

**Matemática del ADN:  
Biofísica de moléculas individuales y mecánica estadística**

ANGEL SÁNCHEZ

Grupo Interdisciplinar de Sistemas Complejos  
Departamento de Matemáticas  
Universidad Carlos III de Madrid

Instituto de Biocomputación y Física de Sistemas Complejos  
Universidad de Zaragoza

`anxo@math.uc3m.es`

6 de marzo de 2007

**Resumen**

En este artículo se discute el modelado matemático de la molécula de ADN, tratando en detalle el modelo de Peyrard-Bishop [Phys. Rev. Lett. 62, 2755 (1989)] como una de las descripciones simplificadas más útiles de dicho sistema. Se abordan tanto las propiedades de las moléculas individuales como las colectivas, haciendo particular hincapié en las herramientas y problemas matemáticos relacionados. En lo referente a dinámica de excitaciones no lineales como modelos de los procesos que tienen lugar en la molécula, se presentan resultados obtenidos con descripciones continuas, más tratables analíticamente, y con descripciones discretas, más realistas como modelos de la molécula de ADN. En cuanto a los aspectos colectivos, se discute en detalle la transición de desnaturalización del ADN, la comparación de los resultados con los datos experimentales, y las mejoras del modelo motivadas por dicha comparación. En la conclusión se mencionan los aspectos que habría que incluir en el modelo en el trabajo futuro.

**Palabras clave :** *ADN, moléculas individuales, biofísica, solitones, mecánica estadística, transiciones de fase*

**Clasificación por materias AMS :** *92C05, 92D20, 35Q51, 35Q80, 70K75, 82B26, 82C22*

## 1. Introducción

Aunque las matemáticas están desde hace tiempo muy relacionadas con la biología, la gran sinergia establecida entre ambas en los últimos años promete dar lugar a un enorme enriquecimiento de ambos campos en las próximas décadas [1, 2]. Una de las áreas de la biología donde esta interacción es más necesaria y relevante es la biología molecular [3], en la que problemas de gran trascendencia práctica (por ejemplo, en el diseño de fármacos) como el del plegamiento de las proteínas concentran los esfuerzos de un gran número de investigadores. Dicho problema es la manifestación más notable de que conocer la composición química de una molécula no permite, al menos por ahora, predecir y entender su función, que aparece como algo emergente de la interacción entre las múltiples unidades que la forman.

De entre las muchas moléculas que intervienen en la vida, el ácido desoxirribonucleico, más conocido por sus siglas ADN, es una de las más importantes, si no la más. Casi cualquier persona hoy en día sabe que el ADN contiene la información genética que permite reproducirse a los organismos y que se transmite a la descendencia para continuar el proceso. Es también bien conocido que el ADN tiene una estructura de doble hélice (véase Fig. 1; véase también en [2, 4] una descripción más detallada) donde unas moléculas llamadas bases constituyen los escalones que unen a las dos hélices, y contienen la mencionada información genética. Sin embargo, el problema de la relación estructura-función se manifiesta ya en el ADN, pese a su aparente simplicidad (comparada con moléculas como las proteínas, compuestas de unidades mucho más heterogéneas). Así, durante la replicación, o copia del ADN para su transmisión a los descendientes, o durante la transcripción, lectura del ADN para su traducción a proteínas, la molécula sufre enormes cambios conformacionales cuya relación con la estructura no es en absoluto evidente.

En este trabajo vamos a considerar un modelo sencillo que pretende aportar información sobre este problema. Al igual que para estudiar cualquier otro sistema, a la hora de plantear un modelo debemos pensar en primer lugar qué pretendemos modelar e identificar los ingredientes mínimos necesarios a tal fin, o de lo contrario correremos el riesgo de complicar el modelo y convertirlo en inútil [5]. En primer lugar, queremos entender los cambios conformacionales que ocurren en los procesos mencionados. Estos procesos son dinámicos, no ocurren instantáneamente sino que tienen velocidades características, y por tanto nuestro modelo habrá de ser también dinámico. En segundo lugar, tanto la replicación como la transcripción involucran movimientos grandes de los componentes de la molécula de ADN, que no se pueden aproximar por pequeñas oscilaciones, es decir, nuestro modelo ha de ser necesariamente no lineal. En tercer lugar, dado que los fenómenos de interés tienen lugar normalmente a la temperatura del cuerpo humano, habrá que incorporar en el modelo efectos térmicos. Por último, pretendemos que el modelo sea lo más general posible, es decir, que no esté referido a una molécula concreta de ADN con una secuencia y tamaño dados, sino que permita extraer ideas aplicables en general a cualquier molécula. Esto, que normalmente molesta mucho a los biólogos, interesados

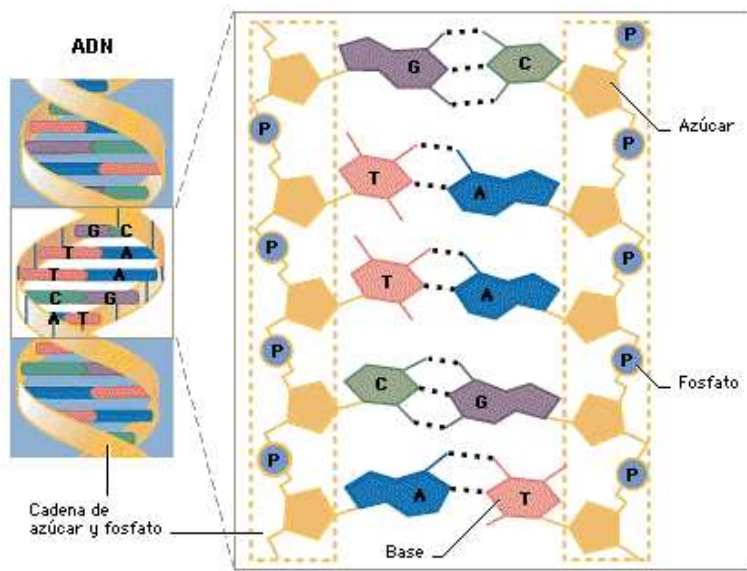


Figura 1: Estructura de la molécula de ADN. Izquierda: representación esquemática de la molécula mostrando su forma de doble hélice. Derecha: detalle ampliado mostrando los distintos componentes de la estructura. Las letras A, C, G y T denotan, respectivamente, las bases adenina, citosina, guanina y timina, y las líneas punteadas representan los enlaces de hidrógeno entre ellas.

en que los matemáticos resolvamos su problema concreto, es una perspectiva más física, más de principios fundamentales, y típicamente exige renunciar a la precisión cuantitativa en aras de una mejor comprensión cualitativa.

## 2. El problema y sus modelos

### 2.1. Fenómenos a modelar

Como acabamos de decir, nuestra meta no es entender absolutamente toda la fenomenología de la molécula de ADN, que es muy compleja y diversa (un resumen bastante conciso es [6]), sino que nos centramos en los movimientos de gran amplitud y en los fenómenos térmicos. De los primeros ya hablado en las páginas de este Boletín R. Álvarez-Nodarse, por lo que no me extenderé sobre ellos. Baste decir que las principales situaciones en las que se producen dichos movimientos son la auto-replicación o copia del ADN y su transcripción a ARN para la posterior traducción de éste en proteínas. El lector puede encontrar más detalles en [2]. En cuanto a los fenómenos térmicos, el más importante es el de la desnaturalización térmica: cuando la temperatura sube por encima de un cierto valor, situado en el rango de 60 - 80 Celsius, la doble cadena se rompe espontáneamente, y sus dos tiras constitutivas quedan libres y pueden separarse por completo. En este caso tenemos también movimientos de gran amplitud, ya que antes de alcanzarse la temperatura de desnaturalización aparecen “burbujas” en la molécula: grupos de bases consecutivas que se separan de sus compañeras dando lugar a aperturas de la molécula, pero sin llegar a romperla por entero. Estas burbujas pueden tener una vida relativamente larga e incluso desplazarse a lo largo de la doble hélice antes de que las bases se vuelvan a enlazar, cerrando la molécula. En cierto sentido son similares a las llamadas burbujas de transcripción, las zonas en las que el ADN se abre localmente para su copia a ARN. La diferencia entre ambos tipos de burbujas es que las de transcripción aparecen por la acción de una enzima, la ARN-polimerasa, que abre la doble hélice para permitir su copia, mientras que las burbujas térmicas aparecen espontáneamente sin ninguna acción ni ayuda externa (salvo, obviamente, la temperatura, que ha de ser suficientemente alta).

### 2.2. Modelos del ADN

Una vez definido claramente el problema que queremos abordar, estamos ya en condiciones de pasar a la fase de modelado. En el trabajo de R. Álvarez-Nodarse ([2], sección 3) se recogen algunos de los modelos que se han ido proponiendo en la literatura, agrupados en dos familias: modelos lineales y no lineales. Dado que queremos ocuparnos de movimientos de gran amplitud, los modelos lineales no nos serán de utilidad, ya que de partida recurren a la hipótesis de pequeños movimientos para linealizar las ecuaciones del péndulo, en las que se basan. Debemos, por tanto, recurrir a modelos no lineales. De estos, los más conocidos y estudiados son tres: el modelo sine-Gordon, el modelo de Yakushevich y el modelo de Peyrard y Bishop.

El modelo de sine-Gordon, propuesto en 1980 por Englander y colaboradores [7] es, en realidad, un “reciclado”: el modelo se plantea de tal manera que se obtiene como descripción de la dinámica del ADN la ecuación de sine-Gordon [8]. No por ello el modelo deja de tener su valor, y de hecho despertó cierto interés inicial, pero resultados experimentales posteriores lo pusieron en cuestión [9]. Más tarde, el modelo fue retomado por Salerno en una versión discreta con la cual intentó entender la relación entre secuencia genética y lugares funcionalmente relevantes de la molécula, tales como los promotores (para entendernos, secuencias cortas que marcan el inicio de un gen). De nuevo, la propuesta tuvo éxito al principio, pero trabajos recientes muestran que ese éxito es producto de un modelado de la secuencia no suficientemente preciso [11]. Así pues, el modelo de sine-Gordon no es demasiado útil para entender la dinámica del ADN, si bien tuvo el mérito de ser el primero y de motivar la introducción de los otros dos.

El modelo de Yakushevich fue propuesto por esta investigadora rusa en 1987, y se puede encontrar discutido en todo detalle en [4]. El modelo es considerablemente más complicado que el de sine-Gordon, y también más realista (aunque está a años luz de la complejidad real del problema), y se basa en considerar el movimiento de las bases como rotaciones en el plano ortogonal al eje de la doble hélice, despreciando cualquier otro movimiento. A diferencia del modelo Peyrard-Bishop, en el cual nos centraremos a continuación, considera las dos cadenas de la doble hélice como independientes, si bien ignora su geometría helicoidal. Aún así, los resultados son francamente buenos, y el modelo llega a predecir el orden de magnitud de varias cantidades físicas relacionadas con esos movimientos de gran amplitud que venimos mencionando. Por tanto, cabe pensar que el modelo identifica los grados de libertad más importantes en el proceso de transcripción. Sin embargo, tiene también sus limitaciones, como ha discutido, entre otros, Gaeta [12], sobre las que se sigue trabajando con distintas generalizaciones en la actualidad.

El modelo restante, objeto central de este trabajo, fue propuesto en 1989 por Peyrard y Bishop [13] y es, con diferencia, el que más exitoso ha resultado hasta la fecha. Ello se debe a que no sólo es un buen modelo dinámico, es decir, una buena descripción de los movimientos en los que estamos interesados, sino que también se aplica satisfactoriamente a los fenómenos térmicos, en particular a la desnaturalización (concretamente, una versión ligeramente modificada por Dauxois [14]). Por otro lado, es el que más comparaciones con resultados experimentales ha sufrido, saliendo airoso (en mayor o menor medida) de todas ellas (véase, por ejemplo, la última en [15], que contiene referencias a otras anteriores). Es precisamente por esta validez general, y a la vez por su simplicidad, que permite un tratamiento matemático (relativamente) asequible y no sólo trabajo de simulación, por lo que será el que vamos a tratar en lo que sigue.

### 3. El modelo de Peyrard y Bishop

#### 3.1. La escala de trabajo

La primera pregunta que uno ha de contestar a la hora de plantear un modelo para los cambios conformacionales básicos del ADN es la de la escala a la que se debe tratar, y ésta va a depender de los fenómenos que queremos modelar. Por ejemplo, si sólo estuviésemos interesados en cómo se comporta la molécula cuando se somete a una fuerza y se estira, hay modelos sencillos, como el llamado *worm-like chain* [16], que ignoran todos los detalles de la molécula y representándola como una cuerda flexible dan muy buenos resultados. Pero este no es nuestro caso, ya que nosotros estamos interesados en propiedades y fenómenos que son característicos del ADN, y modelos como el mencionado predicen las mismas cosas para cualquier polímero. Así pues, tenemos que ir a una escala en la que el detalle de la molécula entre en el modelo. Podríamos entonces pensar en describir el sistema a escala atómica, pero aparte de que esto sería computacionalmente muy costoso, nos sería muy difícil identificar los comportamientos colectivos que involucran simultáneamente a muchos átomos: por ejemplo, la apertura de unos cuantos pares de bases de la doble cadena involucra en torno a 1000 átomos en movimiento coordinado.

En vista de las anteriores reflexiones, la escala apropiada para nuestro modelo debe ser la de la entidad básica que codifica información en el ADN, o sea, el par de bases. Éste es un ejemplo de una *mesoescala* o escala intermedia, que salva los problemas de orden práctico de los modelos con todo detalle a la vez que mantiene la especificidad del modelo para el ADN. Observemos que el mantenernos así en un modelo simple no quiere decir que trabajar con él vaya a ser muy fácil: veremos en lo que sigue que esta simplicidad es suficiente para producir una gran riqueza de resultados y necesitar tratamientos matemáticos cuidadosos.

#### 3.2. Construcción del modelo

Nuestra descripción se va a quedar, pues, al nivel del par de bases, y para ello nuestra variable básica va a ser  $y_n$ , un número real que indica la distancia entre las dos bases que forman el  $n$ -simo par de la cadena de ADN. Supondremos que  $y_n = 0$  es el estado en el que par está cerrado, con las bases situadas a la distancia de equilibrio de sus enlaces químicos. Ello implica que  $y_n$  podrá tomar valores negativos, que indicarían una compresión de los enlaces con respecto a dicha posición de equilibrio. Gráficamente, esta elección para nuestra variable de trabajo se representa en la Fig. 2, donde aparecen las bases como cuadrados, enlazados por líneas a las cadenas azúcar-fosfato, y entre ellas por enlaces de puente de hidrógeno. Por el momento no incluimos la información genética y consideramos todas las bases iguales.

Fijada la variable de trabajo, definiremos el modelo de Peyrard-Bishop

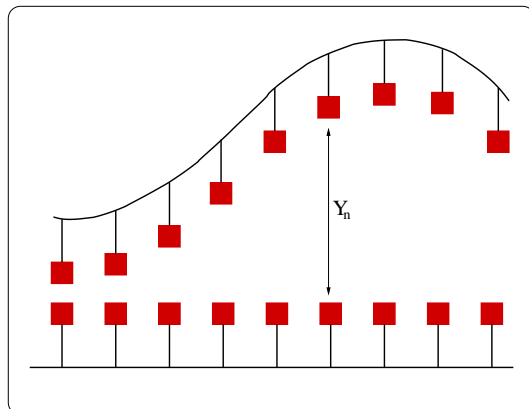


Figura 2: Modelo de Peyrard-Bishop.

mediante el siguiente Hamiltoniano [13]:

$$\mathcal{H} = \sum_{n=1}^N \frac{p_n^2}{2m} + W(y_n, y_{n-1}) + V(y_n), \quad p_n \equiv \frac{dy_n}{dt}, \quad (1)$$

a partir del cuál bastará utilizar las ecuaciones de Hamilton para encontrar las del movimiento de  $y_n$  y su variable canónica conjugada, el momento  $p_n$ . De ello nos ocuparemos más adelante; ahora es importante fijar la elección de las funciones  $W$  y  $V$  que aparecen en la ecuación (1) y entender lo que representa cada uno de dichos términos.

El potencial  $V(y)$  describe la interacción entre las bases de un par. Para este propósito, usaremos un potencial de Morse, cuya expresión es

$$V(y) = D(e^{\alpha y} - 1)^2, \quad (2)$$

y cuya representación gráfica aparece en la Fig. 3. El motivo de esta elección es que representa de manera sencilla los comportamientos fundamentales de los enlaces entre bases. En efecto, si comprimimos el par ( $y < 0$ ) encontramos una resistencia muy fuerte, y que de hecho crece exponencialmente con la cercanía, debido al contacto entre las nubes electrónicas de ambas bases, que se repelen fuertemente. Por otro lado, el estiramiento de los enlaces es lineal para distancias pequeñas y moderadas, hasta llegar un punto en el que los puentes de hidrógeno se rompen y las dos bases dejan de estar conectadas, es decir, ya no se ven y por tanto su potencial de interacción se vuelve independiente de la distancia a que se encuentren. Y finalmente, tiene un único mínimo en  $y = 0$ , es decir, en la posición de equilibrio de los enlaces.

El otro término del Hamiltoniano es el potencial de interacción entre pares de bases vecinos,  $W(y_n, y_{n-1})$ . Este término tiene dos contribuciones importantes: por un lado, está la cadena de azúcar-fosfato, a la que se unen las bases, que es

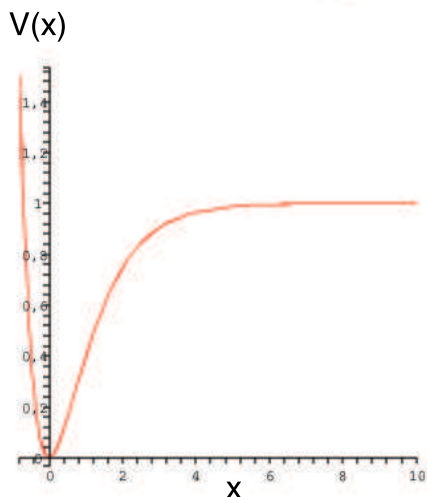


Figura 3: Potencial de Morse.

bastante rígida y sirve de unión entre una base y la siguiente, por lo que si una base se separa de la doble hélice tiende, por acción de esta cadena, a mover a las vecinas. Por otro lado, está la llamada interacción de apilamiento, que proviene del hecho de que las nubes electrónicas de cada par de bases interactúan con las que tienen por encima y por debajo en la doble hélice, ya que están a una distancia de unos pocos Ångstroms y tienen, por tanto, solapamiento (en concreto de los orbitales  $\pi$ ). Además, hay efectos de direccionalidad que hacen que la interacción no sea exactamente simétrica, es decir, que no sea lo mismo tener seguidas una base A y otra T que al revés. Todas estas contribuciones son de difícil modelado, en particular la de apilamiento, y ante esta situación y como forma de avanzar supondremos la expresión más sencilla posible para  $W$ , como es el acoplamiento armónico,  $W(y_n, y_{n-1}) = K(y_n - y_{n-1})^2/2$ . Ésta es obviamente una hipersimplificación, pero confiamos en que la rigidez de la cadena azúcar-fosfato haga que los desplazamientos relativos entre pares de bases vecinos sean pequeños y esta aproximación sea aceptable, aunque sólo los resultados lo dirán.

Quedaría por discutir la elección de los valores concretos de los parámetros que aparecen en nuestro Hamiltoniano,  $m$ ,  $D$ ,  $a$  y  $K$ . En la literatura se han utilizado distintos valores, pero una elección bastante habitual son los resultantes de la comparación de las predicciones del modelo con los experimentos de Campa y Giansanti [17] (lo que ya adelanta que el modelo da buenos resultados). De esta comparación resulta:  $D = 0,03$  eV,  $a = 4,5$  Å<sup>-1</sup>,  $K = 0,06$  eV Å<sup>-2</sup> y  $m = 300$  u.m.a, valores todos ellos bastante razonables: así, por ejemplo,  $D$  está algo por encima de la excitación térmica



típica a temperatura ambiente, lo que implica que el enlace entre pares no estará normalmente roto a dicha temperatura, y  $K$  indica un acoplamiento entre pares vecinos débil, lo que está de acuerdo con experimentos químicos.

En todo caso, comprobado que podemos elegir parámetros físicamente apropiados para nuestro modelo, para los cálculos y el trabajo numérico interesa plantear el problema en términos de cantidades sin dimensiones. Para ello introducimos  $Y = ay$  para la longitud de los enlaces de hidrógeno; medimos la energía en unidades de  $D$ , con lo que el Hamiltoniano será  $\mathcal{H}' = \mathcal{H}/D$ , e introducimos un tiempo adimensional  $\tau = \sqrt{Da^2/mt}$  y la cantidad adimensional  $S = K/Da^2$ , tenemos finalmente

$$\mathcal{H}' = \sum_{n=1}^N \frac{P_n^2}{2} + \frac{S}{2}(Y_n - Y_{n-1})^2 + (e^{-Y_n} - 1)^2, \quad P_n \equiv \frac{dY_n}{d\tau}, \quad (3)$$

expresión con la que trabajaremos de aquí en adelante.

### 3.3. Simulaciones numéricas previas

Una manera sencilla de comprobar la capacidad del modelo para describir el ADN es observar su dinámica y compararla con las propiedades experimentales conocidas. Para ello se recurre a simulación numérica, pero teniendo en cuenta que hay que incluir el efecto de las fluctuaciones térmicas, que no son despreciables a temperatura ambiente. Esto se puede hacer de varias maneras, siendo las más habituales la dinámica de Langevin y la de Nosé. La primera consiste en añadir a la ecuación del movimiento un término de ruido blanco y un término de disipación, estando las magnitudes de ambos ligadas por el teorema de fluctuación-disipación [18], mientras que la segunda es un procedimiento más elaborado [19]. Con cualquiera de ambos métodos, los resultados obtenidos son cualitativamente los mismos que los representados en la figura 4.

Los resultados a temperatura constante [fig. 4(a)], ligeramente inferior a la de desnaturalización, muestran dos tipos de patrones característicos. El más evidente es el de grandes manchas negras, correspondientes a regiones donde la separación entre bases es muy grande para grupos consecutivos de pares. Podemos asociar estas regiones a las burbujas de desnaturalización observadas en los experimentos. Por otro lado, se observan líneas punteadas verticales, debidas a regiones de unos pocos pares que se abren y cierran oscilatoriamente, y que tienen que ver con los modos de vibración localizados llamados *breathers*, que también se detectan en los experimentos. Cuando el sistema se somete a una temperatura variable, creciente desde valores por debajo del de desnaturalización hasta sobrepasarlo [fig. 4(b)], como si realizáramos un experimento de desnaturalización, el modelo da buenos resultados: la desnaturalización viene precedida por sus burbujas, que terminan por invadir toda la muestra, separando ambas cadenas (zona negra superior). Podemos concluir, pues, que el modelo se comporta, al menos cualitativamente, como la molécula de ADN, y que por tanto merece la pena llevar a cabo un estudio más detallado como el que presentamos a continuación.

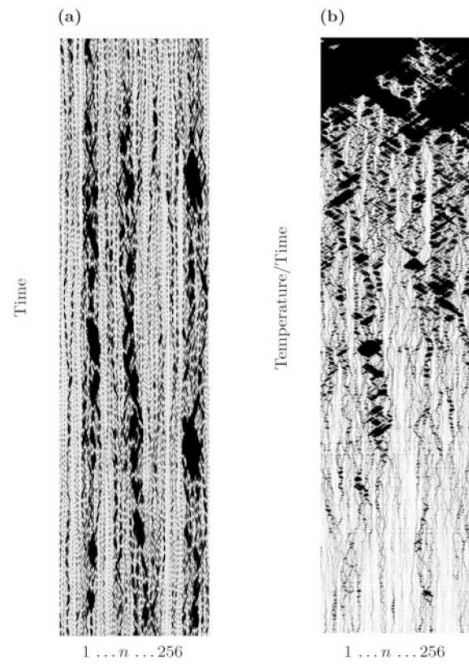


Figura 4: Simulaciones de la dinámica del ADN. El modelo tiene 256 pares de bases, y la distancia entre las bases de cada uno se muestra en una escala de grises que va del blanco (par perfectamente cerrado) al negro (par totalmente abierto). El eje horizontal corresponde a la molécula y el vertical al tiempo. (a) Simulación a temperatura constante de 340 K. (b) Simulación en la que la temperatura crece linealmente más allá de la de desnaturalización.

## 4. Excitaciones no lineales

### 4.1. Ecuaciones del movimiento

Nuestro punto de partida para intentar entender por técnicas analíticas las observaciones que acabamos de describir es la ecuación del movimiento que se deriva del Hamiltoniano (3):

$$\frac{d^2 Y_n}{d\tau^2} = S(Y_{n+1} - 2Y_n + Y_{n-1}) + 2e^{-Y_n}(e^{-Y_n} - 1). \quad (4)$$

Esta ecuación es, en realidad, un sistema de ecuaciones diferenciales ordinarias no lineales acopladas, que no se puede resolver exactamente. Tampoco se pueden linealizar, porque como hemos visto en las simulaciones, observamos movimientos de gran amplitud, que no serían descritos en el límite de pequeño  $Y_n$  (aparte de que como ya se ha dicho, queremos ir más allá de los modelos lineales). Sin embargo, podemos introducir un desarrollo que mantenga las no linealidades de primer orden, definiendo  $Y_n \equiv \epsilon \phi_n$ , con  $\epsilon \ll 1$ , y manteniendo sólo los primeros términos del sumando no lineal:

$$\frac{d^2 \phi_n}{d\tau^2} = S(\phi_{n+1} - 2\phi_n + \phi_{n-1}) - 2 \left( \phi_n - \frac{3}{2}\epsilon\phi_n^2 + \frac{7}{6}\epsilon^2\phi_n^3 \right). \quad (5)$$

En este marco, la solución que generaliza la onda plana de la aproximación lineal es ( $\theta_n \equiv qn - \omega t$ )

$$\phi_n(\tau) = (F_n e^{i\theta_n} + F_n^* e^{-i\theta_n}) + \epsilon(G_n + H_n e^{2i\theta_n} + H_n^* e^{-2i\theta_n}). \quad (6)$$

La elección de los términos adicionales se justificará más adelante, pero su origen es fácil de comprender: como las ecuaciones contienen un término  $\epsilon\phi_n^2$ , la exponencial generará de manera natural su segundo armónico. Además, como la expresión (6) es una onda plana modulada, que en la aproximación lineal sería de amplitud constante, es lógico suponer que los coeficientes  $F_n, G_n$  y  $H_n$  dependerán débilmente de la coordenada espacial a través de variables “lentas”  $X_1 = \epsilon x$ ,  $X_2 = \epsilon^2 x$ , suposición que se extiende a las variables temporales  $T_1 = \epsilon \tau$ ,  $T_2 = \epsilon^2 \tau$  (es decir, hacemos un desarrollo en escalas múltiples). Finalmente, se pasa al continuo en la dependencia espacial, reemplazando las funciones en los sitios  $n \pm 1$  por su desarrollo de Taylor,

$$F_{n\pm 1} = F \pm \epsilon \frac{\partial F}{\partial X_1} \pm \epsilon^2 \frac{\partial F}{\partial X_2} + \frac{\epsilon^2}{2} \frac{\partial^2 F}{\partial X_1^2}. \quad (7)$$

Sustituyendo en las ecuaciones del movimiento, a orden  $\epsilon^0$  la cancelación de los términos en  $e^{\pm i\theta_n}$  muestra que las ecuaciones se satisfacen si se cumple la relación de dispersión de ondas planas,

$$\omega^2 = 2 + 4S \sin^2 \frac{q}{2}. \quad (8)$$

Es interesante observar que esta relación de dispersión corresponde a la red discreta, en contraste con la aproximación continua que hemos hecho para las

funciones  $F$ ,  $G$  y  $H$ ; es decir, pese a esa aproximación, estamos conservando un cierto grado de discretitud, lo cuál es importante para el ADN. Por ello a veces esta aproximación recibe el nombre de semi-discreta [20].

A orden  $\epsilon$ , la cancelación de los términos en  $\exp(\pm i\theta_n)$  conduce a

$$\frac{\partial F}{\partial T_1} + v_g \frac{\partial F}{\partial X_1} = 0, \quad v_g \equiv \frac{S \operatorname{sen} q}{\omega}, \quad (9)$$

donde  $v_g$  es la velocidad de grupo de las ondas con relación de dispersión dada por (8). Por otro lado y al mismo orden, los términos sin exponenciales o con  $\exp(2i\theta_n)$  llevan a

$$G = 3FF^*, \quad (10)$$

$$H = -\frac{1}{2} \frac{F^2}{1 + (8S/3) \operatorname{sen}^4(q/2)}. \quad (11)$$

Llegamos ahora a orden  $\epsilon^2$ , para el cuál se obtiene la ecuación realmente interesante, imponiendo la cancelación de términos en  $\exp(\pm i\theta_n)$ . Tras pasar al sistema de referencia móvil con velocidad  $v_g$ , obtenemos

$$i \frac{\partial F}{\partial \tau_2} + P \frac{\partial^2 F}{\partial \xi_1^2} + Q|F|^2 = 0, \quad (12)$$

donde

$$P = \frac{S\omega^2 \cos q - S^2 \operatorname{sen}^2 q}{2\omega^3}, \quad (13)$$

$$Q = \frac{1}{\omega} \left( 11 - \frac{9}{3 + 8S \operatorname{sen}^4(q/2)} \right). \quad (14)$$

La ecuación (12) no es otra que la famosa ecuación de Schrödinger no lineal (NLS, de sus siglas en inglés), y dado que el producto  $PQ > 0$  cuando  $q < \pi/2$ , tiene soluciones localizadas para vectores de la onda portadora pequeños. De ellas hablaremos en la siguiente subsección.

#### 4.2. *Breathers* y localización de energía

La ecuación NLS tiene soluciones de tipo solitón [21, 22]: soluciones localizadas en el espacio, típicamente exponencialmente, que tienen la propiedad de propagarse manteniendo su forma incluso aunque interaccionen con otros solitones (de ahí su nombre). De hecho, la ecuación NLS es completamente integrable, propiedad que suele llevar asociada la existencia de este tipo de soluciones. Para nuestro problema, ello nos permite escribir  $F$  directamente y, a partir de ahí, obtener  $G$  y  $H$ , con lo que finalmente llegamos a

$$Y_n(t) = 2a_0 \operatorname{sech} \left[ a_0 \sqrt{\frac{Q}{2P}} (n - v_g t - V_e t) \right] \cos(q'n - \omega't) + a_0^2 \operatorname{sech} \left[ a_0 \sqrt{\frac{Q}{2P}} (n - v_g t - V_e t) \right] \{3 - \cos[2(q'n - \omega't)]\}, \quad (15)$$

donde  $a_0$  y  $U_e$  son dos parámetros que caracterizan a la solución solitón de NLS y

$$\omega' = \omega + \frac{v_g U_e}{2P} + \frac{U_e^2}{4P} - \frac{Q a_0^2}{2}, \quad (16)$$

$$q' = q + \frac{U_e}{2P} \quad (17)$$

son la frecuencia y el vector de onda corregidos para la onda portadora. Es importante recordar, visto cómo la hemos obtenido, que desde el punto de vista de nuestro modelo de ADN esta solución es una aproximación. Sin embargo, las simulaciones numéricas del modelo original muestran que es una buena solución, estable, siempre que  $\omega'$  no esté en la banda de frecuencias del modelo discreto linealizado. Ello puede evitarse, por ejemplo, si se toma  $\omega$  cerca de las frecuencias más bajas permitidas,  $\omega = \sqrt{2}$ . Ello corresponde a una solución oscilatoria, de gran amplitud, muy similar a las observadas en la figura 4. Así pues, el *breather* de la ecuación NLS [21], así llamado precisamente por su carácter oscilante en el tiempo, resulta ser una buena explicación de las aperturas espontáneas del ADN, a las que los biólogos también se refieren como *breathing*. Sin embargo, para que esto tenga sentido real, debemos asegurarnos de que estas soluciones aparecen de manera espontánea, es decir, que no es preciso escoger cuidadosamente los datos iniciales para llegar a ellas. En este sentido, podemos estar tranquilos, ya que es conocido que las soluciones extendidas a todo el sistema son inestables, algo conocido como inestabilidad modulacional [23]. El resultado de esa inestabilidad es la formación de un conjunto de solitones aproximados, que, posteriormente, interactúan, y en esa interacción los más grandes crecen a expensas de los más pequeños [24], produciendo así los modos de gran amplitud que nos interesan.

### 4.3. Paredes de dominio

En el límite continuo, la ecuación dinámica resultante del Hamiltoniano (3),

$$\frac{\partial^2 Y}{\partial t^2} - S \frac{\partial^2 Y}{\partial x^2} + \frac{\partial V(Y)}{\partial Y} = 0, \quad \text{con } V(Y) = (e^{-Y} - 1)^2, \quad (18)$$

tiene una solución estática exacta, completamente distinta del *breather*, pero también significativa para la dinámica del ADN. Buscando soluciones estacionarias, se tiene

$$-S \frac{d^2 Y}{dx^2} + \frac{dV(Y)}{dY} = 0, \quad (19)$$

que se resuelve fácilmente por cuadraturas. Multiplicando por  $dY/dx$  e integrando queda

$$-\frac{1}{2} S \left( \frac{dY}{dx} \right)^2 + V(Y) = C, \quad (20)$$

siendo  $C$  una constante determinada por las condiciones de contorno. Imponiendo que los pares de bases están en equilibrio,  $Y = 0$ , cuando  $x \rightarrow -\infty$ ,

resulta  $C = 0$ , y tenemos

$$\frac{dY e^Y}{e^Y - 1} = \sqrt{\frac{S}{2}} dx \Rightarrow \ln(e^Y - 1) = \sqrt{\frac{S}{2}}(x - x_0), \quad (21)$$

y por tanto

$$Y(x) = \ln \left[ 1 + \exp \left( \sqrt{\frac{S}{2}}(x - x_0) \right) \right], \quad (22)$$

donde  $x_0$  es una constante de integración que nos da la posición de la solución. Esta solución describe una configuración donde una parte de la molécula (la situada en  $x < x_0$  está cerrada, mientras que para  $x \gg x_0$  la separación de los pares crece linealmente con el espacio y la molécula está completamente desnaturalizada. Es decir, tenemos una pared de dominio entre dos estados de la molécula de ADN.

Mientras que las soluciones tipo *breather* discutidas antes tienen que ver con las aperturas espontáneas de la doble cadena y con las burbujas de transcripción, soluciones como la que acabamos de encontrar son relevantes para el proceso de replicación del ADN, en el que las dos hebras se separan por completo formando dos nuevas moléculas idénticas a la original. Sin embargo, la solución continua y en un sistema infinito presenta diversos problemas a la hora de utilizarla como modelo de este proceso, de los cuales uno de los más importantes es que su energía total es infinita, siendo, por tanto, inestable: si damos esta solución como dato inicial para la ecuación dependiente del tiempo, veremos que evoluciona a la solución  $Y = 0$ , es decir, a la cadena cerrada por completo. Es, pues, necesario, hacer un estudio más cuidadoso, que presentamos en la siguiente sección.

## 5. Paredes de dominio en el modelo discreto

En esta sección, que sigue básicamente la referencia [25], seguiremos utilizando el modelo PB, pero ahora escribiremos por conveniencia su hamiltoniano de la siguiente forma:

$$H = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \dot{Y}_n^2 + \frac{1}{2R} (Y_{n+1} - Y_n)^2 + V(Y_n) \right\}, \quad (23)$$

escrito en términos de una constante positiva y sin dimensiones,  $R$ , que viene a jugar el papel de una discretización efectiva,  $a = \sqrt{R}$ , de manera que  $R \gg 1$  representa una discretización con paso espacial muy grande mientras que  $R \ll 1$  es el límite continuo.

Igual que en el caso continuo, las soluciones estacionarias del hamiltoniano (23) satisfacen  $\partial H / \partial Y_n = 0$ , lo que conduce a la relación de recurrencia

$$Y_{n+1} = 2Y_n - Y_{n-1} + RV'(Y_n), \quad (24)$$

con  $n = 1, 2, \dots, N$ . Las soluciones de esta relación quedan definidas unívocamente por la condición inicial  $\{Y_0, Y_1\}$ . Restringiéndonos al caso  $Y_0 = 0$ ,

cosa que podemos hacer sin pérdida de generalidad, entonces  $Y_n$  dependerá sólo del valor  $Y_1 = y$ , de manera que la ecuación (24) se puede reescribir en términos de  $y$  introduciendo la notación  $Y_n(y)$  en vez de  $Y_n$ . De ahora en adelante discutiremos el comportamiento de las soluciones  $Y_n(y)$  como función de  $y$ .

La ecuación (24) describe tanto las soluciones estables como las inestables del hamiltoniano (23). Para estudiar la estabilidad, necesitamos el hessiano del sistema, dado por

$$\mathcal{H}_N(y) = \begin{pmatrix} d_1(y) & -1 & 0 & \dots & 0 \\ -1 & d_2(y) & -1 & \dots & 0 \\ 0 & -1 & d_3(y) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_N(y) \end{pmatrix}, \quad (25)$$

donde hemos llamado  $d_n(y) \equiv 2 + RV''(Y_n(y))$ . Llamando  $\Delta_n(y)$  al determinante del menor principal de orden  $n$  del hessiano  $\mathcal{H}_N(y)$ , es decir,

$$\Delta_n(y) = \det(\mathcal{H}_n(y)), \quad (26)$$

las soluciones estables son las que satisfacen  $\Delta_n(y) > 0$  para todo  $n = 1, 2, \dots, N$ . Como el hessiano es una matriz tridiagonal, hay una relación de recurrencia entre los distintos  $\Delta_n$ ,

$$\Delta_{n+1}(y) = d_{n+1}(y)\Delta_n(y) - \Delta_{n-1}(y), \quad (27)$$

con  $\Delta_1 = d_1$  y  $\Delta_2 = d_1 d_2 - 1$ .

La expresión (27) que acabamos de obtener se puede reescribir en función de  $Y'_n(y)$ . Derivando la ecuación (24) con respecto a  $y$  encontramos:

$$Y'_{n+1}(y) = \frac{dY_{n+1}(y)}{dy} = [2 + RV''(Y_n(y))]Y'_n(y) - Y'_{n-1}(y) = d_n(y)Y'_n(y) - Y'_{n-1}(y), \quad (28)$$

con  $Y'_2(y) = \Delta_1(y)$  and  $Y'_3(y) = \Delta_2(y)$ . Por tanto, se ha de cumplir que

$$\Delta_n(y) = Y'_{n+1}(y), \quad (29)$$

para  $n = 1, 2, \dots, N$ , con lo que la región de estabilidad de las soluciones de (24) es la de los puntos que satisfacen  $Y'_n(y) > 0$  para todo  $n = 2, 3, \dots, N + 1$ .

Hasta aquí, no hemos realizado ninguna aproximación, y todos los resultados son exactos para cualquier  $V(Y)$ . A partir de ahora, nos centraremos en el modelo PB eligiendo el potencial de Morse como  $V(Y)$ , tanto para intentar encontrar una expresión analítica de las soluciones de la relación (24) como para analizar su estabilidad en función de la condición inicial  $y$ . Es interesante notar que este formalismo produce soluciones también para el caso del modelo PB continuo, que se pueden encontrar en [25], si bien nosotros nos centraremos en el caso discreto como modelo más realista del ADN.

### 5.1. Soluciones

Como ya hemos dicho, el límite discreto del modelo PB corresponde a  $R \gg 1$ . En este caso, usando una suma telescópica de  $Y_{n+1} - Y_n$ , y teniendo en cuenta las condiciones iniciales  $Y_0 = 0$  e  $Y_1 = y$ , la ecuación (24) se puede reescribir de la forma

$$Y_{n+1}(y) = (n+1)y + R \sum_{k=1}^n (n+1-k) V'(Y_k(y)). \quad (30)$$

Definamos ahora

$$f_k(y) \equiv V'(Y_k(y)) = 2e^{-Y_k(y)} (1 - e^{-Y_k(y)}) = f_1(Y_k(y)). \quad (31)$$

Estas funciones aparecen representadas para distintos valores de  $R$ , dentro del límite discreto, en la figura 5. Como puede verse, las  $f_k$  son muy localizadas, y su solapamiento depende de  $R$ . De hecho, para valores en el límite discreto que nos interesa, este solapamiento es muy pequeño, y podemos calcular la posición de los máximos de las  $f_k(y)$ , y a partir de ahí aproximar  $f_k(y)$  por la primera de ellas,  $f_1(y)$ , poniendo

$$f_k(y) \simeq f_k^{(1)}(y) \equiv f_1(b_k y), \quad (32)$$

con

$$b_n = \frac{1}{2\sqrt{R(R+2)}} \left[ \left( R+1 + \sqrt{R(R+2)} \right)^n - \left( R+1 - \sqrt{R(R+2)} \right)^n \right]. \quad (33)$$

Esta expresión nos proporciona una aproximación analítica de las soluciones para cualquier valor  $y$ , eso sí, tanto más correcta cuanto más discreto sea el caso considerando. Si la sustituimos en la ecuación (30), llegamos a

$$Y_{n+1}^{(1)}(y) = (n+1)y + R \sum_{k=1}^n (n+1-k) f_k^{(1)}(y). \quad (34)$$

lo que constituye una buena aproximación de  $Y_n(y)$  para valores grandes de  $R$ . La figura 6 confirma la precisión de esta aproximación: para  $R \gtrsim 100$ , resulta ser muy buena, mientras que al disminuir  $R$  empeora. Para valores más bajos de  $R$ , se puede mejorar la aproximación recurriendo a la siguiente función,  $f_2(y)$ , en vez de  $f_1(y)$ , como aproximación para el resto de las  $f_k(y)$ , definiendo

$$f_k(y) \simeq f_k^{(2)}(y) \equiv f_2\left(y \frac{b_k}{b_2}\right) \quad (35)$$

para  $k = 3, 4, \dots$ , y aproximando  $Y_n(y)$  por

$$Y_{n+1}^{(2)}(y) = (n+1)y + nRf_1(y) + R \sum_{k=2}^n (n+1-k) f_k^{(2)}(y). \quad (36)$$



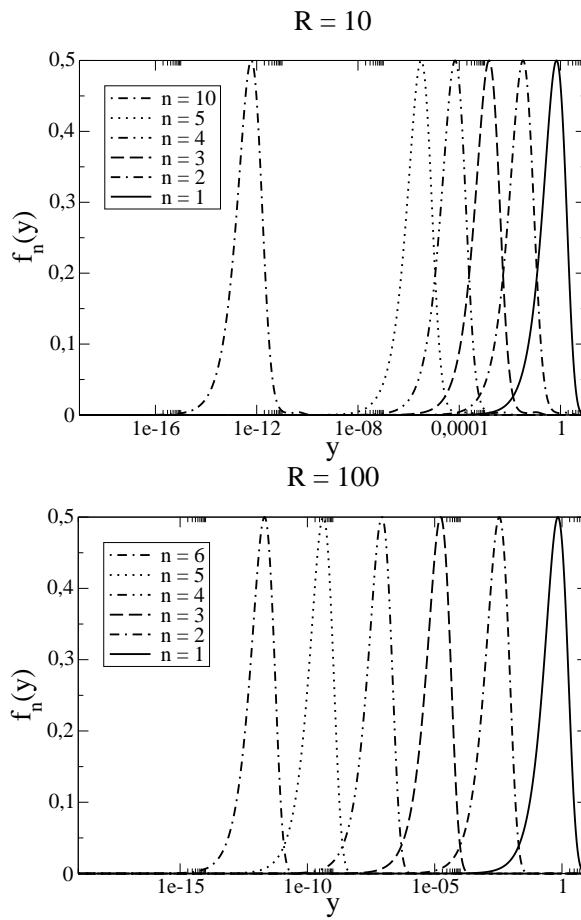


Figura 5: Funciones  $f_n(y)$  para  $R = 10$  (arriba) y  $R = 100$  (abajo), para diferentes valores de  $n$ .

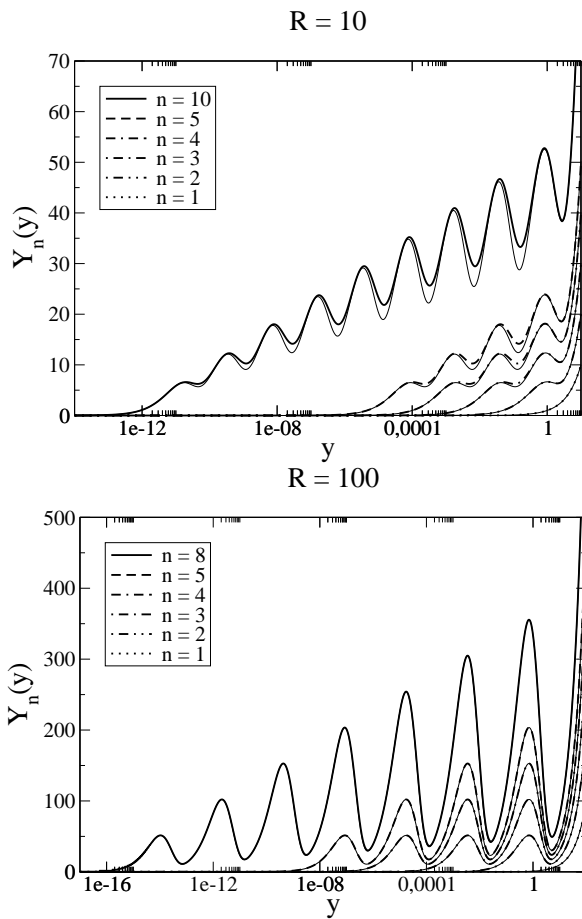


Figura 6: Aproximación  $Y_n^{(1)}(y)$  (véase ecuación (34)) frente al valor exacto  $Y_n(y)$ , para  $R = 10$  (arriba) y  $R = 100$  (abajo), para diferentes valores de  $n$ . Las soluciones exactas aparecen dibujadas en líneas gruesas, mientras que las correspondientes aproximaciones aparecen en líneas finas.

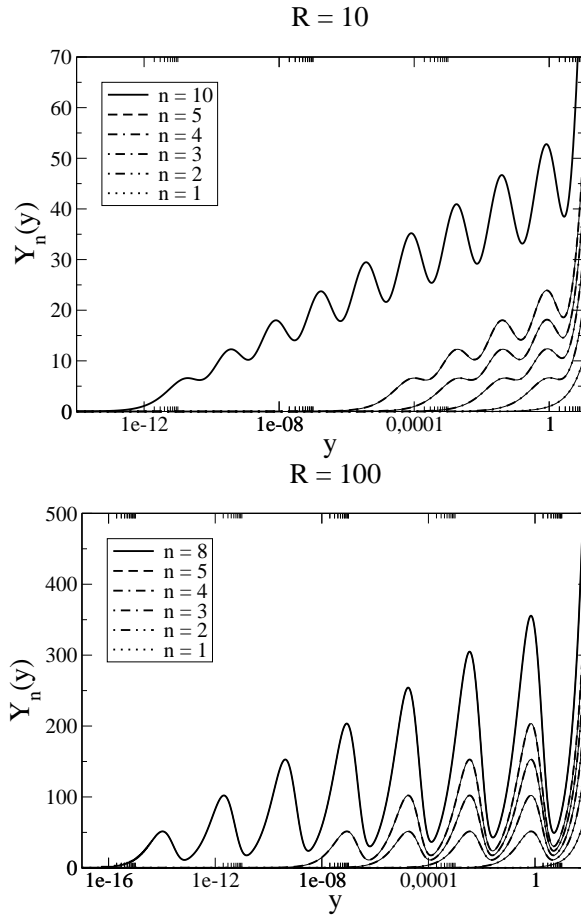


Figura 7: Aproximación  $Y_n^{(2)}(y)$  (véase la ecuación (36)) frente a los valores exactos  $Y_n(y)$ , para  $R = 10$  (arriba) y  $R = 100$  (abajo), para diferentes valores de  $n$ . Las soluciones exactas aparecen dibujadas en líneas gruesas, mientras que las correspondientes aproximaciones aparecen en líneas finas.

Esta aproximación es aún mejor que la dada por  $Y_n^{(1)}(y)$ , e incluso para  $R = 10$  los resultados son muy similares a los exactos (véase la figura 7 para más detalles).

Los errores de estas aproximaciones dependen de los valores de  $R$  y de  $n$ . Por ejemplo,  $Y_n^{(1)}(y)$  es exacta para  $n = 1$  and  $n = 2$ , para cualquier valor de  $R$ , mientras que  $Y_n^{(2)}(y)$  es exacta hasta  $n = 3$  para cualquier  $R$ . Para  $n$  bajos, la mayor diferencia entre el valor exacto  $f_n(y)$  (que se puede obtener numéricamente sin dificultad) y  $f_n^{(1)}$  se sitúa en torno a los máximos de  $f_n(y)$ , con un error máximo  $E_{max}^{(1)} \simeq 0,06$  para  $R = 10$  y  $E_{max}^{(1)} \simeq 0,006$  para  $R = 100$ . Para la aproximación de segundo orden basada en  $f_n^{(2)}$ , el error máximo es  $E_{max}^{(2)} \simeq 2,7 \cdot 10^{-3}$  para  $R = 10$  y  $E_{max}^{(2)} \simeq 3,2 \cdot 10^{-5}$  para  $R = 100$ . Por otro lado, otra región mal aproximada, aunque menos que la que acabamos de mencionar, aparece en torno a los máximos de  $f_{n-1}(y)$ . Estos cálculos se pueden hacer también para órdenes mayores de aproximación,  $f_n^{(k)}(y)$ , y se puede ver que la reducción del error conseguida al utilizar  $k + 1$  en vez de  $k$  es al menos de un orden de magnitud. Hay un límite computacional, dado por la precisión del ordenador, que no nos permite comprobar la validez de esta afirmación más allá de unos ciertos  $k$  y  $n$ , que dependen de  $R$ , pero todo indica que el mismo comportamiento se observará para valores mayores de  $k$  y  $n$ .

Las aproximaciones dadas por las ecuaciones (34) y (36) nos permiten calcular muy aproximadamente  $Y_n(y)$  para cualquier valor de  $n$  e  $y$ . Esto es importante porque el cálculo numérico exacto utilizando la recurrencia (24), (30) tiene problemas serios para valores de  $y$  cercanos a cero, debido a la precisión del ordenador (sobre la que volveremos más adelante). Por tanto, para analizar la estabilidad en el límite discreto, utilizaremos las aproximaciones  $Y_n^{(k)}(y)$  que acabamos de ver. De esta manera, podemos estudiar sistemas mucho mayores que los accesibles por solución numérica directa de las relaciones de recurrencia.

## 5.2. Estabilidad

En esta sección vamos a estudiar la estabilidad de las soluciones estacionarias, centrándonos en el caso de sistemas pequeños, para los que la derivada  $Y_n'(y)$  se puede calcular sin aproximaciones para todo  $n$ , sin incurrir en errores de precisión del ordenador. En la figura 8 se muestra la dependencia de  $\partial Y_n(y)/\partial \log_{10}(y)$  de la condición inicial  $y$ , en escala logarítmica, para distintos valores del tamaño del sistema,  $N$ . El motivo de utilizar  $\partial Y_n(y)/\partial \log_{10}(y)$  en vez de  $Y_n'(y)$  es el de obtener una curva suave: la representación directa de  $Y_n'(y)$  haría imposible observar los intervalos con  $Y_n'(y) > 0$ . Como el signo de ambas derivadas es el mismo para todo  $y > 0$ , se prefiere la escala logarítmica. Con este cambio, para cada  $n$  aparece una estructura “sinusoidal” en la figura 8, con  $n - 1$  máximos y mínimos alrededor de  $Y_n'(y) = 0$ . Resulta así claro que en sistemas de tamaño  $n$  aparece un nuevo intervalo de inestabilidad adicional respecto a los sistemas de tamaño  $n - 1$ , situado a valores bajos de  $y$ . Además, la figura 8 también indica que el conjunto de puntos inestables de un sistema de tamaño  $n$  contiene al de puntos inestables de uno de tamaño  $n - 1$ . Un

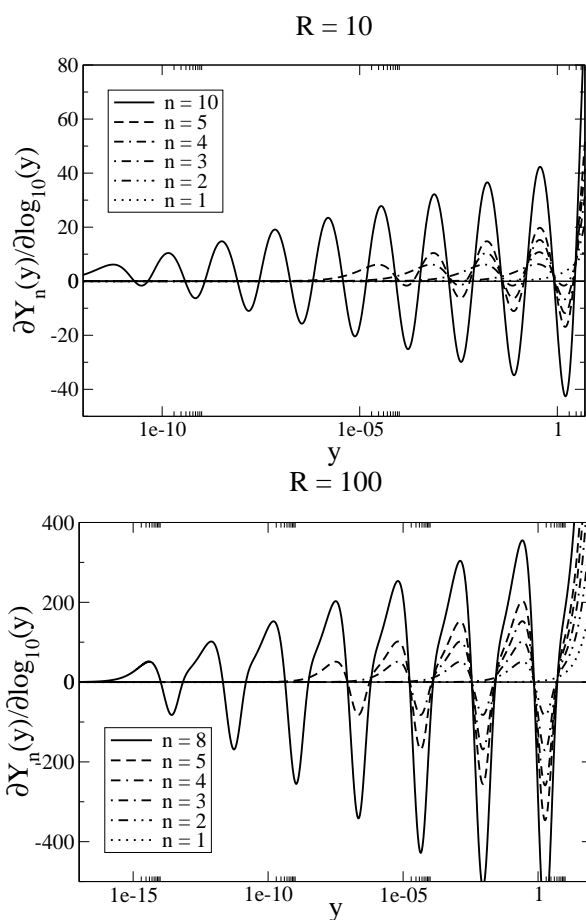


Figura 8: Derivada de  $Y_n(y)$  con respecto a  $\log_{10}(y)$  para distintos valores de  $n$  y  $R = 10$  (arriba) y  $R = 100$  (abajo), siendo el eje  $x$  logarítmico. La región de estabilidad de un sistema de tamaño  $N$  es la intersección de los conjuntos de puntos que satisfacen  $\partial Y/\partial y > 0$  para  $n = 1, 2, \dots, N + 1$ . A partir de las figuras, encontramos que la región de estabilidad corresponde a los puntos que satisfacen la condición para  $n = N + 1$ , ya que la región de estabilidad del sistema de tamaño  $N$  parece estar contenida en la región de estabilidad del de tamaño  $N - 1$  (véase el texto para más detalles).

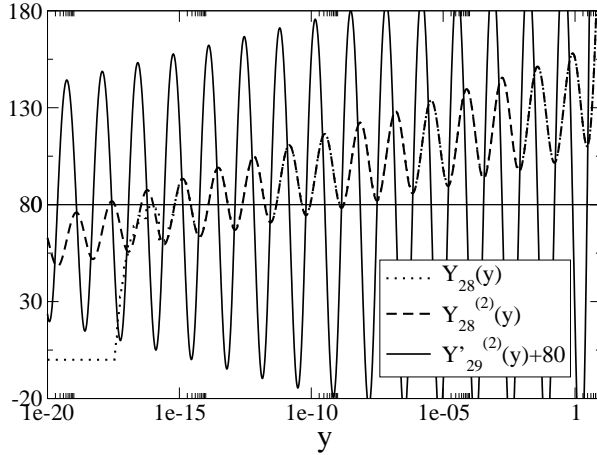


Figura 9: Soluciones estables e inestables en la aproximación  $Y_N^{(2)}$  para un sistema con  $N = 28$  sites,  $Y_0 = 0$  y  $Y_N = 80$ . Las soluciones son las intersecciones de las curvas con la línea  $Y_{28} = 80$ , que marca la condición de contorno derecha. Es preciso estudiar el signo de  $Y'_{29}$  para establecer la estabilidad de las soluciones, por lo que se dibuja  $\partial Y_{29}(y)/\partial \log_{10}(y)$  para mayor claridad. Nuestra aproximación da ocho soluciones estables y siete inestables para estos parámetros, exactamente con en [26].

argumento de plausibilidad en favor de este resultado pasa por analizar los puntos que satisfacen  $Y'_n(y_0) = 0$ , que son los de los extremos de un intervalo de puntos inestables. En ese caso, si  $Y'_{n-1}(y_0) > 0$ , tenemos que  $Y'_{n+1}(y_0) < 0$  (véase la ecuación (28)), y por tanto el intervalo de puntos inestables del sistema de tamaño  $n$  es mayor que el de  $n - 1$ . Esta condición se satisface para todos los nuevos intervalos de inestabilidad que aparecen para cada  $Y_n(y)$ , comenzando con  $Y_2(y)$ , y por inducción se puede aplicar a todos los sistemas. Así pues, los puntos estables de un sistema de tamaño  $n$  son los que satisfacen  $Y'_{n+1}(y) > 0$ .

Como comprobación independiente de la validez de estos resultados, se pueden comparar con los recientemente publicados en [26]. Estudiando el problema discreto estacionario con condiciones de contorno fijas, los autores encuentran ocho soluciones estables y siete inestables para un sistema de tamaño  $N = 28$ . En concreto, las condiciones de contorno utilizadas fueron  $Y_0 = 0$  and  $Y_N = 80$  para  $R = 10,1$ . La solución exacta  $Y_n(y)$  y la aproximada  $Y_n^{(2)}(y)$  de este sistema se recogen en la figura 9. La gráfica deja claro el problema de precisión mencionado con anterioridad, que aparece cuando intentamos calcular la solución numérica exacta para valores bajos de  $y$ . Por otra parte, la solución aproximada  $Y_n^{(2)}(y)$  se comporta perfectamente en todo el rango de  $y$ . Vemos pues que  $Y_n^{(2)}(y)$  da el mismo número de soluciones estables e inestables que el obtenido numéricamente en [26], lo que implica que la estructura de picos de

$Y_n(y)$  da una buena explicación del número y estructura de las soluciones.

## 6. Mecánica estadística de la desnaturalización térmica

Hasta aquí hemos venido estudiando la mecánica de moléculas individuales de ADN. Hemos obtenido soluciones que describen las excitaciones no lineales de la molécula, correspondientes además a distintos procesos que tienen lugar en su función biológica, y que constituyen interesantes ejemplos de estudios en dinámica no lineal. Sin embargo, es preciso ir más allá, puesto que la condición normal para las moléculas es la temperatura corporal, que introduce efectos de fluctuaciones térmicas como ya vimos en las simulaciones. Si queremos abordar esta perspectiva complementaria, tenemos que utilizar el marco de trabajo dado por la mecánica estadística. En esta sección, vamos a ocuparnos del problema de la desnaturalización térmica, que veremos nos proporciona nuevos e interesantes problemas matemáticos.

### 6.1. Transición de desnaturalización

La mecánica estadística es la ciencia que se ocupa de los sistemas de muchas partículas en interacción y a temperatura finita, intentando obtener una descripción colectiva aplicable al sistema como un todo a partir de los grados de libertad microscópicos. De esta manera, se utiliza para fundamentar la termodinámica, que se puede plantear alternativamente como una ciencia axiomática. El lector interesado puede encontrar buenas introducciones en las referencias [27, 28], que no requieren mayores conocimientos de física.

En el formalismo de la mecánica estadística, el objeto central de estudio es la energía libre. Consideramos nuestro sistema en contacto con un baño térmico a temperatura  $T$ , manteniendo el número de partículas (pares de bases) de la molécula constante. Estamos, en el lenguaje de la mecánica estadística, en la colectividad canónica, que nos prescribe comenzar por calcular la llamada función de partición, definida por

$$Z \equiv \int \prod_n dp_n dy_n e^{-\beta H(p_n, y_n)}, \quad (37)$$

donde  $\beta \equiv 1/k_B T$ ,  $k_B$  es la constante de Boltzmann, y  $H$  es, una vez más, el Hamiltoniano del sistema a estudiar, que en nuestro caso será el del modelo PB. Una vez calculada  $Z$ , se puede derivar a partir de ella todas las magnitudes termodinámicas de interés. Éste es en realidad el problema central de la mecánica estadística: en general no se sabe o no se puede calcular la función de partición. Sin embargo, incluso cuando no se sabe, se puede obtener información de ella, como vamos a ver a continuación.

Para calcular la integral (37), vamos a especificar en detalle nuestro sistema. Consideramos una cadena finita de  $N$  pares de bases, con condiciones de contorno periódicas, lo que viene a ser cómo añadir un par de bases ficticio

con índice 0 que hace lo mismo que el par  $N$ . Escribiendo explícitamente el Hamiltoniano, lo que tenemos que calcular es

$$Z = \int \int \prod_{n=1}^N dp_n \prod_{n=0}^N dy_n \exp \left\{ -\beta \left[ \sum_{n=1}^N \frac{p_n^2}{2m} + \frac{1}{2} K(y_n - y_{n-1})^2 + V(y_n) \right] \right\} \delta(y_n - y_0), \quad (38)$$

donde la función  $\delta$  asegura el cumplimiento de las condiciones de contorno periódicas. Las integrales en las variables  $p_n$  son simples integrales gaussianas desacopladas, que se calculan fácilmente, obteniéndose

$$Z = \left( \sqrt{2\pi m k_B T} \right)^N Z_y \quad (39)$$

siendo  $Z_y$  la llamada integral de configuración, dada por

$$Z_y = \int \prod_{n=0}^N dy_n \left[ \prod_{n=1}^N \exp(-\beta f(y_n, y_{n-1})) \right] \delta(y_n - y_0), \quad (40)$$

siendo  $f(y_n, y_{n-1}) \equiv K(y_n - y_{n-1})^2 + V(y_n)$ .

Procederemos a continuación mediante una técnica estándar en este tipo de problemas unidimensionales, conocida como el método del operador de transferencia. En particular, nos centramos en las autofunciones de dicho operador, que se definen por

$$\int dy_{n-1} \exp[-\beta f(y_n, y_{n-1})] \times \phi(y_{n-1}) = e^{-\beta \epsilon_i} \phi_i(y_n). \quad (41)$$

Utilizando el hecho de que las autofunciones de este operador forman una base, podemos escribir  $\delta(y_n - y_0) = \sum_i \phi_i^*(y_N) \phi_i(y_0)$ , con lo que sustituyendo en la expresión (40) e integrando sucesivamente, obtenemos

$$Z_y = \sum_i e^{-\beta N \epsilon_i} \int dy_N |\phi_i(y_N)|^2 = \sum_i e^{-\beta N \epsilon_i}, \quad (42)$$

utilizando  $\phi_i$  normalizadas.

El paso siguiente en mecánica estadística es tomar el llamado límite termodinámico,  $N \rightarrow \infty$ , que es de hecho el único en el que se obtienen transiciones de fase; en sentido estricto, en sistemas de tamaño finito no hay verdaderas transiciones de fase (volveremos sobre la definición concreta de transición de fase más adelante). En este límite, el resultado viene dominado por el menor valor de  $\epsilon_i$ , que denotaremos por  $\epsilon_0$ . Llegamos finalmente a la expresión de la energía libre:

$$f = -\frac{k_B T}{N} \ln Z = \epsilon_0 - \frac{k_B T}{2} \ln(2\pi m k_B T). \quad (43)$$



Cálculos adicionales, que omitimos, muestran que otras magnitudes se pueden obtener también con los autovalores y autofunciones del operador de transferencia. Así, la apertura media de los pares viene dada por

$$\langle y \rangle = \int_{-\infty}^{\infty} dy y |\phi_0(y)|^2, \quad (44)$$

lo que nos lleva a interpretar  $|\phi_0(y)|^2$  como la densidad de probabilidad de la apertura de los pares, mientras que la función de correlación resulta ser

$$\begin{aligned} C(n) &= \langle (y_n - \langle y \rangle)(y_0 - \langle y \rangle) \rangle \\ &= \sum_{i=1}^{\infty} \int_{-\infty}^{\infty} dy y \phi_i^*(y) \phi_0(y) \exp(-\beta(\epsilon_i - \epsilon_0)|n|). \end{aligned} \quad (45)$$

Vemos, pues, que el problema se ha reducido a encontrar los autovalores y autovectores del operador de transferencia. Para ello, hay dos opciones: la primera, atacar el problema numéricamente, discretizando el operador y transformándolo en una matriz, que luego se puede diagonalizar con cualquiera de las rutinas disponibles; la segunda opción es la de renunciar al cálculo exacto e intentar aproximaciones que nos permitan entender cualitativamente el problema. Hablaremos ahora de esta segunda, dejando la alternativa numérica para más adelante.

La aproximación más adecuada es la llamada “de acoplamiento fuerte”, que no es muy correcta para el ADN pero que aún así nos será útil, como veremos. La ecuación que tenemos que resolver para las autofunciones y los autovalores es

$$\int_{-\infty}^{\infty} dx \exp \left\{ -\beta \left[ \frac{1}{2} K(x-y)^2 + V(y) \right] \right\} \phi(x) = \exp(-\beta\epsilon) \phi(y). \quad (46)$$

Si  $K$  es grande, cuando  $y$  difiere de  $x$ , el término  $K(x-y)^2$  crece muy rápidamente, por lo que la integral está dominada por los valores de  $x$  cercanos a  $y$ . Definamos entonces  $x \equiv y + z$  y desarrollemos  $\phi(y+z)$  en potencias de  $z$ :

$$\begin{aligned} \int_{-\infty}^{\infty} dz \exp \left\{ -\beta \left[ \frac{1}{2} K z^2 \right] \right\} \left[ \phi(y) + z\phi'(y) + \frac{1}{2} z^2 \phi''(y) \right] \\ = \exp(-\beta[\epsilon - V(y)]) \phi(y). \end{aligned} \quad (47)$$

Ahora podemos hacer las integrales gaussianas en  $z$ ; la impar se anula y tenemos

$$\sqrt{\frac{2\pi}{\beta K}} \left[ \phi(y) + \frac{1}{2\beta K} \phi''(y) \right] = \exp(-\beta[\epsilon - V(y)]) \phi(y), \quad (48)$$

o, reorganizando los términos,

$$\left[ \phi(y) + \frac{1}{2\beta K} \phi''(y) \right] = \exp \left\{ -\beta \left[ \epsilon + \frac{1}{2\beta} \ln \left( \frac{2\pi}{\beta K} \right) - V(y) \right] \right\} \phi(y). \quad (49)$$

Llamemos  $\tilde{\epsilon} \equiv \epsilon + (1/2\beta) \ln(2\pi/\beta K)$ . Si  $y > 0$ , es decir, cuando las cadenas de la molécula de ADN se separan, que es el proceso que estamos estudiando, el potencial de Morse está acotado por la constante  $D$ . Cuando  $\beta D < 1$ , podemos desarrollar la exponencial, obteniendo

$$\left[ \phi(y) + \frac{1}{2\beta K} \phi''(y) \right] = [1 + \beta V(y) - \beta \tilde{\epsilon}] \phi(y). \quad (50)$$

Llegamos así a una ecuación que es formalmente igual a una ecuación de Schrödinger para una partícula en un potencial de Morse,

$$-\frac{1}{2K} \frac{d^2 \phi(y)}{dy^2} + \beta^2 V(y) \phi(y) = \beta^2 \tilde{\epsilon} \phi(y), \quad (51)$$

cuya solución fue obtenida por el propio Morse.

En cualquier caso, dado que hemos obtenido esta ecuación como una aproximación de nuestro problema de autovalores, no merece la pena detenerse en las soluciones exactas, y algunas ideas generales sobre ecuaciones de Schrödinger, que se pueden encontrar sin dificultad en libros de mecánica cuántica, (como por ejemplo [29]) nos serán suficientes. El problema (51) tiene en general un espectro que consiste en uno o más estados normalizables (espectro discreto) para  $\beta^2 \tilde{\epsilon} < D$  y un espectro continuo (de autofunciones no normalizables) para  $\beta^2 \tilde{\epsilon} > D$ . Lo interesante es que, al contrario del caso en que  $V(y)$  es un pozo cuadrado de extensión finita, el espectro continuo sólo existe si  $V(y)$  es lo bastante profundo. Esto se traduce en que:

- A bajas temperaturas, el mínimo de  $V(y)$ , dado por  $\beta^2 D$ , es grande (recordemos que  $\beta = 1/k_B T$ , siendo  $T$  la temperatura), y tenemos al menos un punto en el espectro discreto. Por tanto, recordando la expresión (44),  $\langle y \rangle$  es finito, y la molécula no está desnaturalizada.
- Al aumentar la temperatura, el mínimo del potencial decrece (en valor absoluto) y a un cierto valor crítico de la temperatura el espectro discreto desaparece, con lo que  $\langle y \rangle$  diverge y la molécula está totalmente separada.

Es importante observar que este resultado, la existencia de una temperatura que separa la fase normal de la fase desnaturalizada, no depende de la forma exacta de  $V(y)$  mientras que sus características básicas respondan a la forma del potencial de Morse. Ello da más validez a nuestro resultado, ya que no conocemos exactamente el potencial que actúa entre los pares de bases y utilizamos el de Morse como una aproximación cualitativa. Para el caso concreto de Morse, se puede ir más allá y calcular la temperatura crítica, que resulta ser

$$T_c = \frac{2\sqrt{2KD}}{ak_B}. \quad (52)$$

## 6.2. Mejoras del modelo

Acabamos de ver que el modelo PB presenta, dentro de las aproximaciones realizadas, una transición de fase de desnaturalización. Este resultado tiene una

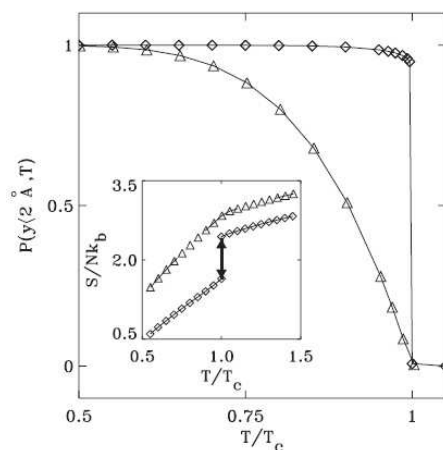


Figura 10: Resultado del cálculo aproximado para la fracción de pares de bases que permanecen cerradas en función de la temperatura para el modelo PB (triángulos) y para la versión mejorada (rombos, véase el texto). El gráfico interior muestra la variación de la entropía con la temperatura para ambos modelos.

importancia que va más allá de la mera descripción del ADN, que discutiremos en la siguiente subsección. Sin embargo, hay que insistir en que es aproximado, y en que por tanto necesitamos analizarlo con todo cuidado.

Un buen test de las aproximaciones realizadas nos lo proporciona el cálculo de la fracción de pares de bases que permanecen ligadas a una temperatura dada. Esta es una cantidad que no está explícitamente recogida en el modelo pero que se puede estimar definiendo un umbral para la separación entre bases por encima del cuál el par se considera roto. Podemos tomar  $y_0 = 2A$  como un valor razonable para la rotura de los enlaces de hidrógeno, aunque el resultado no es muy sensible a este valor.

La figura 10 muestra la dependencia de la fracción de pares de bases cerrados con la temperatura. Observamos que varía suavemente, desde 1 a la mitad de la temperatura crítica a 0 en  $T_c$ . Aquí está el problema: si bien cualitativamente esto es una transición de desnaturalización, cuantitativamente no es un modelo correcto porque la transición de desnaturalización del ADN es mucho más abrupta. Y el problema no se arregla variando los parámetros del modelo, por lo que hay que intentar mejorarlo.

Un punto que admite mejoras claras es el acoplamiento armónico entre pares de bases,  $W(y_n, y_{n-1}) = K(y_n - y_{n-1})^2/2$ . Este término proviene en realidad de un desarrollo del potencial de interacción en torno a su mínimo, es decir, para valores pequeños de  $y_n - y_{n-1}$ . Pero cuando el enlace en el par  $n$  se rompe quedando fijo el  $n-1$ , podemos tener grandes valores para esta diferencia, lo que indica que aquí podemos estar siendo demasiado groseros al modelar. Cuando

uno es más cuidadoso y tiene en cuenta el origen de esta interacción, se llega [14] a una expresión mejorada:

$$W(y_n, y_{n-1}) = \frac{1}{2}K[1 + \rho \exp(-\alpha(y_n + y_{n-1}))](y_n - y_{n-1})^2. \quad (53)$$

El significado de esta expresión es el siguiente: la interacción entre pares de bases es muy fuerte cuando ambos están unidos, debido a la colisión espacial de las nubes de electrones  $\pi$  que forman los enlaces de hidrógeno. Sin embargo, cuando uno de los pares se rompe, el prefactor hace que la interacción entre pares baje de  $K' \approx K(1 + \rho)$  a sólo  $K$ .

Para esta versión mejorada del modelo, los cálculos del apartado anterior no se pueden reproducir, debido a que el término de acoplamiento ya no es sólo función de la diferencia  $y - x$  sino también de la suma. Por tanto, el problema de autovalores se ha de estudiar numéricamente. Como ya dijimos, al discretizar en el espacio, lo que tenemos es una matriz que hay que diagonalizar. Hay que ser cuidadosos en esta discretización; notemos que el núcleo del operador es una exponencial, por lo que sus elementos nunca serán cero, y sin embargo la precisión del ordenador puede hacer que al discretizar tengamos elementos nulos en la matriz. Esto tiene importantes implicaciones (véase el apartado siguiente) y debe evitarse a toda costa. Por otro lado, para mejorar la precisión se pueden utilizar cuadraturas de Gauss-Hermite y un análisis de la dependencia de los resultados con el tamaño y el paso de la discretización [30].

Los resultados de este cambio en el modelo se pueden ver también en la figura 10. La transición de desnaturalización pasa a ser abrupta; en términos clásicos de mecánica estadística, es de primer orden. Esto está en buen acuerdo con las observaciones experimentales. En realidad, si somos un poco más cuidadosos, se ve que el salto no es discontinuo, sino suave, aunque sucede en un intervalo de temperaturas muy estrecho, lo que se suele llamar transición de cuasi-primer orden. Además, la entropía, recogida en el gráfico interior de la figura 10, muestra también un salto, que corresponde a la aparición de calor latente en la transición. La versión del modelo con la interacción entre pares de bases vecinos modificada proporciona, pues, una descripción muy aproximada de la transición de desnaturalización.

### 6.3. ¿Son posibles las transiciones de fase unidimensionales?

Independientemente de su mayor o menor grado de acuerdo con los experimentos, tanto el modelo PB original [13] como el mejorado por Dauxois [14] presentan transiciones de fase. Este resultado es muy interesante porque contrasta vivamente con lo que se suele enseñar en los cursos básicos de mecánica estadística: los modelos en una dimensión espacial con interacciones de corto alcance (en nuestro caso, sólo con los vecinos próximos) no pueden tener transiciones de fase. Dado que esto constituye un problema matemático de gran calibre, parece adecuado dedicarle una breve discusión aunque ello nos lleve a un contexto más general que el de los modelos de ADN.

La definición rigurosa de transición de fase es un concepto puramente matemático: es una no-analiticidad de la energía libre (43). Como mencionamos al introducir estos conceptos, la necesidad de tener una no-analiticidad es la que obliga a tomar el límite termodinámico  $N \rightarrow \infty$ , ya que de lo contrario la función de partición (37), de la que se deriva la energía libre, es una suma finita de exponenciales y es necesariamente analítica. El problema de decidir si un modelo tiene o transición de fase está entonces planteado de una manera no ambigua en el límite termodinámico.

La creencia general de que en una dimensión espacial las transiciones de fase son imposibles se remonta a un teorema que van Hove demostró en 1950 [31] (es también muy conocido un argumento debido a Landau [32], pero que no es riguroso matemáticamente, por lo que no nos detendremos a discutirlo aquí). El trabajo de van Hove se refiere a un sistema de segmentos unidimensionales que no pueden solaparse ni permutar posiciones, que interactúan sólo con sus vecinos inmediatos, y sólo hasta una cierta distancia. Con estas premisas, van Hove escribe la función de partición en términos de un operador de transferencia, al igual que hicimos para el modelo PB, y también igual que antes llega al resultado de que en el límite termodinámico la contribución relevante es la del mayor autovalor. Utilizando la teoría de operadores integrales de Fredholm es entonces capaz de probar que ese autovalor es analítico en la temperatura de forma rigurosa.

El teorema de van Hove es totalmente correcto, pero su mal uso por la comunidad de mecánica estadística ha conducido a la creencia de que es válido para cualquier sistema unidimensional. No es la única vez que ha pasado esto: el mal llamado teorema de localización de Anderson, sobre autofunciones de hamiltonianos con desorden, tiene una historia similar (véase [33] para una discusión accesible de este problema). En realidad, el teorema de van Hove no dice nada sobre el modelo PB porque éste no cumple varias de sus hipótesis, de entre las cuales destaca el hecho de que no sólo hay interacción entre vecinos sino que todos los pares están sometidos al potencial de Morse. Es, pues, posible, que el modelo PB tenga una transición de fase.

Recientemente, se ha probado el resultado más general conocido sobre estos problemas, que va incluso más allá de sistemas unidimensionales, y es válido para aquellos cuya función de partición se puede escribir mediante operadores de transferencia. Sin entrar en detalles técnicos, que se pueden encontrar en [34] el teorema afirma que si el operador de transferencia es lineal, su máximo autovalor es analítico en la temperatura, prohibiendo así la transición de fase en el límite termodinámico. Dado que los resultados numéricos de [30] apuntan muy seriamente a un cruce de autovalores (y por tanto a una pérdida de analiticidad) en el modelo de PB, está claro que debe de incumplir también alguna de las hipótesis de este teorema más general. Éste es todavía un problema abierto de gran interés, que necesita un estudio serio en términos de análisis funcional que lo resuelva de manera definitiva.

## 7. Conclusión

En este artículo hemos revisado con cierto detalle la dinámica no lineal y la mecánica estadística del modelo PB del ADN. Habiendo llegado hasta aquí, podemos concluir que el modelo PB es un gran éxito: Con dos ingredientes básicos, la interacción a próximos vecinos a lo largo de cada una de las dos cadenas y el potencial de Morse para la interacción entre bases de un par, reproducimos mucha fenomenología del ADN. Así, hemos encontrado analíticamente soluciones tipo *breather* que recuerdan tanto a las burbujas de desnaturalización como a las aperturas espontáneas. Tenemos otras soluciones analíticas que describen la separación entre una zona donde la doble cadena está cerrada y otra donde está abierta, análogos del proceso de replicación de la molécula. Hemos podido obtener resultados analíticos también sobre las propiedades térmicas de la molécula, conectados con la transición de desnaturalización, y que incluso nos han permitido apuntar alguna primera mejora del modelo. No sería razonable pedir más a un modelo tan extremadamente simplificado.

Está claro, efectivamente, que el modelo deja fuera prácticamente todo lo que sabemos sobre la molécula. Para empezar, deja fuera la geometría. Si quisieramos incluirla, al menos la llamada estructura secundaria, que en el caso de la doble cadena es la configuración helicoidal, tendríamos que incluir al menos dos grados de libertad por par de bases. Habría que considerar la distancia entre bases como hasta ahora, pero también el ángulo que forma su enlace respecto a un cierto eje de referencia. Se han dado algunos pasos en esta dirección [35, 36], pero el problema se complica sobremanera desde el punto de vista analítico y no es mucho lo que se puede llegar a obtener. Ir aún más allá de la doble hélice e intentar describir los distintos niveles superiores de plegamientos de la molécula (recordemos que en cada una de nuestras células tenemos 2 metros de ADN, luego debe estar bastante bien plegado) se antoja una tarea imposible. No hablaremos, pues, de incluir el ambiente en que se mueve la molécula en la célula, interaccionando con muchas otras moléculas distintas y, sobre todo, con el agua, que aprovecha cualquier oportunidad para colarse en el interior de la doble cadena.

Pero si de algo nos hemos olvidado en este trabajo es del hecho crucial de que el ADN contiene la información genética y es, por tanto, un sistema inhomogéneo. Todo lo que hemos estudiado sobre el modelo PB ha sido bajo la hipótesis de que todos los pares de bases eran iguales, y ciertamente es así en muchas moléculas que se fabrican para propósitos específicos, pero de ninguna manera en nuestras células. Si quisieramos tener en cuenta esta inhomogeneidad la complicación de los cálculos analíticos crece muy rápidamente (no sólo en el modelo PB, sino en general en cualquier modelo no lineal [37]), y de nuevo es poco lo que se puede hacer. Hay algún resultado de tipo perturbativo en modelos más simplificados aún que el PB, como es el caso del modelo de sine-Gordon, citado en la sección 2.2 [11, 38], pero es poco lo que se puede hacer analíticamente hasta el final y se termina evaluando expresiones numéricamente. Para colmo de males, los resultados no son alentadores en cuanto a la utilidad del

modelo para describir la funcionalidad del ADN inhomogéneo, y el modelo PB, que podría solventar las dificultades, se resiste al tratamiento analítico. Desde el punto de vista numérico, y más mecano-estadístico que dinámico, sí se han producido algunas novedades relevantes recientemente, relativas a la utilización del modelo para identificar secuencias de bases con funciones específicas (como por ejemplo los promotores) [39], y a su capacidad para describir experimentos sobre secuencias con propiedades especiales [40] que animan a proseguir los esfuerzos analíticos.

Para concluir estas notas, me parece apropiado un comentario general sobre la filosofía del modelado, que surge en cuanto se discuten los resultados que hemos venido resumiendo con investigadores cercanos al ADN, biólogos o genetistas, por ejemplo. El modelo PB, desde ese punto de vista, es, cuando menos, risible. No tiene química, no tiene geometría, no tiene genética, no tiene ambiente celular, no tiene... ¿Cuál es entonces el interés de trabajar con este modelo? Creo que, respetando al máximo la opinión de esos otros investigadores, el modelo PB aporta muchas cosas importantes. ¿Qué pretendíamos con él? Como dijimos en la introducción, queríamos entender los mecanismos responsables de los cambios conformacionales de la molécula de ADN, dados por movimientos de gran amplitud de las bases. Está claro que si un ordenador del futuro nos permitiera incluir todo lo que sabemos sobre la molécula en nuestro modelo, tendríamos una simulación perfecta de esos cambios. Pero, ¿habríamos aprendido algo? La respuesta ha de ser un rotundo no. Por el contrario, el modelo PB sí nos enseña algo: sabemos ahora que el que los enlaces de hidrógeno entre los pares de bases pierdan su efectividad en cuanto las bases se separan un tanto, y que el esqueleto de azúcar fosfato haga que los pares de bases vecinos estén bastante unidos, es suficiente para observar los movimientos que nos interesan. Por supuesto, los detalles no son correctos. Las aperturas espontáneas no tendrán exactamente el mismo número de bases que nuestros *breathers*, y la replicación viene gobernada por la ADN-polimerasa ligada a la zona donde se está abriendo la doble hélice. Pero sí tenemos una comprensión de la mecánica básica de la molécula, y, lo que es más importante, hemos podido ir más allá de la mera simulación numérica. Hemos obtenido unos resultados analíticos que facilitan la interpretación de las propiedades del modelo, a la vez que nos permiten, por su comparación con las simulaciones, confiar en éstas cuando no podemos tener expresiones analíticas. Consideraciones similares se aplican a toda la discusión sobre las transiciones de fase. Creo entonces que no es aventurado insistir en algo que ya he dicho: el modelo de PB es un éxito en tanto en cuanto consigue plenamente sus objetivos. Y, además, si por el camino nos hemos encontrado con matemática interesante, ¿qué más podemos pedir?

## Agradecimientos

Agradezco a Renato Álvarez Nodarse la invitación a impartir las lecciones en las que se basa este trabajo, dentro del Curso de Verano sobre “Métodos matemáticos en Biología: un viaje de ida y vuelta”, celebrado en la Universidad

Internacional de Andalucía en Baeza, en agosto de 2005. Durante dicho curso aprendí mucho sobre ADN con Antonio Marín y José L. Oliver, a los que también estoy muy agradecido. Por otro lado, la mayor parte de lo que sé sobre ADN lo debo a mi colaboración con Saúl Ares, Sara Cuenda y Niurka R. Quintero; gracias a todos ellos. Gracias por último a José A. Cuesta, con el que trabajé en el problema de las transiciones de fase unidimensionales.

Este trabajo está financiado por los proyectos MOSAICO y NAN2004-9087-C03-03, del Ministerio de Educación y Ciencia, y SIMUMAT-CM de la Comunidad de Madrid.

## Referencias

- [1] J. E. Cohen. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PLoS Biol.* 2(12):e439, 2004.
- [2] R. Álvarez-Nodarse. Modelos matemáticos en biología: un viaje de ida y vuelta. *Bol. Soc. Esp. Mat. Apl.* 35:73–112, 2006.
- [3] A. Neumaier. Molecular modeling of proteins and mathematical prediction of protein structure. *SIAM Rev.* 39:407–460, 1997.
- [4] L. V. Yakushevich. *Nonlinear Physics of DNA*. 2nd edition, Wiley, 2004.
- [5] R. M. May. Uses and abuses of mathematics in biology. *Science* 303:790–793, 2004.
- [6] C. R. Calladine y H. R. Drew. *Understanding DNA: The Molecule and How it Works*. Academic Press, 1992.
- [7] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl y A. Litwin. Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. *Proc. Nat. Acad. Sci. USA* 77:7222–7226, 1980.
- [8] J. Rubinstein. Sine-Gordon equation. *J. Math. Phys.* 11:258–266, 1970.
- [9] M. Frank-Kamenenskii. Physicists retreat again. *Nature* 328:108, 1987.
- [10] M. Salerno. Discrete model for DNA dynamics. *Phys. Rev. A* 44:5292–5297, 1991.
- [11] S. Cuenda, A. Sánchez y N. R. Quintero. Does the dynamics of sine-Gordon solitons predict active regions of DNA? *Physica D* 223:214–221, 2006.
- [12] G. Gaeta. Results and limitations of the soliton theory of DNA transcription. *J. Biol. Phys.* 24:81–96, 1999.
- [13] M. Peyrard y A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.* 62:2755–2758, 1989.



- [14] T. Dauxois, M. Peyrard y A. R. Bishop. Entropy-driven DNA denaturation. *Phys. Rev. A* 47:R44–R47, 1993.
- [15] S. Ares y G. Kalosakas. Distribution of bubble lengths in DNA. *Nano Letters* 7:311–316, 2007.
- [16] C. Bustamante, J. Marko, E. Siggia y B. Smith. Entropic elasticity of  $\lambda$ -phage DNA. *Science*, 265:1599–1600, 1994.
- [17] A. Campa y A. Giansanti. Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNA chains. *Phys. Rev. E* 58:3585–3588, 1998.
- [18] T. Schlick. *Molecular Modeling and Simulation*, pp. 435–438. Springer, 2002.
- [19] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511–519, 1984.
- [20] M. Remoissenet. Low amplitude breather and envelope solitons in quasi-one-dimensional systems. *Phys. Rev. B* 33:2386–2392, 1986.
- [21] A. C. Scott, ed. *Encyclopedia of Nonlinear Science*. Routledge, 2005.
- [22] G. L. Lamb. *Elements of Soliton Theory*. Wiley-Interscience, 1980.
- [23] T. Dauxois, N. Theodorakopoulos y M. Peyrard. Modulational instability: First step toward energy localization in nonlinear lattices. *Nonlinearity* 10:617–630, 1997.
- [24] O. Bang y M. Peyrard. Generation of high energy localized vibrational modes in nonlinear Klein-Gordon lattices. *Phys. Rev. E* 53:4143–4152, 1996.
- [25] S. Cuenda y A. Sánchez. On the discrete Peyrard-Bishop model of DNA: Stationary solutions and stability. *Chaos* 16:23123 1–8, 2006.
- [26] N. Theodorakopoulos, M. Peyrard y R. S. MacKay. Título. *Phys. Rev. Lett.* 93:258101 1–4, 2004.
- [27] D. C. Mattis. *Statistical Mechanics Made Simple*. World Scientific, 2003.
- [28] J. Ortín y J. M. Sancho. *Curso de Física Estadística*. Universitat de Barcelona, 2001.
- [29] A. Galindo y P. Pascual. *Mecánica Cuántica*. Alhambra, 1978.
- [30] N. Theodorakopoulos. Thermodynamic instabilities in one-dimensional particle lattices: a finite size scaling approach. *Phys. Rev. E* 68:026109 1–4, 2003.

- [31] L. van Hove. Sur l'intégrale de configuration pour les systèmes de particules à une dimension. *Physica* 16:137–143, 1950.
- [32] L. D. Landau y E. M. Lifshitz. *Statistical Physics, Part 1*. Pergamon, 1980.
- [33] E. Diez, A. Sánchez y F. Domínguez-Adame. Conjetura que algo queda: el “teorema” de localización de Anderson. *Rev. Esp. Fís.* 13 (4):23-26, 1999.
- [34] J. A. Cuesta y A. Sánchez. General non-existence theorem for phase transitions in one-dimensional systems with short-range interactions, and physical examples of such transitions. *J. Stat. Phys.* 115:869-893, 2004.
- [35] M. Barbi, S. Cocco y M. Peyrard. Helicoidal model for DNA opening. *Phys. Lett. A* 253:358–369, 1999.
- [36] B. Sánchez-Rey, J. F. R. Archilla, F. Palmero y F. Romero. Breathers in a system with helicity and dipole interaction. *Phys. Rev. E* 66:017601 1–4, 2002.
- [37] A. Sánchez y L. Vázquez. Nonlinear wave propagation in disordered media. *Int. J. Mod. Phys. B* 5:2825-2882, 1991.
- [38] S. Cuenda y A. Sánchez. Nonlinear excitations in DNA: Aperiodic models vs actual genome sequences. *Phys. Rev. E* 70:51903 1–8, 2004.
- [39] C. H. Choi, G. Kalosakas, K. O. Rasmussen, M. Hiromura, A. R. Bishop y A. Usheva. DNA dynamically directs its own transcription initiation. *Nuc. Acids Res.* 32:1584–1590, 2004.
- [40] S. Ares, N. K. Voulgarakis, K. O. Rasmussen y A. R. Bishop. Bubble nucleation and cooperativity in DNA melting. *Phys. Rev. Lett.* 94:035504 1–4, 2005.