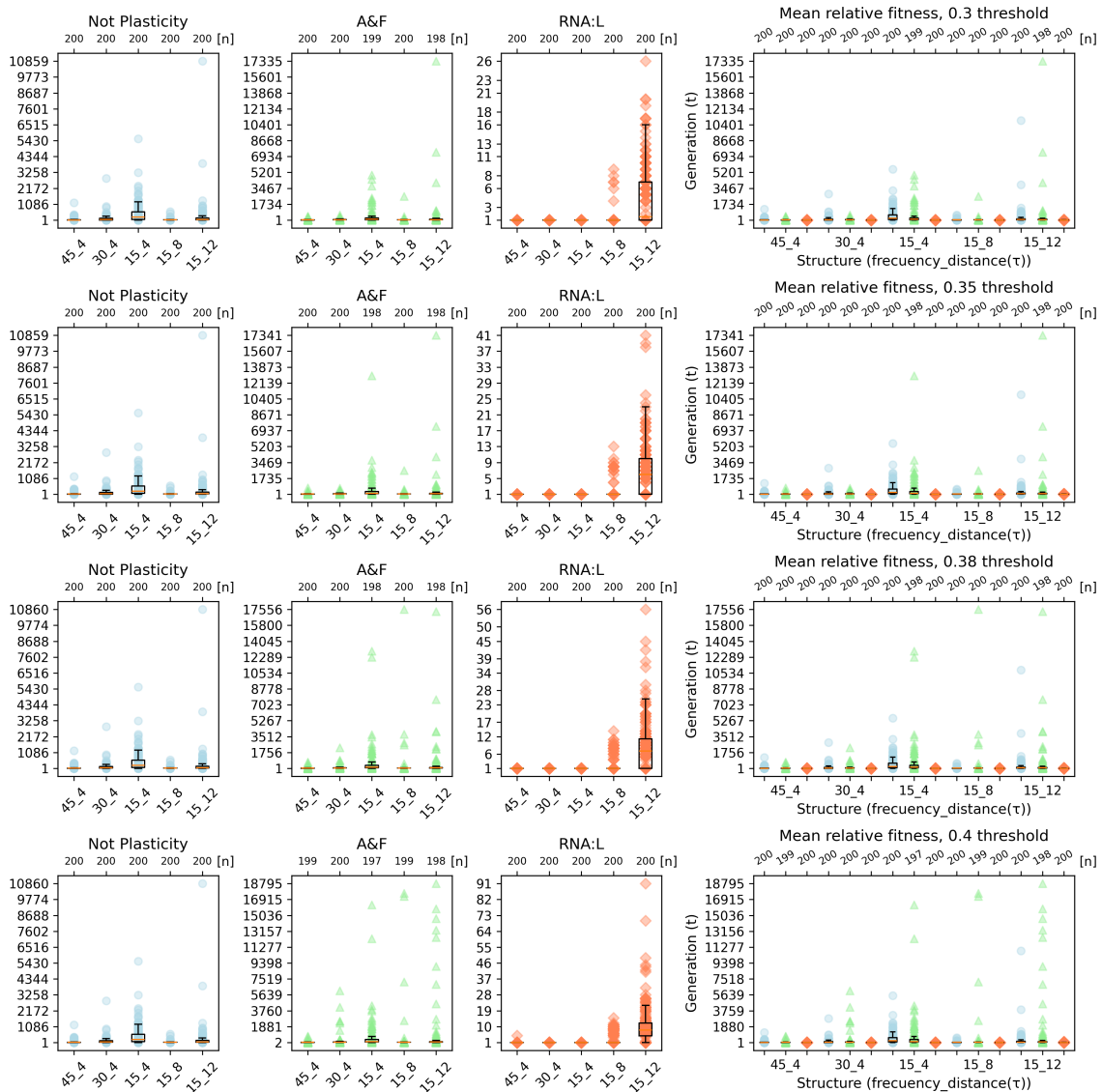
- [1] Julian S. Huxley. *Evolution: The Modern Synthesis*. Definitive Edition. Cambridge, MA: MIT Press, 2009.
- [2] Roger T Hanlon. «Cephalopod dynamic camouflage». En: *Current Biology* 17.11 (2007), R400-R404. DOI: 10.1016/j.cub.2007.03.034.
- [3] Mason D Ramirez y Todd H Oakley. «Eye-independent, light-activated chromatophore expansion (LACE) and expression of phototransduction genes in the skin of Octopus bimaculoides». En: *Journal of Experimental Biology* 218.10 (2015), págs. 1513-1520. DOI: 10.1242/jeb.118703.
- [4] Alejandro M. Serrano et al. «Following the star: Inflorescence heliotropism». En: *Environmental and Experimental Botany* 147 (2018), págs. 75-85. DOI: 10.1016/j.envepbot.2017.11.007. URL: <https://www.sciencedirect.com/science/article/pii/S0098847217302800>.
- [5] A. Shibata et al. «Flower movement induced by weather-dependent tropism satisfies attraction and protection». En: *Nature Communications* 16 (2025), pág. 4132. DOI: 10.1038/s41467-025-59337-6. URL: <https://doi.org/10.1038/s41467-025-59337-6>.
- [6] P. R. Rigby e Y. Sakurai. «Temperature and feeding related growth efficiency of immature octopuses *Enteroctopus dofleini*». En: *Aquaculture Science* 52.1 (2004), págs. 29-36. DOI: 10.11233/aquaculturesci1953.52.29.
- [7] Christina L. Conrath y M. E. Connors. «Aspects of the reproductive biology of the North Pacific giant octopus (*Enteroctopus dofleini*) in the Gulf of Alaska». En: *Fisheries Bulletin* 112.4 (2014), págs. 253-260.
- [8] Reid S. Brewer y Brenda L. Norcross. *Temperature and size-dependent growth and movement of *Enteroctopus dofleini* in the Bering Sea*. Inf. téc. North Pacific Fishery Management Council, 2012.
- [9] Reid S. Brewer, Brenda L. Norcross y Ellen Chenoweth. «Temperature- and size-dependent growth and movement of the North Pacific giant octopus (*Enteroctopus dofleini*) in the Bering Sea». En: *Marine Biology Research* (2017). DOI: 10.1080/17451000.2017.1309436.
- [10] Samuel M. Scheiner y R. F. Lyman. «The genetics of phenotypic plasticity. I. Heritability». En: *Journal of Evolutionary Biology* (1989).

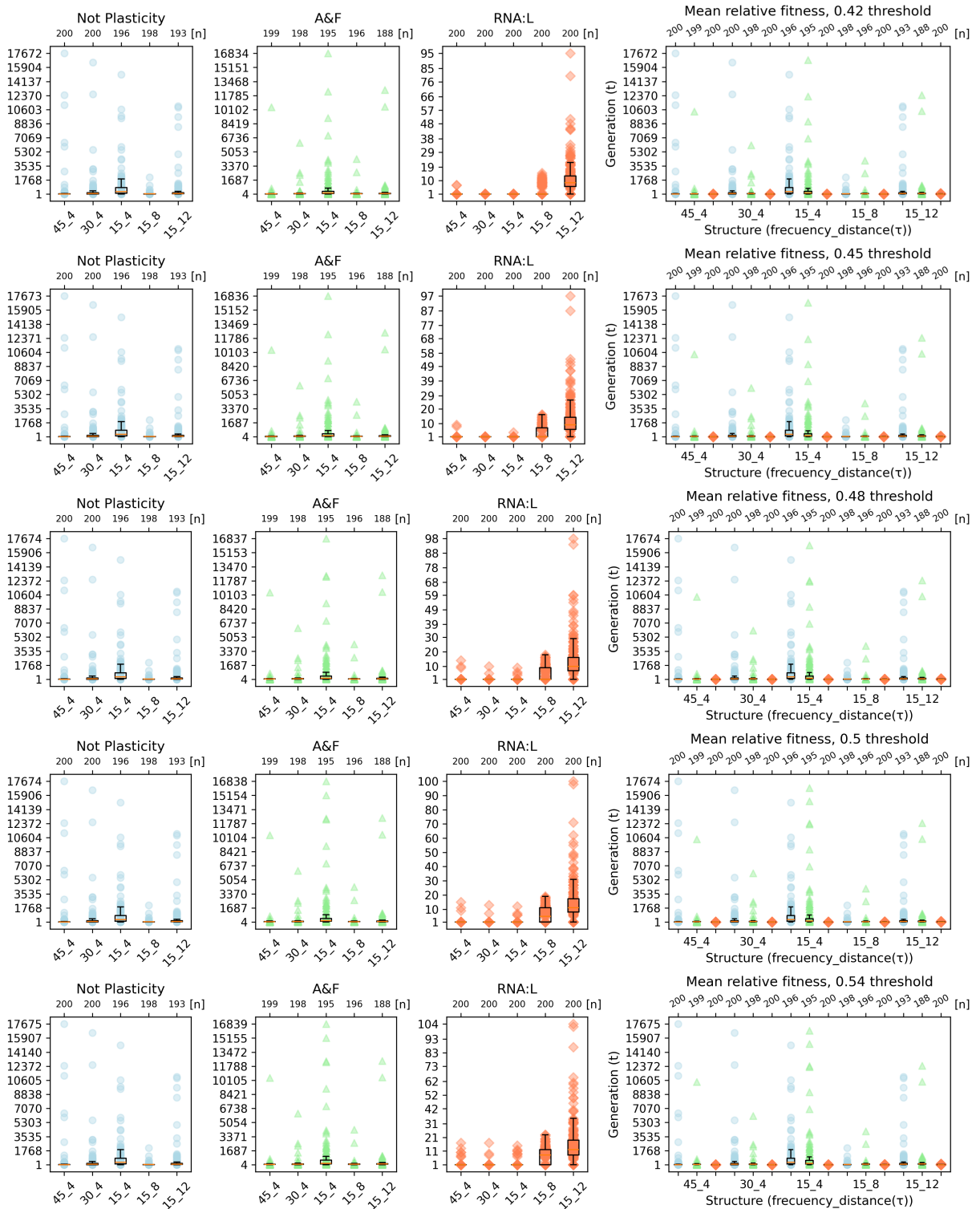
- [11] Ben K. D. Pearce et al. «Origin of the RNA world: The fate of nucleobases in warm little ponds». En: *Proceedings of the National Academy of Sciences* 114 (2017), págs. 11327-11332.
- [12] Nicolas Galtier, Nicolas Tourasse y Manolo Gouy. «A nonhyperthermophilic common ancestor to extant life forms». En: *Science* 283.5399 (1999), págs. 220-221.
- [13] C. Briones, M. Stich y S. C. Manrubia. «The dawn of the RNA World: toward functional complexity through ligation of random RNA oligomers». En: *RNA* 15 (2009), págs. 743-749.
- [14] Armen Y. Mulkidjanian et al. «Origin of first cells at terrestrial, anoxic geothermal fields». En: *Proceedings of the National Academy of Sciences* 109 (2012), E821-E830.
- [15] Victor W. Rodwell et al. «Estructura y función de los ácidos nucleicos». En: *Bioquímica ilustrada de Harper*. McGraw-Hill, 2022. Cap. 34.
- [16] Victor W. Rodwell et al. «Nucleótidos». En: *Bioquímica ilustrada de Harper*. McGraw-Hill, 2022. Cap. 32.
- [17] Shai Rouskin et al. «Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo». En: *Nature* 505.7485 (2014), págs. 701-705.
- [18] Andreas Wagner. «Mutational robustness accelerates the origin of novel RNA phenotypes through phenotypic plasticity». En: *Biophysical Journal* 106.4 (2014), págs. 955-965. DOI: 10.1016/j.bpj.2014.01.003. URL: <https://doi.org/10.1016/j.bpj.2014.01.003>.
- [19] John Mattick y Paulo Amaral. *RNA, the Epicenter of Genetic Information: A new understanding of molecular biology*. Abingdon, UK: CRC Press, 2022. DOI: 10.1201/9781003109242. URL: <https://www.ncbi.nlm.nih.gov/books/NBK595939/>.
- [20] Thomas Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford University Press, 1996.
- [21] Lauren W. Ance y Walter Fontana. «Plasticity, evolvability, and modularity in RNA». En: *Journal of Experimental Zoology* 288.3 (2000), págs. 242-283. DOI: 10.1002/1097-010X(20001015)288:3<242::AID-JEZ5>3.0.CO;2-0.
- [22] Antonio Loreto, Edgardo Ugalde y Carlos Espinosa-Soto. «RNA-ligand complexes and the attenuation of neutral confinement in the evolution of RNA secondary structures». En: *bioRxiv* (2025). DOI: 10.64898/2025.12.19.695547.
- [23] R. B. Lyngsø y C. N. S. Pedersen. «Pseudoknots in RNA Secondary Structures». En: *Journal of Computational Biology* 7.3-4 (2000), págs. 409-427. DOI: 10.1089/106652700750050807. URL: <https://doi.org/10.1089/106652700750050807>.
- [24] Michael Zuker y Patrick Stiegler. «Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information». En: *Nucleic Acids Research* 9.1 (1981), págs. 133-148. DOI: 10.1093/nar/9.1.133. URL: <https://doi.org/10.1093/nar/9.1.133>.

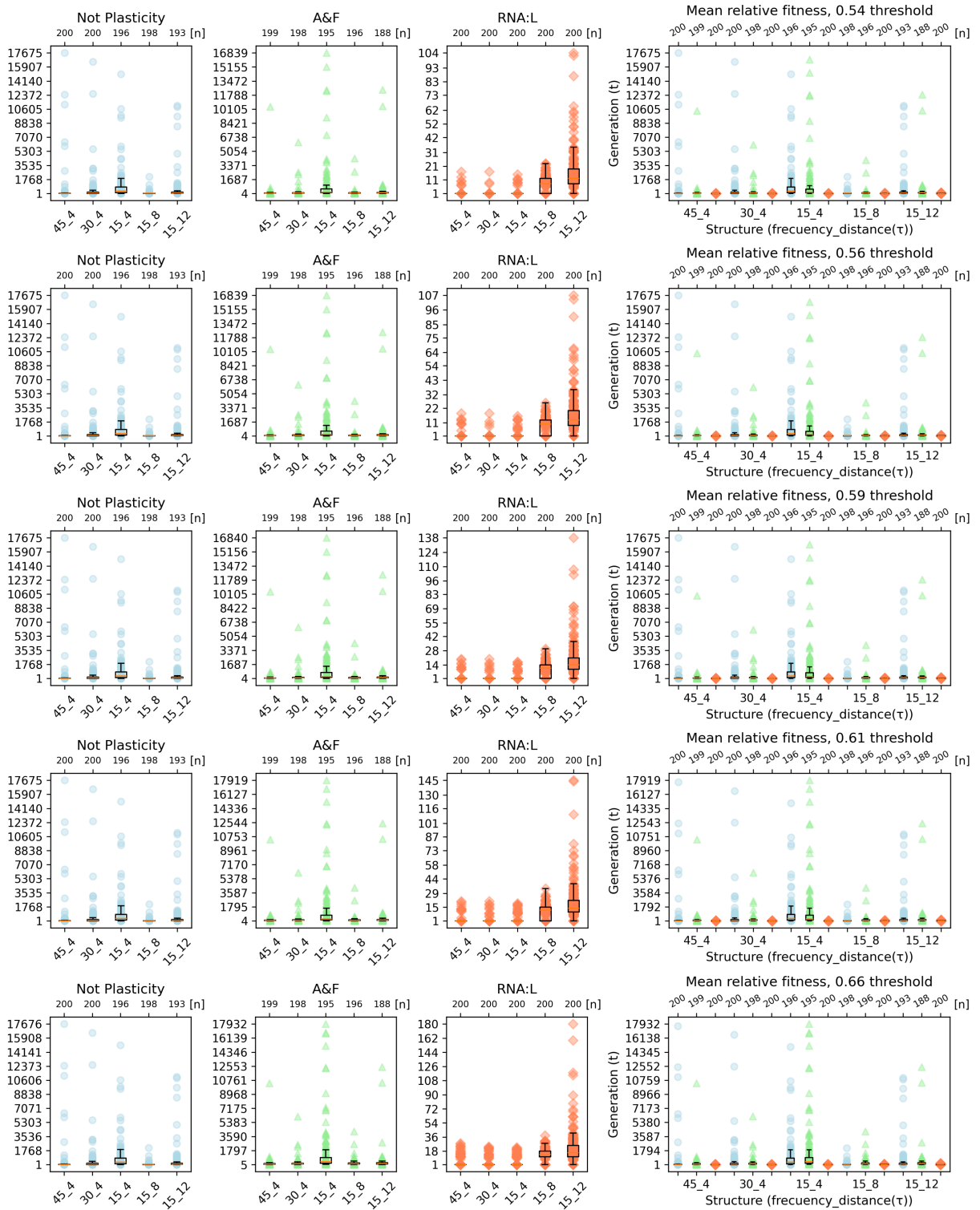
- [25] Sean R. Eddy. «What is dynamic programming?» En: *Nature Biotechnology* 22.7 (2004), págs. 909-910. DOI: 10.1038/nbt0704-909. URL: <https://doi.org/10.1038/nbt0704-909>.
- [26] Stefan Wuchty et al. «Complete suboptimal folding of RNA and the stability of secondary structures». En: *Biopolymers* 49.2 (1999), págs. 145-165. DOI: 10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-A. URL: [https://doi.org/10.1002/\(SICI\)1097-0282\(199902\)49:2%3C145::AID-BIP4%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0282(199902)49:2%3C145::AID-BIP4%3E3.0.CO;2-A).
- [27] Ronny Lorenz, Stephan H. Bernhart y C. et al. Höner Zu S. «ViennaRNA Package 2.0». En: *Algorithms for Molecular Biology* 6.1 (2011), pág. 26. DOI: 10.1186/1748-7188-6-26. URL: <https://doi.org/10.1186/1748-7188-6-26>.
- [28] Douglas H. Turner y David H. Mathews. «NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure». En: *Nucleic Acids Research* 38 (2010), págs. D280-D282. DOI: 10.1093/nar/gkp892. URL: <https://doi.org/10.1093/nar/gkp892>.
- [29] Edward F. Redish y Janice C. Redish. *Living Physics: An introductory physics course for life-science students*. Vol. 2. Toronto, Canadá: Top Hat, 2023.
- [30] Benjamin D. Geller et al. «Entropy and spontaneity in an introductory physics course for life science students». En: *American Journal of Physics* 82.5 (2014), págs. 394-402. DOI: 10.1119/1.4870389. URL: <https://doi.org/10.1119/1.4870389>.
- [31] Benjamin W. Dreyfus et al. «Chemical energy in an introductory physics course for the life sciences». En: *American Journal of Physics* 82.5 (2014), págs. 403-411. DOI: 10.1119/1.4870391. URL: <https://doi.org/10.1119/1.4870391>.
- [32] Jean-Marc Dubois, Gilles Ouanounou y Béatrice Rouzair-Dubois. «The Boltzmann equation in molecular biology». En: *Progress in Biophysics and Molecular Biology* 99.2-3 (2009), págs. 87-93. DOI: 10.1016/j.pbiomolbio.2009.07.001. URL: <https://pubmed.ncbi.nlm.nih.gov/19616022/>.
- [33] K. Shapiro y K. Zhang. «Comparing multiple RNA secondary structures using tree comparisons». En: *Bioinformatics* 6.4 (1990), págs. 309-318. DOI: 10.1093/bioinformatics/6.4.309.
- [34] I. L. Hofacker et al. «Fast Folding and Comparison of RNA Secondary Structures». En: *Monatshefte für Chemie / Chemical Monthly* 125.2 (1994), págs. 167-188. DOI: 10.1007/BF00818163.
- [35] G. J. Szöllosi e I. Derényi. «Congruent evolution of genetic and environmental robustness in micro-RNA». En: *Molecular Biology and Evolution* 26.4 (2009), págs. 867-874. DOI: 10.1093/molbev/msp008.

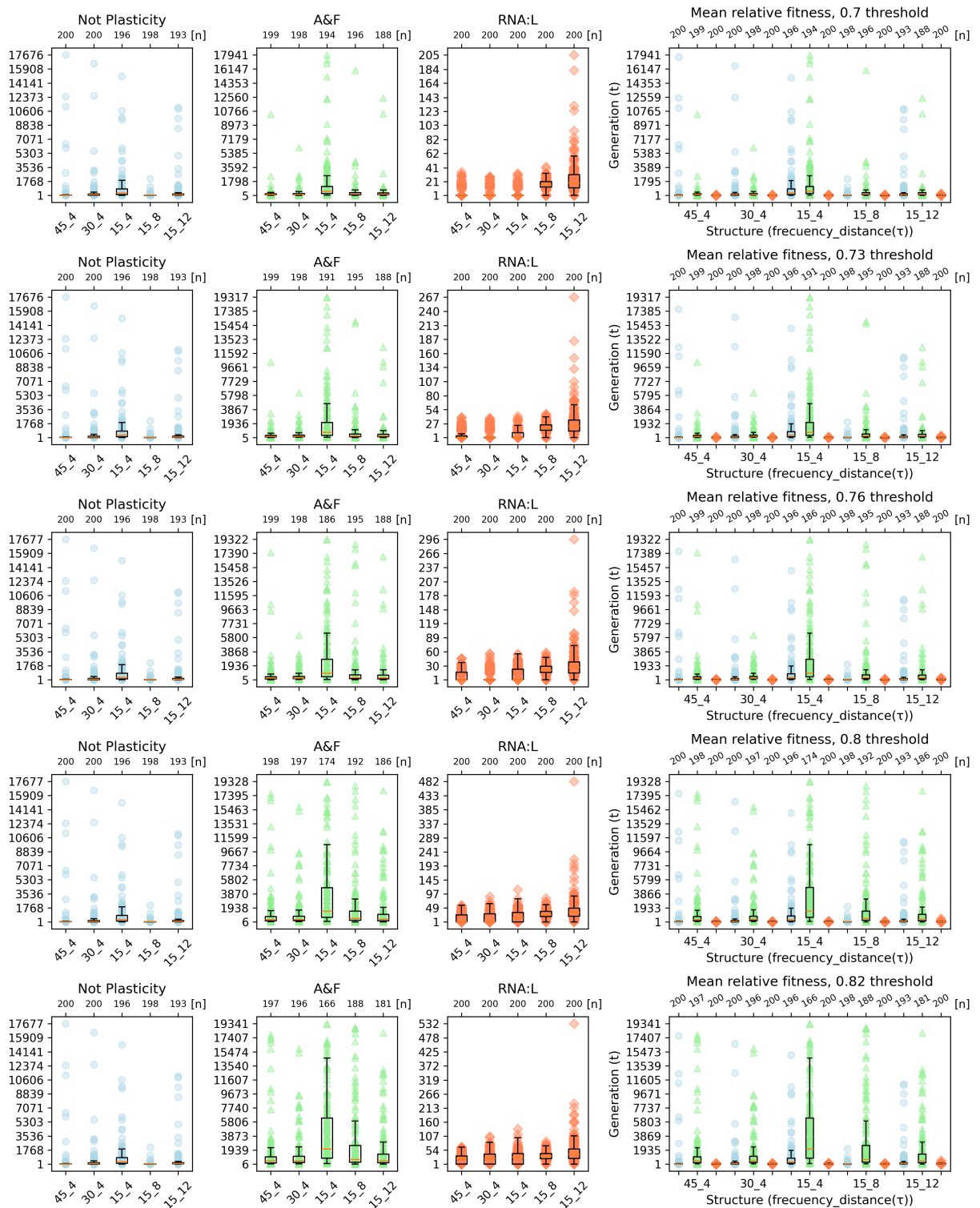
Apéndice

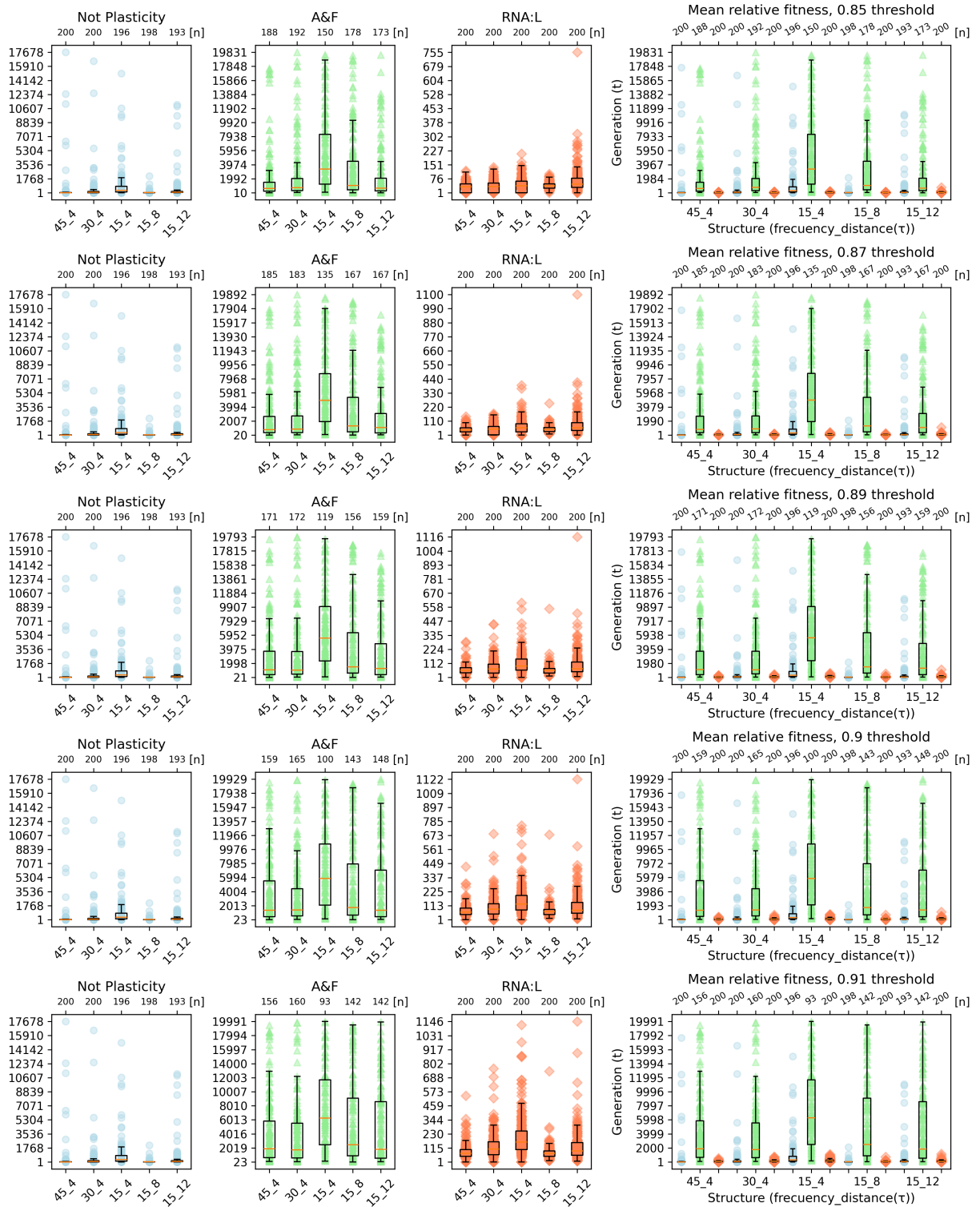
Umbrales explorados, adecuación relativa promedio



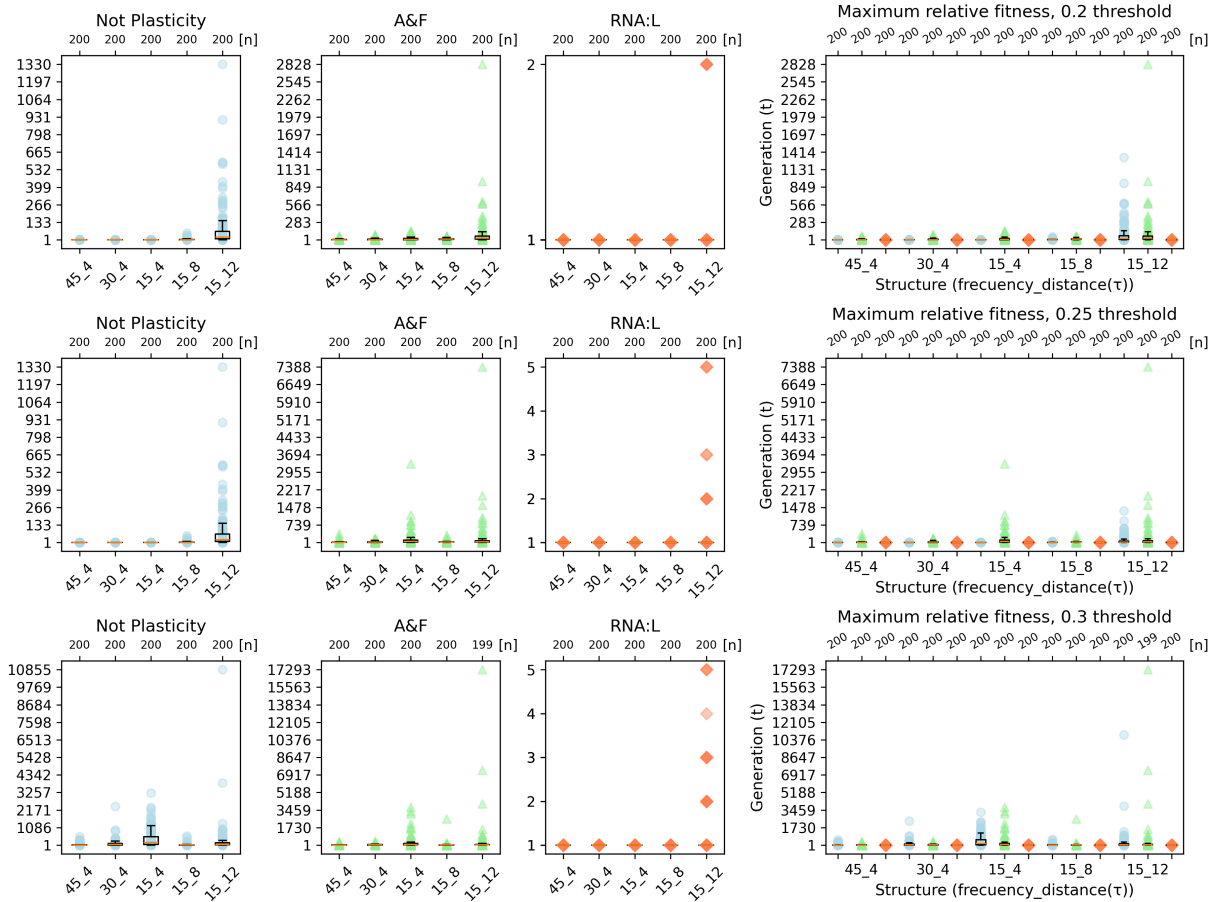


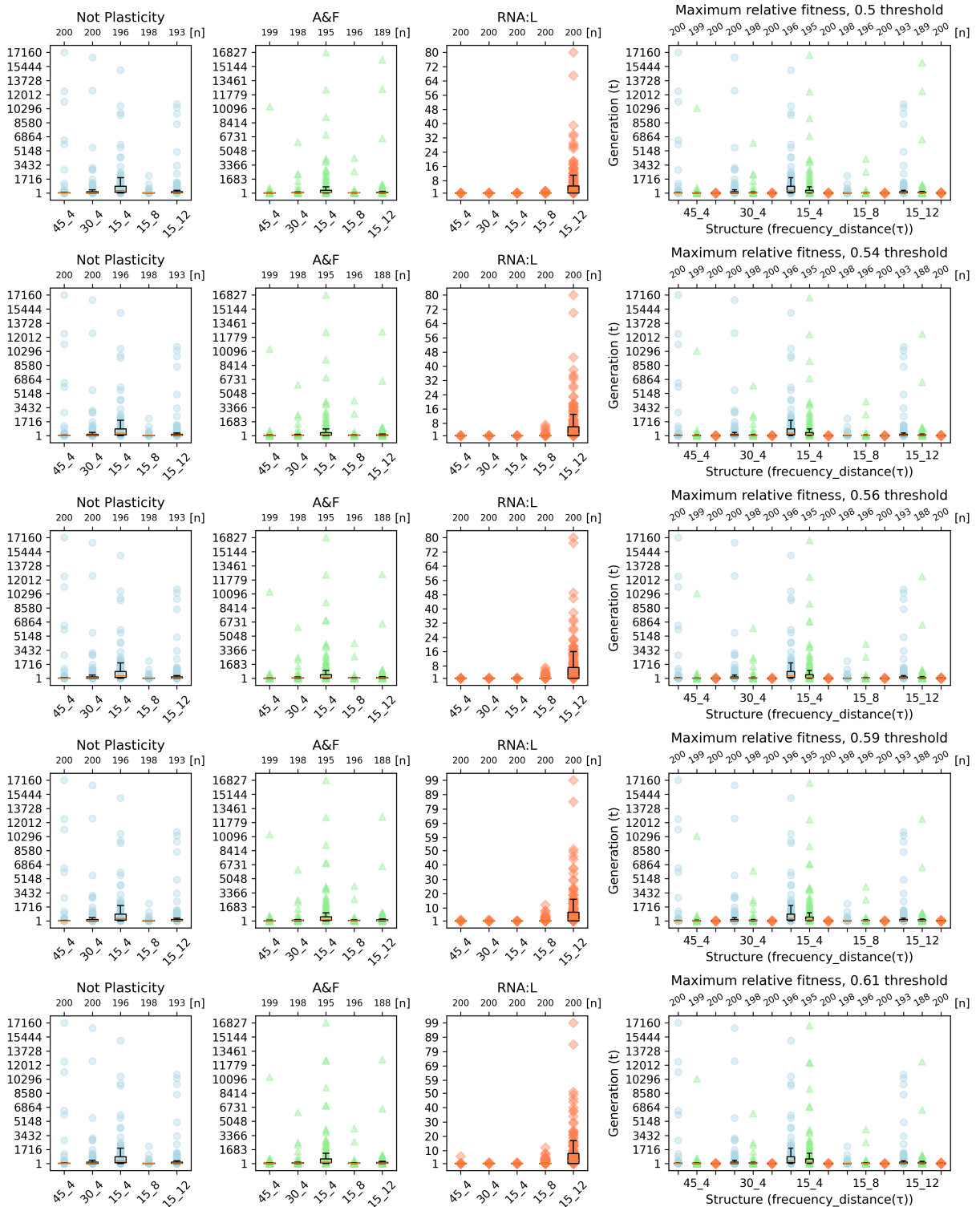


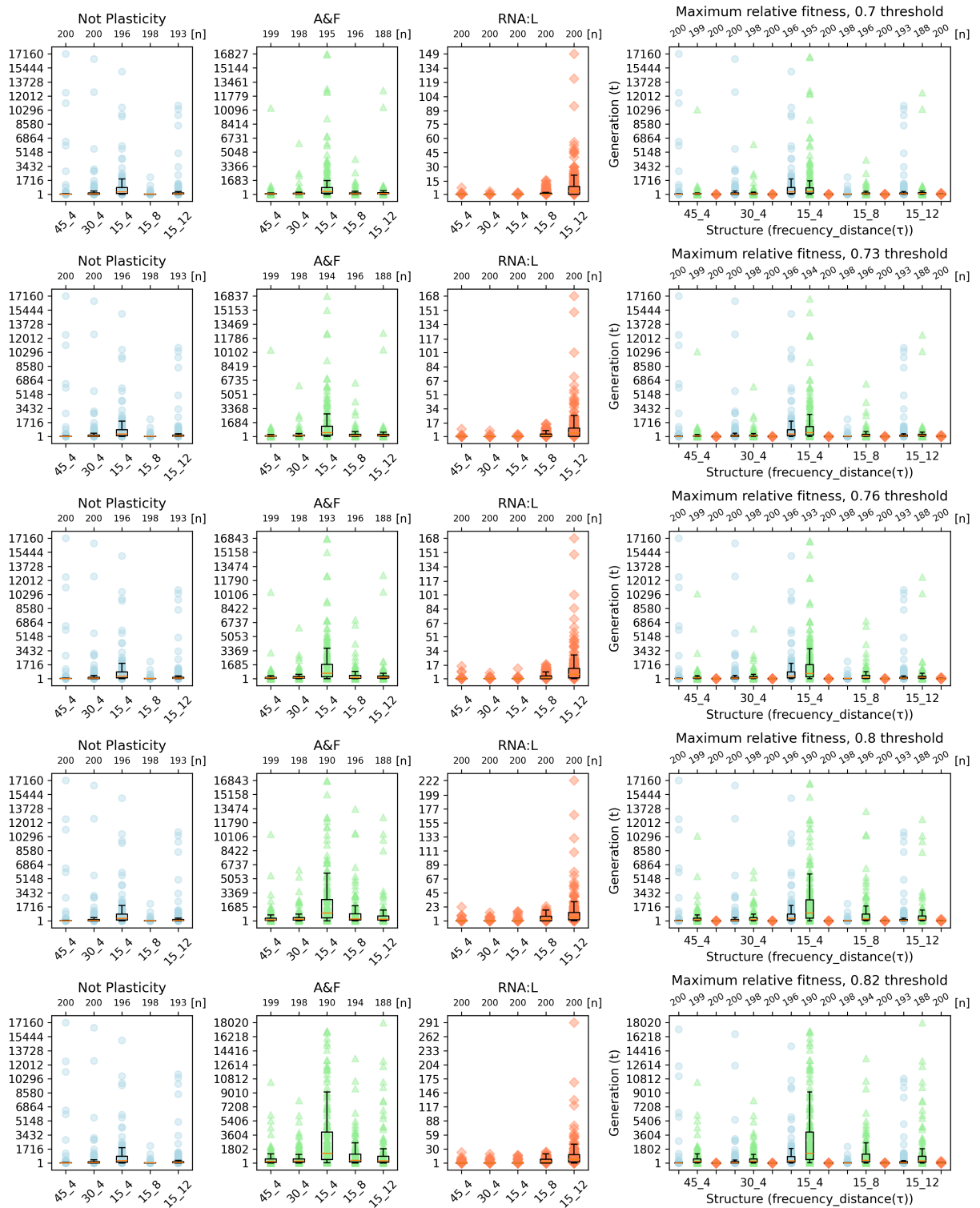


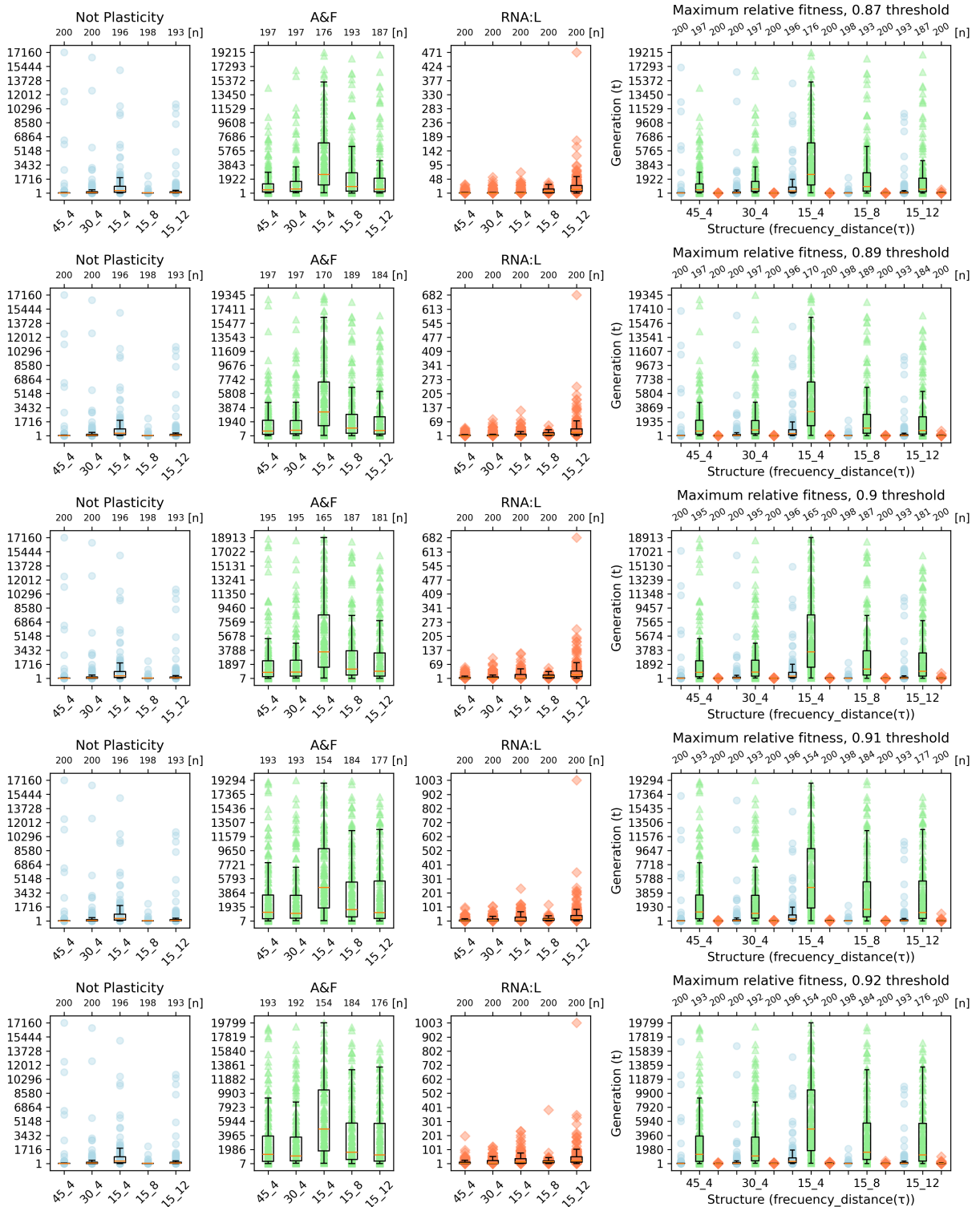


Umbrales explorados, adecuación relativa máxima

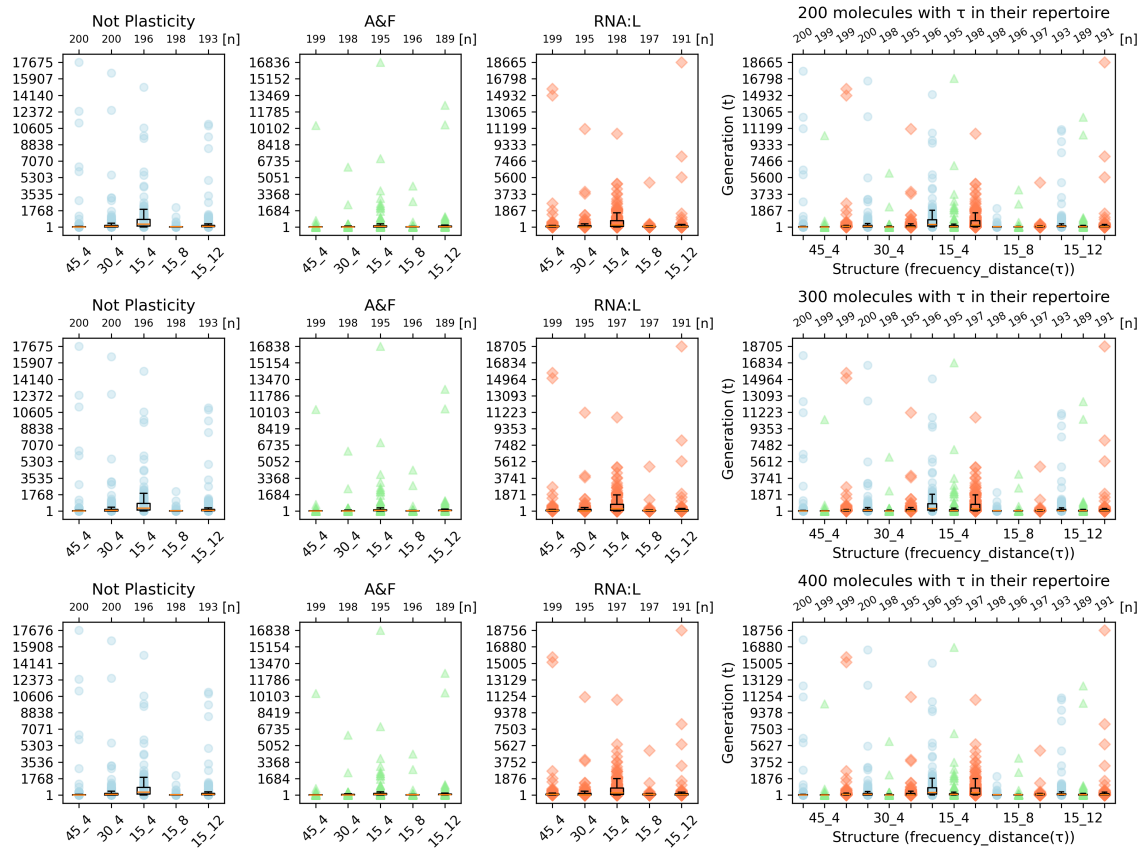


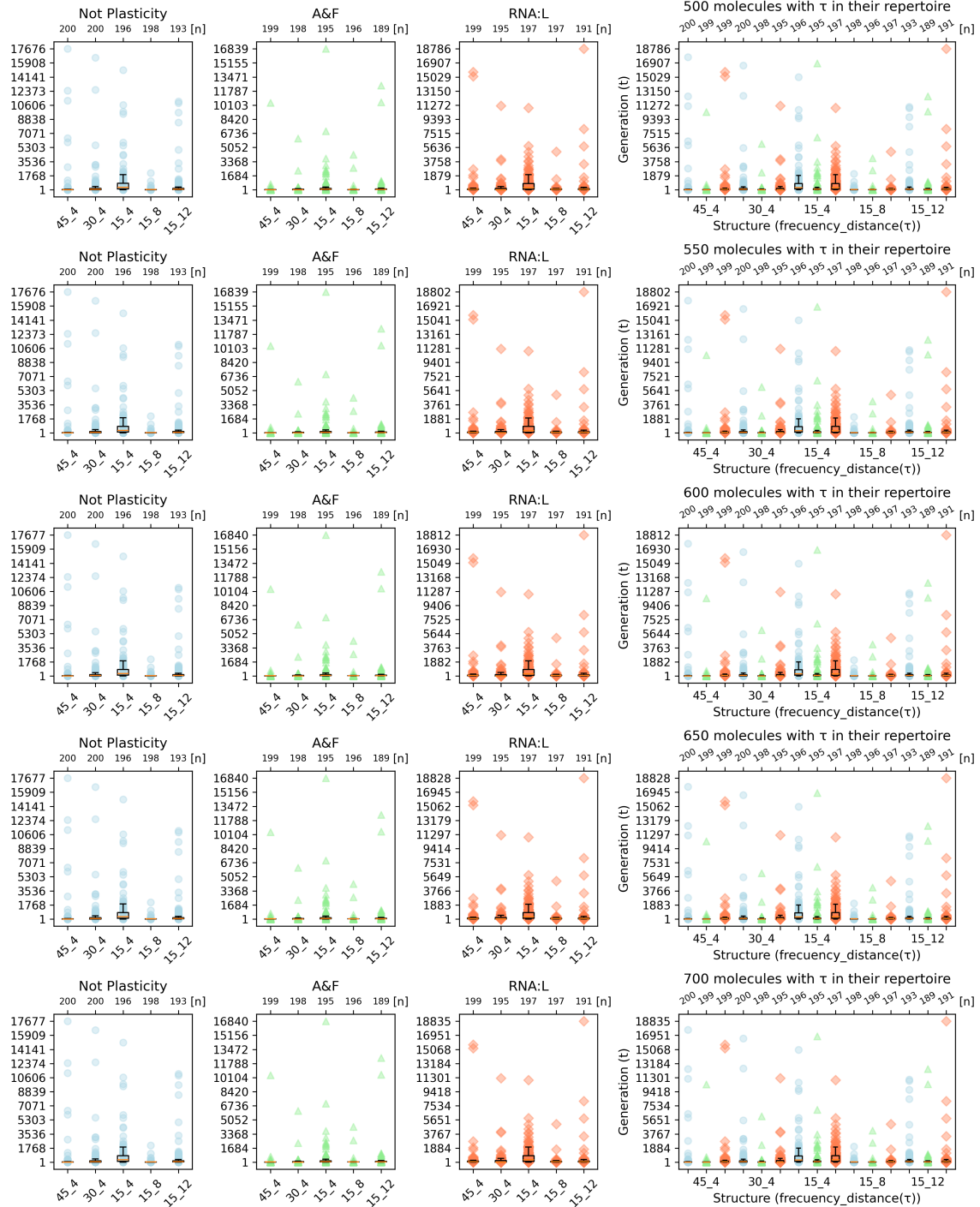


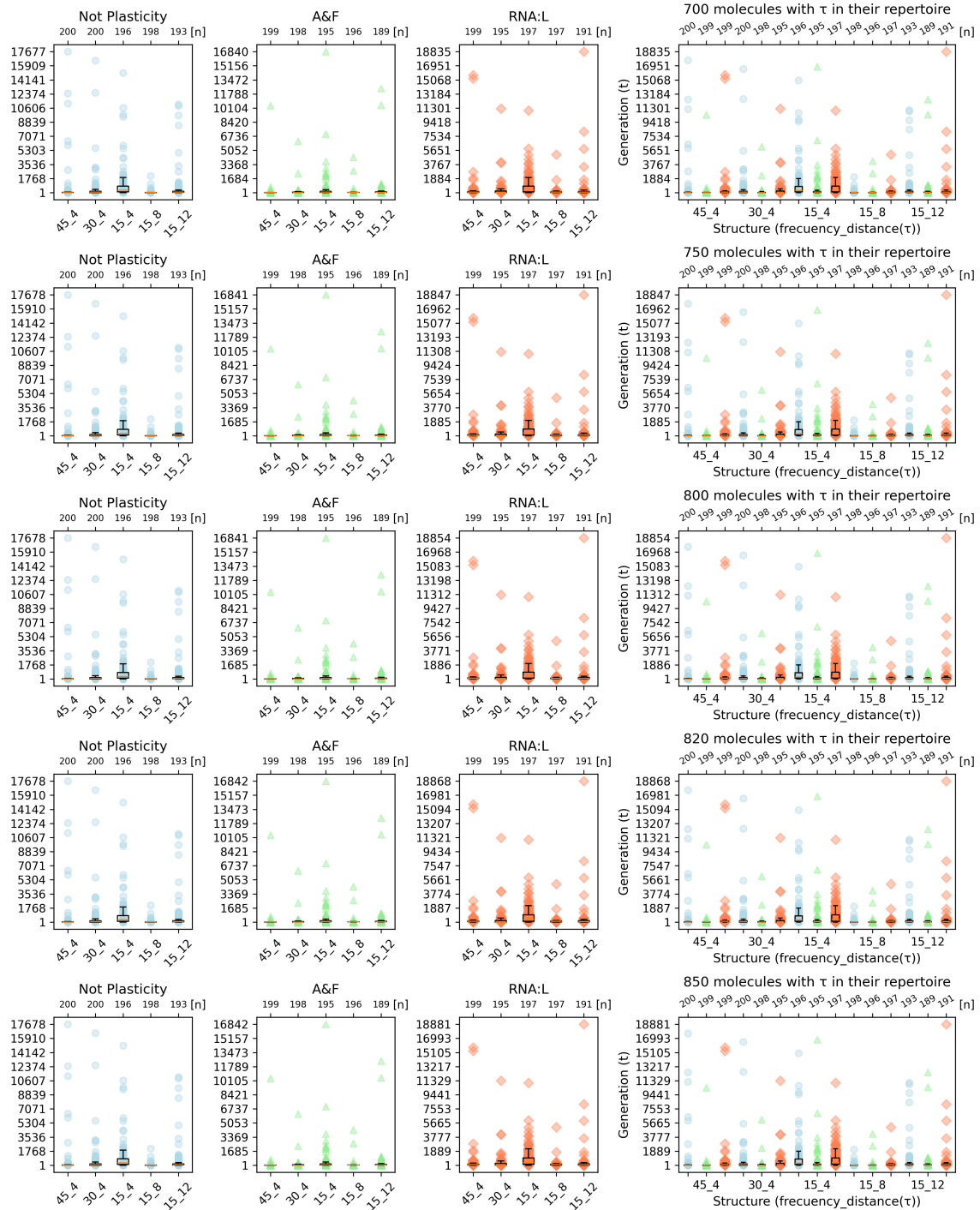


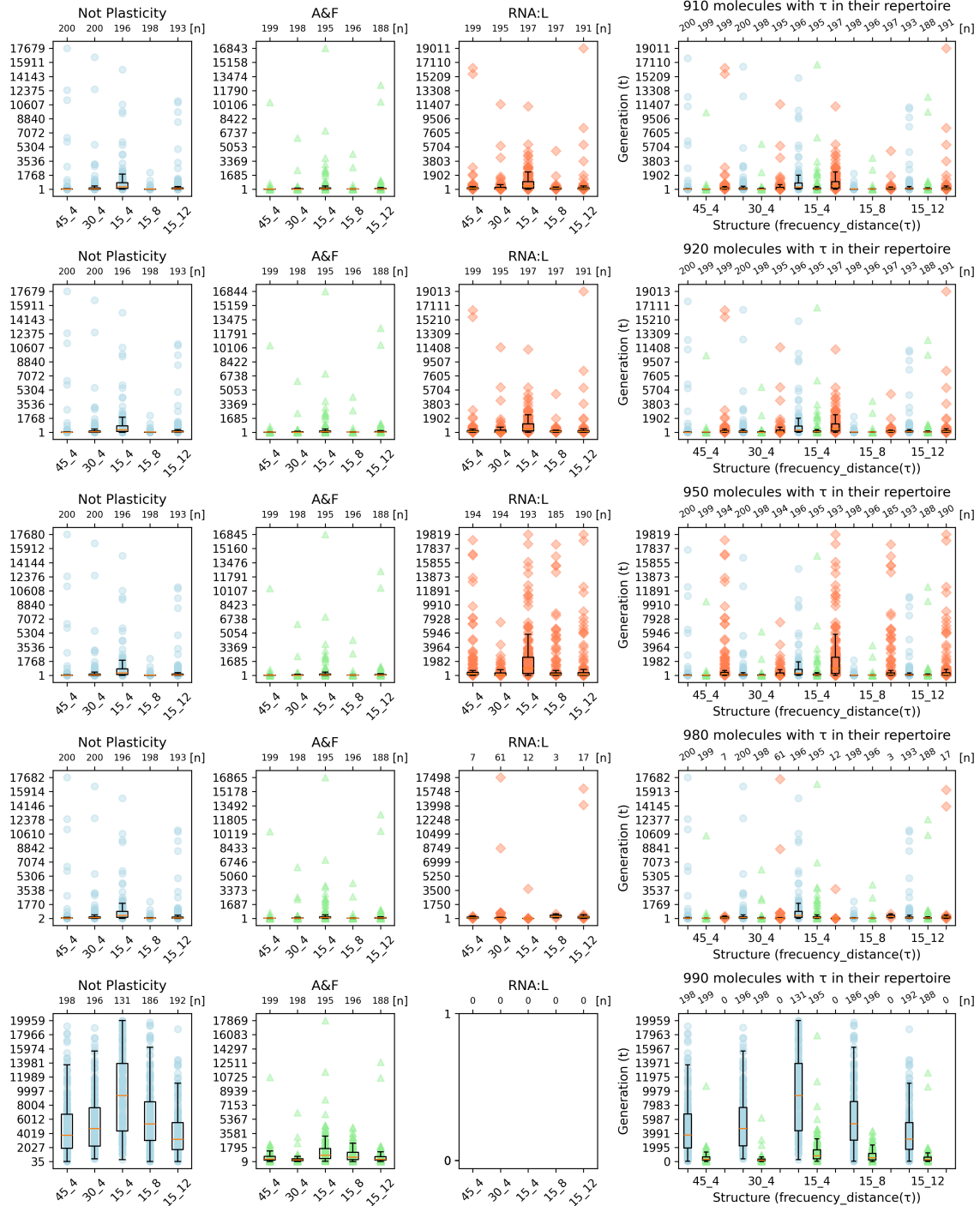


Umbrales explorados, cantidad de genotipos que producen el fenotipo destino como estructura de mínima energía

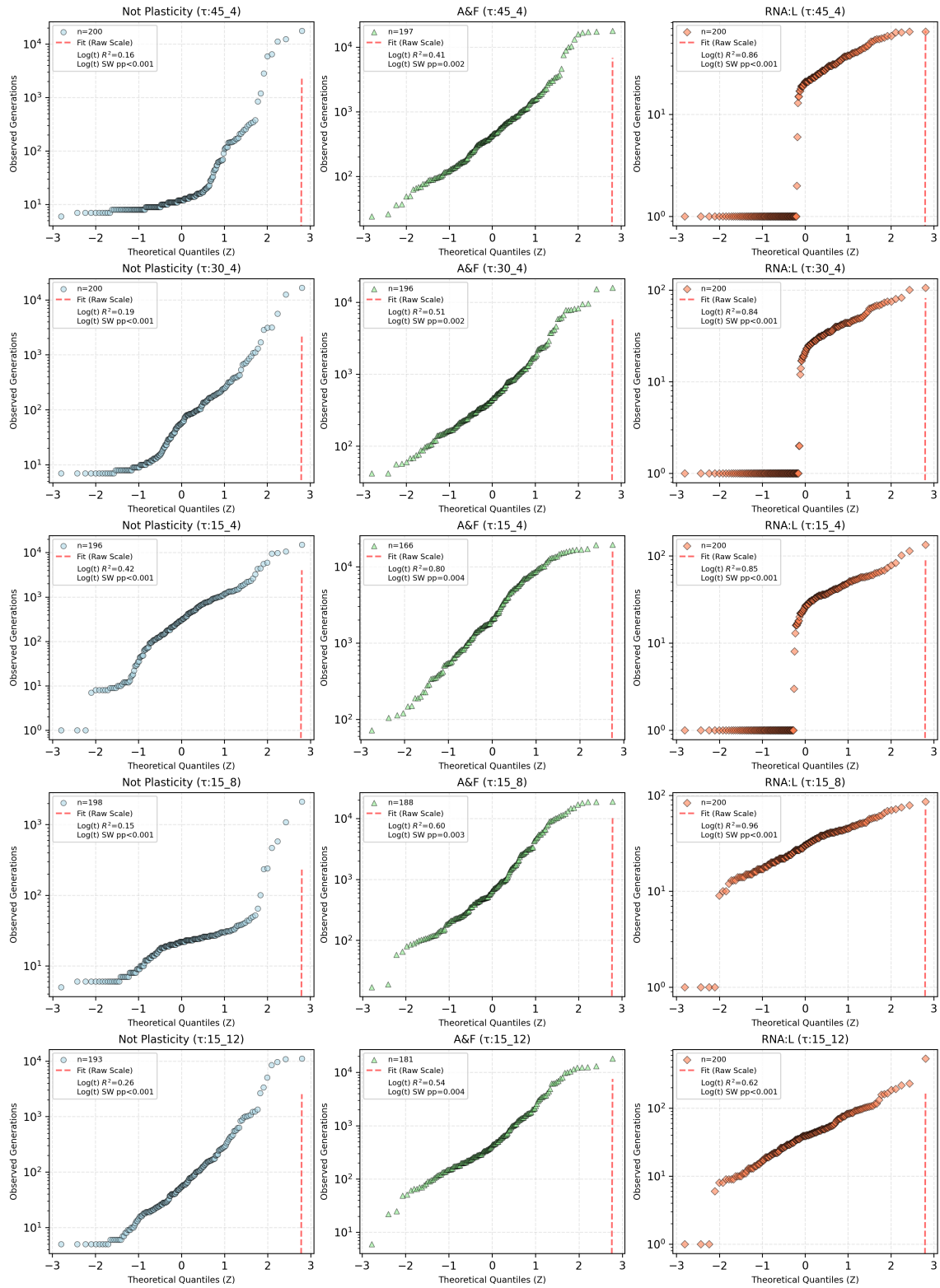




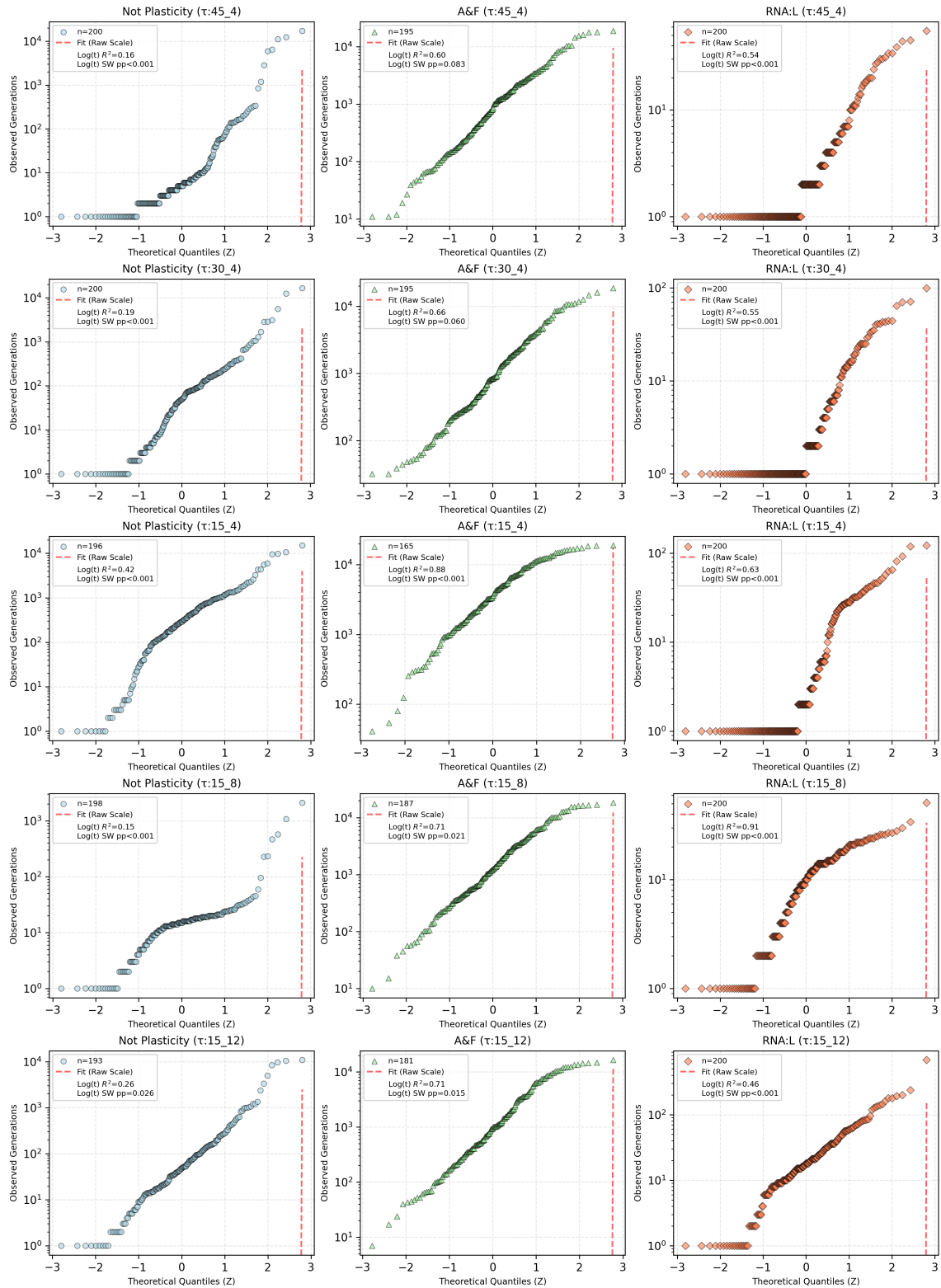




Gráficos QQ, adecuación relativa promedio



Gráficos QQ, adecuación relativa máxima



Gráficos QQ, cantidad de genotipos que producen el fenotipo destino como estructura de mínima energía

