



00382  
7

**UNIVERSIDAD NACIONAL  
AUTONOMA DE MEXICO**

**FACULTAD DE CIENCIAS  
DIVISION DE ESTUDIOS DE POSGRADO**

**POTENCIALES ALEATORIOS  
Y LA ESTRUCTURA DEL  
CODIGO GENETICO**

**T E S I S**

**QUE PARA OBTENER EL GRADO ACADEMICO DE**

**DOCTOR EN CIENCIAS  
(FISICA)**

**P R E S E N T A**

**MAXIMINO ALDANA GONZALEZ**

**DIRECTORES DE TESIS:**

**DR. GUSTAVO MARTINEZ MEKLER  
DR. GERMINAL COCHO GIL**

75/1982



posgrado en ciencias físicas  
u.n.a.m.

*México, D. F.*

*Noviembre 2000*



Universidad Nacional  
Autónoma de México



**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.



posgrado en ciencias físicas  
u n a m

OFICIO: PCF/164/2000  
ASUNTO: Designación de jurado

**ING. LEOPOLDO SILVA GUTIÉRREZ**  
DIRECTOR GENERAL DE LA ADMINISTRACIÓN ESCOLAR  
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
P R E S E N T E

At'n: **Biol. Francisco Javier Incera Ugalde**  
Jefe de la Unidad de Administración del Posgrado

El Comité Académico del Posgrado en Ciencias Físicas en su sesión del 4 de julio de 2000 ha designado como Jurado del alumno **ALDANA GONZÁLEZ MAXIMINO** con número de cuenta **8503349-6**, para dictaminar si el trabajo desarrollado como tesis titulado: *"Potenciales aleatorios y la estructura del código genético"*, dirigida por el Dr. Germinal Cocho Gil, tiene los méritos para obtener el grado de **DOCTOR EN CIENCIAS (FÍSICA)** conforme al plan de estudios 582.

PRESIDENTE:	DR. GERMINAL COCHO GIL
SECRETARIO:	DR. HERNÁN LARRALDE RIDAURA
VOCAL:	DR. CARLOS VILLARREAL LUJÁN
VOCAL:	DR. ANTONIO LAZCANO ARAUJO
VOCAL:	DR. GUSTAVO CARLOS MARTÍNEZ MEKLER
SUPLENTE:	DR. LUIS MOCHÁN BACKAL
SUPLENTE:	DR. FRANCISCO XAVIER SOBERÓN MAINERO

**A T E N T A M E N T E**

**"POR MIRAZA HABLARÁ EL ESPÍRITU"**

Ciudad Universitaria, D.F., a 16 de agosto de 2000.

**EL COORDINADOR DEL POSGRADO EN CIENCIAS FÍSICAS**

**DR. JOSÉ LUIS BOLDÚ OLAIZOLA**

JLBO/yhq

c.c.p. - Cada miembro del Síno  
c.c.p. - Expediente.

*A Juanita,  
porque es el hermoso pilar en  
el que está soportada mi vida.*

*A mi Mamá,  
porque no le queda de otra  
más que conformarse con  
el hijo que le tocó*

*A Andrés,  
que ya es universitario  
y está más alto que yo*

*A Claudia,  
para que se anime*

# Índice General

<b>Agradecimientos</b>	<b>3</b>
<b>Introducción</b>	<b>7</b>
<b>1 El Flujo de la Información Genética</b>	<b>13</b>
1.1 Las moléculas biológicas . . . . .	15
1.2 DNA: la molécula principal . . . . .	16
1.3 RNA: una molécula muy versátil . . . . .	24
1.4 Proteínas: el “hardware” celular . . . . .	29
1.5 El Dogma Central de la Biología Molecular . . . . .	33
<b>2 El Código Genético</b>	<b>37</b>
2.1 Las “palabras” del Código . . . . .	38
2.2 tRNA: la molécula adaptadora . . . . .	42
2.3 Ribosomas: fábricas de proteínas . . . . .	45
2.4 ¿Por qué tres? . . . . .	49
2.5 Algunas ideas sobre el origen del Código . . . . .	52
<b>3 La vida en la Tierra... ¿se originó en la Tierra?</b>	<b>57</b>
3.1 ¿Qué se entiende por “origen de la vida”? . . . . .	59
3.2 Dificultades con el origen terrestre de la vida . . . . .	60
3.3 El escenario extraterrestre . . . . .	67
3.4 No todo está resuelto . . . . .	72
<b>4 Potenciales Aleatorios</b>	<b>75</b>
4.1 El Modelo . . . . .	77
4.2 Los Potenciales . . . . .	81
4.3 Mínimos de energía . . . . .	87
4.4 Distribución de distancias entre mínimos de energía . . . . .	94
4.5 Distribución de distancias para secuencias reales . . . . .	107

<b>5</b>	<b>Dinámica</b>	<b>115</b>
5.1	La ecuación de movimiento . . . . .	116
5.2	Ordenes de magnitud . . . . .	118
5.3	Periodo tres en secuencias aleatorias . . . . .	124
5.4	Periodo 3 en secuencias reales . . . . .	132
5.5	Algunas consideraciones teóricas . . . . .	137
<b>6</b>	<b>Interacción con objetos extendidos</b>	<b>145</b>
6.1	Equivalencia entre objeto extendido y partícula puntual . . . . .	146
6.2	Potenciales aleatorios correlacionados . . . . .	150
6.3	Propiedades de integrabilidad . . . . .	154
6.4	El Teorema de Límite Central . . . . .	158
6.5	Condición de Lindeberg . . . . .	161
6.6	Potenciales bien comportados . . . . .	163
6.7	Función de probabilidad conjunta del potencial aleatorio . . . . .	167
6.8	Dinámica e invariancia de escala . . . . .	174
6.9	Distancia promedio entre mínimos de energía . . . . .	178
	<b>Conclusiones</b>	<b>187</b>
<b>A</b>	<b>Probabilidad de polimerización</b>	<b>193</b>
<b>B</b>	<b>Distribución de distancias para el caso <math>m = \infty</math></b>	<b>197</b>
<b>C</b>	<b>Espectro de potencias para una cadena extendida</b>	<b>203</b>
	<b>Bibliografía</b>	<b>209</b>

# Agradecimientos

*La gratitud es una enfermedad  
de los perros.*

Iósev Stalin.

Yo soy perro... y me siento enfermo. Y es que estoy profundamente agradecido con todas las personas que me ayudaron a realizar mis estudios. Hace once años, cuando ingresé a la carrera de Física en la Facultad de Ciencias de la UNAM, no tenía idea de lo maravilloso que iba a ser el estudiar. Y no me refiero sólo a la “maravilla del conocimiento científico” (aunque también a ésto). Lo que ha hecho que mis estudios sean maravillosos, son algunas de las personas que he conocido y con las que he convivido, y a las que ahora quiero expresar mi agradecimiento.

Para comenzar, agradezco infinitamente a mis asesores, los doctores Gustavo Martínez y Germinal Cocho, quienes propusieron el tema de esta tesis doctoral que me ha mantenido ocupado por poco más de cuatro años. Trabajando con ellos he aprendido más de lo que creí que se pudiera aprender en un doctorado. Pero sobre todo, agradezco la amistad que me han brindado. Han sido mis maestros y mis amigos, y me han apoyado no sólo en la parte académica, sino también en la parte personal. Gustavo y Germi se han preocupado no sólo de que aprenda, sino también de que me sienta bien aprendiendo, y me han hecho sentir que mi trabajo es algo importante. Considero que el haber trabajado y convivido con Germinal y Gustavo, ha sido una de las cosas más afortunadas que me han ocurrido. Muchas gracias a ambos, de todo corazón.

Otro acontecimiento verdaderamente afortunado, es el haber conocido a Hernán Larrarde, con quien también estoy profundamente agradecido. Hernán es, en muchos aspectos, fuente de inspiración inagotable. Trabajar

con él es interesante, motivante, divertido, y nunca ha mostrado, ni por asomo, un dejo de egoísmo, ni con su conocimiento, ni con su tiempo, ni con su amistad (ni con sus cigarros). Además, su contribución a este trabajo, y en general, a mi formación académica, es simplemente invaluable. Hernán es un buen maestro y un gran amigo, al que le debo mucho más de lo que pudiera retribuirle. ¡Gracias Dr. Larralde!

También deseo expresar mi más sincero agradecimiento a los sinodales que revisaron mi tesis, los doctores Antonio Lazcano, Carlos Villarreal, Xavier Soberón y Luis Mochán, cuyos comentarios y correcciones, lejos de ser fastidiosos, me ayudaron a comprender y estructurar mucho mejor los resultados de mi investigación. Antes de entregar mi tesis a los sinodales, me sentía temeroso. Por un lado, mi formación no es en biología, por lo que no estaba seguro de no decir disparates en la fundamentación biológica del trabajo. Y sí los dije. Agradezco encarecidamente a los doctores Antonio Lazcano y Xavier Soberón la bondad y paciencia que mostraron ante mis disparates.

Por otro lado, también estaba temeroso en la parte física y matemática del trabajo, pues *dado que mi formación es en Física, no se supone que yo pueda decir disparates en esta área*. Pero también los dije. La emoción de obtener nuevos resultados, y las ganas de que estos resultados coincidieran con lo que yo esperaba, me condujeron, en algunas ocasiones, a formular interpretaciones equivocadas. Afortunadamente, los físicos profesionales con los que colaboré, y en particular los que revisaron mi tesis detenidamente, me mostraron que las cosas no siempre salen bien desde el principio, y cuando sí salen bien, no siempre se entienden.

Gracias a todos mis sinodales, cuya actitud siempre constructiva, contribuyó a una mejor comprensión e interpretación del trabajo.

La UNAM es una institución realmente increíble. Entre otras cosas, el haber ingresado a la UNAM me permitió conocer a personas como las que he mencionado. Pero aparte, la UNAM me ha brindado las posibilidades de realizar mis estudios en condiciones que en otras instituciones ni siquiera se sueñan. Estoy agradecido con la UNAM, con sus profesores, con sus alumnos, con los cuerpos directivos y administrativos que, tal vez sin saberlo, han contribuido en mi formación académica. En particular, agradezco al Centro de Ciencias Físicas y al Instituto de Física el haberme proporcionado los recursos físicos y humanos que me permitieron realizar mis estudios de posgrado. También quiero agradecer a la Dirección General del Personal Académico (DGAPA), que me otorgó el apoyo económico, por medio de una beca de doctorado, sin el cual no hubiera podido ni siquiera escribir estas líneas.

Realmente no tengo palabras para agradecer a mi familia, a mi mamá y a mis hermanos (Andrés y Claudia), quienes siempre han creído en mí. Y a



Juanita, especialmente a Juanita, quien ha pagado con creces la fechoría de haberse casado conmigo, y aún así, sigue a mi lado. Por eso, es a ellas y a él a quines dedico mi trabajo, porque no encuentro otra forma de mostrarles todo lo que les debo.

La lista es realmente inacabable. Como mencionaba, en la UNAM he conocido a muchas personas que de alguna u otra forma han colaborado conmigo y de las que he aprendido. Por eso, no me queda más que decir

¡Gracias a todas las personas  
que me han ayudado!

¡GRRAUUFF!



# Introducción

En este trabajo abordamos algunos aspectos del origen del código genético. Uno de los objetivos es proponer una explicación de por qué los codones, las unidades básicas de información en los procesos de síntesis de proteínas, están compuestos por tres bases y sólo tres bases. Desde que en la década de los 60's se descubrió la estructura del código genético, la pregunta "¿por qué tres?" ha estado flotando en el aire sin que hasta la fecha se le haya dado una respuesta satisfactoria. El problema se ha abordado desde diversos puntos de vista, que van desde consideraciones fisicoquímicas de la interacción codón-anticodón, o codón-aminoácido, hasta consideraciones de optimización de la codificación de la información genética. No obstante, el problema aún sigue abierto.

Nosotros atacamos este problema desde un punto de vista diferente, considerando las propiedades dinámicas de máquinas moleculares primitivas unidimensionales. Hasta donde sabemos, es la primera vez que se trata de explicar la estructura de tripletes del código genético en base a consideraciones dinámicas. Lo que hacemos es proponer un modelo en el que una partícula puntual cargada interactúa con un polímero cuyos monómeros también están cargados, pero aleatoriamente. En nuestro modelo, la partícula y el polímero representan a las moléculas precursoras del ribosoma y el RNA mensajero, que son las moléculas que en la actualidad llevan a cabo los procesos de traducción de la información genética (síntesis de proteínas). Comparado con el sistema real ribosoma-mRNA, el modelo que proponemos es sumamente sencillo, y vale la pena discutir por qué esperamos que este modelo pueda darnos resultados que algo tengan que ver con la realidad.

Actualmente, el ribosoma se nos presenta como una molécula muy grande y complicada con varios sitios activos y decenas de proteínas enzimáticas entrando y saliendo. Un sistema así muy difícilmente puede ser modelado de forma realista por medio de una partícula puntual. No obstante, el hecho de que el código genético sea el mismo (o casi el mismo) en todos los organismos vivos que se conocen, desde las bacterias hasta

los mamíferos, hace pensar que la estructura de tripletes del código quedó establecida en las etapas iniciales del origen de la vida, es decir, cuando apenas se estaban llevando a cabo los procesos protobióticos y prebióticos que eventualmente condujeron a los primeros organismos unicelulares. En tales etapas primitivas del origen de la vida, el ribosoma seguramente no era la molécula tan complicada que ahora es.

El punto de vista que adoptamos en este trabajo es entonces el siguiente: la estructura de tripletes del código genético quedó establecida en etapas muy tempranas del origen de la vida. Por lo tanto, aun cuando los procesos de traducción de la información genética adquirieron cada vez mayor complejidad a lo largo de la evolución, tanto en las interacciones como en la organización de las moléculas involucradas, dichos procesos se construyeron sobre la base de una estructura preexistente de tripletes de la información genética, la cual se mantiene hasta nuestros días.

Bajo esta perspectiva, el modelo de máquina molecular que proponemos trata de rescatar los elementos esenciales de la interacción del ribosoma con el mRNA, pero no como se da actualmente, sino como probablemente ocurrió en el origen de la vida. Los resultados de este modelo indican que, efectivamente, hay una preferencia dinámica por el tres, es decir, la partícula se mueve a lo largo del polímero en pasos que, estadísticamente, son de tres monómeros en tres monómeros. Lo anterior es así porque el potencial de interacción entre la partícula y el polímero presenta mínimos separados, en promedio, cada tres monómeros. Esta regularidad en los potenciales se ve reflejada en el comportamiento dinámico de la partícula. Cabe mencionar que la preferencia estadística por el tres que resulta en el modelo, tiene lugar para una amplia gama de potenciales de interacción electrostáticos, de tal suerte que los resultados obtenidos son más o menos generales y no dependen fuertemente de un potencial particular o de un ajuste fino de parámetros.

También es necesario enfatizar que nuestro modelo no se contrapone en lo absoluto con las otras teorías que se han elaborado para explicar la estructura del código genético. Por el contrario, se suma a ellas pero desde una perspectiva completamente diferente, contribuyendo a un mejor entendimiento de la estructura del código genético.

Este trabajo consta esencialmente de tres partes. La primera parte, conformada por los capítulos 1, 2 y 3, es una introducción a la fenomenología biológica de nuestro problema. Aquí describimos a grandes rasgos los procesos de almacenamiento, transcripción y traducción de la información genética, lo cual sienta las bases para entender la estructura del código genético y por qué es importante encontrar una explicación para dicha estructura. En particular, en el capítulo 2 se plantea el problema medular de nuestra investigación (“¿por qué tres?”) y se describen las principales

teorías con las que se cuenta actualmente para explicar la estructura del código genético. Como veremos, tales teorías están más enfocadas a tratar de comprender la especificidad de los aminoácidos con los codones, que en explicar la estructura de tripletes del código. Quedará claro, por tanto, que el punto de vista que adoptamos en este trabajo es radicalmente diferente a los que se han adoptado antes, no sólo por los métodos de análisis utilizados, sino porque también estamos tratando de explicar algo que en general se había dado por hecho.

Debido a que nuestro interés está orientado hacia el *origen* de la estructura del código genético, debemos proponer un modelo compatible con las condiciones físicas que prevalecían en el origen de la vida. Esto no resulta nada fácil, porque no se sabe bien a bien cuáles eran dichas condiciones. Sin embargo, adoptaremos la teoría que propone que los procesos prebióticos y protobióticos que eventualmente condujeron a los organismos vivos, se llevaron a cabo fuera de la Tierra, en el espacio exterior, probablemente en el medio interestelar o en superficies de cometas o de meteoritos. En el capítulo 3 exponemos los argumentos que conducen a formular esta hipótesis sobre el origen extraterrestre de la vida. El escenario extraterrestre presenta ciertas ventajas respecto de un escenario terrestre. En particular, la unidimensionalidad de los procesos de almacenamiento, replicación, transcripción y traducción de la información genética, es una consecuencia natural del entorno en el cual se originaron dichos procesos. De ahí viene precisamente la unidimensionalidad del modelo que proponemos.

La primera parte está un poco larga (ocupa casi 70 páginas). Esto es así intencionalmente porque, además de ser una introducción, puede servir también como una referencia rápida para los estudiantes de física que se quieran dedicar a problemas de biología celular, en particular, a los relacionados con el flujo de la información genética.

En la segunda parte, que consta de los capítulos 4 y 5, presentamos el modelo de la partícula cargada interactuando con el polímero cuyos monómeros están cargados aleatoriamente. La carga en los monómeros la escogemos al azar porque suponemos que en el origen de la vida las moléculas precursoras del RNA aún no portaban ninguna información, y por lo tanto la secuencia de bases a lo largo de estas cadenas era más o menos aleatoria. Estudiando las propiedades estadísticas de los potenciales de interacción entre la partícula y el polímero vemos que, bajo condiciones muy generales, la distancia promedio entre mínimos consecutivos en el potencial está muy cercana a tres (en unidades de la longitud de los monómeros), aunque la distancia más probable es dos. Sin embargo, la situación cambia cuando en lugar de asignar las cargas de los monómeros de manera aleatoria, las asignamos en correspondencia con la secuencia de bases de una cadena genética codificadora real (utilizamos secuencias de *E. coli* y de

*Drosophila*, por ejemplo). En este caso tanto la distancia promedio como la distancia más probable entre mínimos consecutivos en el potencial están en tres, lo cual sugiere un proceso de "refinamiento del 3" a lo largo de la evolución.

Las regularidades estadísticas presentes en los potenciales de interacción se traducen en regularidades dinámicas cuando consideramos el movimiento de la partícula a lo largo del polímero. Así, en el capítulo 5 veremos como el espectro de potencias de la velocidad de la partícula presenta un pico muy marcado en periodo 3, lo cual sugiere que el movimiento de la partícula se da en pasos de tres en tres a lo largo del polímero. Esta "periodicidad" en tres se acentúa aún más cuando las cargas de los monómeros en el polímero se asignan en correspondencia con secuencias genéticas reales, lo que está en perfecto acuerdo con el proceso de "refinamiento del 3" al que nos referíamos antes. Como se verá en este refinamiento no ocurre cuando se utilizan secuencias intergénicas no codificadoras para asignar las cargas a lo largo del polímero.

Los resultados obtenidos en la segunda parte de este trabajo apuntan en la dirección de que el movimiento de la partícula a lo largo del polímero se da, en promedio, en pasos de tres monómeros. Por lo tanto, un código genético en el que los codones estén formados por tres bases se ve favorecido por las propiedades dinámicas de las máquinas moleculares primitivas. El hecho de que, para las secuencias genéticas codificadoras se acentúe en gran medida esta regularidad de tres, mientras que para las secuencias no codificadoras lo anterior no ocurra, sugiere que los procesos de síntesis de proteínas se construyeron tomando como base una dinámica similar a la exhibida en nuestro modelo.

El problema de investigación sobre la estructura del código genético nos condujo a trabajar con potenciales aleatorios. Este tipo de potenciales también son de gran interés en otras áreas de la Física, tales como máquinas moleculares o transporte a través de medios amorfos. Por lo tanto, y ya entrados en materia, nos dedicamos a estudiar las propiedades estadísticas de este tipo de potenciales, lo cual es el tema de la tercera parte de este trabajo, que consiste sólo del capítulo 6 (el último). La motivación inicial para este estudio provino de sustituir a la partícula puntual que interactúa con el polímero, por un objeto extendido, ya que el ribosoma, después de todo, no es puntual, sino que tiene cierto tamaño. Como veremos, el tener un objeto extendido interactuando con el polímero es completamente equivalente a tener una partícula puntual interactuando con el mismo polímero, pero de tal manera que la distancia entre la partícula y el polímero debe ser muy grande.

Así, el objeto extendido conduce a un problema similar al que se estudió en la segunda parte. Sin embargo, debido a que en el sistema objeto-

polímero aparecen correlaciones de largo alcance en los potenciales de interacción, mientras que en el sistema partícula-polímero tales correlaciones no están presentes, existen diferencias importantes en la dinámica de ambos sistemas. Los resultados encontrados al considerar la dinámica del objeto extendido a lo largo del polímero, no necesariamente están ligados a la estructura del código genético. Por ejemplo, encontramos que los potenciales de interacción presentan propiedades de escalamiento, es decir, tienen la misma estructura estadística a todas las escalas. Lo anterior significa que un objeto extendido (interactuando con el polímero), "sentirá" el mismo potencial independientemente del tamaño de dicho objeto. Además, mostramos que los potenciales electrostáticos aleatorios generados por redes unidimensionales de partículas puntuales cargadas, presentan una estadística gaussiana correlacionada, independientemente del tipo de partículas que conforman la red (iones, dipolos permanentes, dipolos inducidos, etc.). Como veremos, algunas de las regularidades dinámicas a las que nos hemos referido son consecuencia directa de las correlaciones de largo alcance de los potenciales de interacción.

Terminamos este trabajo con una sección de conclusiones, en la cual se mencionan, principalmente, algunos problemas que aún faltan por resolver y que pueden considerarse como la continuación natural de nuestra investigación.

Vale la pena mencionar que los resultados que presentamos en este trabajo han dado lugar a la publicación de tres artículos de investigación en revistas de corte internacional [1], [2], [3], a un artículo en memorias de congreso internacional (con arbitraje) [4] y a un artículo de divulgación nacional [5].





# Capítulo 1

## El Flujo de la Información Genética

A mediados del siglo XX un grupo de biólogos, físicos, químicos y médicos sentaron las bases de la genética molecular al descubrir que la molécula de ácido desoxirribonucleico (DNA) es la portadora de la información genética de los organismos vivos [6]. A partir de entonces es común aceptar que las características fenotípicas de los seres vivos están codificadas (o especificadas) en los cromosomas, los cuales no son más que moléculas muy largas de DNA enrolladas sobre sí mismas, y que por medio de la transferencia del DNA de la célula madre a las células hijas en la división celular, tales características fenotípicas se heredan de un organismo a sus descendientes. Así, todos los rasgos físicos que caracterizan a una especie, como el tamaño, color de ojos, tipo de piel, tersura del pelo, etc., están “escritos”, de alguna manera, en el DNA del banco genético de la especie. Más aún, las funciones metabólicas internas de un organismo, como la respiración o la degradación de los alimentos, también están especificadas en el DNA celular.

Estos descubrimientos pusieron de manifiesto que en el DNA se encuentra la información de la estructura y funcionamiento de los seres vivos. El problema de determinar cómo dicha información está contenida en el DNA no se ha resuelto completamente, pero se han dado avances sustanciales en su solución. En particular, se conoce ahora con gran precisión la respuesta a las tres preguntas siguientes, las cuales podríamos considerar como fundamentales en el sentido de que representan los primeros pasos en la solución del problema del desciframiento de la información genética:

- ¿Qué información contiene el el DNA?

- ¿Cómo está codificada la información genética en el DNA?
- ¿Cómo se lee dicha información?

Cuando decimos, por ejemplo, que el DNA contiene la información del color de los ojos, ¿a qué nos referimos exactamente? Claramente, en los cromosomas no está escrita la frase "HACER UN OJO CAFE", al menos no de esta manera. Más bien, lo que contiene el DNA es la información de las moléculas protéicas que conforman el ojo, incluyendo aquellas que le dan el color. Esta información debe estar codificada de alguna forma en el DNA, es decir, escrita en algún tipo de "lenguaje molecular" que la maquinaria celular puede leer y comprender. Tal lenguaje molecular se conoce con el nombre de *Código Genético*, y fue descifrado gracias a los trabajos pioneros que realizaron Marshall Nirenberg y J. Heinrich Matthaei en la década de los años 60. Podemos decir que dicho código es el lenguaje que la Naturaleza utiliza para construir a los seres vivos, escribiendo en el DNA la información de las moléculas protéicas que los conforman.

Los mecanismos esenciales de codificación y lectura de la información genética, así como los que se refieren a la síntesis de las moléculas protéicas especificadas por el DNA (las que "conforman el ojo"), están bien comprendidos y forman la base de la biología molecular. No obstante, el paso que nos lleva de la comprensión de estos mecanismos a la formación de un ser vivo, no se ha dado aún. Sabemos qué tipo de moléculas son especificadas por la información genética contenida en el DNA, pero no podemos predecir de antemano para qué sirven ni cómo se organizan. También sabemos cómo está estructurada y cómo se lee dicha información, pero no conocemos el por qué de esta estructura ni el por qué de los mecanismos de lectura. Tampoco se entiende completamente cómo se construyeron tales mecanismos en el origen de la vida, y aunque se tienen modelos que tratan de explicar el origen y evolución de los procesos de codificación, lectura y expresión de la información genética, no han resultado del todo satisfactorios. No está por demás enfatizar el hecho de que dar respuesta a estas interrogantes conllevará a una mejor comprensión del origen, estructura y funcionamiento de los seres vivos.

A lo largo de este trabajo veremos un modelo que pretende explicar algunos aspectos del origen del Código Genético. Comenzaremos en este capítulo describiendo brevemente la fenomenología biológica relacionada con el almacenamiento de la información genética en el DNA, la manera en que dicha información está codificada y los mecanismos moleculares encargados de leerla y hacerla operativa. No abundaremos demasiado en la estructura de los ácidos nucleicos (DNA y RNA) ni de las proteínas, porque existe una gran cantidad de literatura al respecto [7] [8] [9]. Simplemente

se describen aquí los fundamentos biológicos necesarios para plantear el problema del origen del Código Genético y comprender su importancia en el entendimiento del origen de los seres vivos.

## 1.1 Las moléculas biológicas

En términos generales, las células están compuestas de cuatro tipos diferentes de moléculas orgánicas: azúcares, ácidos grasos, proteínas y ácidos nucleicos, cada uno de los cuales juega un papel fundamental en la estructura y funcionamiento celulares. Los azúcares, por ejemplo, son moléculas cuya función principal es el almacenamiento de energía química que puede ser utilizada para dirigir alguna reacción metabólica que de otra forma sería irrealizable. Los ácidos grasos, por otro lado, también sirven para almacenar energía química, pero su función principal dentro de la célula está en la construcción de membranas (como fosfolípidos).

Sin duda alguna, las proteínas son las moléculas de mayor diversidad y versatilidad dentro de la célula. Tales moléculas son las encargadas de regular la actividad, tanto funcional como estructural, de los seres vivos. Dentro de la célula se producen miles de reacciones químicas cada segundo, reacciones que garantizan la vida y la reproducción celular. Estas reacciones tienen que llevarse a cabo de manera organizada, de tal forma que los productos de una reacción sean el sustrato de la reacción siguiente. Al conjunto de todas estas reacciones químicas se le conoce con el nombre de *metabolismo celular*. Las proteínas son las encargadas de asegurar que el metabolismo celular se lleve a cabo correctamente. En otras palabras, debido a la gran diversidad y versatilidad de las proteínas, estas moléculas pueden catalizar una reacción química, indicando el comienzo y el final de la reacción. Pueden servir como catalizadores de muy alta eficiencia, o como inhibidores, de tal manera que la reacción se realice sólo bajo determinadas circunstancias. Pueden sintetizar otras moléculas, o degradarlas justo cuando ya han cumplido su función. Así, por ejemplo, hay proteínas que transportan oxígeno a través de la membrana celular mientras que otras dirigen las reacciones químicas que oxidan los azúcares en las mitocondrias (ciclo de Krebs). Por cada tarea que la célula requiere para mantenerse viva y reproducirse, existe una proteína (o un grupo de proteínas) encargada de realizarla, asegurando que la "maquinaria celular" funcione bien.

Las células no solamente son "máquinas biológicas" muy complejas y organizadas, sino que además son máquinas que se construyen a sí mismas. Las células se reproducen teniendo descendientes, y la información del metabolismo celular debe pasar de la célula madre a las células hijas íntegramente. Los ácidos nucleicos (DNA y RNA) son las moléculas encar-

gadas de almacenar esta información. Como veremos más adelante, en los ácidos nucleicos está contenida la información, codificada en un alfabeto de cuatro letras, de las características estructurales y conformacionales de las proteínas, las cuales controlan a su vez el metabolismo celular. Si hacemos una analogía con las computadoras, podríamos considerar a los ácidos nucleicos como el software (los programas e instrucciones que debe seguir la computadora para realizar alguna tarea determinada), mientras que las proteínas son el hardware (los circuitos y elementos físicos que hacen operativas las instrucciones de los programas).

## 1.2 DNA: la molécula principal

El DNA (ácido desoxirribonucleico) es el componente principal de los cromosomas, y es en todos los organismos vivos <sup>1</sup>, el portador de la información genética. De acuerdo con este cometido, el DNA no está sometido, como la mayoría de las demás macromoléculas de la célula, a un ciclo metabólico en el cual se sintetizan y destruyen constante y paralelamente moléculas de DNA. Los cromosomas se crean en el nacimiento mismo de la célula, y permanecen casi inalterados hasta que la célula muere.

Por su constitución química, el DNA es un polímero lineal extremadamente estable, compuesto por ácido fosfórico, desoxirribosa (pentosa) y cuatro tipos de bases nitrogenadas: adenina (A), timina (T), citosina (C) y guanina (G). Una base nitrogenada (A, T, C o G) unida a un grupo fosfato y a un azúcar se llama *nucleótido*, de tal manera que son cuatro tipos de nucleótidos los que se polimerizan para formar la cadena lineal de DNA. La Fig1.1 muestra la estructura general de los nucleótidos, y la Fig1.2 las cuatro bases nitrogenadas. El grupo fosfato y el azúcar son los mismos en cada nucleótido, y sólo se diferencian en la base nitrogenada que portan. Por tal motivo, identificaremos a los nucleótidos con las mismas letras A, T, C y G con las que identificamos a las bases.

En 1953, J. D. Watson y F. Crick, basándose en las fotografías de rayos X tomadas por M. Wilkins y R. Franklin, descubrieron que el DNA está compuesto por dos cadenas poliméricas de nucleótidos enrolladas una alrededor de la otra, dando lugar a una estructura de doble hélice (como escalera de caracol).

---

<sup>1</sup>Hay algunos virus en los que el material genético está compuesto de RNA en lugar de DNA. Un ejemplo típico de estos *retrovirus*, es el VIH, el virus que causa el SIDA. Sin embargo, a los virus no se les considera como organismos vivos.

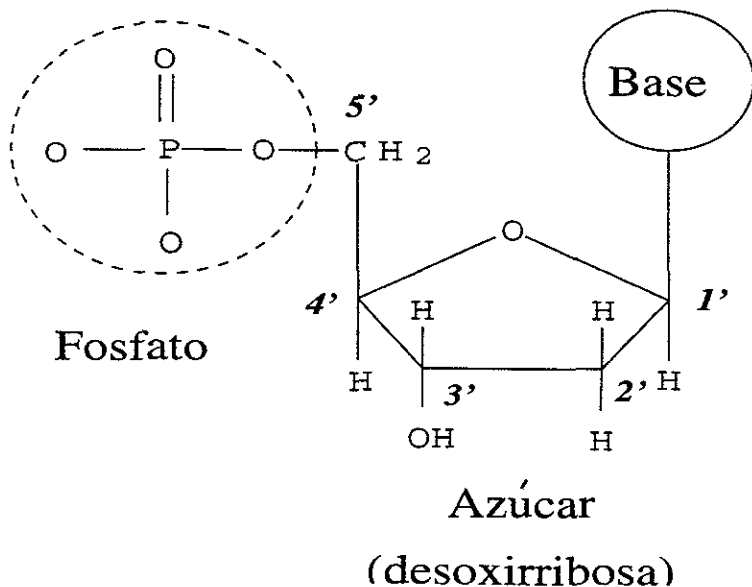


Figura 1.1: Estructura general de los nucleótidos del DNA: un grupo fosfato, un azúcar y una base nitrogenada. Los 5 átomos de carbono en el azúcar están numerados de acuerdo con su posición alrededor del anillo, comenzando con el 1' que es el átomo donde se liga la base, y terminando en el 5' que está ligado al grupo fosfato.

En cada cadena, los nucleótidos se polimerizan de manera que el grupo fosfato de un nucleótido se liga covalentemente al carbono 3' del azúcar del siguiente nucleótido. Por otro lado, lo que mantiene unidas a ambas cadenas es la interacción complementaria entre las bases nitrogenadas de los nucleótidos: la adenina en una de las cadenas se liga a la timina en la otra cadena por medio de dos puentes de hidrógeno, mientras que la citosina en una cadena se liga a la guanina en la otra cadena a través de tres puentes de hidrógeno. La "complementariedad" consiste, precisamente, en que la adenina siempre se liga a la timina (A-T) y la guanina siempre se liga a la citosina (C-G), como puede verse en la Fig.1.2. Aunque existen interacciones no complementarias entre las bases, (como por ejemplo la interacción U-G que se da en el RNA de transferencia), tales interacciones son muy débiles y no ocurren frecuentemente en la molécula de DNA. A las interacciones complementarias habituales A-T y C-G se les llama también interacciones de tipo Watson-Crick.

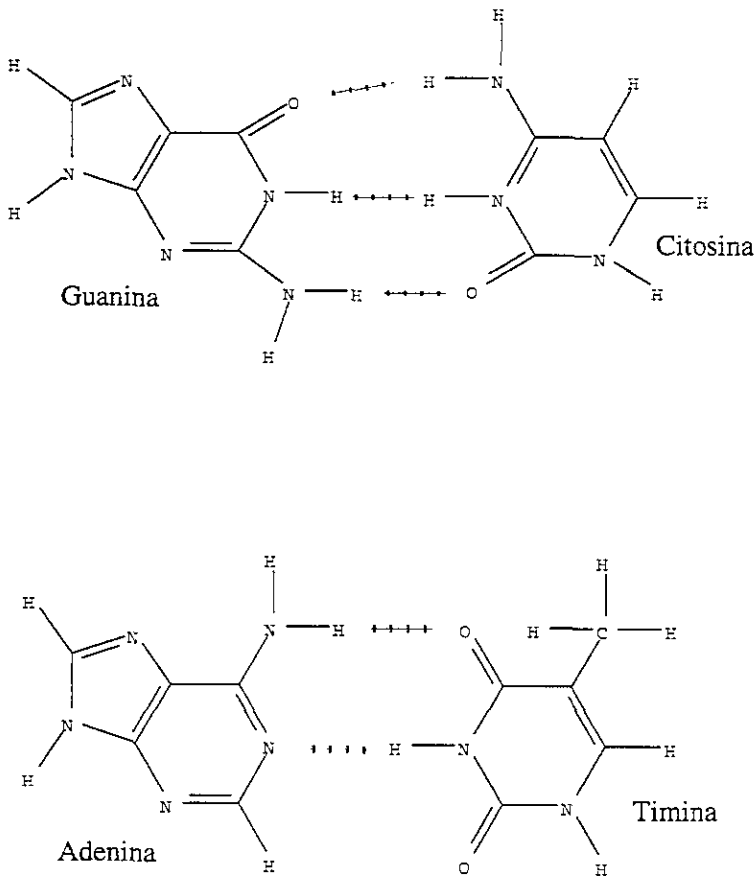


Figura 1.2: Las cuatro bases nitrogenadas del DNA y las interacciones complementarias entre ellas. Como puede verse, la guanina se liga a la citosina por medio de tres puentes de hidrógeno, mientras que la adenina y la timina se ligan a través de sólo dos puentes de hidrógeno.

Por otro lado, la Fig1.3 muestra un esquema del principio estructural de la molécula de DNA. Puede verse de esta figura que las dos cadenas que forman la estructura de doble hélice en el DNA deben ser complementarias para que puedan enlazarse. En otras palabras, si en una de las cadenas se encuentra una A, en la cadena complementaria se debe encontrar una T en el sitio correspondiente, mientras que si se encuentra una C, en la cadena complementaria debe estar presente una G, y viceversa.

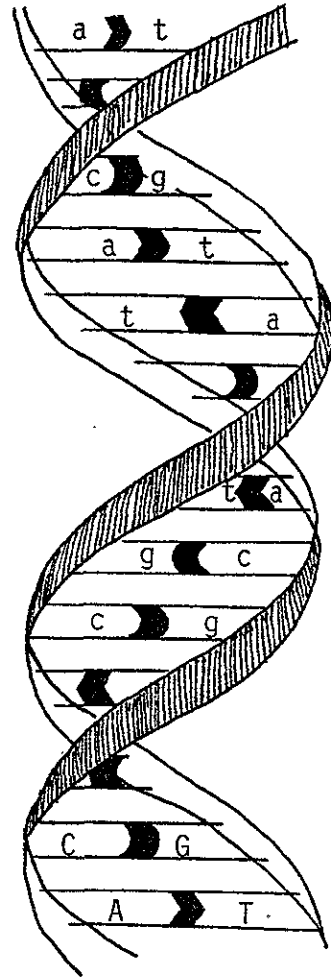
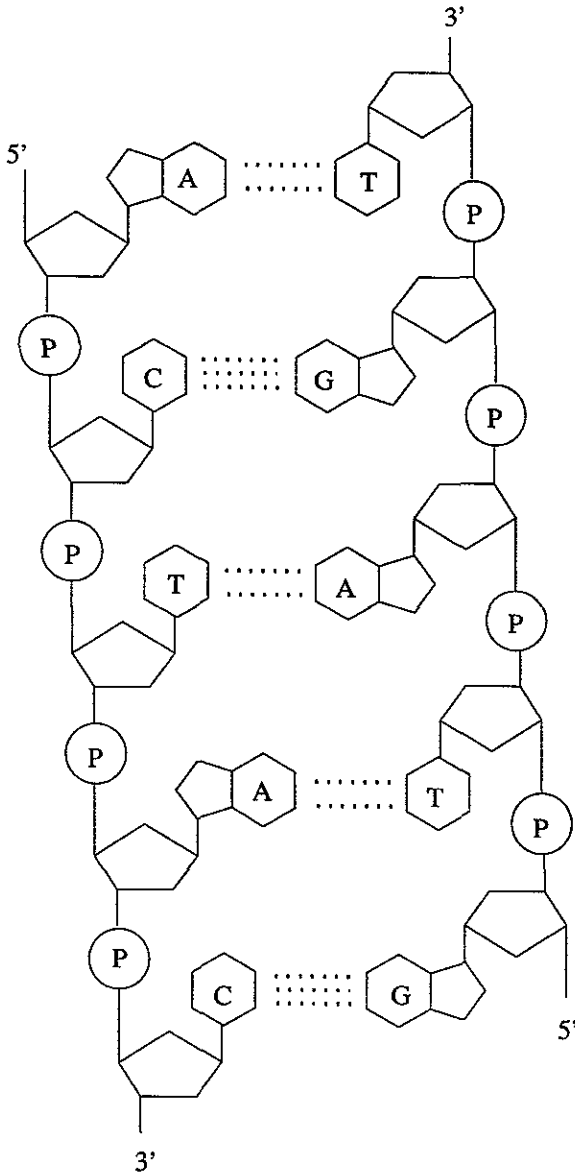
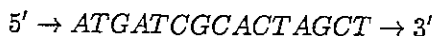
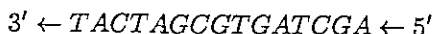


Figura 1.3: Principio estructural de la molécula de DNA, en la cual, dos cadenas de nucleótidos se enlazan por medio de las interacciones complementarias entre las bases. Nótese que las cadenas tienen orientaciones opuestas o antiparalelas (orientación de los azúcares).

Debido a la complementariedad de las dos cadenas, basta con que nos refiramos a una sólo de ellas, ya que si en un trozo de una de las cadenas la secuencia de bases es, por ejemplo



sabemos que en el trozo correspondiente de la cadena complementaria la secuencia de bases deberá ser



Notemos además que en las dos secuencias anteriores, el sentido de las flechas está intercambiado. Lo que ocurre es que la dirección de los nucleótidos en una cadena es opuesta a la dirección que estos tienen en la cadena complementaria, tal y como puede apreciarse en la Fig1.3. El principio de complementariedad entre las dos cadenas que constituyen al DNA resultó ser de importancia fundamental para el almacenamiento y la transmisión de la información genética en las células.

Es solamente en cuanto a su constitución química que al DNA se le puede dar un nombre. Sin embargo, en la Naturaleza existen tantas moléculas de DNA como organismos vivos, siendo la *secuencia de bases* a lo largo de la cadena lo que marca la diferencia entre una molécula y otra. El DNA de cada organismo vivo se caracteriza por tener una longitud y una secuencia de bases particulares. Aparentemente, entre más complejo es un organismo, más larga es su molécula de DNA. Así, el DNA de la bacteria *Escherichia coli*, por ejemplo, contiene aproximadamente  $4.72 \times 10^6$  pares de bases, mientras que la molécula de DNA de los seres humanos contiene  $3.3 \times 10^9$  pares de bases<sup>2</sup>. La tabla I muestra la longitud de la molécula de DNA para diferentes organismos representativos.

Cabe mencionar que no siempre los organismos más complejos contienen moléculas de DNA mayores que las de los organismos menos complejos. Aunque generalmente ocurre así, resulta que hay excepciones. Tal es el caso, por ejemplo, de la salamandra, un anfibio que habita en climas templados y que contiene 27 veces más DNA que los seres humanos. También se sabe que algunas plantas con flores y algunos moluscos contienen incluso hasta 10 veces más DNA que los mamíferos. Esto ha originado un gran debate respecto a lo que se entiende por "organismos más complejos", y basta decir por el momento que la cuestión no ha sido del todo aclarada.

---

<sup>2</sup>Hablamos de "pares de bases" porque en la molécula de DNA hay dos cadenas. En cada una de las cadenas del DNA de *E. coli* hay  $4.72 \times 10^6$  bases.



TABLA I

Organismo	Longitud del Genoma (pares de bases)
<b>Virus</b>	
Hepatitis B	$3 \times 10^3$
HIV (SIDA)	$9.8 \times 10^3$
Fago $\lambda$	$4.9 \times 10^4$
Varicela	$1.24 \times 10^5$
Citomegalovirus	$2.29 \times 10^5$
<b>Bacterias</b>	
Micoplasmas	$6 - 12 \times 10^5$
Arqueobacterias	$19 - 30 \times 10^5$
<i>Bacillus subtilis</i>	$4 \times 10^6$
<i>Escherichia coli</i>	$4.72 \times 10^6$
<b>Organismos Superiores</b>	
Gusano	$1.1 \times 10^8$
<i>Drosophila</i>	$1.7 \times 10^8$
Ratón	$3 \times 10^9$
Ser Humano	$3.3 \times 10^9$

No solamente es la longitud de la molécula de DNA la que varía de un organismo a otro, sino que también cambia la secuencia de bases a lo largo de la cadena. De hecho, es precisamente la forma en la que están ordenadas las cuatro bases A, T, C y G a lo largo de la molécula de DNA la que determina la información genética del metabolismo celular, así como en un libro, la secuencia de letras y signos de puntuación es la que determina la información que lleva consigo dicho libro. No obstante, existen diferencias entre la forma de codificar la información en el DNA y la forma en que se codifica en un libro. Por ejemplo, en el DNA se utiliza un "alfabeto" de sólo cuatro letras (las cuatro bases nitrogenadas) para escribir los mensajes genéticos, y además, en el DNA no hay signos de puntuación.

Existe otra diferencia importante entre el lenguaje que utiliza la Naturaleza para escribir los mensajes genéticos y el lenguaje que utilizamos nosotros para comunicarnos y escribir libros. En nuestro lenguaje, digamos el español, las palabras tienen longitud variable, es decir, el número de letras que conforman cada palabra cambia, según la palabra de que se trate. Así, encontramos palabras tan cortas como "yo", o palabras tan largas como "quimicomedicamentoso". Sin embargo, *las palabras con las cuales está escrita la información genética en el DNA tienen todas la misma longitud, a saber, tres bases por palabra*. Cuando a finales de la década de los sesenta, Marshall Nieremberg descubrió esta propiedad de las "palabras genéticas", se consideró más bien como una curiosidad evolutiva. No obstante, explicar

esta propiedad no ha resultado nada fácil, y de hecho, hasta la fecha, 30 años después de los descubrimientos de Nieremberg, no se ha podido dar una explicación satisfactoria a la pregunta de por qué las palabras que utiliza la Naturaleza para cifrar la información genética del metabolismo celular en los organismos vivos, están todas constituidas por tres bases y sólo tres bases. Este problema constituye el núcleo del presente trabajo y, naturalmente, volveremos a él en capítulos posteriores.

Hasta el momento, hemos visto que en el DNA está escrita la información del metabolismo celular, pero, ¿de qué información se trata exactamente. La respuesta es la siguiente:

*en la molécula de DNA se encuentra, principalmente, la información de las proteínas que utiliza la célula para llevar a cabo sus funciones metabólicas.*

Es necesario señalar que, además de la información de las proteínas, el DNA contiene la información de ciertos RNA catalíticos y ribosomales que sirven para regular los procesos de edición y traducción del mensaje genético. Además, el DNA contiene también secuencias reguladoras que, aunque no codifican para ninguna proteína ni para RNA, sirven para regular el inicio y el final de los procesos de transcripción. Sin embargo, la gran mayoría del DNA funcional de la célula se utiliza para la construcción de las proteínas metabólicas. Vale la pena en este momento hacer un comentario respecto a la "funcionalidad" del DNA. Resulta que poco más del 95% del genoma de las células eucariontes (células con núcleo) no tiene ninguna función, o por lo menos, no se le ha encontrado ninguna hasta el momento. A lo largo de la molécula de DNA de una célula eucariote existen grandes pedazos, llamados *secuencias intergénicas*, que no codifican para proteínas ni para ninguna otra molécula. Intercalados entre estas enormes secuencias intergénicas, existen pedazos que sí codifican para proteínas o para secuencias de RNA reguladoras. A tales pedazos codificadores, o funcionales, se les llama *genes*. En muchas ocasiones, para especificar una proteína determinada se requiere de varios genes. Por ejemplo, para codificar a la proteína hemoglobina, responsable de fijar y transportar el oxígeno en la sangre de los seres humanos, son necesarios dos genes.

Pero incluso si analizamos un gen con detenimiento, nos encontramos con que dentro del gen, también hay pedazos que no codifican, intercalados entre pedazos que sí codifican. A los pedazos no codificadores dentro de un gen se les llama *intrones*, mientras que a los pedazos codificadores se les denomina *exones*. La situación se representa esquemáticamente en la Fig1.4. Claramente, en el momento en que la información contenida en el DNA va a ser utilizada para construir una proteína, es necesario eliminar

todas las partes del gen que no llevan consigo ninguna información (intrones), y quedarse solamente con las regiones codificadoras (exones). A este proceso de “depuración” de la información genética se le conoce con el nombre de *edición*. Más adelante veremos someramente cómo se lleva a cabo la edición del mensaje genético. Por el momento, basta con enfatizar que la cantidad de DNA en forma de secuencias intergénicas e intrones llega a superar el 95% del material genético total de las células eucariontes. En otras palabras, más del 95% del DNA que contienen nuestras células no tiene, aparentemente, ninguna función.

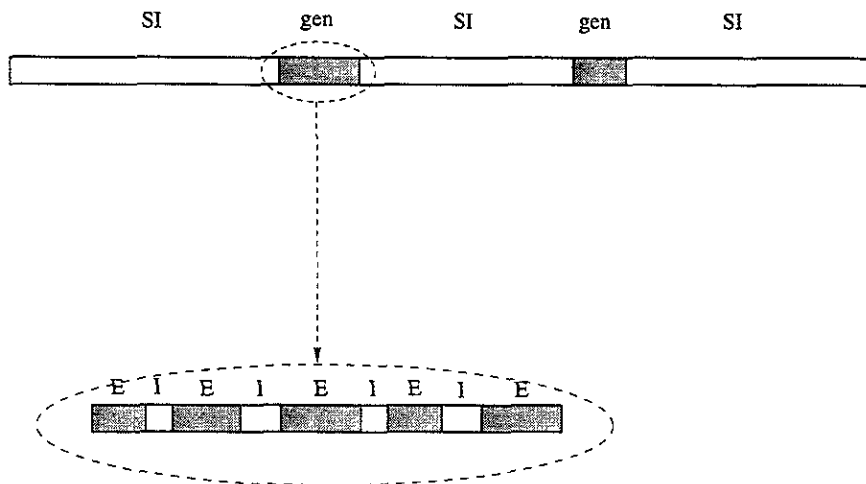


Figura 1.4: Dibujo esquemático del genoma de los organismos eucariontes. A lo largo de la molécula de DNA, existen pedazos no codificadores (secuencias intergénicas SI), intercalados entre pedazos codificadores (genes). A su vez, los genes presentan esta misma estructura, es decir, un gen contiene segmentos codificadores (exones E) intercalados entre segmentos no codificadores (intrones I).

Realmente, lo que ocurre es que no se le ha encontrado ninguna utilidad a este 95% de DNA, lo cual no significa que no la tenga. Sin embargo, hay evidencia a favor de que efectivamente este DNA no tiene ninguna función y que se encuentra en la célula como un remanente de desperdicio evolutivo. Pensemos por ejemplo en la salamandra, que tiene 27 veces más DNA que los seres humanos. ¿Para qué necesita tanto DNA un organismo como la salamandra? Al parecer, dicho organismo no tiene ninguna función biológica especial que le haga requerir tanto material genético. Cuestiones como esta nos hacen pensar que una gran parte de las secuencias

intergénicas y de los intrones no tienen ninguna función específica asignada dentro de la maquinaria celular. Tal vez, en algún momento de la evolución sirvieron para algo, pero después fueron reemplazadas por secuencias más cortas y más eficientes. Sin embargo, el problema se torna aún más paradójico si consideramos que ni los virus ni las bacterias contienen secuencias intergénicas, por lo que casi todo el genoma de estos organismos es codificador. La presencia de material genético no funcional en las células eucariontes sigue siendo un enigma por resolver.

Antes de que nos adentremos en el tema de cómo está codificada la información en el DNA y de qué manera se convierte dicha información en proteínas, es necesario que nos detengamos primero a revisar un poco de la estructura y funcionamiento de otra molécula muy parecida al DNA, pero al mismo tiempo, también muy diferente, a saber, el RNA.

### 1.3 RNA: una molécula muy versátil

A pesar de la importancia fundamental que tiene para los organismos vivos, el DNA sólo sirve para almacenar información. Como se mencionaba en la sección precedente, el DNA no está sujeto a ciclos metabólicos en los cuales se formen y se destruyan nuevas moléculas de DNA. Tampoco es una molécula catalizadora ni nada por el estilo y, aparte de los procesos de duplicación y transcripción, prácticamente no participa en ninguna reacción metabólica celular. El DNA es una molécula muy estable químicamente. Tal vez por eso fue "seleccionada" por la Naturaleza para ser la portadora perenne de la información genética de todos los organismos vivos.

En las células eucariontes, el DNA se encuentra dentro del núcleo celular, y nunca sale de ahí<sup>3</sup>. También se encuentra un poco de DNA en las mitocondrias (los "pulmones" celulares), y en los plástidos de células eucariontes fotosintéticas, pero en mucho menor cantidad. Por otro lado, la síntesis de proteínas se lleva a cabo en los ribosomas, los cuales se encuentran en el retículo endoplasmático, en el exterior del núcleo de la célula. Debido a que el lugar en donde se almacena la información genética (núcleo) y el lugar en donde se utiliza dicha información para construir proteínas (retículo endoplasmático), son dos lugares físicamente diferentes, y además separados por una membrana, evidentemente debe existir algo que transporte la información de un lugar a otro. Ese "algo" es precisamente la molécula de RNA (ácido ribonucleico).

---

<sup>3</sup>En las células procariontes (sin núcleo), el DNA se encuentra difundido a través de toda la célula. No obstante, al igual que en las células eucariontes, sólo participa en los procesos de replicación y transcripción.

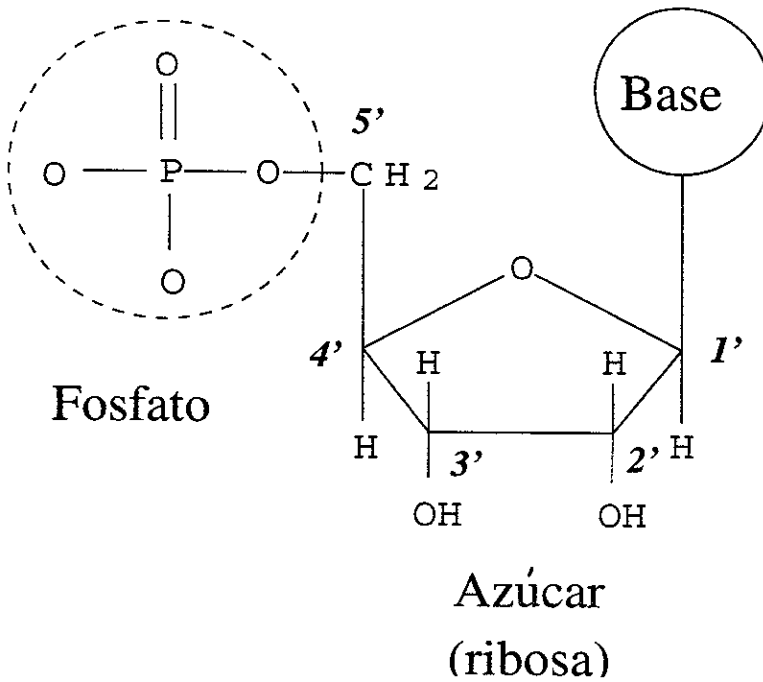


Figura 1.5: Estructura general de un nucleótido del RNA. Nótese que en la posición 2' del azúcar se encuentra un grupo OH, en lugar de un simple hidrógeno H, como en el caso del DNA.

El RNA es un polímero lineal que también está formado por cuatro tipos diferentes de nucleótidos: uracilfosfato (U), adenosinofosfato (A), citosinofosfato (C) y guanosinofosfato (G). Estos nucleótidos, como muestra la Fig1.5, tienen una estructura muy similar a los nucleótidos del DNA: un azúcar, un grupo fosfato y una base nitrogenada. De hecho, también hay interacciones complementarias entre ellos: A-U y C-G, como se muestra en la Fig1.6. Sin embargo, a pesar de la semejanza entre estos nucleótidos (A, U, C, G) y los correspondientes del DNA (A, T, C, G), existe una diferencia sustancial entre ellos: el azúcar de los nucleótidos del DNA no es la misma que la de los nucleótidos del RNA. En el DNA, el azúcar es una *desoxirribosa*, mientras que en el RNA el azúcar es una *ribosa*. Si inspeccionamos detenidamente las figuras 1.1 y 1.5, vemos que la diferencia entre estos dos azúcares radica en el carbono 2': la desoxirribosa tiene un simple hidrógeno (H) ligado a su carbono 2', mientras que la ribosa tiene ligado un grupo hidroxilo (OH) <sup>4</sup>.

<sup>4</sup>Aunque no es el caso para el RNA, vale la pena mencionar que algunas veces los gru-

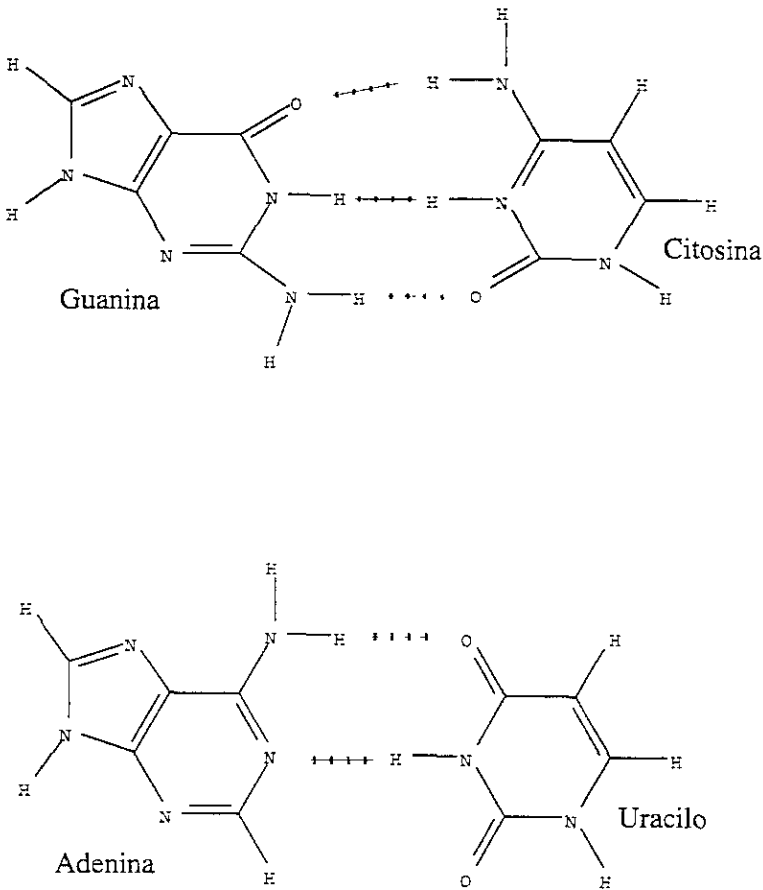


Figura 1.6: Las cuatro bases nitrogenadas que conforman al RNA y las interacciones complementaria entre ellas: A-U y C-G.

Aparte de lo anterior, existen otras diferencias notables entre el DNA y el RNA. Por ejemplo, el RNA en general es unicatenario, es decir, está compuesto por una sólo cadena que se enrolla sobre sí misma, como se ve esquemáticamente en la Fig1.7, mientras que el DNA es bicatenario. Además, el RNA es más susceptible a la hidrólisis que el DNA, lo que significa que tiende fácilmente a deshacerse en el agua. De hecho, los investigadores es-

pos hidroxilo (OH) pueden tener consecuencias catastróficas para los seres humanos. Por ejemplo, un sólo grupo OH de más en el aminoácido tirosina y de menos en el aminoácido fenilalanina, producen un mal funcionamiento en la enzima fenilalanina hidroxilasa, presente en el riñón, trayendo como consecuencia retraso mental en las personas afectadas.

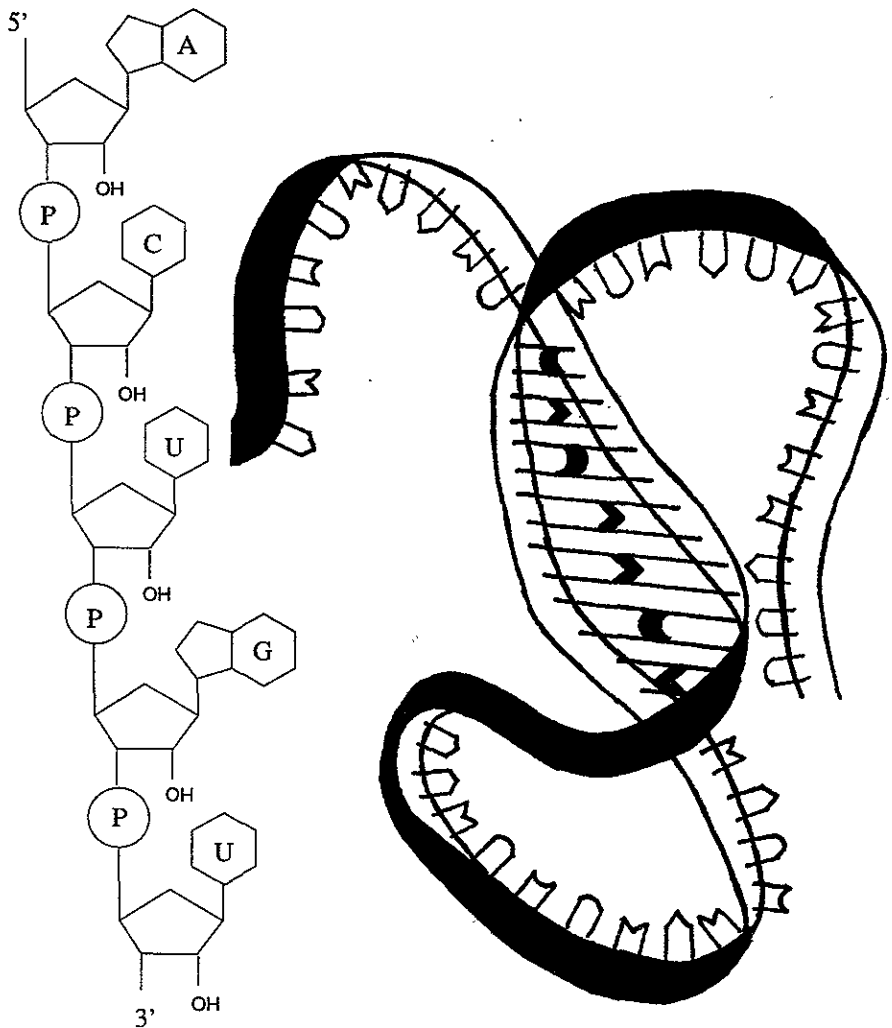


Figura 1.7: Diagrama esquemático de la estructura del RNA. En general, el RNA es una molécula unicatenaria, es decir, formada sólo por una cadena de nucleótidos, que puede plegarse sobre sí misma. Nótese la presencia del grupo OH en el carbono 2' del azúcar.

pecialistas que se ocupan de la síntesis de segmentos cortos de RNA tienen al agua por su enemigo natural y la tratan como tal, de manera que extremen en lo posible las precauciones para mantenerla lejos y excluirla de sus aparatos y reacciones. Los genotecnólogos califican a los nucleótidos y oligonucleótidos de RNA como “en extremo inconstantes químicamente” [10].

Resulta que, al contrario del DNA, la molécula de RNA es muy versátil, ya que participa en numerosas reacciones metabólicas. En la célula se han identificado varias funciones diferentes para las moléculas de RNA, y consecuentemente, a éste se le clasifica de acuerdo a la función que realiza dentro de la maquinaria celular:

- El **mRNA** o *RNA mensajero*, que transporta la información genética desde el núcleo celular hasta los ribosomas.
- El **rRNA** o *RNA ribosomal*, que constituye aproximadamente el 65% del material del que están compuestos los ribosomas (el otro 35% está constituido por proteínas).
- El **tRNA** o *RNA de transferencia*, que interviene directamente en la síntesis de proteínas añadiendo un aminoácido a la cadena polipeptídica en crecimiento.
- El **snRNA** o *RNA nuclear pequeño*, que participa decisivamente en la edición del mensaje genético, cortando los pedazos no codificadores de los genes (intrones).
- La **RNA telomerasa**, implicada en la replicación del DNA en los extremos de los cromosomas.
- El **RNA antisentido**, implicado en la regulación genética.

Como podemos ver, el RNA desempeña numerosas funciones en el metabolismo celular, todas ellas relacionadas directamente con la lectura e interpretación del mensaje genético contenido en el DNA. Más aún, se sabe ahora que incluso ciertas moléculas de RNA son capaces de generar otras moléculas de RNA a partir de sí mismas, o de degradarlas, sin la ayuda de ninguna enzima protéica. [11]. En lo que sigue abundaremos un poco más sobre las propiedades del RNA; por el momento, basta con mencionar que en la actualidad existe la hipótesis, cada vez más generalizada y fundamentada, de que el RNA jugó un papel preponderante en el origen de los seres vivos.



## 1.4 Proteínas: el "hardware" celular

Las proteínas también son polímeros lineales, formados a partir de 20 tipos diferentes de monómeros llamados *aminoácidos*, los cuales se van enlazando uno seguido de otro para construir la cadena protéica, o cadena peptídica. En la Fig1.8 se muestra la estructura general de los aminoácidos con los que se construyen las proteínas. Como puede verse en la figura, cada aminoácido está compuesto por un átomo de carbono central, al que se ligan un átomo de hidrógeno (H), un grupo amino ( $\text{NH}_2$ ), un grupo carboxilo ( $\text{COOH}$ ) y un radical  $R$  que es el que le confiere su identidad química al aminoácido. El proceso de polimerización de los aminoácidos se conoce desde 1900 gracias a los trabajos pioneros del químico alemán Emil Fischer, y consiste en que el grupo amino de un aminoácido reacciona con el grupo carboxilo de otro aminoácido en una reacción de deshidratación, en la que se libera una molécula de agua ( $\text{H}_2\text{O}$ ). El enlace covalente que se forma, denominado *enlace peptídico*, se muestra esquemáticamente en la Fig1.9. Lo importante es que los aminoácidos se van enlazando uno seguido de otro para formar la cadena protéica: dos aminoácidos enlazados constituyen un dipéptido, tres un tripéptido, etc. Si la cadena tiene más de 10 aminoácidos, se denomina polipéptido.

Aún cuando los aminoácidos se van enlazando uno seguido de otro de forma lineal para formar la cadena polipeptídica, resulta que las proteínas no son moléculas lineales. De hecho, se reconocen cuatro niveles de estructura protéica:

- **Estructura primaria.** Es la secuencia lineal de los aminoácidos que conforman la proteína. Esta secuencia está determinada por la secuencia de bases a lo largo del gen (o genes) de DNA que codifican para la proteína en cuestión.
- **Estructura secundaria.** Es la configuración espacial que adquieren los aminoácidos en la proteína debido a la interacción entre aminoácidos vecinos estrechamente ligados en la cadena protéica. En 1951, Linus Pauling y Robert Corey descubrieron que hay dos tipos de estructuras secundarias: la *hélice  $\alpha$*  y la *lámina  $\beta$*  (Fig1.10). En la primera, los aminoácidos se van enlazando formando una hélice, extendiendo sus grupos  $R$  hacia el exterior de ésta. Es como una varilla en la cual la estructura de hélice se estabiliza por medio de puentes de hidrógeno entre aminoácidos vecinos. Por otro lado, en la lámina  $\beta$  varias cadenas polipeptídicas se colocan paralela o antiparalelamente una al lado de otra, y se enlazan por medio de puentes de hidrógeno entre los aminoácidos vecinos, formando una estructura que se asemeja a un plano zigzagante.

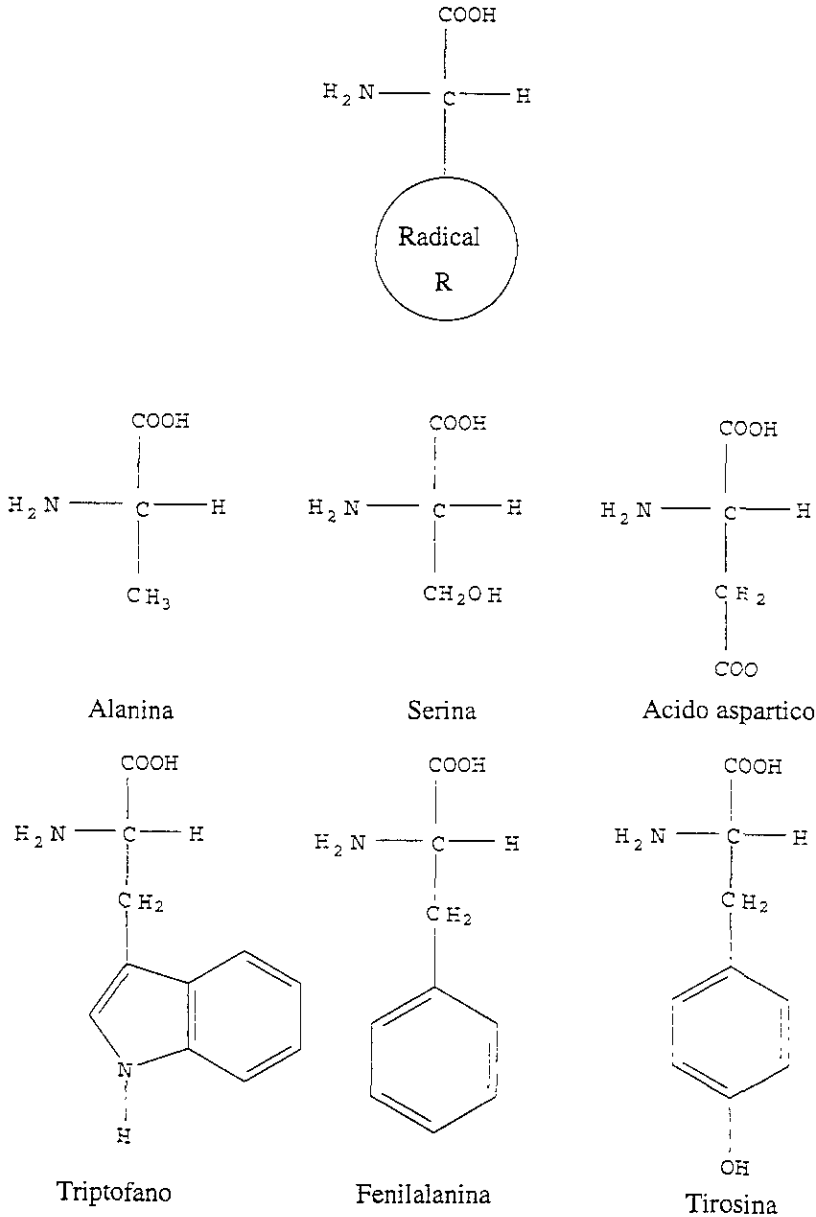


Figura 1.8: En la parte superior de la figura se muestra la estructura general de los aminoácidos: un carbono central unido a un grupo amino (NH<sub>2</sub>), un grupo carboxilo (COOH), un hidrógeno (H) y un radical (R). Lo que cambia entre los diferentes aminoácidos es el radical R, como puede observarse con los seis aminoácidos que se muestran.

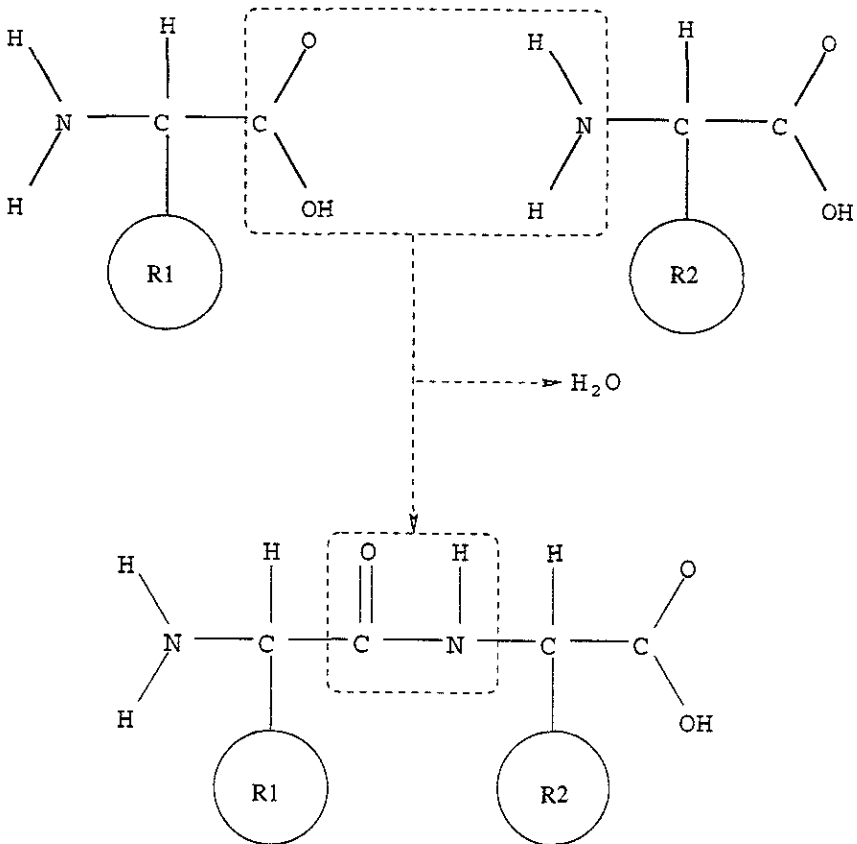


Figura 1.9: Formación del enlace peptídico entre dos aminoácidos como consecuencia de una reacción de deshidratación.

- **Estructura terciaria.** Es la configuración tridimensional que adquiere la proteína completa. Las hélices  $\alpha$  y las láminas  $\beta$  se pliegan sobre sí mismas o unas alrededor de las otras, y como consecuencia, partes que estaban muy alejadas en la cadena polipeptídica quedan unidas por medio de enlaces de hidrógeno. En gran medida, la funcionalidad de una proteína depende de su estructura tridimensional.
- **Estructura cuaternaria.** Es la estructura que se obtiene cuando varias proteínas se ligan entre sí para formar una macromolécula más grande, como ciertas enzimas. Este tipo de macromoléculas se denomina *oligómeros*, y cada cadena protéica de las que se conforma el

oligómero se llama *protómero* o, más comúnmente, *subunidad*.



Figura 1.10: Estructura secundaria de las proteínas: hélice alfa (izquierda) y lámina beta (derecha). Generalmente, todas las proteínas presentan ambas estructuras secundarias, ya que contienen dominios con hélices  $\alpha$  y dominios con láminas  $\beta$ .

Vale la pena mencionar que de estas cuatro estructuras protéicas, la estructura terciaria representa todo un misterio. Actualmente, no se sabe a ciencia cierta como predecir la forma tridimensional de una proteína a partir de la secuencia de aminoácidos que la conforman. Claramente, dicha estructura tridimensional depende del tipo de aminoácidos de los que está hecha la proteína y de cómo están secuenciados a lo largo de la cadena protéica, pero no se sabe aún cómo se da esta dependencia. Debido a que la secuencia de bases a lo largo de un gen de DNA determina unívocamente la secuencia de aminoácidos en la proteína codificada por este gen, lo anterior es equivalente a decir que no se sabe de qué manera la información contenida en el DNA determina la estructura y funcionalidad de las proteínas. Este problema, conocido como el **problema del plegamiento de las proteínas**, es uno de los más grandes retos de la biología molecular contemporánea.

Sin embargo, es claro que la estructura tridimensional de una proteína (y por lo tanto su funcionalidad), depende directamente de la secuencia de

los aminoácidos que la conforman, la cual, a su vez, está determinada por la secuencia de bases en el gen que codifica para dicha proteína. Es en este sentido en el que se dice que la información del metabolismo celular está contenida en el DNA, ya que las funciones metabólicas de las células se realizan por diferentes tipos de proteínas, cada una encargada de llevar a cabo una función específica.

## 1.5 El Dogma Central de la Biología Molecular

El proceso global que lleva la información del metabolismo de la célula, desde el DNA hasta las proteínas, se denomina **flujo de la información genética**. Hasta antes de 1960, la comunidad científica aceptaba que el flujo de la información genética en los organismos vivos consistía, básicamente, en cuatro etapas o procesos: **replicación, transcripción, edición y traducción**.

1. **Replicación (DNA  $\Rightarrow$  DNA)**. En la división celular, el DNA de la célula madre se duplica, dando como resultado dos juegos idénticos de cromosomas. De esta manera, tanto la célula madre como la célula hija contienen la misma información genética después de la división. En este proceso, una molécula de DNA se copia idénticamente a otra molécula de DNA.
2. **Transcripción (DNA  $\Rightarrow$  RNA)**. Es la primera etapa en el proceso de síntesis de una proteína. Aquí, una de las cadenas del gen de DNA que codifica para la proteína, se copia (o "transcribe") íntegramente a una molécula de RNA que porta la misma información que el gen original. La única diferencia entre la cadena de DNA y la cadena transcrita de RNA es que la primera se "escribe" con los cuatro símbolos *A*, *T*, *C* y *G*, mientras que la segunda se escribe con los símbolos *A*, *U*, *C* y *G* (la *U* en el RNA reemplaza a la *T* en el DNA). No obstante, las dos cadenas llevan la misma información.
3. **Edición (RNA  $\Rightarrow$  RNA)**. Esta etapa consiste en eliminar las secuencias no codificadoras (intrones) de la molécula de RNA generada en la transcripción. Como ya hemos mencionado, un gen de DNA contiene intrones y exones. En la transcripción, se copia dicho gen íntegramente a una molécula similar de RNA, incluyendo tanto las partes no codificadoras como las que sí codifican. Para que la información relevante en el RNA transcrito pueda ser leída por los

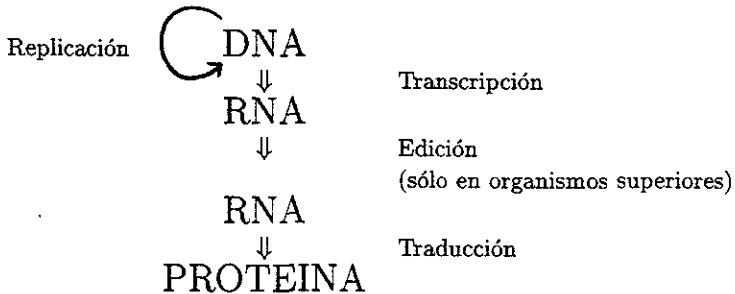
ribosomas, produciendo así la proteína, los intrones se deben “cortar”, quedando únicamente una cadena de RNA con puros exones. La molécula resultante de RNA con la información ya depurada (sin intrones), se denomina *RNA mensajero*, o más brevemente, *mRNA*. Esta es la molécula que finalmente sale del núcleo celular, dirigiéndose al retículo endoplasmático, para ser leída por los ribosomas. Un hecho interesante que vale la pena mencionar, es que la molécula encargada de “cortar” los intrones del RNA transcrito y unir después los exones formando el mRNA, es también una molécula de RNA, llamada *RNA nuclear pequeño* o *snRNA*.

4. **Traducción (RNA  $\Rightarrow$  Proteína).** Esta es la etapa final en la construcción de una proteína. En el retículo endoplasmático, los ribosomas se ligan a la cadena de mRNA y viajan a lo largo de ella, leyendo la información que trae consigo la secuencia de bases. Dependiendo de cómo estén secuenciadas las bases a lo largo del mRNA, los ribosomas van construyendo la cadena protéica, añadiendo uno por uno los aminoácidos especificados por el mRNA. Una vez que el ribosoma llega al final del mensaje, se desliga del mRNA y libera a la proteína ya formada, la cual es transportada a un lugar específico de la célula, donde realizará alguna función determinada.

Es importante mencionar que el proceso de edición no siempre está presente. El DNA de los virus no contiene secuencias intergénicas ni intrones, sino que todo el material genético es codificador. Por lo tanto, cuando un gen de estos organismos se transcribe, el RNA transcrito ya está listo para ser leído por los ribosomas, es decir, directamente de la transcripción emerge el mRNA.

De cualquier forma, haya o no haya edición del mensaje genético, los pasos anteriores en la manipulación de la información genética fueron tan importantes en la Biología Molecular, que en su conjunto llegaron a conocerse como “El Dogma Central de la Biología Molecular”. Antes de la década de los 60's se pensaba que el camino natural que tenía que recorrer la información genética para poder expresarse, era el descrito por estos cuatro procesos, los que podemos resumir esquemáticamente como:

## Dogma Central de la Biología Molecular antes de 1960.



es decir, el DNA puede copiarse en otro DNA (en la replicación celular), o puede copiarse en RNA (en la transcripción), para finalmente traducirse en proteína (en la traducción).

Nótese que en el esquema anterior, la información genética no puede fluir en ninguno de los siguientes sentidos: RNA→DNA, Proteína→RNA, Proteína→DNA. Antes de la década de los 60's, no se conocían organismos en los cuales la información contenida en una cadena de RNA sirviera para elaborar una cadena de DNA portadora de la misma información. Tampoco se conocían (ni se conocen todavía) organismos en los que la información genética fluya de proteínas a los ácidos nucleicos. Sin embargo, las cosas cambiaron drásticamente después de 1960.

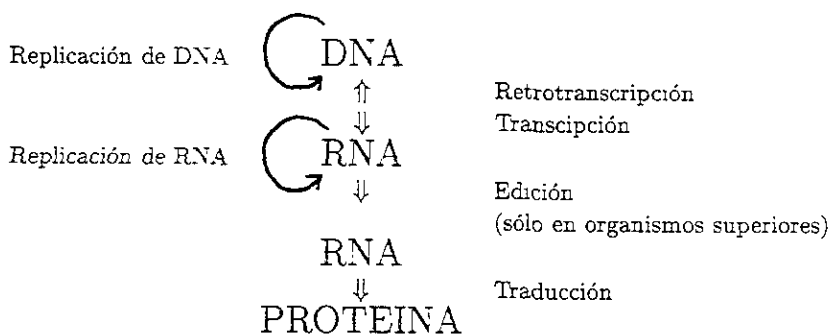
Ya desde 1962, estudiando virus que producían tumores cancerígenos en las aves, Howard Temin, de la universidad de Wisconsin, propuso la existencia de virus cuyo material genético está compuesto de RNA. Esto fue un choque tremendo para la comunidad científica de la época, ya que hasta entonces, el material genético de **todos** los organismos que se conocían estaba compuesto de DNA. Temin propuso que estos virus extraños cuentan con una enzima llamada *DNA polimerasa-RNA dirigida*, o mas simplemente *retrotranscriptasa*, que es capaz de hacer copias de DNA a partir de una cadena molde de RNA. La cadena de DNA así generada puede insertarse en el genoma de la célula infectada por el virus, y permanecer ahí durante años (como es el caso del virus que causa el SIDA).

La hipótesis de la retrotranscriptasa no fue probada experimentalmente sino hasta 1970, de modo simultáneo e independiente, por el propio Temin y por David Baltimore, quienes acuñaron el nombre de *retrovirus* para estos virus cuyo material genético consta de RNA y no de DNA como en los virus "normales". Más aún, en la misma década de los años 60's, Norman R. Pace y Sol Spiegelman demostraron que el material genético

del fago Q $\beta$  (un cierto virus que ataca a la bacteria *E. coli*), estaba hecho de RNA, y además, podía aislarse y replicarse *in vitro*. Seis años más tarde descubrieron que la replicación dependía de una enzima, la *RNA polimerasa-RNA dirigida*, (o más sucintamente, la *RNA replicasa*), que es capaz de transcribir moléculas de RNA copiándolas a otras moléculas complementarias también de RNA. A estos virus se les llamó *reovirus*.

Después de estos descubrimientos (y muchos otros que siguieron posteriormente), quedó claro que la información genética también puede fluir siguiendo las rutas RNA $\rightarrow$ DNA y RNA $\rightarrow$ RNA. Fue necesario entonces la reformulación del Dogma Central de la Biología Molecular para que éste tomara en cuenta los mecanismos de replicación de los retrovirus y los reovirus. Así, la extensión posterior del Dogma Central de la Biología Molecular sobre el flujo de la información genética puede resumirse como se muestra en el siguiente esquema:

### Dogma Central de la Biología Molecular en la Actualidad



Esta nueva concepción del flujo de la información genética trae como consecuencia un hecho de suma importancia: el RNA ya no es una molécula "subordinada" del DNA, sino que ahora ambas moléculas están al mismo nivel en el almacenamiento y procesamiento de la información genética. De hecho, veremos en el capítulo 3 que actualmente existe gran evidencia en favor de que, en el origen de la vida, fue el RNA el que jugó un papel preponderante, y no el DNA. Por el momento, concluiremos este capítulo haciendo una observación también muy importante: *la información genética no fluye de las proteínas a los ácidos nucleicos (DNA o RNA)*. Las proteínas siempre están al final del flujo de la información genética, y nunca al principio. Volveremos a este hecho en el capítulo siguiente.



## Capítulo 2

# El Código Genético

Ya desde 1932 George Beadle, analizando moscas *Drosophila* mutantes, propuso la hipótesis de que los genes especifican enzimas. Poco tiempo más tarde, el propio Beadle, en colaboración con Edward Tatum, se dedicó a estudiar las mutaciones que ocurrían en el hongo *Neurospora crassa* (el moho rosa del pan) cuando era irradiado con rayos X. Beadle y Tatum observaron que los hongos mutantes perdían la capacidad de llevar a cabo ciertas reacciones metabólicas, y demostraron que esto se debía a que los hongos mutados no sintetizaban las enzimas que participaban en dichas rutas metabólicas. Los numerosos experimentos que realizaron, en los que identificaron claramente las enzimas faltantes, proporcionaron pruebas experimentales convincentes de que **un gen especifica una enzima**. Con algunas modificaciones, este concepto se convirtió en un importante principio de la genética.

Ahora sabemos que lo que algunos genes especifican son proteínas o pedazos de proteínas (más precisamente, especifican cadenas polipeptídicas), que pueden o no ser enzimáticas<sup>1</sup>. El elaborado sistema físico-químico de las células tiene el potencial de producir una variedad casi ilimitada de moléculas de proteína a partir de la información contenida en los ácidos nucleicos. Además, este sistema debe ser muy preciso, ya que una proteína "mal hecha" puede traer como consecuencia enfermedades letales para los organismos. Tal es el caso de la *anemia falciforme*, una enfermedad de los seres humanos causada por la alteración de sólo uno de los casi 600 aminoácidos que conforman a la hemoglobina.

Debido a que las proteínas son el producto final en la expresión de la información genética, y son también las responsables de las características

---

<sup>1</sup>Otros genes especifican diferentes tipos de moléculas de RNA, como el rRNA o el tRNA.

fenotípicas de los seres vivos, una pregunta de fundamental importancia es cómo los ácidos nucleicos, con sólo cuatro tipos diferentes de nucleótidos, pueden especificar para proteínas, compuestas por veinte tipos diferentes de aminoácidos. En este capítulo veremos los mecanismos por medio de los cuales la información genética contenida en los ácidos nucleicos se convierte finalmente en proteínas. Al hacer esto, analizaremos la estructura del código genético y discutiremos los problemas que se presentan para explicar el origen de dicha estructura, en particular, el que nos ocupa a nosotros: ¿por qué las palabras del código están formadas por tres y sólo tres bases?

## 2.1 Las “palabras” del Código

El código genético es el “diccionario” que utiliza la Naturaleza para traducir la información genética y convertirla en proteínas. Inicialmente se pensaba que el DNA era capaz de especificar proteínas por sí mismo. Sin embargo, esta hipótesis llegó a ser insostenible, y en 1961 Francis Crick, François Jacob, Jacques Monod y otros autores propusieron la existencia del RNA mensajero (mRNA), el intermediario entre el DNA y las proteínas.

Una vez que quedó claro que el mRNA era la molécula directamente involucrada en la síntesis de las proteínas, el siguiente paso fue saber cómo se llevaba a cabo este proceso. La pregunta central radicaba en cómo sólo 4 letras (las cuatro bases del mRNA) pueden especificar para 20 palabras (los veinte aminoácidos de las proteínas). Sorprendentemente, la respuesta (correcta) a esta pregunta la dio por primera vez Sidney Brenner en 1961, utilizando argumentos teóricos verdaderamente sencillos.

Brenner dedujo que **tres bases en el mRNA especifican un aminoácido en la proteína**. Es decir, las “palabras” que especifican aminoácidos están compuestas por tres “letras”. A estas “palabras” que especifican aminoácidos también se les llama *codones*, y de ahora en adelante utilizaremos este término para referirnos a las “palabras genéticas”. El argumento de Brenner para concluir que los codones están compuestos por tres bases es el siguiente.

Si los codones estuvieran compuestos por sólo dos letras, se podrían formar solamente  $4^2 = 16$  combinaciones utilizando las cuatro bases del RNA, y 16 codones no alcanzan para especificar 20 aminoácidos. Si los codones estuvieran formados por tres letras, entonces existirían  $4^3 = 64$  combinaciones. Por otro lado, si los codones estuviesen formados por 4 letras, entonces el número de combinaciones posibles sería  $4^4 = 256$ . Aún cuando un código de tripletes proporciona 64 codones, evidentemente más de los 20 que se requieren para especificar a los aminoácidos, es mucho más sencillo que un código en el que los codones estuvieran formados por cuatro

(o más) letras, ya que en este último caso se tendría un número excesivo de codones (más de 256) para especificar sólo 20 aminoácidos.

Por lo tanto, de acuerdo con el argumento de Brenner, **el número óptimo de bases por codón, para especificar 20 aminoácidos, es tres**. Por supuesto, en su razonamiento Brenner estaba haciendo una hipótesis que, *a priori*, no podía saber si era válida o no, a saber, que *todos los codones contienen el mismo número de letras*. Muy bien pudo haber ocurrido que algunos aminoácidos estuvieran especificados por codones con tan sólo dos letras, mientras que otros fueran especificados por codones de tres, cuatro o cinco letras, etc. Además, el argumento de Brenner depende fuertemente de que son cuatro y sólo cuatro los tipos diferentes de bases que conforman al DNA. No obstante, trabajos posteriores demostraron que Brenner estaba en lo cierto respecto a que todos los codones están formados por tres letras.

El hecho es que para 1961 se contaba con la hipótesis de que tres bases en el mRNA especificaban un aminoácido en la cadena protéica, aunque no se sabía de qué forma se daba esta especificidad. Surgieron entonces muchas interrogantes. Por ejemplo, 20 de los 64 codones especificaban aminoácidos, ¿y los otros 44? ¿Estos 44 codones no tenían significado? ¿O bien, el código era degenerado y varios codones servían para especificar el mismo aminoácido? Otra interrogante era si los codones a lo largo del mRNA se intersectaban (solapaban) o si se seguían uno después del otro sin solaparse. Además, no se sabía si el código era distinto para diferentes organismos, o si era universal, en el sentido de que todos los organismos vivos tuvieran el mismo código.

El propio Brenner y Francis Crick dieron varios argumentos teóricos en favor de que el código no es solapado. Entre ellos, destaca la observación, hecha por Crick, de que los aminoácidos no reaccionan directamente con los ácidos nucleicos. Por lo tanto, debía existir una **molécula adaptadora** que, por un lado se ligara a un codón en el mRNA, y por el otro se ligara al aminoácido especificado por este codón. Si esto era cierto, los codones en el mRNA no podían intersectarse, ya que de haberlo hecho las moléculas adaptadoras quedarían una encima de otra, lo cual no parecía plausible. Esta predicción de Crick fue comprobada más tarde, cuando se descubrió la existencia de la molécula adaptadora, la cual, además, resultó estar hecha de RNA. A esta molécula se le conoce ahora como *RNA de transferencia*, o simplemente, como *tRNA*.

La serie de descubrimientos que se originaron a partir de la hipótesis de Brenner es realmente inacabable. El lector interesado puede acudir al libro de Judson [6] para abundar en la historia del desciframiento del código genético. Basta con decir aquí que la naturaleza del código quedó resuelta gracias a los trabajos pioneros que realizaron Marshall Nirenberg

y J. Henrich Matthai. Gracias a estos trabajos, y a otros que vinieron posteriormente, sabemos ahora que efectivamente el código genético está constituido por codones de tres bases que no se solapan; que es un código degenerado, puesto que varios codones especifican el mismo aminoácido; que todos los 64 codones tienen significado; y que es universal, en el sentido de que la mayoría de los organismos vivos utilizan el mismo código.

La tabla I muestra el código genético tal y como se le conoce en la actualidad. Podemos ver que el código es degenerado. Por ejemplo, los codones *UCU*, *UCC*, *UCA* y *UCG* especifican todos el aminoácido serina. Como puede verse, todos estos tripletes se diferencian sólo en la última letra. Si inspeccionamos detenidamente la tabla I, veremos que este es el caso en general: *casi siempre, los codones que especifican el mismo aminoácido se diferencian en la última posición*. Este interesante patrón indujo a Crick a formular la **hipótesis del tambaleo** (o *wobble*), a la cual volveremos en la siguiente sección.

También vemos en la tabla I que el código no contiene ambigüedades, en el sentido de que cada triplete especifica un único aminoácido. Además, todos los codones especifican para algún aminoácido, excepto los tripletes *UAA*, *UAG* y *UGA*, los cuales indican el fin del mensaje genético. Por eso están marcados con "stop" en la tabla. Otro triplete de especial importancia es el *AUG*, el cual, además de especificar para el aminoácido metionina, indica también el inicio del mensaje genético (otro codón que algunas veces indica el inicio del mensaje genético, es el codón *GUG*, que normalmente especifica el aminoácido valina).

La tabla I representa el diccionario utilizado por la célula para traducir la información genética contenida en el mRNA, convirtiéndola en proteínas<sup>2</sup>. De acuerdo con este código, la secuencia genética

*AUG.UCG.GGA.UAUGCU.ACC.GGA.AUA.UAG*

se traduciría en la secuencia polipeptídica

**met - ser - gly - tyr - ala - thr - gly - ile - stop**

<sup>2</sup>Como el mRNA es la molécula directamente involucrada en la síntesis de proteínas, el código genético se establece utilizando las cuatro bases del mRNA. Sin embargo, debemos recordar que el mRNA no es más que una copia de un pedazo de DNA, y sirve como mensajero para llevar la información genética desde el núcleo celular hasta los ribosomas.

TABLA I

	U	C	A	G	
U	UUU <i>phe</i>	UCU <i>ser</i>	UAU <i>tyr</i>	UGU <i>cys</i>	U
	UUC <i>phe</i>	UCC <i>ser</i>	UAC <i>tyr</i>	UGC <i>cys</i>	C
	UUA <i>leu</i>	UCA <i>ser</i>	UAA <b>stop</b>	UGA <b>stop</b>	A
	UUG <i>leu</i>	UCG <i>ser</i>	UAG <b>stop</b>	UGG <i>trp</i>	G
C	CUU <i>leu</i>	CCU <i>pro</i>	CAU <i>his</i>	CGU <i>arg</i>	U
	CUC <i>leu</i>	CCC <i>pro</i>	CAC <i>his</i>	CGC <i>arg</i>	C
	CUA <i>leu</i>	CCA <i>pro</i>	CAA <i>gln</i>	CGA <i>arg</i>	A
	CUG <i>leu</i>	CCG <i>pro</i>	CAG <i>gln</i>	CGG <i>arg</i>	G
A	AUU <i>ile</i>	ACU <i>thr</i>	AAU <i>asn</i>	AGU <i>ser</i>	U
	AUC <i>ile</i>	ACC <i>thr</i>	AAC <i>asn</i>	AGC <i>ser</i>	C
	AUA <i>ile</i>	ACA <i>thr</i>	AAA <i>lys</i>	AGA <i>arg</i>	A
	AUG <b>met</b>	ACG <i>thr</i>	AAG <i>lys</i>	AGG <i>arg</i>	G
G	GUU <i>val</i>	GCU <i>ala</i>	GAU <i>asp</i>	GGU <i>gly</i>	U
	GUC <i>val</i>	GCC <i>ala</i>	GAC <i>asp</i>	GGC <i>gly</i>	C
	GUA <i>val</i>	GCA <i>ala</i>	GAA <i>glu</i>	GGA <i>gly</i>	A
	GUG <i>val</i>	GCG <i>ala</i>	GAG <i>glu</i>	GGG <i>gly</i>	G

Un hecho verdaderamente sorprendente es que, salvo muy pocas excepciones<sup>3</sup>, todos los organismos vivos utilizan el mismo código, desde los virus hasta los seres humanos, pasando por arqueobacterias, bacterias, hongos, algas, plantas, insectos, reptiles, mamíferos, etc. Francis Crick concluyó que esta "universalidad" del código genético es una evidencia clara de que todos los organismos vivos provenimos de un ancestro común, una **célula precursora** que hace miles de millones de años estuvo bajo condiciones muy favorables que le permitieron multiplicarse y proliferar por todo el planeta, dando lugar a todos los organismos que se conocen ahora (y a los que se extinguieron sin que llegáramos a conocerlos).

<sup>3</sup>En las mitocondrias celulares el código está ligeramente alterado, ya que ahí el codón *UGA*, que normalmente provoca la terminación del mensaje genético, es utilizado para especificar el aminoácido *triptofano*. Se han encontrado también codones alterados del código en la bacteria *Mycoplasma capricolum* y en genes nucleares de los protistas ciliados *Paramecium*, *Tetrahymena* y *Stylonychia*. Sin embargo, estos cambios son pequeños (aunque no dejan de ser importantes). Mientras no se conozcan más ejemplos, estos cambios pueden considerarse como alteraciones pequeñas en el código universal.

## 2.2 tRNA: la molécula adaptadora

La hipótesis formulada por Crick sobre la existencia de la molécula adaptadora fue confirmada experimentalmente por Robert Holley y sus colaboradores en 1965. A partir de entonces, quedó establecido que dicha molécula es una cadena de RNA (el tRNA al que nos referimos antes), compuesta por aproximadamente 80 nucleótidos, que se pliega sobre sí misma dando una estructura tridimensional en forma de "L" (ver la Fig2.1). En uno de los extremos de la L (el extremo "grande") se liga un aminoácido particular, mientras que en el otro extremo (el extremo "pequeño"), hay una secuencia característica de tres bases denominada *anticodón*, y es precisamente el anticodón el que se une al RNA mensajero.

Aminoácido

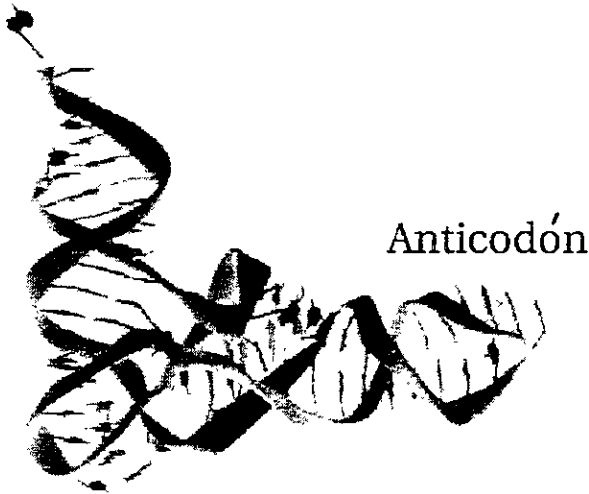


Figura 2.1: Dibujo esquemático del tRNA, en donde se muestra su estructura tridimensional en forma de "L". En uno de los extremos está el anticodón, mientras que en el otro extremo se liga el aminoácido correspondiente.

Hay por lo menos una molécula de tRNA por cada aminoácido; algunos aminoácidos poseen dos o más tRNA's específicos. De hecho, ahora se sabe que en las células eucariontes existen entre 40 y 50 tRNA's diferentes. A los tRNA's se les denota según el aminoácido que portan, de acuerdo a la notación tRNA<sup>aa</sup>, siendo "aa" el nombre del aminoácido. Así, al tRNA portador del triptofano se le denomina tRNA<sup>trp</sup>, mientras que al tRNA portador de la alanina se le designa por tRNA<sup>ala</sup> etc. El anticodón del tRNA<sup>aa</sup> es un triplete de bases complementarias<sup>4</sup> al codón que especifica el aminoácido "aa" en el mRNA. Por ejemplo, el anticodón del tRNA<sup>glu</sup> es CUU, que es un triplete complementario al codón GAA que codifica la glutamina; el anticodón del tRNA<sup>val</sup> es CAG, que es complementario al codón GUC que especifica el aminoácido valina, etc.

El hecho importante es que el aminoácido transportado en el extremo "largo" del tRNA depende del anticodón en el extremo "corto". De qué manera el anticodón hace que el tRNA reconozca al aminoácido que debe transportar, es un problema que no se ha resuelto satisfactoriamente. Sobre todo, porque se ha visto que un tRNA en particular puede funcionar con diferentes aminoácidos. Por ejemplo, en 1966 F. Chapeville realizó un experimento en el cual, al tRNA<sup>cys</sup> se le cambió artificialmente el aminoácido cisteína (que normalmente lleva cargado) por el aminoácido alanina. Se produjo así la molécula *ala*-tRNA<sup>cys</sup>, que tiene el anticodón correspondiente a la cisteína pero que porta el aminoácido alanina. Se observó que estos tRNA's modificados funcionaban como si nada hubiera pasado, e intervenían normalmente en la síntesis de proteínas. Lo único que ocurría es que en la proteína sintetizada, el aminoácido cisteína estaba reemplazado por el aminoácido alanina. Experimentos posteriores demostraron resultados similares con diferentes tRNA's modificados. Sin embargo, en la célula los tRNA's siempre (o casi siempre) se cargan con el aminoácido correcto<sup>5</sup>.

Debido a que los enlaces de hidrógeno entre las bases se establecen sólo si éstas son complementarias, el anticodón en el tRNA solamente se une al codón correspondiente en el mRNA. La Fig.2.2 muestra esquemáticamente cómo los tRNA's se van ligando a la molécula de RNA mensajero para formar la proteína.

Vale la pena hacer notar que existe una discrepancia entre el número de codones y el número de tRNA's. Como sabemos, hay 61 codones que especifican aminoácidos. Sin embargo, solamente existen entre 40 y 50 tRNA's distintos en las células eucariontes. Si la interacción entre el mRNA y el tRNA es a través del codón y el anticodón, se esperaría que existieran

<sup>4</sup>Complementarias en el sentido de enlaces Watson-Crick. Recordemos que la A y la T son complementarias, así como la G y la C.

<sup>5</sup>Algunas veces existen "errores" en la carga de los tRNA's, lo que origina proteínas "defectuosas"

tantos de unos como de otros. Para cada codón en el mRNA, debería existir un tRNA con el anticodón correspondiente. De esta forma se garantizaría que los 20 aminoácidos estuvieran especificados por los 61 codones. Sin embargo, esto no es así.

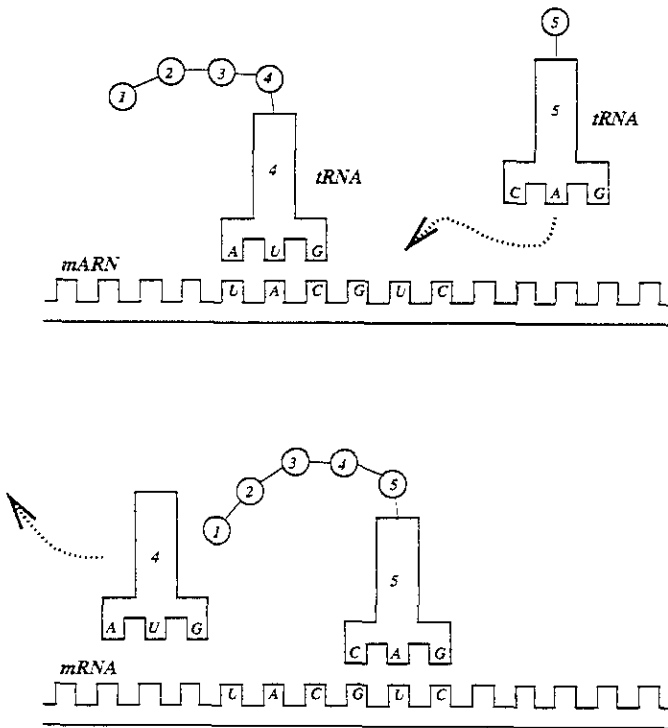


Figura 2.2: Representación esquemática de la interacción complementaria del RNA de transferencia con el RNA mensajero. El tRNA se liga por medio del anticodón al mRNA en el sitio correspondiente. Los tRNA's cargados con su respectivo aminoácido, se van uniendo uno seguido de otro, a la cadena de mRNA, añadiendo su correspondiente aminoácido a la cadena polipeptídica en formación, después de lo cual se liberan.



Una posible explicación al problema anterior radica en la *hipótesis del tambaleo*, formulada por Crick en 1966. Como hemos visto, cuando hay varios codones que especifican el mismo aminoácido, en general estos codones sólo difieren en la tercera base. Esto significa que, en general, bastan sólo dos bases (las dos primeras) para especificar un aminoácido, siendo la tercera base redundante. Así, aparentemente, las dos primeras bases del codón son más importantes que la tercera para atraer el tRNA correcto. Crick concluyó que el puente de hidrógeno en la tercera posición de la interacción codón-anticodón no necesita ser tan específico como en las dos primeras posiciones, y propuso un nuevo conjunto de reglas de apareamiento de bases (diferentes a las interacciones complementarias) para la tercera posición del codón. De acuerdo con estas reglas, una U en la tercera posición del anticodón puede aparearse con una A o una G en la tercera posición del codón, mientras que una G puede aparearse con una U o una C.

Esta relajación en el apareamiento de bases, o “tambaleo”, permite que el anticodón de un único tipo de tRNA se aparee con más de un codón del mRNA. Aplicando estas reglas del tambaleo, se necesitan por lo menos 30 tipos diferentes de tRNA's para reconocer a los 61 codones codificadores. Si no hay más, el tambaleo puede considerarse como una forma de economizar óptimamente a los tRNA's, lo cual, a su vez, equivale a optimizar la información genética. Las estimaciones experimentales actuales dan como resultado entre 30 y 40 tipos diferentes de tRNA en bacterias, y hasta 50 tipos diferentes en células animales y vegetales.

## 2.3 Ribosomas: fábricas de proteínas

Las moléculas que directamente traducen la información del mRNA y la convierten en proteínas, son los ribosomas, a los cuales “les entra” por un lado, el mRNA, y “les sale”, por el otro, la proteína especificada. Los ribosomas son moléculas compuestas por RNA (60%) y por proteínas (40%); están formados por dos subunidades, una grande y otra pequeña, denominadas *60S* y *40S* respectivamente (debido a su velocidad de sedimentación, dada en unidades de *Svedbergs*; mayor S implica mayor masa)<sup>6</sup>. Normalmente, las dos subunidades del ribosoma se encuentran separadas en el citoplasma, y sólo se unen para leer la información en la cadena de mRNA. Una vez que se ha terminado la lectura del mensaje genético, las

<sup>6</sup>No todos los ribosomas son iguales. En las bacterias el ribosoma completo sedimenta a 70S, mientras que la tasa de sedimentación de las subunidades son de 50S y 30S. Los ribosomas de las células eucariontes de los mamíferos sedimentan a 80S y sus subunidades lo hacen a 60S y 40S.

dos subunidades se vuelven a separar.

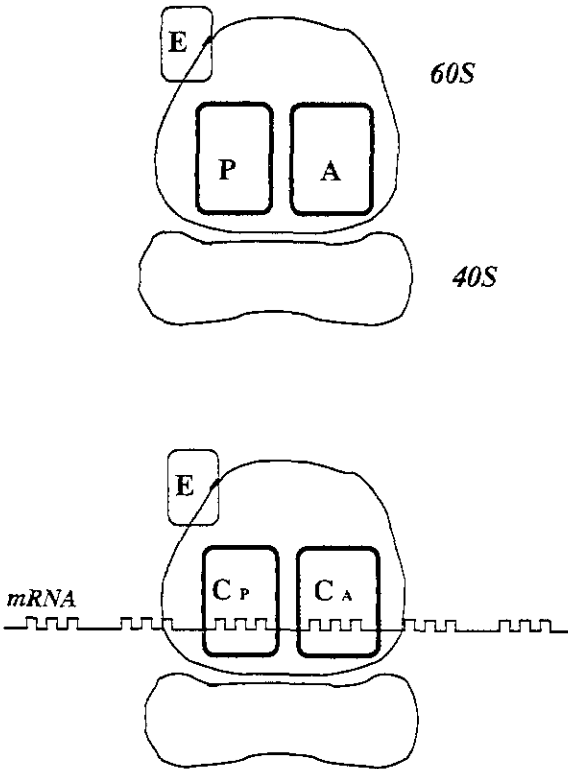


Figura 2.3: Representación esquemática del ribosoma. En los sitios activos P y A quedan expuestos dos codones consecutivos del mRNA, a los que se ligan los tRNA's con el correspondiente anticodón. El sitio E es a través del que se expulsan a los tRNA's que ya han liberado su aminoácido.

El ribosoma contiene tres sitios activos, a saber, el sitio **P** en el que se encuentra un tRNA de transferencia con la cadena polipeptídica en formación, el sitio **A**, en el que entra un tRNA cargado con el nuevo aminoácido que se va a añadir a la cadena, y el sitio **E** que es por donde se expulsan a los tRNA's que ya han liberado su aminoácido<sup>7</sup>. Cuando el ribosoma está ligado al mRNA, quedan expuestos dos codones consecutivos, uno en

<sup>7</sup>Cabe mencionar que estos "sitios" a los que nos referimos, más que ser lugares físicos dentro del ribosoma, son estados configuracionales de esta molécula. Sin embargo, para nuestros propósitos, es útil imaginarnos a estos estados configuracionales como si fueran lugares separados y bien definidos dentro del ribosoma.

el sitio P y otro en el sitio A (Fig2.3). En lo que sigue llamaremos  $C_P$  al codón del mRNA que está expuesto en el sitio P, y  $C_A$  al que queda expuesto en el sitio A.

El proceso de traducción, que se ilustra esquemáticamente en la Fig2.4, comienza cuando el ribosoma se liga al mRNA justo en el codón de inicio AUG (que se encuentra cerca del extremo 5' de la cadena de mRNA), de tal manera que dicho codón queda expuesto en el sitio P. En el sitio A queda expuesto el codón consecutivo al codón de inicio. Una vez ligado el ribosoma al mRNA, en el sitio P entra un tRNA cuyo anticodón es complementario al codón  $C_P$ , mientras que en el sitio A entra un tRNA cuyo anticodón es complementario al codón  $C_A$ . Estos dos tRNA's están cargados cada uno con su respectivo aminoácido, y ya dentro del ribosoma, dichos aminoácidos se ligan covalentemente formando un dipéptido. Después de esto, el tRNA del sitio P es expulsado del ribosoma (a través del sitio E), el cual avanza un codón en la dirección 3' del mRNA. De esta manera, el tRNA que antes estaba en el sitio A se mueve al sitio P, y el sitio A queda liberado. Un nuevo tRNA cuyo anticodón es complementario al codón  $C_A$  recién liberado, se introduce al sitio A, y otra vez, el aminoácido de este nuevo tRNA se une covalentemente con el dipéptido del sitio P, formando un tripéptido (que permanece ligado al tRNA del sitio A). De nuevo, el tRNA del sitio P es expulsado, el ribosoma avanza otro codón liberando el sitio A, en el cual vuelve a entrar otro tRNA con anticodón complementario al codón  $C_A$  recién liberado, se forma entonces un cuatripéptido y todo el ciclo se repite. Así, la cadena polipeptídica se va construyendo en el ribosoma de acuerdo a la secuencia de los codones  $C_A$  del RNA mensajero que se van liberando en el sitio A.

Es importante mencionar que hasta hace poco, se pensaba que la molécula encargada de ligar los aminoácidos covalentemente dentro del ribosoma era una enzima proteínica. Sin embargo, Thomas A. Steitz y su grupo de la Universidad de Yale y del Howard Hughes Medical Institute (New Haven), mostraron en el año 2000 que la unión covalente de los aminoácidos la lleva a cabo *el RNA del ribosoma, sin la ayuda de ninguna enzima proteínica* [12], [13]. Este descubrimiento es de la mayor importancia, porque muestra una vez más la gran versatilidad del RNA, apoyando así la hipótesis de que el RNA fue la molécula de mayor importancia en el origen de la vida. Volveremos a este punto en el capítulo siguiente.

El proceso de traducción se termina cuando en el sitio A queda expuesto alguno de los codones de término (*UAA, UAG o UGA*). Cuando esto ocurre, el sitio A se "bloquea" impidiendo la entrada de otro tRNA, y el ribosoma se desprende del mRNA y de la proteína recién formada, la cual ahora viajará a algún lugar determinado de la célula para comenzar con sus funciones de proteína.

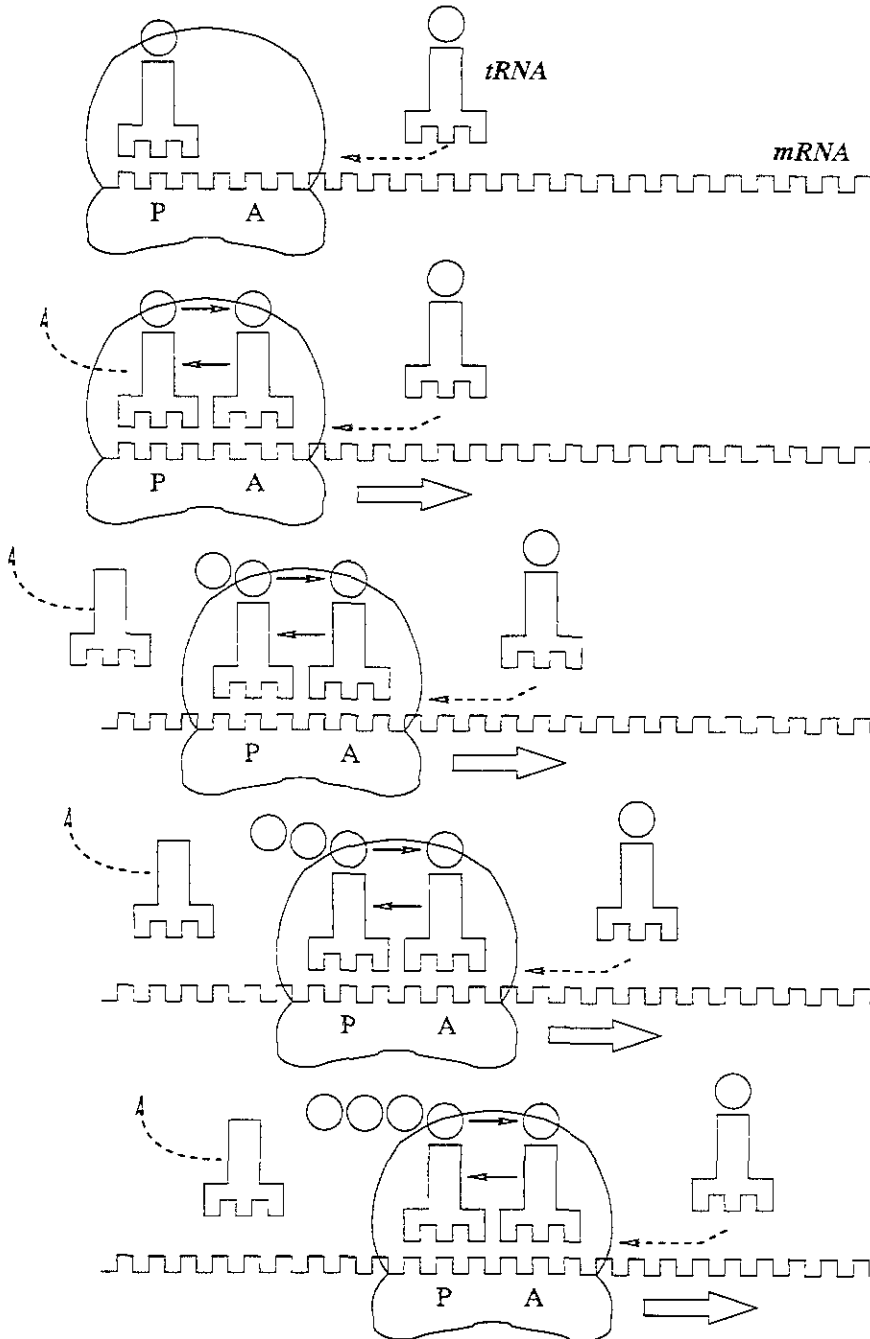


Figura 2.4: Representación esquemática del proceso de síntesis de proteínas en el ribosoma. Cada que se añade un aminoácido a la cadena polipeptídica en formación, el ribosoma avanza tres bases en el mRNA, capturando en el sitio A una tRNA cargada con un aminoácido.

El proceso de “reconocimiento de codones” a través del cual la información genética se transfiere del mRNA a las proteínas está basado en la formación de enlaces complementarios entre el anticodón del tRNA y el codón correspondiente del mRNA. Sin embargo, la manera como se van ordenando los tRNA's a lo largo del RNA mensajero está dirigida por los ribosomas. Estas moléculas operan con una eficiencia sorprendente: en un segundo un solo ribosoma puede añadir hasta 20 aminoácidos a la cadena polipeptídica en formación, de tal suerte que en 30 segundos se puede formar una proteína como la hemoglobina, que consta de unos 600 aminoácidos. El hecho importante es que en cada paso, el ribosoma avanza un codón (es decir, tres bases), dejando entrar al tRNA específico del codón  $C_A$ , cargado con su correspondiente aminoácido.

## 2.4 ¿Por qué tres?

Efectivamente, ¿por qué los codones están formados por tres bases? De acuerdo con el argumento de Brenner, esto es porque con tres bases por codón se “optimiza” el número de palabras que especifican aminoácidos: hay las suficientes para cubrir los 20 aminoácidos, pero no hay tantas como para tener una degeneración excesiva. Si bien es cierto que con este argumento lógico Brenner llegó a “predecir” la estructura de tripletes del código genético, evidentemente no es una respuesta satisfactoria a la pregunta de por qué tres.

Vale la pena mencionar que el argumento de Brenner, inicialmente formulado como un criterio de plausibilidad para dilucidar la estructura del código genético, ha llegado a convertirse en la explicación “natural” que presentan algunos libros (muy respetables) de Bioquímica y Biología Molecular para justificar la estructura en tripletes del código [14]. Sin embargo, no es difícil darse cuenta de que dicho argumento, aún cuando pueda ser “lógico” y “natural”, no constituye de ninguna manera una explicación completa de por qué la información en el mRNA está codificada en tripletes de bases<sup>8</sup>. Los problemas que se presentan con este argumento de “optimización” de la información genética son varios:

- Se supone desde el principio que *todas* las palabras del Código están formadas por el mismo número de bases. ¿Por qué? ¿Por qué no pueden haber palabras más largas que otras?
- Es cierto que tres bases por codón hacen más eficiente la codificación de 20 aminoácidos. Pero entonces, ¿por qué sólo 20 aminoácidos? En

---

<sup>8</sup>El argumento de Brenner es en buena medida una de esas ideas afortunadas que no tenían por qué funcionar, pero que sin embargo, funcionó.

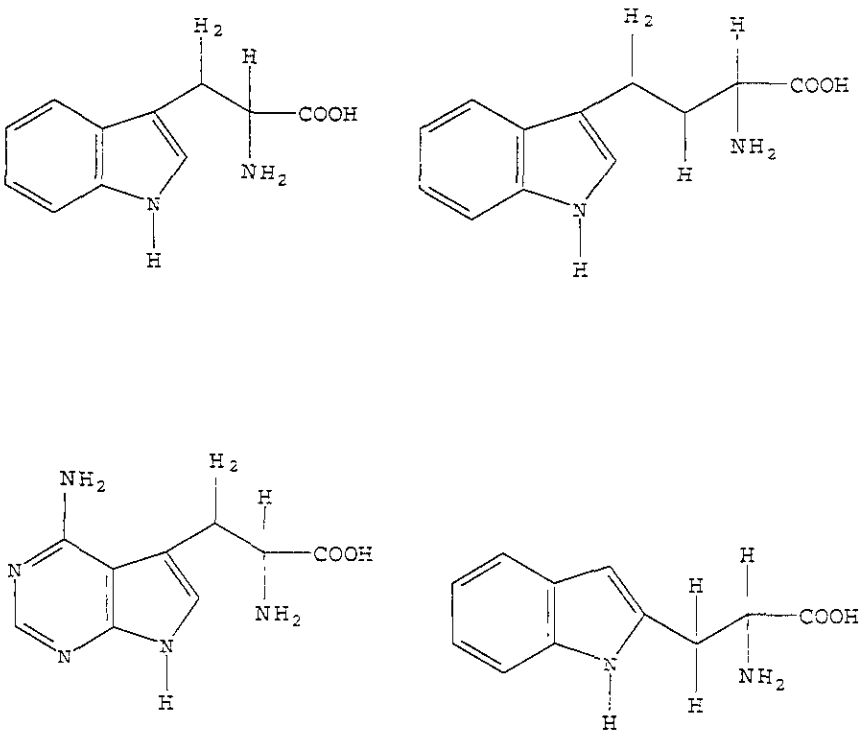


Figura 2.5: De los cuatro aminoácidos mostrados, solamente el primero (triptofano) es sintetizado por las células, a pesar de que todos ellos tienen propiedades químicas muy similares.

la Naturaleza existen mucho más aminoácidos que los 20 que utiliza la célula. Por ejemplo, de los cuatro aminoácidos mostrados en la Fig.2.5 solamente el primero (triptofano) es sintetizado por las células, aunque los cuatro tienen propiedades químicas muy similares. Así, la respuesta basada en la "eficiencia" sólo desplaza la pregunta de por qué tres a por qué 20.

- Se cree que debido a sus afinidades químicas (polaridad, hidrofobicidad, neutralidad, etc.), se requieren menos de 20 aminoácidos para que las células lleven a cabo sus funciones metabólicas [24]. Si la naturaleza estuviera "buscando" un proceso de optimización, entonces probablemente con menos de 15 aminoácidos hubiera sido suficiente, para lo cual se requeriría de sólo dos bases por codón.

- Si la naturaleza no está “buscando” optimizar el número de aminoácidos, ¿por qué sí lo está haciendo con el número de bases por codón?
- El argumento de Brenner depende fuertemente de que sean cuatro, y solamente cuatro, las bases que conforman el DNA. ¿Por qué cuatro?
- La explicación basada en la “optimización” del número de bases por codón para especificar 20 aminoácidos, sugiere que en origen de la vida, cuando apenas se estaban llevando a cabo los procesos prebióticos y protobióticos que eventualmente condujeron a los seres vivos, las proteínas eran las moléculas de mayor importancia, a las cuales después se les acoplaron los ácidos nucleicos, estructurando la información genética para especificar los 20 aminoácidos de las proteínas ya existentes. Sin embargo, ahora se cuenta con evidencia (y cada vez hay más) de que fueron los ácidos nucleicos, en particular el RNA, las moléculas prebióticas precursoras de los organismos vivos, y no las proteínas.

El problema en el último punto es que si tomamos el argumento de Brenner como una “explicación”, estamos asumiendo tácitamente que en el proceso evolutivo surgieron primero las proteínas, formadas por 20 aminoácidos, y después aparecieron el DNA y el RNA, cuya información tuvo que amoldarse y estructurarse a los 20 aminoácidos ya existentes. Lo anterior no parece estar de acuerdo con el flujo de la información genética *observado* en los organismos vivos ni con la evidencia cada vez más contundente de que la evolución de las moléculas biológicas se dió precisamente en el sentido opuesto, comenzando en los ácidos nucleicos y terminando en las proteínas<sup>9</sup>.

Por otro lado, la suposición de que la naturaleza “optimizó” la estructura de la información genética utilizando tripletes, tampoco es cierta. En 1997, Andrew J. Doig de la universidad de Manchester, demostró que se puede construir un código mucho más eficiente y con menos redundancia si los codones tienen longitudes variables de entre 2 a 4 bases [15]. De hecho, él en su artículo le llama a este código “El Código Perfecto”. Si es o no “perfecto” no tiene relevancia. Lo importante es que existe un código más eficiente que el que utiliza la naturaleza en el sentido de que con menos codones se puede codificar más información. ¿En qué momento deja la naturaleza de “optimizar” sus procesos?

Explicar la estructura del código genético es un problema abierto que aún no tiene solución. No sólo está la cuestión de por qué los codones están

---

<sup>9</sup>En el capítulo siguiente veremos la hipótesis del mundo del RNA, formulada independientemente por L. Orgel, C. R. Woese y F. Crick.

formados por tres bases. Hay otras preguntas también, como por ejemplo: ¿por qué hay 20 aminoácidos? ¿Cómo se determinó qué codón específica para qué aminoácido? ¿Por qué hay tres codones de término y uno (a veces dos) codones de inicio? ¿Qué determina la degeneración del código? ¿Por qué a veces hay 6 codones que especifican un aminoácido (leucina, serina) y a veces sólo hay uno (triptofano)? ¿Por qué la degeneración se da casi siempre en la tercera base del codón, y no en la primera o en la segunda?

Ninguna de las preguntas anteriores cuenta todavía con respuestas satisfactorias. Algunas veces la gente dice que la naturaleza “trabaja” con lo que tiene. Esto, junto con un poco de azar y con la selección natural dirigiendo los procesos evolutivos, es más que suficiente para explicar las estructuras y el funcionamiento de los seres vivos. Probablemente sea así. No obstante, si bien es cierto que el azar puede ser una componente importante en la determinación de los procesos biológicos, veremos en los siguientes capítulos que, respecto a la estructura de tripletes del código genético, los “datos” que arrojó la Naturaleza estaban “cargados”. Es decir, la estructura de tripletes era más una necesidad que una casualidad. Por lo pronto, en la siguiente sección discutiremos algunos de los trabajos recientes que se han hecho referentes al origen del código genético.

## 2.5 Algunas ideas sobre el origen del Código

El problema del origen del código genético ha sido abordado desde diversos puntos de vista. De acuerdo con Massimo Di Giulio [16], las teorías sobre el origen del código que en la actualidad cuentan con mayor aceptación son: la **teoría estereoquímica**, la **teoría fisicoquímica**, la **teoría de reducción de ambigüedades**, y la **teoría de coevolución**.

- **Teoría estereoquímica.** Esta teoría sostiene que la estructura del Código Genético se debe a las interacciones estereoquímicas entre los codones, anticodones y los aminoácidos. En otras palabras, esta hipótesis establece que la lisina, por ejemplo, está codificada por los codones *AAA* y *AAG* debido a que existe una correlación estereoquímica entre estos codones y la lisina. Se han propuesto muchos modelos que establecen un cierto tipo de afinidad estereoquímica entre codones, anticodones y aminoácidos, y al parecer, muchos de ellos están sustentados por evidencia experimental [17]. Sin embargo, todos ellos se resumen en el modelo de “llave-cerradura”, que establece que en el origen de la vida se dieron las interacciones entre codones, anticodones y aminoácidos de la misma manera en la cual una llave interacciona con su correspondiente cerradura.



- **Teoría fisicoquímica.** Esta teoría sostiene que la fuerza motriz detrás de la estructura del código genético subyace en la tendencia de reducir las distancias fisicoquímicas entre aminoácidos codificados por codones que se diferencian en una base. En particular, los trabajos de Sonneborn identifican una presión selectiva tendiente a reducir los efectos deletéreos de las mutaciones como el factor esencial que determina la posición de los aminoácidos en la tabla del código genético [18]. Por otro lado, Woese sugiere que la estructura del código se debe a la presión selectiva tendiente a reducir los errores en la traducción de los mensajes genéticos [19].
- **Teoría de reducción de ambigüedades.** Esta teoría, propuesta por Woese y Fitch en 1966 [20], es esencialmente la misma que la anterior, y sugiere que grupos de codones similares fueron asignados a grupos de aminoácidos también similares estructural y químicamente, y que la estructura del código genético se debe a la tendencia evolutiva de reducir las ambigüedades entre y dentro de grupos de aminoácidos afines. Esta hipótesis parece estar apoyada por los trabajos experimentales de Weber y Lacey [21], en los cuales se muestra evidencia de que existen similitudes fisicoquímicas entre los codones que especifican para aminoácidos del mismo grupo.
- **Teoría de coevolución.** Esta teoría fue formulada inicialmente por Nieremberg en 1963, quien reconoció la existencia de contigüidad entre codones que codifican para aminoácidos sintetizados a partir de un precursor común<sup>10</sup>. Por otro lado, en 1973 Dillon propuso que la asignación de codones a los aminoácidos pudo haber estado guiada por las relaciones biosintéticas entre los aminoácidos. Sin embargo, no fue sino hasta 1975 cuando Wong formula claramente la hipótesis de coevolución [22], [23] la cual establece que la estructura de codones del código genético es, sobre todo, una "huella digital" de las rutas biosintéticas de los aminoácidos. Consecuentemente, el origen del código genético podría dilucidarse sobre la base de las relaciones biosintéticas entre los aminoácidos y sus precursores, las cuales son claramente visibles en las rutas metabólicas existentes actualmente en las células. En otras palabras, la hipótesis de coevolución sugiere que inicialmente, solamente se codificaban los aminoácidos precursores, y que conforme estos aminoácidos gradualmente dieron lugar a otros (los aminoácidos producto), algunos de los codones en el "dominio"

<sup>10</sup>Muchos aminoácidos se sintetizan a partir de otros. Por ejemplo, por medio de una desulfurización reductiva (o sea, quitando un azufre utilizando una base), la cisteína se convierte en alanina. El aminoácido original se denomina *precursor*, y al aminoácido final se le llama *producto*.

de los precursores se transfirieron (o fueron cedidos) a los aminoácidos producto.

Implícitamente, las tres primeras teorías establecen que el código genético se estructuró en una etapa muy temprana del origen de la vida, ya que apelan a un determinismo fisicoquímico de interacción entre moléculas. Por otro lado, la hipótesis de coevolución es compatible con una estructuración tardía del código, puesto que dicha estructuración tuvo lugar paralelamente con el desarrollo de las rutas biosintéticas de los aminoácidos. Ambos esquemas no son del todo irreconciliables, y de hecho, Massimo Di Giulio propone una manera de reconciliarlos. Sin embargo, algo que salta a la vista inmediatamente es que en todas estas teorías los codones formados por tres bases se dan por supuestos desde el inicio. Es decir, dichas teorías están más enfocadas a explicar por qué tal o cual codón especifica para tal o cual aminoácido, tratando de hallar una explicación del agrupamiento de los aminoácidos con propiedades fisicoquímicas similares en la tabla del código. No obstante, no se cuestiona el por qué de la estructura de tripletes del código, sino que esta propiedad se supone desde el principio.

Trabajos similares, como los de P. Béland y T. F. Allen [24], donde se propone que en el origen del código genético las secuencias de mRNA podían leerse en ambas direcciones (del extremo 5' al extremo 3' y viceversa) codificando la misma proteína, también dan por sentado que los codones están compuestos por tres bases y sólo tres bases. Esta suposición no necesariamente está mal. De hecho, así es como funciona el código genético. Pero si se quiere tener una explicación satisfactoria del origen de la estructura del código, también es necesario explicar por qué las palabras codificadoras están compuestas por tres letras.

Hasta donde sabemos, el único trabajo que apunta en esta dirección es uno publicado en 1997 por W. Lee y L. Luo [25]. Haciendo un análisis de Fourier, estos autores encuentran que una secuencia genética formada por cuatro letras presenta cierto tipo de periodicidades, y bajo condiciones especiales de correlación entre las letras a lo largo de la secuencia, se obtiene una regularidad con periodo 3 en el espectro de Fourier de las correlaciones. También se presentan otros periodos (2 y 4), pero el periodo dominante es el 3. Estos autores atribuyen la estructura de tripletes del código genético a esta periodicidad de 3 que encontraron en el espectro de Fourier. En este trabajo, Lee y Luo hacen un análisis simbólico de las secuencias genéticas, olvidándose por completo de todo tipo de interacciones fisicoquímicas entre bases, codones, anticodones y aminoácidos. Es difícil creer que la estructura del código genético se deba a propiedades de simbología más que a interacciones físicas y químicas entre moléculas, aunque todo es posible mientras no se demuestre lo contrario.

Como podemos ver, explicar la estructura del código genético es un problema verdaderamente multifacético y muy complejo. Nuestra intención aquí es dar una posible explicación de "¿por qué tres?", basada en consideraciones dinámicas del movimiento del ribosoma a lo largo del RNA mensajero. El punto de vista que adoptamos es radicalmente diferente. En los trabajos descritos en esta sección (que son relativamente recientes), podemos ver que se está buscando la explicación de la estructura del código genético en las interacciones entre codones, anticodones y aminoácidos, en las rutas biosintéticas de éstos últimos, en las presiones selectivas que reducen la ambigüedad y optimizan la información, y hasta en aspectos de simbología matemática. No obstante, la interacción entre el ribosoma y el RNA mensajero no ha sido suficientemente considerada.

En este trabajo proponemos que la estructura de tripletes del código genético tiene que ver con aspectos dinámicos de la interacción entre el ribosoma y el RNA mensajero. Como veremos, nuestro modelo no se contrapone en lo absoluto a los otros modelos y teorías existentes, sino que, junto con ellos, brinda una explicación más sólida de por qué la naturaleza codifica sus mensajes con palabras formadas por tres letras.

Es necesario aclarar que en este trabajo no estamos interesados en por qué tal o cual codón especifica para tal o cual aminoácido. La pregunta que nos hacemos es mucho más sencilla: ¿por qué tres? A pesar de que esta pregunta representa sólo una pequeña parte del problema del origen del código genético, veremos que ya de por sí es lo suficientemente complicada como para dar origen a una tesis doctoral.



## Capítulo 3

# La vida en la Tierra... ¿se originó en la Tierra?

El problema del origen de la vida ha inquietado al hombre prácticamente desde que éste tomó consciencia de que está vivo. Esta inquietud estriba no sólo en mera curiosidad, sino en que el entender nuestros orígenes, de dónde venimos y por qué somos, puede ayudar a entender nuestro futuro: a dónde vamos y qué seremos. A lo largo de la historia de la humanidad, la pregunta “¿cuál fue el origen de la vida?” ha tenido múltiples respuestas que varían de cultura en cultura y de época en época. Tales respuestas se han dado desde diferentes puntos de vista, que van desde lo mitológico hasta lo científico. Sin embargo, aún cuando algunos de estos puntos de vista llegan a ser contradictorios, todos ellos comparten una característica en común: *en general, se asume que la vida en la Tierra se originó en la Tierra misma.* Por alguna razón, la gente se ha sentido más cómoda suponiendo que nuestros orígenes tuvieron lugar aquí mismo, en nuestra propia casa. Así, por ejemplo, casi todas las corrientes religiosas y mitológicas asumen que “los Cielos” están dominados por los dioses, mientras que la Tierra es el lugar destinado para todos los mortales, ya sean plantas, animales o seres humanos, y que tales mortales fuimos “creados” aquí desde el principio.

Esta tendencia a suponer, o mejor dicho, a asumir que la vida en la Tierra se originó en la Tierra, no es particular de la religión o la mitología, sino que también ha permeado en las ideas científicas a tal grado, que se le ha llegado a dar un nombre: se le conoce como *hipótesis endógena*. Por ejemplo, A. I. Oparin en Rusia y J. B. S. Haldane en Inglaterra, propusieron independientemente en la década de 1920 un escenario en el que las primeras moléculas orgánicas útiles para la vida se crearon en la superficie de la Tie-

rra a partir de compuestos de carbono y nitrógeno relativamente simples. De acuerdo con el modelo de Oparin y Haldane, estos compuestos orgánicos adquirieron cada vez mayor complejidad, y eventualmente evolucionaron hasta dar origen a los primeros organismos unicelulares, ocurriendo todo en los mares primitivos de la Tierra. Años más tarde, las ideas de estos dos investigadores inspiraron a S. L. Miller y H. C. Urey, de la Universidad de Chicago, quienes realizaron un experimento en el que simulaban las condiciones primitivas de la Tierra para producir moléculas orgánicas complicadas. Este histórico experimento marcó un hito en el desarrollo de las teorías del origen de la vida, ya que posteriormente muchos otros investigadores realizaron experimentos similares, aunque más sofisticados, para producir moléculas orgánicas más complicadas y en mayores cantidades que las que obtuvieron Miller y Urey, pero siempre con la idea de obtener dichas moléculas a partir de compuestos que se encontraran bajo condiciones físicas y químicas muy parecidas a las que prevalecían en la Tierra primitiva. Es decir, tanto Oparin y Haldane, como Miller y Urey, y muchos otros investigadores posteriores que realizaron experimentos del mismo estilo, han dado por sentado que el origen de la vida en la Tierra tuvo lugar en la Tierra misma.

Solamente algunos escritores de ciencia ficción, y algunos científicos arriesgados (como por ejemplo Fred Hoyle), habían imaginado que los primeros procesos biológicos que eventualmente condujeron a los seres vivos, se pudieron haber llevado a cabo fuera de la Tierra, es decir, en el espacio exterior. Sin embargo, hasta antes de la década de 1980 estas hipótesis no habían sido más que especulaciones sin fundamento. Pero en los últimos 20 años, se ha acumulado evidencia sustancial que sugiere que los primeros procesos que originaron la vida en la Tierra, no se dieron en la Tierra misma, sino que se llevaron a cabo en el espacio exterior. Lo anterior no significa, de ninguna manera, que del espacio exterior hayan "llovido" peces o dinosaurios a nuestro planeta; es simplemente que ahora se cuenta con evidencia de que en el espacio exterior existen las condiciones para la formación de moléculas orgánicas con un alto grado de complejidad, mientras que en la Tierra primitiva tales condiciones parecen no haber existido.

El problema del origen de la vida no se ha resuelto aún. En particular, la cuestión de si la vida se originó en la Tierra o en el espacio exterior sigue siendo tema de gran debate y controversia. No obstante, los trabajos experimentales y teóricos que favorecen el origen extraterrestre de la vida cada vez son más numerosos y están más fundamentados. Por el contrario, los argumentos que favorecen el origen terrestre de la vida ya no son tan contundentes como lo eran hace 20 años. En este trabajo adoptaremos el origen extraterrestre de la vida como una hipótesis razonable, aunque no conclusiva. Para esto, en el presente capítulo veremos los problemas que se

presentan cuando suponemos la vida se originó en la Tierra, y cómo algunos de estos problemas se resuelven, en principio, suponiendo un escenario extraterrestre para dicho origen. Además, veremos un poco de la evidencia que se ha encontrado y que apoya este escenario extraterrestre.

### 3.1 ¿Qué se entiende por “origen de la vida”?

Delos capítulos precedentes queda claro que en el metabolismo celular hay una relación muy estrecha entre ácidos nucleicos y proteínas: los primeros contienen la información para construir a las segundas, mientras que las segundas ayudan a la conservación y replicación de los primeros. De acuerdo con algunos autores [26], [27], la vida consiste en la conservación y replicación de los ácidos nucleicos, y toda la maquinaria celular está montada con este único fin. Según este punto de vista, cualquiera que sea la función particular de una proteína, de una enzima o de un RNA estructural, dicha función está enmarcada dentro de un proyecto más general, y es el de lograr la conservación y replicación de los ácidos nucleicos, en especial el DNA.

Si adoptamos este punto de vista, entonces el “origen de la vida” adquiere un sentido muy bien definido. Por origen de la vida entenderemos lo siguiente: *¿cómo surgieron los ácidos nucleicos y las proteínas a partir de compuestos inorgánicos simples, y cómo se estableció la relación metabólica entre estos dos tipos de moléculas que trajo como consecuencia la conservación y replicación de los ácidos nucleicos?* En esta “definición” de origen de la vida se distinguen claramente dos tipos de procesos:

- (a) el surgimiento de las macromoléculas orgánicas y
- (b) el establecimiento de relaciones fisicoquímicas -o metabólicas- entre tales macromoléculas.

A estos dos tipos de procesos se les denomina **prebióticos** y **protobióticos**, respectivamente, y gran parte de las investigaciones sobre el origen de la vida consisten en dilucidar cómo se llevaron a cabo dichos procesos. Cuando decimos que hay evidencia del origen extraterrestre de la vida, a lo que nos referimos es a que hay evidencia de que tanto los procesos prebióticos como los protobióticos tienen lugar en el espacio exterior, en particular en el polvo interestelar y en superficies de meteoritos y cometas. Esto, por sí sólo, no es un argumento para suponer que la vida no se pudo haber originado aquí en la Tierra. En todo caso, la existencia de procesos prebióticos y protobióticos en el espacio exterior indica que la vida es tan “prolífica”, que además de en la Tierra se puede dar en otros lugares. Sin embargo, en la sección siguiente veremos que se tienen algunos problemas

(serios) si partimos del hecho de que el origen de la vida se dió aquí en la Tierra.

## 3.2 Dificultades con el origen terrestre de la vida

### El problema del tiempo.

Hasta hace no mucho se creía que los primeros organismos vivos (organismos unicelulares como bacterias) aparecieron apenas hace 600 millones de años. Esta creencia estaba basada en la edad de los fósiles de bacteria más antiguos que se habían encontrado. Sin embargo, en 1992 los paleontólogos encontraron en Sudáfrica y en el oeste de Australia, fósiles de cianobacterias de aproximadamente 3600 millones de años de antigüedad (con más precisión, estos fósiles tienen una edad de  $3556 \pm 32$  millones de años). Más aún, en Isua, Groenlandia, se encontraron estructuras de carbono que muy probablemente pudieron ser producidas por organismos vivos (bacterias), ¡que datan de hace 3800 millones de años! [28], [29]

Por otro lado, es bien sabido que la edad de la Tierra es de 4500 millones de años. Sin embargo, en sus inicios la Tierra estaba muy caliente y era un lugar totalmente inhóspito para la vida. De hecho, hay evidencia de que la Tierra fue bombardeada por meteoritos y asteroides durante sus primeros 500 millones de años. Estos impactos contribuían a mantener la temperatura de la superficie de la Tierra lo suficientemente alta como para abortar cualquier intento de formación de moléculas orgánicas complejas, de tal suerte que las condiciones en la Tierra para que se pudieran llevar a cabo los procesos prebióticos y protobióticos no se dieron sino hasta hace 4000 millones de años, a lo más.

Las consideraciones anteriores hacen evidente que el tiempo que transcurrió desde que la Tierra se enfrió lo suficiente como para albergar moléculas orgánicas complejas, hasta que aparecieron las primeras bacterias, fue tan sólo de 100 millones de años cuando mucho. Si los procesos prebióticos y protobióticos se llevaron a cabo en la Tierra, entonces en 100 millones de años (o tal vez menos), se paso de una "sopa" de compuestos inertes, como metano, amoníaco e hidrógeno, a una "sopa" llena de bacterias auto replicantes, con membrana, ácidos nucleicos, ribosomas, proteínas y todo lo demás.

Algunos autores consideran que esta ventana de tiempo es demasiado estrecha como para que se crearan estructuras con tan alto grado de complejidad a partir de compuestos simples (tan sólo los dinosaurios dominaron la Tierra por casi 180 millones de años).



### El problema de la quiralidad.

Casi todos los aminoácidos son moléculas quirales, es decir, presentan una asimetría definida respecto a izquierda y derecha. Así, para cada aminoácido existe su "imagen especular", tal y como la mano izquierda es la imagen especular de la mano derecha (o viceversa). A los aminoácidos con quiralidad izquierda se les denomina *L-aminoácidos*, y a los de quiralidad derecha se les designa por *D-aminoácidos*. Dos aminoácidos que son imágenes especulares uno del otro se denominan *enantiómeros*, y tienen las mismas propiedades químicas. Por eso, la designación "izquierda" o "derecha" no tiene nada que ver con la química de los aminoácidos, sino se debe a que estas moléculas presentan actividad óptica: una muestra de *L-aminoácidos* desvia la luz polarizada hacia la izquierda, mientras que una muestra de *D-aminoácidos* lo hace hacia la derecha.

Con excepción de la glicina (que no es quiral), todos los demás aminoácidos que utiliza la célula se pueden presentar en las dos variedades, *L* y *D*. Sin embargo, el hecho es que, con poquísimas excepciones, todos los organismos vivos que se conocen utilizan sólo *L-aminoácidos*. Además, en los casos en que se utilizan aminoácidos tipo *D*, éstos provienen de precursores tipo *L*. Lo anterior resulta bastante extraño, sobre todo si se toma en cuenta que cuando los aminoácidos son sintetizados artificialmente en los laboratorios, se producen en muestras racémicas, es decir, muestras que contienen aproximadamente la misma cantidad de aminoácidos de tipo *L* que de tipo *D*. No obstante, por alguna razón, la vida en la Tierra ha "escogido" a los *L-aminoácidos* para construir las proteínas que utilizan los seres vivos. Probablemente esta "elección" no fue más que resultado del azar. Sin embargo, como veremos un poco más adelante, otra vez sucede que la Naturaleza juega con dados cargados.

Vale la pena mencionar que en el experimento de Miller-Urey, y en todos los experimentos posteriores de este tipo que tratan de generar aminoácidos a partir de compuestos químicos simples, simulando las condiciones primitivas de la Tierra, han aparecido siempre muestras racémicas de aminoácidos cuando tales experimentos se realizan a temperatura ambiente. No obstante, Mayo Greenberg ha demostrado recientemente [30] que, si durante la síntesis artificial de aminoácidos se radía al sistema con luz ultravioleta polarizada circularmente en algún sentido, ya sea izquierdo (*L*) o derecho (*D*), y si la temperatura del sistema es lo suficientemente baja (cercana al cero absoluto), entonces la muestra final de moléculas orgánicas contiene más aminoácidos de un tipo que de otro, dependiendo del sentido de polarización de la luz ultravioleta.

De los experimentos de Greenberg (y de otros parecidos) se puede concluir que para romper la simetría en la producción racémica de aminoácidos, se requieren básicamente dos cosas: i) un agente externo asimétrico que

actúe sobre el sistema mientras se sintetizan los aminoácidos (como la luz ultravioleta polarizada), y ii) que la temperatura del sistema sea muy baja ( $\sim 10$  °K), para que las fluctuaciones térmicas no "opaquen" la influencia del agente externo.

De acuerdo con las observaciones de yacimientos minerales de miles de millones de años de antigüedad, se sabe que la superficie de la Tierra, en sus orígenes, estaba muy caliente (los minerales se encuentran fundidos). Por lo tanto, en la Tierra primitiva no se tuvieron las condiciones para romper la simetría en la producción racémica de aminoácidos, lo que contrasta con el hecho observado actualmente de que la gran mayoría de los aminoácidos que naturalmente existen en nuestro planeta, son de tipo L.

### **El problema de la polimerización.**

Claramente, la vida no pudo originarse si antes no hubieran existido moléculas en forma de cadenas poliméricas extensas, tales como los ácidos nucleicos y las proteínas. Así, el proceso de polimerización de monómeros es una condición indispensable para el surgimiento de la vida en la Tierra. No obstante, algunos autores consideran que en la Tierra primitiva no se tenían las condiciones para que dicha polimerización se llevara a cabo.

En primer lugar, debido a la gran capacidad de reacción del oxígeno, la polimerización de moléculas orgánicas se debe realizar en un medio químico reductor, es decir, en un medio prácticamente sin oxígeno y con mucho hidrógeno. En un ambiente oxidante, la polimerización de monómeros orgánicos *generalmente no puede efectuarse*, ya que el oxígeno se liga prematuramente a los sitios activos de polimerización de los monómeros, bloqueándolos e impidiendo que puedan enlazarse. De hecho, una de las hipótesis de trabajo de la teoría de Oparin-Haldane y de los experimentos de tipo Miller-Urey, es que la atmósfera primitiva de la Tierra era reductora. Sin esta condición, la polimerización de moléculas orgánicas simplemente no se puede dar.

Sin embargo, algunos modelos recientes de la formación de la Tierra han sugerido que, en lugar de ser reductora, la atmósfera primitiva era más bien medianamente oxidante, con abundancia de  $\text{CO}_2$ ,  $\text{N}_2$  y  $\text{H}_2\text{O}$ . Lo anterior parece estar parcialmente confirmado por observaciones recientes realizadas en yacimientos volcánicos al norte de África que datan de hace más de 4000 millones de años, en donde se ha encontrado abundante oxígeno "prehistórico" (es decir, con composición isotópica muy diferente a la observada actualmente) atrapado en los minerales que conforman dichos yacimientos [31]. Así, aparentemente la atmósfera primitiva no albergaba las condiciones reductoras para que se diera la polimerización de moléculas orgánicas, y por lo tanto, para que surgiera la vida.

Por otro lado, para que se formen polímeros, las moléculas que los con-

forman (monómeros) deben ser *bifuncionales*, es decir, deben tener dos sitios activos de enlace. Sólo de esta manera se pueden encadenar los monómeros uno seguido de otro, como se muestra en la Fig3.1a. Los aminoácidos y los nucleótidos son monómeros bifuncionales. Pero además de moléculas bifuncionales, también existen moléculas *monofuncionales*, las cuales sólo tienen un sitio activo de enlace (Fig3.1b). Claramente, las moléculas monofuncionales no se pueden polimerizar. Más aún, estas moléculas, si se encuentran presentes en el medio, detienen la polimerización de monómeros bifuncionales, ya que si en un extremo del polímero en formación se liga una molécula monofuncional, como se muestra en la Fig3.1d, dicho extremo queda bloqueado.

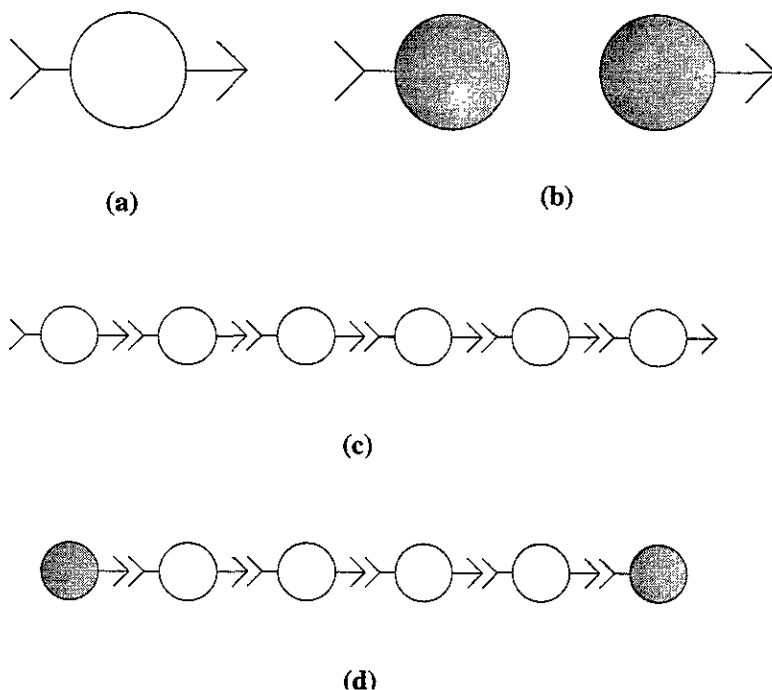


Figura 3.1: Representación esquemática del tipo de moléculas que aparecen en los experimentos de tipo Miller-Urey. (a) Moléculas bifuncionales, con dos sitios activos de polimerización. (b) Moléculas monofuncionales, con sólo un sitio activo de polimerización. (c) Las moléculas bifuncionales se unen por “atrás” y por “delante”, formando un polímero. (d) Las moléculas monofuncionales detienen la polimerización.

En los experimentos de tipo Miller-Urey, además de obtenerse moléculas bifuncionales, como aminoácidos, en la mezcla de compuestos orgánicos resultante también se encuentran moléculas monofuncionales, ¡en mayor abundancia que las bifuncionales! Por ejemplo, en un experimento típico de Miller-Urey, se obtienen 1.6 Mol % de aminoácidos (principalmente glicina y alanina), pero aproximadamente 3.6 Mol % de moléculas monofuncionales, como ácido fórmico, ácido acético y ácido propiónico. Además se producen aminas monofuncionales, como metilamina y etilamina, en proporciones de 5.5 Mol %. Según las condiciones del experimento de Miller, no se encuentran nucleósidos ni nucleótidos, por lo que, si es que hay, su concentración está con seguridad por debajo del límite de comparabilidad, que es de 0.003 Mol % (en esta cantidad fue hallada adenina al irradiarse con electrones las mezclas gaseosas de Miller). En la siguiente tabla se muestra un resultado característico de los experimentos de tipo Miller-Urey [10].

Moléculas Bifuncionales		Moléculas Monofuncionales	
	Cantidad relativa (Mol %)		Cantidad relativa (Mol %)
glicina	1.00	ácido fórmico	3.00
alanina	0.54	ácido acético	0.30
ácido glicólico	0.90	ácido propiónico	0.30
ácido láctico	0.73	metilamina	5.00
ácido succínico	0.50	etilamina	0.50
nucleósidos	< 0.003		
<b>TOTAL</b>	<b>3.97</b>	<b>TOTAL</b>	<b>9.10</b>

En el Apéndice A se muestra que, si en una mezcla se tienen moléculas monofuncionales y bifuncionales en proporciones  $p$  y  $q$ , respectivamente, con  $p + q = 1$ , y si estas moléculas se están enlazando aleatoriamente, entonces la probabilidad  $P(L)$  de que se forme un polímero de longitud  $L$  antes de que sus dos extremos queden bloqueados, está dada por

$$P(L) = (L - 1)p^2q^{L-2} \quad (3.1)$$

lo cual nos lleva a que la longitud promedio  $\bar{L}$  de los polímeros formados en dicha mezcla es

$$\bar{L} = \frac{2}{1 - q}$$

Para la tabla mostrada anteriormente, se tiene  $p = 0.69$ ,  $q = 0.31$ , de donde resulta que la longitud promedio de los polímeros formados es  $\bar{L} \approx 3$ .

Más aún, de la ecuación (3.1) se obtiene que la probabilidad de obtener un polímero cuya longitud sea mayor o igual que  $N$  está dada por

$$P[L \geq N] = q^N \left( 1 + N \frac{p}{q} \right). \quad (3.2)$$

Si  $q = 0.31$  y  $N = 100$ , entonces  $P[L \geq 100] \simeq 3 \times 10^{-49}$ . Es decir, bajo las condiciones del experimento que arrojó los resultados de la tabla anterior, muy difícilmente se obtendrían polímeros largos. Si este tipo de experimentos refleja en algo la síntesis de aminoácidos y nucleótidos que tuvo lugar en la Tierra primitiva, entonces podemos concluir que la polimerización de tales moléculas no pudo haber conducido a la formación de cadenas largas de DNA ni de proteínas, sin las cuales la vida no se puede originar.

Finalmente, hay otro problema con la polimerización que tiene que ver con el agua, pero esto lo abordaremos en un apartado posterior.

### **El problema del huevo y la gallina: el mundo del RNA.**

Este problema no tiene que ver propiamente con el origen terrestre de la vida, sino con la pregunta "¿quién fue primero, los ácidos nucleicos o las proteínas?" El problema queda mejor explicado en palabras de Leslie E. Orgel [32]:

"Cualquiera que trate de resolver este problema (el de la interrelación entre ácidos nucleicos y proteínas) inmediatamente llega a una paradoja. Actualmente los ácidos nucleicos son sintetizados sólo con la ayuda de las proteínas, y las proteínas son sintetizadas sólo si su información está contenida en los ácidos nucleicos. Es extremadamente improbable que las proteínas y los ácidos nucleicos, ambos estructuralmente muy complejos, surgieran espontáneamente en el mismo lugar y al mismo tiempo. Sin embargo, también parece imposible tener los unos sin los otros. Llegamos así al problema del huevo y la gallina: ¿quién fue primero, los ácidos nucleicos o las proteínas? A primera vista podríamos concluir que la vida nunca pudo haberse originado por medios químicos".

Los primeros indicios para la solución de este problema surgieron en 1981, cuando Thomas R. Cech descubrió un tipo de RNA con capacidades enzimáticas, a los que llamó ribozimas [33], [34]. Después de esto, se descubrieron muchos otros tipos de RNA con capacidad enzimática, como por ejemplo el snRNA encargado de la edición del mensaje genético, o el rRNA que cataliza la unión covalente de aminoácidos en el ribosoma.

El primer impulso fue argüir que unos RNA primitivos con capacidad enzimática no necesitarían proteínas catalíticas auxiliares. Esta sugerencia

teórica se ha fortalecido por los trabajos posteriores del propio Cech y de otros autores, quienes han conseguido RNA's sintetizados artificialmente con capacidad de formar réplicas de otros RNA's precursores. Además, el uso de nuevas técnicas de "evolución dirigida in vitro" ha permitido obtener RNA's enzimáticos mutantes que catalizan su propia autoreplicación, en la total ausencia de proteínas. Estos RNA's catalíticos tienen más de 150 nucleótidos, pero hay viroides de plantas con ribozimas de sólo 52 nucleótidos, y se ha demostrado que bastan con cadenas de RNA de sólo 19 nucleótidos para tener actividad enzimática. Además, como ya se ha mencionado, T. A. Steitz y su equipo han mostrado que la síntesis de proteínas que se lleva a cabo en los ribosomas está totalmente catalizada por el RNA ribosomal, y no por las proteínas [12], [13].

La gran versatilidad química del RNA llevó a F. Crick, L. Orgel y C. R. Woese a postular la existencia de un mundo prebiótico basado esencialmente en las propiedades de esta molécula. es decir, un *mundo del RNA*, como lo llegara a bautizar Walter Gilbert [37], [38]. La evolución de este mundo del RNA hacia otro que, como el actual, incluyera a las proteínas. no parece presentar dificultades insalvables, sobre todo si tomamos en cuenta los trabajos de H. F. Noller y de J. A. Priccirilli, que muestran que una enzima primitiva de RNA pudo ser la encargada de unir el RNA de transferencia con su correspondiente aminoácido [36].

Aunque no todo está resuelto y aún hay temas de controversia respecto al mundo del RNA, esta hipótesis se vislumbra como muy prometedora. En particular, el problema del huevo y la gallina queda resuelto: primero fue el mundo del RNA, y después se incorporaron a éste las proteínas. Sin embargo, vale la pena preguntarse si es factible la presencia de un mundo de RNA en la Tierra primitiva. Al parecer, la respuesta es no.

### El problema de la hidrólisis.

Los aminoácidos no se pueden polimerizar en presencia de agua. Recordemos que el enlace peptídico que une dos aminoácidos se da con la liberación de una molécula de agua:  $\text{COOH} + \text{NH}_2 \rightarrow \text{CONH} + \text{H}_2\text{O}$ . En equilibrio, la reacción inversa (hidrólisis) también tiene lugar:  $\text{CONH} + \text{H}_2\text{O} \rightarrow \text{COOH} + \text{NH}_2$ , es decir, en un medio acuoso no controlado, el agua destruye el enlace peptídico. Así, en un sistema en equilibrio, este enlace de deshidratación que liga dos aminoácidos se da tantas veces como la reacción inversa de hidrólisis que los desliga, de tal manera que en presencia de agua se crean y se destruyen paralelamente enlaces peptídicos, impidiendo la formación de cadenas protéicas largas. En las células esto no ocurre porque cuentan con mecanismos muy eficientes para controlar la concentración de agua en cada lugar donde se lleva a cabo una reacción de polimerización.

La hidrólisis también ocurre con los nucleótidos. Los experimentos de

L. Orgel han confirmado que en concentraciones de agua superiores al 99 Mol por ciento, no se produce polimerización de nucleótidos, porque la rotura de los enlaces por hidrólisis ocurre de diez a cien veces más aprisa que las reacciones de polimerización [39]. Además, como ya se ha mencionado, el RNA es muy susceptible a deshacerse en agua.

Si pensamos que la Tierra primitiva estaba rebosante de agua, con sus mares y lagos en constante formación y evolución, entonces queda claro que el mundo del RNA nunca hubiera podido existir... al menos no en la Tierra. Más bien, lo que parece que nunca existió fueron los mares llenos hasta el tope de compuestos orgánicos prebióticos, ese "caldo orgánico" de Oparin en el que se llevaron a cabo los procesos protobióticos y prebióticos que eventualmente dieron origen a los seres vivos.

Las consideraciones precedentes hacen ver al origen de la vida en la Tierra más como una casualidad que como una consecuencia de la interacción y organización de la materia orgánica, ya que para que los procesos prebióticos y protobióticos se llevaran a cabo en la Tierra primitiva, se requería de condiciones físicas y químicas muy improbables: una (auto)organización muy rápida de la materia orgánica; un ambiente químico reductor; temperaturas extremadamente bajas; ausencia de agua; concentraciones comparativamente bajas de moléculas monofuncionales, etc. Algunos autores han imaginado que tales condiciones se pudieron haber dado, por "accidente", en algún nicho escondido de la Tierra primitiva [40], tal como una burbuja atmosférica atrapada en algún yacimiento mineral submarino, o algo por el estilo. Puede ser que haya ocurrido así. No obstante, hipótesis como ésta inevitablemente nos conducen a pensar que la vida es producto de un "accidente" afortunado que ocurrió hace 4000 millones de años en alguna parte de nuestro planeta. Como veremos en la siguiente sección, una hipótesis más plausible es suponer que los primeros procesos prebióticos y protobióticos se llevaron a cabo en el espacio exterior.

### 3.3 El escenario extraterrestre

En 1864 cayó un meteorito de gran tamaño en el pueblo de Orgueil, cerca de Montauban, Francia. Este meteorito era particularmente extraño por su alta concentración de carbono y arcillas, lo cual indujo a los geoquímicos a realizar análisis exhaustivos de sus constituyentes químicos. En 1963, I. R. Kaplan encontró, mezclados con el polvo negro que había raspado de la superficie del meteorito, una multitud de aminoácidos que, hasta antes de esto, eran particulares de los organismos vivos (glicina, alanina, valina, prolina, ácido aspártico, ácido glutámico). De hecho, los encontró en mayores cantidades incluso que las obtenidas en los experimentos del tipo



Figura 3.2: Microfotografía de la superficie de un pedazo de meteorito. Las arcillas minerales de las que se compone forman estructuras laminares entremezcladas que dan origen a canales unidimensionales, en los que se concentra la materia orgánica, favoreciendo así la formación de polímeros orgánicos largos. Las flechas indican la presencia de partículas de ferromagnetita incrustadas en los canales del meteorito.



Miller-Urey. Además, se encontraron dos de las cuatro bases nitrogenadas (alanina y guanina) que conforman al DNA y al RNA. Quedaba la posibilidad de que, después de haber estado en el suelo terrestre y de haber estado guardado durante casi 100 años en el Museo de Historia Natural de la calle Buffon de París, el meteorito se hubiera contaminado por compuestos orgánicos atmosféricos o museológicos. Sin embargo, los análisis efectuados en 1965 por K. A. Kvenvolden y sus colegas demostraron que esto no fue así, ya que la muestra de aminoácidos extraídos del meteorito era más o menos racémica, es decir, contenía aminoácidos de tipo L y de tipo D aproximadamente en la misma proporción (había un ligero exceso de aminoácidos tipo L), mientras que en la Tierra no existen, de forma natural, aminoácidos de tipo D.

Quedó claro, por tanto, que las moléculas orgánicas extraídas del meteorito habían sido sintetizadas fuera de la Tierra. Descubrimientos similares se suscitaron posteriormente utilizando otros meteoritos. Además, después de esto algunos astrónomos y astroquímicos se dedicaron particularmente a buscar materia orgánica en el espacio exterior, y encontraron que en los cometas y en las nubes de gas interestelar también hay grandes concentraciones de moléculas orgánicas, en particular aminoácidos y bases nitrogenadas [41], [42]. Fue evidente entonces que en el espacio exterior también se dan las condiciones de síntesis de moléculas orgánicas que en la Tierra encontramos íntimamente ligadas a la vida. Estos descubrimientos, por sí mismos, no decían mucho acerca del origen de la vida en la Tierra; en todo caso, mostraban que la síntesis de material orgánico se puede dar en cualquier parte, tanto en la Tierra como fuera de ella. Pero lo importante era que la existencia de material orgánico sintetizado en el espacio exterior mostraba que la *hipótesis endógena* (es decir, la hipótesis de que los procesos prebióticos y protobióticos tuvieron lugar en la Tierra primitiva), no era la única posible.

Sin embargo, en la década de 1980 se realizó un descubrimiento de fundamental importancia para el escenario extraterrestre del origen de la vida [41]. François Rober y su grupo encontraron que la materia orgánica en los meteoritos no está distribuida aleatoriamente en ellos. Como muestra la Fig3.2, las arcillas de las que está hecho el meteorito forman estructuras laminares entremezcladas, creando canales (cuasi) unidimensionales. Así, François Robert y su grupo demostraron que la materia orgánica de los meteoritos se alberga, precisamente, en estos canales, ¡en forma de polímero orgánico! Además, en los canales se encuentran incrustadas partículas de ferromagnetita con las cuales el polímero orgánico puede interactuar. Estos trabajos quedaron reforzados con los de Pierre Laszlo y Graham Cairns-Smith [43], [44], quienes mostraron que ciertas superficies minerales, como arcillas o piritas, son poderosos catalizadores para la polimerización

de monómeros orgánicos, como los aminoácidos y los nucleótidos. En particular, las arcillas presentes en los meteoritos son catalizadores altamente eficaces en la polimerización de estas moléculas. Por lo tanto, la probabilidad de formación de polímeros orgánicos, como polipéptidos, es muy alta en los meteoritos, ya que, por un lado, el "choque" de monómeros en los canales del meteorito se da con mucho más frecuencia que si estuvieran regados en un mar tridimensional, y por otro, las arcillas del meteorito catalizan las reacciones de policondensación.

Así mismo, Mayo Greenberg, Max P. Bernstein, F. Krueger y otros autores [45] [42] [46], han sugerido que la materia orgánica observada en el polvo interestelar, puede estar contenida dentro de filamentos tubulares de hielo y silicatos, en los que también se favorece la formación y polimerización de moléculas orgánicas.

Todos estos descubrimientos muestran que los procesos protobióticos también ocurren en el espacio exterior, y que probablemente, a través de los asteroides y meteoritos que continuamente bombardearon la superficie de la Tierra durante los primeros millones de años, se adquirió la materia orgánica de la que están hechos ahora los organismos vivos.

Por otro lado, vale la pena mencionar que actualmente se tienen posibles evidencias de que existió vida fuera de la Tierra. En 1996, David S. McKay y su grupo en la NASA [47], [48], estudiando un meteorito que provenía de Marte y que cayó en la Antártica hace trece mil años, encontraron estructuras carbonatadas de forma globular, que probablemente son fósiles de bacteria, de 3600 millones de años de antigüedad. Haciendo análisis químicos, encontraron que estas estructuras globulares, mostradas en la Fig3.3, están formadas por compuestos de carbono llamados PAHs (hidrocarburos policíclicos aromáticos), que son producidos típicamente en los procesos de fosilización de las bacterias. Si los glóbulos carbonatados del meteorito efectivamente son fósiles de bacteria, es evidente que no provienen de la Tierra. Sin embargo, existe mucha controversia respecto a si dichas estructuras son en realidad fósiles de bacterias o no. Aparentemente, esta hipótesis cuenta cada vez con menos aceptación, pero la mencionamos aquí para hacer notar que actualmente, una buena parte de la comunidad científica considera que hay otras posibilidades reales aparte de la hipótesis endógena.

Con la evidencia que se tiene actualmente, suponer que fue en el espacio exterior donde se dieron los primeros procesos de síntesis y organización de la materia orgánica que posteriormente dieron origen a los seres vivos, no es una idea descabellada (como lo era hace 20 años), sino que por el contrario, resulta ser una hipótesis bastante plausible. Pero independientemente de lo "plausible" que pueda ser esta hipótesis, el hecho es que el suponer que la vida en la Tierra se originó en la Tierra misma, nos conduce a una serie de

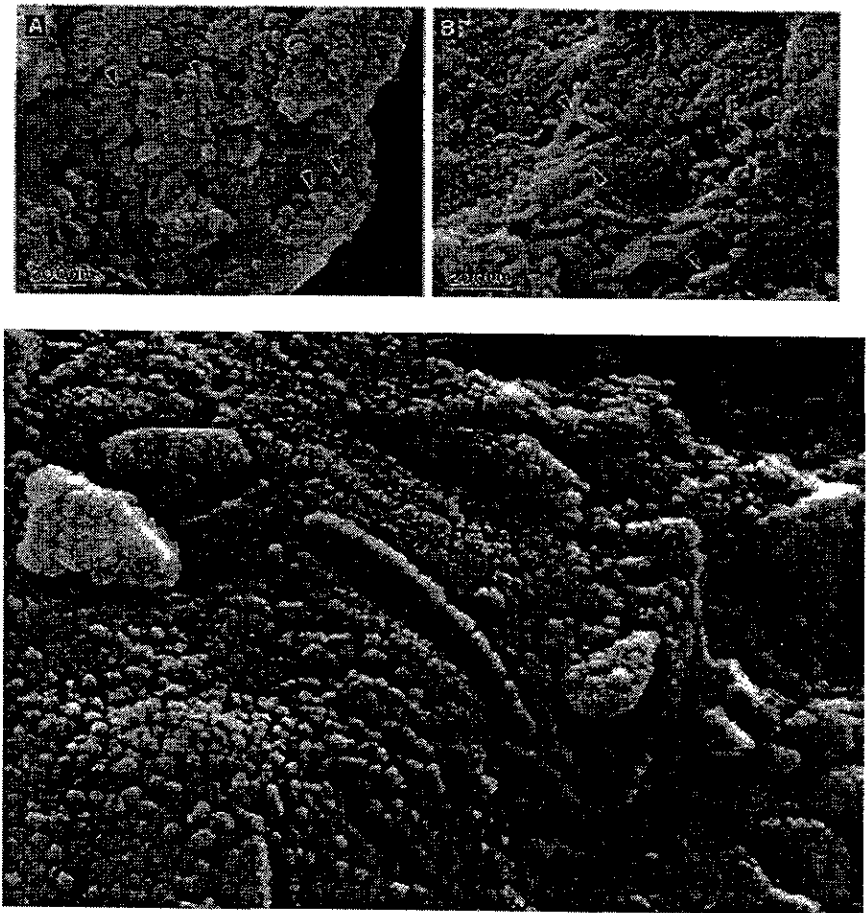


Figura 3.3: Microfotografía de las estructuras globulares que se encontraron en el meteorito marciano, y que datan de hace 3600 millones de años. Se cree que estas estructuras probablemente son fósiles de bacteria.

dificultades que no se han sabido resolver. Sin embargo, tales dificultades no se presentan si suponemos válido el escenario extraterrestre del origen de la vida:

- En el espacio exterior los procesos prebióticos y protobióticos han tenido poco más de la mitad de la edad del Universo para generarse, a partir de que se dio la nucleosíntesis de elementos pesados en el interior de las estrellas (alrededor de 10 mil millones de años). Estos procesos pudieron darse en el polvo interestelar, en superficies de meteoritos o en cometas. Por lo tanto, no es necesario suponer condiciones de autoorganización de la materia extremadamente rápidas para llegar a los organismos vivos.
- En la síntesis extraterrestre de moléculas orgánicas, existen las condiciones para romper la simetría de las *muestras racémicas de aminoácidos*, tal y como M. Greenberg y otros autores han señalado. Por un lado, existe la radiación ultravioleta circularmente polarizada proveniente de los pulsares, y por otro, la temperatura es lo suficientemente baja ( $\sim 10$  °K) como para no “opacar” el efecto de esta radiación. Como ya hemos mencionado, estas son precisamente las condiciones para obtener muestras asimétricas de enantiómeros en la síntesis de aminoácidos. De hecho, se ha encontrado en algunos meteoritos que, efectivamente, hay considerablemente más aminoácidos de un tipo que de otro [49].
- En el espacio exterior abunda el hidrógeno y casi no hay oxígeno libre. Por lo tanto, las reacciones de formación de moléculas orgánicas se llevan a cabo en ambientes reductores.
- En el espacio exterior casi no hay agua, y la poca que hay está congelada. Así, no se tienen las reacciones de hidrólisis que rompen los enlaces de polimerización de las moléculas orgánicas. Por el contrario, las arcillas y silicatos presentes en los meteoritos, cometas y polvo interestelar favorecen tales reacciones de polimerización. Consecuentemente, los procesos bioquímicos basados en el RNA sí se pueden llevar a cabo en estos medios.

### 3.4 No todo está resuelto

Aún cuando el escenario extraterrestre del origen de la vida tiene varias ventajas respecto del escenario terrestre, ya que resuelve algunos problemas importantes que surgen en este último, aún no es definitivo, y existe mucha

controversia al respecto. Los dos problemas más importantes (aunque no los únicos) que se dan en el escenario extraterrestre, son los siguientes.

- No se ha encontrado ribosa (el azúcar del RNA) para la formación de cadenas largas de ácidos nucleicos, ni en los meteoritos ni en el polvo interestelar. Aún cuando se sabe desde hace ya bastante tiempo que la ribosa puede sintetizarse fácilmente a través de reacciones químicas entre moléculas de formaldehído, las cuales sí se encuentran presentes en el espacio exterior, resulta que tales reacciones conducen a una mezcla de azúcares en la cual la ribosa siempre aparece en cantidades muy pequeñas. Lo anterior es un argumento en contra del mundo del RNA, debido a que los otros azúcares producidos en estas reacciones se pueden combinar con las bases nitrogenadas para formar productos que inhiben la síntesis y replicación del RNA. Si se tiene la ribosa en cantidad y pureza suficientes, y se añaden las bases nitrogenadas, entonces las arcillas catalizan las reacciones de síntesis de nucleótidos. Sin embargo, no se ha encontrado ribosa en el medio extraterrestre, aunque se sabe que hay abundancia de HCHO, que es precursor de azúcares.

Algunos autores han sugerido una posible solución a este problema, argumentando que en el mundo del RNA, la ribosa estaba sustituida por alguna otra clase de molécula más fácil de sintetizar (probablemente aminoácidos, como muestra la Fig.3.4). Sin embargo, aún cuando esto haya ocurrido así, todavía se tendría que resolver la cuestión de cómo estas moléculas fueron reemplazadas después por la ribosa.

- Si efectivamente la materia orgánica en la Tierra llegó del espacio exterior, viajando sobre los meteoritos y asteroides que bombardearon nuestro planeta, ¿cómo “sobrevivió” a los impactos sin desintegrarse?

Aparte de lo anterior, pueden presentarse otras dificultades, y seguramente se presentarán. El problema del origen de la vida no se ha resuelto aún, ni con el escenario terrestre ni con el extraterrestre. Probablemente, a lo más que se llegue sea a proponer escenarios plausibles para los procesos prebióticos y protobióticos, sin que seamos capaces de saber a ciencia cierta lo que realmente ocurrió hace 4000 millones de años sobre la superficie de nuestro planeta. Sin embargo, de las investigaciones actuales sobre el origen de la vida podemos concluir que, si bien el azar sigue jugando un papel importante en el escenario extraterrestre, los datos de la Naturaleza están cargados en favor de los organismos vivos si llevamos el origen de la vida al espacio exterior, mientras que con la hipótesis endógena, están en nuestra contra.

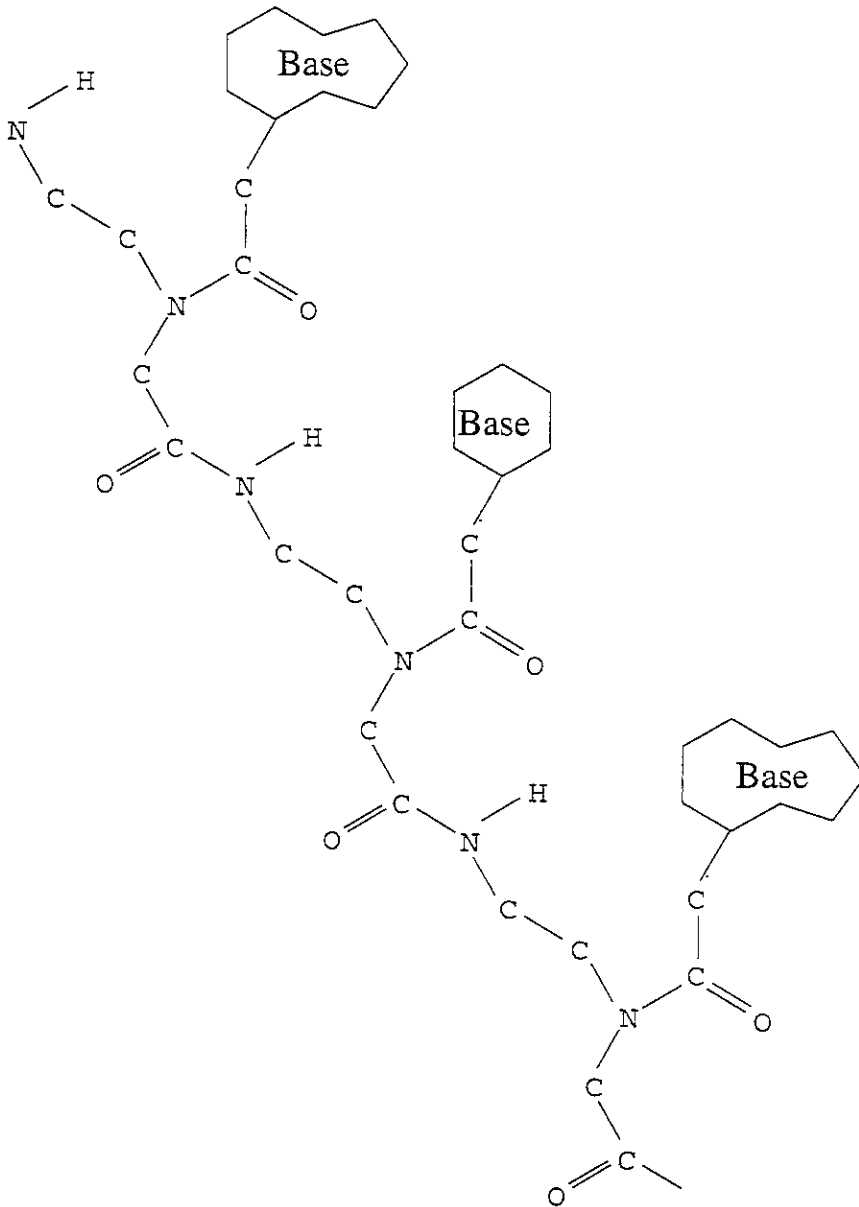


Figura 3.4: Posible molécula de “peptidil-RAN” en la que los azúcares han sido sustituidos por aminoácidos (péptidos). Se cree que en tiempos prebióticos, polímeros de este estilo se pudieron haber formado y replicado mucho más eficientemente que el RNA, el cual evolucionó a partir de ellos.

## Capítulo 4

# Potenciales Aleatorios

En el capítulo precedente vimos que los primeros procesos prebióticos que eventualmente dieron origen a los seres vivos probablemente se llevaron a cabo en un escenario extraterrestre. Si esto fue así, entonces dichos procesos prebióticos debieron darse en estructuras esencialmente *unidimensionales*, ya que la polimerización de la materia orgánica tiene lugar dentro de los filamentos tubulares de polvo interestelar, o bien, dentro de los canales de arcilla de las superficies de meteoritos y cometas. Esta hipótesis es consistente con el hecho observado actualmente en los seres vivos de que los procesos fundamentales de almacenamiento y manipulación de la información genética son procesos secuenciales realizados por moléculas cuya estructura primaria es, en esencia, unidimensional:

- el almacenamiento y la codificación de la información genética se hacen a lo largo de polímeros lineales, a saber, el DNA y el RNA;
- la perpetuación y transmisión del mensaje genético, de una generación a otra, se lleva a cabo copiando secuencialmente una molécula lineal de DNA en otra idéntica (o casi idéntica; algunas veces hay errores en la replicación);
- en la lectura del mensaje genético, el ribosoma se mueve secuencialmente a lo largo de la molécula lineal de mRNA;
- la expresión del mensaje genético también es un proceso secuencial unidimensional, ya que las proteínas se construyen añadiendo aminoácidos, uno a uno, a una cadena polipeptídica lineal (la cual después se pliega sobre sí misma, dando lugar a estructuras tridimensionales).

Claramente, si analizamos con todo detalle la estructura molecular de estos procesos, nos damos cuenta de que estrictamente no son unidimensionales, ya que a escala molecular el ribosoma, por ejemplo, tiene una estructura tridimensional que le permite alojar a dos RNA's de transferencia (cargados con sus respectivos aminoácidos), los cuales también son tridimensionales, etc. Sin embargo, las características fundamentales de estos procesos secuenciales siguen siendo unidimensionales. Para ilustrar lo anterior, podemos hacer una analogía entre los procesos de codificación, almacenamiento y expresión de la información genética en la célula, y los correspondientes procesos "informáticos" en un aparato reproductor de casetes (es decir, una grabadora). Aún cuando la grabadora es un aparato tridimensional, con una estructura muy compleja, la información que queremos escuchar se graba de forma secuencial a lo largo de una cinta magnética unidimensional. Para reproducir esta información, o para grabarla en otro casete, la cinta corre a través de la cabeza lectora de la grabadora, la cual va decodificando secuencialmente la información almacenada. En este sentido, el principio de funcionamiento de una grabadora es unidimensional, aún cuando la estructura necesaria para procesar la información sea tridimensional. Lo mismo pasa en una célula: los procesos fundamentales de almacenamiento, codificación y traducción del mensaje genético son procesos secuenciales llevados a cabo por moléculas cuyo principio de funcionamiento es, en esencia, unidimensional.

En base a las consideraciones anteriores, el *modelo que proponemos* para estudiar el movimiento del ribosoma a lo largo del mRNA es muy simple, y está esquematizado en la Fig4.1. Básicamente, consiste en un polímero unidimensional formado por  $N$  monómeros cargados, interactuando con una molécula  $M$  también cargada. A dicha molécula la llamaremos igualmente "partícula de prueba", o simplemente "partícula". El polímero juega el papel del RNA mensajero, mientras que la molécula hace las veces del ribosoma. Nuestro objetivo central es analizar el movimiento de la partícula a lo largo del polímero, e investigar bajo qué condiciones (si es que las hay), se favorece un movimiento en pasos de tres monómeros. Vale la pena mencionar por qué con un modelo tan simple, en el que el ribosoma se sustituye por una partícula puntual, esperamos obtener resultados que algo tengan que ver con la realidad.

Como ya hemos mencionado, estamos interesados en dar una explicación del *origen* de la estructura de tripletes del código genético, remontándonos a los posibles procesos prebióticos y protobióticos que tuvieron lugar hace miles de millones de años. Esto, ya de entrada, nos mete un poco en el terreno de la especulación, porque no se sabe a ciencia cierta cómo es que ocurrieron tales procesos. Sin embargo, con toda seguridad, en el origen de la vida la estructura y las relaciones biosintéticas de y entre las moléculas



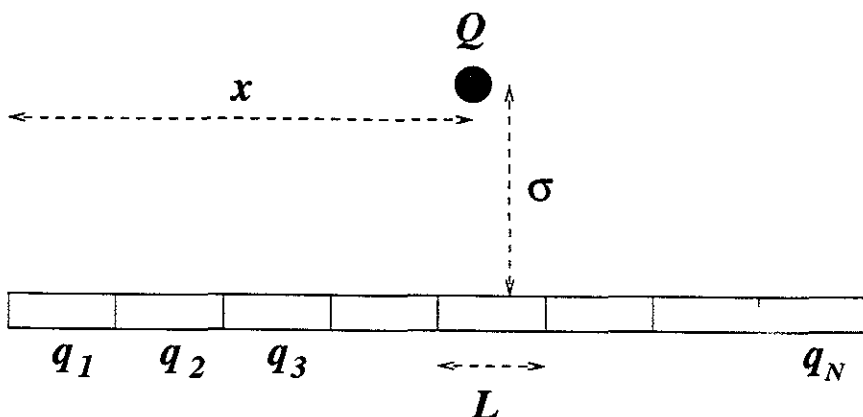


Figura 4.1: En este diagrama se ilustra el sistema que vamos a estudiar: una partícula cargada  $Q$  interactúa con un polímero cuyos monómeros están cargados aleatoriamente. Por simplicidad, supondremos que todos los monómeros tienen la misma longitud  $L$ . La partícula de prueba  $Q$  se mantiene siempre a la misma distancia  $\sigma$  perpendicular al polímero, pero puede moverse a lo largo de él;  $x$  es la posición de la partícula a lo largo del polímero respecto de uno de sus extremos.

biológicas no eran tan complicadas como lo son actualmente. Por lo tanto, debemos considerar a nuestro modelo como una primera aproximación a lo que realmente pudo haber ocurrido en el origen de la vida. Como veremos, aún cuando nuestros resultados no representan de ninguna manera la explicación absoluta, si nos muestran que los procesos por medio de los cuales se estructuró el código genético estaban sesgados favoreciendo una estructura de tripletes.

En este capítulo estableceremos los potenciales de interacción entre la partícula y el polímero, y estudiaremos algunas de sus propiedades estadísticas, mientras que en el capítulo siguiente analizaremos las propiedades dinámicas del movimiento de la partícula a lo largo del polímero.

## 4.1 El Modelo

Como mencionábamos, el modelo que estudiaremos consiste en una partícula cargada  $M$  interactuando con un polímero unidimensional cuyos monómeros están cargados y tienen todos la misma longitud  $L$ . Debemos señalar que por "carga", no nos referimos solamente a carga coulombiana. Utilizamos

este término en un sentido genérico para referirnos a carga coulombiana, momentos dipolares, momentos dipolares inducidos, o a cualquier otra propiedad de las moléculas que les permita interactuar. Supondremos que todos los monómeros que conforman el polímero tienen el mismo tipo de carga, es decir, todos son iones, o son dipolos permanentes, o son moléculas polarizables sobre las que se pueden inducir momentos dipolares, etc. De esta forma, todos los monómeros en el polímero son de la misma naturaleza, y sólo se diferencian en el valor de la carga que portan, el cual es una variable aleatoria. El asignar aleatoriamente las cargas de los monómeros a lo largo del polímero toma en cuenta el hecho de que, muy probablemente, en el origen de la vida los polímeros orgánicos no portaban todavía ninguna información, sino que los monómeros estaban distribuidos más o menos al azar a lo largo de tales polímeros.

En general, la carga en cada monómero no va a estar distribuida uniformemente, sino que estará distribuida de acuerdo con alguna función de densidad lineal de carga. Si llamamos  $\lambda_n(x)$  a la densidad lineal de carga del  $n$ -ésimo monómero, entonces en nuestro modelo  $\lambda_n(x)$  estará dada por

$$\lambda_n(x) = q_n f(x - (n-1)L), \quad x \in [(n-1)L, nL] \quad (4.1)$$

donde  $q_n$  es una variable aleatoria que puede ser discreta o continua, mientras que  $f(x)$  es una función predeterminada tal que  $\int_0^L f(x)dx = 1$ . De esta manera, la forma de la densidad de carga, dada por la función  $f(x)$ , es la misma en cada monómero, pero el valor  $q_n$  de la carga total cambia aleatoriamente de un monómero a otro<sup>1</sup>. Algunos casos de interés, esquematizados en la Fig4.2, son los siguientes:

- **Carga simétrica.** En este caso la función  $f(x)$  es simétrica respecto al centro del monómero (Fig4.2a y b).
- **Carga asimétrica.** En este caso hay más carga de un lado del monómero que de otro (Fig4.2c).

Debido a que la función  $f(x)$  es la misma para todos los monómeros, en el caso de carga asimétrica la asimetría estará orientada en la misma dirección a lo largo de todo el polímero (Fig4.2d). Esta situación es la que se observa en las moléculas tubulares proteínicas de tubulina y actina, en las que los aminoácidos polares que las conforman se orientan todos en la misma dirección [55].

<sup>1</sup>Notemos que si  $f(x) = \delta(x)$ , entonces los monómeros no son extendidos, sino que son partículas puntuales, separadas por una distancia  $L$ . Este caso lo trataremos en un capítulo posterior.

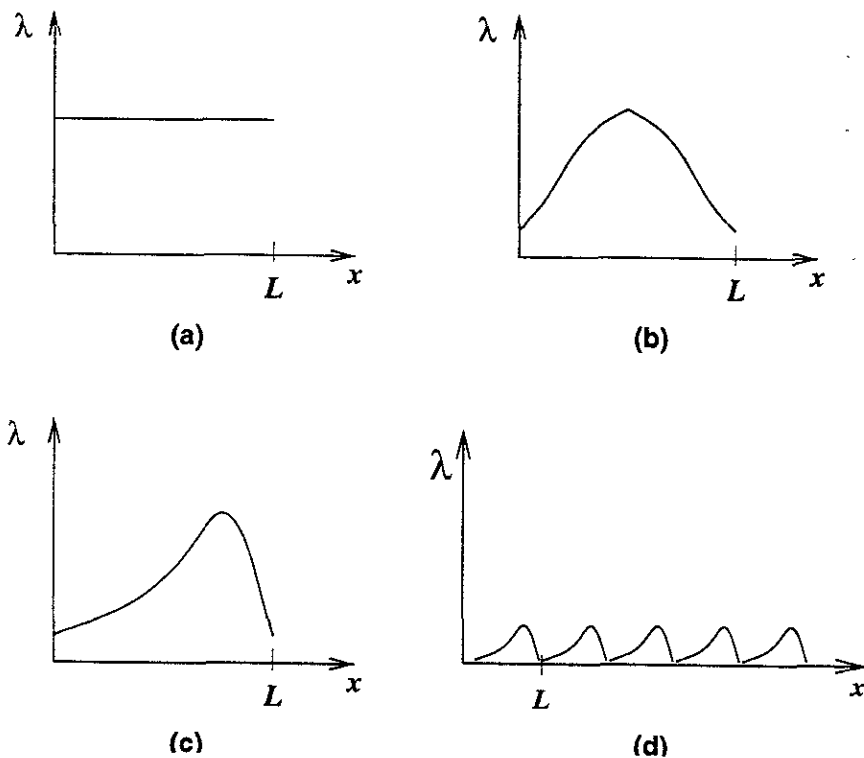


Figura 4.2: Diferentes casos que se pueden presentar de la densidad de carga de los monómeros. (a) Carga uniforme. (b) Carga simétrica no uniforme. (c) Carga asimétrica. En este último caso, supondremos que la asimetría de la carga está orientada en la misma dirección a lo largo de todo el polímero, como se ilustra en (d).

ESTA TESIS NO SALE  
DE LA BIBLIOTECA

En cuanto a las variables aleatorias  $q_n$ , supondremos que su valor está gobernado por una función densidad de probabilidad  $P(q)$ , que es la misma para todos los monómeros, y que satisface las siguientes propiedades:

$$(a) \langle q_n q_m \rangle = 0$$

(b)  $P(q) = 0$  para  $|q| > R$ , donde  $R$  es un número real positivo.

Por lo demás, la función  $P(q)$  es totalmente arbitraria. La propiedad (a) nos dice que las cargas de los monómeros son estadísticamente independientes, de tal suerte que la carga en un monómero no depende de la carga de los demás. Lo anterior no necesariamente es cierto, ya que como sabemos, "cargas opuestas se atraen y cargas iguales se repelen". Hacemos esta suposición como una primera aproximación a nuestro problema, que esperamos sea más o menos válida en tanto que las cargas de los monómeros no sean muy grandes. Para esto, la propiedad (b) establece que hay una probabilidad nula de tener monómeros con cargas excesivamente grandes (más grandes que  $R$ ), tal y como efectivamente ocurre en la naturaleza.

Aunque en nuestro análisis basta con que la función  $P(q)$  satisfaga las dos condiciones anteriores, en general vamos a trabajar con cargas discretas. Esta elección corresponde al hecho de que las moléculas pueden ionizarse perdiendo un electrón, o perdiendo dos electrones, o tres electrones, pero no perdiendo 1.5 electrones. Así, en este capítulo y en el siguiente, supondremos que la función  $P(q)$  está dada por

$$P(q) = \frac{1}{m} \sum_{k=1}^m \delta(q - \xi_k) \quad (4.2)$$

es decir, las cargas  $q_n$  pueden tomar cualquiera de los  $m$  valores  $\{\xi_1, \xi_2, \dots, \xi_m\}$  con la misma probabilidad  $1/m$  ( $\delta(x)$  es la función delta de Dirac). Si  $m = 1$ , entonces todas las cargas son iguales, y el problema es poco interesante<sup>2</sup>. Si  $m = 2$ , entonces hay dos valores de carga que se van alternando aleatoriamente a lo largo del polímero. Este caso correspondería, por ejemplo, a un "DNA binario", compuesto por sólo dos bases diferentes. En realidad, como sabemos, el DNA está compuesto por cuatro bases diferentes (A, T, C, G), situación que en nuestro modelo queda representada haciendo  $m = 4$ . Sin embargo, no nos limitaremos al caso  $m = 4$ , ya que en el origen de la vida la naturaleza pudo haber dispuesto de más de 4 bases (o de menos) para construir a las moléculas portadoras de la información genética.

<sup>2</sup>Pará que el polímero pueda codificar información se necesitan, por lo menos, dos bases diferentes.

Finalmente, para terminar con la descripción del modelo, hace falta decir que, aún cuando la partícula de prueba  $M$  puede moverse en la dirección  $x$  (esto es, a lo largo del polímero), está constreñida a permanecer a una distancia fija  $\sigma$  perpendicular al polímero. De esta manera, nuestro modelo es unidimensional. Esta suposición corresponde al hecho de que el ribosoma interactúa con el RNA mensajero por medio de interacciones electrostáticas de corto alcance (puentes de hidrógeno), de tal manera que la distancia de interacción entre el ribosoma y el mRNA se mantiene más o menos constante. Típicamente, la distancia de interacción de los puentes de hidrógeno es del orden  $\sigma \sim L/2$ , donde  $L$  es la longitud de un nucleótido ( $L \sim 1\text{nm}$ ). De ahora en adelante, haremos  $L = 1$ , es decir, tomaremos la longitud de los monómeros como nuestra unidad espacial de medida.

## 4.2 Los Potenciales

De acuerdo con el modelo que presentamos en la sección anterior, el potencial de interacción entre la partícula de prueba  $M$  y el  $n$ -ésimo monómero del polímero, está dado por

$$V_{\sigma,n}^{\alpha}(x) = KQ \int_{n-1}^n \frac{\lambda_n(x')}{[(x-x')^2 + \sigma^2]^{\alpha/2}} dx' \quad (4.3)$$

donde  $Q$  es la carga de la partícula de prueba,  $K$  una constante que depende del sistema de unidades, y  $x$  la posición de la partícula  $M$  respecto del polímero (en la expresión anterior ya hemos hecho  $L = 1$ ). El parámetro  $\alpha$  caracteriza el tipo de interacción que estamos utilizando:  $\alpha = 1$  corresponde a interacción puramente coulombiana,  $\alpha = 2$  corresponde a una interacción ión-dipolo,  $\alpha = 3$  representa una interacción ión-dipolo inducido, y así sucesivamente [50]. Vale la pena resaltar el hecho de que el parámetro  $\alpha$  es el mismo para todos los monómeros, es decir, estamos suponiendo que todos los monómeros del polímero son del mismo tipo.

El potencial de interacción total  $V_{\sigma}^{\alpha}(x)$  entre la partícula de prueba y el polímero es la suma de los potenciales (4.3) producidos por cada uno de los monómeros:

$$V_{\sigma}^{\alpha}(x) = \sum_{n=1}^N V_{\sigma,n}^{\alpha}(x) = \sum_{n=1}^N KQ \int_{n-1}^n \frac{\lambda_n(x')}{[(x-x')^2 + \sigma^2]^{\alpha/2}} dx' \quad (4.4)$$

La Fig4.3 muestra la forma de estos potenciales utilizando diferentes tipos de densidades de carga  $\lambda_n(x)$ , (es decir, diferentes funciones  $f(x)$ )

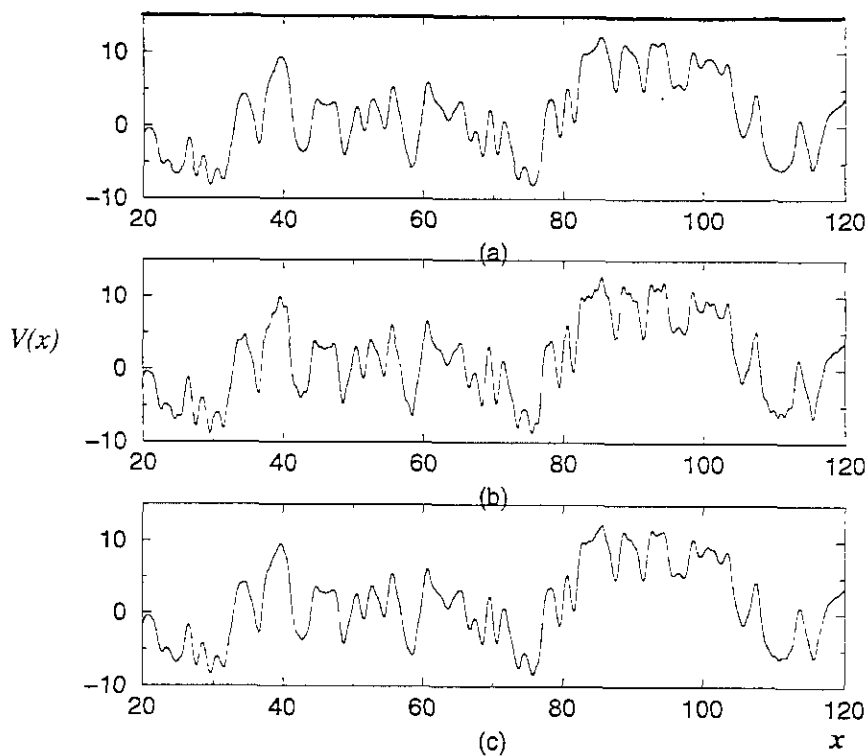


Figura 4.3: Potenciales de interacción utilizando diferentes tipos de densidades de carga para los monómeros. (a) Carga uniforme, (b) carga simétrica no uniforme y (c) carga asimétrica. Vemos que la estructura del potencial es esencialmente la misma en todos los casos. Para calcular estas gráficas se utilizó  $\alpha = 1$  y  $\sigma = 0.5$ .

—ver la ecuación (4.1)). Los casos que se muestran en esta figura son más o menos representativos de los diferentes tipos de densidad de carga que se pueden dar, y son los siguientes:

- Carga uniforme (Fig4.3a):

$$f(x) = 1, \quad x \in [0, 1] \quad (4.5)$$

- Carga simétrica no uniforme (Fig4.3b):

$$f(x) = 2 - 4 \left| x - \frac{1}{2} \right|, \quad x \in [0, 1] \quad (4.6)$$

- Carga asimétrica (Fig4.3c):

$$f(x) = \frac{1}{2} + x, \quad x \in [0, 1] \quad (4.7)$$

Las gráficas de la Fig4.3 se construyeron asignando los valores  $KQ = 1$ ,  $\sigma = 0.5$  y  $\alpha = 1$  (caso coulombiano). Además, la distribución de probabilidad  $P(q)$  que se utilizó para asignar los valores de las cargas  $q_n$  fue

$$P(q) = \frac{1}{6} \sum_{k=-3}^3 \delta(q - k) \quad (4.8)$$

Es decir, las variables  $q_n$  podían tomar cualquiera de los seis valores  $\pm 1, \pm 2, \pm 3$  con la misma probabilidad de  $1/6$ . Como puede verse en la Fig4.3, los potenciales generados con las diferentes densidades de carga  $\lambda_n(x)$  son esencialmente los mismos. Debido a lo anterior, en este capítulo y en el siguiente utilizaremos la densidad de carga (4.5), es decir, supondremos que la carga en cada monómero está distribuida uniformemente. Como veremos más adelante, los resultados de nuestro modelo no dependen fuertemente de la función  $P(q)$  que usemos, siempre y cuando se satisfagan las condiciones (a) y (b) de la sección anterior.

Utilizando entonces la densidad de carga uniforme, el potencial  $V_\sigma^\alpha(x)$  queda como

$$V_\sigma^\alpha(x) = KQ \sum_{n=1}^N q_n \int_{n-1}^n \frac{dx'}{[(x-x')^2 + \sigma^2]^{\alpha/2}} \quad (4.9)$$

Vale la pena ahora analizar el comportamiento de estos potenciales como función de los parámetros  $\alpha$  y  $\sigma$ . Para  $\alpha = 1$ ,  $\alpha = 2$  y  $\alpha = 3$  los potenciales de interacción están dados, respectivamente, por

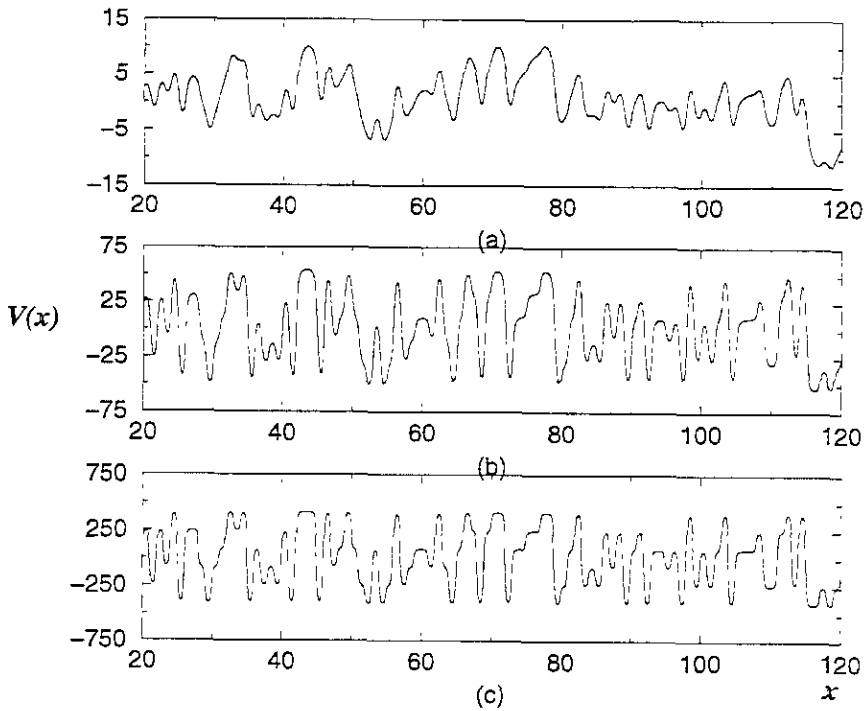


Figura 4.4: Potenciales de interacción con  $\sigma$  fija ( $\sigma = 0.5$ ) y variando el valor de  $\alpha$ . (a)  $\alpha = 1$ , (b)  $\alpha = 3$ , y (c)  $\alpha = 5$ . Las tres gráficas fueron generadas utilizando la misma realización aleatoria de cargas en el polímero. Nótese que conforme  $\alpha$  adquiere valores más grandes, la distribución de mínimos no cambia.



$$V_{\sigma}^1(x) = KQ \sum_{n=1}^N q_n \ln \left[ \frac{n-x + \sqrt{\sigma^2 + (n-x)^2}}{n-1-x + \sqrt{\sigma^2 + (n-1-x)^2}} \right] \quad (4.10)$$

$$V_{\sigma}^2(x) = \frac{KQ}{\sigma} \sum_{n=1}^N q_n \left\{ \tan^{-1} \left( \frac{n-x}{\sigma} \right) - \tan^{-1} \left( \frac{n-1-x}{\sigma} \right) \right\} \quad (4.11)$$

$$V_{\sigma}^3(x) = \frac{KQ}{\sigma^2} \sum_{n=1}^N q_n \left\{ \frac{(n-x)/\sigma}{[1 + (\frac{n-x}{\sigma})^2]^{3/2}} - \frac{(n-1-x)/\sigma}{[1 + (\frac{n-1-x}{\sigma})^2]^{3/2}} \right\} \quad (4.12)$$

En la Fig4.4 mostramos la gráfica de estos potenciales, utilizando la misma realización de las cargas aleatorias  $q_n$  en los tres casos, y los valores  $KQ = 1$ ,  $\sigma = 0.5$ . Podemos ver de esta figura que, cuando  $\alpha$  aumenta, los potenciales se vuelven más “cuadrados”, pero la estructura de máximos y mínimos se conserva. Es decir, los máximos y mínimos del potencial se vuelven más profundos y pronunciados, pero se quedan en el mismo lugar.

Por otro lado, la Fig4.5 muestra la gráfica del potencial (4.12) caso dipolar- para diferentes valores de  $\sigma$ . Vemos que cuando  $\sigma \rightarrow 0$ , ocurre otra vez que los máximos y mínimos de lo potenciales se vuelven más profundos y cuadrados, pero no cambian de posición (respecto del eje  $X$ ).

Las observaciones anteriores sugieren que la distribución estadística de mínimos en los potenciales dados en la expresión (4.4) es invariante respecto de los parámetros  $\sigma$  y  $\alpha$ , al menos para valores de  $\sigma$  suficientemente pequeños<sup>3</sup>. Notemos, por ejemplo, que en las Figs 4.3, 4.4 y 4.5, la distancia promedio entre mínimos consecutivos en el potencial es aproximadamente 3.3 (en unidades de  $L$ ). Esto puede tener consecuencias importantes en el movimiento de la partícula a lo largo del polímero, tomando en cuenta los resultados sobre máquinas moleculares de Marcelo O. Magnasco y Mark M. Millonas [53], [54], quienes en 1994 demostraron que si una molécula  $M$  se encuentra sujeta a un potencial unidimensional  $V(x)$ , el cual es periódico y asimétrico en cada periodo, como se muestra en la Fig4.6, y si además la molécula está sujeta a fluctuaciones aleatorias que no sean ruido blanco gaussiano, entonces dicha molécula se moverá en una dirección preferente (ya sea a la “izquierda” o a la “derecha”) a lo largo del potencial. La molécula se moverá de tal manera que el tiempo que tarda en un mínimo es considerablemente mayor que el tiempo que tarda en cualquier otro punto del potencial (en los máximos, por ejemplo). En otras palabras, la molécula se moverá en una dirección preferente, “brincando” de mínimo en mínimo.

<sup>3</sup>Si  $\sigma$  es muy grande (comparado con  $L$ ), entonces los mínimos se van suavizando y eventualmente desaparecen. El caso  $\sigma \rightarrow \infty$  lo trataremos en el capítulo 6.

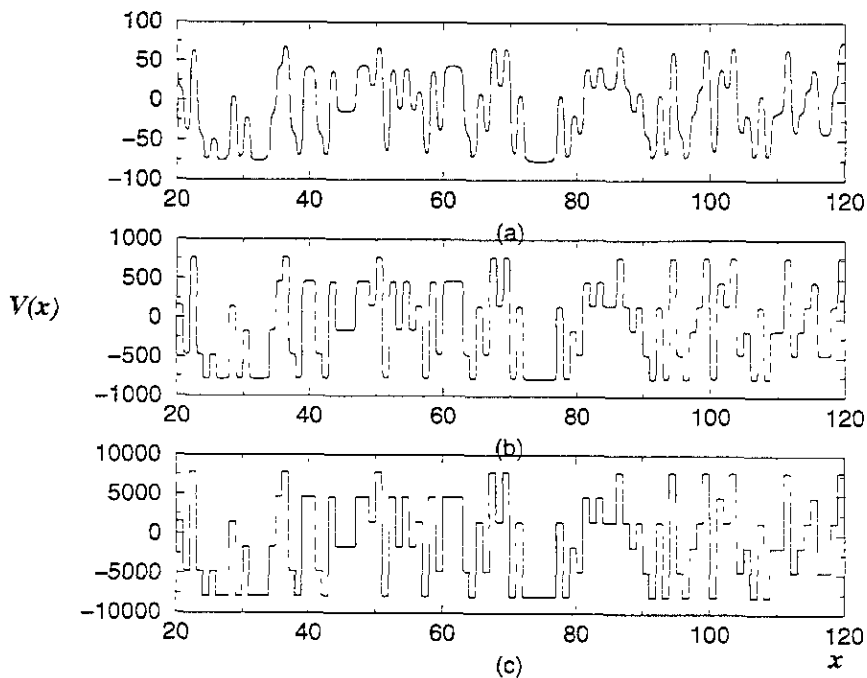


Figura 4.5: Potenciales de interacción para  $\alpha$  fijo ( $\alpha = 2$ ), pero ahora variando el valor del parámetro  $\sigma$ . (a)  $\sigma = 0.1$ , (b)  $\sigma = 0.01$  y (c)  $\sigma = 0.001$ . Notemos otra vez que la distribución de mínimos no cambia. Lo que ocurre al hacer que  $\sigma \rightarrow 0$  es que los mínimos (y los máximos) se vuelven cada vez más “cuadrados”.

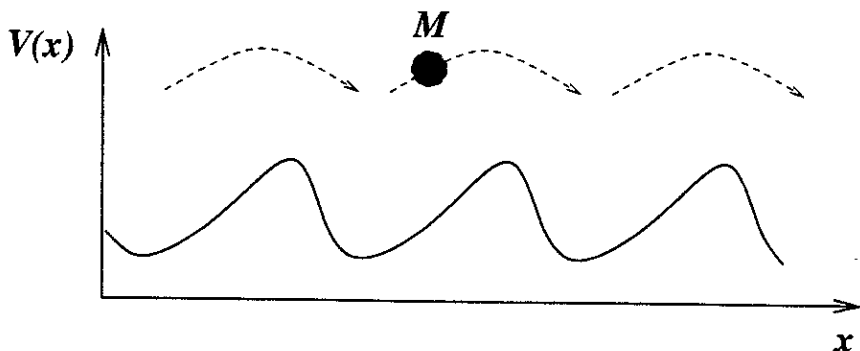


Figura 4.6: Una molécula  $M$  interactuando con un potencial  $V(x)$  asimétrico, y que está sujeta a un ruido externo coloreado, se moverá en una dirección preferente a lo largo del potencial, “brincando” de mínimo en mínimo.

Nosotros no estamos trabajando con potenciales periódicos, sino con potenciales aleatorios. Sin embargo, en la página anterior hacíamos la observación de que, aparentemente, la distancia promedio entre mínimos consecutivos en los potenciales está alrededor de 3 (en unidades de  $L$ ), y que esta distancia es independiente del valor de los parámetros  $\sigma$  y  $\alpha$ . Si esto es cierto, los trabajos de Magnasco y Millonas sugieren que la molécula  $M$  se va a mover a lo largo del polímero, en promedio, brincando en pasos de tres monómeros, siempre y cuando la distancia promedio entre mínimos consecutivos en el potencial esté cerca de 3.

Por lo tanto, resulta muy importante estudiar bajo qué condiciones los mínimos consecutivos en el potencial están separados, en promedio, por tres monómeros. Este es el tema de la siguiente sección.

### 4.3 Mínimos de energía

Las figuras 4.4 y 4.5 sugieren que la distribución de mínimos en el potencial se mantiene inalterada cuando variamos los parámetros  $\sigma$  y  $\alpha$  (dentro de cierto rango de valores de  $\sigma$ ). Lo que cambia al variar estos parámetros es la forma del potencial, el cual se vuelve más cuadrado y con mínimos más profundos. Por lo tanto, con el objetivo de encontrar la distribución de distancias entre mínimos consecutivos podemos trabajar con potenciales escalonados, como el que se muestra en la Fig4.7. Este tipo de potenciales escalonados puede considerarse como el caso límite  $\sigma \rightarrow 0$  de los potenciales suaves dados en la expresión (4.4). Para la clase particular de interacciones

que estamos considerando, los diferentes valores que adquieren los potenciales escalonados están en correspondencia con los valores  $\{\xi_1, \xi_2, \dots, \xi_m\}$  que adquieren las cargas  $q_n$  a lo largo del polímero. Así, una carga  $q_n$  de valor grande corresponderá a un escalón alto en el potencial, mientras que una carga  $q_n$  de valor pequeño corresponderá a un escalón también pequeño. Como cada una de las cargas  $q_n$  puede tomar  $m$  valores al azar, tendremos escalones de  $m$  tamaños diferentes en el potencial, secuenciados en correspondencia con los valores de las cargas  $q_n$ .

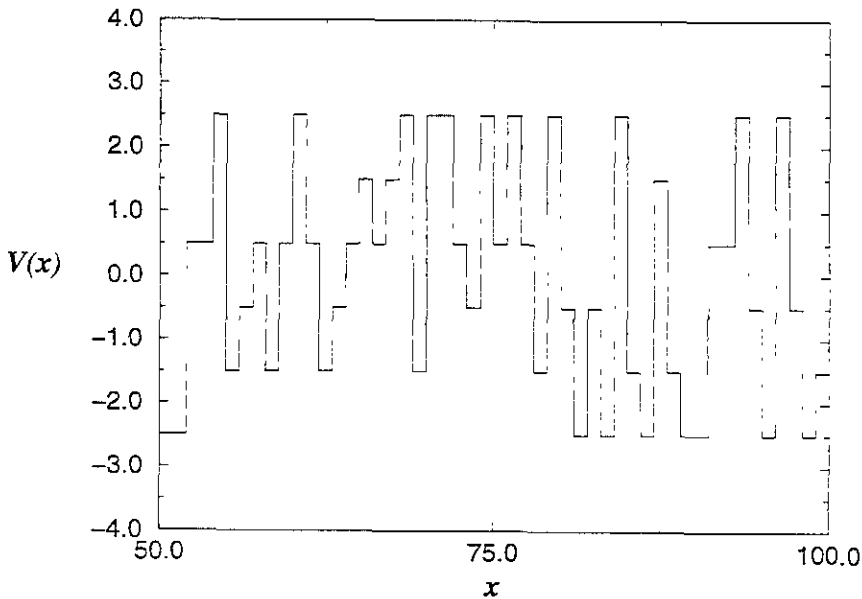


Figura 4.7: Para calcular la distribución de mínimos de energía a lo largo del potencial, es conveniente trabajar con potenciales escalonados como el que se muestra en esta figura. Este tipo de potenciales se obtienen de los potenciales suaves dados en la ecuación (4.4) en el límite  $\sigma \rightarrow 0$ .

Como se observa en las Figs. 4.4 y 4.5, para calcular la distribución de distancias entre mínimos consecutivos en el potencial, los potenciales escalonados son completamente equivalentes a los potenciales continuos dados en (4.4). Por lo tanto, supondremos que el potencial  $V_n(x)$  generado por el  $n$ -ésimo monómero del polímero es un potencial escalón de la forma:

$$V_n(x) = \begin{cases} V_n & x \in [(n-1), n] \\ 0 & x \notin [(n-1), n] \end{cases} \quad (4.13)$$

donde  $V_n$  es una variable aleatoria discreta que puede tomar los valores  $\{v_1, v_2, \dots, v_m\}$  con la misma probabilidad. De esta manera, el potencial escalonado con el que trabajaremos está dado por

$$V(x) = \sum_{n=1}^N V_n [H(x - (n-1)) - H(x - n)] \quad (4.14)$$

donde  $H(x)$  es la función de Heaviside:

$$H(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

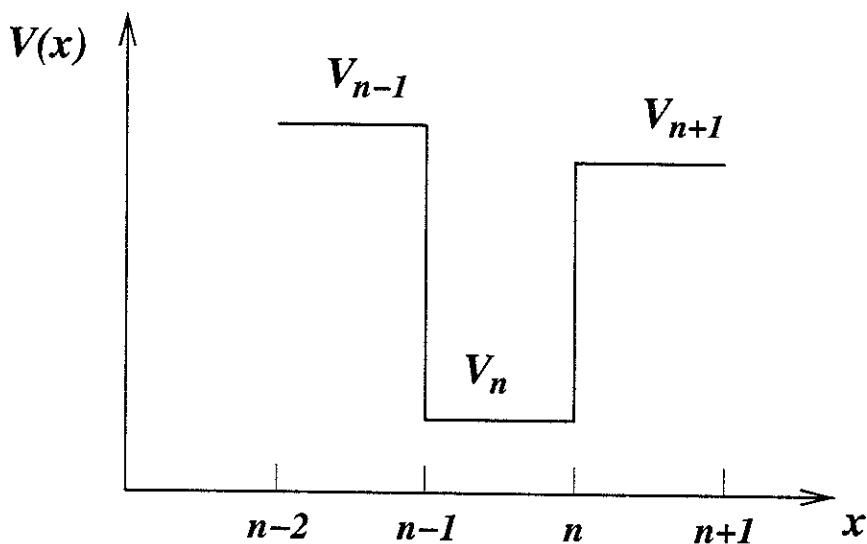


Figura 4.8: Esquema de mínimo simple. Para que el potencial  $V_n$  corresponda a un mínimo, es necesario que su valor sea menor al valor de los potenciales vecinos,  $V_{n-1}$  y  $V_{n+1}$ . Por convención, tomamos el valor "mínimo" del potencial  $V_n$  en el centro del intervalo, es decir, en  $x = n - (1/2)$ .

Para que el potencial  $V(x)$  tenga un mínimo simple en la posición  $x = n - 1/2$ , es necesario que el potencial  $V_n$  generado por el  $n$ -ésimo monómero sea menor que los potenciales  $V_{n-1}$  y  $V_{n+1}$  generados por los monómeros vecinos, es decir,  $V_n$  corresponde a un mínimo simple siempre y cuando

$$V_n < V_{n-1} \quad y \quad V_n < V_{n+1}$$

La situación se ilustra en la Fig4.8. Para calcular la probabilidad de que  $V_n$  corresponda a un mínimo simple, supondremos que los valores  $\{v_1, v_2, \dots, v_m\}$  están ordenados de tal forma que

$$v_1 < v_2 < \dots < v_m \quad (4.15)$$

Si el potencial  $V_n$  adquirió el valor  $v_1$  (el menor de todos los posibles valores), entonces será un mínimo si y sólo si los potenciales  $V_{n-1}$  y  $V_{n+1}$  adquirieron alguno de los  $m - 1$  valores restantes  $v_2, v_3, \dots, v_m$ , que son más grandes que  $v_1$ . Si llamamos  $P_1$  a la probabilidad de que  $V_n$  corresponda a un mínimo simple justo cuando adquirió el valor  $v_1$ , entonces se tiene:

$$\begin{aligned} P_1 &= \text{Prob}(V_{n-1} = v_2, v_3, \dots, v_m) \times \text{Prob}(V_n = v_1) \\ &\times \text{Prob}(V_{n+1} = v_2, v_3, \dots, v_m) \\ &= \frac{m-1}{m} \times \frac{1}{m} \times \frac{m-1}{m} \\ &= \frac{(m-1)^2}{m^3} \end{aligned}$$

Por otro lado, si  $V_n$  adquirió el valor  $v_2$ , para que sea mínimo simple es necesario que los potenciales  $V_{n-1}$  y  $V_{n+1}$  hayan adquirido alguno de los  $m - 2$  valores restantes  $v_3, v_4, \dots, v_m$  que están por arriba de  $v_2$ . Si  $P_2$  es la probabilidad de que  $V_n$  corresponda a un mínimo simple justo cuando adquirió el valor  $v_2$ , entonces se tiene

$$\begin{aligned} P_2 &= \text{Prob}(V_{n-1} = v_3, v_4, \dots, v_m) \times \text{Prob}(V_n = v_2) \\ &\times \text{Prob}(V_{n+1} = v_3, v_4, \dots, v_m) \\ &= \frac{m-2}{m} \times \frac{1}{m} \times \frac{m-2}{m} \\ &= \frac{(m-2)^2}{m^3} \end{aligned}$$

Continuando con este proceso, vemos que la probabilidad  $P_k$  de que el potencial  $V_n$  corresponda a un mínimo simple justo cuando adquirió el valor  $v_k$ , está dada por

$$P_k = \frac{(m - k)^2}{m^3}$$

Por lo tanto, la probabilidad  $P_s$  de que el potencial  $V_n$  sea un mínimo simple, independientemente del valor que haya adquirido, es la suma de todas las  $P_k$ 's, lo cual da como resultado

$$P_s = \sum_{k=1}^{m-1} P_k = \frac{m(m-1)(2m-1)}{6m^3} \quad (4.16)$$

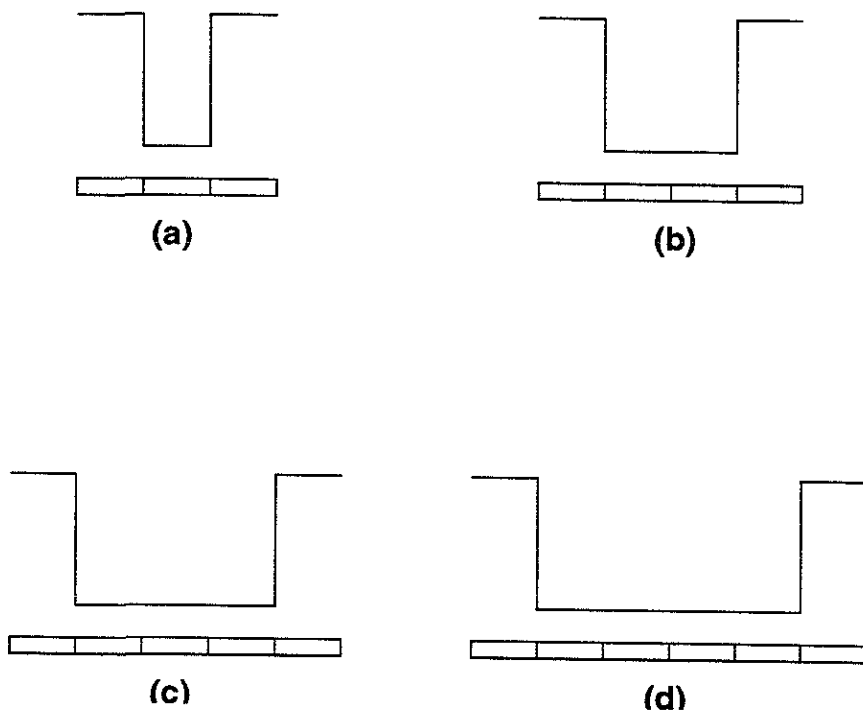


Figura 4.9: Representación esquemática de mínimos planos. (a) Mínimo simple, (b) mínimo plano binario, (c) mínimo plano ternario, etc. Los mínimos planos resultan cuando varias cargas consecutivas a lo largo del polímero adquieren el mismo valor.

Hasta el momento no hemos tomado en cuenta la posible presencia de mínimos planos en el potencial, como los que se muestran en la Fig4.9. Así, además de tener mínimos simples, podemos tener mínimos planos binarios, o ternarios, o cuaternarios, etc. La probabilidad de tener un mínimo simple es  $P_s$ ; la probabilidad de tener un mínimo plano binario es  $P_s \times \frac{1}{m}$ ; la probabilidad de tener un mínimo plano ternario es  $P_s \times \frac{1}{m^2}$ ; etc. Por lo tanto, la probabilidad  $P_{\min}$  de que  $V_n$  corresponda a un mínimo, simple, o binario, o ternario, o del que sea, está dada por

$$\begin{aligned} P_{\min} &= P_s \times \left[ 1 + \frac{1}{m} + \frac{1}{m^2} + \cdots + \frac{1}{m^{N-2}} \right] \\ &= P_s \times \frac{1 - (1/m)^{N-1}}{1 - (1/m)} \end{aligned}$$

Si el número  $N$  de monómeros en el polímero es muy grande, entonces el término  $(1/m)^{N-1}$  que aparece en la expresión anterior puede desprejarse. Tomando en cuenta la expresión (4.16) para  $P_s$ , obtenemos que la probabilidad  $P_{\min}$  de que  $V_n$  corresponda a un mínimo (simple, binario, ternario, etc.) está dada por

$$P_{\min} = \frac{2m - 1}{6m} \quad (4.17)$$

Como todos los potenciales  $V_n$  son equivalentes, la expresión anterior da la probabilidad de ocurrencia de mínimos en el potencial, en el caso en que las cargas  $q_n$  sean discretas y estén distribuidas de acuerdo con la ecuación (4.2).  $P_{\min}$  es entonces la densidad de mínimos en el potencial, y  $1/P_{\min}$  es la distancia promedio  $\bar{d}$  entre mínimos consecutivos:

$$\bar{d} = \frac{6m}{2m - 1} \quad (4.18)$$

La Fig4.10 muestra la gráfica de  $\bar{d}$  como función de  $m$ . Como podemos observar, excepto para el caso  $m = 2$ , en el cual se tiene  $\bar{d} = 4$ , resulta que la distancia promedio entre mínimos de energía consecutivos a lo largo del polímero es una cantidad muy cercana a 3. En particular, para el caso  $m = 4$  (DNA) la distancia promedio es  $\bar{d} \approx 3.4$ . Además, conforme  $m$  aumenta, la distancia promedio  $\bar{d}$  se aproxima más y más a 3, de tal manera que  $\lim_{m \rightarrow \infty} \bar{d} = 3$ . El resultado contenido en la ecuación (4.18) se resume en la siguiente afirmación::



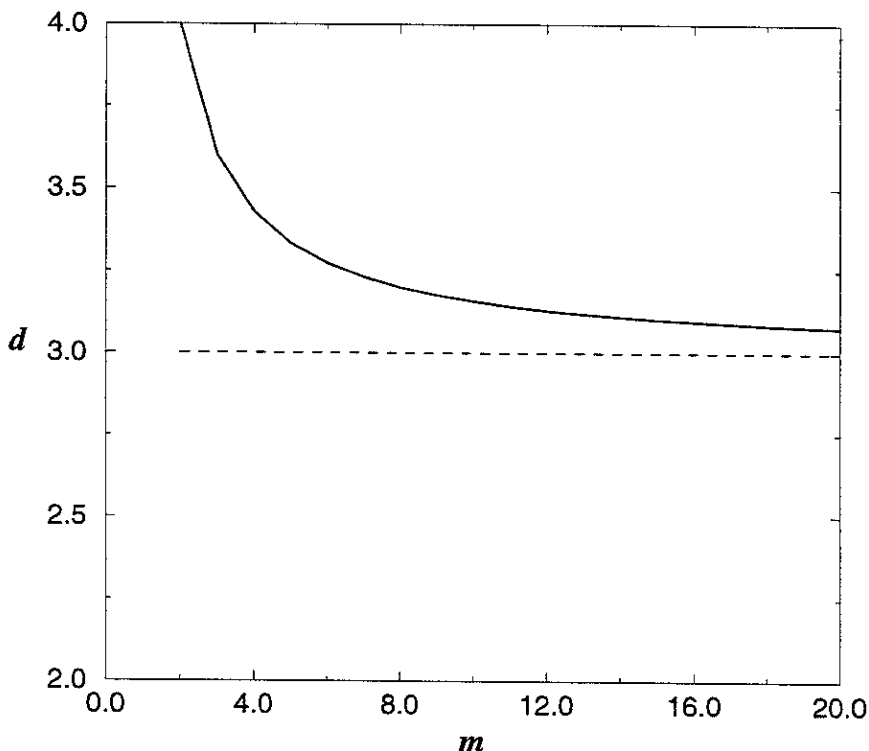


Figura 4.10: Gráfica de la distancia promedio entre mínimos de energía consecutivos como función del número  $m$  de cargas diferentes a lo largo del polímero. Noté que conforme  $m$  aumenta, la distancia promedio  $\bar{d}$  tiende a tres. Para  $m = 4$  se tiene  $\bar{d} \simeq 3.4$ .

*Basta con que existan más de dos tipos diferentes de monómeros, para que la distancia promedio entre mínimos de energía consecutivos a lo largo del polímero sea un número menor que cuatro y mayor que tres.*

El resultado anterior sugiere que ya desde el origen de la vida había una “predilección energética” por el 3. Puesto en otras palabras, si nuestro modelo (esquematizado en la Fig4.1) refleja en algo las condiciones primitivas de la interacción entre el ribosoma y el RNA mensajero (o las moléculas

equivalentes en aquella época), entonces, independientemente del número  $m$  de bases con las cuales se hubiese construido el mRNA, el número 3 estaba privilegiado, en el sentido de que la distancia promedio entre mínimos consecutivos de la energía de interacción estaba muy cerca de tres. Esta regularidad en los potenciales de interacción, se traduciría en una regularidad en la dinámica del sistema, ya que según Magnasco y Millonas, el ribosoma primitivo se movería, preferentemente, de mínimo en mínimo. Debido a esto, la maquinaria actual de síntesis de proteínas probablemente se construyó tomando como base un sistema (originalmente muy simple), que ya se movía preferentemente de tres en tres. Así, la estructura de tripletes del código genético pudiera ser un reflejo de esta "dinámica primitiva" que preferenciaba el número tres.

Sin embargo, hasta el momento sólo hemos analizado las propiedades promedio de los potenciales de interacción, y no hemos analizado la dinámica del sistema. Este será el tema del siguiente capítulo, en donde veremos que, efectivamente, la regularidad en los potenciales de interacción se ve reflejada en una regularidad del movimiento de la partícula a lo largo del polímero. Por el momento, en la sección siguiente nos dedicaremos al cálculo de la distribución de las distancias entre mínimos consecutivos en la energía potencial, y veremos cómo la naturaleza, de alguna forma, ha optimizado las secuencias genéticas de los organismos vivos para preferenciar, aún más, al número 3.

#### 4.4 Distribución de distancias entre mínimos de energía

Aparte del valor promedio  $\bar{d}$  de las distancias entre mínimos consecutivos de energía, también es de interés la distribución de estas distancias, es decir, la probabilidad  $P(d)$  de que dos mínimos consecutivos se encuentren separados una distancia  $d$ . Para calcular esta probabilidad, seguiremos trabajando con el modelo de potenciales escalonados dado por las ecuaciones (4.13) y (4.14). Por lo tanto, tenemos una serie de valores de energía  $\{v_1, v_2, \dots, v_m\}$  distribuidos aleatoriamente a lo largo del polímero, como se muestra en la Fig4.7. Debido a la presencia de mínimos planos en este tipo de potenciales, tomaremos la convención de medir la distancia entre dos mínimos cualesquiera desde sus centros, como se indica esquemáticamente en la Fig4.11. Esto trae como resultado que la distancia  $d$  sólo pueda adquirir valores enteros o semienteros.

Por otro lado, es claro que la probabilidad  $P(d)$  dependerá también del número  $m$  de valores posibles de energía. Por lo tanto, de ahora en

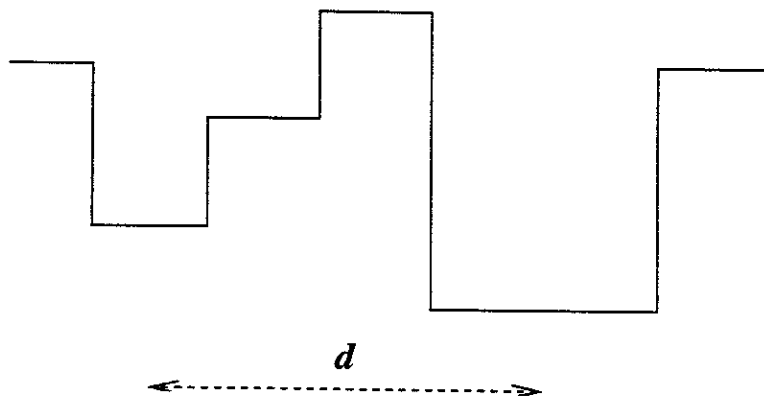


Figura 4.11: Por convención, la distancia entre dos mínimos planos consecutivos la mediremos a partir del centro de cada uno de ellos.

adelante designaremos como  $\mathbf{P}_m(d)$  a la probabilidad de tener dos mínimos de energía consecutivos separados por una distancia  $d$ , cuando hay  $m$  valores posibles de energía.

Para calcular  $\mathbf{P}_m(d)$  conviene definir dos eventos,  $\mathcal{A}$  y  $\mathcal{B}$ , y una restricción  $\mathcal{R}$ , de la manera siguiente:

- Evento  $\mathcal{A}$ : hay un mínimo de energía en la posición  $x_A$ .
- Evento  $\mathcal{B}$ : hay un mínimo de energía en la posición  $x_B$ .
- Restricción  $\mathcal{R}$ : no hay ningún mínimo de energía para posiciones  $x$  tales que  $x_A < x < x_B$ .

Con estas definiciones, si las posiciones  $x_A$  y  $x_B$  están separadas por una distancia  $d$ , entonces  $\mathbf{P}_m(d)$  es una probabilidad condicional dada por

$$\mathbf{P}_m(d) = \mathbf{P}_m(\mathcal{B} \mid \mathcal{A}; d)_{\mathcal{R}} \quad (4.19)$$

es decir,  $\mathbf{P}_m(d)$  es la probabilidad de tener un mínimo en  $x_B$ , dado que se tuvo un mínimo en  $x_A$ , y que no hay ningún mínimo en medio. Hemos escrito explícitamente el parámetro  $d = x_B - x_A$  dentro de la probabilidad condicional, para indicar que dicha probabilidad es función de este parámetro. Además, la restricción  $\mathcal{R}$  garantiza la consecutividad de los mínimos en  $x_A$  y  $x_B$ .

De acuerdo con la definición de probabilidad condicional, la ecuación anterior puede escribirse como

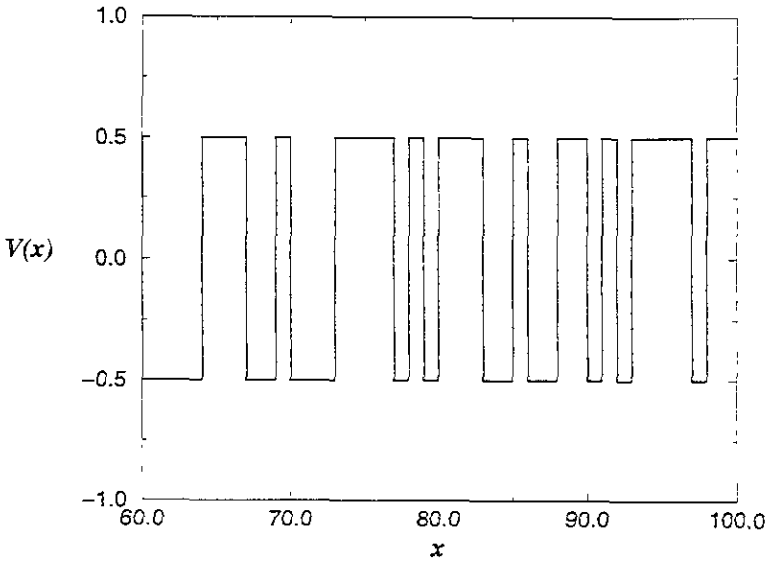


Figura 4.12: Potencial escalonado generado cuando  $m = 2$ . En este caso, los dos valores  $v_1$  y  $v_2$  que se van alternando son  $-0.5$  y  $+0.5$ , respectivamente.

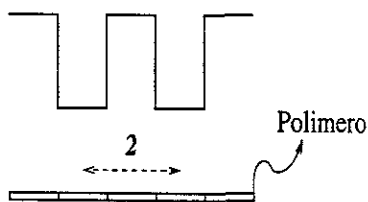
$$P_m(d) = \frac{P_m(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}}}{P_m(\mathcal{A})}$$

donde  $P_m(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}}$  es la probabilidad de tener mínimos en  $x_A$  y en  $x_B$  separados una distancia  $d$ , sin ningún mínimo en medio, mientras que  $P_m(\mathcal{A})$  es simplemente la probabilidad de tener un mínimo en  $x_A$ . Debido a que  $x_A$  es un punto arbitrario,  $P_m(\mathcal{A})$  es la probabilidad de tener un mínimo, en el lugar que sea, a lo largo del polímero. Esta última probabilidad está dada por la ecuación (4.17). Por lo tanto,  $P_m(d)$  queda como

$$P_m(d) = \frac{6m}{2m-1} P_m(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}} \quad (4.20)$$

Lo que resta por calcular es entonces  $P_m(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}}$ . La idea central es contar el número de configuraciones en las que se tienen dos mínimos consecutivos separados una distancia  $d$ , y luego dividir entre el número total de configuraciones. Ilustraremos este cálculo para el caso más fácil, que es  $m = 2$ . Los casos  $m = 3$  y  $m = 4$  son similares.

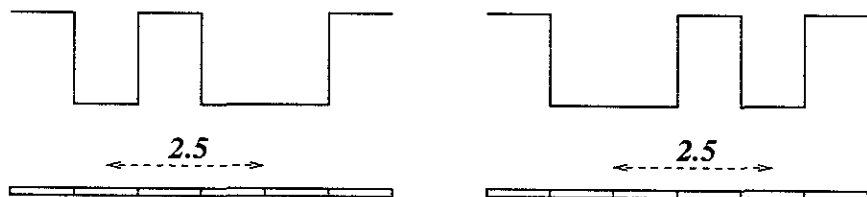
Si  $m = 2$ , entonces sólo se tienen dos valores de energía,  $\{v_1, v_2\}$  que se van alternando aleatoriamente a lo largo del polímero, como se muestra en la Fig4.12. Notemos que, de acuerdo con la convención de medir la distancia entre dos mínimos desde su parte media, la distancia más pequeña que puede haber entre dos mínimos consecutivos es  $d = 2$ , lo cual se puede obtener sólomente con la siguiente configuración:



La probabilidad  $\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 2)_{\mathcal{R}}$  en este caso es igual  $1/2^5$ :

$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 2)_{\mathcal{R}} = \frac{1}{2^5}$$

Por otro lado, las siguientes dos configuraciones, y solamente estas dos, nos dan una distancia entre mínimos de 2.5:



Por lo tanto, la probabilidad  $\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 2.5)_{\mathcal{R}}$  está dada por

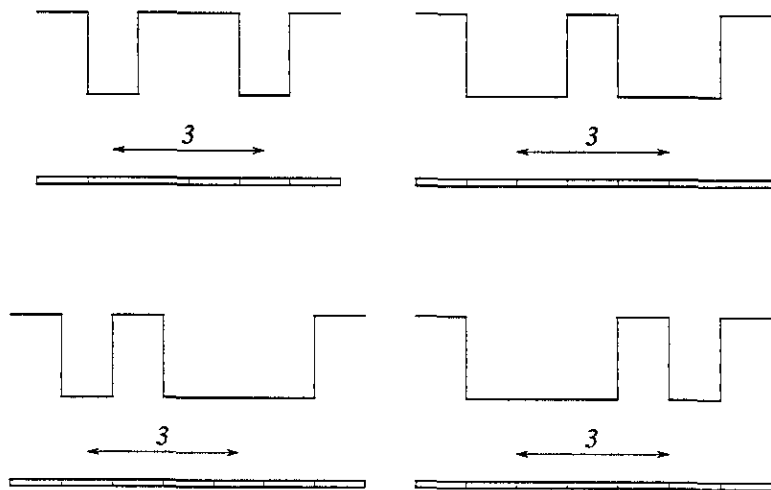
$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 2.5)_{\mathcal{R}} = \frac{1}{2^6} + \frac{1}{2^6} = \frac{2}{2^6}$$

En la siguiente página se muestran las cuatro configuraciones con las que se obtiene  $d = 3$  y las seis configuraciones que conducen a  $d = 3.5$ . Como puede verse,  $\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 3)_{\mathcal{R}}$  y  $\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 3.5)_{\mathcal{R}}$  están dadas por

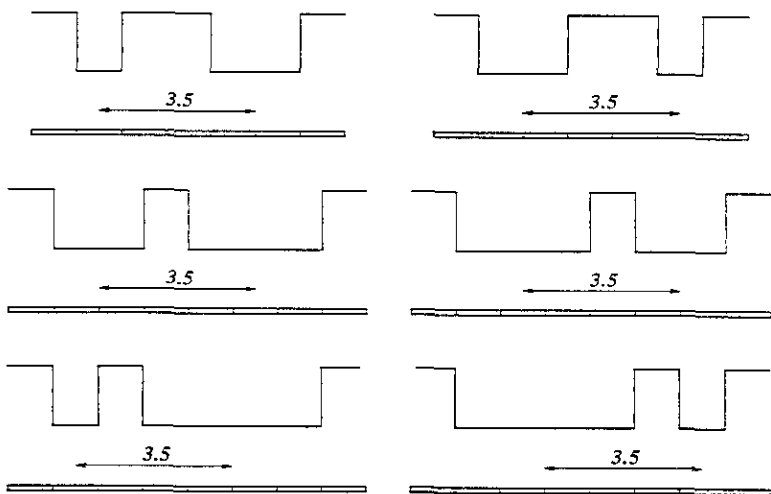
$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 3)_{\mathcal{R}} = \frac{1}{2^5} + \frac{3}{2^6}$$

$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; 3.5)_{\mathcal{R}} = \frac{2}{2^6} + \frac{4}{2^7}$$

Configuraciones en las que se tiene  $d = 3$



Configuraciones en las que se tiene  $d = 3.5$



Siguiendo con este tipo de conteo, rápidamente se llega a la conclusión de que cuando  $d$  es entero,  $\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}}$  está dada por

$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}} = \frac{1}{2^{d+3}} \sum_{k=0}^{d-2} \frac{2k+1}{2^k}, \quad d = 2, 3, 4, \dots \quad (4.21)$$

mientras que si  $d$  es un número semientero se tiene

$$\mathbf{P}_2(\mathcal{B} \cap \mathcal{A}; d)_{\mathcal{R}} = \frac{2}{2^{d'+4}} \sum_{k=0}^{d'-2} \frac{2k+1}{2^k}, \quad \begin{array}{l} d = 2.5, 3.5, 4.5, \dots \\ d' = \text{Int}[d] \end{array} \quad (4.22)$$

Insertando los resultados anteriores en la ecuación (4.20), y haciendo  $m = 2$ , obtenemos:

$$\boxed{m = 2}$$

$$\mathbf{P}_2(d) = 4 \frac{1}{2^{d+3}} \sum_{k=0}^{d-2} \frac{2k+1}{2^k}, \quad d = 2, 3, 4, \dots \quad (4.23)$$

$$\mathbf{P}_2(d) = 4 \frac{2}{2^{d'+4}} \sum_{k=0}^{d'-2} \frac{2k+1}{2^k}, \quad \begin{array}{l} d = 2.5, 3.5, 4.5, \dots \\ d' = \text{Int}[d] \end{array} \quad (4.24)$$

Esto determina completamente la distribución de distancias entre mínimos consecutivos de energía a lo largo del polímero, al menos para el caso  $m = 2$  que estamos considerando. En la Fig4.13a se muestra la gráfica de  $\mathbf{P}_2(d)$ , utilizando las ecuaciones (4.23) y (4.24). Un hecho que llama la atención de esta gráfica es que el valor máximo de  $\mathbf{P}_2(d)$  se encuentra en  $d = 3$ . Es decir, aún cuando la distancia promedio es  $\bar{d} = 4$ , resulta que la distancia más probable es 3. Esto es un hecho muy favorable para el problema que estamos atacando, ya que si la partícula de prueba se mueve de mínimo en mínimo a lo largo del polímero, incluso para  $m = 2$  la partícula se moverá de tres en tres la mayor parte del tiempo, aunque en promedio lo haga de cuatro en cuatro.

Por otro lado, recordemos que las ecuaciones (4.23) y (4.24) las obtuvimos para potenciales escalonados. Sin embargo, la Fig4.13b muestra la distribución de distancias entre mínimos consecutivos de energía para los potenciales continuos dados en la expresión (4.10), obtenida numéricamente. Para esto, dejamos que las cargas  $q_n$  toman los valores  $\pm 1$  con la misma

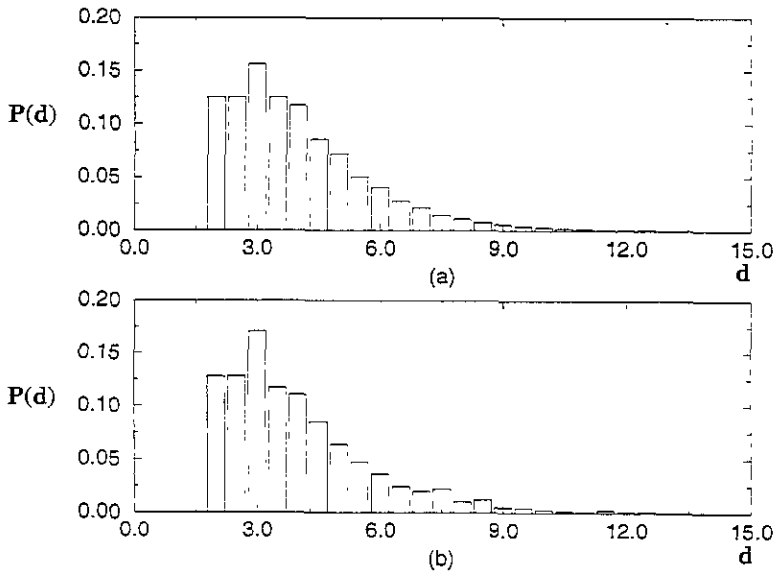


Figura 4.13: Distribución de distancias entre mínimos consecutivos de energía para el modelo de potenciales escalonados con  $m = 2$ . (a) Gráfica de las ecuaciones analíticas (4.23) y (4.24). (b) Histograma obtenido numéricamente a partir de una realización del potencial continuo (4.10) con  $\sigma = 0.5$  ( $\alpha = 1$ ). Como puede apreciarse, el modelo de potenciales escalonados es bastante bueno para describir la estructura de máximos y mínimos de potenciales continuos del tipo dado en (4.4).



probabilidad, de tal manera que sólo se tuvieran dos "estados" de energía potencial para cada monómero. Vemos que el acuerdo con el resultado teórico es muy bueno.

Vemos también que el modelo de potenciales escalonados que hemos estado utilizando es adecuado para describir el tipo de potenciales continuos dados en la expresión (4.4). Lo anterior se debe a que los potenciales escalonados se obtienen de los potenciales continuos a través de un proceso de límite ( $\sigma \rightarrow 0$ , o bien  $\alpha \rightarrow \infty$ ), en el cual los mínimos del potencial no cambian de lugar. Esto ya lo habíamos observado en las Figs 4.4 y 4.5, y fue lo que nos motivó a utilizar el modelo de potenciales escalonados para determinar la distribución de distancias entre mínimos. Sin embargo, no está por demás enfatizar la consistencia de nuestra suposición.

Se puede aplicar la misma idea de contar configuraciones en los casos  $m = 3$  y  $m = 4$  con el fin de obtener  $P_3(d)$  y  $P_4(d)$ , respectivamente. De hecho, lo ideal sería encontrar la expresión general de  $P_m(d)$  para  $m$  arbitrario, pero esto aún no se ha podido hacer. Los resultados para  $m = 3$  y  $m = 4$  son los siguientes:

$$\boxed{m = 3}$$

$$P_3(d) = \frac{18}{5} \frac{1}{3^{d+3}} \sum_{k=0}^{d-2} \frac{2k+1}{3^k} N_3[d-k], \quad d = 2, 3, 4, \dots \quad (4.25)$$

$$P_3(d) = \frac{18}{5} \frac{2}{3^{d'+4}} \sum_{k=0}^{d'-2} \frac{2k+1}{3^k} N_3[d-k], \quad d = 2.5, 3.5, 4.5, \dots, \quad d' = \text{Int}[d] \quad (4.26)$$

$$N_3[d] = 1 + 2d + 2d^2 \quad (4.27)$$

$$\boxed{m = 4}$$

$$P_4(d) = \frac{24}{7} \frac{1}{4^{d+3}} \sum_{k=0}^{d-2} \frac{2k+1}{4^k} N_4[d-k], \quad d = 2, 3, 4, \dots \quad (4.28)$$

$$P_4(d) = \frac{24}{7} \frac{2}{4^{d'+4}} \sum_{k=0}^{d'-2} \frac{2k+1}{4^k} N_4[d-k], \quad d = 2.5, 3.5, 4.5, \dots, \quad d' = \text{Int}[d] \quad (4.29)$$

$$N_4[d] = -2 + \frac{1}{8} \{30d + 73d^2 + 22d^3 + 3d^4\} \quad (4.30)$$

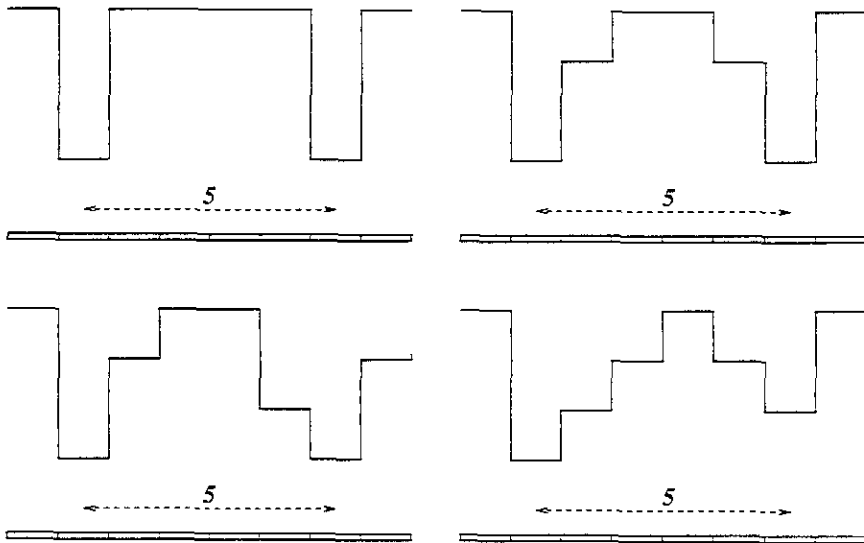


Figura 4.14: Cuatro configuraciones de mínimo simple en las que se tiene  $d = 5$ . Esta “degeneración de mínimo simple” se presenta siempre que  $m > 2$ .

Al comparar las ecuaciones (4.23), (4.25) y (4.28) puede verse que tienen una estructura muy similar. Sin embargo, en las expresiones (4.25) y (4.28) aparecen los polinomios  $N_3[d]$  y  $N_4[d]$ , respectivamente. Esto se debe a que para  $m > 2$  hay varias maneras de tener dos mínimos simples consecutivos separados por la misma distancia. Por ejemplo, las cuatro configuraciones que se muestran en la Fig4.14 presentan mínimos simples separados por una distancia  $d = 5$ . Claramente, aparte de las cuatro configuraciones mostradas, hay un gran número de otras configuraciones que presentan mínimos simples separados 5 unidades, y este número crece conforme aumenta  $m$ . Así, los polinomios  $N_3[d]$  y  $N_4[d]$  toman en cuenta esta “degeneración de mínimo simple”; es decir, estos polinomios dan el número de todas las posibles configuraciones de mínimos simples separados por una distancia fija  $d$ . En el caso  $m = 2$  no se presenta esta degeneración de configuraciones de mínimo simple, ya que sólo hay una manera en la que se pueden tener dos mínimos simples separados una distancia fija  $d$ .

La dificultad de calcular  $\mathbf{P}_m(d)$  para  $m$  arbitraria radica precisamente en encontrar la degeneración de mínimo simple  $N_m[d]$ . Como puede inferirse de las expresiones (4.27) y (4.30), la generalización no parece ser trivial.

Otro caso de interés se presenta cuando  $m \rightarrow \infty$ , ya que en algunas situaciones, el número de valores que puede adquirir el potencial escalona-

do es mucho mayor que el número de valores de carga en los monómeros. Por ejemplo, K. J. Breslauer y sus colaboradores idearon un método para determinar la estabilidad de la molécula de DNA a partir de la secuencia de bases que la conforman [51]. Este modelo está basado en las energías de interacción entre *parejas de bases* a lo largo del DNA. Breslauer midió experimentalmente las energías de interacción entre todas las parejas de bases que se pueden formar, que son 16, y mostró que a partir de estos valores se puede predecir con buena aproximación la energía total que se necesita para separar a las dos cadenas que conforman la molécula de DNA. Aún cuando 16 está muy lejos de  $\infty$ , es un número lo suficientemente grande como para que los resultados con  $m = 16$  y con  $m = \infty$  no difieran apreciablemente.

En el Apéndice B mostramos cómo obtener la distribución  $P_\infty(d)$  de distancias entre mínimos consecutivos del potencial cuando  $m = \infty$ . El resultado es el siguiente:

$$\boxed{m = \infty}$$

$$P_\infty(d) = 3 \frac{2^d}{(d+3)!} (d^2 + d - 2) \quad (4.31)$$

En la Fig4.15 se muestran las gráficas de  $P_3(d)$ ,  $P_4(d)$  y  $P_\infty(d)$ . De esta figura puede observarse un hecho interesante: *el valor más probable en estas distribuciones ahora está en  $d = 2$* . Esto contrasta con el caso  $m = 2$ , donde se tenía que el valor más probable era  $d = 3$  (Fig4.13). Así, nuestros resultados parecen indicar que si el polímero está compuesto por más de dos tipos diferentes de monómeros, la distancia *más probable* entre mínimos consecutivos de energía es *dos*, aún cuando la distancia *promedio* esté cerca de *tres*. Lo anterior implicaría que la partícula, en su movimiento a lo largo del polímero, va brincando de dos en dos la mayor parte del tiempo, aunque en promedio, brinca de tres en tres.

No estamos seguros del significado que podría tener el hecho de que la distancia más probable caiga en dos, mientras que la distancia promedio esté en tres. Tal vez esto sea una consecuencia de la simplicidad de nuestro modelo. Después de todo, incluso el ribosoma primitivo no era una partícula puntual, sino que tenía un cierto tamaño. De hecho, existen entre 20 y 30 puntos de contacto simultáneamente entre el ribosoma actual y el RNA mensajero [9]. Lo anterior significa que el ribosoma mantiene una *interacción extendida* con el mRNA. Por lo tanto, un modelo más adecuado para describir el movimiento del ribosoma a lo largo del mRNA sería el mostrado en la Fig4.16, en el que se tienen varias partículas unidas rígidamente entre sí interactuando con el polímero. En este sistema de "multi-interacción", esperaríamos que el movimiento colectivo de las partículas (ribosoma) se diera de acuerdo con un *potencial promedio*, es decir, el

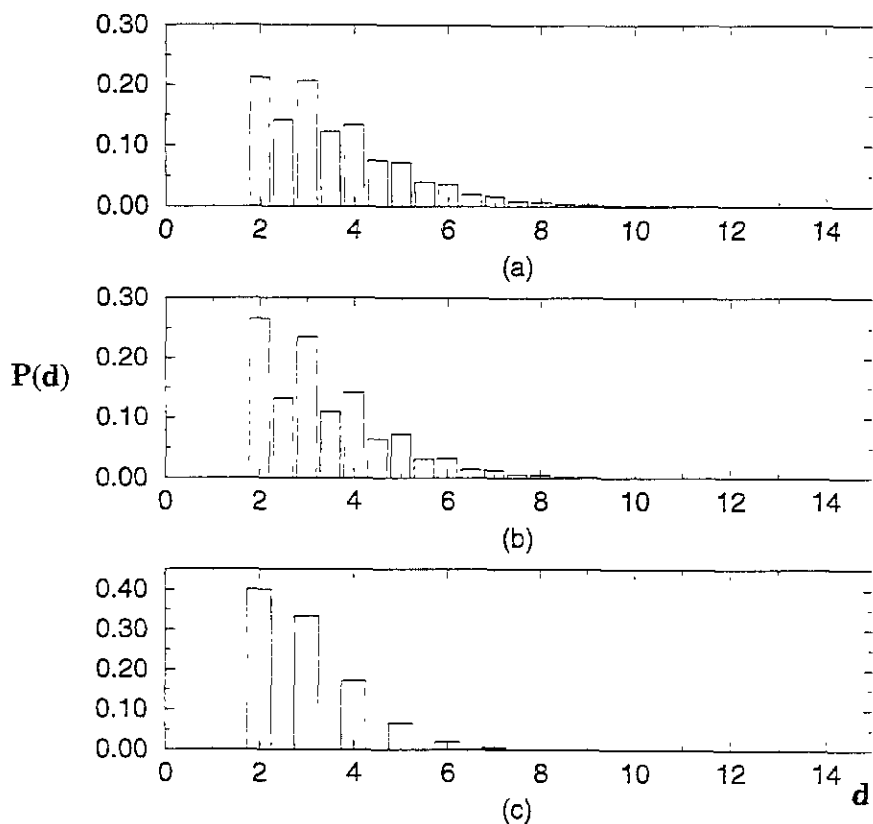


Figura 4.15: Distribuciones de distancia entre mínimos consecutivos de energía para los casos (a)  $m = 3$ , (b)  $m = 4$  y  $m = \infty$ . Nótese que en estas gráficas el valor más probable de la distribución se tiene en  $d = 2$ . Esto podría estar relacionado con la *hipótesis del tambaleo* formulada por Crick.

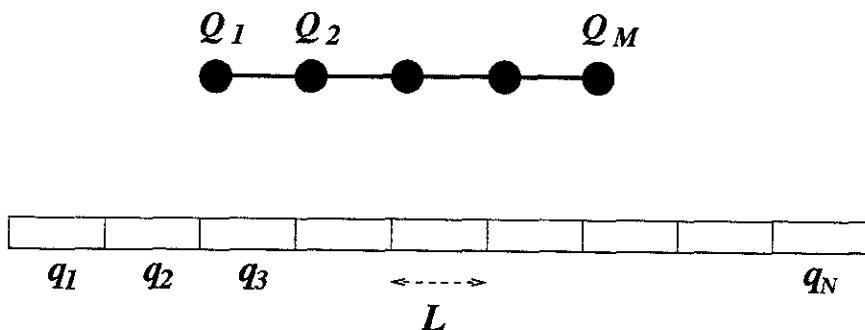


Figura 4.16: Un modelo más realista para describir la interacción del ribosoma primitivo con el mRNA. En este modelo, una cadena extendida interactúa con el polímero en varios puntos de contacto, produciéndose así una interacción colectiva entre varias partículas.

potencial que sienten todas las partículas, y no de acuerdo con el potencial local que siente cada una de ellas. De esta manera, cuando hay varios puntos de contacto, lo más razonable es suponer que el movimiento del sistema se va a dar en pasos cuya longitud sea la distancia promedio entre los mínimos consecutivos de la energía de interacción local (la que siente cada partícula), y no en pasos cuya longitud sea la distancia más probable. Si esto es cierto, el ribosoma se moverá en pasos de tres en tres la mayor parte del tiempo, de acuerdo con el promedio, y no en pasos de dos en dos.

Para corroborar nuestra suposición, simulamos el sistema mostrado en la Fig4.16, con un polímero compuesto por 15000 monómeros, interactuando con una cadena extendida formada por 20 partículas rígidamente unidas y separadas una de otra por una distancia 1 (en unidades de la longitud de los monómeros). Las cargas tanto en los monómeros como en las 20 partículas eran aleatorias y podían tomar los valores  $\pm 1, \pm 2, \pm 3$  con la misma probabilidad. El potencial de interacción entre las partículas y los monómeros fue coulombiano, es decir con  $\alpha = 1$ . En la Fig4.17a se muestra la gráfica del potencial, y en la Fig4.17b la distribución de distancias entre mínimos consecutivos. Como puede observarse, para este sistema la distancia más probable en la distribución de mínimos se ha desplazado considerablemente hacia el tres, lo cual sugiere que, efectivamente, la interacción de la cadena con el polímero se lleva a cabo de acuerdo con el promedio de las interacciones locales. Debemos mencionar que, para el caso mostrado en la Fig4.17, la distancia promedio entre mínimos de energía consecutivos resultó ser  $\bar{d} = 4$ .

Entrando un poco en la especulación, los resultados anteriores podrían

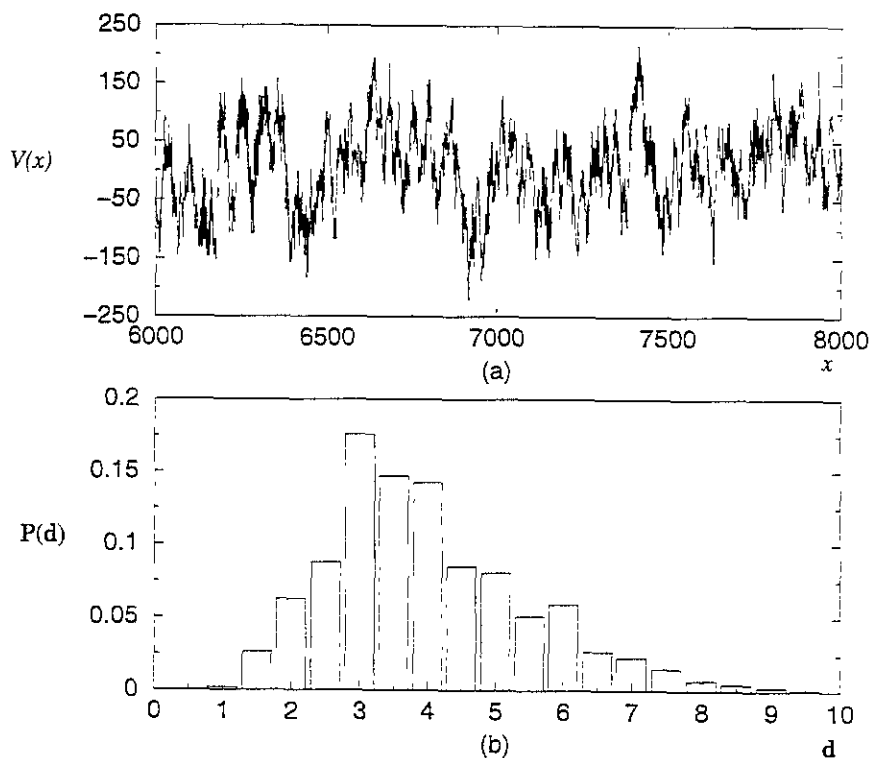


Figura 4.17: (a) Potencial de interacción entre 20 partículas rígidamente unidas interactuando coulombianamente con un polímero compuesto por 1500 monómeros. El cálculo se hizo para el caso  $m = 6$ . (b) Distribución de distancias entre mínimos de energía consecutivos en el potencial anterior. Nótese que la distancia más probable está ahora en  $d = 3$ . La distancia promedio resultó ser  $\bar{d} = 4$ .

ser la explicación de la hipótesis del tambaleo formulada por Crick (ver Capítulo 2). Podemos considerar que el RNA de transferencia mantiene una interacción *local* codón-anticodón con el RNA mensajero, por lo que, de acuerdo con los resultados de la Fig4.15, trataría de moverse la mayor parte del tiempo en pasos de dos monómeros. Por otro lado, el ribosoma mantiene una interacción extendida con el mRNA, y consecuentemente, se mueve preferentemente en pasos de tres monómeros. Este esquema sugiere, entonces, que la especificidad codón-anticodón, y por lo tanto, la especificidad codón-aminoácido, está determinada la mayoría de las veces por sólo dos bases, como efectivamente ocurre en el código genético, mientras que la estructura de tripletes del código queda determinada por la interacción ribosoma-mRNA.

## 4.5 Distribución de distancias para secuencias reales

Hasta el momento hemos trabajado con polímeros cuya composición de carga es aleatoria. Sin embargo, las secuencias genéticas reales distan mucho de tener una composición aleatoria de bases. Por tal motivo, en esta sección asignaremos las cargas de los monómeros a lo largo del polímero, no de manera aleatoria, sino en correspondencia con la secuencia de bases del genoma de algunos organismos, y analizaremos las propiedades del potencial así generado.

Como las secuencias aleatorias están compuestas por cuatro tipos diferentes de bases, entonces en nuestro modelo habrá cuatro valores posibles de carga para cada monómero. Establezcamos la siguiente correspondencia entre bases y valores de carga:

Base	Carga	
A	-2	
U	-1	(4.32)
C	+1	
G	+2	

Con esta asignación, si estamos considerando una cadena genética cuya secuencia de bases es, por ejemplo

A U G A U C G C A G

entonces la correspondiente secuencia de cargas a lo largo del polímero será

-2 -1 +2 -2 -1 +1 +2 +1 -2 +2

De esta manera, construimos polímeros cuya secuencia de cargas  $\{q_n\}$  está en correspondencia con la secuencia de bases de alguna cadena genética.

En las Figuras 4.18 y 4.19 mostramos la distribución de distancias entre mínimos consecutivos de energía, utilizando secuencias genéticas de diversos organismos. Los potenciales de interacción utilizados fueron coulombianos (ecuación (4.10)), con  $\sigma = 0.5$ , y las secuencias genéticas se tomaron del Gene Bank de 1999 tratando de escoger organismos representativos de los diferentes tipos de seres vivos: virus, bacterias, hongos, plantas y animales. Como puede verse, en todos los casos la distancia promedio entre mínimos de energía calculada numéricamente utilizando las distribuciones mostradas está alrededor de 3.4, como cabría esperar de acuerdo con la expresión (4.18) para el caso  $m = 4$ . Pero además, observamos una marcada tendencia en las distribuciones de secuencias reales a disminuir la probabilidad en  $d = 2$  y aumentarla en  $d = 3$ , de tal manera que ahora la distancia más probable es  $d = 3$  y no  $d = 2$  (como ocurre en el caso de secuencias aleatorias).

Cabe mencionar que en las secuencias genéticas de organismos reales, las bases  $A$ ,  $T$ ,  $C$  y  $G$  no se encuentran en la misma proporción. Por ejemplo, la siguiente tabla muestra la proporción de bases en las secuencias genéticas que se utilizaron para calcular los histogramas de las figuras 4.18 y 4.19:

Organismo	Proporción de bases				Longitud
	A	T	C	G	
<i>Drosophila</i>	0.314454	0.218096	0.237838	0.229612	4255
<i>N. crassa</i>	0.363546	0.361135	0.138156	0.137163	7050
Influenza	0.301982	0.317721	0.177017	0.203281	10547
<i>M. tuberculosis</i>	0.172274	0.161519	0.339270	0.326936	20270
Tabaco	0.306896	0.314642	0.192460	0.186002	15939
Humano	0.239025	0.232984	0.265203	0.262788	4966

Como puede verse de la tabla anterior, el DNA de algunos organismos, como el virus de la influenza y el hongo *Neurospora crassa*, tiene una proporción de bases marcadamente diferente del azar (en donde todas las bases estarían en una proporción de  $1/4$ ), mientras que en otros organismos (humano) la composición de bases se asemeja al azar. Con el objetivo de comprobar que los resultados mostrados en las figuras 4.18 y 4.19 no dependen de la proporción de bases en las secuencias genéticas, sino de la forma en que dichas bases están secuenciadas, generamos secuencias aleatorias de la misma longitud y con la misma proporción de bases que las mostradas en la tabla anterior, y calculamos numéricamente la distribución de las distancias entre mínimos consecutivos en los potenciales correspondientes a estas



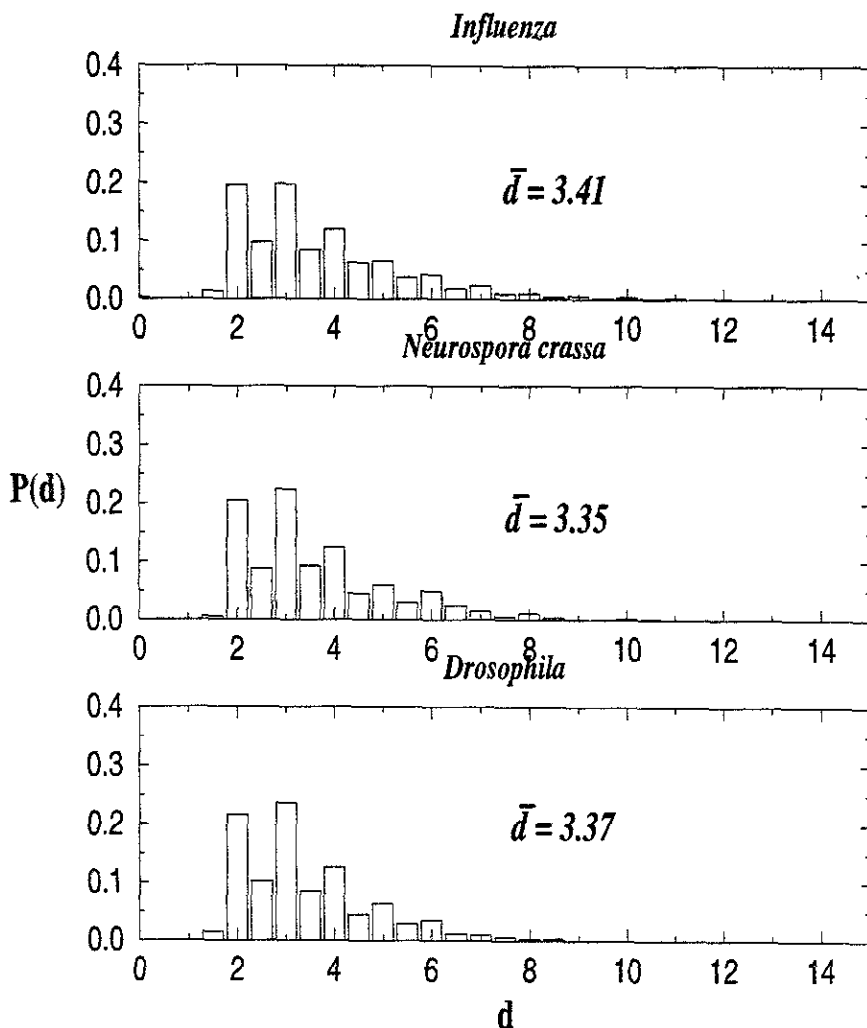


Figura 4.18: Distribución de distancias entre mínimos consecutivos de energía cuando se pone en correspondencia el valor de las cargas a lo largo del polímero con la secuencia de bases del genoma de algún organismo real. En este caso se utilizaron secuencias del virus que causa la influenza, del hongo *Neurospora crassa* y un pedazo del cromosoma 12 de la mosca *Drosophila*. Comparando con la Fig.4.15, puede observarse una clara tendencia en el genoma de los organismos a disminuir la probabilidad en  $d = 2$  y aumentarla en  $d = 3$ .

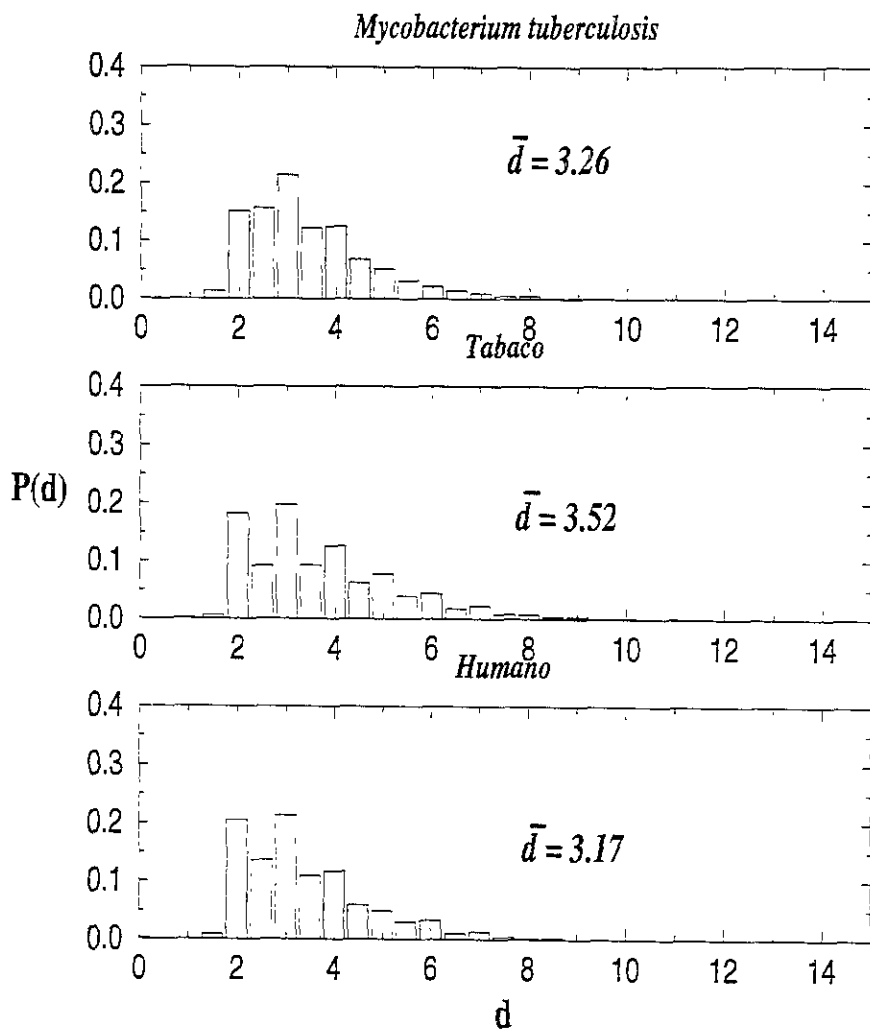


Figura 4.19: Distribución de distancias entre mínimos de energía consecutivos para secuencias reales. En estos casos utilizamos secuencias genéticas del bacilo de la tuberculosis (*Mycobacterium tuberculosis*), de la planta de tabaco y del cromosoma 20 del ser humano. Se observa otra vez un aumento considerable en la probabilidad para  $d = 3$ .

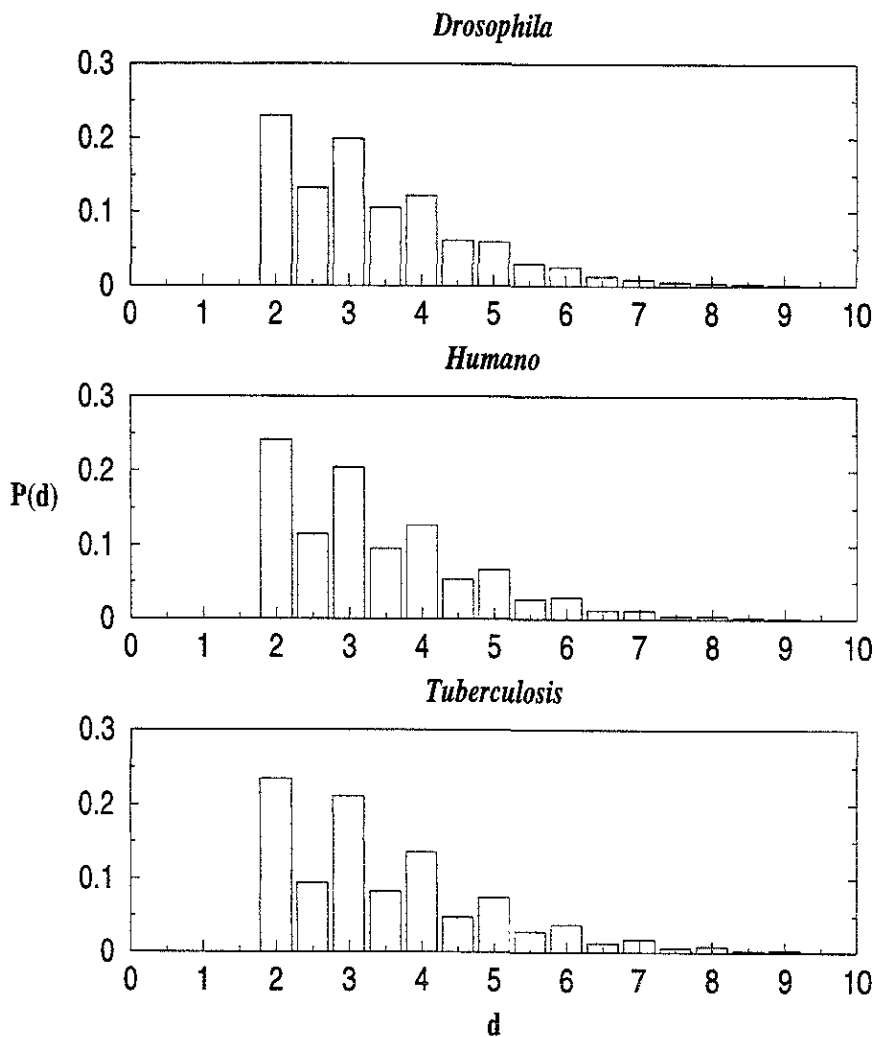


Figura 4.20: Distribución de distancias entre mínimos consecutivos de energía en secuencias aleatorias con la misma composición y longitud que las utilizadas para calcular los histogramas de las figuras 4.18 y 4.19. Obsérvese que al “aleatorizar” las secuencias de los organismos reales, se presenta otra vez el comportamiento azaroso mostrado en la Fig4.15.

secuencias aleatorias. Los resultados se muestran en la Fig4.20. Al comparar la figura anterior con la Fig4.15, vemos que la proporción de bases en las secuencias genéticas de los organismos reales no es la causa de que aumente la probabilidad en  $d = 3$  y disminuya en  $d = 2$ .

Las consideraciones anteriores sugieren que, a lo largo de la evolución, las secuencias genéticas de los organismos vivos se construyeron de tal manera que se optimizó el hecho de que la distancia promedio entre mínimos consecutivos de energía estaba, ya de entrada, alrededor de tres. En otras palabras, el que la distancia promedio fuese tres (o un número muy cercano a tres), no solamente influyó en la evolución de la maquinaria celular de síntesis protéica, sino que también influyó en la estructura de las secuencias codificadoras, optimizando esta distancia promedio de tal suerte que también la distancia más probable entre mínimos consecutivos de energía resultó ser tres.

Independientemente de si asociamos o no un proceso de "optimización del tres" a la evolución, es claro que en las secuencias genéticas de los organismos vivos, se selecciona mucho mejor el número tres, que en las secuencias aleatorias. Es necesario señalar que este resultado no depende fuertemente de la asignación particular de cargas dada en (4.32). Hemos trabajado con otras asignaciones de carga diferentes; por ejemplo, el valor de carga asignado a cada una de las bases  $A$ ,  $U$ ,  $C$  y  $G$  puede ser un número elegido aleatoriamente entre  $-1$  y  $1$ , y las distribuciones de distancias así obtenidas siguen siendo las mismas que las mostradas en las Figuras 4.18 y 4.19. Sin embargo, para que los resultados se mantengan, es necesario pedir que las cargas asignadas a las bases no estén extraordinariamente desproporcionadas. Por ejemplo, la siguiente asignación de cargas

Base	Carga	
$A$	$-2$	
$U$	$-1$	(4.33)
$C$	$+100$	
$G$	$+20000000$	

evidentemente conduciría a resultados muy diferentes, ya que hay una carga que "hiperdomina" a todas las demás. Afortunadamente, casos como estos no se presentan en la naturaleza. Por ejemplo, los momentos dipolares medidos experimentalmente para las cuatro bases del RNA, aunque diferentes, son todos del mismo orden de magnitud [52].

Para concluir este capítulo sobre potenciales, es necesario hacer una observación más. Las Figuras 4.18 y 4.19, que muestran la distribución de distancias entre mínimos consecutivos de energía para secuencias de diversos organismos, indican que, tanto la distancia promedio como la distancia más probable, andan cerca de tres. Sin embargo, estas distribuciones

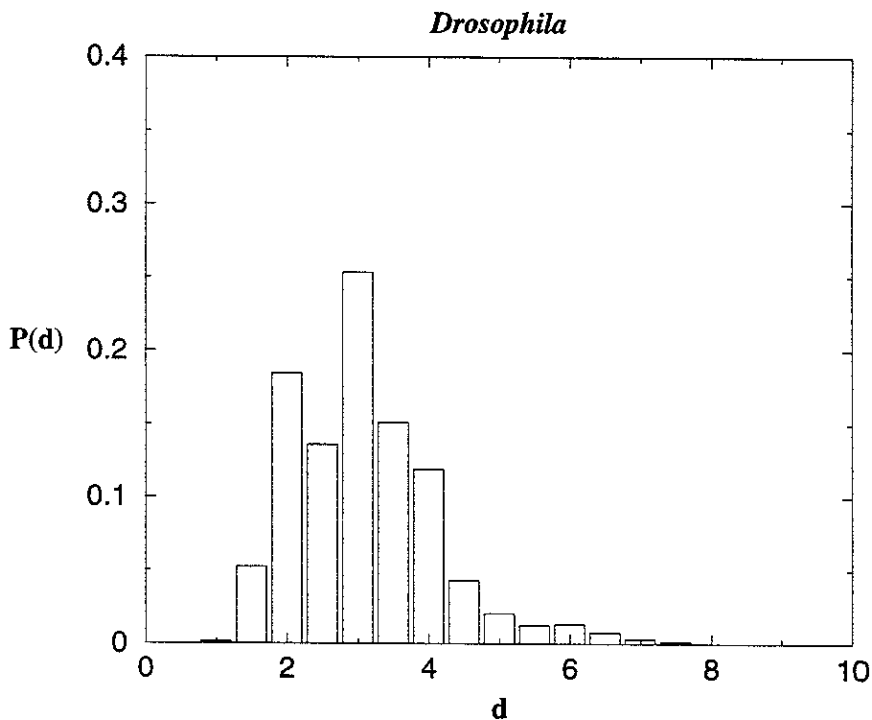


Figura 4.21: Distribución de distancias entre mínimos consecutivos en el potencial de interacción de una cadena extendida formada por 20 partículas, interactuando con un polímero cuya secuencia de cargas estaba en correspondencia con la secuencia de bases de un pedazo de cromosoma de la mosca *Drosophila*. Obsérvese como ha aumentado considerablemente la probabilidad en  $d = 3$  respecto de la distribución que se muestra en la Fig. 4.18, lo cual indica otra vez que la cadena representa mejor al ribosoma que una partícula puntual.

también indican que existe una probabilidad, considerablemente alta, de obtener mínimos de energía separados por otras distancias, es decir, las distribuciones tienen "colas". Desde el punto de vista de nuestro modelo, esto significaría que el ribosoma podría moverse en pasos de cuatro o de cinco bases a lo largo del mRNA, lo cual, desde luego, no se observa. El ribosoma *siempre* se mueve en pasos de tres bases a lo largo del RNA mensajero. Esto lo único que refleja es la simplicidad de nuestro modelo. En la Fig4.21 se muestra la distribución de distancias para la secuencia de la mosca *Drosophila* cuando dejamos que el polímero interactúe con 20 partículas cargadas aleatoriamente y rígidamente unidas. Vemos que con este tipo de interacción extendida, se acentúa aún más la probabilidad en tres, lo cual indica que la cadena extendida representa mejor al ribosoma que la partícula puntual. Sin embargo, el ribosoma actual es una molécula muy compleja, y en la interacción con el mRNA intervienen muchos otros factores que no hemos ni siquiera considerado (reacciones enzimáticas, complejos de iniciación, elongación y término, hidrólisis de GTP, etc). Aún así, con el trabajo que hemos realizado hasta ahora podemos ver que, incluso con un modelo simple de ribosoma (una partícula puntual), se obtienen elementos indicando que el ribosoma primitivo debió haberse movido, preferentemente, en pasos de tres bases en tres bases.

## Capítulo 5

# Dinámica

Hemos visto que el potencial de interacción entre la partícula y el polímero presenta un cierto tipo de regularidad estadística, en el sentido de que los mínimos consecutivos del potencial están separados por una distancia promedio de aproximadamente tres monómeros a lo largo del polímero. También vimos que esta regularidad se acentúa si ponemos la distribución de cargas a lo largo del polímero en correspondencia con las secuencias genéticas de organismos reales, ya que en tal caso no sólo la distancia promedio, sino también la distancia más probable entre mínimos consecutivos del potencial, se encuentran alrededor de tres. Argumentábamos en el capítulo anterior que esta regularidad espacial en los potenciales podría traducirse en una regularidad dinámica en el movimiento de la partícula, lo que en nuestro caso significa que la partícula se movería dando brincos cuya longitud es, aproximadamente, de tres monómeros.

Por lo tanto, lo que nos proponemos en este capítulo es estudiar el movimiento de la partícula a través del polímero, para ver si efectivamente dicho movimiento presenta una regularidad estadística “alrededor del tres”. Problemas de este estilo, en los que se estudia el movimiento de una molécula a lo largo del potencial generado por un polímero, ya se han estudiado, hasta cierto grado, en otros contextos, de los cuales el más relevante para nosotros es el de Máquinas Moleculares en biología [55], [56], [57]. En los últimos 20 años se ha mostrado que una gran cantidad del transporte intracelular (transporte de nutrientes, por ejemplo), se lleva a cabo por proteínas “motoras” que se mueven a lo largo de filamentos tubulares de la misma forma en que lo hace un tren a lo largo de las vías. Ejemplos de lo anterior son el sistema quinesina-tubulina, en el cual una molécula de quinesina se mueve a lo largo de un filamento de tubulina, o el sistema miosina-actina, donde la miosina se mueve a lo largo de filamentos

de actina. Por supuesto, también el sistema ribosoma-mRNA que nosotros estamos considerando cae dentro de la misma categoría de máquinas moleculares, pero con algunas variaciones importantes, como veremos.

En los sistemas quinesina-tubulina o miosina-actina, lo esencial es que los filamentos tubulares que hacen las veces de las vías del tren, son polímeros más o menos rígidos que producen potenciales periódicos dirigidos, es decir, que en cada periodo presentan una asimetría siempre en la misma dirección (como en la Fig4.2d). De esta manera, una proteína motora siempre se moverá hacia un mismo lado a lo largo del filamento. Otro punto importante en estos modelos es que el sistema puede estar en dos o más estados de potencial generados, por ejemplo, por la presencia o la ausencia de moléculas catalizadoras como el ATP o el GTP. Las fluctuaciones térmicas, aunadas al efecto que estas moléculas catalizadoras tienen sobre el sistema, hacen que las moléculas motoras se desplacen a lo largo de los potenciales, produciéndose así, bajo determinadas condiciones, una corriente neta. Por lo tanto, aún cuando los potenciales son periódicos, el transporte de las proteínas motoras es un fenómeno aleatorio, y lo importante es encontrar la corriente neta inducida por los cambios de estado del potencial y por el ruido térmico.

La periodicidad y asimetría de los potenciales estudiados en los sistemas miosina-actina o quinesina-tubulina se desprenden directamente de las propiedades estructurales de los filamentos tubulares sobre los cuales se lleva a cabo el transporte. Tales filamentos están compuestos, generalmente, por aminoácidos polares secuenciados periódicamente. Sin embargo, en el sistema ribosoma-mRNA que nosotros estamos estudiando, la "vía del tren" (es decir, el mRNA) presenta una composición de bases más aleatoria que periódica. Debido a esto, los potenciales de interacción son aleatorios y no periódicos como en los sistemas de proteínas motoras, y nuestro interés no está en ver si se produce o no una corriente neta, sino suponer que dicha corriente existe y descubrir si hay periodicidades o regularidades de algún tipo en la dinámica del sistema. Comenzaremos en la siguiente sección el estudio de la dinámica del modelo presentado en el capítulo anterior, estableciendo la ecuación que gobierna el movimiento de la partícula a lo largo del polímero.

## 5.1 La ecuación de movimiento

El análisis de la dinámica del sistema estará basado en la ecuación de movimiento de Newton  $F = ma$ , para lo cual deberemos establecer la fuerza neta  $F$  que actúa sobre la molécula. Supondremos que sobre la partícula actúa una fuerza externa  $F_{ex}(t)$  que puede depender del tiempo, y cuyo ob-



jetivo es forzar al sistema haciendo que la partícula "explore" el potencial conforme se mueve sobre el polímero. Supondremos además que entre la partícula y el polímero existe una fuerza de fricción  $F_f = -\gamma v$ , siendo  $\gamma$  el coeficiente de fricción y  $v = dx/dt$  la velocidad relativa entre la partícula y el polímero. Por supuesto, la fuerza  $F_p(x) = -\partial V(x)/\partial x$  proveniente del potencial aleatorio  $V(x)$  también está presente. En el caso más general, existirá una fuerza aleatoria  $\eta(t)$  dependiente del tiempo, producida por las fluctuaciones térmicas del entorno en el que se encuentre inmerso el sistema. La ecuación que determina la dinámica del sistema es entonces

$$m \frac{d^2 x}{dt^2} + \gamma \frac{dx}{dt} = -\frac{\partial V(x)}{\partial x} + F_{ex}(t) + \eta(t).$$

La ecuación anterior es la que normalmente se utiliza para describir la dinámica de las máquinas moleculares [55], [56], [59]. Sin embargo, para nuestros propósitos tal ecuación es demasiado general. Por un lado, recordemos que en el escenario que estamos proponiendo, la partícula y el polímero se encuentran en el espacio exterior, ya sea en el polvo interestelar o en superficies de meteoritos o cometas. En estas condiciones, la temperatura es muy baja ( $T \sim 10$  °K) como para que las fluctuaciones térmicas tengan un efecto apreciable en la dinámica del sistema. Por lo tanto, en nuestro análisis haremos  $\eta(t) = 0$ .

Por otro lado, en los sistemas moleculares como las proteínas motoras a las que nos referimos al inicio de este capítulo, la fuerza de fricción es excesivamente grande, y podríamos suponer que lo mismo ocurre con el ribosoma y el RNA mensajero. Esto es cierto en general para los sistemas celulares en los que una molécula se desplaza a lo largo de un filamento tubular [60]. En tales circunstancias la dinámica del sistema se lleva a cabo en un *régimen sobreamortiguado* y el término de inercia  $m d^2x/dt^2$  no aparece en la ecuación de movimiento. La razón de lo anterior es que en los sistemas sobreamortiguados, los efectos de inercia prácticamente no existen. Consideremos, por ejemplo, una canica moviéndose en un medio extremadamente viscoso, como la miel. En tanto apliquemos una fuerza a la canica, ésta se moverá. Si aplicamos una fuerza dos veces más grande, la canica se moverá dos veces más rápido. Pero tan pronto como dejemos de aplicar la fuerza, la canica llegará al reposo casi instantáneamente. Así, en los sistemas sobreamortiguados, la velocidad (mas no la aceleración) es proporcional a la fuerza, y la ecuación de movimiento se convierte entonces en  $F = \gamma v$ , siendo  $\gamma$  la constante de proporcionalidad (coeficiente de fricción).

Dadas las consideraciones anteriores, la ecuación de movimiento que estudiaremos es

$$\gamma \frac{dx}{dt} = -\frac{\partial V(x)}{\partial x} + F_{ex}(t). \quad (5.1)$$

La fuerza externa  $F_{ex}(t)$  es necesaria para forzar al sistema, obligando a la partícula a moverse en lugar de que se quede “atrapada” en alguno de los mínimos del potencial. Dicha fuerza podría estar producida por la radiación electromagnética que permea el espacio interestelar, o podría representar la acción de una “bomba química” como el ATP que periódicamente entra al sistema provocando el movimiento, tal y como ocurre en las máquinas moleculares actuales. Dado que el escenario extraterrestre del origen de la vida es todavía muy incierto en cuanto a los procesos prebióticos y proto-bióticos, no trataremos de establecer la naturaleza particular de la fuerza  $F_{ex}(t)$ . No obstante, independientemente de su naturaleza, es evidente que dicha fuerza debió estar presente en las máquinas moleculares primitivas, ya que las máquinas moleculares que ahora se mueven, en algún momento debieron comenzar a moverse. En nuestro modelo, supondremos que dicha fuerza externa existe y que es la que está forzando al sistema.

## 5.2 Ordenes de magnitud

En esta sección, daremos una forma explícita del potencial aleatorio  $V(x)$ . Al hacer esto también podremos calcular el orden de magnitud de las cantidades físicas involucradas, como son las energías de interacción y la magnitud de la fuerza externa, con lo cual estableceremos la factibilidad de nuestro modelo.

Aún cuando podríamos trabajar con los potenciales que presentamos en el capítulo anterior (ver ecuaciones (4.10), (4.11) y (4.12)), nuestro objetivo ahora es trabajar con un potencial más “realista” que describa el tipo de interacciones que se dan en las moléculas biológicas. Este tipo de interacciones consiste fundamentalmente en interacciones dipolares y dipolares inducidas. Por lo tanto, consideremos el caso en el que los monómeros del polímero son barras unidimensionales de longitud  $L$  cargadas uniformemente, cada uno con una densidad lineal de carga aleatoria  $\lambda_i$ , mientras que la partícula de prueba  $M$  es una molécula polarizable sobre la cual se puede inducir un momento dipolar. Supongamos que la partícula  $M$  se encuentra en una posición horizontal  $x$  y a una distancia  $\sigma$  perpendicular al polímero, como se muestra en la Fig5.1. La energía de interacción está dada por

$$V(x) = -\frac{1}{2} \vec{p}(x) \cdot \vec{E}(x)$$

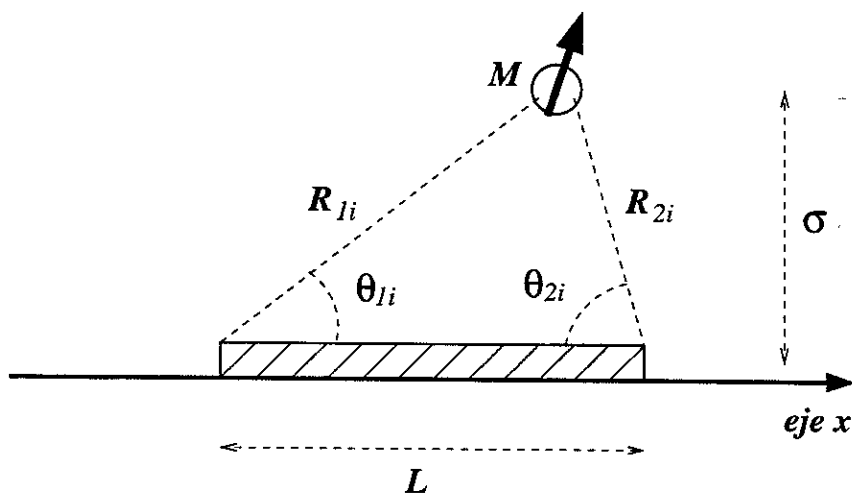


Figura 5.1: Esquema que ilustra los parámetros involucrados en la ecuación (5.4). En este caso, cada uno de los monómeros está cargado uniformemente con una densidad lineal de carga  $\lambda_i$ , mientras que la partícula de prueba  $Q$  es una partícula polarizable sobre la cual se induce un momento dipolar (representado con una flecha sobre la partícula).

donde  $\vec{E}(x)$  es el campo eléctrico generado por todo el polímero justo en la posición  $(x, \sigma)$  en la que se encuentra la partícula, y  $\vec{p}(x)$  es el momento dipolar inducido por dicho campo<sup>1</sup> sobre  $M$ . En primera aproximación, podemos suponer que  $\vec{p}(x) = \alpha \vec{E}(x)$ , donde  $\alpha$  es la constante de polarizabilidad eléctrica de la partícula. Por lo tanto, la energía de interacción  $V(x)$  queda como

$$V(x) = -\frac{\alpha}{2} \vec{E}(x) \cdot \vec{E}(x) = -\frac{\alpha}{2} [\vec{E}(x)]^2 \quad (5.2)$$

El campo eléctrico  $\vec{E}(x)$  producido por todo el polímero es la suma de los campos individuales  $\{\vec{E}_i(x)\}$  generados por cada uno de los monómeros:

$$\vec{E}(x) = \sum_{i=1}^N \vec{E}_i(x) \quad (5.3)$$

<sup>1</sup>El campo eléctrico  $\vec{E}$  y el momento dipolar  $\vec{p}$  son funciones tanto de  $x$  como de  $\sigma$ . Sin embargo, debido a que consideramos  $\sigma$  constante, sólo escribimos la dependencia en la posición horizontal  $x$ .

Para el sistema que estamos tratando, el campo  $\vec{E}_i(x)$  generado en el punto  $(x, \sigma)$  por el  $i$ -ésimo monómero del polímero está dado por

$$\vec{E}_i(x) = K\lambda_i \left[ \left( \frac{1}{R_{2i}} - \frac{1}{R_{1i}} \right) \hat{x} + \frac{1}{\sigma} (\cos \theta_{1i} + \cos \theta_{2i}) \hat{y} \right] \quad (5.4)$$

donde  $K$  es una constante que depende del sistema de unidades (la constante de Coulomb),  $\lambda_i$  es la densidad lineal de carga del monómero. mientras que los parámetros  $R_{1i}$ ,  $R_{2i}$ ,  $\theta_{1i}$  y  $\theta_{2i}$  son los que se muestran en la Fig.5.1. Evidentemente, estos parámetros son funciones de  $x$ , aunque no lo escribimos aquí explícitamente para no complicar más la notación. Sustituyendo las expresiones (5.3) y (5.4) en la ecuación (5.2) obtenemos

$$V(x) = -\frac{K^2\alpha}{2} \left\{ \sum_{i=1}^N \lambda_i \left[ \left( \frac{1}{R_{2i}} - \frac{1}{R_{1i}} \right) \hat{x} + \frac{1}{\sigma} (\cos \theta_{1i} + \cos \theta_{2i}) \hat{y} \right] \right\}^2 \quad (5.5)$$

Para seguir adelante, supondremos que los monómeros pueden estar cargados con una carga electrónica, o con dos cargas electrónicas, o con tres, o en general. con un múltiplo entero de cargas electrónicas. Por lo tanto, la densidad de carga  $\lambda_i$  de cada monómero la escribiremos como

$$\lambda_i = \lambda_0 \beta_i$$

donde  $\beta_i$  es un número entero aleatorio adimensional que puede tomar cualquiera de los  $m$  valores  $\{1, 2, 3, \dots, m\}$  con la misma probabilidad. mientras que  $\lambda_0$  es una medida de la densidad de carga efectiva de los monómeros cuando éstos están ionizados con una sólo carga electrónica. La energía potencial de interacción  $V(x)$  entre la partícula y el polímero queda entonces como

$$V(x) = -\frac{K^2\lambda_0^2\alpha}{2} \left\{ \sum_{i=1}^N \beta_i \left[ \left( \frac{1}{R_{2i}} - \frac{1}{R_{1i}} \right) \hat{x} + \frac{1}{\sigma} (\cos \theta_{1i} + \cos \theta_{2i}) \hat{y} \right] \right\}^2 \quad (5.6)$$

Con el objetivo de trabajar con una ecuación de movimiento adimensional. que fácilmente pueda meterse a la computadora sin tener que preocuparnos de los valores reales de los parámetros involucrados (los cuales son generalmente muy pequeños), expresaremos el potencial en forma adimensional. En primer lugar tomaremos a  $L$  como nuestra unidad de longitud, lo que es equivalente a dividir todas las distancias que aparecen en la ecuación anterior entre  $L$ . Por lo tanto, si hacemos

$$\begin{aligned} y &= x/L, \\ \rho_{1i} &= R_{1i}/L, \\ \rho_{2i} &= R_{2i}/L, \\ h &= \sigma/L, \end{aligned} \tag{5.7}$$

entonces la expresión (5.6) se reduce a

$$V(x) = -\frac{K^2 \lambda_0^2 \alpha}{2L^2} \left\{ \sum_{i=1}^N \beta_i \left[ \left( \frac{1}{\rho_{2i}} - \frac{1}{\rho_{1i}} \right) \hat{x} + \frac{1}{h} (\cos \theta_{1i} + \cos \theta_{2i}) \hat{y} \right] \right\}^2 \tag{5.8}$$

Lo importante de la expresión anterior es que lo que quedó elevado al cuadrado entre los paréntesis angulares, es una función que no tiene dimensiones físicas. A esta función la denotaremos por  $\mathcal{V}(y)$  y nos referiremos a ella como el *potencial adimensional*:

$$\mathcal{V}(y) = \left\{ \sum_{i=1}^N \beta_i \left[ \left( \frac{1}{\rho_{2i}} - \frac{1}{\rho_{1i}} \right) \hat{x} + \frac{1}{h} (\cos \theta_{1i} + \cos \theta_{2i}) \hat{y} \right] \right\}^2 \tag{5.9}$$

de tal forma que el potencial de interacción  $V(x)$  puede escribirse como

$$V(x) = -V_0 \mathcal{V}(y) \tag{5.10}$$

donde  $V_0$  es una constante que determina la amplitud del potencial, y está dada por

$$V_0 = \frac{K^2 \lambda_0^2 \alpha}{2L^2} \tag{5.11}$$

Podemos ahora adimensionalizar la ecuación de movimiento (5.1) si definimos los parámetros adimensionales  $\mathcal{F}$  y  $\tau$  como

$$\mathcal{F} = \frac{2L^3 q}{K^2 \lambda_0^2 \alpha} F_{ex}, \tag{5.12}$$

$$\tau = \frac{K^2 \lambda_0^2 \alpha}{2\gamma L^4} t, \tag{5.13}$$

con lo cual, la ecuación de movimiento se transforma en

$$\frac{dy}{d\tau} = -\frac{\partial\mathcal{V}(y)}{\partial y} + \mathcal{F} \quad (5.14)$$

No está por demás enfatizar que  $y$  es la variable de posición horizontal medida en unidades de  $L$  ( $y = x/L$ ), mientras que  $\tau$  es un parámetro adimensional que sustituye al tiempo.  $\mathcal{F}$  es la amplitud de la fuerza externa, pero ya sin dimensiones. Los valores reales de la fuerza  $F_{ex}$  y del tiempo  $t$  pueden calcularse a partir de las ecuaciones (5.12) y (5.13) como funciones de  $\mathcal{F}$  y  $\tau$  y de los demás parámetros que aparecen en dichas ecuaciones ( $L$ ,  $\alpha$ ,  $\gamma$  y  $\lambda_0$ ). La Tabla I muestra los valores típicos de estos parámetros, medidos en sistemas similares al que estamos estudiando (los sistemas actina-miosina o dineína-tubulina, por ejemplo) [50], [57].

Tabla I	
$L \sim$	1nm
$\alpha \sim$	$10^{-40} \text{C}^2 \text{m}^2 \text{J}^{-1}$
$\gamma \sim$	$10^{-5} \text{kg/s}$
$\sigma \sim$	0.5nm
$\lambda_0 \sim$	$1.6 \times 10^{-10} \text{C/m}$

En la Fig5.2a mostramos la gráfica del potencial adimensional  $\mathcal{V}(y)$  generado por un polímero compuesto de 100 monómeros<sup>2</sup>. Por otro lado, en la Fig5.2b se muestra la derivada espacial de este potencial. Para obtener estas gráficas las variables aleatorias  $\beta_i$  podían adquirir cualquiera de los cuatro valores  $\{1, 2, 3, 4\}$  con la misma probabilidad (caso  $m = 4$ ), e hicimos  $h = 0.5$ . Como puede observarse en la figura, tanto la mayor diferencia de potencial  $\Delta\mathcal{V}_{\max}$  como el valor más grande de la derivada valen aproximadamente 80:

$$\Delta\mathcal{V}_{\max} \sim 80$$

$$\left| \frac{\partial\mathcal{V}}{\partial y} \right|_{\max} \sim 80$$

Por lo tanto, para que el polímero no se quede atrapado en un pozo de potencial, la amplitud de la fuerza externa que forza al sistema tendría que ser del orden  $\mathcal{F} \sim 80$ . A partir de estos números, podemos estimar el orden de magnitud de la diferencia máxima de potencial  $\Delta\mathcal{V}_{\max}$  y de la fuerza externa  $F_{ex}$ , utilizando las ecuaciones (5.8) y (5.12) y los valores de la Tabla I, con lo que se obtiene:

<sup>2</sup>Obsérvese que hay 31 mínimos en un intervalo de 100 monómeros, por lo que la distancia promedio entre mínimos es  $\bar{l} \simeq 3.22$

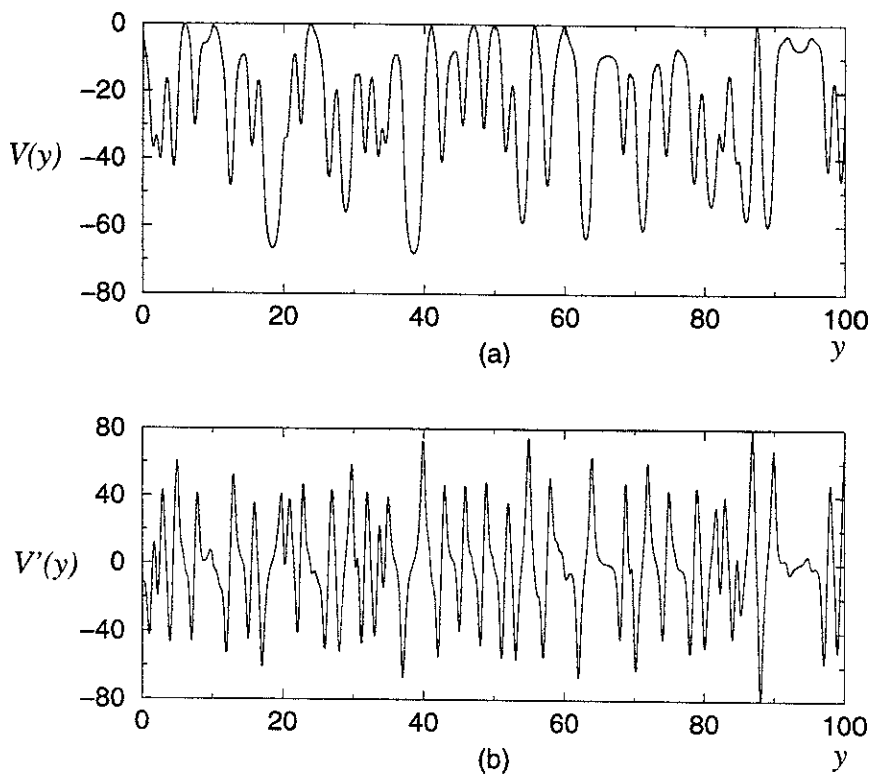


Figura 5.2: (a) Gráfica de una realización típica del potencial adimensional dado en la ecuación (5.9). Obsérvese que las diferencias de potencial más grandes son aproximadamente  $\Delta V_{\max} \sim 80$ . Además, nótese que hay 31 mínimos a lo largo de 100 monómeros, lo que corresponde a una distancia promedio  $\bar{d} = 3.22$  entre mínimos consecutivos. (b) Derivada espacial del potencial adimensional mostrado en la gráfica anterior. Los valores de la derivada espacial del potencial están comprendidos entre  $-80$  y  $+80$ .

$$\begin{aligned}\Delta V_{\max} &\sim 8.3 \times 10^{-21} \text{J} \sim 10^{-20} \text{J} \\ F_{ex} &\sim 8.3 \times 10^{-12} \text{N} \sim 10 \text{pN}\end{aligned}$$

Para tener parámetros de comparación, debemos hacer notar que la energía promedio del enlace de hidrógeno es  $3.3 \times 10^{-20} \text{ J}^{-1}$ . Por otro lado, en trabajos recientes se han medido experimentalmente las fuerzas necesarias para forzar sistemas moleculares similares al que estamos considerando [56],[58]. En particular, para los sistemas miosina-actina y quinesina-tubulina, se ha encontrado que el valor de tales fuerzas es de 5 pN. Como puede verse, los valores de  $\Delta V_{\max}$  y  $F_{ex}$  que hemos obtenido caen dentro del orden de magnitud de las correspondientes cantidades físicas que ocurren en las máquinas moleculares actuales. En este sentido, el modelo que estamos considerando es factible.

Cabe mencionar que en la actualidad, las fuerzas externas que forzan a estas máquinas las proveen “bombas químicas” cuyo funcionamiento está basado en la desulfuración de las moléculas de ATP y GTP. Probablemente en el origen de la vida existían bombas químicas similares a las actuales, o tal vez el movimiento de tales máquinas se comenzó a dar por otro tipo de agentes externos aún no identificados. En nuestro trabajo, como ya hemos dicho, daremos por hecho que tales fuerzas externas están presentes en el sistema, independientemente de su naturaleza.

Finalmente, de la ecuación 5.13 y de los valores de la tabla I, se desprende que una unidad del tiempo adimensional  $\tau$  equivale a un intervalo de tiempo real  $\Delta t$  dado por

$$\Delta t \sim 10 \text{seg} \quad (5.15)$$

### 5.3 Periodo tres en secuencias aleatorias

Ya que hemos visto que los parámetros involucrados en el modelo son físicamente razonables, resolveremos numéricamente la ecuación de movimiento (5.14) utilizando el potencial adimensional (5.9). Para esto supondremos que la fuerza externa  $\mathcal{F}(\tau)$  es constante:  $\mathcal{F} = 80$ .

En la Fig5.3a mostramos la gráfica de la velocidad de la partícula  $v(x)$  como función de la posición. Esta gráfica la obtuvimos utilizando un polímero compuesto por 100 monómeros<sup>3</sup>, cada uno con una carga aleatoria  $\beta$ , cuya función de distribución fue  $P(\beta) = \frac{1}{4} \sum_{l=1}^4 \delta(\beta - l)$  (cuatro valores

<sup>3</sup>Recordemos que L. Orwell mostró experimentalmente que polímeros de RNA compuestos por aproximadamente 90 bases ya mostraban actividad autocatalítica.



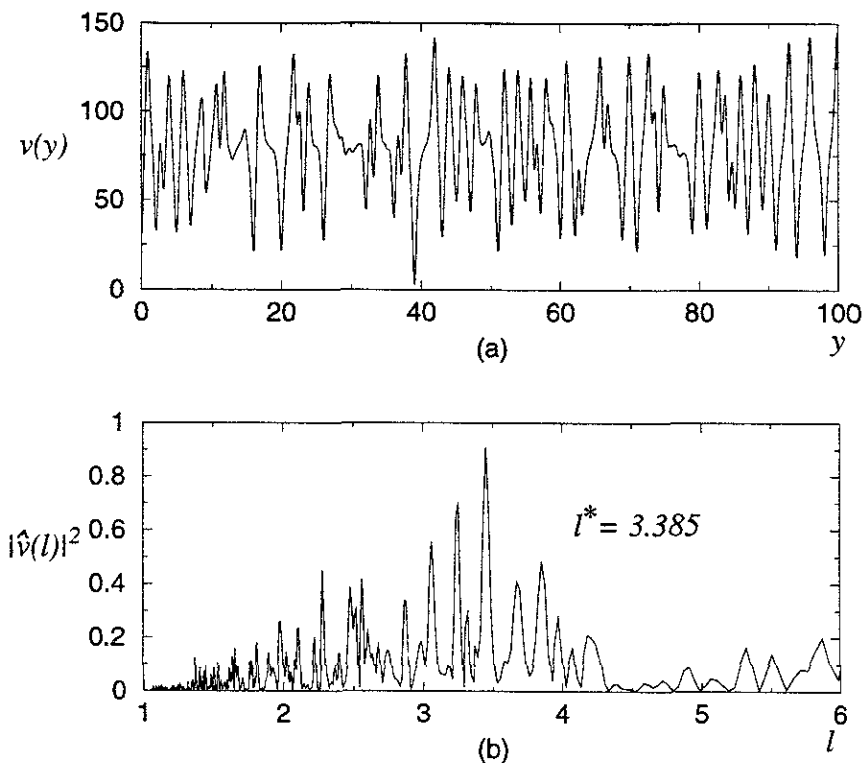


Figura 5.3: (a) Velocidad de la partícula como función de su posición  $y$  a lo largo del polímero. Se utilizaron 4 valores de carga para los monómeros (caso  $m = 4$ ). (b) Espectro de potencias de la velocidad anterior. Por claridad, el espectro de potencias se presenta como función de la variable  $l = 2\pi/k$ , siendo  $k$  la variable natural de la transformada de Fourier. Obsérvese que el pico más grande en el espectro corresponde a una periodicidad espacial  $l^* = 3.385$ .

de carga,  $m = 4$ ). Además, elegimos  $h$ , la distancia adimensional perpendicular entre la partícula y el polímero, como  $h = 0.5$ . Cabe mencionar que en la simulación numérica, la partícula recorrió 100 monómeros en un intervalo de tiempo  $\Delta\tau \sim 3$ , el cual, de acuerdo con la expresión (5.15), equivale a un intervalo de tiempo real  $\Delta t \sim 30$  seg. Lo anterior da una velocidad  $v \sim 3$  nm/seg para la partícula a lo largo del polímero<sup>4</sup>.

Con el objetivo de ver si hay regularidades en la dinámica del sistema, tomamos la transformada de Fourier de la velocidad de la partícula como función de la posición. La Fig5.3b muestra el espectro de potencias  $|\hat{v}(k)|^2$  de esta velocidad, obtenido numéricamente utilizando el método de Lomb [61]<sup>5</sup>. Como puede verse en la figura, el espectro de potencias muestra claramente un pico sobresaliente correspondiente a una regularidad espacial cuyo "periodo" es  $l^* = 3.385$ <sup>6</sup>. Ponemos "periodo" entre comillas porque realmente no tenemos una periodicidad en el sentido estricto de la palabra, es decir, en el sentido de que cada 3.385 monómeros el movimiento de la partícula se repite sistemáticamente (como puede observarse en la Fig5.3a). Lo que tenemos en cambio, es un movimiento en el que la velocidad presenta un comportamiento más o menos regular, y esta regularidad se repite aproximadamente cada 3.385 monómeros.

Lo anterior no es una propiedad particular del potencial de interacción (5.9) que estamos considerando. Por ejemplo, la Fig5.4 muestra resultados similares pero ahora utilizando el potencial coulombiano dado en la ecuación (4.10) del capítulo anterior. En este caso utilizamos la distribución de cargas  $P(\beta) = \frac{1}{4} \sum_{l=1}^4 \delta(\beta - l)$  para los monómeros (caso  $m = 4$ ) y  $h = 0.5$ .

<sup>4</sup>La velocidad a la que se mueve la quinesina a lo largo de los filamentos de tubulina es aproximadamente de 100 nm/seg a concentraciones saturadas de ATP, con un gasto de 700 ATP's por segundo [57], [58]. Por otro lado, los ribosoma actuales dentro de la célula añaden entre 20 y 30 aminoácidos por segundo a la cadena peptídica en formación, lo cual equivale a una velocidad aproximada de 50nm/seg a lo largo del mRNA mensajero

<sup>5</sup>La transformada de Fourier discreta de una serie de puntos  $\{x(t_1), x(t_2), \dots, x(t_N)\}$ , donde los valores  $t_1, t_2, \dots, t_N$  están igualmente espaciados, se define como

$$\hat{x}(k) = \frac{1}{\sqrt{N}} \sum_{j=1}^N x(t_j) \exp(-ij k)$$

donde  $i = \sqrt{-1}$  y  $k = \frac{2\pi}{N}, 2\frac{2\pi}{N}, 3\frac{2\pi}{N}, \dots, N\frac{2\pi}{N}$ . El espectro de potencias es la gráfica de  $|\hat{x}(k)|^2$  contra  $k$ , y muestra las regularidades que hay en la serie original de puntos  $\{x(t_1), x(t_2), \dots, x(t_N)\}$ , si es que dichas regularidades existen. El método de Lomb es una generalización de la transformada anterior en el caso en que los puntos  $t_1, t_2, \dots, t_N$  no están igualmente espaciados.

<sup>6</sup>Para mayor claridad, hemos expresado las gráficas de los espectros de potencias como función de la variable  $l = 2\pi/k$ , siendo  $k$  la variable natural de la transformada de Fourier. De esta manera, el valor  $l^*$  en el cual aparece el pico más grande en el espectro, da directamente la periodicidad espacial.

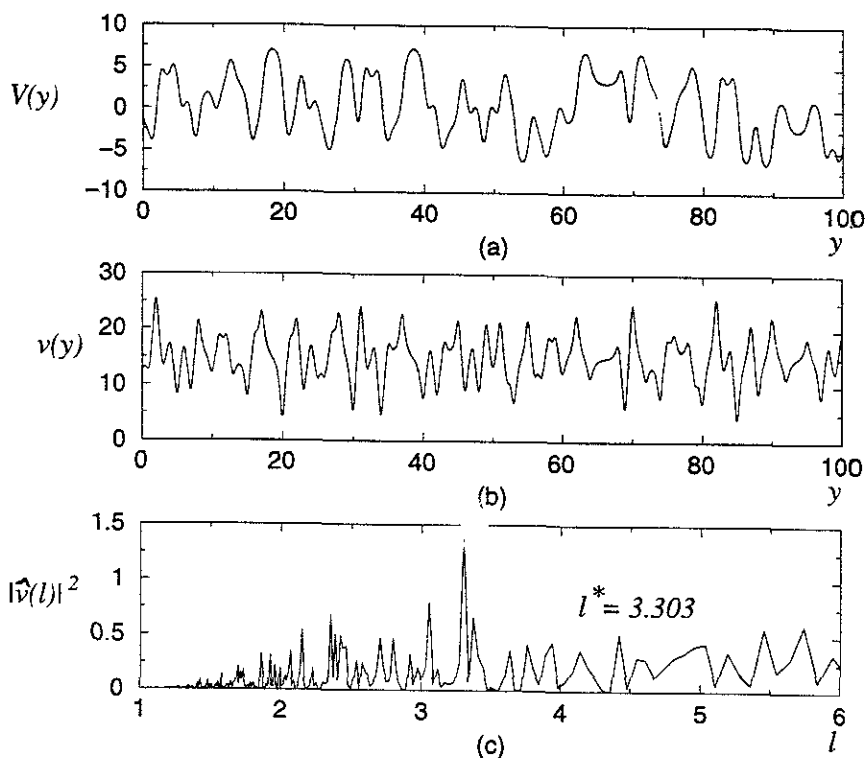


Figura 5.4: (a) Gráfica del potencial coulombiano dado en la ecuación (4.10) del capítulo anterior. Se utilizaron cuatro valores de carga para los monómeros (caso  $m = 4$ ). (b) Velocidad de la partícula como función de su posición a lo largo del polímero. (c) Espectro de potencias de la velocidad anterior. Nótese que en este caso el pico más grande en el espectro aparece en  $l^* = 3.303$ .

Como podemos observar de la Fig5.4, se obtiene nuevamente un pico que sobresale en el espectro de potencias justo en  $l^* = 3.303$ . Hemos trabajado con diferentes tipos de potenciales, suponiendo que la partícula de prueba es un ión, o un dipolo permanente, o un dipolo inducido, y con diferentes tipos de densidades de carga en los monómeros (uniforme, simétrica no uniforme y asimétrica) y en todos estos casos el espectro de potencias de la velocidad muestra un pico sobresaliente cerca del periodo 3. Debemos señalar que, aún cuando el pico cerca del periodo 3 siempre está presente,

en ocasiones (pocas) ocurre que hay otros picos también dominantes, tal y como se muestra en la Fig 5.5. Esto ocurre así porque estamos trabajando con procesos aleatorios es los cuales los valores particulares de las cantidades involucradas presentan fluctuaciones. No obstante, el caso genérico en los espectros de potencias es cuando el pico cerca del periodo 3 es el dominante.

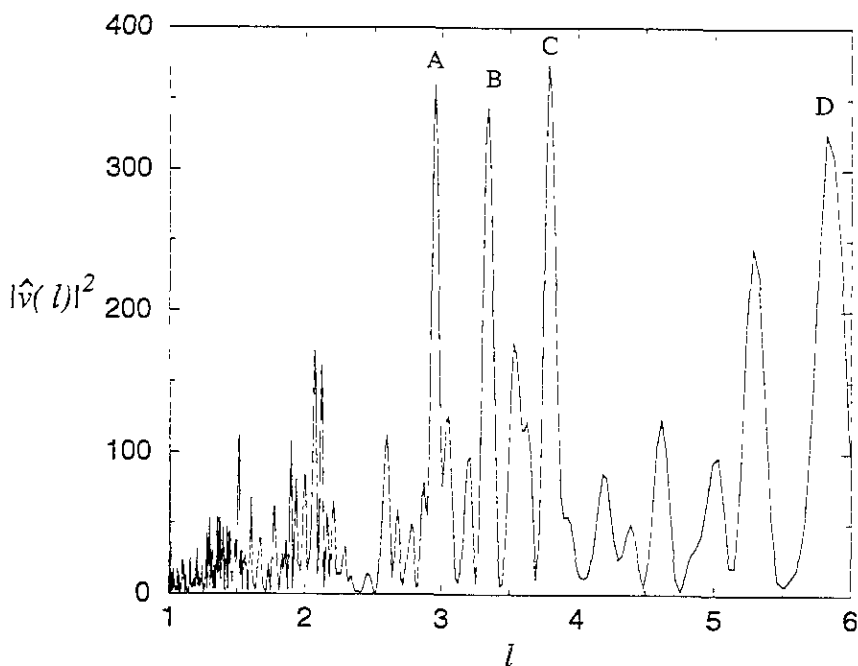


Figura 5.5: Caso en el que aparecen fluctuaciones alrededor del periodo tres. Como puede observarse, hay tres picos dominantes, marcados como A, B y C. Las periodicidades correspondientes a estos picos son:  $l_A^* = 2.95$ ,  $l_B^* = 3.34$  y  $l_C^* = 3.78$ . El pico en D, que también podría considerarse dominante, corresponde a  $l_D^* = 5.82$ .

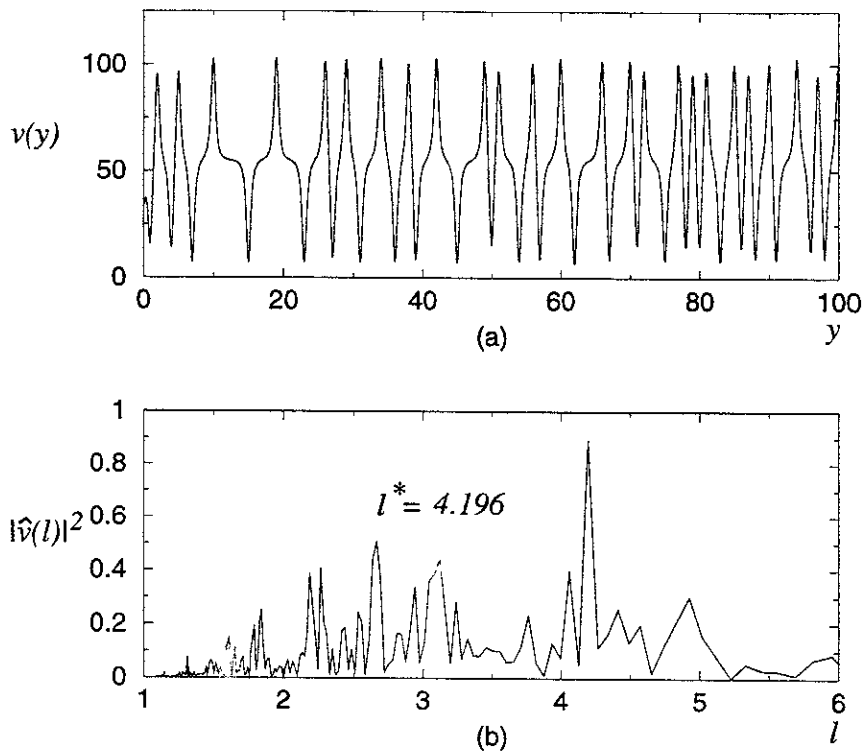


Figura 5.6: (a) Velocidad de la partícula como función de la posición. Se utilizó el potencial adimensional (5.9) y se asignaron sólo dos valores de carga a los monómeros (caso  $m = 2$ ). (b) Espectro de potencias de la velocidad. El pico más grande está en  $l^* = 4.196$ .

La presencia de este pico en el espectro de potencias de la velocidad claramente refleja que, aún cuando el potencial de interacción entre la partícula y el polímero es aleatorio, la dinámica generada por dicho potencial presenta cierto tipo de regularidades estadísticas. Nuestra interpretación es que *el pico en el espectro de potencias contiene la información de la distancia promedio que hay entre los mínimos consecutivos del potencial*. Para ver que esto es así, en la Fig5.6a mostramos la gráfica de la velocidad de la partícula a lo largo de un polímero cuyos monómeros adquieren sólo dos valores de carga,  $-1$  y  $+1$  (caso  $m = 2$ ). Para generar esta gráfica también utilizamos el potencial adimensional (5.9) con  $h = 0.5$ .

Por otro lado, la Fig5.6b muestra el espectro de potencias de la velocidad. Vemos que otra vez se tiene un pico en el espectro, pero en esta ocasión dicho pico se encuentra en  $l^* = 4.196$ .

Recordemos que la distancia promedio entre mínimos consecutivos del potencial es función del número  $m$  de cargas diferentes que pueden adquirir los monómeros (ecuación (4.18)):

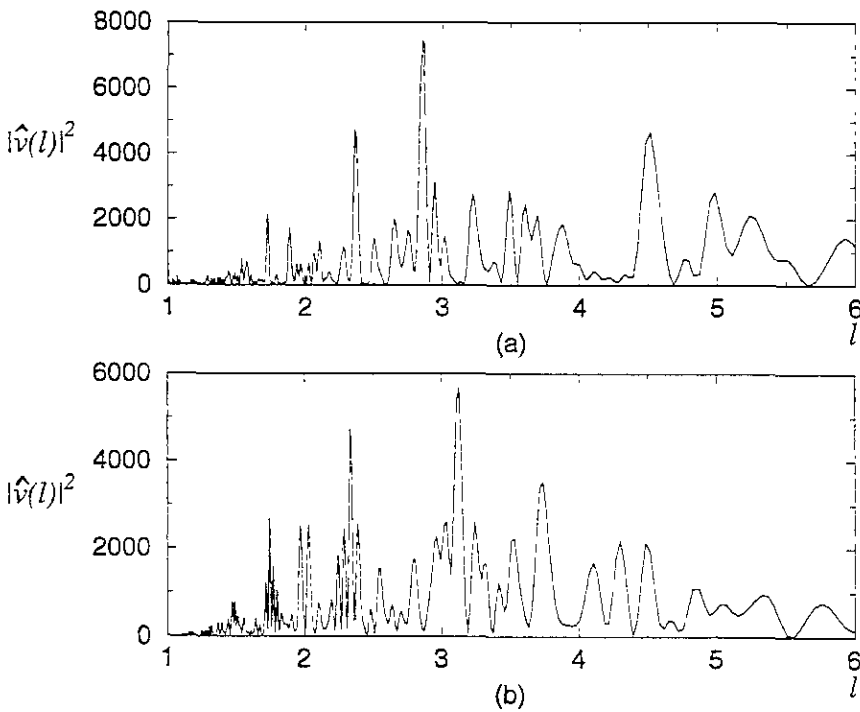


Figura 5.7: Espectros de potencias de la velocidad de la partícula utilizando dos asignaciones diferentes a los valores de carga en los monómeros. (a) Los valores de carga utilizados fueron  $-2, -1, 1, 2$ . (b) Los valores carga se seleccionaron aleatoriamente entre  $-1$  y  $1$ , resultando en este caso particular  $\{-0.31879, 0.35657, 0.57293, 0.89451\}$ .

$$\bar{d} = \frac{6m}{2m - 1}$$

Para  $m = 2$  se tiene  $\bar{d} = 4$ , mientras que para  $m = 4$  la distancia promedio es  $\bar{d} = 3.4$ . Comparando estos valores de la distancia promedio con los valores de periodicidad  $l^*$  obtenidos en los espectros de potencias de las figuras Fig5.3, Fig5.4 (caso  $m = 4$ ) y Fig5.6 (caso  $m = 2$ ), podemos concluir que, efectivamente, el pico en el espectro de potencias de la velocidad muestra que existe una periodicidad en el movimiento de la partícula dada por la distancia promedio entre mínimos consecutivos del potencial. La imagen que surge inmediatamente es la de un movimiento en el que la partícula va “brincando” de mínimo en mínimo, en pasos cuya longitud es, aproximadamente, de tres monómeros.

Vale la pena enfatizar que la ocurrencia de un pico cerca del periodo 3 en el espectro de potencias, no depende fuertemente de los valores particulares que adquieren las cargas. Hasta el momento, hemos trabajado con los valores particulares  $\{1, 2, 3, 4\}$ . Sin embargo, en la Fig5.7 se comparan los espectros de potencias de la velocidad utilizando diferentes valores de las cargas. Para el espectro mostrado en la Fig5.7a, los valores utilizados fueron  $\{-2, -1, 1, 2\}$ , mientras que para el espectro de la Fig5.7b, los cuatro valores de las cargas se escogieron aleatoriamente entre  $-1$  y  $1$ , resultando en este caso particular  $\{-0.31879, 0.35657, 0.57293, 0.89451\}$ . La asignación aleatoria de cargas a lo largo del polímero fue la misma en ambos casos. Como podemos observar de la figura, los espectros de potencias son muy semejantes.

Para concluir esta sección, en la Fig5.8 mostramos el espectro de potencias de la velocidad a lo largo del polímero, de una cadena compuesta por 20 partículas cargadas aleatoriamente y rígidamente unidas (Fig4.16 del capítulo 4). Esta gráfica se generó utilizando el potencial adimensional (5.9) con  $h = 0.5$  y los cuatro valores de carga  $\{\pm 1, \pm 2\}$  tanto en las partículas como en los monómeros. Vemos que también en este caso aparece un pico bien definido en  $l^* = 3.4$ .

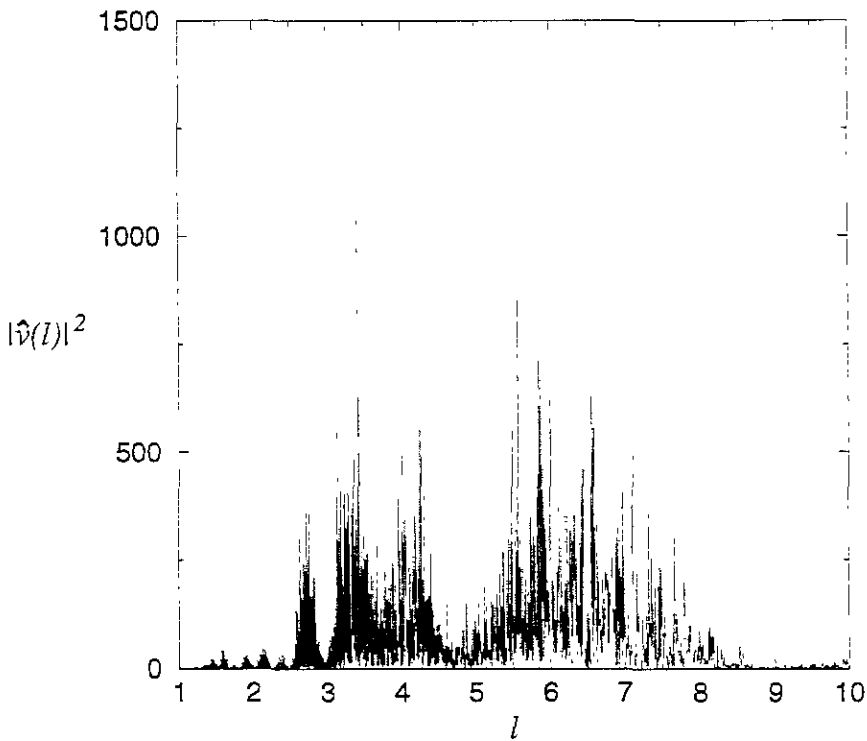


Figura 5.8: Espectro de potencias de la velocidad de una cadena compuesta por 20 partículas cargadas aleatoriamente y rígidamente unidas. El pico corresponde a una periodicidad espacial  $l^* = 3.4186$ . El potencial utilizado en este caso fue el dado en la ecuación (5.9) con  $h = 0.5$ , mientras que los valores de las cargas, tanto en los monómeros como en las 20 partículas fueron  $\{\pm 1, \pm 2\}$ .

## 5.4 Periodo 3 en secuencias reales

En la sección anterior estudiamos la dinámica de partículas moviéndose a lo largo de polímeros cuyos monómeros estaban cargados aleatoriamente. Sin embargo, también podemos asignar las cargas  $\{\beta_i\}$  de los monómeros en correspondencia con la secuencia de bases de una cadena genética real, tal



y como hicimos al final del capítulo anterior para encontrar la distribución de mínimos. Para esto, estableceremos la siguiente correspondencia entre bases y cargas:

Base	Carga $\beta_i$
A	+4
T	+3
C	+2
G	+1

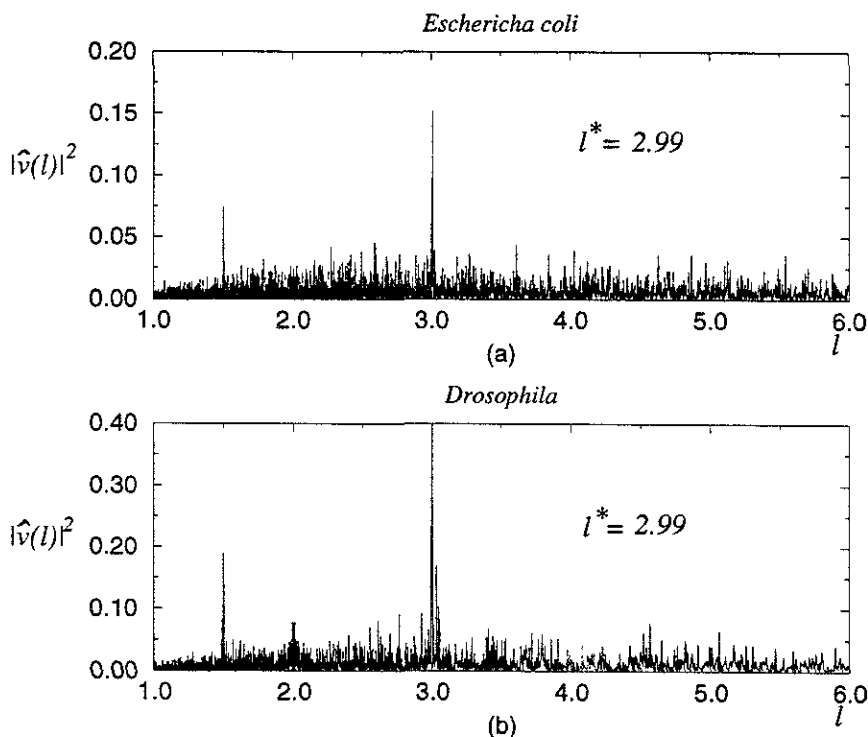


Figura 5.9: Espectro de potencias de la velocidad de la partícula a lo largo del polímero utilizando secuencias genéticas codificadoras de organismos reales, de aproximadamente 10000 bases de longitud. (a) *Escherichia coli*, (b) *Drosophila*. Obsérvese como el pico en el espectro de potencias ha aumentado considerablemente, además de que corresponde a una periodicidad espacial casi igual a 3.

De esta manera, las cargas  $\beta_i$  de los monómeros a lo largo del polímero ya no serán aleatorias, sino que estarán en correspondencia uno a uno con la secuencia de bases de una cadena genética real. El objetivo es ver si bajo estas condiciones hay algún cambio en la dinámica del polímero. Seguiremos trabajando con el potencial adimensional dado en la ecuación (5.9), tomando  $h = 0.5$ .

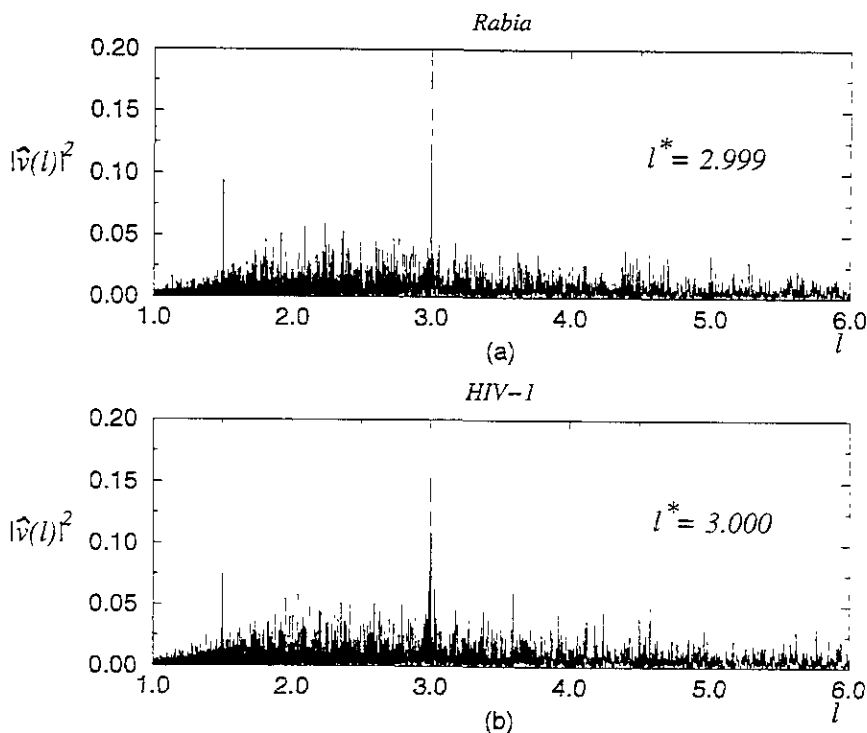


Figura 5.10: Igual que la figura anterior, pero esta vez utilizando secuencias genéticas de (a) el virus de la rabia y (b) el virus del SIDA VIH-1.

Los resultados obtenidos utilizando secuencias genéticas *codificadoras* de diferentes organismos se muestran en las figuras Fig5.9, Fig5.10 y Fig5.11. Dichas figuras muestran el espectro de potencias de la velocidad de la partícula a lo largo del polímero. Podemos observar dos cosas: en primer lugar, el pico en el espectro de potencias es mucho más grande que el que se obtenía con secuencias aleatorias. Esto refleja que existe una regularidad (o periodicidad) mucho mejor definida en la dinámica generada por poten-

ciales producidos por secuencias genéticas codificadoras. En segundo lugar, vemos que con estas secuencias genéticas la periodicidad espacial  $l^*$  determinada por el pico en el espectro de potencias está mucho más cercana a 3 que en el caso aleatorio. Así, la dinámica que se obtiene con las secuencias genéticas codificadoras no sólo presenta una periodicidad mejor definida, sino que además, dicha periodicidad está en  $l^* = 3$  casi exactamente.

Por otro lado, en la Fig.5.12 mostramos el espectro de potencias de la velocidad, pero esta vez utilizando secuencias genéticas *no codificadoras* de organismos reales (secuencias intergénicas). Como puede observarse, la estructura de tales espectros se asemeja mucho más a la estructura que se tiene con secuencias aleatorias, que a la que se tiene con secuencias genéticas codificadoras.

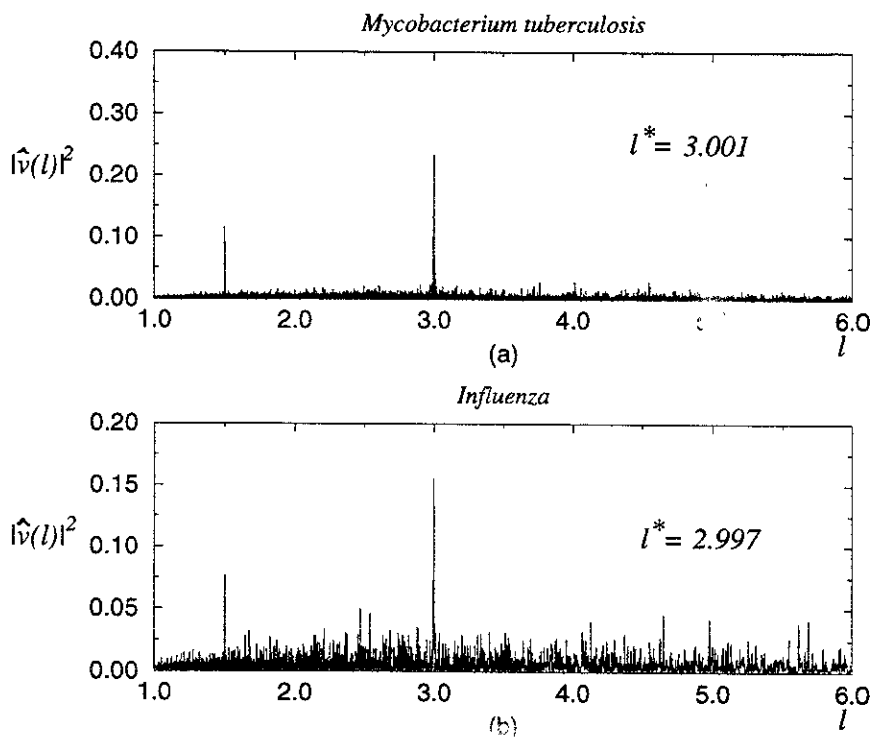


Figura 5.11: Espectro de potencias de la velocidad del polímero, pero ahora utilizando secuencias genéticas del bacilo de la tuberculosis y del virus de la influenza.

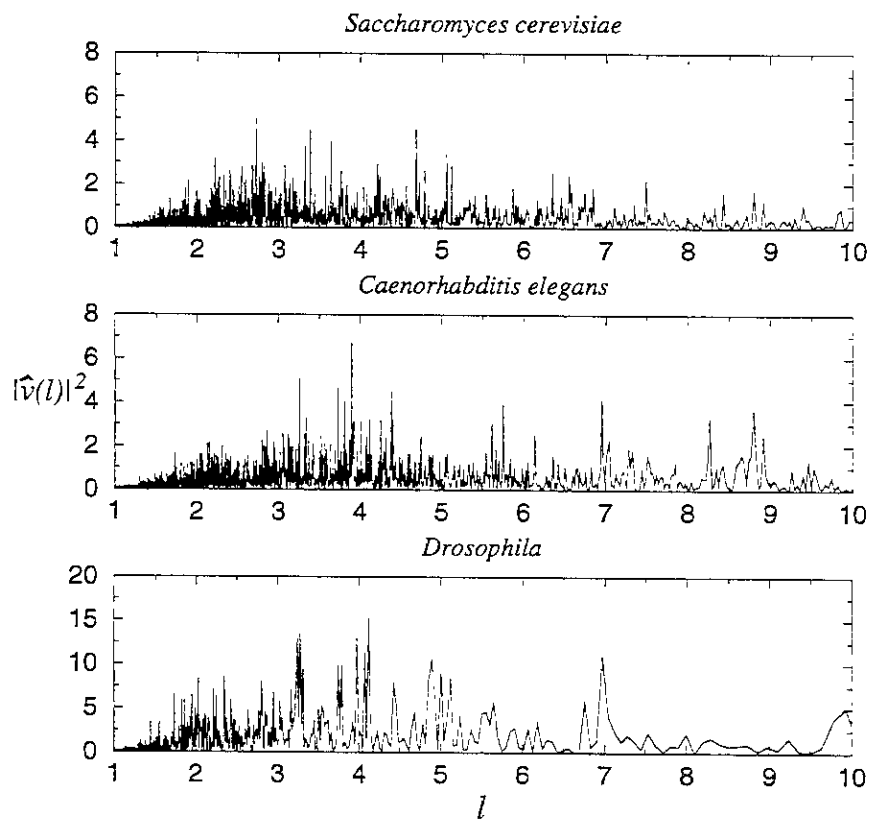


Figura 5.12: Espectro de potencias de la velocidad de la partícula utilizando trozos de secuencias intergénicas de organismos reales. (a) *Saccharomyces cerevisiae* (levadura), (b) *Caenorhabditis elegans* (gusano), (c) mosca *Drosophila*. Los trozos de secuencias intergénicas utilizados contenían aproximadamente 3000 bases, y el potencial que se utilizó fue el dado en la ecuación (5.9) con  $h = 0.5$ .

Con base en los resultados presentados en este capítulo (y en el anterior), podemos concluir que en el origen de la vida, cuando las secuencias genéticas aún no portaban ninguna información, existía una regularidad dinámica alrededor de tres en el sentido de que el ribosoma primitivo se movía a lo largo del RNA mensajero en pasos cuya longitud era, en promedio, de tres monómeros. Esta regularidad es la que se refleja en el pico del espectro de potencias de la velocidad cuando utilizamos secuencias de carga aleatorias a lo largo del polímero. Posteriormente, cuando las secuencias genéticas adquirieron mayor complejidad y se estructuró en ellas la información necesaria para autorreplicarse y construir otras moléculas, los mecanismos de codificación y lectura del mensaje genético se construyeron basándose en una dinámica que preferenciaba el movimiento de tres en tres. Lo que queda claro a partir de nuestros resultados es que dichos mecanismos de alguna manera "refinaron" las regularidades en la dinámica aleatoria ya existente, en el sentido de que acentuaron aún más la periodicidad del movimiento del ribosoma a lo largo del RNA mensajero, de tal manera que para las secuencias con información genética útil (codificadoras), dicha periodicidad quedó casi exactamente en tres. En las secuencias intergénicas, la periodicidad alrededor de 3 no se refinó porque dichas secuencias no participan en la dinámica de los procesos de codificación y lectura del mensaje genético.

Nuestra tesis principal puede resumirse de la siguiente manera:

*los mecanismos de codificación y lectura de la información genética se construyeron tomando como base una dinámica que ya de entrada preferenciaba el movimiento del ribosoma en pasos de tres monómeros a lo largo del RNA mensajero, y a lo largo de la evolución esta preferencia se refinó.*

## 5.5 Algunas consideraciones teóricas

En esta sección discutiremos bajo qué condiciones es válido el análisis de Fourier que realizamos en las dos secciones anteriores. Debido a que en nuestro modelo todos los monómeros son de la misma naturaleza, el potencial de interacción entre la partícula de prueba y el  $n$ -ésimo monómero en el polímero es de la forma

$$V_n(x) = q_n g_\sigma(x - n)$$

donde  $g(x)$  es una función predeterminada cuya forma depende sólo del tipo de interacción que estemos considerando, y  $q_n$  es una variable aleatoria que

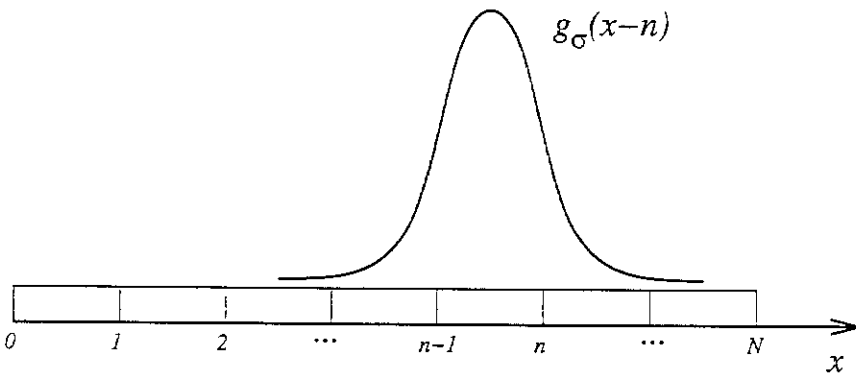


Figura 5.13: Esquema ilustrando la forma del potencial  $g_{\sigma}(x-n)$  producido por el  $n$ -ésimo monómero del polímero. Los potenciales individuales generados por cada uno de los demás monómeros tienen la misma forma que el mostrado aquí, sólo cambia su altura, la cual es proporcional a la carga  $q_n$  que porta cada monómero.

tiene que ver con la carga del monómero. Por lo tanto, el potencial de interacción entre todo el polímero y la partícula de prueba está dado por

$$V(x) = \sum_{n=1}^N q_n g_{\sigma}(x-n) \quad (5.16)$$

Como ejemplo de este tipo de potenciales podemos referirnos a las ecuaciones (4.10), (4.11) y (4.12) del capítulo 4. El potencial  $V(x)$  dado en la ecuación (5.8) de este capítulo es ligeramente diferente, ya que la suma está elevada al cuadrado. Sin embargo, los resultados que obtendríamos con este potencial son similares a los que se obtienen con potenciales de la forma dada en (5.16), en tanto las variables aleatorias  $\beta_i$  que aparecen en la expresión (5.8) sean estadísticamente independientes.

La ecuación de movimiento queda entonces como

$$\gamma v(x) = -\frac{\partial}{\partial x} \sum_{n=1}^N q_n g_{\sigma}(x-n) + F_0 \quad (5.17)$$

donde hemos supuesto que la fuerza externa  $F(t)$  que forza al sistema es constante e igual a  $F_0$ . Tomando la transformada de Fourier de la ecuación anterior respecto a la posición  $x$  obtenemos:

$$\gamma \hat{v}(k) = -(ik)\hat{g}_\sigma(k) \sum_{n=1}^N q_n \exp(-ink) + F_0 \sqrt{2\pi} \delta(k) \quad (5.18)$$

donde  $\hat{v}(k)$  y  $\hat{g}_\sigma(k)$  son las transformadas de Fourier de las funciones  $v(x)$  y  $g_\sigma(x)$ , respectivamente:

$$\begin{aligned} \hat{v}(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v(x) \exp(-ikx) dx \\ \hat{g}_\sigma(k) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g_\sigma(x) \exp(-ikx) dx \end{aligned}$$

El último término de la ecuación (5.18) es solamente la transformada de Fourier de la fuerza constante  $F_0$ . Dicho término sólo contribuye al espectro de potencias de la velocidad con un pico (infinito) en  $k = 0$ , y no tiene nada que ver con la estructura del potencial aleatorio. Por lo tanto, podemos eliminar este término de nuestro análisis, quedándonos solamente con la parte que está directamente relacionada con el potencial aleatorio. Sin embargo, debemos tener presente que la fuerza  $F_0$  debe estar presente en la ecuación de movimiento para forzar al sistema. La parte relevante de la transformada de Fourier de la velocidad en nuestro análisis es entonces

$$\hat{v}(k) = -\frac{1}{\gamma} (ik)\hat{g}_\sigma(k) \sum_{n=1}^N q_n \exp(-ink)$$

El espectro de potencias  $|\hat{v}(k)|^2$  se obtiene al multiplicar la expresión anterior por su complejo conjugado:

$$|\hat{v}(k)|^2 = \frac{1}{\gamma^2} k^2 |\hat{g}_\sigma(k)|^2 \sum_{n=1}^N \sum_{n'=1}^N q_n q_{n'} \exp[-i(n - n')k]$$

Para seguir adelante, tomaremos el promedio de la expresión anterior en el ensemble de todas las posibles realizaciones de las cargas  $\{q_n\}$ , lo que es equivalente a promediar utilizando la función de probabilidad  $P(q)$  que determina el valor de las variables  $\{q_n\}$ . Esto lo haremos suponiendo que la función  $P(q)$  satisface las propiedades (a) y (b) enunciadas en la sección 4.1 del capítulo anterior. De hecho, lo único que necesitamos es que las cargas  $\{q_n\}$  sean estadísticamente independientes, de tal manera que<sup>7</sup>  $\langle q_n q_{n'} \rangle = 0$  si  $n \neq n'$ , y que la función de probabilidad  $P(q)$  tenga

<sup>7</sup>Los paréntesis angulares  $\langle \dots \rangle$  indican promedio respecto a la función de probabilidad  $P(q)$ :  $\langle f \rangle = \int_{-\infty}^{+\infty} f(q) P(q) dq$ .

segundo momento bien definido:  $\langle q^2 \rangle = \int_{-\infty}^{+\infty} q^2 P(q) dq < \infty$ . Además, para simplificar el cálculo, supondremos que las cargas  $q_n$  tienen promedio cero:  $\langle q_n \rangle = 0$ . Esta suposición no es esencial y físicamente equivale a fijar el cero de la energía potencial. Bajo estas condiciones, el promedio del espectro de potencias queda como

$$\langle |\hat{v}(k)|^2 \rangle = \frac{1}{\gamma^2} N \langle q^2 \rangle k^2 |\hat{g}_\sigma(k)|^2 \quad (5.19)$$

Podemos observar que la naturaleza aleatoria de las cargas a lo largo del polímero, determinada por la función de probabilidad  $P(q)$ , sólo aparece en el espectro de potencias de la velocidad a través del segundo momento  $\langle q^2 \rangle$ , como un factor multiplicativo. Por lo tanto, la *forma* del espectro de potencias es la misma independientemente de si las cargas de los monómeros son variables discretas o continuas, en tanto la función de probabilidad  $P(q)$  que las determina tenga segundo momento. Lo que es más importante, es que la estructura del espectro de potencias está determinada por la función  $\hat{g}_\sigma(k)$ , es decir, por la transformada de Fourier del potencial producido por *un solo* monómero. Como ejemplo, consideremos el potencial coulombiano dado en la ecuación (4.10) del capítulo anterior. Para este potencial se tiene

$$g_\sigma(x) = \ln \left[ \frac{1 - x + \sqrt{\sigma^2 + (1 - x)^2}}{-x + \sqrt{\sigma^2 + x^2}} \right] \quad (5.20)$$

cuya transformada de Fourier está dada por[62]

$$\hat{g}_\sigma(k) = \sqrt{\frac{2}{\pi}} \frac{i}{k} K_0(\sigma k) [e^{-ik} - 1]$$

donde  $K_0(\sigma k)$  es la función de Bessel de segunda especie de orden cero<sup>8</sup>. Por lo tanto, en este caso el valor promedio del espectro de potencias de la velocidad queda como

$$\langle |\hat{v}(k)|^2 \rangle = \frac{32}{\gamma^2 \pi} N \langle q^2 \rangle [K_0(\sigma k) \sin(k/2)]^2 \quad (5.21)$$

En la Fig 5.14 se muestra la gráfica de la expresión anterior junto con el espectro de potencias calculado numéricamente para una realización particular del potencial coulombiano (5.20). Podemos observar de la gráfica que la expresión (5.21) nos da una versión “suavizada” del espectro de potencias. De hecho, la suavización es tal, que se pierde completamente el

<sup>8</sup>La razón por la que en esta última sección no utilizamos el potencial  $V(x)$  dado en la expresión (5.8), que fue con el que estuvimos trabajando a lo largo de este capítulo, es porque su transformada de Fourier es muy difícil de obtener analíticamente.



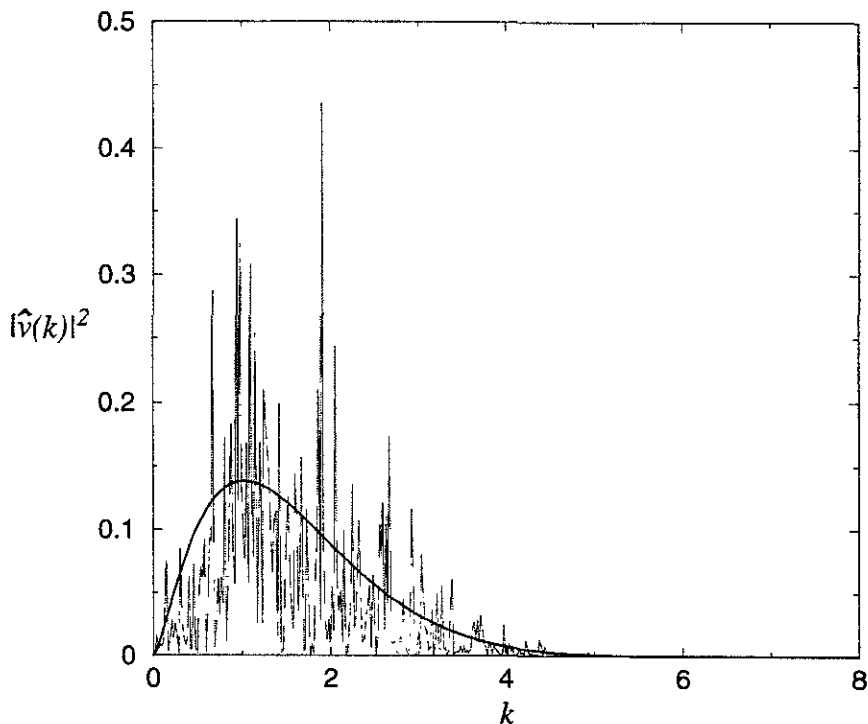


Figura 5.14: Espectro de potencias de la velocidad de la partícula para el potencial coulombiano dado en la ecuación (5.20). Se muestra la gráfica de la expresión teórica (5.21), así como el espectro de potencias calculado numéricamente para una realización particular del potencial.

pico en periodo 3. La razón de lo anterior radica en que el promedio del espectro de potencias dado en la expresión (5.19) depende únicamente del potencial generado por un sólo monómero, mientras que la estructura de mínimos en el potencial tiene que ver con por lo menos tres monómeros. Es claro que al hacer el promedio del espectro de potencias tal y como lo hicimos, en algún lado se pierde la información de la estructura de máximos y mínimos del potencial detectada por la transformada discreta de Fourier.

Esta “pérdida de información” al pasar de la transformada de Fourier discreta a la continua, queda mejor reflejada cuando se considera el espectro de potencias de la velocidad de una cadena extendida formada por partículas rígidamente unidas, moviéndose a lo largo del polímero.

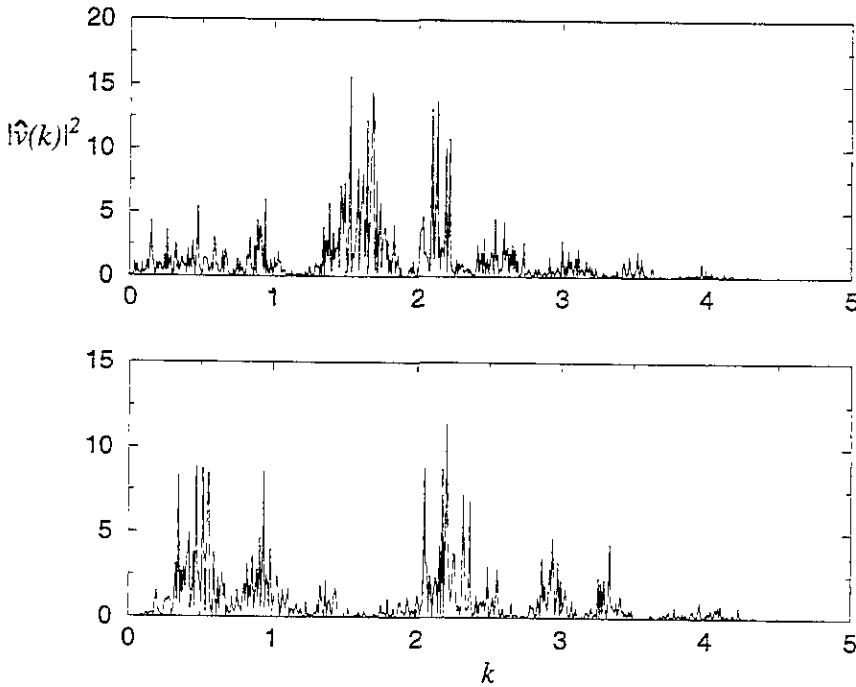


Figura 5.15: Espectros de potencias para dos realizaciones aleatorias de la velocidad de una cadena extendida a lo largo del polímero. El potencial utilizado fue coulombiano. Obsérvese que la estructura de estos espectros es notablemente diferente a la predicha por la ecuación (5.21), cuya gráfica es la curva continua de la Fig5.14.

En el Apéndice C se muestra que el espectro de potencias promedio de la velocidad de la cadena está dado por

$$\langle |\hat{v}(k)|^2 \rangle = \frac{1}{\gamma^2} MNq^2p^2k^2 |\hat{g}_\sigma(k)|^2 \quad (5.22)$$

donde  $M$  es el número de partículas que conforman la cadena,  $N$  es el número de monómeros en el polímero, y  $q^2$  y  $p^2$  son los segundos momentos de las cargas en el polímero y la cadena, respectivamente. Podemos ver que la única diferencia con la ecuación (5.21) se tiene en el factor  $Mp^2$ . Por lo tanto, de acuerdo con el análisis presentado en el Apéndice C, el valor promedio del espectro de potencias está determinado únicamente por

el potencial de interacción de un solo monómero con una sola partícula. Sin embargo, en la Fig5.15 se muestra el espectro de potencias para dos realizaciones de la velocidad de la cadena extendida, calculado numéricamente utilizando la transformada de Fourier discreta. Para generar esta gráfica se utilizó el potencial coulombiano (5.20), una cadena compuesta por 20 partículas y un polímero de 657 monómeros. Como puede apreciarse, la estructura del espectro numérico mostrado en la Fig5.15 difiere notablemente de la dada por la ecuación (5.22), cuya gráfica es la curva continua de la Fig5.14 (salvo por un factor de escala que tiene que ver con el producto  $Mp^2$ ).

Los resultados precedentes muestran que el promedio analítico del espectro de potencias realizado para obtener las ecuaciones (5.19) y (5.22) es demasiado burdo y no toma en cuenta la estructura de máximos y mínimos del potencial. Hay información que se pierde al hacer este promedio y no hemos entendido por qué. Sin embargo, en el siguiente capítulo veremos que la estructura "suave" del espectro de potencias está determinada por las correlaciones de largo alcance de los potenciales de interacción, lo cual tiene que ver más bien con el movimiento de objetos extendidos *continuos* a lo largo del polímero.



## Capítulo 6

# Interacción con objetos extendidos

En los capítulos anteriores se presentó un modelo en el cual un polímero interactuaba con una partícula puntual, o bien, con un objeto extendido compuesto por partículas puntuales rígidamente unidas. En este capítulo modificaremos un poco el modelo anterior, sustituyendo a la partícula puntual por un objeto extendido *continuo*. El objetivo es estudiar la dinámica generada por los potenciales aleatorios, en el caso cuando prevalece en el sistema una *interacción colectiva*, es decir, cuando diferentes partes del polímero interactúan simultáneamente con el objeto extendido. En el presente capítulo nos desligaremos del problema de la estructura del código genético, y estudiaremos las propiedades de los potenciales de interacción entre el polímero y el objeto extendido casi sin hacer referencia al problema particular que se ha estado tratando.

Este cambio de perspectiva en el trabajo se debe a que los potenciales aleatorios como los que se han utilizado en el modelo, también se utilizan ampliamente en otras áreas de la Física, Química y Biología, ya que son parte integral en el modelamiento de sistemas desordenados [63]. Aparecen en la descripción de fenómenos de localización, transporte anómalo, estados vítreos, transporte intracelular y máquinas moleculares, por mencionar sólo algunos ejemplos. Sin embargo, a pesar de esta gama tan amplia de aplicación de los potenciales aleatorios, sus propiedades estadísticas normalmente se postulan como parte del modelo, sin considerar si tales propiedades son físicamente realizables o no. Por ejemplo, en algunos casos se habla de potenciales unidimensionales “ $\delta$ -correlacionados” [64], [65], es decir, potenciales cuyo valor en un punto no depende en absoluto del va-

lor que puedan adquirir en algún otro punto diferente. Como se verá más adelante, tales potenciales no parecen ser físicamente realizables.

Por otro lado, ha llegado a ser evidente que las correlaciones de largo alcance de los potenciales juegan un papel crucial en los procesos bajo consideración. En muchos sistemas físicos, dichas correlaciones están presentes simplemente porque las partículas interactúan a través de interacciones de largo alcance. Por lo tanto, aún distribuciones aleatorias de cargas completamente descorrelacionadas, pueden dar lugar a perfiles de potencial correlacionados espacialmente.

En el capítulo 4 se presentó un análisis estadístico de los potenciales aleatorios, en el caso particular en el que  $\sigma$ , la distancia perpendicular entre la partícula y el polímero, era muy pequeña. El límite  $\sigma \rightarrow 0$  conduce a trabajar con potenciales escalonados, es decir, con variables aleatorias discretas. Por otro lado, en este capítulo se considerará la situación opuesta, es decir, el caso en el que  $\sigma$  es muy grande ( $\sigma \rightarrow \infty$ ). Como se verá, el tener un objeto extendido interactuando con un polímero es completamente equivalente a tener una partícula puntual interactuando con el mismo polímero, pero en este último caso, la distancia perpendicular entre la partícula y el polímero debe ser muy grande.

## 6.1 Equivalencia entre objeto extendido y partícula puntual

Consideremos el sistema mostrado en la Fig6.1a, en el que un polímero cuyos monómeros están cargados aleatoriamente, interactúa con un objeto extendido de longitud  $L_0$ . Supóngase para concretar, que la carga tanto en los monómeros como en el objeto extendido es coulombiana y se encuentra distribuida uniformemente. Sea  $\lambda_n$  la densidad lineal de carga eléctrica del  $n$ -ésimo monómero del polímero, y  $\lambda_0$  la densidad lineal de carga del objeto extendido. Si  $x$  denota la posición horizontal del objeto extendido respecto del polímero, entonces el potencial de interacción  $V_n(x)$  entre el  $n$ -ésimo monómero y el objeto está dado por

$$V_n(x) = K\lambda_n\lambda_0 \int_{x-L_0/2}^{x+L_0/2} \int_{n-1}^n \frac{1}{\sqrt{(x' - x'')^2 + \sigma^2}} dx' dx''$$

donde se ha supuesto que la longitud de los monómeros es  $L = 1$ . El potencial de interacción total  $V_0(x)$  entre todo el polímero y el objeto extendido es la suma de los potenciales individuales  $V_n(x)$ , con lo que resulta

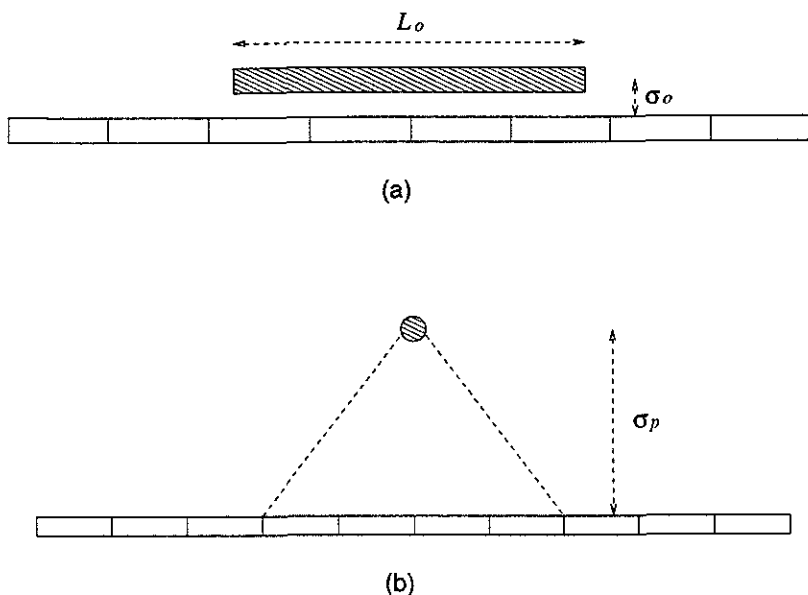


Figura 6.1: (a) Diagrama esquemático mostrando la interacción colectiva entre un polímero y un objeto extendido de longitud  $L_o$ . La distancia perpendicular entre ambos es  $\sigma_o$ . (b) Un sistema equivalente al anterior, en el cual una partícula puntual interactúa con el mismo polímero, solo que ahora la distancia perpendicular  $\sigma_p$  entre la partícula y el polímero es mucho más grande que en el caso anterior:  $\sigma_p \gg \sigma_o$ . La distancia  $\sigma_p$  es, en general, función de la longitud  $L_o$  del objeto extendido.

$$V_o(x) = K\lambda_o \sum_{n=1}^N \lambda_n \int_{x-L_o/2}^{x+L_o/2} \int_{n-1}^n \frac{1}{\sqrt{(x' - x'')^2 + \sigma^2}} dx' dx''$$

lo cual puede escribirse como

$$V_o(x) = \int_{x-L_o/2}^{x+L_o/2} \left[ \sum_{n=1}^N K\lambda_o \lambda_n \int_{n-1}^n \frac{1}{\sqrt{(x' - x'')^2 + \sigma^2}} dx' \right] dx'' \quad (6.1)$$

Nótese que lo que quedó dentro de los corchetes en la expresión anterior, puede considerarse como el potencial de interacción entre todo el polímero

y una *partícula puntual* hipotética de carga  $\lambda_o dx''$  que se encuentra ubicada en la posición  $x''$ . Si llamamos  $V_p(x'')$  a este potencial hipotético, entonces el potencial  $V_o(x)$  queda como

$$V_o(x) = \int_{x-L_o/2}^{x+L_o/2} V_p(x'') dx''$$

La ecuación anterior no es más que el principio de superposición. Puede verse de esta ecuación que el objeto extendido "siente" directamente la interacción de varios monómeros en el polímero, a saber, los monómeros que se encuentran comprendidos en el intervalo  $[x - L_o/2, x + L_o/2]$  y que están directamente debajo del objeto. Por lo tanto, el colocar un objeto extendido en lugar de una partícula puntual es, de alguna manera, equivalente a aumentar el rango de la interacción.

Este efecto de "aumentar el rango de la interacción" puede obtenerse también con una partícula puntual simplemente alejando la partícula del polímero en dirección perpendicular. En otras palabras, si la partícula se encuentra muy cerca del polímero ( $\sigma \rightarrow 0$ ), entonces sólo "siente" al monómero que está justo debajo de ella. Esta es la razón por la que los potenciales se vuelven escalonados cuando  $\sigma \rightarrow 0$ . Si se aleja a la partícula en dirección perpendicular al polímero, entonces "sentirá" la influencia de más monómeros, no sólo el que se encuentra justo por debajo. Por lo tanto, cabría esperar que el colocar un objeto extendido de longitud  $L_o$  interactuando con el polímero, sea equivalente a tener una partícula puntual interactuando con el mismo polímero, pero ahora la partícula se debe encontrar a una distancia perpendicular  $\sigma_p$  lo suficientemente grande. El valor de  $\sigma_p$  dependerá, por supuesto, del tamaño  $L_o$  del objeto extendido, de tal manera que entre más grande sea  $L_o$ , más grande tendrá que ser también  $\sigma_p$ . La situación se ilustra esquemáticamente en la Fig6.1b.

Claramente, al alejar a la partícula puntual del polímero, el potencial de interacción disminuye en intensidad, así que la amplitud del potencial deberá reescalarsse, multiplicándola por alguna constante adecuada, para que sea equivalente al potencial del objeto extendido. En la Fig6.2a se muestra el potencial de interacción entre un polímero y un objeto extendido. En este caso, el objeto extendido tenía una longitud  $L_o = 20$ , y estaba colocado a una distancia perpendicular  $\sigma_o = 0.5$ . Por otro lado, en la Fig6.2b se muestra el potencial de interacción entre una partícula puntual con el mismo polímero que en el caso anterior. Para obtener este potencial, colocamos a la partícula puntual a una distancia  $\sigma_p = 5.5$ , y reescalamos la amplitud del potencial por un factor de 50. La Fig6.2c muestra la superposición de los dos potenciales. Como puede verse, salvo por algunos "chipotitos", los dos potenciales son casi el mismo, lo cual confirma que un objeto extendido es,



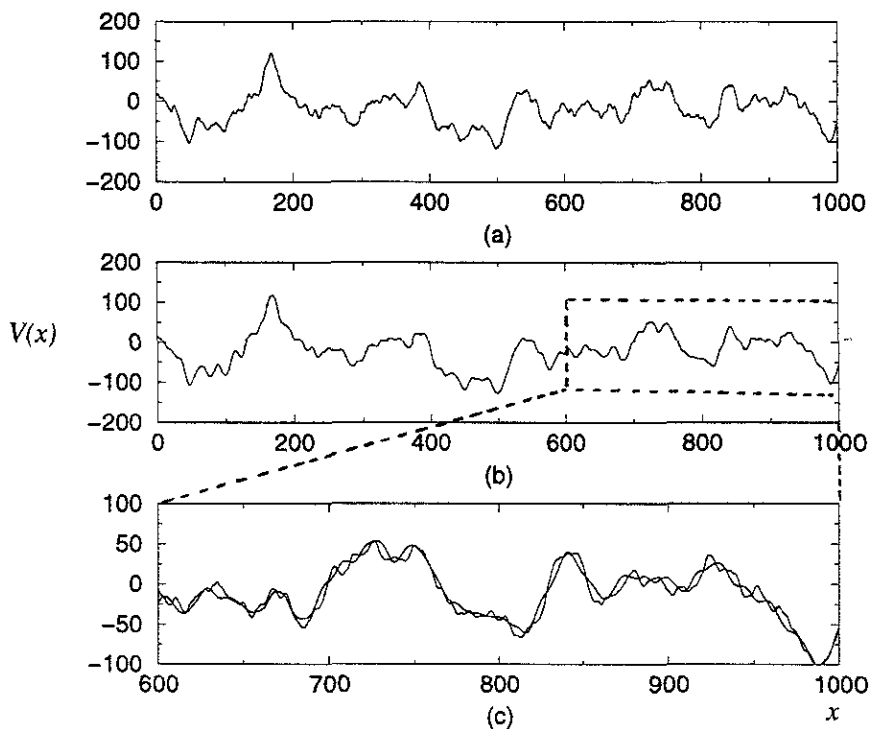


Figura 6.2: (a) Potencial de interacción entre un polímero y un objeto extendido de longitud  $L_o = 20$ , separados por una distancia perpendicular  $\sigma_o = 0.5$  (b) Potencial de interacción entre una partícula puntual y el mismo polímero que en el caso anterior (misma distribución de cargas). Aquí, la partícula se encontraba a una distancia  $\sigma_p = 5.5$  perpendicular al polímero, y el potencial se ha reescalado por un factor de 50. (c) Superposición (ampliada) de los dos potenciales anteriores.

con buena aproximación, equivalente a una partícula puntual pero colocada más lejos.

Las consideraciones anteriores hacen ver que el problema del objeto extendido se convierte en el problema tratado en los capítulos precedentes, donde se tenía una partícula puntual interactuando con el polímero. No obstante, los resultados obtenidos en los capítulos anteriores no son válidos para este nuevo caso, porque los potenciales tratados con anterioridad no presentaban correlaciones de largo alcance, mientras que ahora dichas correlaciones sí están presentes.

Antes de comenzar con el análisis para  $\sigma$  grande, es necesario hacer una puntualización más. Si la partícula está muy lejos del polímero en dirección perpendicular, entonces no importa si los monómeros son barritas de longitud  $L = 1$  o si son partículas puntuales separadas por una distancia  $L = 1$ . Desde lejos, una barrita se "ve" como si fuera una partícula puntual. Este resultado, que es una consecuencia del desarrollo multipolar a distancias grandes de los potenciales electrostáticos, queda ejemplificado en la Fig6.3, donde se muestra el potencial de interacción entre el polímero y la partícula puntual para dos casos diferentes: (a) los monómeros del polímero son barritas cargadas de longitud  $L = 1$ , y (b) los monómeros son partículas puntuales separadas una distancia  $L = 1$ . Para generar estas gráficas se utilizó una distancia  $\sigma = 10$  (que todavía está muy lejos de infinito). Como puede observarse de la figura, ambos potenciales son prácticamente idénticos. Debido a que a lo largo de este capítulo se supondrá que la distancia  $\sigma$  adquiere valores suficientemente grandes, utilizaremos el modelo mostrado en la Fig6.4, en el cual tanto los monómeros del polímero como la partícula de prueba son partículas puntuales cargadas. Además, para evitar que el polímero completo se convierta en una partícula puntual cuando  $\sigma$  es mucho muy grande, trabajaremos con un polímero infinito.

## 6.2 Potenciales aleatorios correlacionados

Como vimos en el capítulo 4, el potencial de interacción  $V(r)$  entre dos partículas cargadas  $q_1$  y  $q_2$  separadas por una distancia  $r$  puede representarse por medio de una ley de potencias de la forma siguiente:

$$V(r) = K \frac{q_1 q_2}{r^\alpha}$$

donde  $K$  es una constante. El exponente  $\alpha$  establece el tipo de interacción bajo consideración:  $\alpha = 1$  corresponde a interacción coulombiana,  $\alpha = 2$  a interacción dipolar, etc. Recordemos que  $q_1$  y  $q_2$  no se refieren solamente a carga coulombiana, sino que  $q_1$  y  $q_2$  pueden ser también momentos dipolares, o momentos dipolares inducidos, etc.

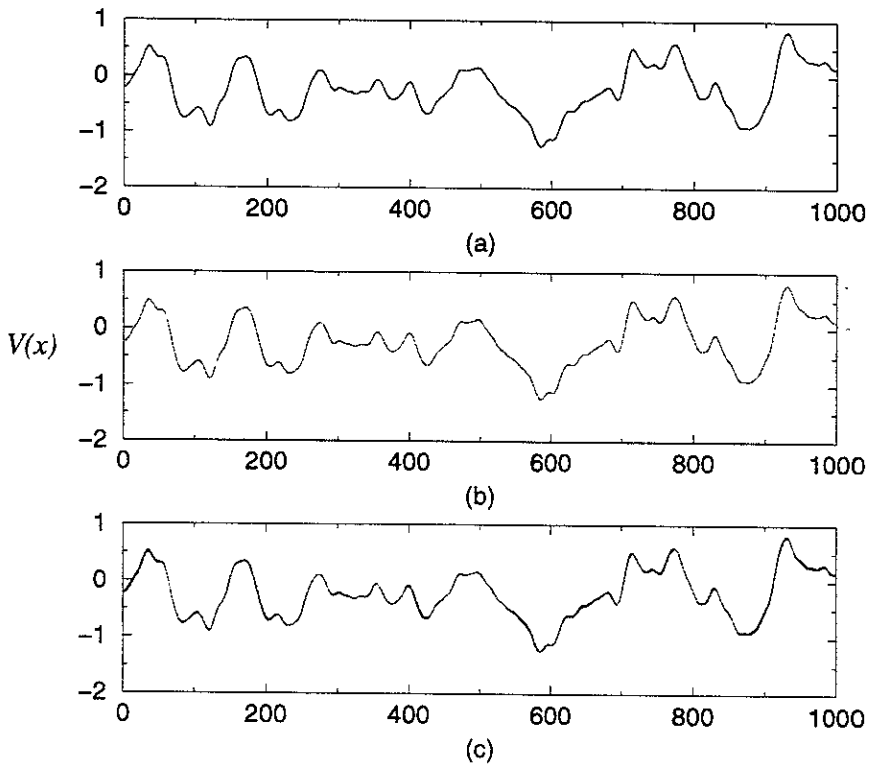


Figura 6.3: (a) Potencial de interacción generado por un polímero cuyos monómeros son barritas cargadas uniformemente de longitud  $L = 1$ . (b) Potencial generado por un polímero con la misma distribución de cargas que en el caso anterior, pero ahora los monómeros son partículas puntuales separadas por una distancia  $L = 1$ . (c) Superposición de las dos gráficas.

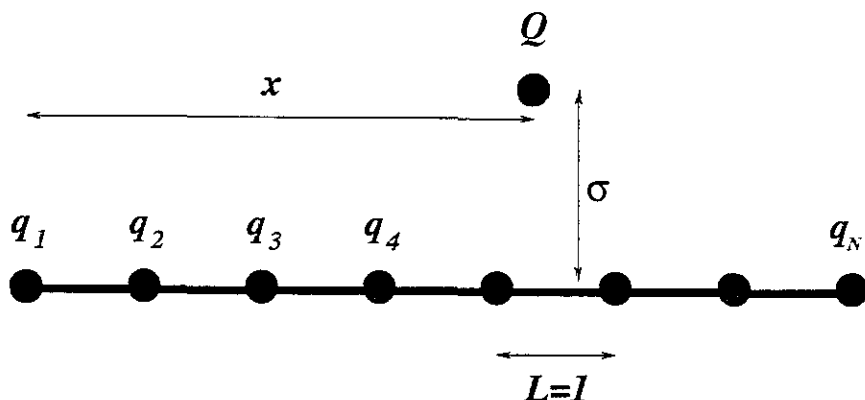


Figura 6.4: Diagrama esquemático del sistema con el que se trabajará en este capítulo.

En el sistema que se muestra en la Fig.6.4 supondremos que todos los monómeros son del mismo tipo, es decir, todos son iones, o todos son dipolos, o todos son dipolos inducidos, y que lo único que cambia de un monómero a otro es el valor de la “carga”  $q_n$ , la cual es una variable aleatoria con función de probabilidad  $P(q)$ . El potencial de interacción entre el  $n$ -ésimo monómero del polímero  $q_n$  y la carga de prueba  $Q$  está dado entonces por

$$V_{\sigma,n}^{\alpha}(x) = a_n g_{\sigma}^{\alpha}(x - n) \quad (6.2)$$

donde  $a_n = KQq_n$  y  $g_{\sigma}^{\alpha}(x)$  es una función definida como

$$g_{\sigma}^{\alpha}(x) \equiv \frac{1}{[x^2 + \sigma^2]^{\alpha/2}}. \quad (6.3)$$

Sumando las contribuciones de cada uno de los monómeros obtenemos que el potencial de interacción total  $V(x)$  entre la partícula de prueba  $Q$  y todo el polímero está dado por

$$V(x) = \sum_{n=-\infty}^{\infty} V_{\sigma,n}^{\alpha}(x) = \sum_{n=-\infty}^{\infty} a_n g_{\sigma}^{\alpha}(x - n) \quad (6.4)$$

(hemos supuesto que el polímero es infinito). Obsérvese que las variables aleatorias  $\{a_k\}$  tienen la misma función de probabilidad  $P(q)$  que las cargas  $\{q_n\}$ , excepto por un factor de normalización que siempre puede tomarse

igual a uno. Por lo tanto, de ahora en adelante no haremos distinción entre estos dos conjuntos de variables aleatorias.

Para hacer análisis vamos a requerir que la función  $P(q)$  satisfaga las propiedades (a) y (b) establecidas en el capítulo 4. Esencialmente, estas propiedades garantizan que las cargas  $\{q_n\}$  sean estadísticamente independientes<sup>1</sup>, propiedad que evidentemente se hereda a las variables  $\{a_n\}$ , y que la función  $P(q)$  tenga segundo momento bien definido. Como veremos, lo anterior será de fundamental importancia para la convergencia de la suma que aparece en la ecuación (6.4). Por lo demás, la función  $P(q)$  será completamente arbitraria.

En este punto vale la pena mencionar que el parámetro  $\sigma$ , además de ser la distancia perpendicular entre la partícula de prueba y el polímero, puede considerarse como la escala de longitud a la cual los potenciales individuales  $V_{\sigma,n}^\alpha(x)$  contribuyen apreciablemente al potencial total de interacción  $V(x)$  en una posición particular  $x$  (en otras palabras,  $\sigma$  es la escala de longitud en la cual la función  $g_\sigma^\alpha(x)$  es apreciablemente diferente de cero). En la Fig6.5 se muestran cuatro realizaciones del potencial total  $V(x)$  dado en la expresión (6.4) para diferentes valores de  $\sigma$  y  $\alpha = 4$ . Esas gráficas las construimos con la distribución de probabilidad  $P(a) = \frac{1}{6} \sum_{m=1}^3 \delta(a \pm m)$ . Puede observarse de esta figura que los potenciales se suavizan conforme el valor de  $\sigma$  aumenta (nótese la diferencia de escalas en las gráficas). Esta "suavización" de los potenciales cuando  $\sigma$  adquiere valores cada vez más grandes es consecuencia de las correlaciones de largo alcance de los potenciales individuales  $V_{\sigma,n}^\alpha(x)$  generados por cada uno de los monómeros. En otras palabras, cuando el valor de  $\sigma$  aumenta, más y más monómeros contribuyen al potencial total  $V(x)$  en una posición particular  $x$ . Lo anterior puede expresarse diciendo que, cuando la partícula  $Q$  se aleja del polímero, dicha partícula es incapaz de "ver" la estructura fina del potencial, resolviendo sólo la estructura "macroscópica" o de "grano gordo". Por otro lado, si la partícula está muy cerca del polímero, entonces sólo "siente" al monómero que está justo por debajo de ella. Esta propiedad de suavización del potencial cuando  $\sigma$  crece, es la que permite sustituir al objeto extendido por una partícula puntual: el objeto extendido también suaviza el potencial.

---

<sup>1</sup>Si las cargas  $q_n$  a lo largo del polímero estuviesen correlacionadas con una longitud de correlación  $\xi$  finita, siempre puede "partirse" al polímero en bloques de tamaño  $\xi$  los cuales no estarán ya correlacionados. Cada bloque puede interpretarse entonces como una nueva partícula puntual cuya carga es la suma de las cargas puntuales que conforman al bloque, obteniendo así otra vez una situación en la cual las nuevas cargas de bloque a lo largo del polímero no están correlacionadas.

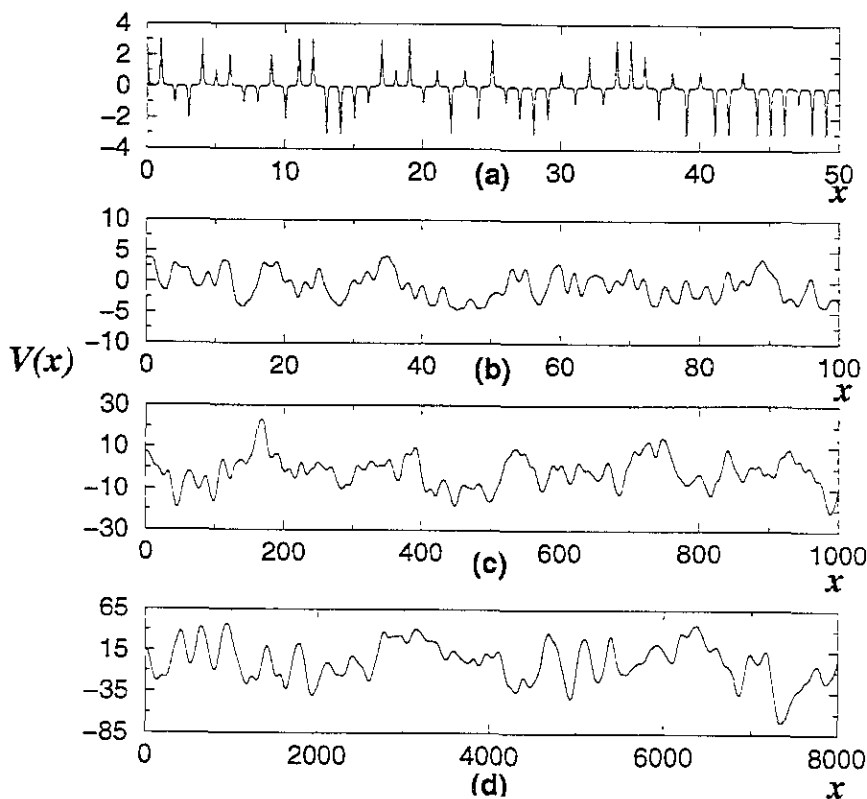


Figura 6.5: Potenciales de interacción para diferentes valores de  $\sigma$  y con  $\alpha = 4$ . (a)  $\sigma = 0.1$ . (b)  $\sigma = 1$ . (c)  $\sigma = 10$ . (d)  $\sigma = 100$ . En todos los casos la amplitud del potencial se ha escalado por el factor  $\sigma^\alpha$ . Nótese que los potenciales se “suavizan” cuando  $\sigma$  aumenta.

### 6.3 Propiedades de integrabilidad

Es claro que, dada una posición fija  $x$ , el potencial total de interacción  $V(x)$  es la suma de un número muy grande de variables aleatorias independientes, a saber, los potenciales individuales  $\{V_{\sigma,n}^\alpha(x)\}$  generados por cada uno de los monómeros. Estos potenciales individuales están caracterizados cada uno por su propia función de probabilidad. Por lo tanto, esperamos que en algún límite apropiado, el Teorema de Límite Central (TLC) pueda aplicarse con el objetivo de encontrar la función de probabilidad conjun-

ta  $P(V(x_1), V(x_2), \dots, V(x_r))$  que caracteriza completamente al potencial  $V(x)$ . Para que esto ocurra, es necesario que la suma que aparece en la expresión (6.4) se mantenga finita para todo valor de  $x$ . Por supuesto, los potenciales  $V_{\sigma, n}^{\alpha}(x)$  dados en las ecuaciones (6.2) y (6.3) cumplen con el hecho de que la suma  $\sum_{n=-\infty}^{\infty} V_{\sigma, n}^{\alpha}(x)$  se mantiene finita para todo valor de  $x$ , siempre y cuando  $\alpha > 1$ . Sin embargo, en general, para otro tipo de potenciales lo anterior no necesariamente se cumple. Es importante entonces discutir bajo qué condiciones dicha suma se mantiene finita.

Consideremos potenciales aleatorios que pueden escribirse de la forma

$$V(x) = \sum_{n=-\infty}^{\infty} a_n g(x - n) \quad (6.5)$$

donde  $a_n$  son variables aleatorias independientes y  $g(x)$  es una función pre-determinada, no necesariamente del tipo dado en la ecuación (6.3). Estos potenciales tendrán o no significado dependiendo de las propiedades de integrabilidad de la función  $g(x)$ . Para nuestros propósitos, será conveniente distinguir entre tres casos diferentes.

1.  $\int_{-\infty}^{\infty} |g(x)| dx < \infty$ . En este caso,  $g(x)$  es absolutamente integrable y no se tiene ningún problema con la suma de la ecuación (6.5), debido a que dicha suma permanece finita para cualquier valor de  $x$ . En efecto, es fácil ver que  $V(x)$  está acotado, ya que

$$\begin{aligned} |V(x)| &= \left| \sum_{n=-\infty}^{\infty} a_n g(x - n) \right| \\ &\leq R \sum_{n=-\infty}^{\infty} |g(x - n)| \\ &\simeq R \int_{-\infty}^{\infty} |g(x)| dx < \infty \end{aligned}$$

donde  $R$  es una cota superior para las variables aleatorias  $a_n$  (propiedad (b) de la función  $P(a)$  —capítulo 4). En particular, nótese que las funciones  $g_{\sigma}^{\alpha}(x)$  dadas en la expresión (6.3) son absolutamente integrables si  $\alpha > 1$ . Nos referiremos al caso en el que  $\int_{-\infty}^{\infty} |g(x)| dx < \infty$  como el *caso de potencial acotado*.

2.  $\int_{-\infty}^{\infty} |g(x)| dx$  no converge pero  $\int_{-\infty}^{\infty} [g(x)]^2 dx < \infty$ . Dado que estamos suponiendo  $\langle a_n \rangle = 0$ , el potencial aleatorio  $V(x)$  no diverge para ningún valor de  $x$ , pero como  $g(x)$  no es absolutamente integrable,

$V(x)$  puede adquirir valores muy grandes. Sin embargo,  $V(x)$  tiene un *valor típico* dado por su varianza, la cual es finita:

$$\begin{aligned}
 \text{Var}[V(x)] &= \langle [V(x)]^2 \rangle \\
 &= \left\langle \left[ \sum_{n=-\infty}^{\infty} a_n g(x-n) \right]^2 \right\rangle \\
 &= \left\langle \sum_{n=-\infty}^{\infty} a_n g(x-n) \sum_{n'=-\infty}^{\infty} a_{n'} g(x-n') \right\rangle \\
 &= \sum_{n=-\infty}^{\infty} \langle a_n^2 \rangle [g(x-n)]^2 + \\
 &\quad + \sum_{n \neq n'} \langle a_n a_{n'} \rangle g(x-n) g(x-n') \\
 &= \Delta^2 \sum_{n=-\infty}^{\infty} [g(x-n)]^2 \\
 &\approx \Delta^2 \int_{-\infty}^{\infty} [g(x)]^2 dx < \infty
 \end{aligned}$$

donde se ha puesto  $\langle a_n^2 \rangle = \Delta^2$  y se ha utilizado el hecho de que las variables  $a_n$  son estadísticamente independientes:  $\langle a_n a_{n'} \rangle = 0$  si  $n \neq n'$ . Debido a que en este caso el potencial aleatorio  $V(x)$  tiene un valor típico, nos referiremos a esta situación como el *caso de potencial localizado*.

3.  $\int_{-\infty}^{\infty} |g(x)|^2 dx$  no converge. Si  $g(x)$  no es una función cuadrado integrable, entonces la suma en (6.5) conduce a divergencias. Esto es debido al hecho de que los potenciales individuales  $V_n(x) = a_n g(x-n)$  decaen tan despacio conforme  $x \rightarrow \pm\infty$ , que las contribuciones provenientes de las fluctuaciones de carga en el sistema no se cancelan lo suficientemente rápido como para asegurar la convergencia. Lo anterior puede visualizarse fácilmente en el caso en el que  $g(x)$  es constante. En esta situación, la superposición de  $N$  de estos potenciales individuales generados por las cargas aleatorias conduce a la misma "caminata aleatoria" en todo punto del espacio, cuyo valor típicamente diverge como  $\sqrt{N}$ . Otro ejemplo que conduce a divergencias lo constituye el caso en el que la partícula de prueba  $Q$  interactúa con cada uno de los monómeros en el polímero por medio de resortes de Hooke cuya constante elástica  $k_n$  es una varia-



ble aleatoria. En tal caso el potencial total  $V(x)$  estaría dado por  $V(x) = \sum_{n=-\infty}^{\infty} (k_n/2) [(x-n)^2 + \sigma^2]$ . Evidentemente, la suma en la expresión anterior diverge para todo valor de  $x$ , debido a que los potenciales individuales  $V_n(x) = (k_n/2) [(x-n)^2 + \sigma^2]$  crecen con la distancia.

El caso en el que  $g(x)$  no es cuadrado integrable puede caracterizarse todavía en dos partes, correspondiendo cada una a diferentes comportamientos físicos:

- *Potencial rugoso.* En muchas situaciones, las cantidades de interés físico son las fuerzas y no los potenciales en sí mismos. Si las fuerzas derivadas de los potenciales aleatorios son cuadrado integrables, entonces el teorema de límite central (TLC) puede aplicarse a las fuerzas en lugar de a los potenciales. La ecuación (6.5) se sustituye entonces por una expresión equivalente en la que aparecerían las derivadas de la función  $g(x)$ . Por otro lado, hay situaciones en las que los perfiles de potencial son particularmente importantes para el problema físico en consideración. Tal es el caso del transporte browniano y otros problemas de difusión relacionados. En estas situaciones, el tamaño y distribución de las barreras de potencial son los elementos importantes de la descripción. Por lo tanto, en dichas circunstancias estamos interesados en caracterizar las propiedades estadísticas de las "diferencias de potencial", lo que podemos hacer considerando como las variables aleatorias relevantes para el análisis a las cantidades  $W_n(x)$  definidas como  $W_n(x) = V_n(x) - V_n(y)$ , donde  $V_n(x)$  son los potenciales individuales generados por cada uno de los monómeros del polímero. Lo anterior es equivalente a colocar el cero de la energía potencial en la posición arbitraria  $x = y$ . Si las variables aleatorias  $W_n(x)$  resultan ser cuadrado integrables, puede aplicarse el TLC a estas variables. Una propiedad general que resulta bajo tales circunstancias es que los potenciales aleatorios así generados ya no son "regulares", sino que ahora son potenciales "rugosos" cuando los analizamos a escalas grandes de longitud, en el sentido de que  $\langle [V(x_1) - V(x_2)]^2 \rangle \sim |x_1 - x_2|^\nu$ , donde  $\nu$  es un exponente de escalamiento que depende del tipo particular de interacción que estemos considerando.
- *Potenciales no diferenciables.* Cuando ni las derivadas ni las diferencias de potencial son cuadrado integrables, los potenciales aleatorios generados no son derivables en ningún punto, lo cual conduce a situaciones que físicamente no tienen sentido. Lo an-

terior ocurre cuando los potenciales crecen lo suficientemente rápido con la distancia, como en el ejemplo de los osciladores armónicos referido anteriormente. Debido a que este tipo de potenciales no tienen significado físico, no los consideraremos en este trabajo.

De las consideraciones anteriores, queda claro que para poder hacer un análisis estadístico de los potenciales aleatorios  $V(x)$  basado en el TLC, es necesario que los potenciales individuales  $V_n(x)$  sean acotados o localizados. No obstante, lo anterior sólo es una condición necesaria, más no suficiente, ya que en la siguiente sección veremos que otro caso en el que el TLC puede *no ser aplicable*, es cuando el alcance de los potenciales es extremadamente corto. Por el momento, basta con mencionar que la clase de potenciales definida en las expresiones (6.4) y (6.3) da lugar a las cuatro clasificaciones de potenciales que acabamos de hacer, dependiendo del valor de  $\alpha$ : si  $\alpha > 1$ , entonces los potenciales son acotados; si  $1/2 < \alpha \leq 1$ , se tiene el caso de potenciales localizados, y si  $-1/2 < \alpha \leq 1/2$ , los potenciales son rugosos. Para  $\alpha \leq -1/2$  se tienen potenciales no diferenciables. En lo que resta de este trabajo supondremos que nuestros potenciales son acotados o localizados. El caso de potenciales rugosos se reduce a los dos casos anteriores si en lugar de considerar a los potenciales tal cual, consideramos a las fuerzas o a las diferencias de potencial.

## 6.4 El Teorema de Límite Central

Nuestro objetivo ahora es calcular la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2), \dots, V(x_r))$  que da la probabilidad de que el potencial  $V(x)$  adquiera diferentes valores determinados en los  $r$  puntos  $x_1, x_2, \dots, x_r$ . El propósito es mostrar que en un límite apropiado, puede aplicarse el TLC de tal suerte que el potencial  $V(x)$  resulta ser un proceso gaussiano. Para esto, trabajaremos otra vez con la clase de potenciales definida en las ecuaciones (6.4) y (6.3). Debe notarse que, en general, las variables aleatorias  $V_{\sigma,n}^\alpha(x)$  no están distribuidas de la misma manera, sino que las funciones de probabilidad que caracterizan dichas variables son diferentes. Si llamamos  $P_n(v)$  a la función de probabilidad asociada al potencial individual  $V_{\sigma,n}^\alpha(x)$ , entonces de la ecuación (6.2) se sigue que  $P_n(v)$  y  $P(a)$  están relacionadas a través de la siguiente ecuación:

$$P_n(v) = \frac{1}{g_{\sigma,n}^\alpha(x)} P\left(\frac{v}{g_{\sigma,n}^\alpha(x)}\right) \quad (6.6)$$

donde por simplicidad en la notación, se ha definido  $g_{\sigma,n}^\alpha(x)$  como

$$g_{\sigma,n}^{\alpha}(x) = g_{\sigma}^{\alpha}(x - n)$$

En la expresión (6.6)  $P(a)$  es la función de probabilidad asociada a las variables aleatorias  $\{a_n\}$ . Esta expresión muestra que las variables aleatorias  $V_{\sigma,n}^{\alpha}(x)$  no están distribuidas de la misma manera. Por lo tanto, si queremos determinar la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2), \dots, V(x_r))$  que caracteriza al potencial total  $V(x)$ , debemos utilizar la versión de Lindeberg del TLC, la cual dice lo siguiente [66]:

### Condición de Lindeberg.

**Teorema 1** Sean  $\xi_1, \xi_2, \dots, \xi_N$  variables aleatorias independientes cuyas funciones de probabilidad asociadas son  $P_1(\xi), P_2(\xi), \dots, P_N(\xi)$ , respectivamente. Supongamos que estas funciones de probabilidad son tales que el valor promedio  $\langle \xi_n \rangle = 0 \forall n = 1, 2, \dots, N$ , y que las varianzas están dadas por  $\text{Var}(\xi_n) = \delta_n$ . Sea  $S_N^2$  la suma

$$S_N^2 = \delta_1^2 + \delta_2^2 + \dots + \delta_N^2 \quad (6.7)$$

Si para todo número  $t > 0$  se satisface la siguiente igualdad:

$$\lim_{N \rightarrow \infty} S_N^{-2} \sum_{n=1}^N \int_{|\xi| > t S_N} \xi^2 P_n(\xi) d\xi = 0 \quad (6.8)$$

entonces la función de probabilidad de la suma normalizada  $\chi_N$  dada por

$$\chi_N = \frac{\xi_1 + \xi_2 + \dots + \xi_N}{S_N} \quad (6.9)$$

tiende a una distribución gaussiana con promedio cero y varianza unitaria conforme  $N \rightarrow \infty$ .

En la siguiente sección veremos qué tipo de potenciales satisfacen la condición de Lindeberg y la forma que adquiere dicha condición para el caso particular de los potenciales que estamos trabajando. Por el momento, discutiremos ahora a qué nos referimos por “límite apropiado” en el cual el TLC puede aplicarse. En primer lugar, notemos que la condición de Lindeberg requiere que el número de términos que contribuyen apreciablemente a la suma (6.9) sea muy grande (formalmente, infinito). En nuestro caso, las variables aleatorias  $\xi_n$  que aparecen en la condición de Lindeberg corresponden a los potenciales individuales  $V_{\sigma,n}^{\alpha}(x)$  generados por cada monómero, mientras que la suma  $\chi_N$  hace las veces del potencial total  $V(x)$ . Por lo tanto, para el problema que nos ocupa, la condición de Lindeberg requiere que

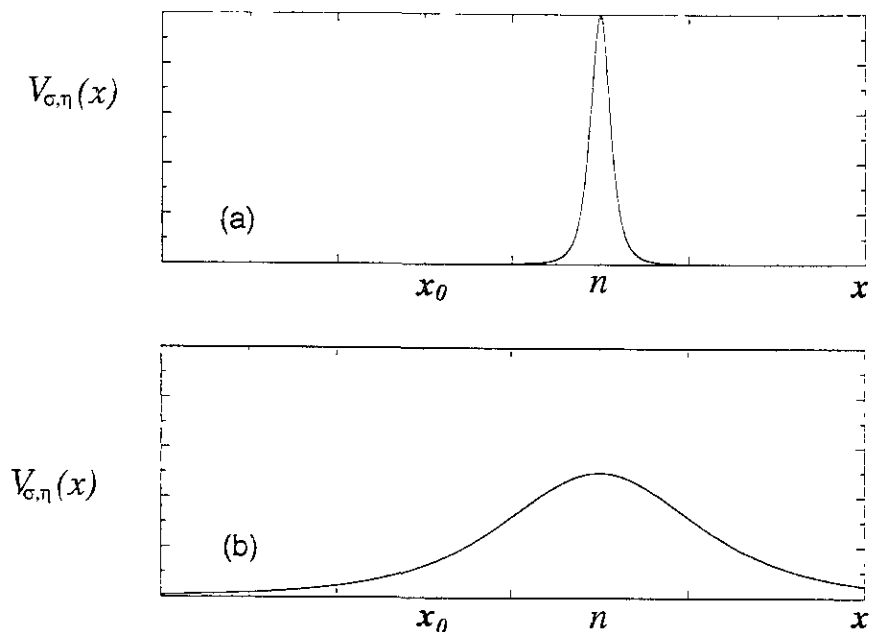


Figura 6.6: Gráfica del potencial individual  $V_{\sigma,n}^\alpha(x)$  generado por el  $n$ -ésimo monómero del polímero. (a) Si el alcance del potencial es muy corto, o si  $\sigma$  adquiere un valor muy pequeño, entonces la contribución de  $V_{\sigma,n}^\alpha(x)$  en la posición  $x = x_0$  es completamente despreciable, en tanto  $n$  esté lo suficientemente lejos de  $x_0$ . (b) Si aumentamos el alcance del potencial, o si  $\sigma$  adquiere valores más y más grandes, el potencial se “ensancha”, contribuyendo apreciablemente en la posición  $x_0$ .

el número de términos que contribuyen apreciablemente a la suma (6.4) sea infinito. Debemos observar que el tener un número infinito de monómeros en el polímero no necesariamente garantiza que el número de términos que contribuyen apreciablemente al potencial total también sea infinito. La situación se ilustra esquemáticamente en la Fig.6.6, donde se muestra la gráfica del potencial individual  $V_{\sigma,n}^\alpha(x)$  generado por el  $n$ -ésimo monómero del polímero, para dos valores diferentes del parámetro  $\sigma$ . Puede verse de la figura que si  $\sigma$  es muy pequeño, la contribución de  $V_{\sigma,n}^\alpha(x)$  puede hacerse completamente despreciable en la posición  $x_0$  simplemente tomando  $n$  lo suficientemente lejos de  $x_0$ . Cuando decimos “completamente despreciable” a lo que nos referimos es a que  $V_{\sigma,n}^\alpha(x)$  adquiere un valor muy pequeño en  $x_0$  comparado con el valor del potencial producido por el monómero que

se encuentra más próximo a  $x_0$ . Bajo estas circunstancias, la suma en la expresión (6.4) sólo contendrá un número pequeño de términos apreciablemente diferentes de cero, sin importar el tamaño del polímero.

Por otro lado, si  $\sigma$  es muy grande, el potencial  $V_{\sigma,n}^\alpha(x)$  se ensancha y su contribución en  $x_0$  es comparable a la que proviene del monómero más cercano a  $x_0$ . Lo anterior significa que cuando la partícula  $Q$  se encuentra muy cerca del polímero, sólo "siente" al monómero que está justo por debajo de ella, mientras que si  $Q$  está lo suficientemente lejos del polímero, entonces se origina una *interacción colectiva* entre la partícula y los monómeros, con más y más términos contribuyendo apreciablemente a la suma (6.4) cuando  $\sigma$  aumenta.

De las consideraciones anteriores es claro que, para poder aplicar el TLC, debemos requerir que el polímero esté formado por un número infinito de monómeros, y además, que  $\sigma$  adquiera un valor lo suficientemente grande como para que la suma en la expresión (6.4) contenga un número también muy grande de términos.

## 6.5 Condición de Lindeberg

En la sección anterior establecimos la condición de Lindeberg (6.8) para un conjunto de variables aleatorias  $\xi_1, \dots, \xi_N$ . Nuestro objetivo ahora es aplicar esta condición al potencial total  $V(x)$ . Para esto, notemos primero que, para una posición fija  $x$ , la varianza  $\delta_n^2(x)$  de la variable aleatoria  $V_{\sigma,n}^\alpha(x)$  está dada por

$$\begin{aligned} \delta_n^2(x) &= \int_{-\infty}^{\infty} [V_{\sigma,n}^\alpha(x)]^2 P_k(V_{\sigma,n}^\alpha(x)) dV_{\sigma,n}^\alpha(x) \\ &= \int_{-\infty}^{\infty} [a_n g_{\sigma,n}^\alpha(x)]^2 \frac{1}{g_{\sigma,n}^\alpha(x)} P\left(\frac{V_{\sigma,n}^\alpha(x)}{g_{\sigma,n}^\alpha(x)}\right) d[a_n g_{\sigma,n}^\alpha(x)] \\ &= [g_{\sigma,n}^\alpha(x)]^2 \int_{-\infty}^{\infty} a^2 P(a) da \\ &= \Delta^2 [g_{\sigma,n}^\alpha(x)]^2 \end{aligned}$$

donde  $\Delta^2$  es, como antes, el segundo momento de la distribución de probabilidad  $P(a)$ . Para llegar al resultado anterior se hizo uso de la ecuación (6.6), que da la función de probabilidad  $P_n(v)$  asociada al potencial individual  $V_{\sigma,n}^\alpha(x)$  en términos de la distribución  $P(a)$ . De lo anterior se sigue que la suma  $S_N^2$  de las varianzas  $\delta_n^2$  que aparece en la condición de Lindeberg, expresión (6.7), está dada por

$$S_N^2 = \Delta^2 \sum_{n=-N}^N [g_{\sigma,n}^\alpha(x)]^2$$

En el límite  $N \rightarrow \infty$  y para valores muy grandes de  $\sigma$ , la suma de la expresión anterior puede sustituirse por una integral, con lo que se obtiene

$$J_0^2 \equiv \lim_{N \rightarrow \infty} S_N^2 = \Delta^2 \int_{-\infty}^{\infty} [g_\sigma^\alpha(x)]^2 dx \quad (6.10)$$

Consecuentemente, la condición de Lindeberg (6.8), aplicada al potencial aleatorio  $V(x)$ , se transforma en

$$\lim_{\sigma \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{S_N^2} \sum_{n=-N}^N \int_{|v| > t S_N} v^2 P_n(v) dv = 0 \quad (6.11)$$

donde también se ha tomado el límite  $\sigma \rightarrow \infty$  de acuerdo con las consideraciones de la sección anterior, en donde se vaía que para tener un número muy grande de términos contribuyendo apreciablemente a la suma (6.4), es necesario que  $\sigma$  adquiriera valores muy grandes. Sustituyendo las ecuaciones (6.6) y (6.10) en la última ecuación, la condición de Lindeberg para potenciales en cuestión queda finalmente como

$$\lim_{\sigma \rightarrow \infty} \frac{1}{J_0^2} \int_{-\infty}^{\infty} dx [g_\sigma^\alpha(x)]^2 \int_{|a| > t J_0 / g_\sigma^\alpha(x)} a^2 P(a) da = 0 \quad (6.12)$$

donde otra vez se ha sustituido la suma que aparece en la ecuación (6.11) por una integral. La condición de Lindeberg significa es que la suma  $J_0^2$  (y por lo tanto también el potencial aleatorio  $V(x)$  dado en la ecuación (6.4)) es la suma de muchas contribuciones, cada una de ellas despreciable en comparación con el valor de la propia  $J_0^2$ . La condición de Lindeberg también impone tres restricciones a nuestros potenciales individuales  $V_{\sigma,n}^\alpha(x)$ , a saber:

- (i) la función  $g_\sigma^\alpha(x)$  debe ser cuadrado integrable;
- (ii)  $g_\sigma^\alpha(x)/J_0 \rightarrow 0$  cuando  $\sigma \rightarrow \infty$ ;
- (iii)  $P(a)$  debe tener un segundo momento bien definido.

Por esta razón se ha asumido que los potenciales  $V_{\sigma,n}^\alpha(x)$  son acotados o localizados, y que  $P(a)$  tiene segundo momento bien definido, o bien, es acotada en el sentido de que  $P(a) = 0$  si  $|a| > R$  (ver la propiedad (b)

del capítulo 4). Por otro lado, es claro que para potenciales rugosos, la condición de Lindeberg se aplicaría a las derivadas del potencial (es decir, a las fuerzas) o a las diferencias de potencial.

Si la condición de Lindeberg se satisface, el TLC garantiza que la distribución de probabilidad  $P(V(x))$  que caracteriza al potencial total  $V(x)$  será una distribución gaussiana. Es importante, por tanto, estudiar qué clase de potenciales satisfacen la condición de Lindeberg (6.12), ya que para este tipo de potenciales la caracterización estadística queda ya determinada por el TLC, lo cual facilita el análisis de los potenciales aleatorios.

## 6.6 Potenciales bien comportados

No es difícil mostrar que los potenciales definidos en las ecuaciones (6.2) y (6.3) satisfacen la condición de Lindeberg (6.12) en tanto  $\alpha > 1/2$ , mientras que si  $-1/2 \leq \alpha < 1/2$  dicha condición la satisfacen las derivadas del potencial. Más aún, es posible demostrar que la condición de Lindeberg también se cumple para una amplia gama de potenciales “apantallados” de la forma

$$V_{\sigma,n}^{\alpha\beta}(x) = a_n g_{\sigma}^{\alpha\beta}(x) \tag{6.13}$$

donde la función  $g_{\sigma}^{\alpha\beta}(x)$  está dada por

$$g_{\sigma}^{\alpha\beta}(x) = \frac{1}{[x^2 + \sigma^2]^{\alpha/2}} \exp \left[ -(\gamma/2) (x^2 + \sigma^2)^{\beta} \right] \tag{6.14}$$

siempre y cuando  $0 < \beta < 1$ , y para todo valor de  $\alpha$ . El factor exponencial

$$\exp \left[ -(\gamma/2) (x^2 + \sigma^2)^{\beta} \right]$$

permite la posibilidad de que el potencial entre la partícula de prueba y el polímero esté “apantallado”. En particular, si  $\beta = 1/2$ , la expresión (6.14) se convierte en un potencial de Yukawa, el cual normalmente se considera de corto alcance. No obstante, de lo anterior se obtiene que incluso el potencial de Yukawa tiene un alcance lo suficientemente largo como para satisfacer la condición de Lindeberg.

Por otro lado, si  $\beta \geq 1$  el alcance del potencial de interacción  $V_{\sigma,n}^{\alpha\beta}(x)$  se hace extremadamente pequeño, y bajo estas circunstancias el TLC ya no es válido. Lo anterior se debe a que en este caso la escala espacial en la cual los potenciales individuales  $V_{\sigma,n}^{\alpha\beta}(x)$  son apreciablemente diferentes de cero es tan pequeña, que para una posición fija  $x$ , la suma  $\sum_{n=-\infty}^{\infty} V_{\sigma,n}^{\alpha\beta}(x)$  nunca contiene un número infinito de términos apreciablemente diferentes de cero,

sino que sólo contiene un número pequeño de tales términos (aquellos que se generan por los monómeros que se encuentran en la vecindad inmediata de la partícula de prueba  $Q$  y que por lo tanto contribuyen apreciablemente al potencial total).

Las consideraciones precedentes hacen ver que la condición de Lindeberg puede no cumplirse ya sea porque los potenciales de interacción no decaen a cero con la distancia lo suficientemente rápido (o son crecientes con la distancia, como en el ejemplo de los osciladores armónicos), o bien porque tales potenciales decaen extremadamente rápido con la distancia y por lo tanto son de muy corto alcance, lo cual implica que para cualquier posición  $x$  dada, sólo un número finito de términos contribuyen al potencial total  $V(x)$ .

Para demostrar que los potenciales apantallados definidos en (6.13) y (6.14) efectivamente satisfacen la condición de Lindeberg, debemos demostrar que

$$\lim_{\sigma \rightarrow \infty} I(\sigma) = 0$$

donde la integral  $I(\sigma)$  está definida de acuerdo con la expresión (6.12):

$$I(\sigma) = J_0^{-2} \int_{-\infty}^{\infty} dx [g_\sigma(x)]^2 \int_{|a| > t J_0 / g_\sigma(x)} a^2 P(a) da \quad (6.15)$$

donde los índices  $\alpha$  y  $\beta$  de la función  $g_\sigma(x)$  se han eliminado para simplificar la notación. Para las funciones  $g_\sigma(x)$  definidas en la expresión (6.14),  $J_0^2$  está dada por

$$\begin{aligned} J_0^2 &= \Delta^2 \int_{-\infty}^{\infty} \left[ \frac{\exp [-(\gamma/2) (x^2 + \sigma^2)^\beta]}{[x^2 + \sigma^2]^{\alpha/2}} \right]^2 dx \\ &= \Delta^2 \sigma^{1-2\alpha} \int_{-\infty}^{\infty} \frac{\exp [-\gamma \sigma^{2\beta} (1+t^2)^\beta]}{[1+t^2]^\alpha} dt \end{aligned} \quad (6.16)$$

Si  $\sigma$  es muy grande, entonces la exponencial en la integral anterior se hace cero muy rápidamente, y la contribución importante a esta integral se obtiene dentro de un intervalo muy pequeño alrededor del origen. Por lo tanto, para valores grandes de  $\sigma$  se puede aproximar  $J_0^2$  como

$$J_0^2 \approx \Delta^2 \sigma^{1-2\alpha} \int_{-\epsilon}^{\epsilon} \frac{\exp [-\gamma \sigma^{2\beta} (1+t^2)^\beta]}{[1+t^2]^\alpha} dt$$



donde  $\varepsilon \ll 1$ . Dentro del intervalo  $[-\varepsilon, \varepsilon]$  puede desarrollarse en serie de Taylor el exponente  $(1+t^2)^\beta$ , manteniendo sólo los términos hasta el segundo orden. Por otro lado, en vista de que  $\varepsilon$  es muy pequeño, dentro del intervalo en consideración es posible tomar simplemente  $(1+t^2)^{-\alpha} = 1$ , lo cual conduce a <sup>2</sup>

$$\begin{aligned} J_0^2 &\approx \Delta^2 \sigma^{1-2\alpha} e^{-\gamma\sigma^{2\beta}} \int_{-\varepsilon}^{\varepsilon} e^{-\gamma\beta\sigma^2 t^2} dt \\ &\approx \Delta^2 \sqrt{\frac{\pi}{\gamma\beta}} \sigma^{1-\beta-2\alpha} e^{-\gamma\sigma^{2\beta}} \end{aligned} \tag{6.17}$$

Ahora debemos mostrar que  $\lim_{\sigma \rightarrow \infty} g_\sigma(x)/J_0 = 0$ , tal y como lo indica la condición (ii) de la sección anterior. Para esto, notemos que para  $\sigma$  muy grande, el cuadrado de la función  $g_\sigma(x)$  puede aproximarse por

$$[g_\sigma(x)]^2 \approx \sigma^{-2\alpha} \frac{\exp[-\gamma\sigma^{2\beta}(1 + \beta(x^2/\sigma^2) + \dots)]}{1 + \alpha(x^2/\sigma^2) + \dots}$$

De la expresión anterior y de (6.17) obtenemos que

$$\begin{aligned} \left[ \frac{g_\sigma(x)}{J_0} \right]^2 &\approx \frac{1}{\Delta^2} \sqrt{\frac{\gamma\beta}{\pi}} \sigma^{2\alpha+\beta-1} e^{\gamma\sigma^{2\beta}} \sigma^{2\alpha} \times \\ &\quad \times \frac{\exp[-\gamma\sigma^{2\beta}(1 + \beta(x^2/\sigma^2) + \dots)]}{1 + \alpha(x^2/\sigma^2) + \dots} \\ &\approx \frac{1}{\Delta^2} \sqrt{\frac{\gamma\beta}{\pi}} \sigma^{\beta-1} \frac{\exp[-\gamma\sigma^{2\beta}(\beta(x^2/\sigma^2) + \dots)]}{1 + \alpha(x^2/\sigma^2) + \dots} \end{aligned}$$

y es fácil ver que la expresión anterior se anula cuando  $\sigma \rightarrow \infty$  para cualquier valor de  $x$  y  $\beta \geq 0$ . Por lo tanto, el valor  $J_0/g_\sigma(x)$  se hace mayor que cualquier número real simplemente tomando  $\sigma$  lo suficientemente grande.

Por otro lado, estamos suponiendo que el segundo momento de  $P(a)$  está bien definido, por lo que  $P(a)$  debe caer a cero más rápido que  $1/|a|^3$  conforme  $|a| \rightarrow \infty$ . Por lo tanto, puede suponerse, sin pérdida de generalidad, que  $P(a) \sim 1/|a|^{3+\varepsilon}$  cuando  $|a| \rightarrow \infty$ , para algún número  $\varepsilon > 0$ . Lo

<sup>2</sup>No se puede hacer lo mismo con el factor  $(1+t^2)^\beta$ , es decir, no se puede tomar simplemente  $(1+t^2)^\beta = 1$ , ya que dicho factor se encuentra como exponente de la función exponencial, la cual es una función extremadamente creciente.

anterior representa, claro está, una cota superior para el comportamiento asintótico de  $P(a)^3$ .

Como se ha visto, si  $\sigma$  es muy grande entonces  $J_0/g_\sigma(x)$  también es muy grande, por lo que puede utilizarse el comportamiento asintótico  $1/|a|^{3+\varepsilon}$  de la función de probabilidad  $P(a)$  para calcular el valor de la integral  $I(\sigma)$ . lo que conduce a

$$\begin{aligned} I(\sigma) &\approx J_0^{-2} \int_{-\infty}^{\infty} dx [g_\sigma(x)]^2 \int_{|a| > t J_0/g_\sigma(x)} a^2 \frac{1}{|a|^{3+\varepsilon}} da \\ &= \frac{2t^{-\varepsilon}}{\varepsilon} \int_{-\infty}^{\infty} \left[ \frac{g_\sigma(x)}{J_0} \right]^{2+\varepsilon} dx \\ &= \frac{2t^{-\varepsilon}}{\varepsilon} J_0^{-(2+\varepsilon)} \int_{-\infty}^{\infty} [g_\sigma(x)]^{2+\varepsilon} dx \\ &= \frac{2t^{-\varepsilon}}{\varepsilon} J_0^{-(2+\varepsilon)} \sigma^{1-2\alpha'} \int_{-\infty}^{\infty} \frac{\exp[-\gamma' \sigma^{2\beta} (1+t^2)^\beta]}{[1+t^2]^{\alpha'}} dt \end{aligned}$$

donde  $\alpha' = \alpha(1 + \varepsilon/2)$  y  $\gamma' = \gamma(1 + \varepsilon/2)$ . Nótese que la integral que aparece en la última igualdad es exactamente la misma que la integral de la expresión (6.16), sólo que los parámetros  $\alpha$  y  $\gamma$  están reemplazados por  $\alpha'$  y  $\gamma'$ , respectivamente. Esta integral está dada en (6.17). (con  $\alpha$  y  $\gamma$  sustituidos por  $\alpha'$  y  $\gamma'$ ). Por lo tanto, la dependencia asintótica de  $I(\sigma)$  con  $\sigma$  queda como

$$I(\sigma) \approx \frac{2t^{-\varepsilon}}{\Delta^{2+\varepsilon} \varepsilon \sqrt{1 + \varepsilon/2}} \left( \frac{\gamma\beta}{\pi} \right)^{\varepsilon/4} \sigma^{(\beta-1)\varepsilon/2} \quad (6.18)$$

De la expresión anterior se desprende que, efectivamente,  $\lim_{\sigma \rightarrow \infty} I(\sigma) = 0$  siempre y cuando  $\beta < 1$ . Por lo tanto, si  $0 \leq \beta < 1$ , los potenciales apantallados definidos en (6.13) y (6.14) satisfacen la condición de Lindeberg, tal y como queríamos demostrar. La demostración de que los potenciales no apantallados definidos en las expresiones (6.2) y (6.3) también satisfacen dicha condición en tanto  $\alpha > 1/2$ , es completamente análoga a la anterior (aunque un poco más simple), por lo cual la omitimos aquí. Resumimos nuestros resultados en la siguiente tabla, que muestra el rango de los parámetros  $\alpha$  y  $\beta$  que aparecen en los potenciales apantallados para los

<sup>3</sup>Si  $P(a)$  tiene soporte compacto, es decir,  $P(a) = 0$  para  $|a| > R$ , tal y como lo establece la propiedad (b) del capítulo 4, entonces el análisis que estamos haciendo se cumple trivialmente. Al asumir que  $P(a) \sim 1/|a|^{3+\varepsilon}$  cuando  $|a| \rightarrow \infty$  estamos considerando el caso más general posible para el cual el segundo momento de la distribución  $P(a)$  está bien definido.

cuales la condición de Lindeberg se satisface. Obsérvese que los potenciales definidos en (6.2) y (6.3) son un caso particular de los potenciales apanallados definidos en (6.13) y (6.14) cuando  $\beta = 0$ . Por lo tanto, puede considerarse a éstos últimos (o a una combinación lineal de ellos) como la clase más general de potenciales de interacción que pueden generarse entre partículas cargadas.

$V(x) = \frac{1}{[x^2 + \sigma^2]^{\alpha/2}} \exp\left(-\frac{\gamma}{2}(x^2 + \sigma^2)^\beta\right)$	
Se satisface la condición de Lindeberg	
$0 < \beta < 1$ y cualquier $\alpha$	Potencial acotado
$\beta = 0$ y $\alpha > 1$	Potencial acotado
$\beta = 0$ y $1/2 < \alpha \leq 1$	Potencial localizado
$\beta = 0$ y $-1/2 < \alpha \leq 1/2$	Potencial rugoso
No se satisface la condición de Lindeberg	
$\beta > 1$ y cualquier $\alpha$	Potencial de muy corto alcance
$\beta = 0$ y $\alpha \leq -1/2$	Potencial no diferenciable

Ya que se ha demostrado que la condición de Lindeberg se satisface para la clase de potenciales con los que estamos trabajando, el TLC garantiza que la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2), \dots, V(x_r))$  que caracteriza al potencial aleatorio total  $V(x)$ , será una distribución gaussiana para valores de  $\sigma$  suficientemente grandes. Esto no quiere decir que la distribución conjunta no pueda ser gaussiana si la condición anterior no se satisface, debido a que dicha condición es suficiente pero no necesaria. De hecho, si la condición de Lindeberg no se cumple, el potencial aleatorio  $V(x)$  puede o no estar descrito por una distribución gaussiana. Sin embargo, en lo que resta de este trabajo supondremos que nuestros potenciales sí satisfacen la condición de Lindeberg, ya que como hemos visto, dicha condición se satisface por una gama muy amplia de potenciales (todos los potenciales electrostáticos producidos por partículas puntuales la satisfacen).

## 6.7 Función de probabilidad conjunta del potencial aleatorio

En esta sección presentamos un formalismo sencillo para calcular la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2), \dots, V(x_r))$  explícitamente.

Con el objetivo de hacer nuestro análisis lo más simple posible, enfocaremos nuestra atención al cálculo de la probabilidad conjunta solamente de dos puntos,  $\mathbf{P}(V(x_1), V(x_2))$ . Todos los resultados presentados pueden fácilmente generalizarse a más de dos puntos.

Comenzaremos suponiendo que el polímero está compuesto por un número finito de monómeros, digamos  $N$ . El elemento central de nuestro análisis es que el potencial total  $V(x)$  es una variable aleatoria que consiste en la suma de los potenciales individuales  $V_{\sigma,n}^\alpha(x)$  generados por los  $N$  monómeros del polímero, y que dichos potenciales individuales también son variables aleatorias independientes cuya función de probabilidad está dada por (6.6). La idea general es establecer una relación de recurrencia para la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2))$ , etiquetando a esta función con el número  $N$  de monómeros en el polímero, y después resolver la ecuación de recurrencia obtenida en el límite  $N \rightarrow \infty$  y para valores muy grandes de  $\sigma$ . Entonces, el TLC garantiza que la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2))$  así obtenida será una gaussiana, de la cual sólo se tendrán que determinar los parámetros que la caracterizan.

En lo que sigue, será conveniente distinguir entre el potencial aleatorio total generado por un polímero compuesto de  $N - 1$  monómeros, y el potencial total generado por un polímero compuesto de  $N$  monómeros. Por lo tanto, llamaremos  $V_{N-1}(x)$  y  $V_N(x)$  a los potenciales aleatorios generados por los polímeros compuestos por  $N - 1$  monómeros y  $N$  monómeros, respectivamente. Para comenzar, supongamos que el polímero consiste de  $N - 1$  monómeros cuyas cargas son  $a_1, a_2, \dots, a_{N-1}$ . Estas cargas generan el potencial total de interacción  $V_{N-1}(x)$  dado por

$$V_{N-1}(x) = \sum_{n=1}^{N-1} a_n g_{\sigma,n}^\alpha(x) \quad (6.19)$$

el cual está caracterizado por la función de probabilidad conjunta

$$\mathbf{P}_{N-1}(V_{N-1}(x_1), V_{N-1}(x_2))$$

Cuando añadimos un monómero más con carga  $a_N$  al final del polímero, como se muestra en la Fig.6.7, dicho monómero contribuye al potencial total con el término  $V_{\sigma,N}^\alpha(x) = a_N g_{\sigma,N}^\alpha(x)$ , de tal forma que ahora el potencial de interacción está dado por

$$\begin{aligned} V_N(x) &= V_{N-1}(x) + a_N g_{\sigma,N}^\alpha(x) \\ &= \sum_{n=1}^N a_n g_{\sigma,n}^\alpha(x) \end{aligned} \quad (6.20)$$

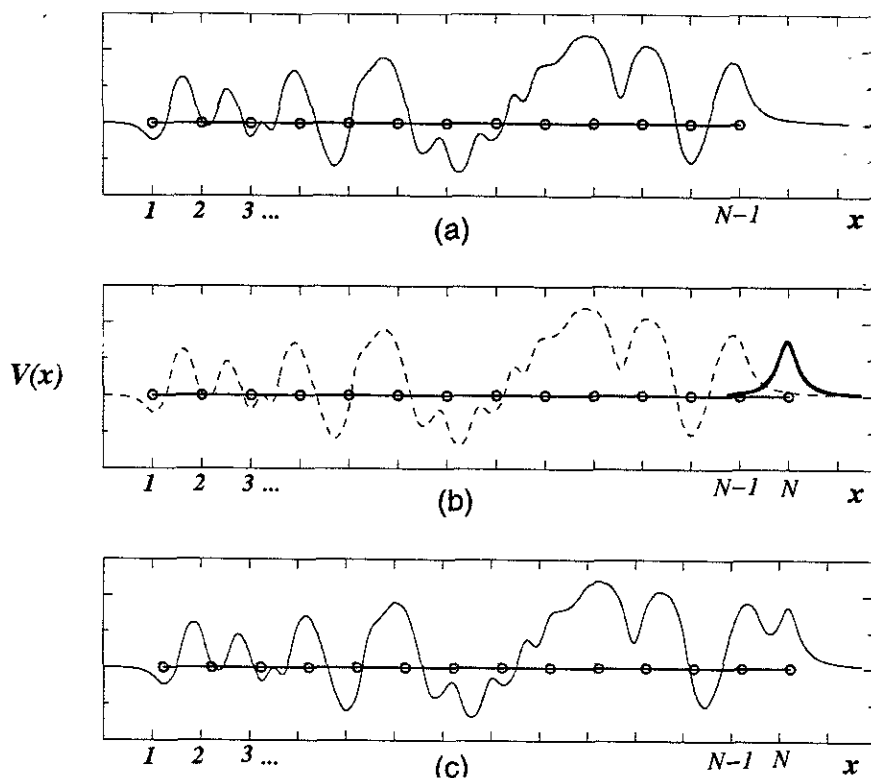


Figura 6.7: (a) Potencial generado por el polímero compuesto de  $N - 1$  monómeros. (b) Un nuevo monómero se añade en la posición  $x = N$ . Este nuevo monómero contribuye al potencial total con el término  $V_{\sigma,N}^{\alpha}(x)$  (curva sólida). (c) El potencial resultante después de la incorporación del  $N$ -ésimo monómero. Como puede observarse, la curva es diferente de la mostrada en (a); por lo tanto, la función de probabilidad también cambia.

El nuevo término añadido al polímero cambia la forma del potencial de interacción, de tal manera que ahora la probabilidad conjunta del potencial es

$$P_N(V_N(x_1), V_N(x_2))$$

Si el potencial  $V_N(x)$  adquiere los valores específicos  $V_N(x_1)$  y  $V_N(x_2)$  en las posiciones  $x_1$  y  $x_2$ , respectivamente, entonces el potencial  $V_{N-1}(x)$  *debió haber adquirido* los valores

$$V_{N-1}(x_1) = V_N(x_1) - aNg_{\sigma,N}^{\alpha}(x_1)$$

y

$$V_{N-1}(x_2) = V_N(x_2) - a_N g_{\sigma, N}^\alpha(x_2)$$

en las correspondientes posiciones  $x_1$  y  $x_2$ . La probabilidad para que el potencial  $V_{N-1}(x)$  adquiriera los valores antes mencionados en  $x_1$  y  $x_2$  está dada por la función de probabilidad conjunta  $\mathbf{P}_{N-1}(V_{N-1}(x_1), V_{N-1}(x_2))$ , que en este caso toma la forma

$$\begin{aligned} \mathbf{P}_{N-1}(V_{N-1}(x_1), V_{N-1}(x_2)) &= \\ &= \mathbf{P}_{N-1}(V_N(x_1) - a_N g_{\sigma, N}^\alpha(x_1), V_N(x_2) - a_N g_{\sigma, N}^\alpha(x_2)) \end{aligned} \quad (6.21)$$

Se puede obtener  $\mathbf{P}_N(V_N(x_1), V_N(x_2))$  a partir de la ecuación anterior multiplicando dicha ecuación por  $P(a_N)$  y después sumando sobre todos los valores posibles de  $a_N$  que hacen que el potencial  $V_N(x)$  adquiriera los valores particulares  $V_N(x_1)$  y  $V_N(x_2)$  en las posiciones  $x_1$  y  $x_2$ , respectivamente. Tomando en cuenta que  $P(a)$  es la misma para todas las cargas  $a_n$ 's en el polímero, la relación entre  $\mathbf{P}_N(V_N(x_1), V_N(x_2))$  y  $\mathbf{P}_{N-1}(V_{N-1}(x_1), V_{N-1}(x_2))$  está dada entonces por

$$\begin{aligned} \mathbf{P}_N(V_N(x_1), V_N(x_2)) &= \int_{-\infty}^{\infty} \mathbf{P}_{N-1}(V_{N-1}(x_1), V_{N-1}(x_2)) P(a) da = \\ &= \int_{-\infty}^{\infty} \mathbf{P}_{N-1}(V_N(x_1) - a g_{\sigma, N}^\alpha(x_1), V_N(x_2) - a g_{\sigma, N}^\alpha(x_2)) P(a) da \end{aligned} \quad (6.22)$$

Esta última ecuación puede resolverse más fácilmente en el espacio de Fourier, porque allí la integral de convolución se transforma en un simple producto. Por lo tanto, si llamamos  $\xi_1$  y  $\xi_2$  a las variables conjugadas (en el espacio de Fourier) de las variables aleatorias  $V_N(x_1)$  y  $V_N(x_2)$  respectivamente, tomando la transformada de Fourier de la ecuación (6.22) obtenemos la siguiente expresión equivalente

$$\widehat{\mathbf{P}}_N(\xi_1, \xi_2) = \widehat{\mathbf{P}}_{N-1}(\xi_1, \xi_2) \widehat{P}(g_{\sigma, N}^\alpha(x_1)\xi_1 + g_{\sigma, N}^\alpha(x_2)\xi_2) \quad (6.23)$$

Para resolver esta ecuación de recurrencia necesitamos conocer  $\widehat{\mathbf{P}}_1(\xi_1, \xi_2)$ , es decir, necesitamos conocer la función de probabilidad conjunta para el caso en el que hay sólo un monómero en el polímero. En tal caso, con sólo un monómero en el polímero colocado en  $x = 1$ , el potencial de interacción total  $V(x)$  está dado simplemente por

$$V(x) = V_1(x) = a_1 g_{\sigma,1}^\alpha(x) \quad (6.24)$$

Por lo tanto, el potencial total  $V(x)$  tendrá el valor  $V_1(x_1)$  en la posición  $x_1$  si y sólo si la variable aleatoria  $a_1$  adquirió el valor

$$a_1 = \frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)} \quad (6.25)$$

La probabilidad de que  $a_1$  tome este valor en particular está dada por  $P(a_1) = P(V_1(x_1)/g_{\sigma,1}^\alpha(x_1))$ , y la probabilidad de que el potencial total adquiera el valor  $V_1(x_1)$  está dada por la ecuación (6.6) con  $n = 1$  y  $v = V_1(x_1)$ :

$$P_1(V_1(x_1)) = \frac{1}{g_{\sigma,1}^\alpha(x_1)} P\left(\frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)}\right) \quad (6.26)$$

Una vez que el potencial total  $V(x)$  ha adquirido el valor  $V_1(x_1)$  en la posición  $x = x_1$ , su valor en otra posición diferente  $x = x_2$  está restringido, y deberá ser

$$V_1(x_2) = a_1 g_{\sigma,1}^\alpha(x_2) = \frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)} g_{\sigma,1}^\alpha(x_2) \quad (6.27)$$

y la probabilidad condicional  $P(V_1(x_2) | V_1(x_1))$  de que esto ocurra, es decir, la probabilidad condicional de que el potencial  $V(x)$  adquiera el valor  $V_1(x_2)$  en la posición  $x_2$  dado que ha adquirido el valor  $V_1(x_1)$  en la posición  $x_1$ , está dada por

$$P(V_1(x_2) | V_1(x_1)) = \delta\left(V_1(x_2) - \frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)} g_{\sigma,1}^\alpha(x_2)\right) \quad (6.28)$$

donde  $\delta(\cdot)$  es la función delta de Dirac. De las ecuaciones (6.26) y (6.28) se desprende que la probabilidad conjunta  $\mathbf{P}_1(V_1(x_1), V_1(x_2))$  es

$$\mathbf{P}_1(V_1(x_1), V_1(x_2)) = \frac{1}{g_{\sigma,1}^\alpha(x_1)} P\left(\frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)}\right) \delta\left(V_1(x_2) - \frac{V_1(x_1)}{g_{\sigma,1}^\alpha(x_1)} g_{\sigma,1}^\alpha(x_2)\right) \quad (6.29)$$

Tomando la transformada de Fourier de la expresión anterior obtenemos

$$\hat{\mathbf{P}}_1(\xi_1, \xi_2) = \hat{P}(g_{\sigma,1}^\alpha(x_1)\xi_1 + g_{\sigma,1}^\alpha(x_2)\xi_2) \quad (6.30)$$

Esta es la condición inicial de la ecuación de recurrencia (6.23), cuya solución es entonces

$$\widehat{\mathbf{P}}_N(\xi_1, \xi_2) = \prod_{n=1}^N \widehat{P}(g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2) \quad (6.31)$$

Por cuestiones de simetría y de claridad en lo que sigue, enumeraremos a los monómeros del polímero desde  $-m$  hasta  $m$ , en lugar de hacerlo de 1 a  $N$ , donde claramente  $N = 2m + 1$ . Después de todo, vamos a tomar el límite  $N \rightarrow \infty$ , lo que es equivalente a tomar el límite  $m \rightarrow \infty$ . Los límites anteriores significan que cada vez tomamos en cuenta la contribución de más y más monómeros en el polímero, y cuando éste llegue a ser infinito, no importa como enumeremos a los monómeros. Con esta "renumeración" de los monómeros, la expresión (6.31) queda como

$$\widehat{\mathbf{P}}_m(\xi_1, \xi_2) = \prod_{n=-m}^m \widehat{P}(g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2) \quad (6.32)$$

Recordemos que estamos suponiendo que las funciones  $g_{\sigma,n}^\alpha(x)$  se anulan cuando  $\sigma$  se hace cada vez más grande (ver las expresiones (6.3) y (6.14)). Por lo tanto, si  $\sigma$  es muy grande, el argumento  $\lambda = g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2$  de la función  $\widehat{P}(\lambda)$  en la ecuación (6.32) es muy pequeño, y se puede desarrollar esta función en serie de potencias de su argumento  $\lambda$ , manteniendo los términos hasta el segundo orden:

$$\begin{aligned} \widehat{P}(g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2) &\simeq 1 - i\mu [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2] \\ &\quad - \frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2 + \dots \end{aligned} \quad (6.33)$$

donde  $\mu$  y  $\Delta^2$  son el primer y segundo momento de la función de probabilidad  $P(a)$ , respectivamente:  $\mu = \langle a \rangle$  y  $\Delta^2 = \langle a^2 \rangle$ . Como estamos suponiendo que  $P(a)$  tiene promedio cero (propiedad (a) en el capítulo 4), entonces  $\mu = 0$ , por lo que la ecuación (6.33) queda como

$$\widehat{P}(g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2) \simeq 1 - \frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2 + \dots$$

Sustituyendo este valor de  $\widehat{P}(g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2)$  en la ecuación (6.32) obtenemos

$$\widehat{\mathbf{P}}_m(\xi_1, \xi_2) \simeq \prod_{n=-m}^m \left( 1 - \frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2 \right) \quad (6.34)$$



En este punto vale la pena enfatizar que estamos trabajando bajo la hipótesis de que el teorema de límite central es válido, debido a que nuestros potenciales satisfacen la condición de Lindeberg. Por lo tanto, aun cuando tenemos un desarrollo sólo hasta el segundo orden en la ecuación (6.34), el TLC garantiza que el término  $1 - \frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2$  que aparece en dicha ecuación necesariamente proviene del desarrollo en serie de una distribución gaussiana, es decir, de una distribución de la forma

$$\exp\left(-\frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2\right)$$

Por lo tanto, de acuerdo con el TLC, la ecuación (6.34) puede escribirse como

$$\begin{aligned} \widehat{\mathbf{P}}_m(\xi_1, \xi_2) &\simeq \prod_{n=-m}^m \exp\left(-\frac{1}{2}\Delta^2 [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2\right) \\ &= \exp\left(-\frac{1}{2}\Delta^2 \sum_{n=-m}^m [g_{\sigma,n}^\alpha(x_1)\xi_1 + g_{\sigma,n}^\alpha(x_2)\xi_2]^2\right) \end{aligned} \quad (6.35)$$

La expresión anterior es claramente la función característica de una distribución gaussiana, como esperábamos. En el límite  $m \rightarrow \infty$  y para  $\sigma$  muy grande, la suma en la última ecuación puede sustituirse por una integral, con lo cual dicha función característica adquiere la siguiente forma final

$$\widehat{\mathbf{P}}(\xi_1, \xi_2) = \exp\left(-\frac{1}{2} [J_0^2 \xi_1^2 + J_0^2 \xi_2^2 + 2J_{1,2}^2 \xi_1 \xi_2]\right) \quad (6.36)$$

donde  $J_0^2$  y  $J_{1,2}^2$  están definidas como

$$J_0^2 = \Delta^2 \int_{-\infty}^{\infty} [g_\sigma^\alpha(x)]^2 dx \quad (6.37)$$

$$J_{1,2}^2 = \Delta^2 \int_{-\infty}^{\infty} g_\sigma^\alpha(z - x_1) g_\sigma^\alpha(z - x_2) dz \quad (6.38)$$

Tomando la transformada de Fourier inversa de la ecuación (6.36) obtenemos la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2))$  que queríamos encontrar:

$$P(V(x_1), V(x_2)) = \frac{1}{2\pi\sqrt{J_0^4 - J_{1,2}^4}} \times \\ \times \exp \left[ -\frac{1}{2} \frac{J_0^2 V(x_1)^2 + J_0^2 V(x_2)^2 - 2J_{1,2}^2 V(x_1)V(x_2)}{J_0^4 - J_{1,2}^4} \right] \quad (6.39)$$

El resultado anterior muestra explícitamente que el proceso que estamos considerando es un proceso gaussiano caracterizado por  $J_0^2$  y por  $J_{1,2}^2$ . De hecho,  $J_{1,2}^2$  es la función de correlación del potencial aleatorio, la cual claramente sólo es función de la distancia de separación  $|x_1 - x_2|$  entre los puntos  $x_1$  y  $x_2$ . La correlación a distancias grandes del potencial puede obtenerse tomando la transformada de Fourier de la ecuación (6.38) y analizando después su comportamiento para números de onda pequeños. Debido a que  $J_{1,2}^2$  está dada a través de una convolución, su transformada de Fourier es muy fácil de obtener:

$$\widehat{J_{1,2}^2}(\lambda) = \Delta^2 [\widehat{g_\sigma^\alpha}(\lambda)]^2 \quad (6.40)$$

El comportamiento detallado dependerá, por supuesto, de la naturaleza de la interacción, pero en términos generales, puede decirse que para los potenciales acotados o localizados  $\widehat{J_{1,2}^2}(\lambda) \rightarrow const$  cuando  $\lambda \rightarrow 0$ . Por otro lado, para potenciales rugosos,  $\widehat{J_{1,2}^2}(\lambda) \sim \lambda^{-\nu}$  para  $\lambda$  pequeña, donde el exponente  $\nu$  es una medida de la "rugosidad" del potencial. Ya se ha visto antes que para potenciales rugosos es más conveniente trabajar con las derivadas del potencial o con las diferencias de potencial, en lugar de trabajar con el potencial en sí. No es difícil demostrar que para los potenciales  $V_{\sigma,n}^\alpha(x)$  definidos en las expresiones (6.2) y (6.3), el valor del exponente  $\nu$  está dado por  $\nu = 1 - 2\alpha$  cuando  $-1/2 < \alpha \leq 1/2$ , y por lo tanto el comportamiento asintótico del potencial aleatorio total  $V(x)$  para distancias grandes está dado por

$$\langle [V(x_1) - V(x_2)]^2 \rangle \rightarrow \begin{cases} f(\alpha) |x_1 - x_2|^{1-2\alpha} & \alpha \leq 1/2 \\ const & \alpha > 1/2 \end{cases}$$

cuando  $|x_1 - x_2| \rightarrow \infty$ , donde  $f(\alpha)$  es una función que se anula en  $\alpha = 0$ .

## 6.8 Dinámica e invariancia de escala

En esta sección ilustramos como los resultados obtenidos en la sección anterior pueden utilizarse para determinar algunas de las propiedades estadísticas de la dinámica de la partícula moviéndose a lo largo de un

polímero compuesto por  $N = 2m + 1$  monómeros ( $m \rightarrow \infty$ ). Para esto, retomaremos el modelo presentado en el capítulo 5, en el que una fuerza externa  $F$  actúa sobre la partícula para forzar al sistema, impidiendo que se quede atrapado en alguno de los mínimos del potencial aleatorio<sup>4</sup>. Restringiremos nuestro análisis al caso de *fuerzas acotadas*, el cual es análogo al caso de potenciales acotados discutido antes en este mismo capítulo, solo que ahora trabajaremos con fuerzas en lugar de potenciales. En tales circunstancias, siempre es posible escoger el valor de la fuerza  $F$  lo suficientemente grande como para evitar que la partícula se quede atrapada en un mínimo de potencial.

En principio, la fuerza  $F$  puede depender del tiempo, pero en nuestro análisis consideraremos otra vez que  $F$  es una fuerza constante. La dinámica de la partícula en el régimen sobre amortiguado está descrita, como antes, por la ecuación

$$\gamma v = -\frac{\partial V(x)}{\partial x} + F$$

o tomando en cuenta que el potencial aleatorio  $V(x)$  está dado por la expresión (6.4), la ecuación de movimiento es entonces

$$\gamma v = -\frac{\partial}{\partial x} \left[ \sum_{n=-m}^m a_n g_\sigma^\alpha(x-n) \right] + F \quad (6.41)$$

La Fig6.8a muestra la gráfica de la velocidad de la partícula como función de la posición a lo largo del polímero, utilizando los valores de los parámetros  $\alpha = 4$  y  $\sigma = 10$ . Por otro lado, la Fig6.8b muestra el espectro de potencias de la velocidad anterior calculado numéricamente. Como puede verse de la figura, el espectro de potencias muestra una “panza” sobresaliente, lo cual indica que existen ciertas regularidades estadísticas en la dinámica de la partícula, a pesar de la aleatoriedad del potencial de interacción. Como veremos a continuación, esta “panza” está relacionada directamente con las correlaciones de largo alcance del potencial aleatorio. Para ver que efectivamente esto es así, tomamos la transformada de Fourier de la ecuación (6.41). Los pasos a seguir son exactamente los mismos que ya hicimos en la última sección del capítulo 5. De hecho, la ecuación (6.41) que vamos a transformar es exactamente la misma que la ecuación (5.17) del capítulo anterior, sólo cambia la forma de la función  $g_\sigma^\alpha(x)$ . Por lo tanto, el valor promedio del espectro de potencias de la velocidad está dado, como antes, por la siguiente expresión:

<sup>4</sup>Debido a que vamos a considerar un polímero infinito, en esta sección supondremos que es la partícula la que se mueve, y no el polímero.

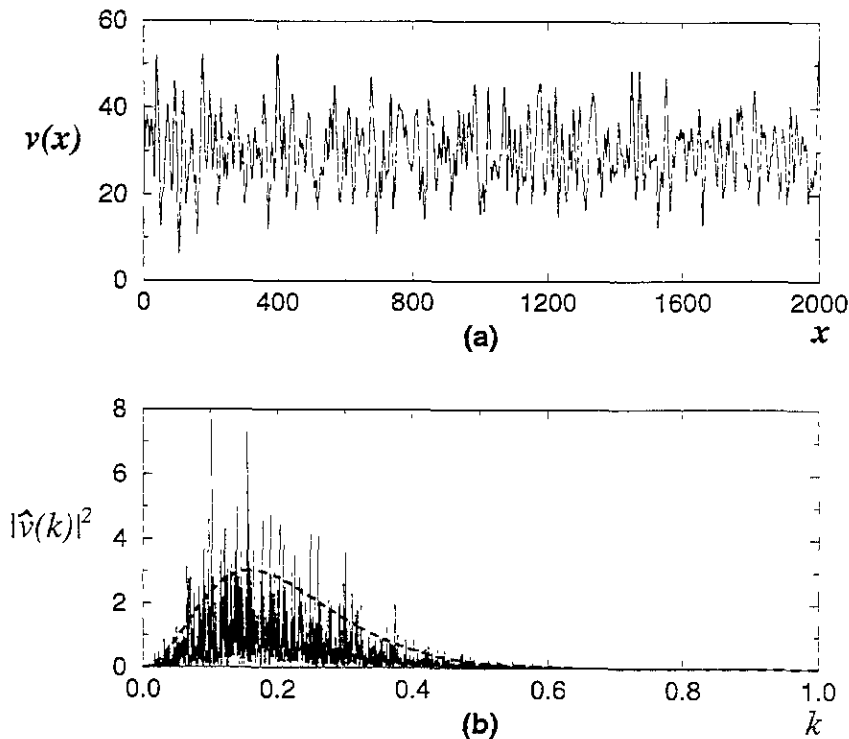


Figura 6.8: (a) Velocidad de la partícula a lo largo del polímero, como función de la posición. (b) Espectro de potencias de la gráfica anterior. La curva punteada es el valor promedio del espectro de potencias,  $\langle |\hat{v}(k)|^2 \rangle$ , dado en la ecuación (6.44).

$$\langle |\hat{v}(k)|^2 \rangle = \frac{N\Delta^2}{\gamma} k^2 [\hat{g}_\sigma^\alpha(k)]^2 \quad (6.42)$$

o tomando en cuenta la ecuación (6.40), obtenemos

$$\langle |\hat{v}(k)|^2 \rangle = \frac{N}{\gamma} k^2 \widehat{J}_{1,2}^2(k) \quad (6.43)$$

La ecuación anterior claramente muestra que las regularidades estadísticas en la dinámica de la partícula, exhibidas en el espectro de potencias de la velocidad, son una consecuencia directa de las correlaciones de largo alcance del potencial, dadas por la función  $\widehat{J}_{1,2}^2(k)$ . Para la clase de potenciales

acotados  $V_{\sigma,n}^{\alpha}(x)$  definidos en (6.2) y (6.3), con  $\alpha > 1/2$ , el espectro de potencias de la velocidad está dado por

$$\langle |\hat{v}(k)|^2 \rangle = \frac{N8\pi\Delta^2}{[\Gamma(\alpha/2)]^2 2^{\alpha}\sigma^{2\alpha}} \left[ (\sigma k)^{(\alpha+1)/2} K_{(\alpha-1)/2}(\sigma k) \right]^2 \quad (6.44)$$

donde  $K_{\mu}(z)$  es la función de Bessel de segunda especie de orden  $\mu$ , y  $\Gamma(\alpha)$  es la función gama. En la Fig6.8b se muestra la gráfica de la ecuación anterior, y puede verse que dicha gráfica es la envolvente, o la suavización, del espectro de potencias calculado numéricamente. Nótese también de la ecuación anterior que el espectro de potencias es, de hecho, función de la variable escalada  $z = \sigma k$ . Multiplicar en el espacio de Fourier el número de onda  $k$  por  $\sigma$  es equivalente, en el espacio real, a tomar la distancia perpendicular  $\sigma$  entre el polímero y la partícula como nuestra unidad de longitud, en lugar de la distancia  $L$  que hay entre los monómeros. Lo que la ecuación (6.44) nos está diciendo es que si en el espacio real dividimos todas las longitudes espaciales entre  $\sigma$ , y si reescalamos el tamaño de las velocidades por el factor  $\sigma^{2\alpha}$ , entonces la estructura estadística de la velocidad de la partícula a lo largo del polímero será exactamente la misma para todas las escalas de longitud. En otras palabras, *la estructura estadística de la dinámica del sistema es invariante ante transformaciones de escala.*

Lo anterior puede verse claramente en la Fig6.9, que muestra la gráfica de la velocidad de la partícula para diferentes valores del parámetro  $\sigma$ . Para generar estas gráficas utilizamos  $\alpha = 4$  y la distribución de cargas  $P(a) = (1/6) \sum_{m=1}^3 \delta(a \pm m)$ . Como puede observarse de la figura, si  $\sigma$  es lo suficientemente grande, la estructura estadística de la velocidad no cambia cuando  $\sigma$  adquiere valores cada vez más grandes. Por supuesto, esta invariancia de escala también está presente en el potencial de interacción  $V(x)$ , como bien puede observarse en la Fig6.5, y puede tener consecuencias importantes. Recordemos que una partícula alejada del polímero es equivalente a un objeto extendido interactuando con el mismo polímero. Por lo tanto, dicha invariancia de escala significa entonces que el objeto extendido siempre "siente" el mismo tipo de potencial, independientemente de su tamaño. Lo anterior puede ser la explicación del hecho de que la fuerza de fricción entre dos superficies en contacto no depende, dentro de ciertos límites, del tamaño de las superficies en cuestión.

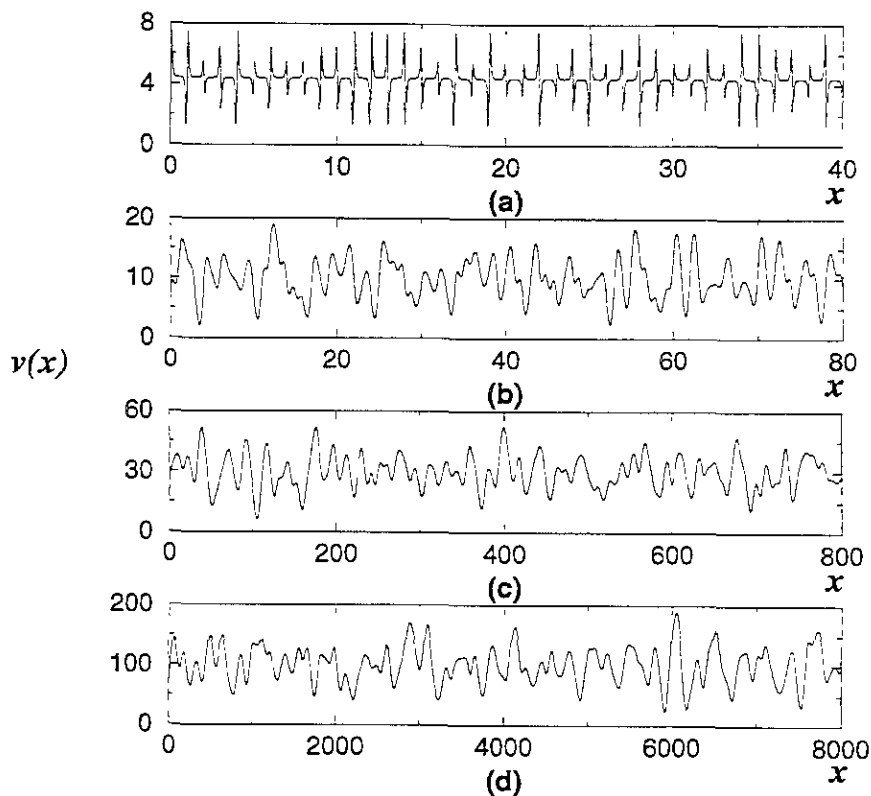


Figura 6.9: Velocidad de la partícula a lo largo del polímero para diferentes valores de  $\sigma$  y  $\alpha = 4$ . (a)  $\sigma = 0.1$ , (b)  $\sigma = 1$ , (c)  $\sigma = 10$  y (d)  $\sigma = 100$ . Nótese como la estructura estadística de la velocidad ya no cambia para valores grandes de  $\sigma$ .

## 6.9 Distancia promedio entre mínimos de energía

Debido a que los potenciales aleatorios con los que hemos estado trabajando son procesos gaussianos, puede utilizarse el método de Rice [67] para encontrar una relación explícita entre algunas de las características estadísticas de los potenciales y las propiedades dinámicas del movimiento de la partícula a lo largo del polímero. Lo que haremos entonces es encon-

trar una expresión para la distancia promedio  $\bar{l}$  entre mínimos consecutivos en el potencial.

Debemos notar que los máximos y mínimos del potencial se caracterizan por tener derivada cero. Por lo tanto, encontrar la distancia promedio entre mínimos consecutivos en el potencial es equivalente a encontrar la distancia promedio entre ceros consecutivos en la derivada del potencial. Por supuesto, los puntos silla del potencial también tienen derivada cero, pero supondremos que el número de dichos puntos silla es despreciable comparado con el número de máximos y mínimos en el potencial. Por lo tanto, lo que haremos es encontrar una expresión que nos de la distancia promedio entre ceros consecutivos de alguna función aleatoria, y después aplicaremos los resultados obtenidos al caso particular de los potenciales con los que estamos trabajando.

Comenzaremos nuestra deducción de la distancia promedio  $\bar{l}$  de la siguiente manera. Supongamos que tenemos una función  $y$  dada por

$$y = F(a_1, a_2, \dots, a_N; x) \quad (6.45)$$

donde las variables  $\{a_n\}$  son variables aleatorias. Para un conjunto de valores particulares de las variables  $\{a_n\}$ , la ecuación anterior da una curva  $y(x)$  cuando graficamos a  $y$  como función de  $x$ . Debido a que las variables  $\{a_n\}$  son aleatorias, la curva  $y(x)$  también será aleatoria (como las gráficas de nuestros potenciales). Supongamos además que conocemos la función de probabilidad  $\mathbf{P}(\xi, \eta; x)$ , donde las variables  $\xi$  y  $\eta$  son

$$\begin{aligned} \xi &= F(a_1, a_2, \dots, a_N; x) \\ \eta &= \frac{\partial}{\partial x} F(a_1, a_2, \dots, a_N; x) \end{aligned}$$

Seleccionemos ahora un pequeño intervalo  $[x_1, x_1 + dx]$  en el eje  $x$  y luego asignemos los valores aleatorios a las variables  $\{a_n\}$ . Queremos calcular la probabilidad de que la curva  $y(x)$  corte al eje  $x$  en algún punto del intervalo  $[x_1, x_1 + dx]$ , para lo cual consideraremos primero el caso en el que dicho corte se hace con pendiente positiva, como se muestra en la Fig6.10. Escogemos  $dx$  lo suficientemente pequeño de tal manera que todas las posibles curvas  $y(x)$  que pasen por este intervalo puedan considerarse como líneas rectas. Llamemos  $x_0$  al punto del intervalo  $[x_1, x_1 + dx]$  en el cual la curva  $y(x)$  corta al eje  $x$ . Si dicha curva adquiere el valor  $\xi$  en  $x_1$ , es decir, si  $\xi = y(x_1)$ , entonces  $x_0$  está dado por

$$x_0 = x_1 - \frac{\xi}{\eta} \quad (6.46)$$

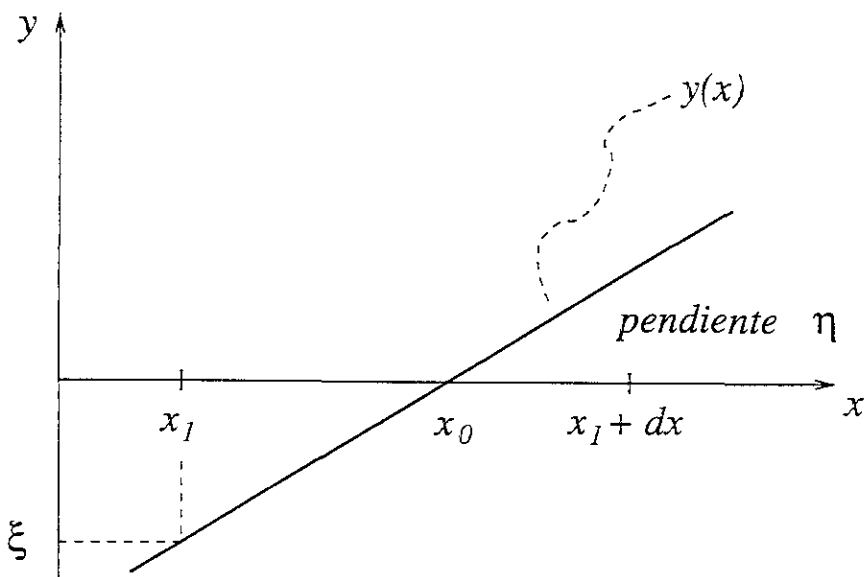


Figura 6.10: Diagrama para calcular la probabilidad de que la curva  $y(x)$  corte al eje  $x$  en el intervalo  $[x_1, x_1 + dx]$ . En este caso la curva corta el eje con pendiente  $\eta$  positiva.

donde  $\eta$  es la pendiente de la curva  $y(x)$  en el intervalo en consideración. Es claro que  $\xi$  y  $\eta$  deben tener signos opuestos. Como el punto  $x_0$  está dentro del intervalo  $[x_1, x_1 + dx]$ , se debe cumplir que

$$x_1 \leq x_1 - \frac{\xi}{\eta} \leq x_1 + dx$$

lo cual puede escribirse como

$$-\eta dx \leq \xi \leq 0 \quad (6.47)$$

Para una curva  $y(x)$  dada, es decir, para un conjunto de valores dado a las variables aleatorias  $\{a_n\}$ ,  $\xi$  y  $\eta$  están determinadas por

$$\begin{aligned} \xi &= F(a_1, a_2, \dots, a_N; x_1) \\ \eta &= \left. \frac{\partial}{\partial x} F(a_1, a_2, \dots, a_N; x) \right|_{x=x_1} \end{aligned}$$



Si estos valores de  $\xi$  y  $\eta$  satisfacen la desigualdad (6.47), entonces la curva  $y(x)$  corta al eje  $x$  con pendiente positiva en el intervalo  $[x_1, x_1 + dx]$ . La probabilidad de que esto ocurra está dada por

$$\int_0^\infty d\eta \int_{-\eta dx}^0 \mathbf{P}(\xi, \eta; x_1) d\xi$$

Como  $dx$  es muy pequeño, la función  $\mathbf{P}(\xi, \eta; x_1)$ , puede aproximarse en el intervalo de integración  $\xi \in [-\eta dx, 0]$ , simplemente por  $\mathbf{P}(\xi, \eta; x_1) \approx \mathbf{P}(0, \eta; x_1)$ , con lo que la integral anterior se transforma en

$$\int_0^\infty d\eta \int_{-\eta dx}^0 \mathbf{P}(\xi, \eta; x_1) d\xi \approx dx \int_0^\infty \eta \mathbf{P}(0, \eta; x_1) d\eta$$

La última expresión da la probabilidad de que la curva  $y(x)$  corte al eje  $x$  en el intervalo  $[x_1, x_1 + dx]$ , haciéndolo con pendiente positiva. Razonomientos similares conducen a que la probabilidad de que dicho corte se haga con pendiente negativa, está dada por

$$-dx \int_{-\infty}^0 \eta \mathbf{P}(0, \eta; x_1) d\eta$$

Por lo tanto, la probabilidad de que  $y(x)$  corte al eje  $x$  en el intervalo  $[x_1, x_1 + dx]$ , independientemente de si el corte se lleva a cabo con pendiente negativa o positiva, es

$$dx \int_{-\infty}^\infty |\eta| \mathbf{P}(0, \eta; x_1) d\eta$$

Esta última expresión da la probabilidad de que la función  $y(x)$  tenga un cero en un intervalo pequeño de longitud  $dx$ . Si integramos dicha expresión sobre un intervalo de longitud 1 a lo largo del eje  $x$ , claramente obtenemos el número esperado de ceros de la curva  $y(x)$  por unidad de longitud, lo cual es precisamente la densidad de ceros de la curva  $y(x)$ . Por lo tanto, si llamamos  $\rho_0$  a esta densidad de ceros de la curva  $y(x)$ , entonces obtenemos

$$\rho_0 = \int_x^{x+1} dx \int_{-\infty}^\infty |\eta| \mathbf{P}(0, \eta; x) d\eta \quad (6.48)$$

Para conectar los resultados anteriores con el problema que planteamos al inicio de esta sección, que es el de determinar la distancia promedio entre mínimos consecutivos en el potencial aleatorio  $V(x)$ , vamos a tomar a la curva  $y(x)$  como la derivada espacial del potencial  $V(x)$ . Esto es porque los máximos y mínimos del potencial están determinados por los ceros de

la derivada  $\partial V(x)/\partial x$  (despreciando los puntos silla). Por lo tanto, con la notación anterior haremos

$$\begin{aligned}\xi &= \frac{\partial V(x)}{\partial x} \\ \eta &= \frac{\partial^2 V(x)}{\partial x^2}\end{aligned}$$

Necesitamos ahora la función  $\mathbf{P}(\xi, \eta; x)$ . Debido a que los potenciales que estamos considerando son procesos gaussianos, entonces la función  $\mathbf{P}(\xi, \eta; x)$  es muy similar a la que ya calculamos en la expresión (6.39), sólo que tenemos que sustituir en dicha expresión a las variables aleatorias  $V(x_1)$  y  $V(x_2)$  por  $\xi$  y  $\eta$ , respectivamente. Además, los parámetros  $J_0^2$  y  $J_{1,2}^2$  también cambian. Haciendo exactamente lo mismo que ya hicimos en la sección 6.7 para encontrar la función de probabilidad conjunta  $\mathbf{P}(V(x_1), V(x_2))$ , es fácil demostrar que en este caso la función de probabilidad  $\mathbf{P}(\xi, \eta)$  está dada por

$$\mathbf{P}(\xi, \eta) = \frac{1}{2\pi\sqrt{ab-c^2}} \exp\left\{-\frac{1}{2} \frac{b\xi^2 + a\eta^2 - 2c\xi\eta}{ab-c^2}\right\} \quad (6.49)$$

donde los parámetros  $a$ ,  $b$  y  $c$  están definidos como

$$a = \Delta^2 \int_{-\infty}^{\infty} \left[ \frac{\partial g_\sigma^\alpha(x)}{\partial x} \right]^2 dx \quad (6.50)$$

$$b = \Delta^2 \int_{-\infty}^{\infty} \left[ \frac{\partial^2 g_\sigma^\alpha(x)}{\partial x^2} \right]^2 dx \quad (6.51)$$

$$c = \Delta^2 \int_{-\infty}^{\infty} \left( \frac{\partial g_\sigma^\alpha(x)}{\partial x} \right) \left( \frac{\partial^2 g_\sigma^\alpha(x)}{\partial x^2} \right) dx \quad (6.52)$$

Como estamos suponiendo que tanto las funciones  $g_\sigma^\alpha(x)$  como todas sus derivadas se anulan para  $x \rightarrow \pm\infty$ , entonces integrando por partes la expresión (6.52) obtenemos inmediatamente que  $c = 0$ . Por lo tanto, la probabilidad  $\mathbf{P}(\xi, \eta)$  queda como

$$\mathbf{P}(\xi, \eta) = \frac{1}{2\pi\sqrt{ab}} \exp\left\{-\frac{1}{2} \left( \frac{\xi^2}{a} + \frac{\eta^2}{b} \right)\right\} \quad (6.53)$$

Combinando la expresión anterior con la ecuación (6.48), obtenemos la densidad de ceros  $\rho_0$ :

$$\begin{aligned}\rho_0 &= \int_x^{x+1} dx \int_{-\infty}^{\infty} |\eta| \frac{1}{2\pi\sqrt{ab}} e^{-\frac{\eta^2}{2b}} d\eta \\ &= \frac{1}{\pi} \sqrt{\frac{b}{a}}\end{aligned}$$

El resultado anterior no depende en lo absoluto del punto  $x$ . Esto se debe a que los potenciales con los que estamos trabajando presentan invariancia traslacional, es decir, presentan la misma estructura estadística a lo largo de todo el polímero. Tomando ahora en cuenta las expresiones (6.50) y (6.51) se obtiene

$$\rho_0 = \frac{1}{\pi} \left( \frac{\int_{-\infty}^{\infty} [\partial^2 g_\sigma^\alpha(x)/\partial x^2]^2 dx}{\int_{-\infty}^{\infty} [\partial g_\sigma^\alpha(x)/\partial x]^2 dx} \right)^{1/2} \quad (6.54)$$

Ahora bien, el teorema de Parseval dice que para cualquier función  $f(x)$ , cuya transformada de Fourier es  $\hat{f}(k)$ , se cumple que

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\hat{f}(k)|^2 dk$$

Por lo tanto, la ecuación (6.54) que da la densidad de mínimos  $\rho_0$  se transforma en

$$\rho_0 = \frac{1}{\pi} \left( \frac{\int_{-\infty}^{\infty} [k^2 \hat{g}_\sigma^\alpha(k)]^2 dk}{\int_{-\infty}^{\infty} [k \hat{g}_\sigma^\alpha(k)]^2 dk} \right)^{1/2} \quad (6.55)$$

Recordemos que el promedio del espectro de potencias de la velocidad,  $\langle |\hat{v}(k)|^2 \rangle$ , precisamente es igual a

$C [k \hat{g}_\sigma^\alpha(k)]^2$ , donde  $C$  es una constante (ecuación (6.42)). Por lo tanto, la ecuación (6.55) puede escribirse en términos de  $\langle |\hat{v}(k)|^2 \rangle$ , obteniendo finalmente que

$$\rho_0 = \frac{1}{\pi} \left( \frac{\int_0^\infty k^2 \langle |\hat{v}(k)|^2 \rangle dk}{\int_0^\infty \langle |\hat{v}(k)|^2 \rangle dk} \right)^{1/2} \quad (6.56)$$

La ecuación (6.56) determina la densidad de ceros en la *derivada de la potencial*. El inverso de esta densidad,  $1/\rho_0$ , da la distancia promedio entre ceros consecutivos en la gráfica de dicha derivada, lo cual es equivalente a la

distancia promedio entre máximos y mínimos consecutivos en el potencial mismo. Como puede verse en todas las gráficas de potenciales aleatorios que se han presentado en este trabajo, los máximos del potencial son completamente equivalentes a los mínimos, y viceversa (no se puede distinguir si las gráficas están de "cabeza" o si están de "pie" a menos que miremos las escalas en los ejes). Por lo tanto, la distancia promedio entre mínimos consecutivos del potencial es el doble de la distancia promedio entre máximos y mínimos. De aquí que si llamamos  $\bar{l}$  a la distancia promedio entre mínimos consecutivos del potencial, se tiene

$$\bar{l} = \frac{2}{\rho_0}$$

lo cual, tomando en cuenta la ecuación (6.56), conduce finalmente a

$$\bar{l} = 2\pi \left( \frac{\int_0^\infty \langle |\hat{v}(k)|^2 \rangle dk}{\int_0^\infty k^2 \langle |\hat{v}(k)|^2 \rangle dk} \right)^{1/2} \quad (6.57)$$

Para la clase de potenciales  $V_{\sigma,n}^\alpha(x)$  definidos en (6.2) y (6.3), el espectro de potencias  $\langle |\hat{v}(k)|^2 \rangle$  está dado en la ecuación (6.44), por lo que para esta clase de potenciales se tiene

$$\bar{l}(\alpha) = 2\pi\sigma h(\alpha) \quad (6.58)$$

donde la función  $h(\alpha)$  está definida como

$$h(\alpha) = \left( \frac{\int_0^\infty [z^{(\alpha+1)/2} K_{(\alpha-1)/2}(z)]^2 dz}{\int_0^\infty [z^{(\alpha+3)/2} K_{(\alpha-1)/2}(z)]^2 dz} \right)^{1/2} \quad (6.59)$$

Obsérvese que  $h(\alpha)$  es una función que sólo depende de  $\alpha$ , es decir, sólo depende del tipo de interacción bajo consideración. Por esta razón se ha escrito explícitamente en la ecuación (6.58) la dependencia de  $\bar{l}$  con  $\alpha$ . La Fig6.11 muestra la gráfica de  $\bar{l}(\alpha)/\sigma$  como función de  $\alpha$ . Haciendo un ajuste de mínimos cuadrados a la curva mostrada en esta figura<sup>5</sup>, obtenemos

$$\frac{\bar{l}(\alpha)}{\sigma} \simeq 5.13\alpha^{-1/2} \quad (6.60)$$

<sup>5</sup>Aun cuando tenemos la expresión exacta para  $\bar{l}$  en (6.58) y (6.59), las funciones de Bessel no son fáciles de integrar analíticamente. Por lo tanto, la gráfica de la Fig6.11 la obtuvimos numéricamente. así como la relación (6.60).

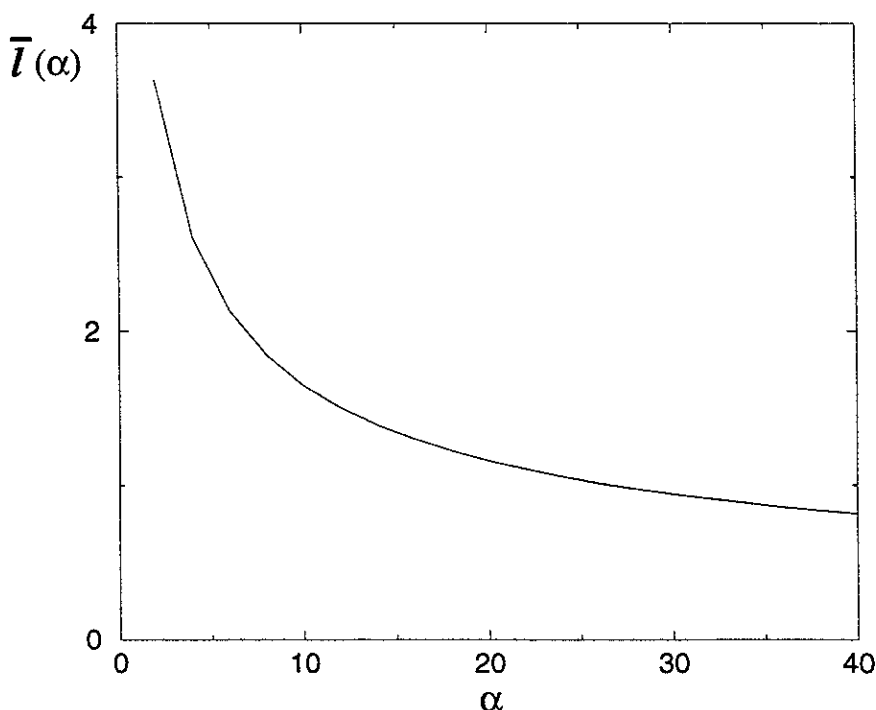


Figura 6.11: Distancia promedio  $\bar{l}(\alpha)$  (en unidades de  $\sigma$ ) entre mínimos consecutivos del potencial, como función de  $\alpha$ .

Vale la pena señalar que  $\bar{l}/\sigma$  es una cantidad *invariante de escala* que sólo depende del tipo de interacción entre los monómeros y la partícula de prueba, determinado por el valor de  $\alpha$ . Lo anterior es cierto siempre y cuando  $\sigma$  sea lo suficientemente grande como para que la aproximación gaussiana en la caracterización de los potenciales sea válida. Notemos, por cierto que si  $2 \leq \alpha \leq 5$ , lo cual abarca un rango amplio de las interacciones que se dan en la naturaleza, entonces la distancia escalada  $\bar{l}/\sigma$  está comprendida entre 2.3 y 3.6, es decir, vuelve a aparecer una vez más que la distancia promedio entre mínimos consecutivos en el potencial se encuentra alrededor de 3.



# Conclusiones

En esta sección discutiremos los resultados obtenidos a lo largo de este trabajo y plantearemos también algunos problemas que quedan por resolver.

El estudio realizado se centró en analizar algunas de las propiedades estáticas y dinámicas de máquinas moleculares primitivas. Para esto, se propuso un modelo de máquina molecular en el cual una partícula cargada se mueve a lo largo de un polímero cuyos monómeros están cargados aleatoriamente. La aleatoriedad en las cargas de los monómeros trae como consecuencia que el potencial de interacción entre la partícula y el polímero también sea aleatorio. Las clase de potenciales de interacción considerados representa, genéricamente, al tipo de potenciales electrostáticos que típicamente ocurren entre partículas cargadas.

Los resultados expuestos en los capítulos 4 y 5 muestran que, aún ante la aleatoriedad de los potenciales de interacción, existen regularidades estadísticas en la estructura espacial de tales potenciales, que se traducen en regularidades dinámicas del movimiento de la partícula. Por una parte, cuando  $\sigma$ , la distancia perpendicular entre la partícula y el polímero, adquiere valores pequeños, la distancia promedio  $\bar{d}$  entre mínimos consecutivos del potencial de interacción está comprendida siempre entre 3 y 4, dependiendo del número  $m$  de monómeros diferentes que conformen al polímero. Para el caso particular  $m = 4$ , la distancia promedio resulta ser  $\bar{d} \sim 3.4$ . Bien podría haber ocurrido de otra manera, es decir, que la distancia promedio estuviera cerca de 10 o de 20, pero no fue así. Está cerca de tres. Además, este resultado es independiente del parámetro  $\alpha$ , que en nuestro modelo determina el alcance de los potenciales de interacción. Lo anterior se debe a que, en el caso  $\sigma \rightarrow 0$ , la distribución de máximos y mínimos del potencial aleatorio está determinada, fundamentalmente, por la distribución de cargas a lo largo del polímero.

Aún cuando la distancia promedio  $\bar{d}$  entre mínimos consecutivos de los potenciales aleatorios está cerca de tres, al analizar la distribución de estas distancias se encontró que el valor más probable ocurre en  $d = 2$  cuando  $m > 2$ , es decir, cuando hay más de dos tipos diferentes de

monómeros (para  $m = 2$  la distancia más probable es  $d = 3$ ). Este resultado podría deberse a que estamos considerando *interacciones locales* entre el polímero y la partícula, al suponer que esta última es puntual. Al sustituir a la partícula por una cadena extendida, compuesta por varias partículas rígidamente unidas y cargadas aleatoriamente, la distribución de distancias entre mínimos consecutivos en el potencial cambia, de tal manera que la distancia más probable para este caso ocurre, generalmente, en  $d = 3$ . Lo anterior sugiere que el movimiento de la cadena se dará de acuerdo con el promedio de las interacciones locales.

Por otro lado, al analizar la dinámica generada por los potenciales aleatorios para el caso de  $\sigma$  pequeña, se encontró que el espectro de potencias de la velocidad de la partícula a lo largo del polímero tiene, generalmente, un pico dominante cerca del periodo 3. Este resultado tampoco depende del valor del parámetro  $\alpha$ , y es aproximadamente independiente de los valores particulares de carga asignados a los monómeros, siempre y cuando dichos valores no estén disparatadamente desproporcionados. También se encontró esta periodicidad cercana a tres en el espectro de potencias de la velocidad de la cadena extendida compuesta por varias partículas.

Es importante resaltar la enorme regularidad que se presenta en la dinámica del sistema cuando la asignación de las cargas a lo largo del polímero se hace en correspondencia con secuencias genéticas *codificadoras* de organismos reales. En esta situación la magnitud del pico en el espectro de potencias de la velocidad de la partícula aumenta considerablemente respecto de lo que se obtiene con la asignación aleatoria de cargas. Además, la periodicidad mostrada por el espectro de potencias se define casi exactamente en  $d = 3$  (las "fluctuaciones" alrededor de  $d = 3$  son extremadamente pequeñas). Adicionalmente, ocurre que para las secuencias genéticas codificadoras, la distancia más probable entre mínimos consecutivos en los potenciales de interacción es  $d = 3$  y no  $d = 2$ , como el en caso aleatorio.

Lo anterior no sucede cuando la asignación de las cargas en el polímero se hace utilizando secuencias intergénicas de organismos reales, ya que en este caso, tanto la distribución de distancias como los espectros de potencias de la velocidad presentan una estructura semejante a la que se tiene con la asignación aleatoria de cargas.

Con base en estos resultados, es posible proponer un escenario prebiótico para explicar la estructura de tripletes del código genético, ya que sugieren que los procesos prebióticos y protibióticos de codificación y traducción de la información genética, se llevaron a cabo tomando como base una dinámica en la que se preferenciaba el movimiento en "pasos de tres en tres" en las máquinas moleculares precursoras de los complejos actuales de síntesis de proteínas (ribosoma-mRNA). Durante el transcurso de la evolu-



ción, esta "preferencia" se refinó en las secuencias genéticas directamente involucradas en los procesos de codificación y lectura del mensaje genético, lo cual no ocurrió en las secuencias intergénicas (que no participan en dichos procesos). Así, la tesis fundamental que proponemos en relación al código genético, es que la preferencia dinámica a la que nos hemos referido, y su posterior refinamiento evolutivo, conllevan a la estructura de tripletes del código.

Vele la pena recalcar que la naturaleza del problema que estudiamos, ha permitido obtener resultados analíticos exactos en la caracterización estadística de los potenciales aleatorios. En los capítulos 4 y 5 se realizó dicho análisis en el *límite microscópico*  $\sigma \rightarrow 0$ . En este límite, la distribución estadística de los potenciales aleatorios está determinada por la distribución de las cargas en el polímero. El problema con potenciales aleatorios continuos se convierte entonces en un problema con variables aleatorias discretas, en el cual no existen correlaciones de largo alcance en los potenciales de interacción.

La situación cambia drásticamente en el *límite macroscópico*  $\sigma \rightarrow \infty$ . Este límite se obtiene cuando consideramos la interacción del polímero con un objeto extendido continuo. En tales circunstancias, la interacción ya no es local, sino que se convierte en una interacción colectiva, en la cual muchas partes del polímero están interactuando con el objeto extendido. El problema aquí es más difícil de tratar que en el límite opuesto  $\sigma \rightarrow 0$ , debido a que las correlaciones de largo alcance de los potenciales de interacción se vuelven relevantes.

En el capítulo 6 se mostró con un ejemplo numérico que un objeto extendido continuo de longitud  $L_o$  interactuando con el polímero, es equivalente a una partícula puntual interactuando con el mismo polímero, pero colocada a una distancia perpendicular  $\sigma_p$  lo suficientemente grande. El potencial de interacción en ambos casos es casi el mismo, una vez que se escala la amplitud de los potenciales adecuadamente. Claramente, el valor de  $\sigma_p$  es función del tamaño  $L_o$  del objeto extendido. Sin embargo, la dependencia funcional entre estas dos variables aún está por determinarse.

El análisis estadístico de los potenciales de interacción presentado en el capítulo 6 muestra que en el límite macroscópico  $\sigma \rightarrow \infty$ , dichos potenciales son procesos gaussianos correlacionados, siendo la función de correlación directamente proporcional al espectro de potencias de la velocidad de la partícula a lo largo del polímero. Así, la dinámica del sistema queda determinada por las correlaciones del potencial aleatorio. Este comportamiento, que es una consecuencia directa del Teorema de Límite Central, es muy general y se cumple para una amplia gama de potenciales. Se demostró que la mayoría de los potenciales electrostáticos que típicamente ocurren en la naturaleza, satisfacen la condición de Lindeberg, la cual establece la validez

del Teorema de Límite Central. Dicha condición requiere que el potencial individual generado por cada monómero sea de cuadrado integrable, y que la función de probabilidad  $P(q)$  que determina la distribución de cargas en los monómeros, tenga segundo momento bien definido.

El cálculo analítico realizado en el capítulo 6 mostró que solamente los potenciales crecientes con la distancia, o los que son de extremadamente corto alcance, no satisfacen la condición de Lindeberg.

La generalidad de nuestros resultados hacen ver que no es posible asignar de forma arbitraria propiedades estadísticas a los potenciales de interacción aleatorios: en el límite  $\sigma \rightarrow 0$ , tales propiedades quedan determinadas por la distribución de cargas a lo largo del polímero, mientras que en el límite opuesto  $\sigma \rightarrow \infty$ , el Teorema de Límite Central restringe a dichos potenciales a ser procesos gaussianos correlacionados. En medio de estos dos extremos, los potenciales son correlacionados, pero no necesariamente obedecen a una estadística gaussiana. Por lo tanto, un potencial aleatorio " $\delta$ -correlacionado"<sup>6</sup> como los que usualmente se utilizan para el modelamiento de transporte a través de sólidos desordenados, no sería físicamente realizable.

En la aproximación gaussiana, es decir, cuando  $\sigma$  es muy grande, fue posible calcular con relativa facilidad la distancia promedio entre mínimos consecutivos del potencial aleatorio. Resultó que dicha distancia es una cantidad invariante de escala cuyo valor sólo depende del parámetro  $\alpha$  que establece el tipo de interacción particular que se esté considerando. Esta propiedad de invariancia de escala se cumple en general para la estructura estadística del potencial aleatorio, como lo sugiere el hecho de que la función de correlación también sea invariante de escala, y trae como consecuencia que la estructura del potencial de interacción que "siente" el objeto extendido, es independiente del tamaño de dicho objeto<sup>7</sup>.

Un problema importante que aún queda por resolver en el límite macroscópico  $\sigma \rightarrow \infty$ , es el encontrar la distribución de distancias entre mínimos y máximos en los potenciales aleatorios gaussianos. El interés en este problema no es puramente matemático: como se ha mostrado en este trabajo, la dinámica de sistemas desordenados como los que se han considerado aquí, está determinada, en gran medida, por la estructura de máximos y mínimos del potencial.

Cabe mencionar que el trabajo desarrollado representa una caracterización exhaustiva de la estructura estadística de los potenciales aleatorios. Dicha caracterización se ha realizado tanto para el caso en el que las co-

<sup>6</sup>Los potenciales  $\delta$ -correlacionados son aquellos para los cuales la función de correlación está dada por  $\langle V(x_1)V(x_2) \rangle = \delta(x_1 - x_2)$ .

<sup>7</sup>Probablemente esta sea la razón de que la fuerza de fricción entre dos superficies en contacto, no dependa del área de dichas superficies

relaciones de largo alcance de los potenciales son despreciables ( $\sigma \rightarrow 0$ ), como para el caso en el que dichas correlaciones juegan el papel fundamental ( $\sigma \rightarrow \infty$ ). Este tipo de análisis puede ser de interés en otras áreas de la Física teórica y experimental.

En lo referente al código genético, consideramos que la relevancia de nuestro trabajo radica en que señala hacia una dirección que no se había considerado antes, a saber, la importancia de los aspectos dinámicos de las máquinas moleculares primitivas en las propiedades del código genético. Tales aspectos aportan diversos elementos que apoyan un escenario primitivo en el que las moléculas precursoras del ribosoma y del mRNA se movían una a lo largo de la otra en pasos cuya longitud promedio era de tres monómeros. El escenario propuesto parece ser factible, y es compatible con el conocimiento actual que se tiene sobre los procesos prebióticos y protobióticos en condiciones extraterrestres. No obstante, inmediatamente surge la pregunta de qué ocurre cuando los sistemas bajo consideración se encuentran sujetos a un ruido térmico de fondo, o bien, para ponerlo en términos de escenarios de origen de la vida, qué ocurre cuando se pasa de temperaturas bajas ( $T \sim 10$  K), propias del escenario extraterrestre, a condiciones de temperatura alta (escenario terrestre) donde las fluctuaciones térmicas ya no son despreciables. Queda también por determinarse con mayor claridad la naturaleza de las fuerzas externas  $F_{ex}$  que operaban sobre las máquinas moleculares primitivas, y dilucidar cómo pudo darse la transición de los mecanismos de locomoción primitivos, a las bombas químicas actuales.

El modelo que se ha presentado es uno de los más simples que se pueden construir para describir máquinas moleculares (incluso las primitivas). Una mejor aproximación consistiría, por ejemplo, en suponer que las partículas de la cadena extendida no se encuentran unidas rígidamente, sino que se unen por medio de resortes hookeanos, confiriéndole cierta flexibilidad a la cadena y dando lugar a una mayor riqueza dinámica, pudiéndose presentar fenómenos de conmensurabilidad y fenómenos de "stick-lip". El problema anterior representa una continuación natural del modelo que hemos estudiado.

Sin embargo, aún ante la simplicidad de nuestro modelo, consideramos que los resultados obtenidos aportan nuevos elementos para entender la estructura del código genético, y nos muestran cómo, en algunas ocasiones, los "datos" con los que juega la naturaleza están "cargados" en lo referente a los procesos de organización de la materia.



## Apéndice A

# Probabilidad de polimerización

Consideremos una caja en la que se encuentran moléculas monofuncionales (M) y bifuncionales (B) en proporciones  $p$  y  $q$  respectivamente, con  $p+q = 1$ . El objetivo es calcular la probabilidad de que se forme un polímero de  $L$  moléculas, quedando bloqueado en sus extremos por moléculas monofuncionales, tal y como se muestre en la FigA.1. Para tal efecto, es útil imaginar que el proceso de polimerización se lleva a cabo extrayendo moléculas aleatoriamente de la caja, una por una, y uniéndolas después para formar el polímero. La polimerización continuará siempre y cuando en el proceso no se hayan extraído las dos moléculas M que bloquean la polimerización. Evidentemente, cuando se extraiga la segunda de estas moléculas la polimerización se detendrá. Llamemos  $M_1$  a la molécula monofuncional que aparece primero, y  $M_2$  a la que aparece hasta el final del proceso.

La probabilidad de que en una realización particular de  $L$  intentos se extraigan 2 moléculas M y  $L - 2$  moléculas B, sin importar el orden en que aparezcan, es  $p^2 q^{L-2}$ . Sin embargo, de todas las realizaciones posibles, sólo interesan aquellas en las que  $M_2$  aparece hasta el final, en el intento  $L$ . La primer molécula  $M_1$  pudo haber aparecido en el primer intento, o en el segundo, o en el tercero, o incluso en el intento  $L - 1$ . Por lo tanto, de todas las realizaciones posibles de extracción de las  $L$  moléculas, sólo hay  $L - 1$  realizaciones en las cuales  $M_2$  aparece hasta el final, justo en el intento  $L$ , mientras que  $M_1$  apareció en cualquiera de los  $L - 1$  primer intentos. La probabilidad  $P(L)$  de que el polímero tenga una longitud de  $L$  moléculas cuando queda bloqueado, está dada entonces por

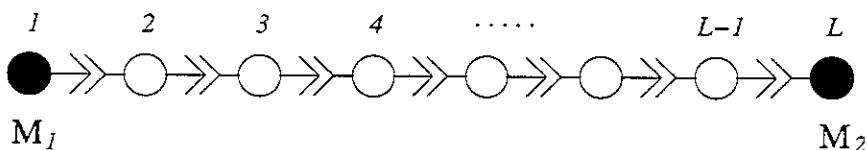
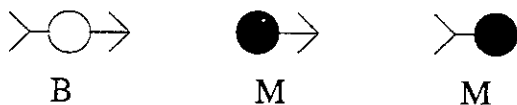


Figura A.1: En la parte superior de la figura se muestra esquemáticamente la estructura de moléculas bifuncionales (B) y monofuncionales (M). En la parte inferior se muestra un polímero compuesto por  $L$  moléculas. Los extremos del polímero están bloqueados por las moléculas monofuncionales  $M_1$  y  $M_2$ , por lo que el proceso de polimerización ya no puede proseguir.

$$P(L) = (L - 1)p^2q^{L-2} \quad (\text{A.1})$$

Notemos que esta probabilidad está normalizada a 1:

$$\sum_{L=2}^{\infty} P(L) = 1 \quad (\text{A.2})$$

lo cual significa que el polímero *debe* tener alguna longitud. La longitud promedio  $\bar{L}$  de los polímeros formados está dada por (nótese que la longitud mínima de los polímeros es  $L = 2$ )

$$\begin{aligned}
 \bar{L} &= \sum_{L=2}^{\infty} LP(L) \\
 &= \sum_{L=2}^{\infty} L(L-1)p^2q^{L-2} \\
 &= p^2 \frac{d^2}{dq^2} \sum_{L=0}^{\infty} q^L
 \end{aligned}$$

$$\begin{aligned}
 &= p^2 \frac{d^2}{dq^2} \left( \frac{1}{1-q} \right) \\
 &= p^2 \frac{2}{(1-q)^3}
 \end{aligned}$$

pero debido a que  $p + q = 1$ , la ecuación anterior se convierte en

$$\bar{L} = \frac{2}{1-q} \quad (\text{A.3})$$

Por otro lado, la probabilidad  $P[L > N]$  de que la longitud del polímero sea *mayor* que  $N$  está dada por

$$P[L > N] = \sum_{L=n+1}^{\infty} P(L)$$

lo que, tomando en cuenta la ecuación de normalización (A.2), puede escribirse como

$$\begin{aligned}
 P[L > N] &= 1 - \sum_{L=2}^N P(L) \\
 &= 1 - \sum_{L=2}^N (L-1)p^2 q^{L-2} \\
 &= 1 - p^2 \frac{d}{dq} \sum_{L=1}^{N-1} q^L
 \end{aligned}$$

Al realizar la suma geométrica y la derivada en la expresión anterior, y tomando en cuenta otra vez que  $p + q = 1$ , se obtiene finalmente

$$P[L > N] = q^N \left( 1 + N \frac{p}{q} \right)$$

que es la ecuación (3.2) del capítulo 3.





## Apéndice B

# Distribución de distancias para el caso $m = \infty$

Sin pérdida de generalidad, supongamos que los potenciales discretos  $v_n$  generados por los monómeros en el polímero, son variables aleatorias continuas uniformemente distribuidas en el intervalo  $[0, 1]$ . Esta suposición es válida porque ahora cada variable  $v_n$  puede adquirir una infinidad de valores con la misma probabilidad. Debido a lo anterior, no es necesario considerar la presencia de mínimos planos en los potenciales. Tales mínimos ocurren con una probabilidad despreciable.

Sea  $P(d)$  la probabilidad de que dos mínimos consecutivos en el potencial estén separados por una distancia  $d$ . Supongamos que se tiene un mínimo en  $v_1$  y que el siguiente mínimo ocurre en  $v_{d+1}$ , de tal manera que la distancia entre estos dos mínimos es  $d$ , como se muestra en la FigB.1. Sea  $P[d \geq l]$  la probabilidad condicional de que el mínimo en  $v_{d+1}$  ocurra a una distancia  $d \geq l$ , dado que se tiene un mínimo en  $v_1$ . La probabilidad  $P(l)$  de que los dos mínimos estén separados justamente por la distancia  $d = l$  está dada entonces por

$$P(l) = P_0 \{P[d \geq l] - P[d \geq l + 1]\} \quad (\text{B.1})$$

donde  $P_0$  es la probabilidad de que  $v_1$  sea un mínimo. Para calcular  $P_0$  notemos que  $v_1$  será un mínimo si y sólo si los valores de los potenciales  $v_0$ ,  $v_1$  y  $v_2$  satisfacen la desigualdad

$$v_0 > v_1 < v_2$$

Debido a que las variables  $v_n$  están distribuidas uniformemente en el

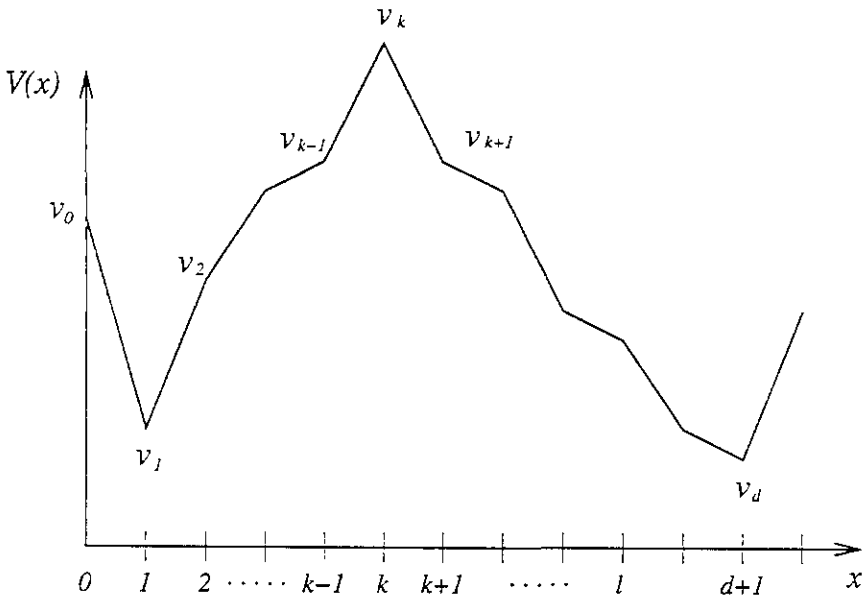


Figura B.1: Dos mínimos consecutivos a lo largo del potencial de interacción discreto, separados por una distancia  $d$ . Se supone que el potencial en cada sitio es una variable aleatoria que puede tomar valores uniformemente distribuidos entre 0 y 1.

intervalo  $[0, 1]$ , la probabilidad de que se satisfaga la desigualdad anterior está dada por la siguiente integral:

$$\begin{aligned} P_0 &= \int_0^1 dv_0 \int_0^{v_0} dv_1 \int_{v_1}^1 dv_2 \\ &= 3. \end{aligned}$$

Por otro lado, dado que se tiene un mínimo en  $v_1$ , el siguiente mínimo caerá a una distancia mayor o igual que  $l$  siempre y cuando los valores de los potenciales  $v_0, v_1, \dots, v_l, v_{l+1}$  satisfagan la desigualdad

$$v_0 > v_1 < v_2 < \dots < v_{k-1} < v_k > v_{k+1} > \dots > v_l > v_{l+1}$$

donde  $k$  es algún entero tal que  $2 \leq k \leq l$ . Nótese que, de acuerdo con la desigualdad anterior,  $v_k$  corresponde a un máximo en el potencial. La situación se ilustra en la FigB.1. Si  $J(k, l)$  denota a la probabilidad de que se satisfaga la desigualdad anterior, entonces se tiene

$$J(k, l) = \int_0^1 dv_0 \int_0^{v_0} dv_1 \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \int_0^{v_k} dv_{k+1} \cdots \int_0^{v_l} dv_{l+1} \quad (\text{B.2})$$

La probabilidad  $P[d \geq l]$  se obtiene sumando las probabilidades  $J(k, l)$  para todos los valores de  $k$  comprendidos entre 2 y  $l + 1$ , es decir, sumando la probabilidad de todas las posibles configuraciones como la mostrada en la FigB.1, en las cuales, el máximo  $v_k$  puede ocurrir en cualquiera de las posiciones  $k = 2, k = 3, \dots, k = l + 1$ :

$$P[d \geq l] = \sum_{k=2}^{l+1} J(k, l) \quad (\text{B.3})$$

Para simplificar la notación, sea  $r = (l + 1) - k$ . Las últimas  $r$  integrales que aparecen en la expresión (B.2) son fáciles de hacer y conducen al siguiente resultado:

$$\int_0^{v_k} dv_{k+1} \cdots \int_0^{v_l} dv_{l+1} = \frac{v_k^r}{r!}$$

por lo que  $J(k, l)$  queda entonces como

$$J(k, l) = \int_0^1 dv_0 \int_0^{v_0} dv_1 \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \frac{v_k^r}{r!} \quad (\text{B.4})$$

Para calcular la integral múltiple en la ecuación anterior, utilizaremos el método de las funciones generadoras. Denotemos con  $I(k, \tau)$  a la siguiente integral múltiple:

$$I(k, \tau) = \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \frac{v_k^\tau}{\tau!} \quad (\text{B.5})$$

de tal manera que  $J(k, l)$  puede escribirse simplemente como

$$J(k, l) = \int_0^1 dv_0 \int_0^{v_0} dv_1 I(k, \tau) \quad (\text{B.6})$$

Se trata de encontrar una función generadora  $F(k, z)$  cuyo desarrollo en serie de potencias de  $z$  esté dado por

$$F(k, z) = \sum_{\tau=0}^{\infty} I(k, \tau) z^\tau \quad (\text{B.7})$$

Al sustituir en este desarrollo de potencias el valor de  $I(k, \tau)$  dado en (B.5) se obtiene

$$\begin{aligned} F(k, z) &= \sum_{\tau=0}^{\infty} \left( \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \frac{v_k^\tau}{\tau!} \right) z^\tau \\ &= \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \sum_{\tau=0}^{\infty} \frac{v_k^\tau}{\tau!} z^\tau \\ &= \int_{v_1}^1 dv_2 \cdots \int_{v_{k-1}}^1 dv_k \exp(zv_k) \end{aligned}$$

Relizando ahora la integral respecto de  $v_k$  se llega a la siguiente relación de recurrencia:

$$F(k, z) = \frac{e^z}{z} R_k - \frac{1}{z} F(k-1, z) \quad (\text{B.8})$$

donde el factor  $R_k$  está definido como

$$\begin{aligned} R_k &= \int_{v_1}^1 dv_2 \cdots \int_{v_{k-2}}^1 dv_{k-1} \\ &= \frac{(1-v_1)^{k-1}}{(k-2)!} \end{aligned} \quad (\text{B.9})$$

La solución de la ecuación de recurrencia (B.8) es

$$F(k, z) = (-1)^{k-1} \sum_{n=0}^{\infty} \frac{(v_1 - 1)^{k-1+n}}{(k-1+n)!} z^n e^z$$

la cual, si se sustituye  $e^z = \sum_{r=0}^{\infty} z^r / r!$  se transforma en

$$F(k, z) = \sum_{r=0}^{\infty} \left[ (-1)^{k-1} \sum_{n=0}^r \frac{(v_1 - 1)^{k-1+n}}{(k-1+n)! (r-n)!} \right] z^r$$

Comparando la expresión anterior con (B.7) se obtiene inmediatamente forma funcional de  $I(k, r)$ :

$$I(k, r) = (-1)^{k-1} \sum_{n=0}^r \frac{(v_1 - 1)^{k-1+n}}{(k-1+n)! (r-n)!}$$

Sustituyendo este valor de  $I(k, r)$  en la expresión (B.6) resulta

$$J(k, l) = (-1)^{k-1} \int_0^1 dv_0 \int_0^{v_0} dv_1 \sum_{n=0}^r \frac{(v_1 - 1)^{k-1+n}}{(k-1+n)! (r-n)!}$$

lo cual, al hacer las integrales, queda como

$$J(k, l) = \sum_{n=0}^r \frac{(-1)^n (k+n)}{(k+n+1)! (r-n)!} \quad (\text{B.10})$$

De las ecuaciones (B.3) y (B.10) se llega a que  $P[d \geq l]$  está dada por

$$P[d \geq l] = \sum_{k=2}^{l+1} \sum_{n=0}^{l+1-k} \frac{(-1)^n (k+n)}{(k+n+1)! (l+1-k-n)!}$$

donde hemos sustituido  $r = l+1-k$ . Con el objetivo de realizar las sumas anteriores, es conveniente hacer el cambio de variable  $m = l+1-k-n$  en la segunda sumatoria, lo cual conduce a

$$P[d \geq l] = \sum_{k=2}^{l+1} \sum_{m=0}^{l+1-k} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)! m!} (-1)^k$$

Notemos que lo único que depende del índice  $k$  en la expresión anterior es el factor  $(-1)^k$ . Sin embargo, para poder efectuar la suma de este término, es necesario invertir el orden de las sumatorias en  $k$  y  $m$ , lo cual da como resultado

$$\begin{aligned}
 P[d \geq l] &= \sum_{m=0}^{l-1} \sum_{k=2}^{l+1-m} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)!m!} (-1)^k \\
 &= \sum_{m=0}^{l-1} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)!m!} \sum_{k=2}^{l+1-m} (-1)^k \\
 &= \sum_{m=0}^{l-1} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)!m!} \frac{1 \pm (-1)^m}{2}
 \end{aligned}$$

donde se escoge el signo más o el signo menos en el último factor, dependiendo de si  $l$  es impar o par, respectivamente. Definamos  $J_1$  y  $J_2$  como

$$\begin{aligned}
 J_1 &= \sum_{m=0}^{l-1} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)!m!} \\
 J_2 &= \sum_{m=0}^{l-1} \frac{(-1)^{l+1-m} (l+1-m)}{(l+2-m)!m!} (-1)^m
 \end{aligned}$$

de tal manera que

$$P[d \geq l] = \frac{J_1 \pm J_2}{2}$$

Al desarrollar explícitamente las sumas que determinan a  $J_1$  y  $J_2$ , se observa que la mayoría de los términos se cancelan, obteniéndose

$$\begin{aligned}
 J_1 &= \frac{1}{2l!} - \frac{1}{(l+2)!} \\
 J_2 &= (-1)^{l+1} \left\{ \frac{2^{l+1}l}{(l+2)!} - \frac{1}{2l!} + \frac{1}{(l+2)!} \right\}
 \end{aligned}$$

por lo que, independientemente de si  $l$  es par o impar, se llega a que

$$P[d \geq l] = \frac{2^l}{(l+2)!}$$

De acuerdo con la ecuación (B.1) y con el resultado anterior, la probabilidad  $P(l)$  de que dos mínimos consecutivos en el potencial estén separados una distancia  $l$ , está determinada por

$$P(l) = 3 \frac{2^l}{(l+3)!} (l^2 + l - 2)$$

que es precisamente la ecuación (4.31) del capítulo 4.

## Apéndice C

# Espectro de potencias para una cadena extendida

Consideremos el sistema mostrado en la FigC.1, en el que un polímero compuesto por  $N$  monómeros interactúa con una cadena extendida formada por  $M$  partículas rígidamente unidas. Las cargas en los monómeros del polímero las denotaremos por  $q_n$  y las cargas que conforman a la cadena por  $p_m$ . Supondremos que las variables aleatorias  $q_n$  y  $p_m$  tienen promedio cero y son estadísticamente independientes, de tal manera que

$$\begin{aligned}\langle q_n q_{n'} \rangle &= q^2 \delta_{n,n'} \\ \langle p_m p_{m'} \rangle &= p^2 \delta_{m,m'} \\ \langle q_n p_m \rangle &= 0\end{aligned}\tag{C.1}$$

donde  $\delta_{n,m}$  es la delta de Kronecker. Además, supondremos que los monómeros en el polímero son todos de tamaño  $L = 1$ , mientras que las distancias  $L_m$  entre las partículas de la cadena extendida son variables aleatorias independientes, distribuidas de acuerdo con alguna función de probabilidad  $P(L)$ . En el caso particular en el que  $P(L) = \delta(L - L_0)$ , dichas distancias adquieren todas el mismo valor  $L_0$ . Si denotamos por  $x$  a la posición del extremo izquierdo de la cadena, entonces el potencial de interacción  $U_{nm}(x)$  entre el  $n$ -ésimo monómero y la  $m$ -ésima partícula está dado por

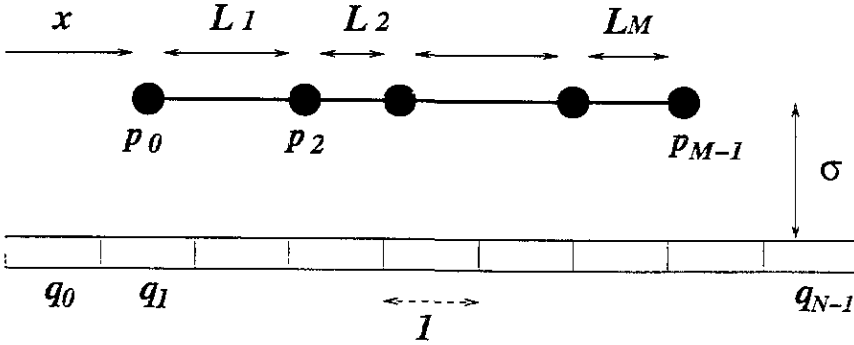


Figura C.1: Cadena extendida formada por  $M$  partículas rígidas unidas, interactuando con un polímero compuesto por  $N$  monómeros. Las distancias  $L_m$  entre las partículas de la cadena son variables aleatorias.  $x$  denota la coordenada de posición del extremo izquierdo de la cadena, mientras que  $\sigma$  es la distancia perpendicular entre el objeto y el polímero

$$U_{nm}(x) = q_n p_m g_\sigma \left( x + \sum_{j=1}^m L_j - n \right)$$

donde  $g_\sigma(x)$  es una función predeterminada que determina el tipo de interacción entre un solo monómero y una sola partícula. El potencial de interacción entre todo el polímero y la cadena extendida es entonces

$$U(x) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} q_n p_m g_\sigma \left( x + \sum_{j=1}^m L_j - n \right) \quad (\text{C.2})$$

La transformada de Fourier de la expresión anterior respecto de la posición  $x$  está dada por

$$\hat{U}(k) = \hat{g}_\sigma(k) \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} q_n p_m \exp \left[ ik \left( \sum_{j=1}^m L_j - n \right) \right]$$

donde  $\hat{g}_\sigma(k)$  es la transformada de Fourier de  $g_\sigma(x)$ . Debido a que la velocidad de la cadena es el negativo de la derivada espacial del potencial (más una fuerza constante externa que fuerza al sistema y que no vamos a considerar), entonces la transformada de Fourier de dicha velocidad está dada por



$$\hat{v}(k) = -ik\hat{g}_\sigma(k) \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} q_n p_m \exp \left[ ik \left( \sum_{j=1}^m L_j - n \right) \right]$$

El espectro de potencias de la velocidad se obtiene al multiplicar  $\hat{v}(k)$  por su complejo conjugado, con lo que resulta

$$\begin{aligned} |\hat{v}(k)|^2 &= k^2 |\hat{g}_\sigma(k)|^2 \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{n'=0}^{N-1} \sum_{m'=0}^{M-1} q_n q_{n'} p_m p_{m'} \times \\ &\times \exp \left[ ik \left( \sum_{j=1}^m L_j - \sum_{j=1}^{m'} L_j + n - n' \right) \right] \end{aligned} \quad (\text{C.3})$$

Queremos calcular ahora el espectro de potencias de la velocidad en el ensemble de todas las posibles realizaciones de cargas  $q_n$  y  $p_m$ , y en el ensemble de todas las posibles realizaciones de distancias  $L_m$ . Hay dos maneras de hacer esto: promediando primero respecto a las cargas y luego respecto a las distancias, o viceversa. Si promediamos primero en el ensemble de cargas, debido a la independencia estadística establecida en las expresiones (C.1), se obtiene

$$\begin{aligned} \langle |\hat{v}(k)|^2 \rangle_{qp} &= q^2 p^2 |k\hat{g}_\sigma(k)|^2 \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{n'=0}^{N-1} \sum_{m'=0}^{M-1} \delta_{n,n'} \delta_{m,m'} \times \\ &\times \exp \left[ ik \left( \sum_{j=1}^m L_j - \sum_{j=1}^{m'} L_j + n - n' \right) \right] \end{aligned}$$

La presencia de las deltas de Kronecker en la expresión anterior hace que sólo sobrevivan los términos diagonales de las sumatorias, obteniéndose

$$\langle |\hat{v}(k)|^2 \rangle_{qp} = NMq^2 p^2 |k\hat{g}_\sigma(k)|^2 \quad (\text{C.4})$$

Por otro lado, si se promedia primero respecto a las distancias  $L_m$  y luego respecto a las cargas se obtiene el mismo resultado. Para demostrar lo anterior, notemos que, debido a la independencia estadística de las variables  $L_m$ , la función de probabilidad conjunta  $P(L_1, L_2, \dots, L_M)$  está dada simplemente por

$$P(L_1, L_2, \dots, L_M) = \prod_{j=1}^M P(L_j)$$

Las variables  $L_m$  aparecen en la expresión (C.3) a través del término  $\exp \left[ ik \left( \sum_{j=1}^m L_j - \sum_{j=1}^{m'} L_j \right) \right]$ . Supongamos primero que  $m > m'$ . Entonces, al promediar dicho término utilizando la función de probabilidad conjunta  $P(L_1, L_2, \dots, L_M)$  se obtiene

$$\begin{aligned} \left\langle \exp \left[ ik \left( \sum_{j=1}^m L_j - \sum_{j=1}^{m'} L_j \right) \right] \right\rangle_L &= \left\langle \exp \left[ ik \left( \sum_{j=m'+1}^m L_j \right) \right] \right\rangle_L = \\ &= \int \int \cdots \int \left\{ \prod_{j=m'+1}^m P(L_j) \exp(ikL_j) \right\} dL_1 \cdots dL_M \\ &= \left[ \int P(L) \exp(ikL) dL \right]^{m-m'} \\ &= [\hat{P}^*(s)]^{m-m'} \end{aligned}$$

donde  $\hat{P}(s)$  es la transformada de Fourier de  $P(L)$ , y  $\hat{P}^*(s)$  es el complejo conjugado de  $\hat{P}(s)$ . Una expresión análoga se obtiene cuando  $m' > m$ . El resultado general es

$$\left\langle \exp \left[ ik \left( \sum_{j=1}^m L_j - \sum_{j=1}^{m'} L_j \right) \right] \right\rangle_L = [\hat{P}^\pm(s)]^{|m-m'|}$$

donde  $\hat{P}^+(s)$  y  $\hat{P}^-(s)$  se refieren respectivamente a  $\hat{P}(s)$  o a  $\hat{P}^*(s)$  según sea  $m > m'$  o bien  $m' > m$ . Por lo tanto, el promedio de la expresión (C.3) respecto de las distancias  $L_m$  está dado por

$$\begin{aligned} \left\langle |\hat{v}(k)|^2 \right\rangle_L &= |k \hat{g}_\sigma(k)|^2 \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{n'=0}^{N-1} \sum_{m'=0}^{M-1} q_n q_{n'} p_m p_{m'} \times \\ &\quad \times [\hat{P}^\pm(s)]^{|m-m'|} \exp[ik(n-n')] \end{aligned}$$

Sin embargo, debido a la independencia estadística de las variables  $q_n$  y  $p_m$ , al promediar la expresión anterior en el ensemble de todas las posibles

realizaciones de cargas, otra vez resulta que sólo sobreviven los términos diagonales  $m = m'$  y  $n = n'$ , con lo cual se obtiene nuevamente la expresión (C.4). Así, el promedio del espectro de potencias de la velocidad de la cadena a lo largo del polímero es

$$\langle |\hat{v}(k)|^2 \rangle = NMq^2p^2 |k\hat{g}_\sigma(k)|^2$$

independientemente de si las cargas  $p_m$  están uniformemente espaciadas o no.



# Bibliografía

- [1] M. Aldana, F. Cázarez-Bush, G. Cocho, G. Martínez Mekler, *Primordial synthesis machines and the origin of the genetic code*. *Physica A*, **257**, (1998), pp. 119-127.
- [2] G. Martínez Mekler, M. Aldana, F. Cázarez-Bush, R. García-Pelayo, G. Cocho, , *Primitive molecular machine scenario for the origin of the three base codon composition* *Origins of Life and Evolution of the Biosphere*, **29**, (1999), pp. 203-214.
- [3] M. Aldana, H. Larralde, G. Martínez Mekler, *Statistical characterization of random potentials*. *Physical Review E*, **61**, 6, (2000), pp. 6136-6148.
- [4] M. Aldana, G. Cocho, G. Martínez Mekler, *On the role of molecular machines in the origin of the genetic code*. *Proceedings of the XV Sitges Conference on Statistical Mechanics: "Statistical Mechanics of Biocomplexity"*, ed. Miguel Rubi. *Lecture Notes in Physics*, Springer-Verlag, Berlin, (1998).
- [5] M. Aldana, G. Cocho, G. Martínez Mekler, *La vida... ¿se originó en la Tierra?. ¿Cómo ves?* **23**, (2000), pp. 10-18.
- [6] H. F. Hudson, *El Octavo Día de la Creación*. CONACyT, México, (1987).
- [7] B. Alberts et. al, *Molecular Biology of the Cell*. Garland Publishing, (1989).
- [8] D. Voet y J. G. Voet, *Bioquímica*. OMEGA, (1992).
- [9] B. G. Lewin, *Genes V*. Oxford University Press, (1994).
- [10] B. Vollmert, *La Molécula y la Vida*. GEDISA, (1998).

- [11] M. Eigen, W. Gardiner, P. Schuster, R. Winkler-Oswatitsch, *The origin of genetic information*. Scientific American, **244**, (1981), pp. 78-94.
- [12] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution*. Science, **289**, (agosto 2000), pp. 905-9020.
- [13] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *The structural basis of ribosome activity in peptide bond synthesis*. Science, **289**, (agosto 2000), pp. 9020.
- [14] W. S. Klug y M. R. Cummings, *Conceptos de genética*. Prentice Hall, (1999).
- [15] A. J. Doig, *Improving the efficiency of the genetic code by varying the codon length: the perfect genetic code*. Jour. Theor. Biol., **188**, (1997), pp. 355-360.
- [16] M. Di Giulio, *On the origin of the genetic code*. Jour. Theor. Biol., **187**, (1997), pp. 573-581.
- [17] M. Di Giulio, *Reflections on the origin of the genetic code: a hypothesis*. Jour. Theor. Biol., **191**, (1998), pp. 191-196.
- [18] T. M. Sonneborn, *Degeneracy of the genetic code, extent, nature and genetic implications*. Evolving Genes and Proteins, (V. Bryson, H. J. Vogel eds.) New York Academic Press, (1965), pp. 377-397.
- [19] C. R. Woese, D. H. Dugre, S. A. Dugre, M. Kondo y W. C. Saxiner *On the fundamental nature and evolution of the genetic code*. Cold Spring Harbor Symp. quant. Biol. **31**, (1966), pp. 723-736.
- [20] W. M. Fitch, *The relation between frequencies of amino acids and ordered trinucleotides*. Jour. Mol. Biol., **16**, (1966), pp. 1-8.
- [21] A. L. Weber y J. C. Lacey Jr., *Genetic code correlations: amino acids and their anticodon nucleotides*. Jour. Mol. Evol., **11**, (1978), pp. 199-210.
- [22] J. T. Wong, *A coevolution theory of the genetic code*. Proc. Natl. Acad. Sci., **72**, (1975), pp. 1909-1912.
- [23] J. T. Wong, *Role of minimization of chemical distances between amino acids in the evolution of the genetic code*. Proc. Natl. Acad. Sci., **77**, (1980), pp. 1083-1086.

- [24] P. Béland y T. F. H. Allen, *The origin and evolution of the genetic code*. Jour. Theor. Biol., **170**, (1994), pp. 359-365.
- [25] W. Lee y L. Luo, *Periodicity of base correlation in nucleotide sequence*. Phys. Rev. E., **56**, (1997), pp. 848-851.
- [26] Richard Dawkins, *El gen egoísta: las bases biológicas de nuestra conducta*. Biblioteca científica Salvat, **9**, (1986).
- [27] R. Dawkins, *El relojero ciego*. Biblioteca de divulgación científica, RBA editores, (1993).
- [28] R. Monastersky, *The rise of life on Earth*. National Geographic, **193**, (1998), pp. 54-81.
- [29] H. D. Holland, *Evidence for life on earth more than 3850 million years ago*. Science, **275**, (1997), pp. 38-39.
- [30] J. M. Greenberg, *Chilarity in interstellar dust and in comets: life from dead stars*. Phys. Origin of Homochilarity in Life, (ed. D. B. Cline), AIP proc., **379**, (1996), pp. 185-210.
- [31] J. S. Levine, *The photochemistry of atmospheres: Earth, the other planets and the comets*. (ed. J. S. Levine), Academic Press, Orlando. (1985), pp. 3-38.
- [32] L. E. Orgel, *The origin of life on the earth*. Scientific American, (octubre, 1994), pp. 53-61.
- [33] T. R. Cech, *The chemistry of self-splicing RNA and RNA enzymes*. Science, **236**, (1987), pp. 1532-1539.
- [34] T. R. Cech, *RNA as an enzyme*. Scientific American, **255**, (1991), pp. 64-75.
- [35] H. F. Noller, V. Hoffarth y L. Zimniak, Science, **256**, (1992), pp. 1416-1419.
- [36] J. A. Priccilli, T. S. McConnell, A. J. Zaug y H. F. Noller, Science, **256**, (1992), pp. 1420-1424.
- [37] W. Gilbert, *The RNA world*. Nature, **319**, (1986), pp. 618-625.
- [38] R. F. Gesteland, *RNA world: the nature of RNA suggest a prebiotic RNA world*. (R. F. Gesteland y J. F. Atkins eds.), Cold Spring Harbor Monograph Series, **24**, Cold Spring Harbor Press, NY. (1993).

- [39] H. L. Sleeper, R. Lohrmann y L. E. Orgel. *Jour. Mol. Evol.*, **13**, (1979), pp. 231-245.
- [40] J. Maddox, *Lo que queda por descubrir*. DEBATE, (1999).
- [41] F. Robert, *Las primeras moléculas orgánicas*. Mundo Científico, **103**, (1991), pp. 640-651.
- [42] M. P. Bernstein, S. A. Sandford y J. Allamandola, *Life's far-flung raw materials*. Scientific American, (julio, 1999), pp. 26-33.
- [43] P. Laszlo, *Las arcillas en química orgánica*. Mundo Científico, **102**, (1991), pp. 552-561.
- [44] J. A. Aguilera, *Luces y sombras sobre el origen de la vida*. Mundo Científico, **136**, (1993), pp. 508-519.
- [45] J. M. Greenberg y Aigen Li, *From cosmic radiation of chiral bioorganics in interstellar dust to comet seeds of life's origins: are  $10^{25}$  chances enough?* Chemical Evolution: Physics of the Origin and Evolution of Life, (J. Chela-Flores y F. Raulin eds.), Kluwer Academic Publishers, Netherlands, (1996), pp. 51-71.
- [46] F. R. Krueger y J. Kissel, *Origins of Life and Evolution of the Biosphere*, **19**, (1989), pp. 87
- [47] E. K. Gibson, D. S. McKay, K. Thomas-Keppta, C. S. Romanek, H. Vali, S. J. Clemett, X. D. F. Chailier, C. R. Maechling, R. N. Zare. *Search for past life on Mars: possible relic activity in Martian meteorite ALH804001*. Science, **273**, (agosto 1996), pp. 924-930.
- [48] E. K. Gibson, D. S. McKay, K. Thomas-Keppta, C. S. Romanek, *The case for relic life in Mars*. Scientific American, (diciembre 1997), pp. 36-41
- [49] J. R. Cronin y S. Pizzarello, *Enantiomeric Excesses in meteoritic amino acids*. Science, **275**, (febrero 1997), pp. 951-955.
- [50] P. W. Atkins, *Fisicoquímica*. Adison-Wesley Iberoamericana, (1991).
- [51] K. J. Breslauer, Ronald Frank, H. Blöcker, L. A. Marry, *Predicting DNA duplex stability from the base sequence*. Proc. Natl. Acad. Sci., **83**, pp. 3746-3750.
- [52] A. D. MacKerell Jr., J. Wiórkiewicz-Kuczera, M. Karplus, *An all-atom empirical energy function for the simulation of nucleic acids*. Jour. Am. Chem. Soc., **117**, (1995), pp. 11946-11975.



- [53] M. O. Magnasco, *Molecular combustion motors*. Phys. Rev. Letters, **72**, (1994), pp. 2656-2659.
- [54] M. M. Millonas, M. I. Dykman, *Transport and current reversal in stochastically driven ratchets*. Phys. Rev. A., **185**, (1994), pp. 65-69.
- [55] F. Jülicher, A. Ajdari, J. Prost. *Modeling molecular motors*. Reviews of Modern Physics, **69**, #4, (1997), pp. 1269-1281.
- [56] N. J. Córdova, B. Ermentrout, G. F. Oster. *Dynamics of single-motor molecules: the thermal ratchet model*. Proc. Natl. Acad. Sci., **89**, (1992), pp. 339-344.
- [57] Martin Bier. *Brownian ratchets in physics and biology*. Contemporary Physics, **38**, #6, (1997), pp. 371-379.
- [58] K. Svoboda, C. F. Schmidt, B. J. Schnapp, S. M. Block. Nature, **365**, (1993), pp. 721-727.
- [59] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, Berlin, 1989).
- [60] T. P. Russell et al. *Direct observation of reptation at polymer interfaces*. Nature, **365**, (1993), pp. 235-236.
- [61] W. H. Press, W. T. Vetterling, S. A. Teukolsky, B. P. Flannery. *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, Oxford. (1992).
- [62] I. S. Gradshteyn, I. M. Ryzhik. *Table of integrals, series and products*. Academic Press.(1965), cuarta edición.
- [63] J. F. Ziman. *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems*. Cambridge University Press, Cambridge, Inglaterra, (1979).
- [64] A. H. Romero y J. M. Sancho. *Brownian motion in short range random potentials*. Phys. Rev. E., **58**, (1998), pp. 2833-2837.
- [65] A. Engel, F. Moss. *Mean first-passage time in random fields*. Phys. Rev. A., **38**, (1988), pp. 571-573.
- [66] W. Feller. *An Introduction to Probability Theory and its Applications*. Segunda Edición, Vol. 2. John Wiley and Sons, New York, (1978).
- [67] S. O. Rice. *Mathematical Analysis of Random Noise*. Bell Syst. Tech. J., **24**, 41, (1945), pp. 46-156.