

4
2 ej.

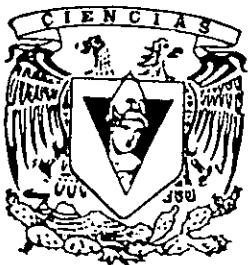


UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

CONTINUIDAD EN EVOLUCION:
UN MODELO MATEMATICO.

T E S I S
QUE PARA OBTENER EL TITULO DE:
M A T E M A T I C A
P R E S E N T A :
ILEANA BORJA TECUATL



DIRECTOR DE TESIS: DR. PEDRO EDUARDO MIRAMONTES VIDAL

MEXICO, D. F.

OCTUBRE 1999

TESIS CON
FALLA DE ORIGEN

280521



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:
 Continuidad en evolución: Un modelo matemático

realizado por Ileana Borja Tecuatl

con número de cuenta 9232772-0 , pasante de la carrera de matemáticas

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de tesis Propietario Dr. Pedro Eduardo Miramontes Vidal

Propietario Dr. Germinal Cocho Gil

Propietario Dr. José Ignacio Barradas Bribiaca

Suplente Dr. Jaime Lagúñez Otero

Suplente Dr. Gustavo Martínez Mekler

Miramontes
P. Cocho
José Ignacio Barradas
Jaime Lagúñez Otero
Gustavo Martínez Mekler

Consejo Departamental de Matemáticas
 Profesor Julio César Guevara Bravo
 Coordinador de la Licenciatura

Ciudad Universitaria, a 30 de septiembre de 1999.

Introducción

Antecedentes¹

El origen de la vida es una cuestión que ha ocupado a las sociedades desde tiempos inmemoriales. En la búsqueda de una respuesta se han generado diversos argumentos, cada cual acorde al tipo de cultura de la sociedad que lo propone y cada uno relativamente satisfactorio en su época. De esta forma tenemos que las respuestas más antiguas al problema del origen de la vida y a su evolución, han sido del tipo religioso o espiritual —principalmente en las primeras sociedades— y que estas respuestas han variado paulatinamente, de forma que ahora también se cuenta con argumentos basados en estudios científicos, que de igual manera han sido apoyados, refutados o anulados a medida que la ciencia sigue su curso.

Desde épocas muy remotas los pueblos creían que la vida surgía de manera espontánea a partir de la materia inerte. Por siglos se consideró posible que el lodo diera origen a peces, que el estiércol generara moscas o que a partir de una roca se formaran escarabajos. Esta creencia es comúnmente conocida como *la generación espontánea* y persistió hasta mediados del siglo XIX.

Alrededor de 1750 Lazzaro Spallanzani demostró que la creencia de la generación espontánea, defendida en ese entonces por un clérigo de nombre John Needham Turberville y apoyada por el zoólogo Georges-Louis Leclerc, Conde de Buffon, era errónea. Spallanzani mostró que si en una botella se introduce algún preparado de plantas, agua y semillas del tipo que sea, luego se cierra a fuego dicha botella, impidiendo que algo penetre en ella y hecho eso se calienta la botella por una hora, entonces muere todo microorganismo encontrado anteriormente en la infusión y nada se encontrará vivo en el interior de la botella aunque se conserve su contenido por mucho tiempo [21].

¹Basado en [6]

Aún con esta demostración la creencia de la generación espontánea contaba todavía con adeptos y Spallanzani tuvo que hacer una serie de pruebas para convencer a sus colegas de que él tenía la razón. Pese a dichos esfuerzos a principios del siglo XIX la generación espontánea, por esas fechas también conocida como *fuera vegetativa*, seguía aún en pie y fue en 1864 que Louis Pasteur, procediendo de manera similar a Spallanzani, probó definitivamente que la generación espontánea no es posible [21].

Louis Pasteur mostró que cuando se hierve un líquido se mata cualquier microorganismo existente en él y lo deja estéril, impidiendo que algo más crezca ahí.

La prueba de Pasteur representó una pauta importante para iniciar una nueva línea de investigación en busca de mejores respuestas al origen de la vida. Sin embargo durante casi un siglo después de su descubrimiento no hubo muchos resultados. Es a partir de 1924 que las investigaciones en torno al origen de la vida retoman fuerza. Desde entonces, científicos de todo el mundo han tratado de argumentar con razonamientos formales el origen de la vida y las especies. Estos argumentos han generado un torrente de información que nos llena de motivos nuevos y cada vez mejor explicados que, empero, son discutidos constantemente.

Algunos investigadores han hecho experimentos basados en materiales presentes en la atmósfera primitiva, a fin de encontrar en materia antigua partículas que puedan representar una explicación al origen de la vida. Por ejemplo, el soviético Alexander Ivanovich Oparin propuso en 1924 que las formas unicelulares actuales deben provenir de moléculas orgánicas simples de la atmósfera de la Tierra en sus orígenes. Utilizando evidencias geológicas estableció condiciones teóricas para la atmósfera primitiva. En 1929 el bioquímico inglés J. B. S. Haldane propuso que la vida en la Tierra debió originarse en la atmósfera antigua (bajo las condiciones predichas por Oparin) debido a la energía en forma de calor y radiación ultravioleta a la que estaba expuesta. Esa radiación la filtra actualmente la capa de ozono.

Con el objeto de probar estas teorías, en 1953 Stanley Miller mezcló gases que según Oparin estaban presentes en la atmósfera primitiva (hidrógeno, metano, amoníaco, dióxido de carbono, agua y nitrógeno), y sometióndolos a descargas eléctricas por una semana encontró como resultado que entre los compuestos producidos había aminoácidos. Con este experimento Miller mostró que un posible origen a la vida era el propuesto por Oparin-Haldane.

Paralelo a estos estudios se encuentran los de aquellos investigadores interesados en explicar la manera en que se hereda la información genética.

la cual constituye un nuevo ser y continua la vida en sí. Parte importante de análisis relacionados a este fin —y un buen impulso para proseguir la búsqueda en términos de lo que se conocía sobre el *ácido desozirribonucléico* (DNA), que para 1950 ya se creía que era la clave de la herencia— fue el trabajo presentado por James D. Watson, Francis Crick y Rosalind Franklin, publicado por primera vez en 1953 y que más tarde (1962) les valió el premio Nobel a los dos primeros. En dicho trabajo sus autores proponen y aclaran parte de los grandes misterios del DNA: su estructura. A partir de entonces los estudios sobre el origen de la vida y las especies, así como su evolución, se han centrado en tratar de conocer cuál fue la primer estructura en la vida: ¿Las proteínas, el RNA o el DNA?

Las *proteínas* son estructuras elementales para que los organismos lleven a cabo sus funciones vitales; el RNA, *ácido ribonucléico*, se encarga de proporcionar los medios para convertir las bases nitrogenadas en *aminoácidos*, los cuales son moléculas que dan lugar a las proteínas. Las *bases nitrogenadas* son las estructuras básicas del DNA junto con el fosfato y la desoxirribosa. Las bases nitrogenadas del DNA son la Adenina, la Timina, la Citosina y la Guanina. Dichas bases se encuentran emparejadas bajos los siguientes principios de complementaridad, la Adenina siempre es pareja de la Timina y la Guanina siempre es pareja de la Citosina. El DNA es la molécula que porta la información genética dentro de las células.

Las *enzimas* son proteínas con capacidades catalíticas; los ácidos nucleicos pueden replicarse a sí mismos. La catálisis de sus funciones y la auto-replicación son dos características con que debió contar la primer estructura en la vida. Así pues, las investigaciones sobre el origen de la vida se centraron en tratar de encontrar qué estructura, proteína o ácido nucleico, podía autocatalizar su replicación.

El DNA, el RNA y los aminoácidos están relacionados mediante los procesos de duplicación, transcripción y traducción. Estos procesos constituyen la manera en que una célula produce una proteína a partir de la información genética contenida en su DNA. Brevemente, este procedimiento se puede resumir como sigue: primero el DNA se duplica gracias a la complementaridad de sus bases nitrogenadas y a la acción de unas proteínas llamadas polimerasas del DNA. Esta duplicación permite que se forme una secuencia de RNA con la misma información que la secuencia de DNA. Ese proceso se llama *transcripción*. La información del RNA se lee y traduce del alfabeto de bases nitrogenadas al alfabeto de aminoácidos. Esta traducción da como resultado una secuencia de aminoácidos llamada proteína. (En el capítulo

uno se explican con detalle estos procedimientos.)

Entre varias teorías sobre la vida precelular se encuentra la teoría del *mundo RNA*, que propone que el RNA fue el medio a través del cual se formaron los primeros seres vivos. Así pues esta teoría propone que el RNA es una estructura que puede replicarse a sí misma y que cuenta con propiedades catalíticas.

La totalidad de la información genética de un organismo se conoce como *genoma*. La parte de DNA que codifica para una misma proteína se llama *gene*. Puesto que los genes requieren enzimas y estas proteínas a su vez requieren para su formación la presencia de genes, el problema de qué fue primero si DNA, RNA o proteínas se convierte en un problema del tipo *¿Qué fue primero el huevo o la gallina?* [6, p. 1] El RNA porta en esencia la misma información genética que el DNA y se ha visto que bajo ciertas circunstancias el RNA tiene características catalíticas, por lo que representa una alternativa para resolver dicho problema. Más aún, el RNA puede ser transcrito en DNA mediante el procedimiento inverso a la transcripción normal. Estos hechos constituyen una buena razón para considerar que la teoría del mundo RNA es el camino original que se siguió para llegar a las células.

En 1961 Marshall Nirenberg y sus colegas descubrieron que el RNA mensajero podía ser transcrito en aminoácidos. Su hallazgo provino de la transcripción de RNA mensajero en el aminoácido fenilalanina [22, p. 213]. Más tarde, en 1968, Francis Crick y Leslie Orgel propusieron que la primer molécula de información fue el RNA, teoría que empieza a tomar fuerza después de que Harry Noller mostró que el RNA ribosomal es fundamental en la traducción del RNA mensajero para formar proteínas.

En 1986 Walter Gilbert adoptó el término *mundo RNA* para describir el tiempo durante el cual el RNA fue la primer molécula catalítica y con información para originar vida. Como resultado importante que favorece esta teoría se tiene en el mismo año que Thomas Cech publicó sus descubrimientos sobre RNA con características catalíticas.

1989 es otro año favorable para la teoría mencionada. Entre los resultados obtenidos ese año se cuentan el de Thomas Cech y Sidney Altman, quienes compartieron el premio Nobel por descubrir RNA catalítico; el de Jack Szostak y su equipo de trabajo, quienes mostraron evidencias de RNA auto-replicante y el de Gerald Joyce, quien empezó experimentos utilizando reacción de polimerasas en cadena para simular la evolución del RNA.

A partir de que fue propuesta y a raíz de los logros que consolidan a la teoría del mundo RNA como entre las más razonables, científicos de todo el

mundo no han dejado de realizar experimentos en su afán por apoyar dicha teoría. Los resultados durante la presente década no se han quedado atrás. Por ejemplo: el equipo de trabajo de Harry Noller presentó en 1992 evidencias sobre la relación directa entre RNA ribosomal y la síntesis de proteínas y en 1994 Charles Wilson creó en el laboratorio de Szostak moléculas de RNA que llevan a cabo funciones dentro de la célula, incluso de mejor manera que las proteínas que usualmente las realizan [14, p. 213].

Todos estos avances en tanto a la biología molecular trajeron consigo el desarrollo de la evolución molecular, la cual relaciona los avances hasta ahora mencionados con los logrados por aquellos investigadores que centraron sus preguntas en tratar de resolver el problema al origen y evolución de las especies.

La evolución molecular y la evolución de las especies

La evolución molecular comprende dos áreas de estudio: la evolución de macromoléculas y la reconstrucción de la historia en la evolución de genes y organismos. La *evolución de macromoléculas* estudia los tipos y patrones de cambio, al paso del tiempo, que hay en el material genético (por ejemplo en secuencias de DNA) y en sus productos (proteínas); la evolución de macromoléculas se encarga también de estudiar los mecanismos responsables de dichos cambios. La otra área, también conocida como *filogenia molecular*, estudia la historia de la evolución de organismos y macromoléculas, utilizando información proveniente de la biología molecular.

En principio podría parecer que estas dos áreas constituyen campos de estudio independientes, pues el objeto de la primera es estudiar las causas y efectos de los cambios evolutivos en las moléculas, mientras que para la segunda las moléculas son herramientas que nos permiten reconstruir la historia de los organismos y sus componentes genéticos. Sin embargo, estas dos disciplinas están íntimamente relacionadas y los progresos en un área favorecen los estudios de la otra [23, p. 1]. Por ejemplo, los conocimientos nuevos en filogenia son esenciales para determinar el orden de los cambios en las moléculas bajo estudio y a su vez, los patrones y tipos de cambio en una molécula dada son cruciales en los intentos por reconstruir la historia de la evolución de un grupo de organismos [idem].

De esa forma, podemos observar que los avances en la evolución molecular representan avances en la solución al problema del origen de las especies y el cómo de la especiación marca pautas para proseguir el estudio de la evolución a nivel micro. Esto último se ve representado por la *paleontología molecular*, área de la biología que estudia las especies extintas a través de los mecanismos provistos por la biología molecular. Por ejemplo, utilizando reacciones de polimerasa en cadena se pueden recuperar secuencias particulares de DNA provenientes de especímenes extintos encontrados en museos, utilizando para ello material orgánico preservado (como piel y músculos o huesos) [23, p.129]. Utilizando ese método y recurriendo a los resultados que ofrece la evolución molecular ha sido posible establecer relaciones filogenéticas de especies extintas, como en el caso del cuaga (mamífero africano extinto) y del lobo marsupial australiano. De igual forma ha sido posible determinar relaciones ascendentes-descendientes entre poblaciones humanas extintas, tal es el caso de los hombres de la ciénaga (*bog people*) de la edad de hierro en Escandinavia y momias egipcias, entre las cuales los resultados provenientes de comparaciones morfológicas habían sido ambiguos [23, p. 130].

Ya que se ha puesto de manifiesto la relación entre evolución a nivel molecular y a nivel macro, es importante que se expongan las principales corrientes en tanto al estudio de la evolución de las especies.

Sin lugar a dudas la teoría sobre el origen de las especies más debatida es la que Charles R. Darwin presentó en 1859. Por mucho tiempo se ha creído que esta teoría explica de manera plausible el origen de las especies y más tarde, con los neodarwinistas actuales, se ha llegado a pensar que es irrefutable [23, p. 38]. Hay corrientes que no están a favor de esa postura ni de los argumentos dados por Darwin, o no del todo, y resultado de esto son nuevas teorías en torno a la especiación.

De entre las diversas teorías sobre la especiación sobresalen:

- El punto de vista más difundido entre los palentólogos, que corresponde al punto de vista darwinista, y que plantea que las nuevas especies provienen de una transformación lenta y progresiva de poblaciones enteras. Esta teoría es conocida como *gradualismo*.
- La teoría de la *especiación geográfica*, propuesta por Sewall Wright, que sugiere una interpretación diferente a la información paleontológica. Esta interpretación considera que

las nuevas especies no evolucionan en la misma área de sus ancestros sino que para que la evolución se lleve a cabo una parte de ellos debe aislarse del resto, generando así un ambiente que obligue al cambio y a partir de ahí la especie evolucione. En esta teoría la evolución no se origina de la transformación lenta de todos sus antepasados.

- La teoría del *equilibrio punteado*. Esta teoría ha sido ampliamente discutida desde que fue propuesta en 1972 por Stephen Jay Gould y Niles Eldredge. Su principal punto de vista consiste en argumentar que la historia de la vida no es mejor explicada como un proceso de cambio continuo a lo largo del tiempo; ellos sugieren que la historia de la vida es un estado de aparente equilibrio que es alterado sólo en ocasiones por sucesos rápidos y espaciados. Ellos rechazan que la evolución sea continua mediante transformaciones graduales.

La teoría del gradualismo y la del equilibrio punteado están basadas en la teoría de la selección natural, propuesta por Darwin; mientras que la especiación geográfica tiene su origen en la propuesta de Sewall Wright, conocida como deriva génica.

La *selección natural* está definida como el cambio en la reproducción, al paso del tiempo, de genotipos o individuos distintos dentro de una población. El cambio en la reproducción es causado por factores como la mortalidad, la fertilidad, la fecundidad y la posibilidad de dejar descendencia. La selección natural está dada por el cambio a través del tiempo en la frecuencia de los alelos (un alelo es cada una de las formas alternativas de un gene particular). Cuando una población consiste de individuos que no difieren unos de otros en ese sentido, es decir que para un gen particular presentan un mismo alelo, entonces no están sujetos a selección natural [23, p. 22].

La *deriva génica* establece que no todos los cambios en la frecuencia de los alelos está dada por selección natural, algunos cambios ocurren por azar. Más aún, esta teoría explica que los cambios en la frecuencia de los alelos tienen un comportamiento acumulativo, es decir que de generación en generación, la frecuencia de un alelo tenderá a desviarse más y más de su frecuencia original, lo que al final resulta en la fijación de un alelo y la pérdida de los otros. Esto ocurre, dice la teoría, a menos que en la población haya una entrada constante de alelos por procesos como mutación o migración, o a

menos que la coexistencia de dos o más alelos en un mismo gen sea activada permanentemente por un tipo de selección balanceado [23, p. 29].

Es importante mencionar que los estudios sobre la evolución de las especies son difíciles de probar debido a que los registros históricos en los que se basan, fósiles en su mayoría, en ocasiones se encuentran incompletos, por lo que impiden establecer de manera precisa cuál fue el procedimiento de especiación que se siguió para el caso en estudio. Existe además la desventaja que la propia evolución de la Tierra representa, ya que la geografía en que se desarrolló la especie analizada puede no ser la misma en la actualidad, lo que obliga a considerar únicamente de manera parcial el factor geográfico y de interacción con el medio que las teorías proponen.

Aun con estas limitantes la teoría del equilibrio punteado ha obtenido una cantidad considerable de seguidores durante los 26 años que han pasado desde su propuesta y, más aún, ha sido aceptada por la mayoría de los paleontólogos [19, p. 223]. Algo que favorece a esta teoría es que no centra su estudio en algún factor especial (como el estudio a través del fósil o de la geografía) sino que plantea un concepto más integral y por ende más flexible. Como sus autores comentan, "su descontento con el gradualismo se debe a que es una teoría rígida, un dogma restrictivo [18, p. 119]." En contraste con otras corrientes, cuyo estudio se enfoca únicamente en las partes de la historia en que hay cambios y desprecian los estados "estacionarios" de la evolución, bajo el argumento de que no representan datos importantes, Gould y Eldredge plantean justamente lo contrario; ellos consideran que esos estados estacionarios en realidad representan bastante información; dentro de ellos hay cambios morfológicos en la especie que son despreciados al no considerarse en los estudios propuestos por otras teorías y que implica recurrir a argumentos rígidos, a partir de la poca información obtenida con los registros históricos del cambio notorio entre una especie y otra, para explicar la especiación.

Otra diferencia importante entre la teoría del equilibrio punteado y las otras teorías planteadas es que los autores de la primera tratan a las especies como unidades reales en la naturaleza, no como elementos arbitrarios en un camino continuo entre individuos y reinos [18, p. 117].

Dadas las características novedosas de la teoría del equilibrio punteado y considerando el rechazo que tuvo, principalmente en sus inicios, Gould y Eldredge trabajaron arduamente a lo largo de varios años para hacer más preciso su planteamiento y enfatizar que es una teoría totalmente comprobable, ventaja importante sobre las demás [18, p. 120]. El resultado es una

teoría que ha adquirido cada vez más seguidores y que ha provisto de mejores argumentos para el cómo de la especiación a la mayoría de los palentólogos, pues es un modelo que es válido para distintas escalas de la evolución. Esto es apoyado por información proveniente de la genética molecular porque se cree que cambios genéticos grandes comunmente son acompañados de procesos de especiación [18, p. 115].

Evolución molecular, ¿continua o discontinua?

Después de este recuento histórico se puede observar que en dos áreas aparentemente distintas (la evolución de las especies y la evolución molecular/biología molecular) se ha logrado obtener argumentos favorables al origen de la especiación o de la vida en sí, respectivamente y, dada su relación intrínseca es de vital importancia que dichos estudios no corran por separado pues, como Walter Fontana y Peter Schuster establecen, “el concepto de evolución no puede ser separado del entendimiento entre la relación genotipo-fenotipo. De hecho está definida por ella [12, p. 1454].”

Siguiendo esa línea de estudiar en paralelo la evolución a nivel molecular y a nivel macro, es importante enfatizar que tal vez en un “análisis final, el puntualismo podría verse como un fenómeno intrínseco a la evolución de una entidad y menos dependiente de fuerzas externas como hasta ahora se ha supuesto [idem].” Puesto que los cambios en el fenotipo, es decir en la forma y estructura de un espécimen, están asociados a cambios en el genotipo, la evolución de las especies puede ser estudiada en términos de la evolución molecular. Así pues, es importante tratar de averiguar si la evolución molecular es gradual o punteada, es decir, si corresponde al tipo de evolución propuesta por el gradualismo o por el equilibrio punteado.

Para distinguir en la evolución molecular entre un cambio continuo o discontinuo (es decir, gradual o punteado —la evolución punteada está dada por transiciones discontinuas), se requiere de una relación de cercanía entre fenotipos. Tal relación esta basada en la probabilidad de que algún fenotipo sea accesible desde otro mediante cambios en el genotipo.

El RNA combina en una sola molécula genotipo (secuencias) y fenotipo (estructura secundaria) haciéndolo ideal para experimentos sobre evolución.

Para establecer la cercanía entre fenotipos se utiliza un modelo matemático que relaciona el RNA con su estructura secundaria mediante la función de plegamiento, que asocia a cada secuencia de RNA de longitud fija, su es-

estructura secundaria de mínima energía. Se definen el espacio de secuencias, que corresponde a un espacio métrico con la métrica de Hamming, y el de estructuras, formas, el cual es un espacio topológico. Se define la topología del espacio de formas y se definen además los conceptos de cercanía en uno y otro espacio.

Una vez definido el concepto de cercanía en cada uno de los espacios, se hace notar mediante la función de plegamiento, que dos secuencias de RNA cercanas implican la cercanía de las estructuras secundarias correspondientes.

Hecho lo anterior se establecen los criterios de continuidad para cada uno de los espacios para finalmente analizar cuándo un proceso de evolución es continuo o discontinuo.

Un resultado bonito e importante es que el espacio de formas tiene asociada una gráfica dirigida, por lo que el concepto de continuidad en un proceso evolutivo puede ser visto en términos de caminos dirigidos en dicha gráfica.

En los próximos capítulos se darán las bases, tanto biológicas como matemáticas, que nos introducirán al planteamiento del modelo sobre continuidad en la evolución.

En el capítulo 1 se presentará la biología del RNA. Se explicará brevemente cuál es el procedimiento que dada una secuencia de DNA induce una secuencia de RNA. Se hablará también de los tipos de RNA que existen, de cuál es su función y su estructura. Se definirá el término "estructura de mínima energía" y se hablará sobre la función de plegamiento.

En el capítulo 2 se abarcarán las nociones sobre Teoría de las Gráficas, necesarias para la comprensión clara de la interpretación de un proceso evolutivo continuo en términos de un camino dirigido. Se definirá aquí lo que es una gráfica y una gráfica dirigida. Se presentarán algunas propiedades de las digráficas, manera en que se llama a las gráficas dirigidas, pero debido a que las gráficas nos sirven en el modelos sólo para entender mejor el concepto de continuidad en evolución, no se ahondará en el estudio de las gráficas, en cambio se darán algunas referencias para aquellos que esten interesados en esta materia.

El capítulo 3 corresponde a la parte de Análisis y Topología. Aquí se presentarán los conceptos de espacio métrico y topología y se empezarán a relacionar con las definiciones utilizadas en el modelo. Se darán ejemplos sencillos de espacios métricos y espacios topológicos para comprender mediante ellos qué es cada uno. Se dan además las definiciones de vecindad, término importante a lo largo del modelo e indispensable para la construcción de los espacios de secuencias y formas, a los que ya se ha hecho mención. Se

presentarán además las vecindades en cada uno de los ejemplos escogidos.

El modelo se planteará y formalizará en el capítulo 4. este capítulo está dividido en 2 partes, la primera nos mostrará el modelo en sí con todas las definiciones que utiliza, la explicación y las conclusiones. La segunda parte es la formalización del modelo. Aquí se darán las demostraciones que prueban que el modelo está bien fundamentado. Se dará la demostración de que el espacio de secuencias es un espacio métrico y se aclarará cuáles son las vecindades en este espacio. Lo mismo se hará para la parte más fuerte del modelo, que es la definición de la topología del espacio de formas.

El quinto y último capítulo planteará una futura línea de investigación. Se hablará brevemente acerca de una estructura secundaria particular del RNA, la de Hammerhead. Se explicará su importancia en las áreas biológica y médica y se planteará una aplicación del modelo a este caso particular.

A partir del primer capítulo y a lo largo de la Tesis se definen los términos biológicos a los que se hace referencia. Varios de ellos se definen dentro del texto en el que se mencionan, tal es el caso de las palabras en negritas. Los términos que se presentan en itálicas se definen al final de la tesis, en el glosario.

Contenido

Introducción	iii
1 Sobre la Biología del RNA	1
2 Un paseo por la Teoría de las Gráficas	11
3 Análisis y Topología	29
4 El modelo y su justificación	43
5 Un caso Particular: HH	57
Bibliografía	61
Glosario	65

Capítulo 1

Sobre la Biología del RNA

Este capítulo presenta las bases biológicas necesarias para la comprensión del modelo que se dará en el capítulo cuatro. La información biológica que se da es en su mayoría sobre el RNA, esto debido a que el modelo centra su aplicación en secuencias y estructuras secundarias de RNA.

1.1 Del DNA al RNA y a las proteínas

Hacia 1945 el gene ya era considerado como unidad fundamental de la herencia, pero se sabía poco acerca de cómo funciona y cuál es su estructura. Los genes sólo podían identificarse por mutaciones que produjeran cambios fenotípicos, es decir visibles en la forma y/o estructura del organismo. Estos cambios variaban desde alteraciones simples como el color de los ojos hasta cambios morfológicos drásticos, como la textura de un chícharo o el tamaño y forma de las alas de una mosca.

A principios de siglo se llevaron a cabo muchos trabajos sobre errores fenotípicos, como el albinismo o la pigmentación en las plantas y los animales. Estos trabajos permitieron comenzar un estudio sistemático que relacionara los factores hereditarios, o **genes**, con las enzimas.

En 1941 George W. Beadle y Edward L. Tatum se plantearon preguntas que los llevaran a comprender de mejor manera la función de los genes. Ellos cuestionaron cuáles son los pasos metabólicos en la producción de las proteínas y qué alteraciones impiden la formación normal de éstas. Sus estudios con el hongo del pan *Neurospora crassa* les permitieron establecer que los genes producen enzimas que actúan directa o indirectamente en la cadena

metabólica de la síntesis de proteínas. Además establecieron que si existe una alteración, **mutación**, que afecte un gene en la secuencia bajo síntesis, ésta se bloquea y el resultado es la ausencia de la proteína deseada.

Así pues, Beadle y Tatum afirmaron que las mutaciones en los genes producen su inactivación o no funcionamiento y, por primera vez, se relacionó la actividad bioquímica de un gene con su estructura molecular [8. p. 22].

Beadle y Tatum acuñaron a partir de entonces la famosa frase *un gene. una enzima*, que se refiere al hecho de que se necesita la acción de un gene para producir una enzima. Esta frase se ha modificado con el tiempo pues ahora se sabe que los genes codifican la formación de **polipéptidos**, moléculas más pequeñas que forman a las proteínas [idem].

Una vez que se mostró evidencia entre la relación de un gene y la síntesis de proteínas, a partir de los años cincuenta se cuestionó qué determina la secuencia de estas últimas y con ello la pregunta es qué determina el desarrollo de uno u otro organismo. La respuesta dada a esta pregunta por la genética clásica es muy formal: "las secuencias de proteínas están determinadas por los genes". Sin embargo, la respuesta a de qué manera está determinada por ellos no es explicada de manera precisa mediante esa área de la Biología [23. p. 18].

De esa forma, todavía faltaban por contestar qué elementos conforman al material genético que contiene a los genes, cómo se duplica para ser transmitido de células madres a células hijas y cuál es el procedimiento que origina las proteínas.

Después de la publicación del trabajo de Watson, Franklin y Crick estas preguntas se resolvieron de una vez por todas. En particular se dió respuesta a qué determina la secuencia de proteínas: "la secuencia de aminoácidos de todas las proteínas de las células está determinada por la secuencia de residuos de una de las dos hebras complementarias del DNA [idem]."

Para ser más precisos observemos lo siguiente:

Según el modelo de Watson, Franklin y Crick, las moléculas de DNA consisten de 2 cadenas de *polímeros*. Cada cadena consiste de 4 tipos de residuos llamados A (**Adenina**), G (**Guanina**), C (**Citosina**) y T (**Timina**). La secuencia de bases puede ser totalmente arbitraria pero las secuencias en ambas cadenas están fuertemente interconectadas debido a un principio de complementaridad: $A \leftrightarrow T$; $T \leftrightarrow A$; $G \leftrightarrow C$; $C \leftrightarrow G$; donde $x \leftrightarrow y$ significa: "x es siempre complementario. a y." (figura 1.1)



Figura 1.1: Representación gráfica de una secuencia de DNA

El concepto de que cada proteína consiste de una serie particular de *aminoácidos* data de los años cincuenta, en los que Sanger caracterizó la insulina [22. p. 71]. Esto llevó a los científicos a cuestionarse cómo una secuencia de nucleótidos en el DNA representa una secuencia de aminoácidos en una proteína.

La secuencia de *nucleótidos* en el DNA es importante por su estructura *per se*, ya que codifica la secuencia de aminoácidos que constituye la proteína correspondiente. La relación entre una secuencia de DNA y la secuencia de la proteína generada a partir de la primera, es llamada **código genético** y

es esto lo que pudo ser aclarado a partir del modelo de Watson, Franklin y Crick.

El punto central ahora es cómo un gene da origen a una proteína. Esto ocurre en dos fases. Durante la primera, **transcripción**, una enzima particular reconoce la secuencia de nucleótidos entre los genes y moviéndose a lo largo de un gene crea una copia de éste en la forma de una molécula de RNA.

La estructura química de una molécula de ácido Ribonucleico (RNA) es muy similar a la de una molécula de ácido desoxirribonucleico (DNA). El RNA es también una cadena de polímeros formada por 4 tipos distintos de nucleótidos: A, C, G y U (**Uracilo**). Entre la A, C y G utilizados en el DNA y el RNA no existen cambios significativos, sin embargo, la timina y el uracilo difieren químicamente uno del otro. Aún con ese cambio el RNA está regido por las mismas reglas de complementaridad utilizadas por el DNA, de forma que: $A \leftrightarrow U$; $U \leftrightarrow A$; $G \leftrightarrow C$; $C \leftrightarrow G$; donde $x \leftrightarrow y$ significa: "x es siempre opuesto, complementario, a y."

Existen otras dos diferencias importantes entre el DNA y el RNA: la primera es que la molécula de azúcar utilizada por el RNA es la ribosa y no la desoxirribosa; la segunda es que el RNA en contraste con el DNA está formado por una sola hélice, es decir, que en general el RNA no se encuentra en la forma espiral, o de doble hélice, en que se conoce al DNA.

En conclusión, un gene es copiado de acuerdo a la misma regla de complementaridad que gobierna la replicación del DNA, con la diferencia de que el papel que juega la T en el DNA es hecho ahora por la U en el RNA.

Para recordar el proceso de réplica del DNA basta ver la figura 1.2.

La síntesis del RNA procede de una de las dos hélices complementarias; la enzima que lleva a cabo este proceso es llamada **polimerasa del RNA**. Así pues, la polimerasa del RNA crea una copia de un gene del DNA. Si la célula es *procarionte* dicha copia recibe el nombre de RNA mensajero. En el caso de células *eucariontes* el RNA mensajero procede de un proceso en dos partes.

La primera parte consiste en obtener una secuencia de RNA como se ha explicado hasta ahora. Dicha secuencia está formada por zonas que codifican, es decir que almacenan partes de la información necesaria para producir una proteína, separadas entre sí por zonas que no codifican. Las zonas que codifican se llaman exones y las que no lo hacen se llaman intrones. Este RNA se llama RNA pre-mensajero. En la segunda etapa, unas enzimas llamadas nucleasas, cortan el RNA en secciones de forma que se puedan retirar todos

los intrones para dejar juntos a los exones. Este proceso da origen a una secuencia de RNA que sirve para sintetizar una molécula de proteína. Dicho RNA corresponde al RNA mensajero (figura 1.3). Este procedimiento en dos partes no es necesario en células procariontes pues la copia de RNA obtenida en el primer paso es en sí una copia que codifica.

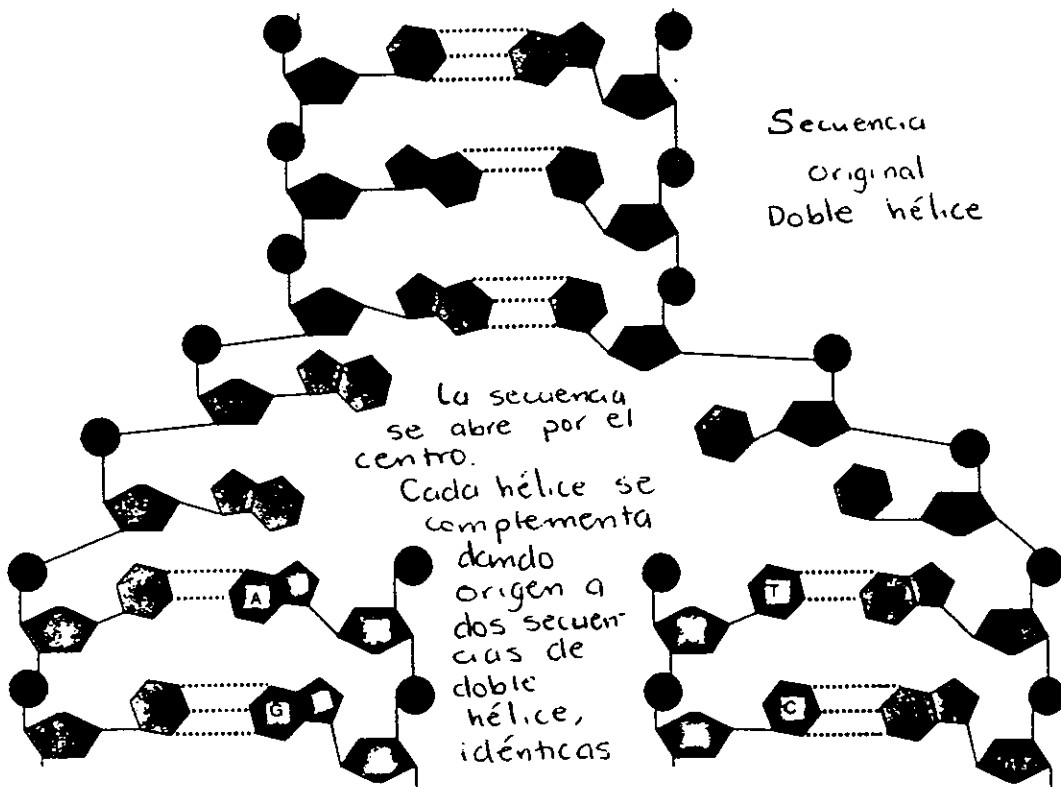


Figura 1.2: Replicación de una secuencia de DNA

El RNA mensajero es utilizado en la segunda fase de la síntesis de proteínas.

la cual es llamada **traducción**. La traducción es un proceso complejo que implica la acción de alrededor de 50 proteínas diferentes y una molécula de RNA. Esta "maquinaria" es llamada **Ribosoma**.

El RNA que utiliza el ribosoma no es el RNA mensajero (mRNA), el cual dirige la síntesis de la proteína dentro del ribosoma, sino el RNA ribosomal (rRNA), el cual es parte integral de la "máquina" a la cual hacemos mención.

Un ribosoma es como una computadora molecular que traduce textos del

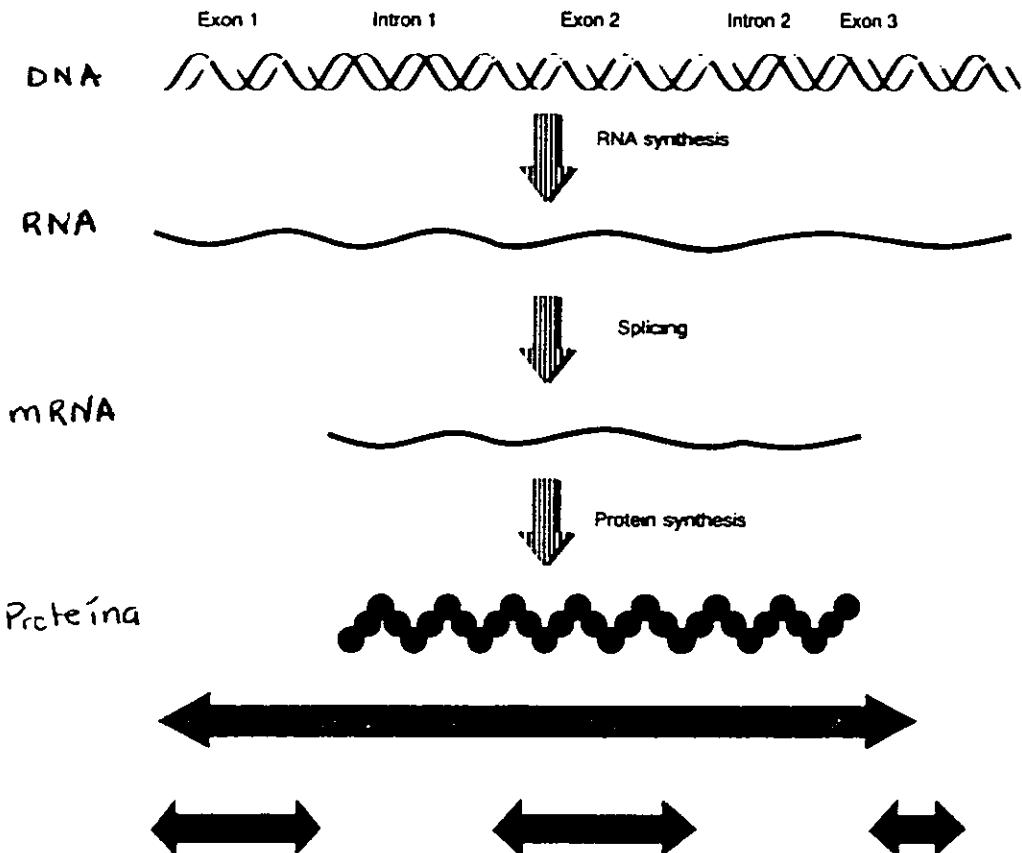


Figura 1.3: Síntesis del RNA en una célula eucarionte

lenguaje de nucleótidos del DNA, o del RNA, al lenguaje de aminoácidos

de las proteínas. Esta "computadora" especializada trabaja de acuerdo a un único programa conocido bajo el nombre de código genético, el cual se muestra en la figura 1.4.

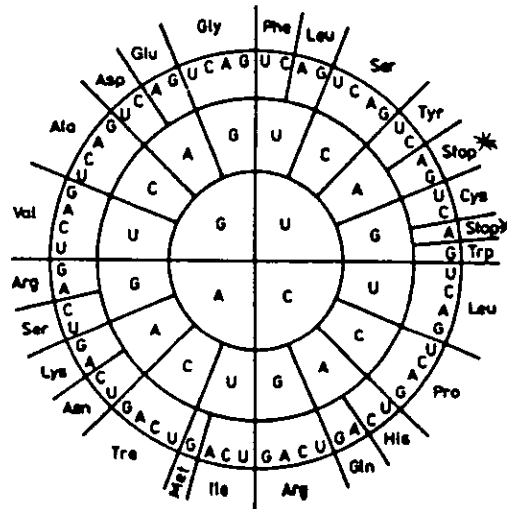


Figura 1.4: El código genético

Aún cuando la traducción está a cargo de los ribosomas, hace falta una pieza esencial para el mecanismo de traducción. Esta pieza es un adaptador-descifrador que relaciona el lenguaje del mRNA con el de las proteínas y asegura el lugar exacto que ocupará cada aminoácido en la proteína. Este adaptador tiene un *anticodón* en el extremo que queda en contacto con el RNA mensajero y, en el otro extremo, el aminoácido correspondiente al *codón* que identifica. Este adaptador es llamado **RNA de transferencia (tRNA)**. Ver la figura 1.5.

Existen en la célula tantos tipos distintos de tRNA como codones hay para los aminoácidos. El RNA de transferencia reconoce su lugar en el RNA mensajero reagrupando localmente en este último una doble hélice, para ello utiliza la interacción del codón-anticodón. Es debido a esto que los aminoácidos que *formarán* la proteína son conectados unos con otros en el

lugar correcto.

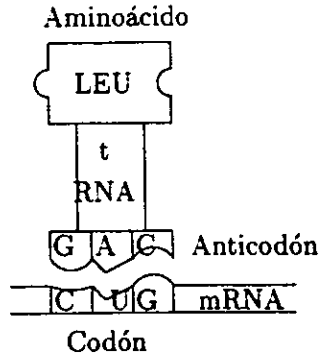


Figura 1.5: RNA de transferencia

El ribosoma es la base del proceso de “ensamble” de las proteínas. El provee dos lugares para dos tRNAs, permite el transporte de energía necesario para llevar a cabo la traducción, permite que la proteína sea ensamblada de manera adecuada y además se mueve a lo largo del mRNA codón por codón para lograr una síntesis de proteína “perfecta”, sin traslapes.

El proceso de traducción es explicado esquemáticamente mediante la ilustración 1.6.

Es así como en términos generales se lleva a cabo una proteína. Una vez explicado este procedimiento y que hemos visto la relación tan importante que hay entre el RNA y la síntesis de una proteína, la que en conjunto con las demás con que cuenta un organismo determinan la formación y desarrollo de éste, estamos en condición de resumir los tipos de RNA que existen y hablar sobre algunas de sus propiedades.

1.2 Distintos tipos de RNA

Como vimos en la sección anterior los genes son transmitidos de generación en generación como secuencias de ácidos nucleicos pero funcionan en la forma

de proteínas.

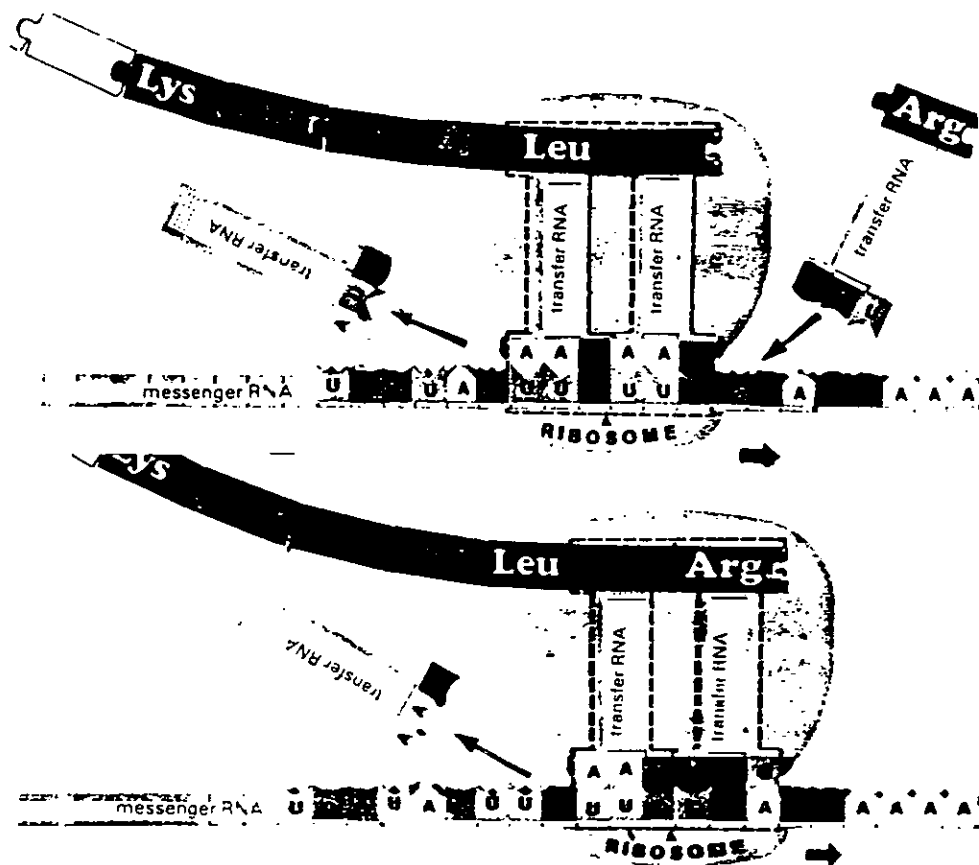


Figura 1.6: El proceso de traducción. Dos tRNA encuentran su lugar en el ribosoma a la vez que los codones y anticodones respectivos forman localmente una secuencia de doble hélice. Se establece la unión entre los dos aminoácidos. El ribosoma se mueve hacia el siguiente codón y libera el primer tRNA, quedando nuevamente dos tRNA en su interior: el proceso se repite hasta terminar la secuencia de RNA mensajero.

Hay tres procesos (replicación, transcripción y traducción) que son los responsables de la herencia de la información genética y de la conversión entre una forma de expresarla (secuencias de nucleótidos) y otra (secuencias de aminoácidos): durante el desarrollo de dichos procesos aparecen varios tipos de RNA de los cuales los principales son:

- El RNA mensajero (mRNA): el cual contiene exactamente la misma información que el segmento de DNA del que proviene. La información que contiene está en forma de nucleótidos.
- El RNA de transferencia (tRNA): está hecho de 20 variedades distintas una para cada aminoácido. Tiene forma de una "L" y carga un triplete de nucleótidos en un extremo y el aminoácido correspondiente al triplete complementario en el otro extremo.
- El RNA ribosomal (rRNA): se asocia con el proceso que indica el momento en que empieza la síntesis de una proteína. Además se ha visto que el RNA ribosomal es también el blanco de algunos antibióticos y otros agentes que inhiben la síntesis de proteínas. Ya que los ribosomas desarrollan el papel central en la preservación de la vida, es muy probable que se hayan conservado sin cambios esenciales durante miles de millones de años, es por ello que son como reliquias de las células primitivas y eso hace que el RNA ribosomal sea utilizado comúnmente para experimentos de evolución molecular.

Dentro de la célula hay distintas formas de RNA, el cual se presenta generalmente como una sola hebra. Sin embargo, tanto el RNA como el DNA (en forma de una sola hélice) pueden generar regiones de doble hélice.

Se puede formar una región doble intramolecular en moléculas de una sola hebra cuando éstas contienen dos secuencias complementarias que emparejan sus bases una con otra utilizando las reglas de complementaridad que rigen el DNA y el RNA. Ver la figura 1.7.

En algunas ocasiones una molécula de una sola hebra puede emparejar sus bases con otra molécula de una sola hebra que sea complementaria, esto da como resultado una región intermolecular doble. Este tipo de emparejamiento de bases no está restringido a interacciones DNA-DNA o RNA-RNA, sino que también puede ocurrir entre una molécula de DNA y otra de RNA.

El concepto de emparejamiento de bases es importante en aquellos procesos que involucran ácidos nucleicos (DNA o RNA). La ruptura de parejas de bases es un aspecto crucial de la función de las moléculas de doble hélice, mientras que la habilidad de generar parejas de bases es esencial para la actividad de ácidos nucleicos de una sola hebra, principalmente del RNA.

Es a partir de esta capacidad de generar parejas de bases dentro de una secuencia de RNA que resultan las diversas estructuras del RNA, de las cuales

hablaremos a continuación.

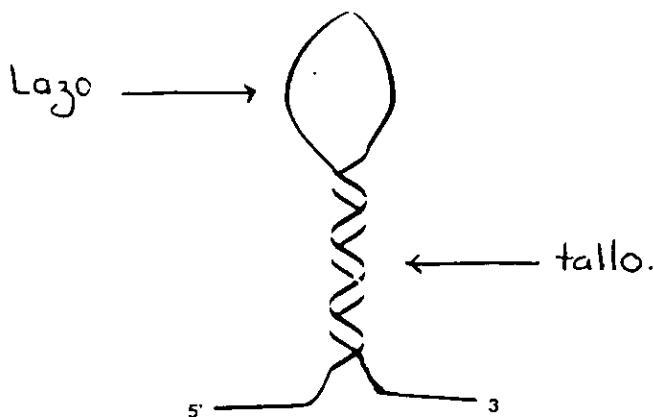


Figura 1.7: Horquilla formada por el emparejamiento de bases de una secuencia de RNA

1.3 Estructuras primaria, secundaria y terciaria

La estructura primaria del RNA es una cadena de polinucleótidos con uniones de azúcar y fosfato, es decir, es una secuencia sencilla de bases nitrogenadas montadas sobre un esqueleto de azúcar y fosfato.

Considerando el RNA en su estructura primaria como una sola hebra, la molécula se desplaza en el espacio y al hacerlo ocasiona que algunas de sus bases se emparejen bajo principios energéticos así como bajo los principios de complementariedad expuestos anteriormente. Cuando esto sucede, la estructura obtenida a partir del emparejamiento de una o más zonas en una secuencia de RNA, recibe el nombre de estructura secundaria de la secuencia.

Cuando una secuencia de bases es seguida de cerca por una secuencia complementaria en la misma molécula, es posible que la cadena se pliegue en

sí misma para generar una estructura doble antiparalela, llamada horquilla (hairpin). Esta horquilla consiste de una región de doble hélice formada por bases emparejadas (tallo) y de un lazo hecho de bases sin emparejar, en un extremo, ver la figura 1.7.

Cuando las zonas complementarias se encuentran relativamente lejos, el efecto producido por su emparejamiento resulta en una horquilla con un lazo muy grande.

La estructura terciaria de una secuencia es creada por uniones de hidrógeno que implican bases que no fueron emparejadas en la estructura secundaria.

Debido a que la capacidad de una molécula de RNA para formar una estructura secundaria es esencial para la actividad de dicha molécula y que la formación de una estructura secundaria ayuda a predecir la estabilidad de la secuencia de RNA, entonces es importante medir la ocurrencia de una estructura secundaria en dicha secuencia.

Hay que recordar que la estabilidad de una doble hélice resulta de las uniones de hidrógeno entre pares complementarios en una secuencia (A-T y G-C) así como de la interacción entre las bases que están "apiladas" una encima de otra a lo largo del eje de la hélice.

La manera más común de medir la ocurrencia de una estructura secundaria está dada por reglas que describen las interacciones de pares de bases. Cuando existe más de una estructura posible la estabilidad relativa a cada una de esas estructuras puede ser evaluada por tales reglas.

Aún cuando este procedimiento para medir la ocurrencia de una estructura secundaria considera que el RNA es una estructura estable y no toma en cuenta la influencia de otras estructuras sobre ella, la aplicación de tales reglas provee un primer paso para calcular la probabilidad de que una estructura particular se forme. Las bases de estas reglas consisten en calcular la *energía libre* de la formación de cada estructura.

La estabilidad de una estructura está determinada por la cantidad de energía liberada. De esta forma, si existen dos estructuras alternativas que tienen energía libre de formación de $\Delta G = -21$ Kcal/mol y $\Delta G = -35$ Kcal/mol, respectivamente, entonces es más probable, intrínsecamente, que esta última sea más estable.

La energía libre global de una estructura secundaria de RNA está dada por la ecuación:

$$\Delta G_{Total} = \sum_{z \in S} \text{energía requerida para sostener una base } z \text{ sin pareja} + \sum_{y \in P} \text{energía liberada por cada emparejamiento } y$$

donde $S = \{\text{bases sin pareja}\}$ y $P = \{\text{parejas de bases}\}$, es decir,

$$\Delta G_{Total} = \sum_{x \in S} \Delta G_x + \sum_{y \in P} \Delta G_y$$

Las reacciones que liberan energía tienen un valor de energía libre negativo. Las reacciones que requieren energía tienen un valor de energía libre positivo. Mantener una base sin pareja requiere energía, así pues, la energía requerida por este proceso tiene constante $\Delta G > 0$, por lo que $\sum_{x \in S} \Delta G_x > 0$. Cada emparejamiento de bases libera energía, así, cada una de esos emparejamientos tiene constante $\Delta G < 0$ y por lo tanto $\sum_{y \in P} \Delta G_y < 0$

La energía libre total (ΔG_{Total}) debe ser negativa en el caso de que una estructura secundaria de RNA sea estable.

1.3.1 Estructura secundaria de mínima energía

Cuando para una secuencia de RNA hay más de una estructura secundaria posible, se determina, por convención, que aquella que es más factible de formarse es la que tiene el valor más pequeño de energía libre total y que por lo tanto es la más estable.

Esto es, dada una secuencia de RNA, τ , y sea $S_h(\tau) = \{\text{estructuras secundarias generadas a partir de } \tau\}$, entonces es más factible que se forme la

$$S \in S_h(\tau) \text{ tal que } \Delta G_{S_{Total}} = \min_{s \in S_h(\tau)} \{ \Delta G_{s_{Total}} \}$$

o bien aquella tal que su valor absoluto de energía libre total sea el máximo de todos los valores absolutos de la energía libre de las estructuras en $S_h(\tau)$, es decir:

$$| \Delta G_{S_{Total}} | = \max_{s \in S_h(\tau)} \{ | \Delta G_{s_{Total}} | \}$$

la razón de esto es que $\Delta G_{s_{Total}} < 0 \forall s \in S_h(\tau)$ y que se considera que entre más estable es una estructura, es más factible que aparezca.

La S descrita aquí es llamada la **estructura secundaria de mínima energía**, asociada a τ .

1.3.2 Función de plegamiento

Basados en los métodos conocidos para calcular la energía libre de una estructura secundaria. Peter Schuster y Walter Fontana desarrollaron un programa

que permite calcular la estructura secundaria de mínima energía, para una secuencia de RNA dada. Un antecedente a este programa es el que presentó Michael Zuker en 1996, el cual también sirve para calcular estructuras secundarias de mínima energía [27].

Dicho programa permite introducir una secuencia arbitraria de RNA y a cambio entrega la estructura bidimensional asociada a dicha secuencia, con la propiedad de que la estructura entregada es la estructura de mínima energía asociada a la secuencia dada.

Sea l una longitud fija. A la asociación hecha entre una secuencia de RNA, τ , de longitud l y su estructura secundaria de mínima energía, S_τ , la llamaremos $F(\tau)$.

Con todas las posibles secuencias de RNA de longitud l definimos el conjunto $S_q = \{\text{secuencias de RNA de longitud } l\}$. Con las estructuras secundarias de mínima energía, a las cuales nos referiremos como **formas**, correspondientes a las secuencias de S_q , definimos el conjunto $S_h = \{\text{formas asociadas a } S_q\}$. Con estos conjuntos definimos $F: S_q \rightarrow S_h$ como la **función de plegamiento** dada por la acción del programa de Walter Fontana y Peter Schuster, para cada una de las secuencias en S_q .

Esta función es de gran importancia para el planteamiento del modelo que trata la tesis. El programa de Schuster y Fontana se puede consultar en [26].

Los capítulos siguientes nos dan las bases matemáticas para el desarrollo del modelo, a continuación se hablará de la teoría de las gráficas.

Capítulo 2

Un paseo por la Teoría de las Gráficas

La construcción de modelos matemáticos puede implicar distintas áreas de la matemática. Frecuentemente se utilizan en Biología gráficas que representen la situación planteada, lo cual se refleja en los ejemplos que se dan a lo largo del capítulo. Aquí se muestran las definiciones más generales dentro de la teoría de las gráficas, incluyendo las digráficas, así como algunas propiedades y aplicaciones. Para un estudio más detallado de la teoría que se presenta en este capítulo se pueden consultar los textos [10] y [20].

Este capítulo nos presenta un panorama útil para la comprensión de la asociación que plantea el modelo entre un camino dirigido entre dos estructuras secundarias de mínima energía de secuencias de RNA y una trayectoria continua en la evolución de dichas estructuras. Si cada estructura secundaria del espacio de formas es vista como un nodo de una gráfica dirigida, entonces, como se verá en el capítulo cuatro, un proceso evolutivo entre dos estructuras secundarias es continuo si existe un paseo dirigido entre esas dos estructuras y al contrario, la evolución entre dichas formas será discontinuo si no existe un paseo dirigido que conecte los nodos respectivos.

2.1 Gráficas

Definición 2.1 (Gráfica) *Una gráfica G es un conjunto finito V distinto del vacío dotado de una relación R , simétrica y antirreflexiva.*

Observación 2.2 *Puesto que R es antirreflexiva entonces ningún elemento de V está relacionado consigo mismo y dado que es simétrica, entonces $\forall (u, v) \in R$ (par ordenado en la relación R), el par $(v, u) \in R$. Denotamos $A = \{\text{pares simétricos en } R\}$.*

En una gráfica, a V se le llama el **conjunto de los vértices** y cada elemento de V es llamado **vértice**. El número de vertices constituye el **orden de G** . De forma análoga, el conjunto A se conoce como **conjunto de aristas** y cada elemento de A , es decir cada conjunto formado por dos pares ordenados simétricos de R , es conocido como **arista**. El número de aristas de G determina el tamaño de G ($|A| = \text{tamaño de } G$, $|V| = \text{orden de } G$.)

Observación 2.3 *una gráfica de orden p y tamaño q es denotada como una gráfica (p, q)*

Ejemplo 2.4 *Sea G definida por $V = \{v_1, v_2, v_3, v_4\}$ junto con la relación*

$$R = \{(v_1, v_2), (v_1, v_3), (v_2, v_1), (v_2, v_3), (v_3, v_1), (v_3, v_2), (v_3, v_4), (v_4, v_3)\}$$

En este caso A está dado por

$$A = \{\{(v_1, v_2), (v_2, v_1)\}, \{(v_1, v_3), (v_3, v_1)\}, \{(v_2, v_3), (v_3, v_2)\}, \{(v_3, v_4), (v_4, v_3)\}\}$$

Dada la observación 2.2, entonces cada vez que $(u, v) \in R$, el conjunto $\{(u, v), (v, u)\}$ es una arista de G , la cual conviene escribir como uv , o bien, vu . Más aún, debido a que el conjunto de aristas esta determinado por la relación R , la cual junto con V define a G , entonces la gráfica puede ser descrita en términos de sus vértices y sus aristas.

Es bastante útil representar una gráfica describiendo los vértices como puntos y las aristas como segmentos de recta que unen los vértices correspondientes. Así pues, del ejemplo anterior tenemos:

Ejemplo 2.5 $V = \{v_1, v_2, v_3, v_4\}$, $A = \{v_1v_2, v_1v_3, v_2v_3, v_3v_4\}$ y G está representada por:

El orden de $G =$ el tamaño de $G = 4$.

Nota: Es conveniente referirse a la representación de G como “ G misma”.

Cabe mencionar que por definición toda gráfica tiene vértices, pues $V \neq \emptyset$, sin embargo es posible que la relación R sea un conjunto vacío, con lo que la gráfica no tendría aristas.

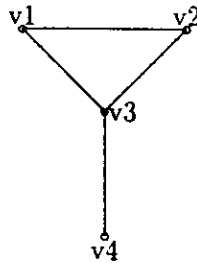


Figura 2.1: Gráfica G

Definición 2.6 (Gráfica trivial) La gráfica $(1,0)$ es llamada gráfica trivial.

Sea $e = uv \in A(G)$ (es conveniente referirse a A como $A(G)$ y a V como $V(G)$), especialmente cuando se está trabajando con varias gráficas a la vez), entonces decimos que e **une** los vértices u y v .

Se dice que dos vértices u y v son **adyacentes** si $\exists e \in A(G)$ tal que $e = uv$ y de manera similar decimos que dos aristas e_1 y e_2 son **incidentes** en v si $\exists u$ y $\exists u'$ tal que $e_1 = uv$ y $e_2 = vu'$, con u y u' diferentes.

Así, en nuestro ejemplo v_1 es adyacente a v_2 , v_2 es adyacente a v_3 y éste es adyacente a v_4 , pero v_4 no es adyacente a v_2 ni a v_1 . Además $e_1 = v_1v_2$ y $e_2 = v_3v_2$ son incidentes en v_2 .

A continuación veremos algunas definiciones relacionadas con las gráficas.

Definición 2.7 (Subgráfica) Sea G una gráfica. Una gráfica H es subgráfica de G si $V(H) \subseteq V(G)$ y $A(H) \subseteq A(G)$.

Ejemplo 2.8 La gráfica H de la figura 2.2, dada por $V(H) = \{v_1, v_2, v_3\}$ y $A(H) = \{v_1v_2, v_1v_3, v_2v_3\}$, es una subgráfica de la gráfica de la figura 2.1.

Definición 2.9 (Paseo) Sean u y v dos vértices de una gráfica G . Un u - v paseo. C en G es una secuencia alternada de vértices adyacentes y las aristas correspondientes, que empieza en u y termina en v , sin importar que los vértices se repitan.

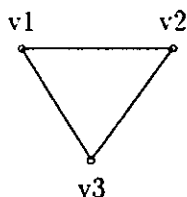


Figura 2.2: Subgráfica de G

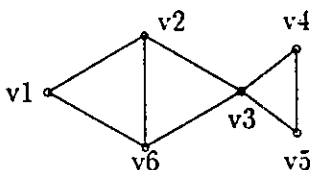


Figura 2.3: Gráfica K

Ejemplo 2.10 Consideremos la gráfica anterior. La secuencia $v_3, v_2, v_1, v_6, v_3, v_4, v_5, v_4$ es un v_3 - v_4 paseo en esa gráfica. Nótese que puesto que los vértices son adyacentes, un paseo queda bien descrito en términos solamente de los vértices pues la arista entre cada par consecutivo de ellos es obvia.

Definición 2.11 (Trayectoria) Una u - v trayectoria en una gráfica, es un u - v paseo que no repite ninguna arista.

Ejemplo 2.12 En la gráfica K de la figura 2.3, el v_3 - v_4 paseo descrito ahí no es una v_3 - v_4 trayectoria; sin embargo, la secuencia $v_3, v_2, v_1, v_6, v_3, v_4$ sí es una v_3 - v_4 trayectoria en la misma gráfica.

Definición 2.13 (Camino) Un u - v camino en una gráfica, es un u - v paseo (o u - v trayectoria) que no repite ningún vértice.

Ejemplo 2.14 De nuevo, utilizando como referencia la gráfica 2.3 tenemos que la secuencia v_3, v_4, v_5 es un camino en dicha gráfica.

Definición 2.15 (Circuito) Una u - v trayectoria en que $u=v$ y tiene al menos tres aristas, es llamada circuito.

Definición 2.16 (Ciclo) Un circuito que no repite ningún vértice, excepto el primero y el último, es llamado ciclo. Un ciclo que recorre todos los vértices de una gráfica se llama ciclo completo.

Ejemplo 2.17 Consideremos la gráfica de la figura 2.4. En esta gráfica la

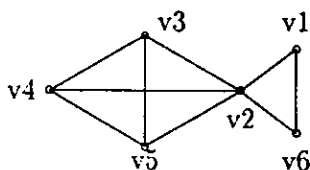


Figura 2.4:

secuencia $v_1, v_2, v_3, v_5, v_2, v_6, v_1$ es un circuito pero no un ciclo. En cambio, la secuencia v_2, v_4, v_3, v_5, v_2 es un ciclo (y también un circuito).

Observación 2.18 Dadas las características de una trayectoria, un camino, un circuito o un ciclo de una gráfica G y debido a que los conjunto de vértices y aristas utilizados por ellos son en sí subconjuntos de $V(G)$ y $A(G)$, respectivamente, entonces cada una de estas secuencias determina una subgráfica de G .

2.1.1 Principales propiedades

Definición 2.19 (Gráfica conexa) Se dice que dos vértices u y v de una gráfica están conectados si bien $u = v$, o si $u \neq v$ y existe una u - v trayectoria en la gráfica. Una gráfica G es conexa si todo par de vértices de G está conectado.

Si no todo par de vértices está conectado, se dice que la gráfica es **disconexa**.

Definición 2.20 (Grado de v) Sea G una gráfica. Sea $v \in A(G)$, el número de aristas que inciden en v es llamado el grado de v en G y se denota como $gr_G(v)$ o, simplemente, $gr(v)$ cuando es claro a qué gráfica se hace referencia.

Ejemplo 2.21 Utilizando la gráfica de la figura 2.4, tenemos que: $gr(v_1) = gr(v_6) = 2$. $gr(v_3) = gr(v_4) = gr(v_5) = 3$ y $gr(v_2) = 5$. Observemos que en esta gráfica la suma de los grados de sus vértices es un número par ($=18$), el cual además es el doble del número de aristas de la gráfica. Esto último no es coincidencia y se refleja en el siguiente teorema, el cual constituye una de las propiedades más generales de las gráficas.

Teorema 2.22 Sea G una (p, q) gráfica con vértices v_1, v_2, \dots, v_p , entonces

$$\sum_{i=1}^p gr_G(v_i) = 2q$$

Es decir, para toda gráfica G , la suma de los grados de sus vértices es igual al doble del número de aristas.

DEMOSTRACIÓN: Puesto que una arista une dos vértices, cuando se suman los grados de éstos cada arista es contada dos veces, una por cada uno de los vértices en que la arista incide. ■

Definición 2.23 (Vértice par (o impar)) Un vértice es llamado par (impar) si su grado es par (impar)

En el ejemplo anterior 2.21 podemos observar que la gráfica a la que se hace mención tiene cuatro vértices impares y dos pares. Esto es consecuencia del teorema 2.22.

Corolario 2.24 Toda gráfica contiene un número par de vértices impares.

DEMOSTRACIÓN: Sea G una gráfica. Si G no tiene vértices impares entonces el resultado se sigue trivialmente (cero es un número par). Supongamos que G tiene j vértices impares, digamos v_1, v_2, \dots, v_j . Si G no tiene vértices pares entonces por el teorema 2.22 tenemos que

$$gr(v_1) + gr(v_2) + \dots + gr(v_j) = 2q$$

que es un número par, pero para cada vértice v_i con $i = 1, 2, \dots, j$, $gr(v_i)$ es un número impar, así pues tenemos una suma de j impares que nos da par por lo que j debe ser par (razón: la suma impar de impares es impar).

De forma análoga, si G también tiene vértices pares, digamos u_1, u_2, \dots, u_m entonces por 2.22 tenemos que

$$(gr(v_1) + gr(v_2) + \dots + gr(v_j)) + (gr(u_1) + gr(u_2) + \dots + gr(u_m)) = 2q$$

Ahora bien puesto que para cada $i = 1, 2, \dots, m$, $gr(u_i)$ es par, entonces el segundo término de la suma anterior es un número par, así pues,

$$gr(v_1) + gr(v_2) + \dots + gr(v_j) = 2q - (gr(u_1) + gr(u_2) + \dots + gr(u_m))$$

que de igual forma es un número par, por lo que se sigue que j debe ser par ■

Definición 2.25 (Gráfica regular) Sea G una gráfica tal que todos sus vértices tienen el mismo grado. digamos r , G es llamada gráfica r -regular, o gráfica de grado r .

Definición 2.26 (Gráfica completa) Una gráfica G es completa si cada par de vértices de G es un par de vértices adyacentes.

Observación 2.27 una gráfica completa de orden p (con p vértices) es $(p-1)$ -regular y se denota como K_p

Ejemplo 2.28 A continuación mostramos algunas gráficas completas:

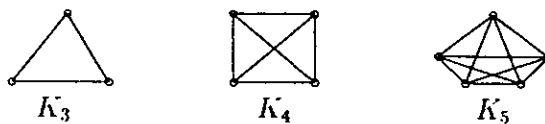


Figura 2.5: Algunas gráficas completas

2.1.2 Las gráficas como modelos matemáticos

Como se dijo al principio del capítulo, es posible resolver problemas haciendo uso de gráficas que representen la situación planteada. Así pues, las gráficas podrían servirnos para solucionar los siguientes problemas abstractos:

Problema 2.29 *Supóngase que hay n islas cerca de la Bahía de los Angeles en Baja California, en el Mar de Cortés. Supóngase también que hay una compañía que da servicio de transporte en barco entre la Bahía y algunas islas, así como entre ciertas islas y otras. Un visitante quiere saber cuál es recorrido que debe hacer para visitar la mayor cantidad de esas islas.*

Esta situación puede ser representada mediante una gráfica en la que los vértices denotan las islas y la bahía (un vértice para cada isla y uno más para la bahía). Dos vértices serían unidos en esta gráfica, si existe una ruta directa en barco que una esos puntos. Así, el visitante podría decidir cual es el paseo que más le conviene de forma que pueda visitar más islas.

Problema 2.30 *En 1859 Sir William Hamilton inventó un juego que utiliza un dodecaedro sólido regular cuyos 20 vértices son bautizados con el nombre de 20 ciudades famosas, respectivamente. El jugador debe viajar "alrededor del mundo" por las aristas del dodecaedro de forma que visite cada ciudad exactamente una vez, pero regresando al lugar de partida.*

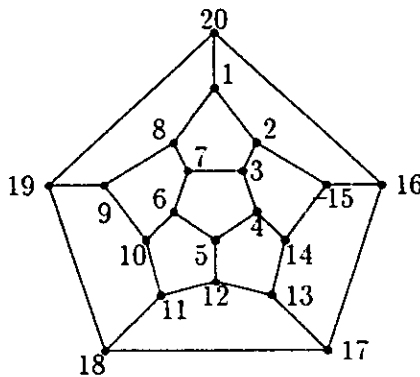


Figura 2.6: Gráfica asociada al dodecaedro

En términos de gráficas, el objetivo del juego es encontrar un ciclo completo en la gráfica que representa un dodecaedro, mostrado en la figura 2.6. Los puntos de la gráfica son denotados como 1, 2, ..., 20 (en lugar de Cd. de México, París, Viena, etc.) con lo que encontrar el ciclo no es muy difícil.

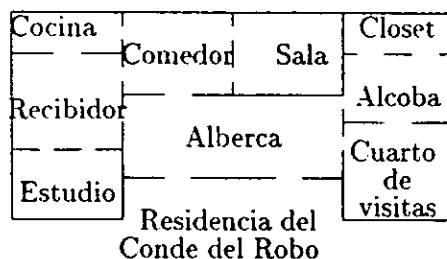


Figura 2.7: Plano de la la residencia del Conde del Robo

Problema 2.31 *Hubo un crimen en la casa del archibillonario Conde del Robo: ¡el Conde fue asesinado! Javier Me Tiche, el detective conocido internacionalmente, notario público, subgerente actual del restaurante más famoso de la ciudad y estudioso de la Teoría de las Gráficas, ha sido llamado para investigar. El amo de llaves asegura que vio al jardinero llegar al jardín, donde está la alberca (lugar en que se encontró el cuerpo del Conde) y poco después de eso lo vio salir por la misma puerta. El jardinero dice, sin embargo, que él no puede ser la persona que describe el amo de llaves, pues advierte que él entró a la casa, pasó por cada puerta exactamente una vez y salió de la casa. Javier Me Tiche estudió cuidadosamente el plano de la residencia, dado en la figura 2.7, y después de unas horas declaró el caso resuelto. ¿Quién mató al Conde?*

Este problema puede ser resuelto, al igual que los anteriores, utilizando la Teoría de las Gráficas; sin embargo, se requiere de más teoría para resolverlo que la que he abarcado hasta ahora. Es una invitación al lector para incursionar en el estudio de esta materia. La información necesaria para resolver este problema puede ser encontrada en cualquier libro de gráficas, pero recomiendo el libro [10] por ser un texto accesible y contener una solución a este problema.

2.2 Gráficas dirigidas

Definición 2.32 (Digráfica) *Una digráfica, o gráfica dirigida, consiste de un conjunto finito V distinto del vacío y una colección de pares ordenados de*

puntos distintos. es decir V está dotado de una relación E antirreflexiva.

Cada punto de V es llamado **vértice** y cada par ordenado (u, v) de E , es llamado **arco o arista dirigida** y usualmente se denota como uv . El arco uv va de u a v y se dice que es **incidente** en dichos vértices. Además se dice que u es **adyacente a v** y que v es **adyacente desde u** .

Definición 2.33 (Grado exterior) Sea $v \in V$, el grado exterior de v , $gr-ext(v)$, es el número de puntos adyacentes desde v .

Definición 2.34 (Grado interior) Sea $v \in V$, el grado interior de v , $gr-int(v)$, es el número de puntos adyacentes a v .

Definición 2.35 (Paseo dirigido) Un paseo en una digráfica, o paseo dirigido, es una secuencia alternada de puntos y arcos, $v_0, a_1, v_1, \dots, a_n, v_n$, en la cual cada arco $a_i = v_{i-1}v_i$.

La **longitud** de un paseo es n , el número de arcos que lo componen.

Definición 2.36 (Paseo cerrado) Un paseo cerrado es un paseo cuyos primer y último puntos son uno mismo.

Definición 2.37 (Paseo completo) Un paseo completo es un paseo que visita todos los vértices de la digráfica.

Definición 2.38 (Trayectoria) Una trayectoria es un paseo que no repite vértices.

Definición 2.39 (Ciclo) Un ciclo es un paseo cerrado, no trivial, que no repite vértices excepto el primero y el último.

Si existe alguna trayectoria de u a v , entonces se dice que v es **accesible** desde u y la **distancia**, $d(u, v)$, de u a v es la longitud de la trayectoria más corta entre u y v . A continuación se dan ejemplos de cada tipo de paseo definido en esta sección. Los ejemplos están basados en la digráfica de la figura 2.8.

Ejemplo 2.40 Paseo dirigido: $v_1, v_3, v_5, v_1, v_3, v_4, v_6$; paseo cerrado: $v_3, v_4, v_6, v_4, v_2, v_1, v_3$; paseo completo: $v_1, v_3, v_5, v_6, v_4, v_2, v_1$; trayectoria: v_4, v_2, v_1, v_5 y ciclo: v_4, v_2, v_1, v_3, v_4

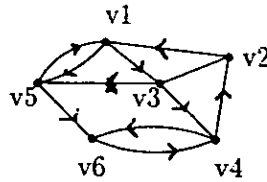


Figura 2.8: Una gráfica dirigida

2.2.1 Principales propiedades

Un paseo en una digráfica es una secuencia dirigida de v_0 a v_n . Entre las propiedades de las digráficas se encuentra la presencia de paseos que no tienen esta característica de dirección pero que son análogos a paseos en una gráfica.

Definición 2.41 (Semipaseo) *Un semipaseo es una secuencia alternada de vértices y arcos, $v_0, a_1, v_1, \dots, a_n, v_n$ pero en la cual cada arco a_i puede ser $v_{i-1}v_i$ o bien $v_i v_{i-1}$.*

De forma análoga se pueden definir una semitrayectoria o un semiciclo. Los siguientes son ejemplos de un semipaseo, una semitrayectoria y un semiciclo, respectivamente. También están basados en la digráfica de la figura 2.8.

Ejemplo 2.42 *Semipaseo: v_1, v_2, v_4, v_3, v_5 ; semitrayectoria: v_4, v_3, v_1, v_5, v_6 y semiciclo: $v_3, v_1, v_2, v_4, v_6, v_5, v_3$.*

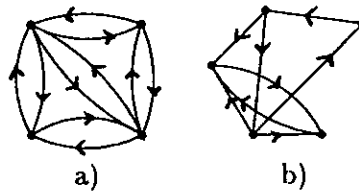


Figura 2.9: a) Digráfica fuerte; b) Digráfica unilateral

Mientras que una gráfica puede ser conexa o disconexa y sólo tiene estas dos posibilidades, una digráfica cuenta con tres formas distintas de conexidad.

Definición 2.43 (Digráfica fuerte) *Una digráfica es conexa fuertemente, o es fuerte, si todo par de puntos en ella es accesible mutuamente, es decir los arcos uv y vu existen para todo par de vértices u y v .*

Definición 2.44 (Digráfica unilateral) *Una digráfica es conexa unilateralmente, o es unilateral, si para todo par de vértices, al menos uno de los*

dos es accesible desde el otro, es decir que para todos u y v vértices, el arco uv existe o bien existe el arco vu .

En la figura 2.9 se muestran un ejemplo de digráfica fuerte y otro de una unilateral.

Definición 2.45 (Digráfica débil) *Una digráfica es conexa débilmente, o es débil, si cualesquiera dos puntos están unidos por un semicamino.*

Es fácil ver que toda digráfica fuerte es una unilateral y que cualquiera unilateral es una digráfica débil, sin embargo los recíprocos no son ciertos.

Una digráfica que no cumple siquiera con ser una digráfica débil, es decir que existan dos puntos para los cuales no haya un semicamino en la digráfica, es llamada digráfica **disconexa**.

En la figura 2.10 se muestran una digráfica débil y una disconexa.

El siguiente teorema¹, ofrece condiciones necesarias y suficientes para los tres tipos de conexidad.

Teorema 2.46 *Una digráfica D es fuerte \Leftrightarrow contiene un paseo cerrado completo, es decir un ciclo que recorra todos los vértices de D .*

Una digráfica D es unilateral \Leftrightarrow contiene un paseo completo.

Una digráfica D es débil \Leftrightarrow contiene un semipaseo completo.

2.2.2 Las digráficas como modelos matemáticos

Al igual que las gráficas, las digráficas también pueden representar situaciones que nos ayuden a solucionar problemas. Para ello es conveniente dibujar una digráfica de forma que los vértices sean expresados como puntos y los arcos como segmentos de línea o curva a los cuales se añade una cabeza de flecha que indica la dirección del arco. Recordemos que el arco uv indica que u se dirige a v . Si en una gráfica los arcos uv y vu existen simultáneamente, entonces se dibujan dos curvas, que no se intersecan, entre u y v y cada arco tiene una cabeza de flecha, cada una en dirección opuesta a la otra.

A continuación veremos un par de ejemplos en los cuales se aplican las digráficas.

¹cuya demostración puede ser consultada en *Structural Models: an introduction to the theory of directed graphs*. Harary, et al. Wiley, New York, 1965

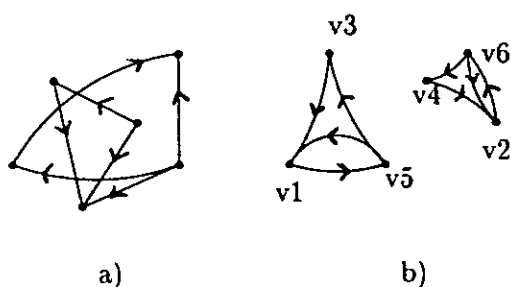


Figura 2.10: a) Digráfica débil; b) Digráfica disconexa

Problema 2.47 *Una ciudad cuenta con varias calles y avenidas de un solo sentido así como con avenidas y calles de doble sentido. Un visitante quiere saber como llegar del Zócalo de la Ciudad al museo de arte sin faltar a las reglas de tránsito. ¿Cómo puede hacer el visitante para resolver su problema sin vagar por toda la ciudad hasta descubrirlo?*

El tránsito admisible de la ciudad puede ser representado mediante una gráfica dirigida en donde las intersecciones entre calles y avenidas importantes se esquematicen por medio de vértices, es decir puntos. En dicha gráfica habría un arco de u a v , si fuera posible manejar, sin infringir la ley.

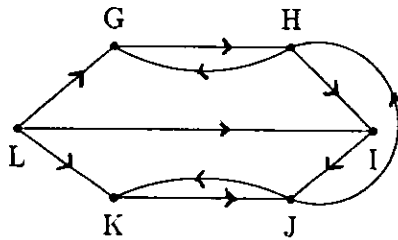


Figura 2.11: Sistema de transporte del problema 2.48

desde el sitio representado por u hasta el sitio asociado a v y en el camino no hubiera un cambio en el sentido de las calles.

Problema 2.48 *El sistema de transporte de cierta ciudad cuenta con seis estaciones, digamos G, H, I, J, K y L . Los trenes circulan de acuerdo a las condiciones siguientes:*

Desde G hasta H . De H a G y de H a I . De I a J . De J a H y de J a K .

De L a G, de L a K y de L a I. De K a J.

Además, es posible transbordar en cualquier estación hacia otro tren.

Dadas las características de este sistema:

¿Cómo se puede llegar de H a J?

¿De las estaciones G, H, I, K y L, a cuáles no se puede llegar desde ninguna otra estación?

¿Desde cuáles de las siguientes estaciones: G, H, J y K, se puede llegar a I con exactamente un cambio de tren?

¿Si la estación L se cierra, cuál de los siguientes paseos no es posible: G a J, J a K, L a K, L a J, L a G?

Este problema puede ser planteado en términos de la gráfica de la figura 2.11. El resto se deja al lector.

2.3 La teoría de gráficas y la evolución molecular

Los modelos matemáticos para genética se han centrado, tradicionalmente, en la genética de poblaciones, utilizando para su estudio bases de la teoría de la probabilidad. Sin embargo, existen áreas dentro de la genética cuyo estudio a través de la probabilidad no es el adecuado, pero que en cambio pueden ser estudiadas mediante la teoría de las gráficas. Estas áreas se refieren a la organización, funcionamiento y evolución de la información genética a nivel molecular. El estudio de dichas áreas utilizando gráficas es razonable cuando se considera a los sistemas genéticos como sistemas discretos, es decir que están formados por una cantidad finita de elementos que interactúan entre sí.

Los matemáticos están descubriendo que varios problemas presentes y pasados dentro de la genética pueden ser formulados en términos de la teoría de las gráficas. Esta aplicación revela a dicha teoría como algo más que una área abstracta de las matemáticas y la enriquece al enfrentarla con nuevos problemas. Así mismo, el uso de métodos matemáticos origina un avance significativo en la comprensión del funcionamiento de los sistemas genéticos a nivel molecular [24, pp. iv, ix].

Existen dos etapas en la aplicación de métodos matemáticos a datos reales. La primera consiste en modelar la situación planteada y la segunda en establecer la manera en que se procesará la información. Esta segunda

etapa surge del hecho de que la información experimental generalmente no es representada adecuadamente en su totalidad por un modelo formal, pues es imposible cubrir con él todas las posibilidades reales de la situación en estudio.

Para los propósitos de varios modelos de gráficas aplicados a la genética a nivel molecular, la primer etapa consiste en plantear la estructura de algún problema de interés y en describir las propiedades de la información que queda representada por el modelo. La segunda etapa consiste en aproximar la información experimental, cuya estructura puede ser complicada, utilizando modelos de estructura sencilla.

El uso de la información genética para estudiar la organización de los cromosomas es uno de los problemas biológicos a los que se ha aplicado la teoría de las gráficas. En este problema se trata de modelar la manera en que están acomodados los defectos genéticos, mutaciones, dentro de un cromosoma. Para ello se utilizan gráficas asociadas a intervalos de \mathbb{R} . Un cromosoma es un conjunto de estructura compleja formado por DNA y proteínas. Los cromosomas se encuentran en el núcleo de la célula y almacenan información genética.

La evolución de una familia de moléculas de hemoglobina es otro problema en el que se aplica la teoría de las gráficas. Para ello se utilizan árboles con pesos en sus aristas. Un árbol es una gráfica conexa sin ciclos. Los árboles se utilizan comúnmente en problemas de optimización de trayectorias y de organización de sistemas de información.

La solución propuesta por Boris Mirkin y Serguei Rodin a estos problemas y a varios más se puede consultar en el libro [24], el cual incluye una amplia explicación de los problemas que plantea; los problemas propuestos en dicho texto se plantean tanto en términos biológicos como en términos de la Teoría de las Gráficas.

El modelo de Walter Fontana y Peter Schuster sobre continuidad en la evolución, no sólo utiliza la Teoría de las Gráficas, también requiere conceptos del área del Análisis Matemático y de la Topología. El siguiente capítulo trata esos temas.

Capítulo 3

Análisis y Topología

Una de las características del modelo que se presentará en el capítulo cuatro es que la teoría necesaria para su justificación no es muy complicada, lo cual lo hace un modelo interesante y accesible para aquellas personas que desean conocer un ejemplo del tipo de trabajo que abarca la Biomatemática. En el capítulo presente se explican las herramientas matemáticas que se utilizan para justificar el modelo. Estas herramientas corresponden a los espacios métricos y a los espacios topológicos, de los cuales se dan ejemplos y algunas de las propiedades principales. Como texto para estudiar estos temas recomiendo el libro [9].

En el modelo se tratan dos tipos distintos de elementos, las secuencias de RNA y sus estructuras secundarias. El conjunto de las estructuras secundarias se asocia con un espacio métrico al cual se le asigna la métrica de Hamming. En este capítulo se muestra la métrica de Hamming para el espacio de secuencias de ceros y unos. El conjunto de estructuras secundarias de mínima energía se organiza en un espacio topológico. Para ello se asigna un conjunto de vecindades para cada elemento del espacio y se prueba que la familia de todos estos conjuntos es un sistema de vecindades en X , lo cual nos determina una topología para el espacio.

Para la comprensión del modelo se requiere la definición de espacio métrico y la definición de la métrica de Hamming, dada en 3.7. Se necesita también la definición de espacio topológico, de vecindad y de sistema de vecindades. Se utiliza el teorema 3.42, que nos dice que: "Dado un conjunto $X \neq \emptyset$, y un sistema de vecindades, \mathcal{N} , de X . Existe una única topología en X que tiene a \mathcal{N} como el sistema de vecindades de dicha topología". Para analizar la continuidad de la evolución mediante la función de plegamiento, se requiere

el criterio de que toda función F es continua si, y sólo, si F preserva abiertos bajo imagen inversa. Si el lector ya conoce estas definiciones y teoremas puede pasar directamente al capítulo cuatro donde encontrará la presentación del modelo. Se recomienda a aquellos lectores que no conozcan alguno de los conceptos indicados, que los revisen en este capítulo antes de leer el desarrollo del modelo.

3.1 Espacios métricos

Definición 3.1 (Métrica) Sea $X \neq \emptyset$, Una métrica en X es una función $d : X \times X \rightarrow \mathbb{R}$ que satisface:

- a) $d(x, y) \geq 0$, $d(x, y) = 0 \Leftrightarrow y = x$
- b) $d(x, y) = d(y, x)$
- c) $d(x, z) \leq d(x, y) + d(y, z)$

Definición 3.2 (Espacio métrico) Un conjunto $X \neq \emptyset$ dotado con una métrica d se llama espacio métrico y se denota como (X, d)

Definición 3.3 (Norma) Una norma en \mathbb{R}^n es una función $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que:

- a) $\| v \| \geq 0$, $\| v \| = 0 \Leftrightarrow v = \bar{0}$
- b) $\forall \alpha \in \mathbb{R}$, $\| \alpha v \| = |\alpha| \| v \|$
- c) $\| v + w \| \leq \| v \| + \| w \|$

Proposición 3.4 Para cualquier norma $\| \cdot \|$ en \mathbb{R}^n , la función $d_{\| \cdot \|} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $d_{\| \cdot \|}(x, y) = \| y - x \|$, es una métrica.

DEMOSTRACIÓN: Sean x, y, z en \mathbb{R}^n ,

- a) $d_{\| \cdot \|}(x, y) = \| y - x \| \geq 0$, por ser norma. $d_{\| \cdot \|}(x, y) = 0 \Leftrightarrow \| y - x \| = 0 \Leftrightarrow y - x = \bar{0} \Leftrightarrow y = x$
- b) $d_{\| \cdot \|}(x, y) = \| y - x \| = 1 \cdot \| y - x \| = |-1| \| y - x \| = \| -1(y - x) \| = \| x - y \| = d_{\| \cdot \|}(y, x)$
- c) $d_{\| \cdot \|}(x, z) = \| z - x \| = \| (z - y) + (y - x) \| \leq \| z - y \| + \| y - x \| = d_{\| \cdot \|}(y, z) + d_{\| \cdot \|}(x, y) = d_{\| \cdot \|}(x, y) + d_{\| \cdot \|}(y, z)$

■

Definición 3.5 (Métrica inducida por una norma) Sea $X = \mathbb{R}^n$ junto con una norma $\| \cdot \|$ cualquiera, la métrica $d_{\| \cdot \|}$ de la proposición anterior se llama métrica inducida por la norma $\| \cdot \|$.

3.1.1 La métrica de Hamming

Proposición 3.6 Sea $X = \{0,1\}^n$ y $d_H : X \times X \rightarrow \mathbb{N}$, dada por

$$d_H(x, y) = \sum_{i=1}^n |y_i - x_i|$$

La función d_H es una métrica.

DEMOSTRACIÓN: Sean x, y, z en X ,

a) $|y_i - x_i| \geq 0 \forall i = 1, \dots, n$, entonces $\sum_{i=1}^n |y_i - x_i| \geq 0 \therefore d_H(x, y) \geq 0$. Además, $d_H(x, y) = 0 \Leftrightarrow \sum_{i=1}^n |y_i - x_i| = 0 \Leftrightarrow |y_i - x_i| = 0 \forall i = 1, \dots, n$ (pues todos los sumandos son no negativos) y esto sucede sí y sólo sí $y_i - x_i = 0$ para todo i , así pues $x_i = y_i \forall i = 1, \dots, n \therefore x = y$

b) $d_H(x, y) = \sum_{i=1}^n |y_i - x_i| = \sum_{i=1}^n |x_i - y_i| = d_H(y, x)$

c) $|z_i - x_i| = |(z_i - y_i) + (y_i - x_i)| \leq |z_i - y_i| + |y_i - x_i| \forall i = 1, \dots, n$. Así pues, la $\sum_{i=1}^n |z_i - x_i| \leq \sum_{i=1}^n |z_i - y_i| + \sum_{i=1}^n |y_i - x_i|$. Por lo tanto, $d_H(x, z) \leq d_H(x, y) + d_H(y, z)$. ■

Definición 3.7 (Métrica de Hamming) Sea $X = \{0,1\}^n$, el espacio de todas las secuencias de longitud n con entradas ceros y unos. $d_H : X \times X \rightarrow \mathbb{N}$ se llama métrica de Hamming en el espacio $\{0,1\}^n$.

3.1.2 Otros ejemplos

Proposición 3.8 Sea $X \neq \emptyset$. El espacio (X, d) donde

$$d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

es un espacio métrico. La métrica d se llama métrica discreta.

DEMOSTRACIÓN: Basta probar que d es métrica. Sean $x, y, z \in X$,

a) $d(x, y) \geq 0$ por definición, además $d(x, y) = 0 \Leftrightarrow x = y$.

b)

$$d(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases} = \begin{cases} 0 & y = x \\ 1 & y \neq x \end{cases} = d(y, x)$$

c) Por casos, si $x = z$ entonces $d(x, z) = 0 \leq d(x, y) + d(y, z)$.

Si $x \neq z$ entonces $d(x, z) = 1$ y

- i) si $y = x$ entonces $y \neq z \therefore d(y, z) = 1 \therefore d(x, z) \leq d(x, z) + d(y, z)$
- ii) si $y = z$ entonces $y \neq x$, entonces $d(x, y) = 1 \therefore d(x, z) \leq d(x, y) + d(y, z)$
- iii) si $x \neq y$ y $y \neq z$, entonces $d(x, y) = d(y, z) = 1$ con lo que $d(x, z) = 1 \leq 2 = d(x, y) + d(y, z) \therefore d(x, z) \leq d(x, y) + d(y, z)$

Por lo tanto d es métrica. ■

Proposición 3.9 Si $X = \mathbb{R}^n$ entonces ninguna norma induce a la métrica discreta.

DEMOSTRACIÓN: Supongamos que $\| \cdot \|$ induce d . Sea $x \neq \vec{0}$, $x \in \mathbb{R}^n$ entonces $\|x\| = \rho > 0$, por las propiedades de la norma. Además $\|\alpha x\| = \alpha x \forall \alpha > 0$, con lo que $\|\alpha x\|$ toma una infinidad de valores, pero $\|\alpha x\| = d(0, \alpha x) = 1 \forall \alpha > 0$, contradicción. Por lo tanto no hay una norma que induzca d . ■

Proposición 3.10 Sea $X = (0, \infty)$, (X, d_1) y (X, d_2) con $d_1(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right|$ y $d_2 = |e^{-x} - e^{-y}|$, son espacios métricos.

DEMOSTRACIÓN: Basta probar que d_1 y d_2 son métricas. Sean $x, y, z \in X$,

I:a) $d_1(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right| \geq 0$ y $d_1(x, y) = 0 \Leftrightarrow \left| \frac{1}{x} - \frac{1}{y} \right| = 0 \Leftrightarrow \frac{1}{x} = \frac{1}{y} \Leftrightarrow x = y$

b) $d_1(x, y) = \left| \frac{1}{x} - \frac{1}{y} \right| = \left| \frac{1}{y} - \frac{1}{x} \right| = d_1(y, x)$

c) $d_1(x, z) = \left| \frac{1}{x} - \frac{1}{z} \right| = \left| \left(\frac{1}{x} - \frac{1}{y} \right) + \left(\frac{1}{y} - \frac{1}{z} \right) \right| \leq \left| \frac{1}{x} - \frac{1}{y} \right| + \left| \frac{1}{y} - \frac{1}{z} \right| = d_1(x, y) + d_1(y, z)$

II:a) $d_2(x, y) = |e^{-x} - e^{-y}| \geq 0$ y $d_2(x, y) = 0 \Leftrightarrow |e^{-x} - e^{-y}| = 0 \Leftrightarrow e^{-x} = e^{-y} \Leftrightarrow -x = -y$, pues e^x es inyectiva. $\therefore d_2(x, y) = 0 \Leftrightarrow x = y$

b) y c) son análogos a los de I. ■

3.1.3 Bolas y conjuntos abiertos

Definición 3.11 (Bola abierta en (X, d)) Sea $X \neq \emptyset$ y $d : X \times X \rightarrow \mathbb{R}$ una métrica en X . Sea $r > 0$ y $x_0 \in X$. El conjunto

$$B_r^d(x_0) = \{x \in X \mid d(x, x_0) < r\}$$

se llama bola abierta, relativa a d , con centro en x_0 y radio r .

Ejemplo 3.12 Si $X = \mathbb{R}^n$ y $d = d_{\|\cdot\|}$, para alguna norma $\|\cdot\|$, $B_r^d(x_0) = \{x \in \mathbb{R}^n \mid \|x - x_0\| < r\}$, para $x_0 \in \mathbb{R}^n$

Ejemplo 3.13 Sea $X \neq \emptyset$, junto con la métrica discreta. Sea $x_0 \in X$ entonces

$$B_r^d(x_0) = \begin{cases} \{x_0\} & 0 < r \leq 1 \\ X & 1 < r \end{cases}$$

DEMOSTRACIÓN: Sea $0 < r \leq 1$, nos interesa $\{x \in X \mid d(x_0, x) < r \leq 1\}$, es decir, nos interesan aquellas x tales que $d(x_0, x) < 1$, lo cual sucede para $x = x_0$. $\therefore B_r^d(x_0) = \{x_0\} \forall r \ni 0 < r \leq 1$.

Sea $r > 1$. $\forall x \in X$, $d(x_0, x) \leq 1 < r \Rightarrow x \in B_r^d(x_0) \forall x \in X$
 $\therefore B_r^d(x_0) = X \forall r > 1$.

■

Ejemplo 3.14 Sean (X, d_1) y (X, d_2) como en 3.10.

$B_1^{d_1}(1) = \{x \in (0, \infty) \mid |\frac{1}{x} - \frac{1}{1}| < 1\} = \{x \in (0, \infty) \mid |\frac{1}{x} - 1| < 1\} = (\frac{1}{2}, \infty)$,
 pues

$$|\frac{1}{x} - 1| < 1 \Leftrightarrow -1 < \frac{1}{x} - 1 < 1 \Leftrightarrow 0 < \frac{1}{x} < 2 \Leftrightarrow \frac{1}{2} < x < \infty$$

Sea $0 < \epsilon < 1$. $B_\epsilon^{d_1}(1) = \{x \in (0, \infty) \mid |\frac{1}{x} - 1| < \epsilon\} = (\frac{1}{1+\epsilon}, \frac{1}{1-\epsilon})$, ya que

$$|\frac{1}{x} - 1| < \epsilon \Leftrightarrow -\epsilon < \frac{1}{x} - 1 < \epsilon \Leftrightarrow 1 - \epsilon < \frac{1}{x} < 1 + \epsilon \Leftrightarrow \frac{1}{1 + \epsilon} < x < \frac{1}{1 - \epsilon}$$

Sea $\epsilon > 0$.

$$B_\epsilon^{d_2}(1) = \{x \in (0, \infty) \mid |e^{-x} - e^{-1}| < \epsilon\} = \begin{cases} (\ln(\frac{\epsilon}{1+\epsilon}), \ln(\frac{\epsilon}{1-\epsilon})) & \text{si } \frac{\epsilon}{1+\epsilon} > 1 \\ (0, \ln(\frac{\epsilon}{1-\epsilon})) & \text{si } \frac{\epsilon}{1+\epsilon} \leq 1 \end{cases}$$

ya que

$$\begin{aligned} |e^{-x} - e^{-1}| < \epsilon &\Leftrightarrow -\epsilon + e^{-1} < e^{-x} < \epsilon + e^{-1} \Leftrightarrow \ln\left(\frac{1}{e} - \epsilon\right) < \ln(e^{-x}) < \ln\left(\frac{1}{e} + \epsilon\right) \\ &\Leftrightarrow \ln\left(\frac{1}{e} - \epsilon\right) < -x < \ln\left(\frac{1}{e} + \epsilon\right) \Leftrightarrow -\ln\left(\frac{1}{e} + \epsilon\right) < x < -\ln\left(\frac{1}{e} - \epsilon\right) \Leftrightarrow \\ &\ln\left(\frac{1}{e} + \epsilon\right)^{-1} < x < \ln\left(\frac{1}{e} - \epsilon\right)^{-1} \Leftrightarrow \ln\left(\frac{1}{e+\epsilon}\right) < x < \ln\left(\frac{1}{e-\epsilon}\right) \end{aligned}$$

y a fin de que $x \in (0, \infty)$, debemos ver si $\ln\left(\frac{1}{e+\epsilon}\right) > 0$, es decir debemos pedir que $\frac{1}{e+\epsilon} > 1$, si no es así, entonces el intervalo solución será $(0, \ln\left(\frac{1}{e-\epsilon}\right))$.

Ejemplo 3.15 (Bolas con la métrica de Hamming) Sea $X = \{0, 1\}^n$ con la métrica de Hamming, sea $x_0 \in X$ y $\epsilon \in \mathbb{N} \cap (0, n]$.

$$B_\epsilon^{d_H}(x_0) = \{x \in X \mid d(x_0, x) < \epsilon\} = \{x \in X \mid \sum_{i=1}^n |x_i - x_{0i}| < \epsilon\}$$

Como $x, x_0 \in \{0, 1\}^n$, entonces cada una de sus entradas es 0 ó 1, con lo que $\forall i = 1, \dots, n$ $x_i, x_{0i} \in \{0, 1\}$, así pues $|x_i - x_{0i}| \in \{0, 1\}$ y más aún, $|x_i - x_{0i}| = 1 \Leftrightarrow x_i \neq x_{0i}$, con lo que

$$\sum_{i=1}^n |x_i - x_{0i}| = \text{número de lugares en que la secuencia } x \text{ difiere de la secuencia } x_0$$

y por lo tanto

$$B_\epsilon^{d_H}(x_0) = \{x \in X \mid x \text{ difiere de } x_0 \text{ en menos de } \epsilon \text{ lugares}\}$$

Observación 3.16 La métrica de Hamming es esencial en la justificación del modelo que se presentará en el próximo capítulo.

Definición 3.17 (Vecindad en (X, d)) Sea (X, d) espacio métrico. Una vecindad de $x \in X$ es un conjunto $U \subseteq X$ para el cual existe un $\epsilon > 0 \ni B_\epsilon^d(x) \subseteq U$.

Definición 3.18 (Conjunto abierto en (X, d)) Sea (X, d) un espacio métrico y sea $G \subset X$ un subconjunto dado. Se dice que G es d -abierto si $\forall x_0 \in G \exists r = r(x_0) > 0 \ni B_r^d(x_0) \subset G$. Es decir, G es d -abierto si para cualquier punto x_0 de G existe una bola alrededor de x_0 que esta contenida en G .

Teorema 3.19 *Sea (X, d) un espacio métrico y sea $R > 0$. Para toda $z \in X$, el conjunto $G = B_R^d(z)$ es un conjunto abierto.*

DEMOSTRACIÓN: P.d. $\forall x_0 \in G \exists r > 0 \ni B_r^d(x_0) \subset G$. Sea $x_0 \in G$, basta probar que si $d(x, x_0) < r \Rightarrow d(x, z) < R$.

Dem.: Sea $r = R - d(x_0, z)$. $r > 0$ ya que $x_0 \in G$ y entonces $d(x_0, z) < R$.

Sea $x \in X \ni d(x, x_0) < r$, es decir $x \in B_r^d(x_0)$, se tiene que

$d(x, z) \leq d(x, x_0) + d(x_0, z) < r + d(x_0, z) = (R - d(x_0, z)) + d(x_0, z) = R$.

$\therefore d(x, z) < R$, por lo que $x \in G$ y en conclusión $B_r^d(x_0) \subset G$.

■

Teorema 3.20 *Sea (X, d) un espacio métrico, son válidas las siguientes afirmaciones:*

- 1) *Si G y H son d -abiertos, entonces $G \cap H$ también es d -abierto.*
- 2) *Si $\{G_\alpha\}_{\alpha \in I}$ es cualquier colección de d -abiertos, entonces*

$$\bigcup_{\alpha \in I} G_\alpha = \{x \in X \mid x \in G_\alpha, \text{ para algún } \alpha \in I\}$$

también es d -abierto.

DEMOSTRACIÓN: 1): Si $G \cap H = \emptyset$, el resultado se sigue por vacuidad. Si $G \cap H \neq \emptyset$, entonces sea $x_0 \in G \cap H$. Como G y H son d -abiertos entonces existe $r_1 > 0 \ni B_{r_1}^d(x_0) \subset G$ y existe $r_2 > 0 \ni B_{r_2}^d(x_0) \subset H$.

Sea $r = \min\{r_1, r_2\}$, entonces $B_r^d(x_0) = B_{r_1}^d(x_0) \cap B_{r_2}^d(x_0) \subset G \cap H$.

$\therefore B_r^d(x_0) \subset G \cap H$ y en conclusión $G \cap H$ es d -abierto.

2): Sea $x \in \bigcup_{\alpha \in I} G_\alpha \Rightarrow \exists \alpha \in I \ni x \in G_\alpha$. Como G_α es d -abierto entonces existe $r = r_\alpha(x) > 0 \ni B_r^d(x) \subset G_\alpha \subset \bigcup_{\alpha \in I} G_\alpha$. $\therefore \bigcup_{\alpha \in I} G_\alpha$ es d -abierto. ■

Corolario 3.21 *Si A_1, \dots, A_n son d -abiertos en X , entonces*

$$\bigcap_{i=1}^n A_i, \text{ también es } d\text{-abierto en } X.$$

3.1.4 Continuidad de funciones

Sean (X, d) y (Y, D) espacios métricos. Sea $f : X \rightarrow Y$ una función.

Definición 3.22 (Imagen directa) Sea $A \subset X$, el conjunto $f(A)$ definido como $f(A) = \{y \in Y \mid y = f(x), \text{ para algún } x \in A\}$ recibe el nombre de imagen directa de A . $f(A)$ es un subconjunto de Y . De manera alternativa se define $f(A) = \{f(x) \mid x \in A\}$.

Definición 3.23 (Imagen inversa) Sea $B \subset Y$, el conjunto $f^{-1}(B)$ definido como $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$ recibe el nombre de imagen inversa de B . $f^{-1}(B) \subset X$.

Definición 3.24 (Función continua en $a \in X$) Se dice que $f : X \rightarrow Y$ es una función continua en $a \in X$ si $\forall \epsilon > 0 \exists \delta = \delta(\epsilon, a) > 0 \ni$ si $a' \in X$ con $d(a, a') < \delta \Rightarrow D(f(a), f(a')) < \epsilon$. Dicho de otro modo, f es una función continua en $a \in X$ si $\forall \epsilon > 0 \exists \delta = \delta(\epsilon, a) \ni f(B_\delta^d(a)) \subset B_\epsilon^D(f(a))$.

Definición 3.25 (Función continua en X) Si $f : X \rightarrow Y$ es continua para toda $a \in X$, se dice que f es continua en X .

Teorema 3.26 Sean (X, d) y (Y, D) espacios métricos y sea $f : X \rightarrow Y$. La función f es continua en $X \Leftrightarrow \forall U \subset Y, D$ -abierto, el conjunto $f^{-1}(U)$ es d -abierto.

DEMOSTRACIÓN: La demostración de este teorema es análoga a la del teorema 3.45 que se dará en la sección 3.2.4.

■

3.2 Topología

Definición 3.27 (Topología) Sea $X \neq \emptyset$ y \mathcal{T} una familia de subconjuntos de X , llamados abiertos. Se dice que \mathcal{T} es una topología para X si:

i) $X, \emptyset \in \mathcal{T}$.

ii) La unión arbitraria de elementos en \mathcal{T} pertenece a \mathcal{T} . Es decir, si $\{A_i\}_{i \in \Omega} \subseteq \mathcal{T}$, entonces $\bigcup_{i \in \Omega} A_i \in \mathcal{T}$.

iii) La intersección finita de elementos en \mathcal{T} pertenece a \mathcal{T} . Es decir, si $\{A_i\}_{i \in \Omega} \subseteq \mathcal{T}$, Ω finito, entonces $\bigcap_{i \in \Omega} A_i \in \mathcal{T}$.

Definición 3.28 (Espacio topológico) *Un conjunto $X \neq \emptyset$ dotado con una topología \mathcal{T} se llama espacio topológico y se denota (X, \mathcal{T}) .*

Teorema 3.29 *Todo espacio métrico (X, d) es un espacio topológico.*

DEMOSTRACIÓN: Sea \mathcal{T} la familia de d -abiertos de X . X y \emptyset son d -abiertos de X (razón: para X cualquier bola sirve; \emptyset es d -abierto por vacuidad.) Así pues, $X, \emptyset \in \mathcal{T}$. Además, por el teorema 3.20, se tiene que si $\{G_\alpha\}_{\alpha \in I}$ es cualquier colección de d -abiertos, entonces $\bigcup_{\alpha \in I} G_\alpha = \{x \in X \mid x \in G_\alpha, \text{ para algún } \alpha \in I\}$ también es d -abierto, por lo que la unión arbitraria de elementos en \mathcal{T} es elemento de \mathcal{T} . Finalmente, por el corolario 3.21, se tiene que la intersección arbitraria de elementos en \mathcal{T} es un elemento de \mathcal{T} . Por lo tanto X es un espacio topológico con topología \mathcal{T} . ■

Teorema 3.30 *Sea $X \neq \emptyset$ y $\{\mathcal{T}_i\}_{i \in I}$ una familia de topologías para $X \implies \bigcap_{i \in I} \mathcal{T}_i$ es una topología para X .*

DEMOSTRACIÓN:

i): Dado que $\{\mathcal{T}_i\}_{i \in I}$ es una familia de topologías, entonces para todo $i \in I$, $X, \emptyset \in \mathcal{T}_i$. Por lo que $X, \emptyset \in \bigcap_{i \in I} \mathcal{T}_i$.

ii): Sea $\{A_j\}_{j \in J, J \subseteq I} \subset \bigcap_{i \in I} \mathcal{T}_i$, entonces $\{A_j\}_{j \in J} \in \mathcal{T}_i \forall i \in I$. Como \mathcal{T}_i es una topología para cada i , entonces $\bigcup_{j \in J} A_j \in \mathcal{T}_i \forall i \in I$. Así pues, $\bigcup_{j \in J} A_j \in \bigcap_{i \in I} \mathcal{T}_i$.

iii): Sea $\{A_j\}_{j \in J, J \text{ finito}} \subset \bigcap_{i \in I} \mathcal{T}_i$, entonces para todo $i \in I$, $\{A_j\}_{j \in J} \in \mathcal{T}_i$. Como \mathcal{T}_i es una topología para cada i y J es finito, entonces $\bigcap_{j \in J} A_j \in \mathcal{T}_i, \forall i \in I$. Así pues, $\bigcap_{j \in J} A_j \in \bigcap_{i \in I} \mathcal{T}_i$.

Por lo tanto $\bigcap_{i \in I} \mathcal{T}_i$ es una topología para X . ■

3.2.1 Algunas topologías

Proposición 3.31 *Sea X un conjunto infinito.*

El conjunto $\mathcal{T} = \{\emptyset\} \cup \{A \subseteq X \mid X \setminus A \text{ es finito}\}$ es una topología.

DEMOSTRACIÓN: i): por definición de \mathcal{T} , $\emptyset \in \mathcal{T}$. Por otro lado, $X \setminus X = \emptyset$, que es finito, entonces $X \in \mathcal{T}$.

ii): Sea $\{A_i\}_{i \in I} \subseteq \mathcal{T} \implies X \setminus A_i$ es finito o \emptyset para cada $i \in I$. Puesto que $X \setminus A = A^c$, entonces A_i^c es finito o vacío para toda i . $(\bigcup_{i \in I} A_i)^c = \bigcap_{i \in I} A_i^c \subseteq A_i^c$. Entonces $(\bigcup_{i \in I} A_i)^c$ está contenido en un conjunto finito o bien es vacío, por lo tanto es finito o \emptyset y en conclusión $(\bigcup_{i \in I} A_i) \in \mathcal{T}$.

iii): Sean $A_1, A_2, \dots, A_n \in \mathcal{T}$. $(\bigcap_{i=1}^n A_i)^c = \bigcup_{i=1}^n A_i^c$. Como A_i^c es finito o \emptyset para $i = 1, 2, \dots, n \implies \bigcup_{i=1}^n A_i^c$ es finito o bien es el conjunto vacío. Por lo tanto $(\bigcap_{i=1}^n A_i)^c$ es finito o \emptyset y en conclusión $(\bigcap_{i=1}^n A_i) \in \mathcal{T}$.

Por lo tanto \mathcal{T} es una topología para X .

■

Definición 3.32 (Topología cofinita) *La topología \mathcal{T} de la proposición anterior se llama topología cofinita.*

Proposición 3.33 *Sea X , conjunto y sea $\mathcal{T} = \{\emptyset, X\}$. \mathcal{T} es una topología para X .*

DEMOSTRACIÓN: i): Es claro que $\emptyset, X \in \mathcal{T}$.

ii): $X \cup \emptyset = X$, por lo que se sigue que la unión arbitraria de elementos de \mathcal{T} es elemento de \mathcal{T} .

iii): $X \cap \emptyset = \emptyset$, por lo que se sigue que la intersección arbitraria de elementos de \mathcal{T} es elemento de \mathcal{T} .

Por lo tanto \mathcal{T} es una topología. \mathcal{T} se llama topología indiscreta.

■

Proposición 3.34 *Sea X un conjunto y $\mathcal{T} = \{A \mid A \subseteq X\}$. (X, \mathcal{T}) es un espacio topológico.*

DEMOSTRACIÓN: i): $X, \emptyset \subseteq X$. Por lo tanto $X, \emptyset \in \mathcal{T}$.

ii): Sea $\{A_i\}_{i \in I} \subseteq \mathcal{T} \Rightarrow A_i \in \mathcal{T} \forall i \in I \Rightarrow A_i \subseteq X \forall i \in I$. Entonces, $\bigcup_{i \in I} A_i \subseteq X$. $\therefore \bigcup_{i \in I} A_i \in \mathcal{T}$.

iii): Sean $A_1, A_2, \dots, A_n \in \mathcal{T} \Rightarrow A_i \subseteq X \forall i = 1, 2, \dots, n$. Entonces, $\bigcap_{i=1}^n A_i \subseteq A_i \subseteq X \Rightarrow \bigcap_{i=1}^n A_i \subseteq X \therefore \bigcap_{i=1}^n A_i \in \mathcal{T}$.

Por lo tanto \mathcal{T} es una topología para X , de donde (X, \mathcal{T}) es espacio topológico. \mathcal{T} se llama topología discreta. ■

3.2.2 Vecindades

Como en el caso de los espacios métricos, se puede definir el concepto de vecindad.

Definición 3.35 (Vecindad) Sea (X, \mathcal{T}) un espacio topológico y sea $x \in X$. Una vecindad de x es un conjunto $U \subseteq X$ tal que $\exists A \in \mathcal{T} \ni x \in A \subseteq U$. Es decir, es un conjunto U para el cual existe un abierto A de \mathcal{T} tal que x está en A y A está contenido en U .

Observación 3.36 Esta definición es consistente con la que se dió para espacios métricos, es decir, una vecindad en espacios métricos es una vecindad en la topología \mathcal{T} del teorema 3.29, definida por la métrica en el espacio.

Teorema 3.37 Sea (X, \mathcal{T}) un espacio topológico. $A \subseteq X$ es abierto si y sólo si A es vecindad de todos sus puntos.

DEMOSTRACIÓN:

\Rightarrow) Si A es abierto, A mismo es vecindad de todos sus puntos.

\Leftarrow) Sea $x \in A$, entonces existe $A_x \in \mathcal{T}$ tal que $x \in A_x \subseteq A$. Entonces $A = \bigcup_{x \in A} A_x$, que es un elemento de \mathcal{T} , ya que es unión de elementos en \mathcal{T} y \mathcal{T} es topología. ■

Observación 3.38 El conjunto de vecindades de un punto x en un espacio topológico X se denota como \mathcal{N}_x^X , o bien como \mathcal{N}_x cuando es claro el espacio en el que se trabaja.

Definición 3.39 (Sistema de vecindades de una topología) Sea X un espacio topológico. El conjunto $\mathcal{N} = \{\mathcal{N}_x \mid x \in X\}$ se llama sistema de vecindades de la topología de X .

Proposición 3.40 Sea X un espacio topológico. El sistema de vecindades \mathcal{N} de la topología de X satisface las siguientes condiciones:

- i) $V \in \mathcal{N}_x, V \subseteq U \Rightarrow U \in \mathcal{N}_x$.
- ii) $V_i \in \mathcal{N}_x, i \in I, I$ finito. $\Rightarrow \bigcap_{i \in I} V_i \in \mathcal{N}_x$.
- iii) $V \in \mathcal{N}_x \Rightarrow x \in V$.
- iv) $U \in \mathcal{N}_x \Rightarrow \exists V \in \mathcal{N}_x$ tal que $U \in \mathcal{N}_y \forall y \in V$.

DEMOSTRACIÓN: i) Sea A abierto, tal que $x \in A \subseteq V \subseteq U$, entonces $x \in A \subseteq U$, A abierto. Por lo tanto $U \in \mathcal{N}_x$.

ii) Si $V_i \in \mathcal{N}_x \Rightarrow \exists A_i$, abierto, tal que $x \in A_i \subseteq V_i \Rightarrow x \in \bigcap_{i \in I} A_i \subseteq \bigcap_{i \in I} V_i, \forall i \in I \Rightarrow x \in \bigcap_{i \in I} A_i \subseteq \bigcap_{i \in I} V_i$, con $\bigcap_{i \in I} A_i$ abierto, por lo tanto $\bigcap_{i \in I} V_i \in \mathcal{N}_x$

iii) Por definición de vecindad $x \in V$.

iv) Sea V abierto tal que $x \in V \subseteq U$, entonces, para todo $y \in V, y \in V \subseteq U$. por lo tanto $U \in \mathcal{N}_y \forall y \in V$.

■

Definición 3.41 (sistema de vecindades) Sea X un conjunto, y para cada $x \in X$ sea \mathcal{N}_x una familia que satisface las condiciones i)-iv) de la proposición anterior. A la colección $\mathcal{N} = \{\mathcal{N}_x \mid x \in X\}$ la llamaremos un sistema de vecindades en X .

Teorema 3.42 Sea X un conjunto y \mathcal{N} un sistema de vecindades de X . Entonces existe una única topología en X que tiene precisamente a \mathcal{N} como su sistema de vecindades.

DEMOSTRACIÓN: Veremos primero que si tal topología existe entonces es única. El teorema 3.37 afirma que un conjunto es abierto si, y sólo si, es vecindad de todos sus puntos. Esto nos permite definir también la topología. Sea $\mathcal{A} = \{A \subseteq X \mid A \in \mathcal{N}_x \forall x \in A\}$. Por construcción \mathcal{A} sería la única topología que determina a estas vecindades. Veamos que efectivamente \mathcal{A} es una topología. $\emptyset \in \mathcal{A}$ por vacuidad, y $X \in \mathcal{A}$ por la hipótesis i). Sea

$\{A_i\}_{i \in I} \subseteq \mathcal{A}$ una colección no vacía de elementos en \mathcal{A} . Si $x \in \bigcup_{i \in I} A_i$, entonces $x \in A_i$ para alguna $i \in I$. $A_i \in \mathcal{N}_x$. Por i), $\bigcup_{i \in I} A_i \in \mathcal{N}_x$, pues $A_i \subseteq \bigcup_{i \in I} A_i$, por lo tanto, $\bigcup_{i \in I} A_i \in \mathcal{A}$. Sea $\{A_i\}_{i \in I} \subseteq \mathcal{A}$, I finito no vacío. Si $x \in \bigcap_{i \in I} A_i$ entonces $x \in A_i$ para toda $i \in I$, por lo tanto $A_i \in \mathcal{N}_x$ para toda $i \in I$. De ii), se sigue que $\bigcap_{i \in I} A_i \in \mathcal{N}_x$, por lo que $\bigcap_{i \in I} A_i \in \mathcal{A}$.

Así tenemos que \mathcal{A} es una topología en X . Falta estudiar cuáles son las vecindades en esta topología. Basta ver que las vecindades en esta topología son precisamente las del sistema de vecindades dado. Sea V una vecindad de x según \mathcal{A} , entonces existe $A \in \mathcal{A}$ tal que $x \in A \subseteq V$. Por definición de \mathcal{A} , $A \in \mathcal{N}_x$, y por i), $V \in \mathcal{N}_x$. Inversamente, supongamos ahora que $U \in \mathcal{N}_x$. Sea $A = \{y \in X \mid U \in \mathcal{N}_y\}$. En particular, $x \in A$ y, por iii), $y \in U$ para todo $y \in A$. Así, $x \in A \subseteq U$. Basta verificar que $A \in \mathcal{A}$.

Sea $y \in A$. Ya que $U \in \mathcal{N}_x$, por definición de A y por iv) sabemos que existe $V \in \mathcal{N}_y \ni U \in \mathcal{N}_z \forall z \in V$. Así, $V \subseteq A$ y por i) A es vecindad de y . Por lo tanto $A \in \mathcal{A}$

■

3.2.3 Continuidad de funciones

Definición 3.43 (Función continua en $x \in (X, \mathcal{T})$) Sean X y Y espacios topológicos. Sea $f : X \rightarrow Y$. La función f es continua en $x \in X$ si dada una vecindad V de $f(x)$, $f^{-1}(V)$ es una vecindad de x . Se dice que f es continua en X , si es continua en cada punto de X .

Proposición 3.44 Sean X y Y espacios topológicos, $f : X \rightarrow Y$ es continua en $x \in X$ si y sólo si dada una vecindad V de $f(x)$, existe una vecindad U de x tal que $f(U) \subset V$.

DEMOSTRACIÓN: Sea $x \in X$ y sea V vecindad de $f(x)$.

\Rightarrow) Como f es continua entonces $U = f^{-1}(V)$ es vecindad de x y puesto que $f(f^{-1}(V)) \subset V$, se tiene que $f(U) \subset V$.

\Leftarrow) Hay que demostrar que existe un abierto A tal que $x \in A \subseteq f^{-1}(V)$. Por hipótesis existe U vecindad de x , tal que $f(U) \subset V$, entonces $\forall u \in U, f(u) \in V \Rightarrow U \subseteq \{x \in X \mid f(x) \in V\} = f^{-1}(V) \therefore U \subseteq f^{-1}(V)$. Además como U es vecindad de x , existe A abierto tal que $x \in A \subseteq U \subseteq f^{-1}(V)$. Por lo tanto $f^{-1}(V)$ es vecindad de x . En conclusión f es continua en x . ■

Teorema 3.45 Sean X y Y espacios topológicos, sea $f : X \rightarrow Y$: f es continua en X si y sólo si para todo abierto B en Y , $f^{-1}(B)$ es abierto en X .

DEMOSTRACIÓN:

\Rightarrow) Sea B abierto en Y , y sea $x \in f^{-1}(B)$. Entonces $f(x) \in B$, y por ser B abierto, es vecindad de $f(x)$. Así pues, como f es continua en x , por ser continua en X , entonces $f^{-1}(B)$ es vecindad de x y como x fue arbitrario, entonces $f^{-1}(B)$ es vecindad de todos sus puntos. Por lo tanto $f^{-1}(B)$ es abierto.

\Leftarrow) Sea $x \in X$ y sea V vecindad de $f(x)$, entonces existe A , abierto en Y , tal que $f(x) \in A \subseteq V$. Como A es abierto en Y , entonces $f^{-1}(A)$ es abierto en X . Además $f^{-1}(A) = \{x \in X \mid f(x) \in A\} \subseteq \{x \in X \mid f(x) \in V\} = f^{-1}(V)$ $\therefore f^{-1}(A) \subseteq f^{-1}(V)$ y dado que $f(x) \in A$, entonces $x \in f^{-1}(A) \subseteq f^{-1}(V)$ y $f^{-1}(A)$ es abierto. Por lo tanto $f^{-1}(V)$ es vecindad de x . Puesto que x fue arbitraria en X , entonces f es continua en X .

■

El siguiente capítulo nos muestra el modelo en sí; la primer parte del capítulo nos habla del planteamiento del modelo y la segunda de su justificación. Ahí entrarán en uso las nociones matemáticas que vimos en el capítulo presente y en el anterior.

Capítulo 4

El modelo y su justificación

Ultimamente se debate en el área de la evolución si la historia de la vida ha sido punteada, a través de transiciones discontinuas en los fenotipos, o bien si las especies han evolucionado de manera gradual y continua.

El modelo que se presenta en este capítulo pretende hacer comprensible y explicar cómo surge en la evolución el concepto de transición discontinua.

Para distinguir entre un cambio continuo y uno discontinuo en la evolución, se necesita de una relación de cercanía entre fenotipos. Esa relación está basada en la probabilidad de que un fenotipo sea accesible desde otro, mediante cambios en el genotipo.

Un gene es la unidad de la herencia. Cada gene es una secuencia de ácidos nucléicos que contiene la información necesaria para formar un precursor de una proteína particular. Los genes pueden sufrir cambios. Dichos cambios se llaman **mutaciones**.

Existen mutaciones que no trascienden al nivel del fenotipo, se llaman **mutaciones silenciosas**. Dentro de ese grupo se encuentran las que producen un cambio en la secuencia de los aminoácidos que codifica el gene alterado, pero que no afectan la actividad de la proteína correspondiente. Este tipo de mutaciones recibe el nombre de mutaciones o **sustituciones neutrales**.

Gracias a las mutaciones neutrales, los fenotipos denotan clases de equivalencia de genotipos. Estas clases de equivalencia permiten establecer relaciones entre fenotipos a fin de definir la vecindad, o conjunto de formas cercanas, de una estructura secundaria de mínima energía, dada.

Puesto que el RNA provee secuencias (genotipo) y estructuras secundarias (fenotipo) en una sola molécula, el espacio de secuencias considerado es el

espacio. S_q , de todas las secuencias de RNA de longitud l , fija. Junto con este espacio se considera el espacio. S_h , de todas las estructuras secundarias de mínima energía asociadas a las secuencias del espacio S_q .

Dentro de estos espacios se estudia la manera en que el plegamiento de secuencias induce una topología en el conjunto S_h .

La relación de cercanía obtenida con dicha topología, es decir las vecindades, sugiere una noción de transformación continua. Dada esa relación de cercanía es posible explicar las transiciones grandes en la trayectoria de la evolución de dos fenotipos, al nivel de estructuras de RNA, si se identifican aquellas transformaciones que son discontinuas irreducibles.

En la primera sección de este capítulo se explica el modelo de manera detallada. En la segunda sección, se da la justificación del modelo.

4.1 El modelo

En esta sección se describe el modelo sin justificar los resultados enunciados, ellos se demuestran en la siguiente sección.

4.1.1 Premisas del modelo

Para el desarrollo del modelo se consideran secuencias de RNA de longitud l , fija, pues como ya se ha dicho, proveen genotipo y fenotipo en una sola molécula, lo cual lo hace ideal para experimentos sobre evolución. El fenotipo de una secuencia está dado por su estructura secundaria de mínima energía.

Para fines de simplicidad las estructuras secundarias de mínima energía recibirán el nombre de formas.

Puesto que el propósito es estudiar la evolución entre dos fenotipos y no entre sus productos finales, se define una estructura secundaria α , fija, como blanco del proceso evolutivo y se estudia cómo una forma cualquiera evoluciona dentro de una población bajo selección hacia el blanco fijado. Cabe mencionar que la evolución de una entidad está considerada aquí en su expresión más simple: la replicación en un ambiente constante.

Se define S_q como el conjunto de todas las secuencias de RNA de longitud l . Así mismo, se define S_h como el conjunto de todas las formas asociadas a las secuencias del conjunto S_q .

Con estos dos conjuntos se define la función $F : S_q \rightarrow S_h$ como la función de plegamiento dada por la acción del programa de Walter Fontana y Peter

Schuster para calcular la estructura secundaria de mínima energía de cada secuencia en S_q . Esta función de plegamiento es la misma que se definió en el capítulo uno.

Se considera, además, que la única fuente de variación que existe en los espacios, está dada por las mutaciones puntuales de las secuencias de S_q . Es decir, que la fuerza que obliga a la evolución es la que imponen las mutaciones puntuales. De hecho, en la naturaleza este tipo de mutaciones es el más común.

4.1.2 Planteamiento y desarrollo

A fin de entender la historia en la evolución de dos fenotipos, se debe organizar el conjunto S_h en una forma tal que refleje la posibilidad de que un fenotipo en S_h sea alcanzado mediante otro a través de cambios en el genotipo. Tal organización está dada por la asignación de un sistema de vecindades para cada entidad en el conjunto.

En el problema que estudiamos hay dos clases de entidades: las secuencias y las formas, las cuales están relacionadas mediante un procedimiento termodinámico de plegamiento, que en nuestro caso está representado por F .

El conjunto S_q está organizado de manera natural en un espacio de vecindades, puesto que las mutaciones puntuales inducen vecindades canónicas. Las **mutaciones puntuales** son aquellas que sólo intercambian una base nitrogenada por otra en un único punto dentro de la secuencia.

La **vecindad canónica** de una secuencia $a \in S_q$ consiste de todas las secuencias que difieren de ella en una base nitrogenada. Esta vecindad se puede interpretar en términos de una métrica particular, la de Hamming. $B_1^{d_H}(a) = \{b \in S_q \mid d_H^{S_q}(a, b) = 1\}$ representa la vecindad canónica de la secuencia $a \in S_q$. La métrica de Hamming nos sirve en situaciones en las que la longitud de la secuencia se conserva y en que las mutaciones puntuales son la única fuente de variación, lo cual se cumple en S_q , pues la longitud de las secuencias es fija y por hipótesis sólo se admiten mutaciones puntuales.

Organizar las formas en un espacio de vecindades es algo problemático, pues en contraste con las secuencias no es sencillo encontrar una métrica que refleje las diferencia entre una forma y otra. Esto se debe a que desde el punto de vista de la evolución no hay procesos físicos que modifiquen la estructura secundaria de una secuencia de manera directa y hereditable. De hecho, la transformación de una forma a otra es el resultado de una serie de cambios en la secuencia de la que proviene.

Así pues, es necesario que la organización de las formas esté relacionada con las secuencias de las que resultan.

Una primera asociación entre secuencias y formas está dada por la función de plegamiento y las mutaciones neutrales, las cuales permiten organizar las secuencias en clases de equivalencia determinadas por las estructuras secundarias.

La relación de equivalencia que rige las secuencias en términos de sus formas está dada de la siguiente manera. Sean α y $\beta \in S_h$ y sean a y $b \in S_q \ni F(a) = \alpha$ y $F(b) = \beta$, entonces $a \equiv b \Leftrightarrow \alpha = \beta$. Esta relación nos indica que dos secuencias están en la misma clase de equivalencia si y sólo si, dan lugar a la misma forma.

Dada una forma $\alpha \in S_h$ y una secuencia, a , que se pliega en α , se define la clase de equivalencia de a , como $[a] = \{s \in S_q \mid s \equiv a\} = \{\text{todas las secuencias que se pliegan en } \alpha\}$.

Esta relación refleja que algunas estructuras secundarias son más frecuentes que otras. Es decir, si el tamaño de la clase de equivalencia de una secuencia es grande, ello se interpreta como que la estructura asociada a dicha clase es frecuente. Una definición de frecuente está dada por la noción de común: Una estructura es **común** cuando más secuencias que el promedio $\frac{\# \text{ total de secuencias}}{\# \text{ total de estructuras}}$ se pliegan en ella.

La propiedad de ser frecuente se hace más precisa al considerar que en el límite de secuencias de longitud muy grande, la fracción de tales estructuras tiende a cero, mientras que la fracción de secuencias que se pliegan en ellas tiende a uno. Es decir, $EF = \{\alpha_i\}_{i \in I}$ es el conjunto de estructuras frecuentes si

$$\lim_{l \rightarrow \infty} \frac{\#EF}{\# \text{ total de estructuras}} = 0 \text{ y } \lim_{l \rightarrow \infty} \frac{\# \text{ secuencias}_{\alpha_i}}{\# \text{ total de secuencias}} = 1$$

donde $\# \text{ secuencias}_{\alpha_i}$ es el número de secuencias que se pliegan en α_i , $i \in I$. Cabe mencionar que el número de estructuras posibles se incrementa al incrementar l , la longitud de las secuencias, es decir, el número de estructuras es una función de l . Lo mismo se cumple para el número de secuencias.

Para hablar de vecindades entre las estructuras secundarias, es necesario definir en qué términos una forma, o estructura secundaria, puede ser cercana a otra. para ello es indispensable definir con anterioridad el concepto de accesibilidad.

Formas accesibles

Si en un proceso de evolución una forma α es seguida de una forma β , entonces β debe ser accesible desde α . El concepto de accesibilidad significa que una secuencia cuya forma es β puede ser alcanzada por una mutación puntual desde otra secuencia cuya forma es α . Nótese que esta noción de accesibilidad no cuantifica distancia, por lo que el sistema de vecindades que obtengamos, o topología, será más débil que el espacio de secuencias, en el sentido de que no podremos cuantificar la distancia entre dos estructuras.

En términos formales el concepto de accesibilidad se define de la siguiente forma: Se dice que β es accesible desde α , o $\beta \leftarrow \alpha$, si existe un par $a, b \in S_q \ni d_H(a, b) = 1$ y $f(a) = \alpha, f(b) = \beta$.

El conjunto de todas las estructuras accesibles desde α se describe como $\Sigma_\alpha = \{\beta \mid \beta \leftarrow \alpha\}$.

Dado el concepto de accesibilidad, se dice que una forma β es cercana a una forma α si es muy probable que β sea accesible desde α .

El problema se transforma entonces en estimar la frecuencia con la que una mutación en la secuencia de α da lugar a la forma β . Cabe aclarar que si una forma α está asociada a una clase de equivalencia muy grande entre secuencias, entonces la cercanía de β a α significa que α dará lugar a β con probabilidad alta cuando se promedia sobre todas las secuencias que se pliegan en α . Esto es, sea $A = \{s \in B_1(a) \mid s \in [b]\}$ con $F(a) = \alpha$ y $F(b) = \beta$, si $\frac{\#A}{\#[a]} \approx 1$, para cualquier a que se pliegue en α , entonces α es cercana a β .

Así pues, definir la vecindad de α en términos de sus elementos cercanos no depende de una secuencia particular y la vecindad se convierte en una propiedad robusta de α .

Sea $S_\alpha = \{s \in S_q \mid s \text{ se pliega en } \alpha\}$, S_α denota el mismo conjunto que $[a]$, pero sin hacer referencia a alguna a particular. El conjunto S_α tiene por frontera al conjunto $B_\alpha \subset S_q$, donde $B_\alpha = \{x \in S_q \mid d_H(x, s) = 1, s \in [a]\} = \cup_{s \in [a]} B_1^H(s)$.

Debido a que no hay recursos disponibles para identificar totalmente los conjuntos S_α , $\alpha \in S_h$, ellos se encuentran por muestreo.

Para encontrar S_α , se generan por plegamiento inverso un conjunto de n secuencias de longitud l que tengan a α como su estructura secundaria de mínima energía. Entre mayor sea n , mejor será la aproximación a S_α .

El **plegamiento inverso** es el proceso que nos permite saber, a partir de las características de una estructura secundaria dada, qué secuencias se

pliegan en ella. En [26] y [27] se pueden encontrar programas que realizan dicho procedimiento.

Recordemos que una secuencia de RNA está formada por una sucesión de 4 posibles bases nitrogenadas, así pues, por cada base puede haber tres mutaciones puntuales distintas. Debido a que una mutación puntual puede ocurrir en cualquiera de las l bases nitrogenadas que conforman una secuencia, entonces cada secuencia de la muestra de S_α tiene $3l$ vecinos. Sea $b \in S_q$, b es vecino de a si $d_H(b, a) = 1$. De esta forma, el conjunto $B_1^H(a)$ tiene cardinalidad $3l$, para cualquier $a \in S_q$, en particular para cualquier secuencia en S_α , y en consecuencia la muestra de el conjunto B_α , generada a partir de la muestra de S_α , tiene cardinalidad $3ln$.

Una vez que se encontraron los $3l$ vecinos de cada secuencia de la muestra de S_α , se obtiene la forma asociada a cada una de esas $3ln$ secuencias. Estas estructuras constituyen una muestra de Σ_α .

Cada estructura β en Σ_α tiene dos multiplicidades asociadas. Una de ellas, $N(\beta, \alpha)$, cuenta el número total de vecindades $B_1^H(s)$, $s \in S_\alpha$, en que la estructura β ocurre por lo menos una vez. Se dice que β ocurre en la vecindad $B_1^H(s)$, si existe $k \in B_1^H(s) \ni k \in S_\beta$, es decir, si hay una secuencia k en la vecindad de s , con s una secuencia que se pliega en α , tal que se pliega en β .

Si se normaliza $N(\beta, \alpha)$ por el tamaño, $M_\alpha = n$, de S_α , se obtiene la frecuencia de vecindades

$$\nu(\beta, \alpha) = \frac{N(\beta, \alpha)}{M_\alpha}$$

Esta frecuencia refleja la probabilidad de que una secuencia k se pliegue en β , donde k es una secuencia generada a partir de una mutación puntual de una secuencia s y s es cualquier secuencia que se pliega en α .

La otra multiplicidad se refiere al número total de ocurrencias, $N_t(\beta, \alpha)$, en B_α de la estructura β . Es decir, refleja el número total de secuencias de B_α que se pliegan en β . Así, cada vecindad de las secuencias en S_α , respectivamente, está ponderada por el número de veces que β ocurre en esa vecindad.

Si se normaliza $N_t(\beta, \alpha)$ por $3lM_\alpha$, es decir, por la cardinalidad de B_α , obtenemos la frecuencia de ocurrencia

$$v(\beta, \alpha) = \frac{N_t(\beta, \alpha)}{3lM_\alpha}$$

Esta frecuencia nos indica la probabilidad de que una secuencia k , entre todos los vecinos posibles de secuencias que se pliegan en α , se pliegue en β .

Estos índices se estiman a partir de la muestra de Σ_α . Dichos números indican lo mismo de distintas maneras, una contando vecindades, otra contando secuencias.

Ahora bien, para definir propiamente las vecindades en el espacio de formas, definimos conjuntos que jugarán un papel similar al de las ϵ – *bolas* del espacio de secuencias.

Sea $\alpha \in S_h$, consideremos los conjuntos

$$\Psi_\epsilon(\alpha) = \{\beta \in \Sigma_\alpha \mid \rho(\beta, \alpha) \geq \epsilon\}$$

donde $1 \geq \epsilon > 0$ y $0 \leq \rho(\beta, \alpha) \leq 1$. ρ denota una medida para la frecuencia de β en la frontera de S_α . Esta frecuencia puede ser $\nu(\beta, \alpha)$ o $v(\beta, \alpha)$ definidas anteriormente. Para el trabajo se considera $\rho = \nu(\beta, \alpha)$ [13, p. 19].

Es fácil ver que si $\epsilon_1 > \epsilon_2$ entonces $\Psi_{\epsilon_1}(\alpha) \subset \Psi_{\epsilon_2}(\alpha)$.

La topología de S_h depende de las formas que contiene $\Psi_\epsilon(\alpha)$, para cualquier $\alpha \in S_h$.

El contenido de estos conjuntos $\Psi_\epsilon(\alpha)$ no es arbitrario, más bien depende del plegamiento del RNA, en este caso depende del algoritmo de Schuster y Fontana. Dicho contenido debe ser obtenido mediante análisis numéricos de $\rho(\beta, \alpha)$. Los análisis realizados revelaron que no todas las estructuras son igualmente importantes, por lo que la búsqueda del contenido de $\Psi_\epsilon(\alpha)$ se limita a aquellas estructuras que son frecuentes [13, 13-18].

Cercanía

Simulaciones hechas por Walter Fontana y Peter Schuster, autores del modelo, revelan que $\exists 1 \geq \delta_\alpha > 0$, que depende de α , tal que δ_α separa el conjunto característico de estructuras β altamente frecuentes en las que α se puede plegar, de las estructuras menos frecuentes. Esto permite describir la estructura topológica del espacio de formas considerando para cada forma α , sólo aquellas estructuras accesibles desde α con una frecuencia de al menos δ_α .

Sea $\Psi_{\delta_\alpha}(\alpha) = \{\beta \in \Sigma_\alpha \mid \rho(\beta, \alpha) \geq \delta_\alpha\}$. $\Psi_{\delta_\alpha}(\alpha)$ es llamado el conjunto característico de α y contiene aquellas estructuras que son accesibles desde α con una probabilidad alta.

Se dice que una forma β es cercana a α si $\beta \in \Psi_{\delta_\alpha}(\alpha)$, es decir, si β es accesible desde α con una probabilidad de al menos δ_α .

Sea $C_\alpha = \{\beta \in S_h \mid \beta \in \Psi_{\delta_\alpha}(\alpha)\}$. C_α es el conjunto de todos los vecinos de α , es decir, todas las formas cercanas a α .

El conjunto de vecindades de α está determinado por $N_\alpha = \{V \subset S_h \mid \exists \epsilon > 0 \ni \delta_\alpha \geq \epsilon \text{ y } \Psi_\epsilon(\alpha) \subset V\}$ y $\mathcal{V} = \{N_\alpha\}_{\alpha \in S_h}$ determina la topología para S_h .

Ahora bien, para hablar de continuidad en la evolución es necesario caracterizar los cambios que hay de una estructura secundaria a otra para saber si es posible lograr ese cambio en un sólo paso o no.

Continuidad

Se dice que dos estructuras secundarias son **comparables** si se puede llegar de una a la otra ya sea removiendo o añadiendo pares de bases. Esta condición está sujeta al hecho de que no se permiten *pseudo-nudos* en las estructuras secundarias.

Por el contrario, dos estructuras son **incomparables** si la interconversión, es decir la conversión de una estructura en la otra, implica ambos actos, esto es, si se necesita añadir y remover pares de bases.

En este último caso, es útil distinguir aquellas interconversiones que son **cambios generalizados**.

Se dice que un cambio de pares de bases es generalizado, si por cada par de bases cambiado, al menos una de las bases se mantiene emparejada.

Esta definición incluye los cambios estándar en los cuales las parejas se recorren en unas cuantas bases. Un cambio generalizado también abarca aquellas transformaciones como el "roll-over", los giros y los dobles giros. Como ejemplo de todos estos cambios se presenta la figura 4.1.

Aquellas estructuras que difieren en cambios generalizados se llaman **shift-incompatibles**.

Por definición de cercanía, β es cercana a α si podemos pasar de α a β en un solo paso, con probabilidad mayor o igual a δ_α .

Hay transiciones de α a β que no son en un solo paso, sino que pueden ocurrir en varias etapas. Para estas transiciones decimos que si β no es cercana a α pero existe una serie de estructuras $\alpha \equiv \alpha_0, \alpha_1, \dots, \alpha_{i-1}, \alpha_i, \dots, \alpha_n \equiv \beta \ni \alpha_i$ sea cercana a α_{i-1} , entonces la transición completa de α a β es **continua**. Claramente, si dos estructuras son cercanas, entonces la transición de α a β es continua.

Con esta definición de continuidad ya se puede hablar claramente de si una sucesión evolutiva de α a β , es decir la sucesión de formas producidas

durante la evolución de α a β , es continua. La definición es exactamente igual. La evolución de una estructura α a una β es continua, si existe una sucesión de estructuras $\alpha \equiv \alpha_0, \alpha_1, \dots, \alpha_{i-1}, \alpha_i, \dots, \alpha_n \equiv \beta$, $\ni \alpha_i$ es cercana a $\alpha_{i-1} \forall i = 1, \dots, n$. Una transición, o proceso evolutivo, de una estructura a otra es discontinuo irreducible, si no existe tal sucesión de estructuras (figura 4.2).

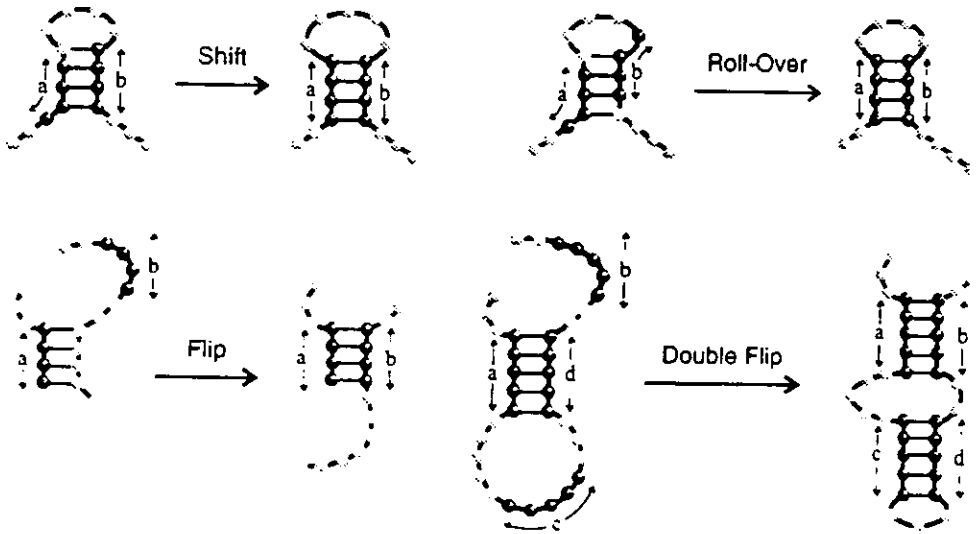


Figura 4.1: Cambios generalizados.

4.1.3 Conclusiones

Entre las conclusiones encontradas por Schuster y Fontana a través de sus análisis numéricos se encuentra el hecho de que la cercanía entre dos estructuras no implica la similitud de las mismas. Más aún, la cercanía no es una relación simétrica. Esto queda ilustrado con la figura 4.3.

Una manera de visualizar la estructura del espacio S_h es a través de una gráfica dirigida. Cada forma en el espacio se representa mediante un nodo. Se asigna un segmento dirigido del nodo α al nodo β si β está en el conjunto característico, C_α , de α , ver la figura 4.4.

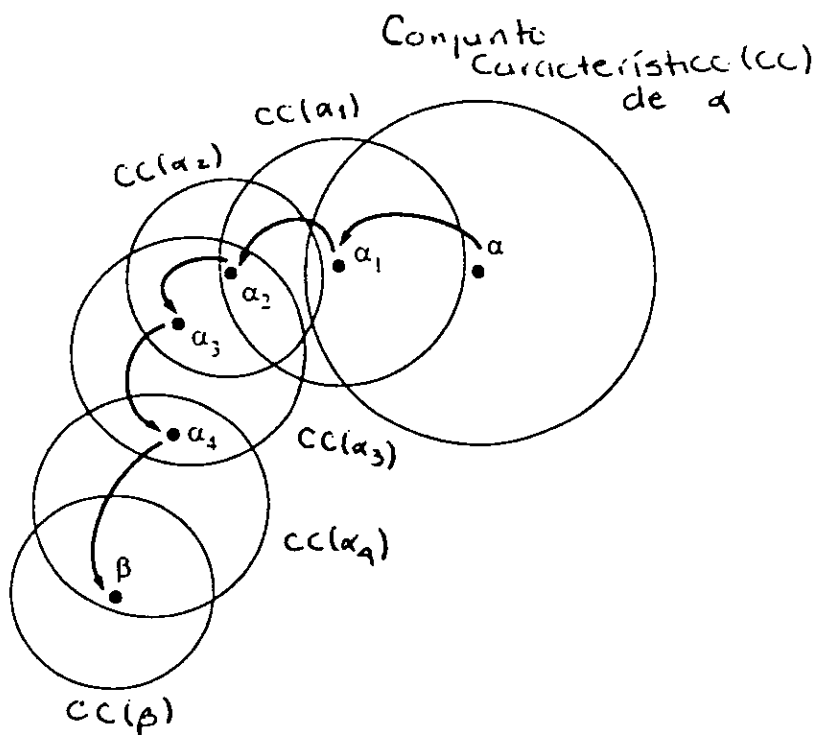


Figura 4.2: Un proceso continuo de evolución

Con esta representación de S_h , se puede pensar en una transformación continua de α a β como un camino dirigido de α a β en la gráfica de S_h .

Una transformación discontinua se puede interpretar como una transición entre nodos de la gráfica para los cuales no existe un paseo dirigido.

Otra conclusión del trabajo de Fontana y Schuster es que las transiciones discontinuas son precedidas por largos periodos de deriva génica. Esto es debido a que una transición discontinua puede ser obtenida con una mutación puntual, sólo si el resto de la secuencia provee el contexto necesario para que ello se logre. Estas secuencias son muy raras y están forzadas severamente

por las condiciones del ambiente. Además cuando un fenotipo está bajo una selección muy fuerte, la deriva génica es el único medio que se tiene para producir el contexto genotípico requerido.

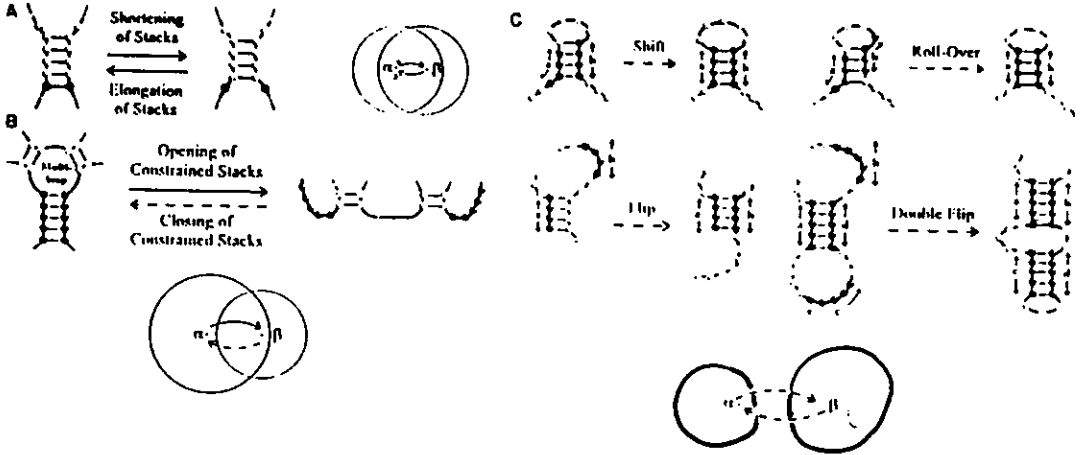


Figura 4.3: A. Acortar o alargar una zona de bases emparejadas es una transición continua. B. Abrir una región de bases emparejadas es un movimiento continuo, mientras que emparejar una secuencia de bases no lo es. C. Los cambios generalizados son transiciones discontinuas.

En la figura 4.5 se muestra una estructura secundaria y su vecindad. Con este ejemplo se puede observar que la cercanía entre formas no implica la similitud de las mismas.

La siguiente sección está destinada a formalizar y demostrar lo que aquí se menciona. La parte principal de dicha sección consiste en demostrar que la familia $\mathcal{V} = \{N_\alpha\}_{\alpha \in S_h}$, donde $N_\alpha = \{V \subset S_h \mid \exists \epsilon > 0 \exists \delta_\alpha \geq \epsilon \text{ y } \Psi_\epsilon(\alpha) \subset V\}$,

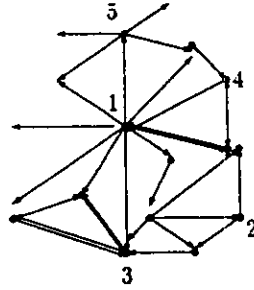


Figura 4.4: Ejemplo de la representación como digráfica de la topología para S_h . El proceso de evolución de 1 a 2 es continuo en ambos sentidos. La evolución de 3 a 1 es discontinua.

es una topología para S_h .

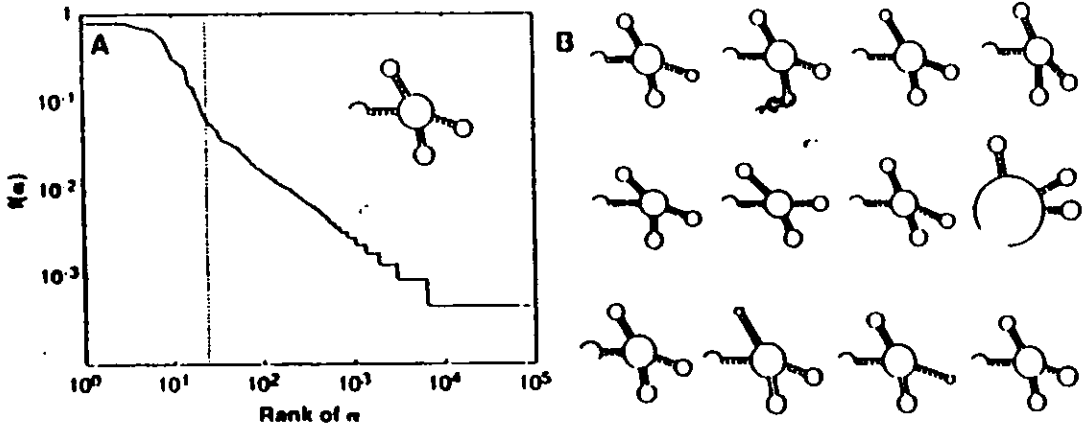


Figura 4.5: Ejemplo de una vecindad

4.2 Justificación

Como ya se mencionó, esta sección está dedicada a la formalización del modelo. Aquí se demuestra que el espacio (S_q, H) es un espacio métrico. H es una métrica de Hamming análoga a la definida en 3.7 del capítulo 3. Se demuestra también que (S_h, \mathcal{V}) es un espacio topológico. $\mathcal{V} = \{N_\alpha\}_{\alpha \in S_h}$. Para esto se utiliza el teorema 3.42 del capítulo 3. Finalmente, se muestra que la función de plegamiento $F : S_q \rightarrow S_h$ es una función continua casi en ninguna parte. Es decir, sólo bajo ciertas condiciones, la imagen inversa de una vecindad en S_h es una vecindad en S_q . Esto expresa que en general la evolución no es continua sino a brinquitos.

4.2.1 El espacio de secuencias S_q

Como ya se definió anteriormente, el espacio S_q es el espacio de todas las secuencias de RNA de longitud l , fija.

En el capítulo tres definimos la métrica de Hamming para el espacio de secuencias de ceros y unos. Dicha métrica cuenta el número de lugares en que dos secuencias difieren. la métrica de Hamming se puede generalizar a otros espacios.

Consideremos $d_i : S_q \times S_q \rightarrow \{0, 1\}$ tal que

$$d_i(a, b) = \begin{cases} 0 & a_i = b_i \\ 1 & a_i \neq b_i \end{cases} \quad \text{Para } i = 1, \dots, l$$

Sea $d_H(a, b) = \sum_{i=1}^l d_i(a, b)$, d_H es la generalización de la métrica de Hamming para el espacio S_q . d_H también cuenta el número de lugares en que la secuencia a y la b difieren.

Para ver que el espacio (S_q, d_H) es un espacio métrico, basta probar que d_H es una métrica.

Lema 4.1 *Para todo $i = 1, \dots, l$, d_i satisface que $d_i(a, b) \leq d_i(a, c) + d_i(c, b)$ para $a, b, c \in S_q$.*

DEMOSTRACIÓN: Sean $a, b, c \in S_q$. Si $d_i(a, c) = 1$ (o $d_i(c, b) = 1$) entonces, $d_i(a, b) \leq 1 \leq 1 + d_i(c, b) = d_i(a, c) + d_i(c, b) \therefore d_i(a, b) \leq d_i(a, c) + d_i(c, b)$ (análogamente si $d_i(c, b) = 1$).

Si $d_i(a, c) = d_i(c, b) = 0 \Rightarrow a_i = c_i = b_i \Rightarrow a_i = b_i \Rightarrow d_i(a, b) = 0 \Rightarrow d_i(a, b) \leq d_i(a, c) + d_i(c, b)$ (de hecho, ambas partes son iguales a cero).

Por lo tanto, $d_i(a, b) \leq d_i(a, c) + d_i(c, b) \forall i = 1, \dots, l$. ■

(S_q, d_H) es un espacio métrico

Proposición 4.2 *La función d_H definida arriba es una métrica.*

DEMOSTRACIÓN: i) Por definición de d_i , $d_i(a, b) \geq 0 \forall i = 1, \dots, l \Rightarrow \sum_{i=1}^l d_i(a, b) \geq 0 \therefore d_H(a, b) \geq 0$ y $d_H(a, b) = 0 \Leftrightarrow \sum_{i=1}^l d_i(a, b) = 0 \Leftrightarrow d_i(a, b) = 0 \forall i = 1, \dots, l \Leftrightarrow a_i = b_i \forall i = 1, \dots, l \Leftrightarrow a = b$

ii)

$$d_i(a, b) = \begin{cases} 0 & a_i = b_i \\ 1 & a_i \neq b_i \end{cases} = \begin{cases} 0 & b_i = a_i \\ 1 & b_i \neq a_i \end{cases} = d_i(b, a) \forall i = 1, \dots, l$$

entonces $\sum_{i=1}^l d_i(a, b) = \sum_{i=1}^l d_i(b, a) \therefore d_H(a, b) = d_H(b, a)$

iii) Por el lema 4.1 tenemos que, $d_i(a, b) \leq d_i(a, c) + d_i(c, b)$ para toda $i = 1, \dots, l$, entonces $\sum_{i=1}^l d_i(a, b) \leq \sum_{i=1}^l d_i(a, c) + \sum_{i=1}^l d_i(c, b)$

$\therefore d_H(a, b) \leq d_H(a, c) + d_H(c, b)$

Por lo tanto, d_H es una métrica. ■

Con esto queda justificado que (S_q, d_H) es un espacio métrico.

4.2.2 El espacio de formas S_h

Dentro de la sección anterior, se definió el conjunto S_h como el conjunto de todas las estructuras secundarias de mínima energía que surgen de las secuencias de RNA del espacio S_q , respectivamente.

A cada α de S_h se le asignó su conjunto de vecindades $\mathcal{N}_\alpha = \{V \subset S_h \mid \exists \epsilon > 0 \ni \delta_\alpha \geq \epsilon \text{ y } \Psi_\epsilon(\alpha) \subset V\}$. Con este conjunto de vecindades, se definió el conjunto $\mathcal{V} = \{\mathcal{N}_\alpha \mid \alpha \in S_h\}$ y se afirmó que constituye una topología para S_h , lo que le daría a S_h su estructura de espacio topológico.

Para probar que \mathcal{V} constituye una topología para S_h , haremos uso de el teorema 3.42, que nos dice que sea \mathcal{N} un sistema de vecindades de X , en este caso de S_h , entonces existe una única topología en X que tiene precisamente a \mathcal{N} como su sistema de vecindades. Por lo que si se prueba que \mathcal{V} es un sistema de vecindades de S_h , habríamos terminado.

Lema 4.3 *Sea $\alpha \in S_h$. Si $\epsilon_1 > \epsilon_2$, entonces $\Psi_{\epsilon_1}(\alpha) \subset \Psi_{\epsilon_2}(\alpha)$.*

DEMOSTRACIÓN: Sea $\beta \in \Psi_{\epsilon_1}(\alpha) \Rightarrow \rho(\beta, \alpha) \geq \epsilon_1 > \epsilon_2 \Rightarrow \rho(\beta, \alpha) \geq \epsilon_2$
 $\Rightarrow \beta \in \Psi_{\epsilon_2}(\alpha) \therefore \Psi_{\epsilon_1} \subset \Psi_{\epsilon_2}$. ■

(S_h, \mathcal{V}) es un espacio topológico

Proposición 4.4 Sea $\alpha \in S_h$, el conjunto de vecindades de α , \mathcal{N}_α , satisface:

i) $V \in \mathcal{N}_\alpha, V \subseteq U \Rightarrow U \in \mathcal{N}_\alpha$.

ii) $V_i \in \mathcal{N}_\alpha, i \in I, I$ finito. $\Rightarrow \bigcap_{i \in I} V_i \in \mathcal{N}_\alpha$.

iii) $V \in \mathcal{N}_\alpha \Rightarrow \alpha \in V$.

iv) $U \in \mathcal{N}_\alpha \Rightarrow \exists V \in \mathcal{N}_\alpha$ tal que $U \in \mathcal{N}_\beta \forall \beta \in V$.

Es decir $\mathcal{V} = \{\mathcal{N}_\alpha\}_{\alpha \in S_h}$ es un sistema de vecindades de S_h .

DEMOSTRACIÓN: i) Si $V \in \mathcal{N}_\alpha, V \subseteq U \Rightarrow \exists \epsilon > 0$, con $\delta_\alpha \geq \epsilon \ni \Psi_\epsilon(\alpha) \subset V \subseteq U \Rightarrow \Psi_\epsilon(\alpha) \subset U \therefore U \in \mathcal{N}_\alpha$.

ii) $V_i \in \mathcal{N}_\alpha, i \in I$, finito \Rightarrow existen $\epsilon_i > 0, \delta_\alpha \geq \epsilon_i, i \in I \ni \Psi_{\epsilon_i} \subset V_i$.
 Sea $\epsilon = \max_{i \in I} \{\epsilon_i\} \Rightarrow \epsilon \geq \epsilon_i \forall i \in I$ entonces, por el lema 4.3 se tiene que $\Psi_\epsilon(\alpha) \subset \Psi_{\epsilon_i}(\alpha) \forall i \in I \Rightarrow \Psi_\epsilon(\alpha) \subset \bigcap_{i \in I} \Psi_{\epsilon_i}(\alpha) \subset \bigcap_{i \in I} V_i \Rightarrow \Psi_\epsilon(\alpha) \subset \bigcap_{i \in I} V_i \therefore \bigcap_{i \in I} V_i \in \mathcal{N}_\alpha$.

iii) $V \in \mathcal{N}_\alpha \Rightarrow \exists \epsilon > 0$, con $\delta_\alpha \geq \epsilon \ni \Psi_\epsilon(\alpha) \subset V, \alpha \in \Sigma_\alpha$ y $\nu(\alpha, \alpha) = \frac{N(\alpha, \alpha)}{M_\alpha} = \frac{M_\alpha}{M_\alpha} = 1 \geq \delta_\alpha \geq \epsilon \Rightarrow \nu(\alpha, \alpha) \geq \epsilon \Rightarrow \alpha \in \Psi_\epsilon(\alpha) \subset V \therefore \alpha \in V$.

iv) Sea $U \in \mathcal{N}_\alpha \Rightarrow \exists \epsilon > 0$, con $\delta_\alpha \geq \epsilon \ni \Psi_\epsilon(\alpha) \subset U$. Sea $A = \{\beta \in \Psi_\epsilon(\alpha) \mid \Psi_{\delta_\beta}(\beta) \subset \Psi_\epsilon(\alpha)\}, \alpha \in A$ ya que por el lema 4.3 se tiene que $\Psi_{\delta_\alpha}(\alpha) \subset \Psi_\epsilon(\alpha)$, pues $\delta_\alpha \geq \epsilon$. $\therefore A \neq \emptyset$. Sea $\epsilon_1 = \min_{\beta \in A} \{\delta_\beta\}$ y sea $V = \Psi_{\epsilon_1}(\alpha), \epsilon_1 \leq \delta_\alpha$, pues $\alpha \in A$, y $V \in \mathcal{N}_\alpha (\Psi_{\epsilon_1}(\alpha) \subseteq V = \Psi_{\epsilon_1}(\alpha))$, además, por el lema 4.3 $\Psi_{\epsilon_1}(\alpha) \subset \Psi_\epsilon(\alpha) \subset U$. De esto último, lo primero ocurre por definición de A y lo segundo por hipótesis. Basta ahora con demostrar que $U \in \mathcal{N}_\beta \forall \beta \in V$.

Sea $\beta \in V = \Psi_{\epsilon_1}(\alpha)$. P.D. $\exists \epsilon_2 > 0, \delta_\beta \geq \epsilon_2, \ni \Psi_{\epsilon_2}(\beta) \subset U$.

Sea $\epsilon_2 = \delta_\beta, \delta_\beta \geq \epsilon_1$ debido a que ϵ_1 es el mínimo sobre las δ_β . Entonces $\Psi_{\epsilon_2}(\beta) = \Psi_{\delta_\beta}(\beta) \subset \Psi_\epsilon(\alpha) \subset U \Rightarrow \Psi_{\epsilon_2}(\beta) \subset U$. $\therefore U \in \mathcal{N}_\beta$. Como β fue arbitraria en V , entonces $U \in \mathcal{N}_\beta \forall \beta \in V$. ■

Con esto queda justificado que S_h es un espacio topológico.

4.2.3 Sobre la continuidad de $F: S_q \rightarrow S_h$

Los abiertos de S_h son conjuntos V tales que existe $\epsilon > 0$, $\delta_\alpha \geq \epsilon \Rightarrow \Psi_\epsilon(\alpha) \subset V$, $\alpha \in S_h$. Esto debido a que \mathcal{V} es topología para S_h .

Los abiertos de S_q , viendo a S_q como espacio topológico (lo cual es posible gracias al teorema 3.29), son conjuntos U tales que existe $\epsilon > 0$ tal que $B_\epsilon(a) \subset U$, con $a \in S_q$.

Sea A abierto de S_h . $F^{-1}(A) = \{b \in S_q \mid F(b) \in A\}$. Sea $\alpha \in A \Rightarrow \exists \epsilon > 0$, con $\delta_\alpha \geq \epsilon \Rightarrow \Psi_\epsilon(\alpha) \subset A$. $F^{-1}(\Psi_\epsilon(\alpha)) \subset F^{-1}(A)$.

F sería continua si existiera un $\delta > 0$ tal que $B_\delta^{d_H}(a) \subset F^{-1}(A)$ con $a \ni F(a) = \alpha$. Este δ podría proceder de aquellas secuencias que bajo F van a dar a $\Psi_\epsilon(\alpha)$, es decir, provendría de $F^{-1}(\Psi_\epsilon(\alpha))$.

$$F^{-1}(\Psi_\epsilon(\alpha)) = \{b \in S_q \mid F(b) \in \Psi_\epsilon(\alpha)\} = \{b \in S_q \mid F(b) \in \Sigma_\alpha y \frac{N(F(b), \alpha)}{M_\alpha} \geq \epsilon\} = \{b \in S_q \mid d_H(b, a) = 1, \frac{N(F(b), \alpha)}{M_\alpha} \geq \epsilon\} \subseteq \{b \in S_q \mid d_H(b, a) = 1\} = B_\delta^{d_H}(a).$$

En el mejor de los casos, $F^{-1}(\Psi_\epsilon(\alpha))$ sería igual a $B_\delta^{d_H}(a)$, pero ello implicaría que todas las secuencias vecinas de a se pliegan en alguna $F(b) \in \Psi_\epsilon(\alpha)$, con probabilidad alta, lo cual casi nunca sucede.

Así pues, se observa que la función de plegamiento es continua casi en ninguna parte. Esto es debido a que las formas frecuentes producidas por mutaciones puntuales de una secuencia que se pliega en α , arbitraria, generalmente no pertenecen al conjunto característico, $\Psi_{\delta_\alpha}(\alpha)$, de α , y para aquellas que lo hacen, no siempre se cumple con una frecuencia alta [12, p. 1454].

Con esto se observa que los vecinos de una estructura secundaria no siempre provienen de secuencias vecinas, lo que indica que una forma que procede a otra en un proceso evolutivo, puede no provenir de una sola mutación puntual, con lo que los cambios en los fenotipos ocurren a "brinquitos", o dicho de otra manera, ocurre de forma punteada.

4.2.4 Mis conclusiones

Mi trabajo en el desarrollo de esta tesis consistió en formalizar el modelo de continuidad en la evolución propuesto por Walter Fontana y Peter Schuster.

Esa formalización consistió en demostrar que es posible asociar el espacio S_q con un espacio métrico, asignar al conjunto S_h una topología y analizar la continuidad de la función de plegamiento a partir de las características de la topología dada para S_h .

Al verificar que el modelo está bien fundamentado, concluyo que es una propuesta confiable de la cual se sigue que la evolución entre dos fenotipos es, la más de las veces, a "brinquitos", lo que indica la presencia de una evolución punteada. Cabe mencionar que debido a que la no continuidad de la función de plegamiento, y por ende la de un proceso de evolución analizado bajo este modelo, está fuertemente determinada por la topología asociada al espacio de estructuras secundarias de mínima energía, si ella se cambia el resultado obtenido no se puede asegurar. Esto no contradice el resultado al que se llega con el modelo planteado por Walter Fontana y Peter Schuster, sólo indica que en cuestión de evolución todavía no se puede esclarecer cabalmente qué medios producen los cambios de una especie a otra. Este modelo muestra la presencia de evolución punteada, lo que indica que el gradualismo no es único medio por el que la evolución se desarrolla.

En este capítulo se habló del modelo desde un punto de vista general. El capítulo siguiente plantea un caso particular al que se puede aplicar el modelo.

Capítulo 5

Un caso particular: HH

El modelo tal cual fue explicado puede ser utilizado para fijar como blanco final a una estructura secundaria particular, con el propósito de caracterizar su vecindad según el modelo indica y analizar si un proceso de evolución dirigido a esa estructura es continuo o discontinuo.

El caso particular que se tratará es el del Hammerhead.

5.1 Hammerhead: una ribozima con “Cabeza de Martillo”

Las ribozimas son secuencias de RNA con capacidad catalítica.

Existe un tipo de ribozima que por sus características especiales para buscar, reconocer, cortar y destruir moléculas de RNA, es conocida como ribozima Cabeza de Martillo (HH, por sus siglas en Inglés).

La ribozima Cabeza de Martillo fue descubierta comparando las secuencias de ciertos viroides y virusoides que presentaron autocatalización como parte de su proceso de replicación [15].

Las ribozimas HH tienen la característica de no contener información para la creación de proteínas, aun cuando ellas están formadas de RNA, pero en cambio actúan como enzimas que cortan moléculas de RNA. Una vez que las ribozimas HH encuentran y reconocen su blanco, la ribozima se pega a él y solamente a él, ignorando cualquier otro tipo de moléculas de RNA.

Hecho eso HH corta a su blanco en pedacitos que son frágiles e inestables. Estos pedazos inestables de RNA son reconocidos por la célula como basura y entonces son destruidos.

Debido a la habilidad de la ribozima HH de reconocer diferencias entre un par de bases de varias secuencias, la ribozima HH permite ser utilizada en terapias génicas.

5.2 La medicina y el HH

La terapia génica es la manipulación del material genético de las células a fin de combatir enfermedades. Este tipo de terapias han sido utilizadas para tratar enfermedades genéticas del sistema inmunológico tal como la deficiencia de ADA (Adenosin DeAminasa); esta deficiencia es una enfermedad congénita en la cual los bebés nacen sin la enzima ADA necesaria para la generación de sus propias células T y B, células de la sangre, lo que ocasiona que los bebés sean incapaces de combatir enfermedades y sufren de inmunodeficiencias graves.

Debido a la habilidad de la ribozima Cabeza de Martillo para reconocer sitios específicos y unirse a una molécula en tales lugares, así como su capacidad para cortar las moléculas a las que se une, la ribozima puede ser adaptada para reconocer ciertas moléculas de interés como virus patológicos, como el VIH (virus de inmunodeficiencia humana, causante del SIDA), o bien podría utilizarse para eliminar genes cancerígenos.

El VIH es, de hecho, un excelente candidato para combatirse a través de terapia génica ya que ésta funciona introduciendo un gene "guardián" en el material genético de las células del paciente, dando así a las células la capacidad de defenderse por sí mismas.

El VIH contiene muchos genes que potencialmente pueden ser blancos de la terapia génica. Además, siendo que el material genético del VIH está compuesto de RNA, la ribozima HH puede ser adaptada para reconocer de manera específica las moléculas de RNA del virus y después de ello cortarlas para que la célula las elimine.

Actualmente existen tres clínicas (dos en EUA y una en Australia) que están utilizando tratamientos génicos a base de la ribozima HH para estudiar, *in vivo*, los efectos del uso de la ribozima en contra del VIH. Los responsables de estos grupos de trabajo opinan que la ribozima HH es una molécula fascinante y que tiene un potencial increíble para ser utilizada en contra del VIH y otras enfermedades.

5.3 Aplicaciones y futuras líneas de investigación

Aun cuando el uso de la ribozima HH ha tenido una buena aceptación entre varias instituciones y se han enfocado a investigar su eficacia en tratamientos génicos contra el cáncer y sobre todo contra el VIH, hasta hace poco la mayoría de los experimentos con ribozimas sólo habían sido *in vitro*.

La notación "Cabeza de Martillo", atribuída a dicha ribozima no solo por sus características ya mencionadas sino también por la forma de su estructura secundaria, está basada en imágenes obtenidas a través de la cristalografía por rayos X [11], ver la figura 5.1; sin embargo *in vivo* algunas ribozimas HH tienen forma de "Y".

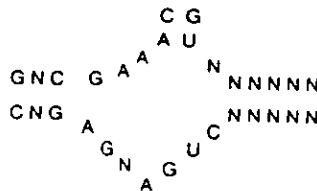


Figura 5.1: La estructura Hammerhead

Este pequeño detalle podría en algún momento desembocar en una perspectiva completamente diferente a la que se tiene ahora con respecto a la estructura, el mecanismo y la manera de "actuar" de la ribozima, especialmente dentro del medio intracelular.

Es debido a esto último que es importante caracterizar la vecindad, según el modelo, de la estructura Cabeza de Martillo pues con ello se puede saber qué tanta probabilidad hay de que una secuencia se pliegue en la forma HH y con ello qué tan sencillo sería encontrarla *in vivo*.

Una vecindad grande nos indicaría que la ocurrencia de esta estructura tiene varias alternativas, indicando una buena posibilidad de que los tratamientos en base a HH funcionen y más aún, si entre los vecinos de HH se encuentra estructuras encontradas comunmente en las células.

Así pues, el modelo representaría un acercamiento al estudio de la ribozima HH por medios alternativos, un modelo matemático.

Este proyecto queda para una investigación posterior, tema de una tesis de maestría.

Bibliografía

- [1] Hammerhead ribozymes: New engineered predators against HIV. Publicado en la red. www.actup99.org/BAR/art082897.html.
- [2] Molecular “fossils” of early life. Publicado en la red. www.eurekaalert.org/releases/gu-mfoel.html.
- [3] Origins of life and RNA world: Never underestimate the power of soup. Publicado en la red. www.postmodern.com/~jka/rnaworld/nfrna/nf-index.html.
- [4] Ribozyme report. The structure and function of the hammerhead ribozyme. Publicado en la red. attila.stevens-tech.edu/~rsamuel/index.html.
- [5] The RNA world. Publicado en la red. www.panspermia.org/rnaworld.html.
- [6] The RNA world: History of an idea, an idea of history. Publicado en la red. www.postmodern.com/~jka/rnaworld/rna/rnaworlded.html.
- [7] Scientist debate RNA’s role at beginning of life on earth. Publicado en la red. www.the-scientist.library.u...du/yr1997/mar/research-970331.html.
- [8] Ana Barahona and Daniel Piñero. *Genética: La continuidad de la vida*. Fondo de Cultura Económica, México, D.F.. 1994. La Ciencia desde México, 125.
- [9] Robert Bartle. *Introducción al Análisis Matemático*. Limusa, México, D.F. 1989.

- [10] Gary Chartrand. *Graphs as Mathematical Models*. Prindle, Weber & Schmidt, incorporated, Boston, 1978.
- [11] Bratty J et al. The Hammerhead RNA domain, a model ribozyme. *Biochim Biophys Acta*, 3:345-359, 1993.
- [12] Walter Fontana and Peter Schuster. "Continuity in evolution: On the nature of transitions". *Science*, 280:1451-1455, 29 de mayo 1998.
- [13] Walter Fontana and Peter Schuster. "Shaping space: The possible and the attainable in RNA genotype-phenotype mapping". *Journal of theoretical biology*, 194(4), 1998.
- [14] Maxim Frank-Kamenetskii. *Unraveling DNA*. VCH publishers, Inc., 1993.
- [15] Ferbeyre G, Smith JM, and Cedergren R. Schistosome satellite DNA encodes active Hammerhead ribozymes. *Molecular Cellular biology*, 7:3380-3388, 1998.
- [16] David Goodsell. *The Machinery of Life*. Springer-Verlag, 1992.
- [17] Stephen Gould and Niles Eldredge. Punctuated equilibria: An alternative to phyletic gradualism. In *Models in Paleobiology*, pages 82-115. Freeman and Cooper, 1972.
- [18] Stephen Gould and Niles Eldredge. "Punctuated equilibria: The tempo and mode of evolution reconsidered". *Paleobiology*, 3:115-151, 1977.
- [19] Stephen Gould and Niles Eldredge. "Punctuated equilibrium comes of age". *Nature*, 366:223-227, 1993.
- [20] Frank Harary. *Graph Theory*. Addison-Wesley, Reading, Mass., 1969.
- [21] Paul de Kruif. *Los Cazadores de Microbios*. Editorial Epoca, S.A., México, D.F., Séptima edición.
- [22] Benjamin Levin. *Genes VI*. Oxford University Press, New York, 1997.
- [23] Wen-Hsiung Li and Dan Graur. *Fundamentals of Molecular Evolution*. Sinauer Associates, Inc.

- [24] Boris Mirkin and Sergeï Rodin. *Graphs and Genes*, volume 11 of *Biomathematics*. springer-Verlag, Heidelberg, 1984.
- [25] Joël de Rosnay. *Biokit: A Journey to Life*. Orgenics ltd, Adama books, New York, 1984.
- [26] Peter Schuster and Walter Fontana. Algoritmo para el plegamiento de secuencias de RNA. Publicado en la red. www.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.
- [27] M. Zuker. Mfold, algoritmo para el plegamiento de secuencias de RNA. Publicado en la red. www-bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html.

Glosario

Aminoácido: Producto inicial de la síntesis de proteínas.

Anticodón: Triplete de nucleótidos complementario a un codón.

Codón: Trinucleótido que corresponde a cada uno de los residuos aminoácidos. Hay varios codones que no corresponden a aminoácidos pero que indican el final de la síntesis de una proteína (Estos se marcan en la figura del código genético con un *). También hay codones que indican el inicio de la síntesis de la proteína, usualmente AUG, GUG o UUG.

Eucarionte: Célula que posee núcleo. Organismo cuyas células tienen núcleo.

Energía libre: Es una constante termodinámica que indica la cantidad de energía requerida para, o liberada por una reacción. Esta cantidad es medida en Kcal/mol y se expresa con el parámetro ΔG . Las reacciones que requieren energía tienen un valor positivo. Reacciones que liberan energía tienen un valor negativo.

Nucleótido: Molécula formada por una base nitrogenada (A, T, C, G), azúcar (deoxirribosa para el DNA, ribosa para el RNA) y fosfato.

Polímero: Compuesto químico que representa una cadena de grupos recurrentes. La más sencilla es $\cdots -CH_2 - CH_2 - CH_2 - \cdots$ que corresponde al polietileno.

Procarionte: Célula que carece de núcleo. Organismo cuyas células no poseen núcleo.

Viroide: Molécula de RNA infecciosa que funciona independientemente, sin ser encapsulada por alguna cubierta de proteínas

Virusoide: Moléculas similares a un viroide pero que son encapsulados por algún virus junto con el genoma viral. Los virusoides no se replican de manera independiente, requieren la asistencia de algún virus. Los virusoides también son conocidos como RNA satélites.