



UNIVERSIDAD CARLOS III DE MADRID
DEPARTAMENTO DE MATEMATICAS

TESIS DOCTORAL

**Secuencia genética y dinámica
de excitaciones no lineales de ADN**

Autora

Sara Cuenda Cuenda

Director

Ángel Sánchez Sánchez

Leganés, Abril 2007

*A mi hermano, por hacer que me interese la ciencia.
A mi padre, por hacerme entenderla.
A mi madre, por apoyarme en todo lo que hago.*

A mi familia.

Agradecimientos

Quiero agradecer esta tesis en primer lugar a mi director, Ángel Sánchez. Aunque las comparaciones son odiosas, a veces son necesarias para darnos cuenta de la calidad de las personas con las que trabajamos. En este sentido Anxo ha sido un magnífico director de tesis, capaz de guiarme, interesarse por mí, darme libertad en lo que hacía, o también la lata para que trabajara, según fuera necesario. Ha sido un privilegio poder trabajar con él en el transcurso de estos años de realización de esta tesis doctoral, y espero poder seguir colaborando y tratando con él en el futuro.

Igualmente creo necesario agradecer a todos los miembros del GISC, aunque particularmente a mis compañeros de la Carlos III, la gran acogida que me dieron y el buen ambiente que han creado, que ha hecho que seamos la envidia de los becarios del Departamento de Matemáticas de la Universidad Carlos III de Madrid. Gracias por atenderme cuando he necesitado ayuda, por dar y soportar las misas semanales y por no haber cogido a ninguna otra becaria en estos cinco años.

También quiero agradecer a los becarios de la Carlos III el buen rato que me han hecho pasar. A Saúl, por la acogida que me dio al empezar el doctorado. A María José, Luis y Nacho por las charlas (in)trascendentes en momentos de baja productividad laboral pero de grandes logros en el entorno social. A los demás, por el tiempo que hemos pasado juntos en las comidas y en alguna merienda: a Matteo, Paco, Alfredo, Roberto, Marina (qué pocos quedamos ya). A los que ya se fueron después de ser doctores. A mis compañeros y al profesor de Tai-Chi, gracias a los cuales me enganché a este (semi)deporte que me ha mantenido más o menos derecha en los últimos años, y a los que abandoné despiadadamente durante la elaboración de esta memoria.

Y por último, quiero agradecer muy especialmente a mi ya marido José Ángel por animarme a empezar esta tesis y por apoyarme y acompañarme en los momentos de debilidad. Gracias por estos años que me has regalado.

A todos ellos les debo la madurez (no digo que sea ni mucha ni poca) que me ha permitido acabar esta tesis. Que lo sepan.

A handwritten signature in blue ink, appearing to read 'Sara Cuenda', enclosed within a light blue oval shape.

Sara Cuenda.

Índice general

1. Introducción	1
I sG continuo	7
2. El modelo homogéneo de sine-Gordon en el continuo.	9
2.1. Introducción histórica del modelo.	9
2.2. Solitones y ondas solitarias.	14
2.3. <i>Kinks, breathers</i> y otras soluciones.	17
3. Perturbaciones inhomogéneas.	27
3.1. Ecuación de Klein-Gordon perturbada.	27
3.2. Estudio con una coordenada colectiva.	29
3.3. Estudio con dos coordenadas colectivas.	34
3.4. La importancia de la anchura del <i>kink</i>	37
II sG discreto	45
4. El modelo homogéneo de sine-Gordon discreto.	47
4.1. El modelo de Frenkel-Kontorova.	47
4.2. Modelo mecánico.	49
4.3. Efectos de la discretización.	53
4.4. El modelo de Englander del ADN.	56
5. Modelo de Salerno del ADN.	59
5.1. Secuencias de ADN.	59
5.2. Secuencias aperiódicas.	65

5.3. El potencial efectivo, V_{eff}	72
6. Aplicabilidad del modelo de Salerno del ADN	85
6.1. Secuencias aperiódicas.	85
6.2. Introducción de ruido.	93
6.3. Secuencias de ADN.	97
6.4. Trabajos posteriores sobre el modelo de Salerno.	101
III Modelo de PB	111
7. El modelo de Peyrard-Bishop del ADN	113
7.1. Introducción del modelo.	114
7.2. Estudio analítico de la pared de dominio.	118
IV Conclusión	133
8. Conclusiones y problemas abiertos.	135
V Apéndices	141
A. Wobbles en el modelo de sG.	143
A.1. Integración de TB a partir de $\phi = 0$	145
A.2. Construcción de un <i>breather</i>	146
A.3. Superposición no lineal de un <i>kink</i> y medio <i>breather</i>	147
A.4. Obtención del <i>wobble</i>	148
B. Nociones de biología y biofísica.	151
B.1. Estructura del ADN	151
B.2. Transcripción del ADN.	154
B.3. El genoma del fago T7.	158
B.4. Algunas cosillas de biofísica.	161
B.5. La base de datos del NCBI.	163
Publicaciones	167
Referencias	169

1

Introducción

Motivación.

Esta tesis doctoral comenzó buscando una respuesta a la siguiente pregunta: ¿Es posible explicar determinados comportamientos dinámicos del ADN mediante un modelo no lineal, determinista, y sencillo? Dichos comportamientos se refieren principalmente a los procesos de transcripción de proteínas por parte de la ARN polimerasa, aunque también engloban fenómenos de replicación del ADN, o procesos de desnaturalización mecánica del ADN. El modelo en cuestión es el modelo de sine-Gordon, un ejemplo típico de modelo no lineal muy estudiado en el pasado, como mostraré en los capítulos introductorios de esta memoria.

La pregunta anterior, aunque sencilla, es tremendamente ambiciosa; una respuesta afirmativa a esa pregunta respondería a muchos biólogos y genetistas que todavía están estudiando los entresijos de la funcionalidad del ADN mediante descripciones exhaustivas de miles de procesos en los que éste se ve involucrado. Paradójicamente, y pese a lo que casi todo el mundo podría pensar, las respuestas iniciales a este problema fueron afirmativas (Englander et al. 1980), aunque dejaron gran cantidad de cuestiones abiertas. Estos resultados, junto con otros obtenidos posteriormente (Salerno 1991; Salerno 1992; Domínguez-Adame et al. 1995), fueron los responsables de que empezara una investigación en la búsqueda de una

explicación a esta aparente relación entre excitaciones no lineales en el modelo de sine-Gordon discreto y codificación de la secuencia genética. Como resultado de este acercamiento inicial publicamos dos artículos (Cuenda y Sánchez 2004b; Cuenda y Sánchez 2004a) que abordan el problema de la dinámica de *kinks* en secuencias genómicas. En ambos trabajos se empieza a percibir una respuesta negativa a la pregunta inicial. Sin embargo, en ellos también se obtiene un resultado optimista que proviene de la utilización del potencial efectivo en secuencias inhomogéneas (Salerno y Kivshar 1994). Dicho potencial efectivo, tal y como se utilizó en sus orígenes, parecía dar una descripción cualitativa del movimiento de *kinks* en secuencias inhomogéneas; en nuestros trabajos resultó ser un gran aliado en el análisis de las secuencias consideradas, permitiéndonos, con el tiempo, analizar las secuencias *a priori* de manera intuitiva, antes de estudiar la dinámica de *kinks* que se mueven sobre ellas. Estos resultados nos llevaron a cambiar ligeramente el rumbo de la investigación que habíamos empezado, para llevarlo al campo de la aplicación del potencial efectivo en el modelo de sine-Gordon continuo. En este sentido obtuvimos resultados numéricos satisfactorios (Cuenda y Sánchez 2005) que apoyaban aún más la utilidad de esta herramienta y nos permitieron entender cada vez mejor la dinámica de *kinks* en el modelo de sine-Gordon inhomogéneo.

Posteriormente aparecieron algunos trabajos sobre otro modelo de ADN, el de Peyrard-Bishop (Peyrard y Bishop 1989), que relacionaban excitaciones no lineales en la doble cadena de ADN con posiciones relevantes de la secuencia en determinados organismos (Choi et al. 2004; Kalosakas et al. 2004). El modelo de Peyrard-Bishop ya había dado buenos resultados en la explicación de la desnaturalización térmica del ADN, y la conexión con elementos funcionales de secuencias genómicas sería un resultado de gran interés. Dada nuestra experiencia, por un lado, con el modelo de sine-Gordon y, por otro, con el potencial efectivo de *kinks*, comenzamos el estudio para la obtención de un potencial efectivo de soluciones estacionarias del modelo de Peyrard-Bishop. Aunque no llegamos a obtenerlo, publicamos un nuevo artículo (Cuenda y Sánchez 2006) sobre el análisis exhaustivo de estas soluciones del modelo en el límite discreto y en el límite continuo.

Poco después, motivados por resultados (Lennholm y Hörnquist 2003; Bashford 2006) que de nuevo relacionaban la dinámica de *kinks* del modelo de sine-Gordon con posiciones relevantes de la secuencia genómica de determinados organismos, y con la colaboración con Niurka R. Quintero, de la Universidad de Sevilla, nos decidimos a realizar un análisis exhaustivo del problema. Para ello, revisamos detalladamente los primeros resultados sobre el modelo (Salerno 1991; Salerno 1992) que inspiraron los trabajos posteriores (Lennholm y Hörnquist 2003;

Bashford 2006), así como estos últimos. Resultado de esta investigación es otro artículo (Cuenda et al. 2006) que creemos que pone punto y final a la búsqueda de una respuesta a la pregunta inicial y que es, lamentablemente, no: no creemos que exista relación entre la dinámica de *kinks* en el modelo discreto de sine-Gordon inhomogéneo y los elementos funcionales del genoma del que se han obtenido las secuencias inhomogéneas. En nuestro estudio creemos dar motivos suficientes que justifican esta respuesta, aunque, para ser honestos, nos hubiese gustado más un resultado en favor del modelo.

Esquema de la memoria.

Una vez vista la motivación de cada trabajo realizado durante esta tesis explicaré cómo se han organizado en esta memoria los resultados obtenidos. Puesto que es una memoria donde se le da gran importancia a resultados previos en los que está basado el trabajo de esta tesis, haré énfasis en los trabajos originales que se han desarrollado durante su elaboración para que estos queden claramente diferenciados. La memoria se divide en tres partes: la que se refiere al modelo de sine-Gordon continuo; la segunda, relativa al modelo de sine-Gordon discreto; y la última, que trata el modelo de Peyrard-Bishop aplicado al ADN.

La primera parte engloba los capítulos 2 y 3. El segundo capítulo es puramente introductorio, y trata del modelo de sine-Gordon homogéneo y continuo. En él se hace un repaso de la ecuación de sine-Gordon en coordenadas características en la geometría diferencial y se introduce la transformación de Bäcklund como método para obtener soluciones nuevas a partir otras conocidas. Posteriormente se introduce la ecuación de sine-Gordon en coordenadas estándar a partir de un Hamiltoniano, desde el punto de vista de la mecánica analítica, y se menciona la transformada espectral inversa como método para obtener soluciones de la ecuación. También se da una idea de lo que es un solitón, y de en qué se diferencia de las ondas solitarias, así como su relación con sistemas completamente integrables. Para ello se citan algunas diferencias entre la ecuación del modelo de sine-Gordon, que posee soluciones de tipo solitón, y el modelo ϕ^4 , que no posee solitones pero sí soluciones de tipo onda solitaria, aunque ambos son modelos de Klein-Gordon no lineales. Finalmente, se obtienen las principales soluciones de la ecuación de sine-Gordon que se utilizarán en la memoria: *kinks*, *breathers* y fonones.

En el tercer capítulo se introducen perturbaciones inhomogéneas en modelos de Klein-Gordon no lineales, como sine-Gordon y ϕ^4 . Posteriormente se muestra la técnica de coordenadas colectivas aplicada a la dinámica del centro del *kink*, se explican resultados anteriores para el caso particular de una inhomogeneidad

de tipo coseno y se detalla la obtención del potencial efectivo. Posteriormente se introduce la anchura del *kink* como segunda coordenada colectiva y se analiza el caso no perturbado. En la última sección se describen los resultados originales de esta tesis (Cuenda y Sánchez 2005) que dan cuenta de la importancia de la anchura del *kink* en el análisis de la perturbación inhomogénea de tipo coseno y su influencia en el fenómeno de competición de escalas.

La segunda parte consta de los capítulos 4, 5 y 6. El capítulo cuarto es introductorio. Comienza con la introducción del modelo de Frenkel-Kontorova que, aunque no es más que el modelo de sine-Gordon discreto, tiene gran valor por ser el redescubrimiento de la ecuación en el ámbito de la física del estado sólido a finales de los años 20. Posteriormente se introduce el sistema de la cadena de péndulos acoplados, también descritos por la ecuación de sine-Gordon, el significado de la discretización efectiva en el modelo en términos físicos, y los efectos que ésta produce al comparar con resultados del modelo continuo. Posteriormente se introduce el modelo de Englader del ADN como analogía al modelo de péndulos descrito anteriormente.

El capítulo quinto comienza introduciendo el modelo de Salerno del ADN, en el que por primera vez se introduce la secuencia genómica para la búsqueda de elementos biológicamente funcionales en ésta a partir de la dinámica de *kinks*. Se realiza una revisión muy exhaustiva de los trabajos de Salerno en promotores en el fago T7 (Salerno 1991; Salerno 1992), en los que se relaciona positivamente la dinámica de *kinks* y los promotores de una secuencia de ADN. También se realiza una revisión parecida con un trabajo posterior de dinámica de *kinks* en secuencias aperiódicas (Domínguez-Adame et al. 1995), que da la posibilidad de relacionar el contenido de información de una secuencia con ciertos aspectos de la dinámica de *kinks*. Posteriormente se introduce el potencial efectivo para *kinks* en modelos discretos, tal y como fue definido por Salerno y Kivshar (1994), y se discute su utilización por parte de los autores de los trabajos anteriores para explicar sus resultados. Posteriormente, como trabajo original de la tesis, se analiza el significado del potencial efectivo, se propone una nueva expresión para él y se realiza un análisis comparativo entre la dinámica de *kinks* en secuencias inhomogéneas y la dinámica de una partícula que se mueve de acuerdo al potencial efectivo.

En el capítulo sexto se describen en detalle varios trabajos originales de la tesis en los que se revisan los resultados obtenidos por los autores mencionados en el capítulo anterior. En la primera parte del capítulo se explican los resultados obtenidos por Domínguez-Adame et al. (1995), basándonos en la importancia que tiene la forma del potencial efectivo alrededor de la posición inicial del *kink*, así como su estructura general de picos y pozos, y se repite un análisis similar

al realizado por los autores, utilizando en este caso secuencias de ADN en lugar de secuencias periódicas (Cuenda y Sánchez 2004b). A continuación se hace un estudio de la influencia del ruido en los resultados obtenidos, tanto de ruido térmico, con simulaciones de dinámica Langevin, como de condiciones iniciales con ruido y dinámica determinista (Cuenda y Sánchez 2004a). También se explican los resultados de Salerno (Salerno 1991; Salerno 1992) en términos de las secuencias que utiliza y se comparan los resultados que se obtienen con las secuencias “buenas”. Finalmente, se hace un estudio de todos los promotores del fago T7 en función del potencial efectivo a raíz de un estudio estadístico de la dinámica de *kinks* (Lennholm y Hörnquist 2003) y se discute la posibilidad de un potencial efectivo para *breathers*, como el propuesto por Bashford (2006).

El séptimo capítulo es la tercera parte de esta memoria. En él se introduce el modelo de Peyrard-Bishop del ADN y los tipos de excitaciones que se han estudiado en él, las de tipo *breather* y las paredes de dominio. Posteriormente se describe un resultado original de esta tesis (Cuenda y Sánchez 2006) en el que se realiza un estudio analítico de las paredes de dominio en sistemas finitos y su estabilidad, tanto en el límite continuo como en el discreto, y que debe ser el punto de partida de una posible generalización del concepto de potencial efectivo.

Finalmente, en las conclusiones resumo las ideas principales de esta tesis y expongo las líneas de investigación que han quedado abiertas o en las que estamos trabajando. Además, esta tesis incluye dos apéndices: el apéndice A es un estudio realizado durante una estancia en la Universidad de Sevilla bajo la dirección de Niurka R. Quintero sobre un tipo de solución de sine-Gordon formada por tres solitones a partir de la transformación de Bäcklund, su comparación con la expresión derivada mediante la transformada espectral inversa y su significado; y en el apéndice B se encuentra toda la información biológicamente relevante que puede ser necesaria para un mejor entendimiento de la segunda parte de la memoria, y que forma parte del trabajo de campo que ha sido necesario realizar para una mejor comprensión del problema del modelado del ADN.

Agradecimientos.

Este trabajo ha sido posible gracias a los proyectos FIS2004-01001 y MOSAICO del Ministerio de Educación y Ciencia, al proyecto SIMUMAT de la Comunidad de Madrid y al proyecto UC3M-FI-05-007 de la Comunidad de Madrid y la Universidad Carlos III de Madrid. También agradezco una beca de doctorado del Departamento de Matemáticas de la Universidad Carlos III de Madrid y una beca/contrato de formación de personal investigador de la Consejería de Educación

de la Comunidad de Madrid y el Fondo Social Europeo que me han proporcionado financiación en los últimos años.



sG continuo

2

El modelo homogéneo de sine-Gordon en el continuo.

En este capítulo se introduce la ecuación de sine-Gordon continua, tanto desde el punto de vista geométrico como desde el punto de vista físico. Se presentan la transformación de Bäcklund, así como el principio de superposición no lineal dado por la identidad de Bianchi, como métodos para obtener soluciones de la ecuación a partir de otras conocidas. Se muestran las principales características de los solitones, así como de sus diferencias con ondas solitarias, y su relación con sistemas completamente integrables y transformada espectral inversa. Por último, se obtienen las expresiones de las principales soluciones que van a utilizarse a lo largo de la memoria de esta tesis doctoral.

2.1. Introducción histórica del modelo.

El modelo de sG en geometría.

La ecuación de sG apareció en la segunda mitad del siglo XIX en el campo de la geometría diferencial, más concretamente en la teoría de superficies pseudo-esféricas (Bour 1862). Se llama superficie hiperbólica a aquella que tiene curvatura negativa en toda la superficie, y superficie pseudoesférica a la que, además,

tiene la misma curvatura en cualquier punto de su superficie. Supongamos una superficie hiperbólica Σ definida por las longitudes de arco de curvas asintóticas¹ (ξ, τ) , y sea ϕ el ángulo formado por dos curvas asintóticas. Entonces, si la curvatura *total* de Σ (dada por el Teorema *Egregium* de Gauss) es $K = -1$, constante en toda la superficie (es decir, la superficie es una pseudoesfera) se obtiene que (Rogers y Schief 2002)

$$\phi_{\xi\tau} = \sin \phi. \quad (2.1)$$

Esta expresión de la ecuación de sG se denomina ecuación característica o ecuación en coordenadas características.

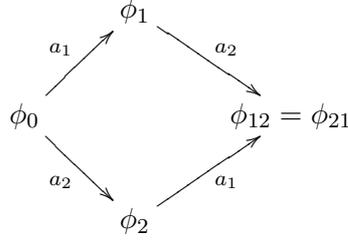
En el siglo XIX, Luigi Bianchi (Bianchi 1879) presentó, en su tesis de habilitación, una manera de construir geoméricamente superficies pseudoesféricas. Poco después, Ralf Josef Bäcklund (Bäcklund 1883) utilizó una simetría de la ecuación de sG para generalizar el resultado de Bianchi y poder construir soluciones de la ecuación (2.1) de manera iterativa con un parámetro libre. Dicha transformación, denominada transformación de Bäcklund (TB), permite generar una nueva superficie pseudoesférica a partir de otra conocida. Así, sea $\phi(\xi, \tau)$ una solución de (2.1), y sea la TB definida por el parámetro a :

$$\begin{aligned} \left(\frac{\psi - \phi}{2}\right)_{\xi} &= a \sin\left(\frac{\psi + \phi}{2}\right), \\ \left(\frac{\psi + \phi}{2}\right)_{\tau} &= \frac{1}{a} \sin\left(\frac{\psi - \phi}{2}\right). \end{aligned} \quad (2.2)$$

La expresión (2.2) define una nueva función $\psi(\xi, \tau)$ que depende de la solución ϕ y del parámetro a . Se puede demostrar fácilmente, calculando las derivadas cruzadas en (2.2), que $\psi(\xi, \tau)$ también es solución de (2.1). Puesto que la nueva solución se ha obtenido a partir de otra existente, ψ se puede entender como una transformación de ϕ ; además, la transformación se ha realizado utilizando el parámetro a , por lo que la transformación de Bäcklund se denota por \mathbb{B}_a , y se escribe $\psi = \mathbb{B}_a(\phi)$. Esta transformación es “reversible”: una vez obtenido ψ , se puede volver a la solución ϕ , salvo constante de integración, cambiando el parámetro a por $-a$ e integrando de nuevo (2.2).

¹Una curva asintótica es una curva siempre tangente a una dirección asintótica de la superficie, y una dirección asintótica es aquella en la que la curvatura normal es cero. En general no tienen por qué existir direcciones asintóticas en cada punto de una superficie (por ejemplo, en cualquier punto de una esfera). Sin embargo, en superficies hiperbólicas, esto es, de curvatura negativa en cada punto de la superficie, puesto que las dos curvaturas principales deben tener signos opuestos, debe existir una dirección en la que la curvatura normal sea cero y por tanto sea una dirección asintótica. Por tanto, todas las superficies hiperbólicas son parametrizables en arcos de longitud de curvas asintóticas.

Posteriormente, Bianchi (1892) demostró su importante *teorema de permutabilidad*. Sea ϕ_0 una solución semilla de la ecuación de sG (2.1), y sean ϕ_1 y ϕ_2 las TB de ϕ_0 via \mathbb{B}_{a_1} y \mathbb{B}_{a_2} , esto es, $\phi_1 = \mathbb{B}_{a_1}(\phi_0)$ y $\phi_2 = \mathbb{B}_{a_2}(\phi_0)$. Sean además $\phi_{12} = \mathbb{B}_{a_2}(\phi_1)$ y $\phi_{21} = \mathbb{B}_{a_1}(\phi_2)$. Entonces, $\phi_{12} = \phi_{21}$. Esto se entiende de una manera más gráfica con un *Diagrama de Bianchi*:



Este importante teorema nos permite relacionar de manera puramente algebraica las cuatro soluciones del diagrama anterior de la forma²:

$$\tan\left(\frac{\phi_{12} - \phi_0}{4}\right) = \frac{a_2 + a_1}{a_2 - a_1} \tan\left(\frac{\phi_2 - \phi_1}{4}\right). \quad (2.3)$$

De esta forma se puede despejar ϕ_{12} en función de las otras tres soluciones,

$$\phi_{12} = \phi_0 + 4 \arctan\left[\frac{a_2 + a_1}{a_2 - a_1} \tan\left(\frac{\phi_2 - \phi_1}{4}\right)\right]. \quad (2.4)$$

Dicha relación representa un principio de superposición no lineal, a diferencia de lo que ocurre en las EDPs lineales, donde se aplica el principio de superposición lineal ya conocido. Los principios de superposición no lineales generalmente van asociados a EDPs que admiten TB que llevan de una solución a otra de la misma ecuación.

El caso $a_2 = -a_1$ es un caso particular de (2.4) que se resuelva fácilmente, ya que da lugar a $\phi_{12} = \phi_0$ (salvo constante de integración), como ya vimos en la ecuación (2.2). Pero resolver el caso $a_2 = a_1$ no es tan sencillo puesto que, para esos parámetros, $\phi_2 = \phi_1$, y por tanto tenemos una indeterminación. La resolución de (2.2), en ese caso, sólo es posible cuando las constantes de integración coinciden, y la dependencia funcional de ϕ_i respecto a a_i , entonces, puede escribirse como $\phi_i(x, t) = f(a_i, x, t)$, puesto que ambos provienen del mismo ϕ_0 . Por tanto, para $a_2 = a_1$ basta con tomar el límite

$$\tan\left(\frac{\phi_{12} - \phi_0}{4}\right) = \lim_{a_2 \rightarrow a_1} (a_1 + a_2) \frac{\tan\left(\frac{f(a_1, x, t) - f(a_2, x, t)}{4}\right)}{a_1 - a_2}, \quad (2.5)$$

²Para una demostración de esta expresión véase, por ejemplo, Rogers y Schief (2002).

que, al calcularlo, se obtiene

$$\phi_{12} = \phi_0 + 4 \arctan \left(\frac{a_1}{2} \frac{\partial f(a_1, x, t)}{\partial a_1} \right). \quad (2.6)$$

Modelos completamente integrables.

Hasta ahora la introducción que he hecho al lector de la ecuación de sG ha sido fundamentalmente geométrica. Sin embargo, la ecuación de sG se ha utilizado en numerosas ocasiones para describir varios fenómenos importantes de física no lineal en campos tan variados como dislocaciones en cristales (Frenkel y Kontorova 1939), propagación de flujos magnéticos en uniones Josephson (Josephson 1965), cadenas de péndulos conectados por muelles (Scott 1969), propagación de pulsos ópticos en medios resonantes (Arecchi et al. 1969), estabilidad de movimientos de fluidos (Scott et al. 1973), en física de partículas elementales (Skyrme 1958), en ferromagnetismo (Enz 1964), en la dinámica de cadenas de ADN (el tema principal de esta tesis), etc. Debido a la gran versatilidad de esta ecuación en la Física, parece necesario introducirla también desde este punto de vista.

Así, según la mecánica analítica, otra forma de obtener la ecuación de sG es la siguiente. Sea el sistema dinámico definido por el Hamiltoniano

$$H[\phi] = \int_{-\infty}^{\infty} dx \left\{ \frac{1}{2} \phi_t^2 + \frac{1}{2} \phi_x^2 + (1 - \cos \phi) \right\}, \quad (2.7)$$

donde $\phi(x, t)$ es un campo escalar que define el estado del sistema. Las ecuaciones de movimiento del sistema se obtienen mediante las ecuaciones de Hamilton

$$p_\phi = \frac{\delta H}{\delta \dot{\phi}}, \quad \dot{p}_\phi = -\frac{\delta H}{\delta \phi}, \quad (2.8)$$

donde el punto significa derivada parcial con respecto al tiempo. El resultado es:

$$\phi_{tt} - \phi_{xx} + \sin \phi = 0, \quad (2.9)$$

que es precisamente la ecuación (2.1) bajo el cambio de coordenadas $x = \xi + \tau$ y $t = \xi - \tau$. Ésta es la expresión estándar de la ecuación.

Un aspecto muy interesante de la ecuación de sG es que es completamente integrable (CI). En dimensión finita, un sistema hamiltoniano de N variables se dice que es CI si admite N constantes de movimiento independientes entre sí. En el caso de sistemas hamiltonianos de dimensión infinita, la generalización implica que deben poseer un número infinito de cantidades conservadas independientes

entre sí. Sin embargo, pese a la dificultad de demostrar la existencia de infinitas cantidades conservadas³, esto no es suficiente para demostrar que un sistema de EDPs sea CI debido a los diferentes cardinales de los infinitos. Así que, ¿qué otras propiedades cumplen los sistemas completamente integrables? Utilizando las herramientas de mecánica analítica, el teorema de Liouville nos dice que una condición necesaria y suficiente para que un sistema sea completamente integrable es que exista una transformación canónica a las variables acción-ángulo en las que el sistema sea completamente separable y las ecuaciones de movimiento puedan entonces ser integradas mediante cuadraturas de manera trivial⁴. En sistemas de dimensión infinita, como sG, la generalización de las variables acción-ángulo y la integración “trivial” de las ecuaciones de movimiento constituye la base física de una importante herramienta matemática, que es la Transformada Espectral Inversa (TEI)⁵.

Sin embargo, pese a esta relación con los sistemas Hamiltonianos, la TEI nació de manera independiente en 1967 de las manos de Clifford S. Gardner, John M. Greene, Martin D. Kruskal, y Robert M. Miura (Gardner et al. 1967) como una herramienta matemática para resolver de manera exacta una ecuación no lineal denominada KdV⁶. En 1972, Vladimir E. Zakharov y Boris V. Shabat (Zakharov y Shabat 1972) utilizaron ese mismo formalismo para resolver la ecuación de Schrödinger no lineal. El turno a la ecuación de sG le llegaría un año después, cuando Mark J. Ablowitz, David J. Kaup, Alan C. Newell y Harvey Segur (Ablowitz et al. 1973) aplicaron el formalismo de Zakharov y Shabat de TEI a la ecuación (2.1). La conexión de la TEI con sistemas Hamiltonianos la daría Peter D. Lax (Lax 1978) de manera posterior a todo el proceso de resolución mediante TEI de las ecuaciones no lineales mencionadas.

Sin dar un análisis detallado⁷, la esencia de la TEI es relacionar la ecuación (2.1) con un par de ecuaciones de autovalores para v_1 y v_2 , lineales,

$$\begin{aligned}\frac{\partial v_1}{\partial x} + i\zeta v_1 &= -\frac{1}{2}\phi_x v_2, \\ \frac{\partial v_2}{\partial x} - i\zeta v_2 &= \frac{1}{2}\phi_x v_1.\end{aligned}\tag{2.10}$$

Dichas ecuaciones, junto con las condiciones de que v_1 y v_2 están relacionadas

³Esto, en general, no resulta fácil de demostrar. Sin embargo, en el caso de sG sí se ha demostrado la existencia de infinitas corrientes conservadas (Yoon 1976).

⁴La explicación y demostración aparece en cualquier libro de mecánica analítica.

⁵En inglés se denomina *Inverse Scattering Transform*, y sus siglas son IST.

⁶El nombre proviene de sus autores, Korteweg-deVries, y se introducirá en la sección 2.2.

⁷Para más información véase Debnath (1997), de donde procede esta información. Una explicación más detallada aparece en Lamb (1980).

en el límite $x \rightarrow \pm\infty$, se pueden considerar como un problema para determinar el autovalor ζ dada una función $q = -\frac{1}{2}\phi_x$ denominada *potencial*, y relacionada a su vez con la solución $\phi(\xi, \tau)$ de la ecuación (2.1). Tras un proceso bastante complicado, la solución final ϕ se obtiene calculando la siguiente expresión:

$$\phi(\xi, \tau) = 4 \arctan \left[\frac{\text{Im}|I - iM|}{\text{Re}|I - iM|} \right], \quad (2.11)$$

donde I es la matriz unidad de orden N , N es el número de solitones de la solución, y M es una matriz $N \times N$ con elementos de la forma

$$M_{jk}(\xi, \tau) = \frac{2m_k}{\zeta_j + \zeta_k} \exp \left[\frac{\zeta_j + \zeta_k}{2} \xi + \frac{\tau}{\zeta_k} \right], \quad (2.12)$$

donde m_i son constantes asociadas a cada solitón que compone la solución final. Dicha solución se construye de manera que, para un *kink* (*antikink*), se tiene que m_k es real y positivo (negativo); para un *breather*, compuesto por un par *kink-antikink*, se necesitan dos constantes complejas, una compleja conjugada de la otra, $m_{b_2} = m_{b_1}^*$. En coordenadas características, la parte real de m_i está relacionada con la posición en el eje real, x_0 , mientras que la parte imaginaria nos da una traslación temporal, t_0 , de cada solitón. Por otro lado, el autovalor ζ_k asociado a un *kink* o un *antikink* es real y positivo, mientras que un *breather* necesita dos autovalores (recordemos que está compuesto por dos solitones), uno complejo conjugado del otro, es decir, $\zeta_{b_2} = \zeta_{b_1}^*$; para conseguir que estén en reposo, es necesario además que $|\zeta_i| = 1$, tanto para *kinks* o *antikinks* como para *breathers*.

2.2. Solitones y ondas solitarias.

Introducción histórica.

Otra propiedad muy importante de la ecuación de sG es que posee soluciones de tipo solitón. Para aclarar ideas, primero intentaré explicar qué es un solitón y qué es una onda solitaria.

Tenemos por un lado el archiconocido “suceso” acontecido a John Scott Russell en el canal de Edinburgo a Glasgow en 1834, en el que observó una gran onda de agua aislada que avanzó varios kilómetros sin cambiar su forma, y a la que llamó “gran onda de traslación” (Russell 1844). Esto nos describe bastante bien el concepto de *onda solitaria*: una joroba, simétrica y aislada, que se mueve a velocidad constante pero que no cambia de forma con el tiempo. Esta idea es claramente diferente del concepto usual de *onda periódica* o *tren de ondas*, típico

de sistemas lineales. Tendrían que pasar más de 60 años para que los holandeses D. J. Korteweg y G. de Vries formularan un modelo matemático que diera explicación al fenómeno observado por Scott Russell (Korteweg y de Vries 1895). Ellos derivaron la famosa ecuación KdV de propagación de ondas en una dirección en la superficie del agua,

$$\phi_t + \phi\phi_x + \phi_{xxx} = 0. \quad (2.13)$$

En esta ecuación no lineal, que describe la evolución temporal a largos tiempos del fenómeno de ondas, existe un balance entre el término no lineal, que tiende a localizar la joroba, y el término dispersivo, que tiende a ensancharla. La ecuación KdV representa el paradigma de los modelos de ecuaciones con ondas solitarias, con un término que tiende a dispersar o ensanchar la solución y otro que tiende a localizarla. En 1965, Martin Kruskal y Norman Zabusky hicieron un estudio numérico de dicha ecuación partiendo de una condición inicial de onda periódica, y observaron cómo la onda se iba deformando hasta llegar a convertirse en una serie de jorobas bien definidas de la forma sech^2 de diferentes alturas y diferentes velocidades, las más altas las más veloces y las más bajas las más lentas, que interactuaban entre sí y luego emergían sin cambio de forma ni amplitud salvo una pequeña traslación. Debido a estas propiedades que asemejaban estos objetos a partículas, Kruskal y Zabusky llamaron a estas ondas solitarias *solitones* (como fonones, protones, etc.).

Propiedades de solitones.

Desde la introducción de los solitones en la literatura científica ha habido una gran cantidad de estudios, tanto analíticos como numéricos, del comportamiento de estos objetos en la ecuación KdV⁸ y también en otros modelos de ecuaciones no lineales que poseen propiedades similares. Han sido necesarias diversas ramas de matemática pura y aplicada para explicar muchas de las novedosas propiedades que han aparecido. Sin embargo, no es fácil dar una definición precisa de lo que es un solitón. Se puede decir que este término se asocia con una solución de una EDP no lineal (o un sistema de ellas) que: (i) representa una onda cuya forma no varía con el tiempo, (ii) está localizada, de manera que decae o se aproxima a un valor constante en el infinito⁹, y (iii) puede interactuar fuertemente con otros

⁸Algunas referencias a estudios de este tipo se pueden encontrar en: Zabusky y Kruskal 1965; Zabusky 1967; Lax 1968; Gardner, Greene, Kruskal, y Miura 1967; Gardner, Greene, Kruskal, y Miura 1974; Hirota 1971; Hirota 1973a; Hirota 1973b; Zabusky y Galvin 1971; Hammack y Segur 1974, y Weidman y Maxworthy 1978.

⁹Además de los solitones de KdV, con forma sech^2 , existen otras ecuaciones con soluciones de tipo solitón con otros perfiles, como sech , o $\arctan(e^{\alpha x})$.

solitones y mantener su forma y su velocidad, salvo una traslación espacial.

En los últimos años, el concepto de solitón ha sido utilizado ampliamente. Por ejemplo, la ecuación de Schrödinger no lineal (SNL), que describe ondas en plasmas, en superconductores, en óptica no lineal, etc., tiene soluciones solitónicas que tienen la forma de una sech que modula una onda viajera monocromática. A diferencia de los solitones de KdV, los solitones de SNL no dependen de la amplitud. Otro modelo es el de sG, del que trata esta tesis, que admite solitones como solución que son *kinks* o *antikinks*, de la forma $\arctan(e^{\pm\alpha x})$, cuya amplitud es independiente de la velocidad. Además de estos, existe una gran variedad de modelos de ecuaciones con soluciones solitónicas de alguno de los tipos mencionados.

Modelos no integrables y ondas solitarias.

Sin embargo, hay que tener cuidado de no generalizar el concepto de solitón al de onda solitaria. De hecho, existen EDPs no lineales con soluciones de tipo onda solitaria pero que no son solitones, en el sentido de que no cumplen las tres características del párrafo anterior (generalmente lo que más limita la definición de solitón es la tercera propiedad que he mencionado). Sin embargo, de las propiedades descritas en los solitones se desprende que se pueden considerar ondas solitarias. Las EDPs no lineales que tienen soluciones de tipo solitón son las completamente integrables, para las que existen herramientas matemáticas como la TB o la TEI con las que obtener soluciones. Así que conceptos que ya he mencionado hasta ahora como TB, TEI, solitones y ecuaciones CI van unidos en un solo paquete, de manera que si encontramos uno de ellos estudiando una ecuación lo más probable es que los otros tres salgan a relucir antes o después¹⁰. A continuación voy a poner un ejemplo muy sencillo de dos ecuaciones, muy parecidas entre sí, pero de las cuales una es completamente integrable mientras que la otra no. Ya conocemos la ecuación de sG en su forma estándar (2.9), y también conocemos el Hamiltoniano (2.7) del que surgen dichas ecuaciones. Pues bien, supongamos que, en lugar del término $(1 - \cos \phi)$ del Hamiltoniano, se escribe uno genérico, de la forma $V(\phi)$, donde V , que es el *potencial* del modelo, es una función bien comportada. Entonces, el Hamiltoniano del modelo es

$$H[\phi] = \int_{-\infty}^{\infty} dx \left\{ \frac{1}{2} \phi_t^2 + \frac{1}{2} \phi_x^2 + V(\phi) \right\}. \quad (2.14)$$

¹⁰Depende de los conocimientos matemáticos de cada uno o de la habilidad de buscar en la literatura para encontrarlos.

Dicho Hamiltoniano (2.14) da lugar al conocido modelo de Klein-Gordon no lineal, cuyas ecuaciones de movimiento son

$$\phi_{tt} - \phi_{xx} + \frac{dV(\phi)}{d\phi} = 0. \quad (2.15)$$

Ahora supongamos que elegimos el potencial $V(\phi) = \frac{1}{4}(1 - \phi^2)^2$, que define el llamado modelo ϕ^4 . Dicho potencial tiene dos mínimos en $\phi = \pm 1$, mientras que el potencial de sG tiene infinitos mínimos de la forma $\phi = 2n\pi$, con $n = 0, \pm 1, \pm 2, \dots$. No sería muy descabellado pensar en la posibilidad de que la ecuación de movimiento asociada al modelo ϕ^4 , que es

$$\phi_{tt} - \phi_{xx} + \phi^3 - \phi = 0, \quad (2.16)$$

poseyera también solitones, como sG. Sin embargo, la ecuación (2.16) posee soluciones de tipo onda solitaria *que no son solitones*, con un perfil de $\pm \tanh$, muy parecido al perfil de los solitones de sG. La principal diferencia surge en la interacción entre dos ondas solitarias, puesto que en ϕ^4 no se cumple la tercera propiedad de los solitones que mencionada anteriormente. Es curioso ver cómo dos ecuaciones tan parecidas con soluciones también parecidas no poseen ambas a la vez soluciones de tipo solitón o de tipo onda solitaria.

Sin embargo, pese a que en este apartado he querido establecer la diferencia entre ondas solitarias y solitones, la sutileza de que una ecuación sea o no completamente integrable y por tanto tenga soluciones de tipo solitón no ha sido un problema de interés a lo largo de la realización de esta tesis. La diferencia entre ambos tipos de soluciones se ha introducido para completar un poco la discusión que estamos haciendo sobre las EDPs no lineales, pero no es relevante a la hora de leer y entender los resultados de esta tesis. De hecho, en la sección 2.3, cuando estudiemos el comportamiento de soluciones de onda solitaria en medios inhomogéneos, trataremos tanto el modelo de sG como el ϕ^4 sin tener en cuenta ni observar diferencias significativas en los resultados de los dos modelos. Estudios como éstos sirven para analizar la dependencia o no de un cierto comportamiento de un modelo con que sea CI o no.

2.3. *Kinks, breathers* y otras soluciones.

Solitones en sG.

Una vez introducida la idea fundamental de lo que es un solitón y una onda solitaria, volvamos al modelo de sG para obtener la expresión de una solución

fundamental en el estudio de la ecuación de sG, el *kink* (y el *antikink*).¹¹ Partiendo de la ecuación de sG estándar (2.9), supongamos el *Ansatz*¹² $\phi(x, t) = \tilde{\phi}(z)$ con $z = x - vt$, es decir, estamos suponiendo que la solución es estacionaria y se mueve con una velocidad constante v . Introduciendo dicho *Ansatz* en (2.9), se obtiene la ecuación diferencial ordinaria (EDO)

$$(1 - v^2)\tilde{\phi}'' = \sin \tilde{\phi}. \quad (2.17)$$

Multiplicando por $\tilde{\phi}'$ a ambos lados de la ecuación e integrando respecto a z , se obtiene que

$$\frac{1}{2}(1 - v^2)(\tilde{\phi}')^2 = A - \cos \tilde{\phi}, \quad (2.18)$$

donde A es una constante de integración. Tomando $A = 1$, $|v| < 1$ e imponiendo las condiciones de contorno $\tilde{\phi} \rightarrow 0 \pmod{2\pi}$ y $\tilde{\phi}' \rightarrow 0$ para $z \rightarrow \pm\infty$, se obtiene que

$$\phi(x, t) = 4 \arctan \left\{ \exp \left(\pm \frac{x - vt - x_0}{\sqrt{1 - v^2}} \right) \right\}. \quad (2.19)$$

La expresión (2.19) es un solitón que, dependiendo del signo que se utilice, se llama *kink*¹³ (para el signo $+$) o *antikink* (para el signo $-$). El aspecto que tienen estas soluciones está representado en la figura 2.1 sin dar una escala en el eje x . Esto es debido a que, para $v = 0$, el *kink* (*antikink*) se encuentra estático en una posición del eje x dada por su centro, x_0 , y con una anchura característica, que es la que tiene en reposo. En el caso $v \neq 0$, tenemos la misma forma de *kink* (o *antikink*) que antes, pero que se mueve con velocidad constante v por el eje x y además se estrecha un poco de acuerdo con el factor de Lorentz $\gamma = (1 - v^2)^{-1/2}$. Se puede pasar de una solución a otra mediante una transformación de Lorentz,

$$x' = \frac{x - vt}{\sqrt{1 - v^2}}, \quad t' = \frac{t - vx}{\sqrt{1 - v^2}}. \quad (2.20)$$

De esta manera podemos pasar de un *kink* estático a otro en movimiento que también es solución de la ecuación. Este estrechamiento del *kink* con la velocidad v se hace crítico en $|v| = 1$, velocidad para la cual su anchura se hace cero (por eso se dice que la velocidad crítica de un *kink* es $|v| = 1$). La invariancia en coordenadas estándar (2.20) equivale a una dilatación en coordenadas características, de la forma:

$$\xi' = a\xi, \quad \tau' = \frac{\tau}{a}, \quad (2.21)$$

¹¹La mayoría de estas notas proceden de Scott (1999) y Debnath (1997).

¹²La palabra *Ansatz* es alemana y significa “enfoque” o “punto de partida”.

¹³La palabra *kink* es inglesa y tiene múltiples acepciones, como “nudo” o “retorcimiento” (en una manguera, por ejemplo), “tortícolis” o “contractura”, “defecto” (que hay pulir) o “manía”.

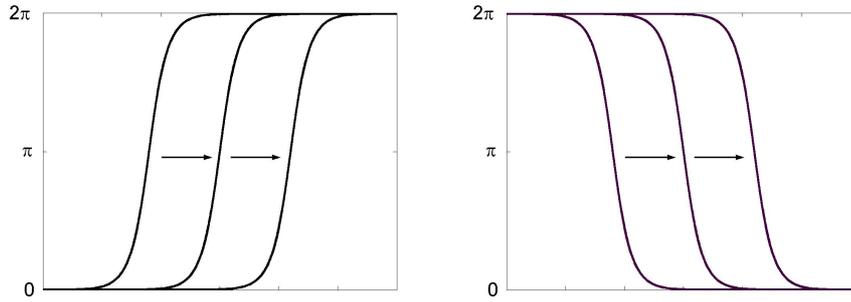


Figura 2.1: Solución tipo *kink* (izquierda) y *antikink* (derecha) de la ecuación de sG. Se ha elegido dibujar las soluciones en el rango $[0, 2\pi]$, aunque cualquier traslación vertical de $\pm 2\pi k$, con k entero, es igualmente válida. Las flechas indican el sentido del movimiento para velocidad no nula mayor que cero. No se ha dibujado la escala espacial porque la anchura del *kink* depende de su velocidad.

con $x = \xi + \tau$, $t = \xi - \tau$, lo mismo para (x', t') y (ξ', τ') , y donde $a = \sqrt{(1-v)/(1+v)}$. Dicha invariancia llevó a Bäcklund a generalizar la obtención de nuevas soluciones de la ecuación de sG, propuesta por Bianchi, mediante el parámetro a para obtener la TB (2.2) discutida en la sección 2.1.

Otra propiedad muy importante de los *kinks* y *antikinks* es que poseen carga topológica. De hecho, a estos solitones se les denomina solitones *topológicos*, a diferencia de los de otras ecuaciones solitónicas como SNL o KdV. Esta propiedad se entiende fácilmente mirando de nuevo la figura 2.1. En ella se observa que el valor de la solución en $x \rightarrow \infty$ y en $x \rightarrow -\infty$ varía en 2π . Puesto que el modelo es infinito, estos valores no varían con la evolución temporal del sistema, por lo que esa diferencia se conserva en el tiempo. Así, se puede definir una *carga topológica*, $q = (\phi(\infty, t_0) - \phi(-\infty, t_0))/2\pi$, que es conservada. Con esta definición, la carga topológica del *kink* es $q = 1$, y la del *antikink* es $q = -1$, y el hecho de ser no nulas les da a estos objetos una gran robustez, puesto que su amplitud nunca puede hacerse cero.

Dos solitones.

Hemos visto hasta ahora la expresión de la solución de la ecuación de sG con un solitón (2.19), pero la ecuación (2.1) acepta soluciones con N solitones. Para obtener las expresiones de estas soluciones, más complicadas, utilizaré la

TB (2.2). En primer lugar voy a mostrar cómo se obtiene la expresión del *kink* (2.19) utilizando esta potente herramienta. Como ya mencionamos en la sección 2.1, la TB permite obtener una nueva solución a partir de otra existente y con un parámetro libre. Supongamos ahora que ponemos en marcha la maquinaria para obtener una solución a partir de la solución trivial $\phi = 0$. La expresión (2.2), para un parámetro a_1 , quedaría entonces así:

$$\begin{aligned} \left(\frac{\psi_K}{2}\right)_\xi &= a_1 \sin\left(\frac{\psi_K}{2}\right), \\ \left(\frac{\psi_K}{2}\right)_\tau &= \frac{1}{a_1} \sin\left(\frac{\psi_K}{2}\right). \end{aligned} \quad (2.22)$$

Esta ecuación puede integrarse fácilmente multiplicando la ecuación de arriba por $d\xi$, la de abajo por $d\tau$ y sumando ambas ecuaciones para obtener la expresión de una diferencial exacta¹⁴. La integración de esta expresión da, en términos de las coordenadas características,

$$\tan\left(\frac{\psi_K}{4}\right) = \exp\left[a_1\xi + \frac{\tau}{a_1} + \delta\right], \quad (2.23)$$

donde δ es una constante de integración. En función de las coordenadas estándar, la expresión anterior se escribe

$$\tan\left(\frac{\psi_K}{4}\right) = \exp\left[\left(a_1 + \frac{1}{a_1}\right)x + \left(a_1 - \frac{1}{a_1}\right)t + \delta\right], \quad (2.24)$$

que se corresponde con la expresión ya dada en (2.19) para

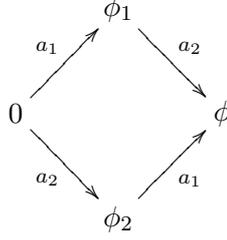
$$v = \frac{1 - a_1^2}{1 + a_1^2}, \quad x_0 = -\frac{2a_1}{1 + a_1^2}\delta. \quad (2.25)$$

Nótese que la expresión del *kink* dada por (2.19) se obtiene haciendo a_1 real con $a_1 > 0$ en (2.24), mientras que la del *antikink* sale con $a_1 < 0$. Además, para $|a_1| = 1$ lo que se obtiene es un *kink* (o un *antikink*) en reposo, y si cambiamos el sentido de movimiento del solitón (esto es, pasamos de $+v$ a $-v$) es equivalente a cambiar el parámetro a_1 por $1/a_1$.

Podemos encontrar la solución con dos solitones (ya sean dos *kinks*, dos *antikinks* o un *kink* y un *antikink*) aplicando de nuevo la TB con otro parámetro a_2 a

¹⁴Para más detalle, véase el apéndice A.

la solución con un solitón (2.24). Para ello utilizaremos el siguiente diagrama de Bianchi



asociado a la identidad de Bianchi (2.4) partiendo de la solución trivial $\phi_0 = 0$,

$$\tan\left(\frac{\phi}{4}\right) = \frac{a_1 + a_2}{a_1 - a_2} \tan\left(\frac{\phi_1 - \phi_2}{4}\right). \quad (2.26)$$

Dicha fórmula nos lleva a la expresión de la solución con dos *kinks* que colisionan en $x = 0$,

$$\tan\left(\frac{\phi_{KK}}{4}\right) = \frac{v \operatorname{sh}\left(\frac{x}{\sqrt{1-v^2}}\right)}{\operatorname{ch}\left(\frac{vt}{\sqrt{1-v^2}}\right)}, \quad (2.27)$$

o un *kink* y un *antikink*,

$$\tan\left(\frac{\phi_{AK}}{4}\right) = \frac{\operatorname{sh}\left(\frac{vt}{\sqrt{1-v^2}}\right)}{v \operatorname{ch}\left(\frac{x}{\sqrt{1-v^2}}\right)}. \quad (2.28)$$

También es posible generar otra solución con $a_2 = a_1$, y el resultado sería

$$\tan\left(\frac{\phi_{a_1 a_1}}{4}\right) = t \operatorname{sech}\left(\pm \frac{x - vt - x_0}{\sqrt{1-v^2}}\right) \quad (2.29)$$

Fijándonos en las expresiones (2.27), (2.28) y (2.29), los dos solitones no pueden estar en reposo, sino que debe haber un movimiento relativo entre ambos, como si interaccionaran entre sí. Esta “interacción” queda más patente en el caso del choque *kink-antikink*, ya que para $a_1 = \beta + i\alpha \in \mathbb{C}$, con $\alpha^2 + \beta^2 = 1$, y $a_2 = 1/a_1$, se obtiene la expresión de un estado *ligado* entre los dos solitones llamado *breather*¹⁵,

$$\tan\left(\frac{\phi_B}{4}\right) = \frac{\beta \sin(\alpha t)}{\alpha \operatorname{ch}(\beta x)}. \quad (2.30)$$

¹⁵Para un cálculo detallado ver el apéndice A.

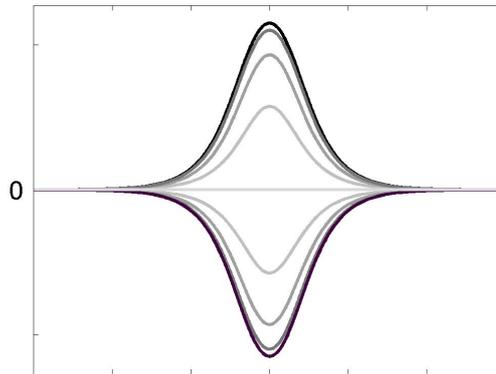


Figura 2.2: Solución tipo *breather* de la ecuación de sG, dibujada para diferentes tiempos, equiespaciados entre sí. No se definen las escalas vertical ni horizontal porque dependen de los parámetros α y β que aparecen en la expresión (2.30) (realmente, sólo de uno de ellos, porque el otro está ligado).

El nombre¹⁶ es muy descriptivo, ya que la gráfica del objeto varía con el tiempo (el *breather* no es estático) como una joroba que se infla y se desinfla, pasando del eje positivo al negativo, de forma cíclica. La figura 2.2 nos da una idea de su aspecto en reposo. Además, el *breather* (2.30) también puede moverse con velocidad constante v sin más que hacer una transformación de Lorentz (2.20), de la misma forma que un (*anti*)*kink* (ya que, al fin y al cabo, está formado por dos *kinks* ligados).

Podemos entender mejor lo que pasa en el sistema si pensamos en términos de energía. Utilizaré el Hamiltoniano del modelo continuo, (2.7), para calcular la energía de un solitón y de un *breather*. Para ello sólo es necesario sustituir la expresión de cada solución en el funcional (2.7) y calcular el resultado (ver Re-moissenet (1994) para un cálculo detallado). La energía de un *kink* (o un *antikink*) es entonces $E_K = 8/\sqrt{1-v^2}$, y la de un *breather* $E_B = 16\beta/\sqrt{1-v^2}$, ambas expresiones para objetos en movimiento. Ahora podemos fijarnos en varios detalles de estas expresiones. Por un lado, la energía del *breather* es menor que la energía de un par *kink-antikink*, por ser siempre $0 < \beta < 1$, lo cual tiene lógica

¹⁶*Breather* se puede traducir como “respirador” o “que respira”, aunque realmente significa “respiro”.

si realmente un *breather* es un estado ligado de un *kink* y un *antikink*. Por otro lado, la energía de ambos objetos es como la de una partícula relativista de masa en reposo $m_K = 8$ para el *kink* (y el *antikink*) y $m_B = 16\beta$ para el *breather*. Esto, unido a la contracción de Lorentz ligada a la invariancia (2.20) de las soluciones ya mencionada, hace que estos objetos tengan un marcado carácter de partícula relativista.

De la misma manera que se construyen soluciones con dos solitones es posible construir soluciones con N solitones, sin más que ir utilizando paso a paso la maquinaria que ofrece la TB a través de la identidad de Bianchi (2.3). Así, por ejemplo, podemos calcular la expresión de la solución de sG con tres solitones estáticos, dos de ellos ligados en forma de *breather* y el tercero como un *kink*. La expresión, calculada en detalle en el apéndice A, es la siguiente:

$$\tan\left(\frac{\phi_W}{4}\right) = \frac{U}{V}, \quad (2.31)$$

con

$$\begin{aligned} U(x, t) &= e^{z_0+2z_1} + \frac{1+\beta}{1-\beta}e^{z_0} - 2\frac{\beta}{1-\beta}e^{z_1} \cos z_2, \\ V(x, t) &= 1 + \frac{1+\beta}{1-\beta}e^{2z_1} - 2\frac{\beta}{1-\beta}e^{z_0+z_1} \cos z_2, \end{aligned} \quad (2.32)$$

y donde $z_0 = x - x_k$, $z_1 = \beta(x - x_b)$ y $z_2 = \alpha(t - t_b)$. Los parámetros x_k y x_b van asociados a traslaciones espaciales del *kink* y del *breather*, respectivamente, mientras que t_b da la traslación temporal del sistema. Aunque este tipo de objetos ya eran conocidos desde los años 80 (Segur 1983), recientemente ha habido estudios numéricos de este tipo de objetos (Kalbermann 2006), llamados *wobbles* o *wobbling kinks*¹⁷, cuyas expresiones se han obtenido a partir de TEI, y no son exactamente iguales a (2.31) aunque ambas definen el mismo objeto (para ver la relación entre las dos formas de escribir los *wobbles*, ver el apéndice A). En el estudio realizado por Segur (1983) se obtiene la expresión analítica del *wobble*, y se hace una leve referencia a su posible inestabilidad, pero con decaimiento algebraico en lugar de exponencial, aunque no hay un análisis detallado del problema. En Kalbermann (2006), sin embargo, se estudian numéricamente algunas perturbaciones del sistema. El estudio no es muy minucioso, y además el método utilizado no parece el adecuado en el caso de decaimiento algebraico, por lo que no se puede decir nada relevante sobre este tipo de soluciones, de las que es difícil encontrar información. Parece lógico pensar que estos enrevesados objetos

¹⁷Del inglés *wobble*, balancear, oscilar.

no hayan encontrado en la literatura la cantidad de adeptos que han encontrado los *kinks*, y en menor medida los *breathers*.

Otras soluciones.

Hasta ahora sólo hemos estudiado soluciones de la ecuación de sG que son solitones o que están formadas por varios solitones. Sin embargo, la ecuación de sG también posee soluciones que no son solitónicas, sino periódicas, de la forma

$$\phi(x, t) = \tilde{\phi}(\theta) = \tilde{\phi}(\theta + 2\pi) \pmod{2\pi}, \quad (2.33)$$

donde $\theta = \omega t - kx$ es la variable de una onda viajera, y $v = \omega/k$ es la velocidad de onda local. Dichas soluciones se obtienen a partir de (2.18) para valores de A distintos de $A = 1$ (que era el caso de los solitones), y su expresión formal es

$$\sqrt{k^2 - \omega^2} \int^{\tilde{\phi}} \frac{dy}{\sqrt{2(A - \cos y)}} = \theta. \quad (2.34)$$

Esta expresión, para $A > 1$, $v < 1$, da una solución que crece monótonamente con θ , y representa un tren de *kinks*, separados unos de otros por una distancia $2\pi/k$; para $-1 < A < 1$, $v > 1$, la solución oscila periódicamente en torno a $\tilde{\phi} = 0$, con periodo $2\pi/k$, amplitud $\arccos(A)$ y relación de dispersión *no lineal*

$$\omega^2 - k^2 = \frac{\pi^2}{4K^2 \left(\sqrt{\frac{1-A}{2}} \right)}, \quad (2.35)$$

donde $K(\cdot)$ es la integral elíptica de primera especie. Conforme la amplitud aumenta de 0 a π (es decir, A disminuye de 1 a -1), la parte de la derecha de la relación (2.35) baja hasta 1; por tanto, para ondas periódicas de baja amplitud, la menor frecuencia permitida es la unidad. Estas soluciones aparecen en situaciones de inestabilidad, y están muy relacionadas con otras soluciones que describiremos a continuación.

Dichas soluciones son los *fonones*, que son soluciones que describen el modelo de sG en el límite lineal, para $\phi \ll 1$, cuando la aproximación $\sin \phi \simeq \phi$ es válida. En ese caso, la ecuación (2.9) pasa a ser la de Klein-Gordon,

$$\phi_{tt} - \phi_{xx} + \phi = 0. \quad (2.36)$$

Utilizando la transformada de Fourier, la solución que se obtiene es una onda monocromática, $\phi(x, t) \propto \exp(i\omega t - ikx)$, con relación de dispersión $\omega^2(k) =$

$1 + k^2$. Estos fonones sólo son soluciones exactas para la ecuación linealizada. Las soluciones exactas de las que son aproximación son las ondas periódicas que he mencionado anteriormente, aunque se suele hablar de unas u otras indistintamente como fonones. Los fonones aparecen a menudo en procesos en los que existe algún tipo de inestabilidad, en estados transitorios a soluciones estables. Los mencionaré más adelante cuando hable de modelos perturbados.

Un estudio más completo de los diferentes tipos de soluciones de la ecuación de sG aparece en Scott (1999), de donde he recogido parte de esta información. Ahí también aparece un estudio de sistemas continuos de tamaño finito en términos de funciones elípticas. En la elaboración de esta memoria no he creído necesario extenderme en este tema porque no ha sido relevante para la realización de la tesis.

3

Perturbaciones inhomogéneas.

En este capítulo se explica cómo introducir perturbaciones en el modelo de Klein-Gordon, y se estudia una perturbación en particular del sustrato de tipo sinusoidal. Dicho sistema aparece bastante en la literatura, entre otras cosas, por mostrar atrapamiento de *kinks* para valores de la longitud de onda de la perturbación del orden de la anchura del *kink*. Para estudiar la dinámica de *kinks* en estos sistemas, primero se utiliza el centro del *kink* como coordenada colectiva, construyendo para ello el potencial efectivo asociado al sistema perturbado. A continuación se introduce, además, de la anchura del *kink* como coordenada colectiva, se explica el caso del modelo sin perturbar y luego se presentan, como parte del trabajo original realizado durante esta tesis doctoral, los resultados numéricos obtenidos en el caso perturbado, tanto para sine-Gordon como para ϕ^4 , que indican la clara influencia de la anchura del *kink* en el atrapamiento de éste.

3.1. Ecuación de Klein-Gordon perturbada.

Cómo introducir una perturbación.

En el apartado 2.2 se introdujo la ecuación de movimiento del modelo de Klein-Gordon no lineal (2.15). Dicha ecuación tiene un término inercial, dado

por la segunda derivada temporal, otro dispersivo, dado por la segunda derivada espacial, y el término que proviene del potencial. Si introducimos más términos, la ecuación de Klein-Gordon no lineal *perturbada* pasaría a escribirse así:

$$\phi_{tt} - \phi_{xx} + \frac{dV(\phi)}{d\phi} + F(x, t, \phi, \phi_t, \phi_x, \dots) = 0, \quad (3.1)$$

donde F representa un término de fuerza genérico, que puede depender de la posición x , del tiempo t , del campo escalar ϕ y de cualquiera de sus derivadas. Los términos perturbativos más corrientes que nos encontramos en la literatura son el término disipativo o de rozamiento, $\gamma\phi_t$, y un término de fuerza externa de la forma $\epsilon f(x, t)g(\phi)$. Los parámetros γ y ϵ sirven para regular la intensidad de estas perturbaciones respecto al resto de los términos de la ecuación (3.1).

Un tipo interesante de modelo perturbado que aparece con frecuencia en la literatura [véanse, por ejemplo, las contribuciones de Sánchez et al. (1992), Sánchez y Bishop (1994), Willis (1997) o Cuenda y Sánchez (2005)] es aquél con una perturbación sinusoidal espacial de la forma $F = \epsilon \frac{dV}{d\phi} \cos(kx)$, de manera que las ecuaciones de movimiento se escriben así,

$$\phi_{tt} - \phi_{xx} + \frac{dV}{d\phi} [1 + \epsilon \cos(kx)] = 0. \quad (3.2)$$

Veamos cuál es el significado de la ecuación (3.2). Recordemos que, para el caso $\epsilon = 0$, el tercer sumando de la ecuación proviene del término potencial, característico de cada modelo, y que da la fuerza que actúa sobre el campo ϕ en la posición x . Dicha fuerza es independiente del acoplamiento que provenga del término dispersivo, dado por la derivada espacial del segundo término. Supongamos que introducimos un parámetro extra que multiplique ese término potencial, de la forma $(1 + \epsilon) \frac{dV}{d\phi}$. En ese caso estaremos aumentando la intensidad de esa fuerza respecto a la de los otros dos términos de la ecuación (3.2), el inercial y el dispersivo, y la ecuación, anteriormente adimensionalizada, dependería de otro parámetro extra. Redefiniendo una nueva escala temporal y espacial, podríamos tener la misma ecuación de sG adimensionalizada y sin perturbar del principio. Esto ha ocurrido porque hemos introducido una variación *global*, que ha cambiado el término potencial de forma *homogénea* en todo el eje x . Pero imaginemos que, en lugar de hacerlo por una constante, multiplicamos el término del potencial por una función que dependa de la posición, $[1 + \epsilon f(x)]$. En ese caso, la intensidad de la fuerza potencial vendrá dada por ese factor, que ahora depende de la posición. Por eso se dice que la perturbación es *inhomogénea*, o que el modelo es inhomogéneo, en contraposición al caso en el que la perturbación es global.

Las perturbaciones inhomogéneas tienen una clara motivación física. En la descripción de algunos procesos mediante un modelo, normalmente se suelen realizar una serie de aproximaciones que, en el caso ideal, nos llevan al modelo homogéneo. Sin embargo, algunas veces esto no es así. Por ejemplo, en uniones Josephson, la inhomogeneidad va asociada a la dependencia con la posición del espesor de la capa de material aislante que separa los dos superconductores de los que está compuesta la unión; en una cadena de péndulos, en la dependencia con la posición de las masas de los péndulos (eligiendo la longitud de éstos para que su momento de inercia no varíe); en el modelo de ADN, en el orden de las bases que nos da una secuencia específica de ADN, etc.

3.2. Estudio con una coordenada colectiva.

El centro del *kink*.

Recordemos ahora las ecuaciones de movimiento del modelo de sG sin perturbar, (2.9), y del modelo ϕ^4 sin perturbar, (2.16); ambos son modelos de Klein-Gordon no lineales, correspondientes a dos elecciones distintas del potencial $V(\phi)$. Estas ecuaciones, como ya indiqué en el apartado 2.2, poseen soluciones topológicas de tipo onda solitaria (solitones en el caso de sG). Dichas soluciones son, para el modelo de sG, la expresión (2.19) y, para el modelo ϕ^4 ,

$$\phi(x, t) = \pm \tanh \left(\frac{x - vt - x_0}{\sqrt{2(1 - v^2)}} \right). \quad (3.3)$$

De ahora en adelante utilizaremos únicamente las soluciones con signo positivo, que nombraremos indistintamente como *kinks*. Estos *kinks* dejan de ser soluciones de las ecuaciones perturbadas (3.1) de los modelos sG y ϕ^4 debido al término extra de fuerza. Sin embargo, debido a su carga topológica conservada, son soluciones extremadamente estables de las ecuaciones originales. En el sistema sin perturbar, ambas soluciones permanecen en reposo para $v = 0$, o se mueven como un objeto rígido a lo largo del eje x para velocidades no nulas. El hecho de que sean tan estables hace pensar que, bajo una perturbación del sistema (como por ejemplo la aparición de más fuerzas en el modelo), la forma del *kink* variará poco y se moverá de manera casi rígida a lo largo del eje x .

Esta hipótesis permite utilizar la técnica de *coordenadas colectivas* con estos objetos para obtener información sobre los modelos perturbados. Esta técnica consiste en la aproximación de un campo escalar $\phi(x, t)$ mediante una serie de variables dependientes del tiempo $Y_i(t)$ (coordenadas colectivas) que se definen en

un *Ansatz* adecuado. De esta forma la ecuación que describe el modelo, que originalmente es una EDP, pasa a ser un sistema de ecuaciones diferenciales ordinarias (EDOs), con tantas ecuaciones como coordenadas colectivas se hayan usado en la aproximación. Esta técnica fue introducida en los modelos de Klein-Gordon no lineales en Fogel et al. (1976), y posteriormente ha sido utilizada en numerosas ocasiones.

Normalmente, y debido a la rigidez de los *kinks*, la coordenada colectiva que se utiliza en este tipo de estudios es su centro, que en el caso del modelo de sG es la posición del eje x en la que la solución corta a la abscisa $y = \pi \pmod{2\pi}$ y en el modelo ϕ^4 es el punto de corte con el propio eje x . El *Ansatz* que se utiliza es, por tanto,

$$\tilde{\phi}(z) = 4 \arctan(e^z) \quad (3.4)$$

para el *kink* de sG y

$$\tilde{\phi}(z) = \tanh(z) \quad (3.5)$$

para el *kink* de ϕ^4 , con $z = (x - X(t))/l_0$, siendo l_0 la anchura en reposo del *kink* y $X(t)$ la posición del centro del *kink*, que depende del tiempo pero cuya velocidad no debe ser muy grande para que la aproximación sea válida, ya que hemos eliminado el factor de Lorentz de la expresión original de los *kinks*.

Trabajos anteriores.

La ecuación (3.2) aparece en la literatura en varias ocasiones, ya que da lugar a comportamientos nada triviales en la propagación de solitones. Así, en Sánchez et al. (1992) se estudió la evolución de *breathers* de sG en este modelo perturbado. Sea λ_B la longitud de onda del *breather*, y sea $\lambda = 2\pi/k$ la longitud de onda de la perturbación; mediante simulaciones numéricas se observó que, en los casos en los que λ_B/λ era muy pequeño o muy grande, la aproximación de una coordenada colectiva predecía correctamente el comportamiento del *breather*. Sin embargo, cuando $\lambda_B/\lambda \simeq 1$, las simulaciones numéricas reflejaron, en algunos casos, la desaparición del *breather* debido a una gran radiación de energía; en otros, la descomposición del *breather* en un par *kink-antikink*; a veces, el *breather* inicial salta unas cuantas barreras de potencial y finalmente se descompone en diferentes excitaciones de tipo *breather*, pero apenas distinguibles de la radiación. Se encontraron comportamientos tan anómalos incluso para valores de ϵ muy pequeños (alrededor de 0.1). Aparecía de esta manera una interesante situación de *competición de escalas*, en la que existe un comportamiento anómalo cuando la longitud de onda de la perturbación aplicada es comparable con la anchura del *breather*.

Todos estos casos fueron encontrados en un estudio sistemático del comportamiento del *breather* para diferentes valores del par $(\epsilon, \lambda_B/\lambda)$, pero sin embargo no se logró dar explicación a este interesante fenómeno.

Otro trabajo muy interesante donde se estudia la perturbación (3.2) mediante una coordenada colectiva¹, en este caso para *kinks* de sG, es Sánchez y Bishop (1994). En él los autores continúan el trabajo realizado para los *breathers* en Sánchez et al. (1992) pero, debido a la mayor sencillez del *kink* frente al *breather*, estudian el comportamiento de *kinks* sometidos a la perturbación para ver si con los resultados se puede aportar algo de luz al problema de los *breathers*. Además, puesto que el *kink* es muy estable y no puede desaparecer, lo que hicieron fue utilizar como condición inicial *kinks* con velocidad no nula colocados en un máximo de la perturbación y estudiar su evolución. En este caso la competición de escalas se manifestó en que el *kink* no se propagaba para ciertos valores de λ , y se decía entonces que el *kink* estaba *atrapado*. En el trabajo realizaron un estudio con coordenadas colectivas, parecido al llevado a cabo en Sánchez et al. (1992) para *breathers*, que de nuevo dio resultados favorables fuera del régimen de competición de escalas. Para explicar el atrapamiento del *kink* en determinados valores de λ , realizaron un análisis de estabilidad lineal numérica y relacionaron el espectro de frecuencias en el que emiten los *kinks*, para diferentes valores de λ , con los resultados de atrapamiento, dando lugar a predicciones cuantitativamente correctas. Sin embargo, pese a sus resultados, el papel de la anchura del *kink*, una de las longitudes involucradas en la competición, no estaba claro, y no había resultados analíticos sencillos.

Posteriormente, Willis (1997) realizó un análisis de coordenadas colectivas del problema (3.2) en el que se incluye, aparte de la posición del centro del *kink*, la anchura del mismo (aunque aquí trabajaron con la pendiente en el centro del *kink*, pero una es la inversa de la otra) y los modos de fonones. Sin embargo, en dicho estudio no hacen ninguna mención al proceso de competición de escalas que se produce en este sistema relacionado con el atrapamiento de los *kinks* para determinados valores de λ . Por eso fue necesaria una nueva aproximación al problema mediante la técnica de coordenadas colectivas, pero utilizando en esta ocasión sólo el centro y la anchura del *kink* para entender la relevancia que la anchura del *kink* tiene en todo el problema. Aunque más adelante se analizará en detalle dicho trabajo, llevado a cabo como trabajo original de parte de esta tesis,

¹Existen otros trabajos en los que se estudia esta misma ecuación, pero utilizan otras técnicas, como técnicas perturbativas con funciones de Green o teoría espectral inversa perturbada; sin embargo, no se preocupan por el fenómeno de competición de escalas, sino sólo de la radiación que emite el *kink* en la ecuación perturbada (Mkrtychyan y Shmidt 1979; Malomed 1990; Malomed y Tribelsky 1990; Fei et al. 1993).

es necesario presentar antes el estudio del problema del *kink* de sG perturbado mediante la ecuación (3.2) utilizando una sola coordenada colectiva.

El potencial efectivo.

A continuación seguiré el trabajo de Sánchez y Bishop (1994) para el modelo de sG. Las ecuaciones de movimiento (3.2) son un caso particular de la ecuación (3.1), para $F(x, \phi) = \epsilon \cos(kx)V'(\phi)$. El Hamiltoniano del que proviene la ecuación (3.2) es entonces

$$H[\phi] = \int_{-\infty}^{\infty} dx \left\{ \frac{1}{2} \phi_t^2 + \frac{1}{2} \phi_x^2 + V(\phi) + \epsilon V(\phi) \cos(kx) \right\}. \quad (3.6)$$

Ahora es necesario utilizar los *Ansätze*² dados por las expresiones (3.4) y (3.5), con $z = (x - X(t))/l_0$, válidos para velocidades pequeñas. Sustituyendo estas expresiones en el Hamiltoniano del problema, (3.6), y calculando la integral término a término, se obtiene finalmente una expresión para el Hamiltoniano que es la energía de una partícula puntual, newtoniana, situada en $X(t)$,

$$E(X, \dot{X}) = \frac{M_0}{2} \dot{X}^2 + M_0 + \epsilon A(k) \cos(kX). \quad (3.7)$$

En esta expresión, M_0 es la “masa en reposo” del *kink*, que proviene de la integración del término cinético

$$M_0 l_0 = \int_{-\infty}^{\infty} dz [\phi'(z)]^2, \quad (3.8)$$

y que vale $M_0 = 8$ en el modelo de sG y $M_0 = 4/3\sqrt{2}$ en ϕ^4 ; l_0 es la anchura en reposo del *kink* y es $l_0 = 1$ en el modelo de sG y $l_0 = \sqrt{2}$ en ϕ^4 ; y $A(k)$ es una función dada por la integral del término perturbativo del Hamiltoniano (3.6),

$$H^{\text{pert}}[\phi] = \int_{-\infty}^{\infty} dx \{ \epsilon V(\phi) \cos(kx) \}, \quad (3.9)$$

que para el modelo de sG vale

$$A(k) = \frac{2k\pi}{\sinh\left(\frac{k\pi}{2}\right)} \quad (3.10)$$

²Plural de *Ansatz*.

y para ϕ^4 es

$$A(k) = \frac{k\pi(k^2 + 2)}{12 \sinh\left(\frac{k\pi}{\sqrt{2}}\right)}. \quad (3.11)$$

En la expresión de la energía (3.7) aparece un término cinético y un término potencial, que es el que rige la dinámica del *kink*. Por ello, se denomina *potencial efectivo* a este término,

$$V_{\text{eff}} = M_0 + \epsilon A(k) \cos(kX). \quad (3.12)$$

La expresión (3.12) está formada por un término que proviene del Hamiltoniano sin perturbar, constante, y otro término que proviene de la perturbación del sistema, que nos da como resultado un potencial sinusoidal $\cos(kX)$ cuya amplitud viene dada por el factor $\epsilon A(k)$.

Esta herramienta tan sencilla nos ha resultado de gran utilidad a lo largo del desarrollo de esta tesis. Por eso es conveniente recordar cuál es el método utilizado para obtener la expresión (3.12): partiendo de las ecuaciones de movimiento perturbadas (3.2), se necesita por un lado el Hamiltoniano exacto del que provienen dichas ecuaciones de movimiento, (3.6), y por otro una solución estable y robusta que utilizaremos como *Ansatz*, (3.4) o (3.5); con estos dos elementos ya se puede obtener la expresión de la energía de una partícula newtoniana, (3.7), sustituyendo la expresión de nuestro *Ansatz* en la expresión del Hamiltoniano del modelo perturbado; el término potencial de la expresión de la energía así obtenida, (3.12), es el potencial que define el movimiento de la partícula newtoniana situada en $X(t)$, y que aproxima el movimiento de nuestro *Ansatz*, supuesto indeformable.

Lógicamente, dada la sencillez del proceso, esta aproximación sólo es válida en determinadas situaciones. En el problema que nos ocupa, dado por la ecuación (3.2), el potencial efectivo (3.12) sólo resulta válido para valores de k muy grandes o muy pequeños. En el caso k grande, o $\lambda = 2\pi/k$ mucho menor que la anchura del *kink*, entonces el *kink* es tan ancho³ que apenas nota la perturbación ($A(k)$ se hace prácticamente cero), y se mueve de forma casi uniforme como si estuviera en un potencial constante; se podría decir que el potencial efectivo es el promedio del potencial en la anchura del *kink* (en el caso del modelo de sG discreto se verá más claramente). En el caso k pequeño, o λ mucho mayor que la anchura del *kink*, éste se ve afectado por una fuerza prácticamente homogénea a lo largo de su estructura

³Puede que no sea tan intuitivo pensar en la anchura de un *kink* como lo es pensar en la anchura del solitón de tipo joroba de la ecuación KdV, que ya mencioné en el apartado 2.2. Para ello es mejor pensar en la derivada del *kink*, que sí es una joroba, de tipo sech en el caso de sG y de tipo sech^2 para ϕ^4 .

central, y los cambios de sentido de la fuerza son tan suaves que no suponen deformaciones drásticas en su estructura. El caso intermedio no es explicable con este modelo porque no se da ninguna de las dos situaciones mencionadas. Por el contrario, el *kink* nota cambios bruscos en la fuerza que actúa sobre él, y que a menudo son en un sentido a un lado de su centro y en sentido contrario en el otro lado, lo que hace que se ensanche y se contraiga demasiado y su anchura no pueda suponerse constante, como en los otros casos. Estos cambios en la anchura del *kink* pueden influir en el potencial efectivo que “ve”, de manera que parece lógico pensar que este potencial efectivo debería depender de la anchura del *kink*, además de depender de su centro.

3.3. Estudio con dos coordenadas colectivas.

La anchura del *kink*.

Con todos estos datos en mente, en Cuenda y Sánchez (2005), uno de los trabajos realizados durante la elaboración de esta tesis doctoral, estudiamos la ecuación de Klein-Gordon perturbada (3.2) mediante el método de las variables colectivas, pero en esta ocasión utilizando, además del centro del *kink* X , su anchura l . Con este estudio quisimos demostrar la importancia del acoplamiento de la anchura del *kink* con la longitud característica de la perturbación como explicación del proceso de “atrapamiento” que sufre el *kink* para determinados valores de k . En dicho estudio utilizamos un enfoque basado en el Lagrangiano, como el desarrollado en Quintero y Zamora-Sillero (2004), y calculamos la expresión del Lagrangiano exacto del que se obtiene la ecuación de movimiento (3.2). En esta memoria, por consistencia con la nomenclatura introducida hasta ahora, y ya que se obtienen resultados equivalentes, he preferido escribir el Hamiltoniano exacto del modelo, dado por la expresión (3.6). La única diferencia con el proceso anterior es que, en este estudio, nuestros *Ansätze*, (3.4) y (3.5), son tales que $z = (x - X(t))/l(t)$, donde ahora $l(t)$, la anchura del *kink*, es otra coordenada colectiva que depende del tiempo. Siguiendo el mismo proceso que antes, introducimos estos *Ansätze* en el Hamiltoniano perturbado (3.6), y calculamos su expresión en términos de X , l y sus derivadas temporales. El resultado que se obtiene es

$$H(X, l, \dot{X}, \dot{l}) = \frac{M_0 l_0}{2l} \dot{X}^2 + \frac{M_0 l_0 \alpha}{2l} \dot{l}^2 + \frac{M_0}{2} \left(\frac{l_0}{l} + \frac{l}{l_0} \right) + \epsilon \cos(kX) \frac{W(kl)}{k}, \quad (3.13)$$

donde M_0 y l_0 son las constantes definidas anteriormente; α es el resultado de la integral

$$\alpha M_0 l_0 = \int_{-\infty}^{\infty} dz z^2 [\phi'(z)]^2, \quad (3.14)$$

cuyo valor es $\alpha = \pi^2/12$ en el modelo de sG y $\alpha = (\pi^2 - 6)/12$ en el modelo de ϕ^4 ; y $W(x)$ es de nuevo una función que proviene de la integración del término perturbativo (3.9) (la única diferencia es que ahora $l \neq l_0$), que vale

$$W(x) = \frac{2\pi x^2}{\sinh\left(\frac{\pi x}{2}\right)} \quad (3.15)$$

en el modelo de sG y

$$W(x) = \frac{\pi x^2 (x^2 + 4)}{24 \sinh\left(\frac{\pi x}{2}\right)} \quad (3.16)$$

en el modelo ϕ^4 .

Puesto que ahora se utilizan dos coordenadas colectivas, la expresión del Hamiltoniano obtenido en (3.13) no es la expresión de la energía de una partícula newtoniana (3.7), como ocurría en el análisis con una coordenada colectiva. El término de potencial del Hamiltoniano depende ahora de dos variables, X y l , y para conocer la evolución temporal del sistema se necesitan las ecuaciones de Hamilton, que son

$$\frac{d}{dt} \left(\frac{\partial H}{\partial \dot{Y}_i} \right) + \frac{\partial H}{\partial Y_i} = 0, \quad (3.17)$$

donde Y_i es cada una de las dos coordenadas colectivas X y l . Las ecuaciones de movimiento vienen entonces dadas por el siguiente sistema de EDOs⁴,

$$\begin{aligned} P &= \frac{M_0 l_0 \dot{X}}{l}, \\ Q &= \frac{\alpha M_0 l_0 \dot{l}}{l}, \\ \dot{P} &= \epsilon \sin(kX) W(kl), \\ \dot{Q} &= -\frac{1}{2M_0 l_0} \left(P^2 + \frac{Q^2}{\alpha} \right) + \frac{M_0 l_0}{2} \left(\frac{1}{l^2} - \frac{1}{l_0^2} \right) \\ &\quad - \epsilon \cos(kX) W'(kl). \end{aligned} \quad (3.18)$$

⁴Este sistema de ecuaciones es el mismo que aparece en Cuenda y Sánchez (2005), derivado a partir del Lagrangiano.

Dos coordenadas colectivas en el modelo sin perturbar.

Para entender mejor este sistema de ecuaciones vamos a ver qué aspecto tiene cuando utilizamos este mismo *Ansatz*, pero esta vez en el modelo sin perturbar [tal y como fue introducido por Rice (1983)]. En ese caso, $\epsilon = 0$ y el sistema (3.18) se convierte en

$$\begin{aligned} P &= \frac{M_0 l_0 \dot{X}}{l}, \\ Q &= \frac{\alpha M_0 l_0 \dot{l}}{l}, \\ \dot{P} &= 0, \\ \dot{Q} &= -\frac{1}{2M_0 l_0} \left(P^2 + \frac{Q^2}{\alpha} \right) + \frac{M_0 l_0}{2} \left(\frac{1}{l^2} - \frac{1}{l_0^2} \right). \end{aligned} \quad (3.19)$$

Este sistema lo podemos integrar de manera exacta utilizando el hecho de que $\dot{P} = 0$, y por tanto P es una constante de movimiento, que podemos utilizar en la integración de l y más tarde en la integración de X . Antes voy a mostrar la solución del caso $\dot{l} = 0$, que se resuelve sin más que despejar la última ecuación del sistema (3.19) para $\dot{Q} = Q = 0$. El resultado, despejando $l = l_c$ (anchura crítica o constante), es

$$l_c = l_0 \sqrt{1 - \dot{X}^2}, \quad (3.20)$$

que es precisamente el factor de Lorentz que nos da la solución general del *kink* que se mueve con velocidad constante en el modelo de Klein-Gordon. La energía correspondiente a dicha solución se obtiene sustituyendo la expresión (3.20) en (3.13), utilizando que $\dot{X} = lP/M_0 l_0$ y recordando que estamos estudiando el caso sin perturbar, y por tanto $\epsilon = 0$. El resultado es

$$E_c = \frac{M_0 l_0}{l_c}. \quad (3.21)$$

Para resolver el sistema (3.19) en el caso general, voy a reescribir la ecuación para \dot{Q} en términos de $y = \sqrt{l}$, que es

$$\alpha M_0 l_0 \ddot{y} = \frac{M_0 l_0}{4} \left(\frac{1}{y^3} - \frac{y}{l_0^2} \right) - \frac{P^2}{4M_0 l_0} y. \quad (3.22)$$

Esta ecuación podemos entenderla como la ecuación de movimiento de una partícula de masa $\alpha M_0 l_0$ que se mueve bajo los efectos de un potencial $U(y)$. Dicho potencial es

$$U(y) = \frac{M_0 l_0}{8y^2} + \frac{M_0 l_0}{8l_c} y^2, \quad (3.23)$$

donde hemos usado que

$$\frac{l_0^2}{l_c^2} = \frac{1}{1 - \dot{X}^2} = \left(1 + \frac{P^2}{M_0^2}\right), \quad (3.24)$$

lo que nos permite escribir la energía de la partícula y de la forma

$$E = \frac{1}{2}\alpha M_0 l_0 \dot{y}^2 + U(y). \quad (3.25)$$

Puesto que la energía es conservada, despejamos la velocidad de la partícula y en función de su energía E y su posición y y, operando, se obtiene la dependencia de la anchura con el tiempo. El cálculo se reduce a una integral inmediata, un arco seno, por lo que la evolución temporal de la anchura del *kink* en esta aproximación, $l(t)$, viene entonces dada por un movimiento sinusoidal alrededor de un valor constante,

$$\frac{l(t)}{l_c} = \frac{E}{E_c} \pm \sqrt{\left(\frac{E}{E_c}\right)^2 - 1} \sin \left[\frac{1}{\alpha^{1/2} l_c} (t - t_0) \right], \quad (3.26)$$

donde se debe verificar $E > E_c$. Esta oscilación hace que la velocidad del *kink* también sea oscilatoria en torno a un valor medio,

$$\dot{X}(t) = \frac{P}{M_0 l_0} l(t). \quad (3.27)$$

Obviamente, esta solución del sistema de EDOs no es la solución exacta de la ecuación de Klein-Gordon que vimos anteriormente (que es (2.28) para sG y (3.3) para ϕ^4). Únicamente para el caso $\dot{l} = 0$ se obtiene una solución exacta de la EDP, que es precisamente el caso de movimiento uniforme del *kink*. Sin embargo, he querido explicar en detalle este formalismo de dos coordenadas colectivas, el centro del *kink* y su anchura, porque es el que hemos utilizado en el estudio de la perturbación (3.2) como parte del trabajo desarrollado durante esta tesis, y que aparece en Cuenda y Sánchez (2005). En la próxima sección se describe en detalle dicho trabajo.

3.4. La importancia de la anchura del *kink*.

Resultados numéricos.

Una vez conocido el resultado del sistema de EDOs (3.18) para el caso $\epsilon = 0$, vamos a ver qué ocurre en el caso perturbado. Lamentablemente no hemos podido

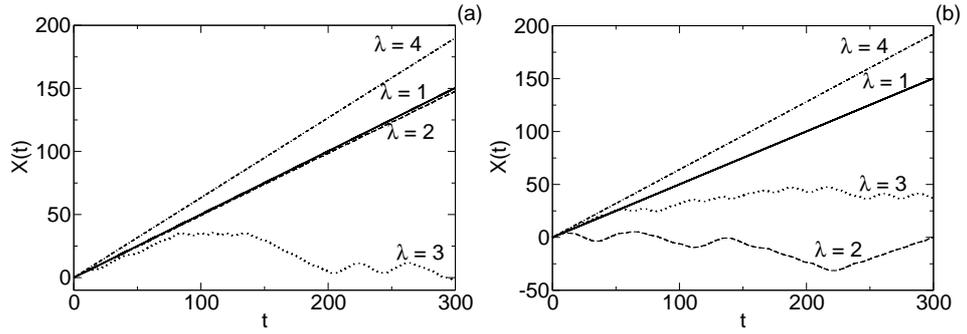


Figura 3.1: Resultado de la integración numérica del sistema de EDOs (3.18) para los modelos de (a) sG y (b) ϕ^4 . En las gráficas aparece la coordenada colectiva del centro del *kink*, $X(t)$, en función del tiempo t para diferentes valores de $\lambda = 2\pi/k$ y para $\epsilon = 0.7$. Las condiciones iniciales son $X(0) = 0$, $\dot{X}(0) = 0.5$, $l(0) = l_0(1 - \dot{X}(0)^2)^{1/2}$ y $\dot{l}(0) = 0$.

resolverlo analíticamente como en el caso anterior, así que hemos utilizado la integración numérica⁵ para estudiar el sistema para diferentes valores de λ . Respecto al valor de ϵ , y puesto que este trabajo estaba motivado por Sánchez y Bishop (1994), consideramos el valor $\epsilon = 0.7$ que se utiliza en el artículo mencionado, y que implica una perturbación del sistema no demasiado pequeña, y unas condiciones iniciales dadas por $X(0) = 0$, $\dot{X}(0) = 0.5$, $l(0) = l_0(1 - \dot{X}(0)^2)^{1/2}$ y $\dot{l}(0) = 0$. Los resultados de dichas simulaciones aparecen en la figura 3.1, en la que se observa la posición de la coordenada colectiva $X(t)$ en función del tiempo t para diferentes valores de la longitud de onda de la perturbación, λ , y para los modelos de sG y ϕ^4 . En ambos casos se observa un régimen de valores de λ en el que el centro del *kink*, X , queda atrapado, de la misma forma en que se describía en Sánchez y Bishop (1994) para el modelo completo⁶ de sG. Analizando los resultados de las simulaciones más detalladamente vemos que se produce el fenómeno de atrapamiento en uno de los primeros pozos del potencial para valores de $\lambda = 3$ en el caso de sG y $\lambda = 2$ y 3 para ϕ^4 .

Para comparar este resultado con el del modelo completo de sG y ϕ^4 , reali-

⁵El método que utilizamos para la integración numérica de (3.18) fue un Runge-Kutta de cuarto orden y paso fijo, obtenido de Press et al. (1992).

⁶Llamaré “modelo completo” o “modelo EDP” al descrito por la ecuación de Klein-Gordon no lineal, en contraposición al “modelo en coordenadas colectivas” o “modelo EDO”, descrito por el sistema de EDOs que proviene de la aproximación en coordenadas colectivas.

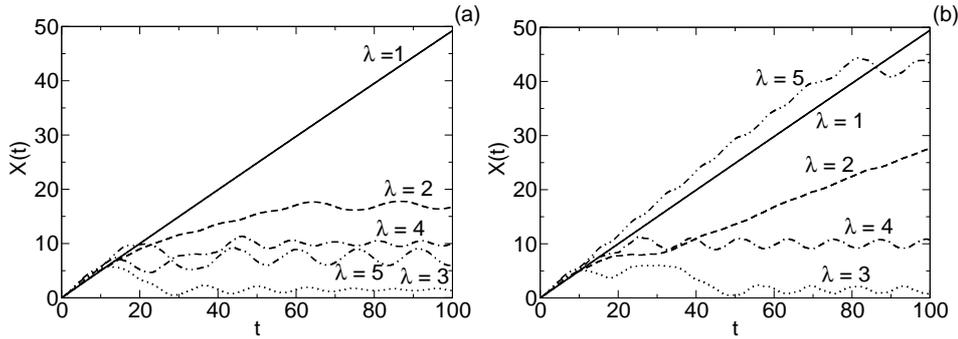


Figura 3.2: Resultado de la integración numérica del sistema completo (3.2) para los modelos de (a) sG y (b) ϕ^4 . En las gráficas aparece la posición del centro del *kink* $X(t)$, calculado como el punto en el que la solución vale $\phi = \pi$ en sG y $\phi = 0$ en ϕ^4 , en función del tiempo t para diferentes valores de $\lambda = 2\pi/k$ y para $\epsilon = 0.7$. Las condiciones iniciales son las expresiones (2.19) y (3.3) para $x_0 = 0$ y $v = 0.5$.

zamos también la integración⁷ de los correspondientes modelos EDP con condiciones iniciales dadas por las expresiones (2.19) y (3.3) con $x_0 = 0$ y $v = 0.5$, equivalentes a las utilizadas en el caso de la integración del modelo EDO. El resultado es la figura 3.2. En ella se muestra la evolución temporal de la posición del centro del *kink*, calculado como el punto en el que la solución vale $\phi = \pi$ en sG y $\phi = 0$ en ϕ^4 , para diferentes valores de λ . El fenómeno de atrapamiento, descrito en Sánchez y Bishop (1994) para sG, se reproduce en estas simulaciones tanto en sG como en ϕ^4 para diferentes longitudes de onda de la perturbación. Si nos fijamos en tiempos de integración más largos del modelo completo, nos encontramos como resultado la figura 3.3, en la que se observa el fenómeno de atrapamiento para prácticamente todas las longitudes de onda que se muestran. Sin embargo, entre todos esos procesos de atrapamiento, hay dos claramente diferenciados: por un lado están los casos en los que el *kink* queda atrapado después de haber avanzado sin problema a lo largo de varias longitudes de onda; por otro lado están los casos en los que el *kink* queda atrapado al poco tiempo de empezar la simulación, en uno de los primeros pozos por los que pasa. La diferencia entre ambos sucesos estriba en que, en los primeros, cuando ya ha avanzado un trecho antes de quedar atrapado, el *kink* ha experimentado procesos radiativos importantes que le han

⁷Para la integración numérica de la EDP de sG y ϕ^4 dada por (3.2) utilizamos un método de Runge-Kutta de cuarto orden y paso fijo, tras discretizar el Laplaciano en diferencias finitas.

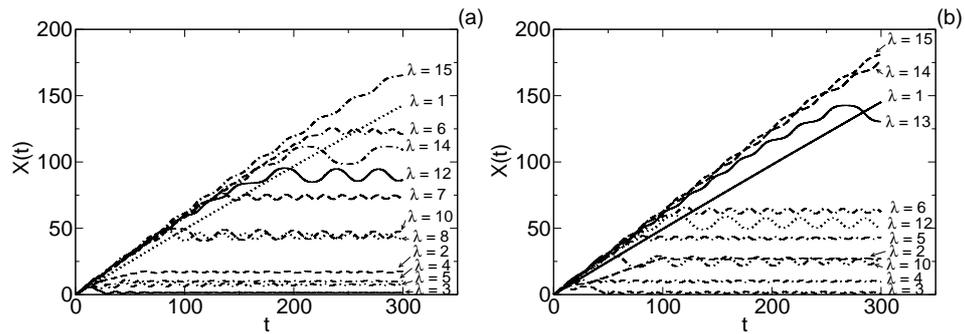


Figura 3.3: Igual que la figura 3.2, pero para tiempos más largos y un rango más amplio de λ .

hecho emitir energía en forma de fonones, perdiendo energía cinética [un estudio de la emisión de radiación en este potencial perturbativo es realizada por Fei et al. (1993)]; esto hace que la aproximación de dos coordenadas colectivas deje de ser fiable por intervenir en el proceso otros mecanismos disipativos. En los segundos, en los que el *kink* se queda atrapado en uno de los primeros pozos del potencial, la pérdida de energía del *kink* por radiación puede considerarse despreciable, y por tanto es otro mecanismo diferente el que hace que el *kink* quede atrapado en un pozo cercano. Teniendo en cuenta esta separación, nosotros nos ocuparemos sólo de los casos en los que no intervienen procesos radiativos, esto es, los casos en los que el *kink* queda atrapado poco después de partir de la condición inicial, que son las simulaciones para λ entre 3 y 5 en el modelo completo de sG y λ entre 3 y 4 en el modelo completo de ϕ^4 . El caso para $\lambda = 2$ es complejo en ambos casos, ya que en sG la velocidad del *kink* descende poco a poco hasta que se queda parado, y en ϕ^4 parece que en un punto se queda casi parado, aunque luego empieza a moverse con velocidad casi constante, aunque menor que la inicial.

Discusión de resultados.

Podemos ahora comparar ambos resultados entre sí, los del modelo completo de Klein-Gordon con los de la aproximación en coordenadas colectivas. La conclusión es clara: la aproximación de dos coordenadas colectivas predice cualitativamente la aparición del fenómeno de competición de escalas del modelo completo en los casos de atrapamiento en los primeros pozos de potencial, aunque no sea capaz de dar datos precisos de los valores de λ para los que se produce dicho proceso.

Lo importante de esta conclusión es que aquí no intervienen en modo alguno fenómenos de radiación, fonones, etc. Como mostré anteriormente, en el estudio de la perturbación (3.2) con una coordenada colectiva no se obtuvo proceso alguno de atrapamiento, ya que el centro del *kink* superaba todas las barreras del potencial para cualquier velocidad inicial no nula [ver la expresión (3.7)]. Al introducir la segunda coordenada colectiva es cuando se observa el fenómeno de competición de escalas al resolver el sistema de EDOs (2.17). Por tanto, la explicación más razonable de la competición de escalas involucrará un fenómeno de acoplamiento de la anchura del *kink*, $l(t)$, con la posición de su centro, $X(t)$. Dicho acoplamiento es crítico para ciertas longitudes de onda de la perturbación, λ , de manera que afecta de forma drástica a la velocidad del centro del *kink*, $\dot{X}(t)$.

Hemos llegado a esta conclusión en el caso en el que la perturbación no es demasiado pequeña, $\epsilon = 0.7$, por haber partido de un trabajo anterior en esta materia (Sánchez y Bishop 1994). Sin embargo, cabe plantearse si este resultado sigue siendo válido para perturbaciones de menor amplitud. Para ello realizamos simulaciones numéricas análogas a las anteriores para el modelo completo de sG (3.2) y su aproximación al sistema de EDOs (3.18) dada por las dos coordenadas colectivas, pero cambiando el valor de ϵ a 0.1. El resultado aparece en la figura 3.4. En ella se observa que la predicción dada por el modelo EDO de sG da resultados cuantitativos en la primera cuarta parte de la simulación. Posteriormente entran en juego fenómenos radiativos que restan energía al *kink*, haciendo que se decelere ligeramente para ciertos valores de λ . Pese a esta diferencia, la predicción del modelo aproximado es muy buena, con un error relativo pequeño. Sin embargo, al igual que en el caso de $\epsilon = 0.7$, lo más importante es la predicción de existencia o no de competición de escalas.

Hemos visto que para $\epsilon = 0.1$ y velocidad inicial igual a 0.5 (ya sea en el modelo EDP o en el EDO) no existe atrapamiento del *kink* en ningún pozo del potencial. Sin embargo, para velocidades iniciales muy pequeñas volvemos a encontrar sucesos de atrapamiento del *kink* en los dos casos, como se puede ver en la figura 3.5. Para conseguir este comportamiento de competición de escalas con $\epsilon = 0.1$ ha sido necesario utilizar una velocidad inicial de 0.005. De nuevo, se puede ver cómo el modelo EDO predice cualitativamente el comportamiento del modelo EDP.

Ante estos resultados, parece natural plantearse cómo puede ser que para un valor tan pequeño de la perturbación, $\epsilon = 0.1$, encontremos resultados que, aunque cualitativamente aceptables, no conllevan una mejora cuantitativa respecto a los resultados de $\epsilon = 0.7$. ¿No tendría que ser mejor aproximación para valores pequeños de ϵ ? La respuesta está en que el método de coordenadas colectivas es más

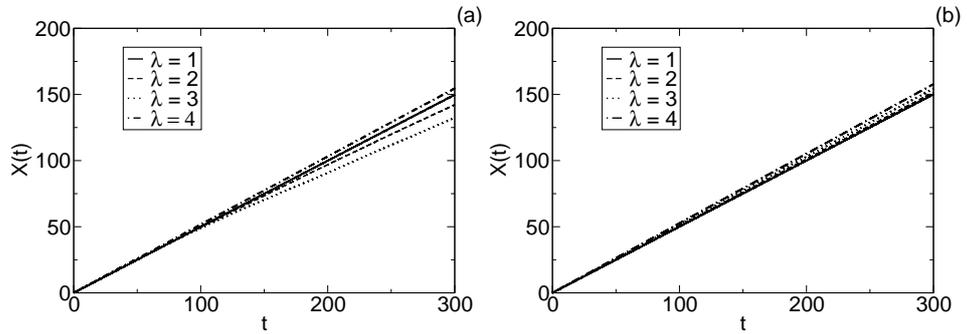


Figura 3.4: Resultado para el modelo de sG de la integración numérica (a) del sistema completo (3.2) y (b) del sistema de dos coordenadas colectivas (3.18) para varios valores de λ y $\epsilon = 0.1$. Las condiciones iniciales son, en (a), la expresión (2.19) con $x_0 = 0$ y $v = 0.5$, y en (b), $X(0) = 0$, $\dot{X}(0) = 0.5$, $l(0) = l_0(1 - \dot{X}(0)^2)^{1/2}$ y $\dot{l}(0) = 0$.

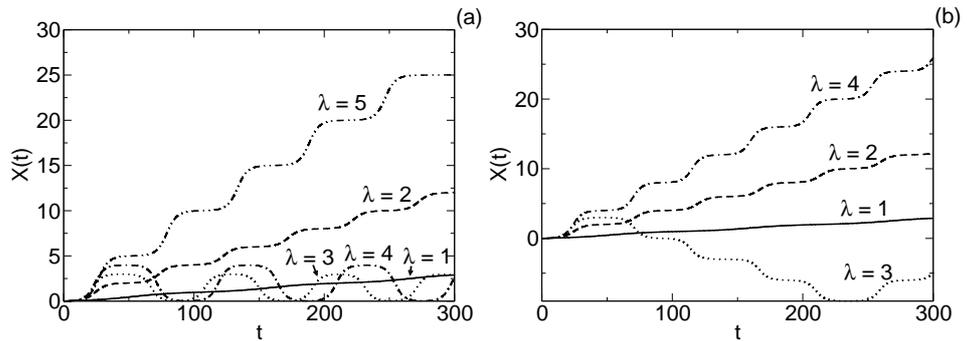
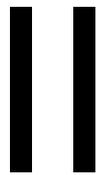


Figura 3.5: Resultado para el modelo de sG de la integración numérica (a) del sistema completo (3.2) y (b) del sistema de dos coordenadas colectivas (3.18) para varios valores de λ y $\epsilon = 0.1$. La condición inicial en (a) es la expresión (2.19) con $x_0 = 0$ y $v = 0.005$, y en (b) es el conjunto $\{X(0) = 0, \dot{X}(0) = 0.005, l(0) = l_0(1 - \dot{X}(0)^2)^{1/2}, \dot{l}(0) = 0\}$.

un método variacional que un método perturbativo. No existe ningún parámetro pequeño que pueda ser usado para controlar la aproximación. Por tanto, al utilizar un valor pequeño de ϵ no estamos garantizando que la aproximación se acerque más al resultado real del problema a estudiar. De hecho, como hemos visto, en el caso $\epsilon = 0$, la introducción de la anchura como coordenada colectiva sólo da la solución exacta para $\dot{l}(t) = 0$, siendo incorrecta en el resto de los casos. La mejor o peor aproximación dependerá de cómo se comporte el sistema en cada caso particular, es decir, de si la solución viajera se acerca más o menos al *Ansatz* que hemos propuesto como aproximación. Lo peor de todo esto es que, normalmente, sólo podremos comprobarlo *a posteriori*, después de haber comparado el resultado de ambos casos. Esto hace que este método, aunque muy útil e intuitivo en algunos casos, tenga que ser utilizado con la preocupación de saber lo que realmente se está estudiando en cada problema. Así, en nuestro caso, este método nos ha dado información de cómo a partir del acoplamiento entre la posición del *kink* y su anchura aparece atrapamiento de *kinks* para diferentes longitudes de onda de la perturbación. Esto nos da información sobre los mecanismos que pueden estar actuando sobre el *kink* en el modelo completo, pero no podemos ir más allá y suponer que estos resultados nos dan una predicción cuantitativa del problema.



sG discreto

4

El modelo homogéneo de sine-Gordon discreto.

En este capítulo se introduce el modelo discreto de sG homogéneo, empezando por su aparición como modelo de interacción de partículas con próximos vecinos y su posterior aplicación a la física del estado sólido. Posteriormente se introduce el modelo de péndulos, muy útil para entender y visualizar gran parte del contenido de esta tesis. A continuación se explica el potencial de Peierls-Nabarro, la principal diferencia con el modelo continuo, que aparece por efecto de la discretización del sistema. Como última aplicación del modelo homogéneo de sG discreto, se introduce el modelo de Engländer del ADN, que será muy utilizado en lo que resta de memoria de tesis.

4.1. El modelo de Frenkel-Kontorova.

Introducción del modelo¹

A finales de los años 20, Ludwig Prandtl (Prandtl 1928) y Ulrich Dehlinger (Dehlinger 1929) presentaron un modelo muy sencillo² que describe la dinámica

¹La mayor parte de las notas de esta sección se han obtenido de Braun y Kivshar (2004).

²Realmente, el trabajo de Prandtl fue realizado en los años 1912 y 1913, pero no fue publicado hasta 1928. Dehlinger realizó su trabajo de forma totalmente independiente y, de hecho, aunque

de una cadena de partículas que interactúan elásticamente con sus próximos vecinos en presencia de un potencial periódico externo. Posteriormente, el modelo fue introducido de manera independiente por Yakov I. Frenkel y Tatyana A. Kontorova (Frenkel y Kontorova 1939). El Hamiltoniano del modelo es

$$H = \sum_n \left\{ \frac{m_a}{2} \left(\frac{dx_n}{dt} \right)^2 + \frac{K}{2} (x_{n+1} - x_n - a_0)^2 + \frac{\epsilon_s}{2} \left[1 - \cos \left(\frac{2\pi x_n}{a_s} \right) \right] \right\}, \quad (4.1)$$

donde m_a es la masa de cada partícula, K la constante elástica de la interacción, a_0 la distancia de equilibrio del potencial de interacción (en ausencia del potencial externo), ϵ_s la amplitud del potencial externo y a_s su periodo. Según el propio Prandtl, dicho modelo retiene las características esenciales de la dinámica de dislocaciones en cristales: los átomos que se encuentran entre dos capas cristalinas perfectas tienen una dirección en la que pueden moverse fácilmente; están elásticamente unidos a posiciones de equilibrio relativas a sus próximos vecinos; además, se encuentran sujetos a fuerzas periódicas en la posición que provienen de las capas cristalinas adyacentes.³

El Hamiltoniano (4.1) se puede dejar en función de un único parámetro que defina el sistema. Para ello, basta con dividir la expresión (4.1) por $\epsilon_s/2$ y reescalar x_n y t para llegar a la expresión

$$H = \sum_n \left\{ \frac{1}{2} \left(\frac{dx_n}{dt} \right)^2 + \frac{1}{2d^2} (x_{n+1} - x_n - a_0)^2 + (1 - \cos x_n) \right\}, \quad (4.2)$$

donde el parámetro $d = \sqrt{\epsilon_s/2K}$ es un parámetro adimensional que representa la discretización del sistema (más adelante describiré en más detalle este parámetro).

Las ecuaciones de movimiento correspondientes a este Hamiltoniano son

$$\frac{d^2 x_n}{dt^2} + \sin x_n - \frac{1}{d^2} (x_{n+1} - 2x_n + x_{n-1}) = 0. \quad (4.3)$$

Mirando la ecuación (4.3), se observa que no incluye de manera explícita el parámetro a_0 del Hamiltoniano. De hecho, la ecuación es la versión discreta de la ecuación de sG (2.9) que estudiamos en el capítulo anterior. En 1948, Alfred Seeger, uno de los alumnos de Dehlinger, llegó a la ecuación (2.9) a partir de (4.3) en su *Diplomarbeit*. Después, en el único libro de geometría diferencial que tenía

las aproximaciones de ambos a la solución de las ecuaciones básicas son totalmente diferentes, Dehlinger reconoció los resultados alcanzados anteriormente por Prandtl.

³Obtenido de Seeger (1980).

en su estantería, encontró referencias⁴ a los trabajos de Bianchi (Bianchi 1879; Bianchi 1892) y Bäcklund (Bäcklund 1883) que ya mencionamos en el capítulo 1. Desde entonces han aparecido muchos trabajos relacionados con este modelo, tanto analíticos como experimentales, que involucran *kinks*, *breathers*, modos normales, aplicaciones... que aún hoy siguen produciéndose.

4.2. Modelo mecánico.

Cadena de péndulos.

Hemos introducido el modelo de sG discretizado utilizando el modelo de Frenkel-Kontorova por motivos históricos, ya que fue la primera aplicación de la ecuación de sG (2.9) a un sistema físico⁵. De hecho, me referiré a él como modelo discreto de sG en lugar de modelo de Frenkel-Kontorova para no complicar la notación.

Sin embargo, para entender mejor la dinámica o el comportamiento del modelo discreto de sG en determinadas situaciones, y las diferencias que existen con el modelo continuo, conviene introducir un modelo mecánico que, además, va a resultar muy útil cuando apliquemos el modelo de sG al ADN.

Este modelo macroscópico lo introdujo Scott (1969). Consiste en una cadena de péndulos acoplados elásticamente entre sí. Para entenderlo, imaginemos una cuerda o goma elástica, de forma cilíndrica y sección transversal despreciable frente a su longitud, con sus extremos sujetos de manera que la goma quede tirante. Podemos imaginar unos mecanismos que sujetan los extremos de manera que la goma tenga la libertad de rotar sobre su eje y las condiciones de contorno las podamos fijar nosotros como queramos. Supongamos ahora que, sobre la goma tensa y en equilibrio, sin torsiones, se pinchan N alfileres de costura mirando hacia abajo a lo largo de la cadena, unos al lado de otros, de manera que todos los alfileres y la goma estén contenidos en un único plano. Los alfileres los supondremos de tal manera que toda su masa esté concentrada en su cabeza, como péndulos. Por ahora nos limitamos a visualizar un sistema finito de N alfileres o péndulos porque es más sencillo, aunque cuando lo hayamos entendido bien podemos dar el paso al infinito. El resultado sería aproximadamente el de la figura 4.1.

⁴Curiosamente, él llamó a las ecuaciones (2.1) y (2.9) ecuaciones de Enneper (Enneper 1870), ya que no conocía el trabajo de Bour (Bour 1862). El nombre de "sine-Gordon" se lo puso Martin Kruskal muchos años después, por la analogía con el modelo de Klein-Gordon (Rubinstein 1970).

⁵En el capítulo 1 vimos que los primeros estudios que lo utilizan, aparte de Frenkel y Kontorova (1939), son de los años 60.

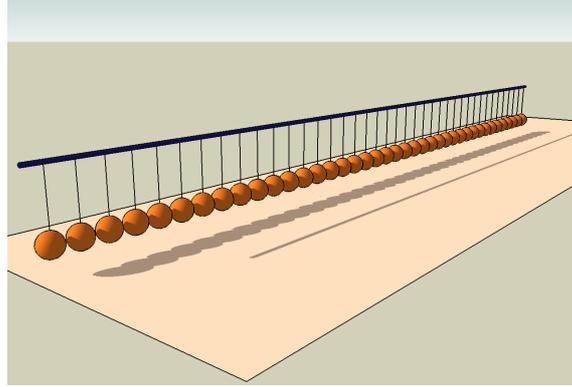


Figura 4.1: Modelo mecánico de una cadena de péndulos acoplados.

Hamiltoniano del modelo.

El Hamiltoniano de este sencillo modelo mecánico es

$$H(\{\phi_n\}) = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} I \dot{\phi}_n^2 + \frac{1}{2} K (\phi_{n+1} - \phi_n)^2 + mgL(1 - \cos \phi_n) \right\}, \quad (4.4)$$

donde ϕ_n es el ángulo respecto a la vertical del péndulo n -ésimo, $\dot{\phi}_n$ es la derivada temporal de ϕ_n , I el momento de inercia de los péndulos, K la constante elástica de la goma en torsión, y mgL la energía potencial de cada péndulo en posición horizontal.

El Hamiltoniano (4.4) se parece mucho al (4.2), para $a_0 = 0$. En este caso, el primer término del Hamiltoniano (4.4) se corresponde de esta manera con un término cinético de cada péndulo relacionado con su momento de inercia. El segundo término nos da la interacción de cada péndulo con sus próximos vecinos, debida a la torsión elástica de la goma. Y por último, el tercer término nos da el potencial gravitatorio que actúa sobre la cabeza de cada péndulo. Lógicamente, la posición de equilibrio del sistema será la que tiene todos los péndulos mirando hacia abajo y la goma sin torsión. Además, y de manera intuitiva, se entiende que si los péndulos no pesan casi nada en comparación con la rigidez de la goma, entonces al hacer girar uno de los péndulos con la mano haremos girar todo el sistema de manera rígida, o con muy poca deformación. De la misma forma, unos péndulos muy pesados y una goma elástica muy débil hará que podamos mover cada péndulo de manera casi independiente con respecto al resto. Los casos interesantes se producen entre medias, cuando el movimiento forzado de unos de

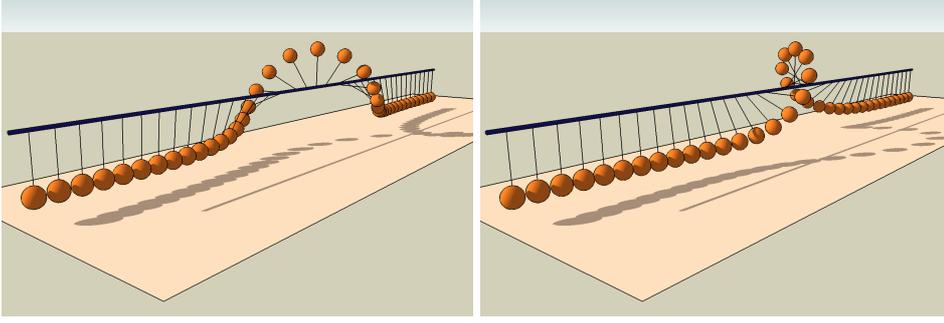


Figura 4.2: Imagen de un *kink* (izquierda) y de un *antikink* (derecha) en el modelo mecánico. El sentido de giro se ha definido suponiendo que el péndulo más cercano es el $n = 0$, el más lejano es el $n = N$ y utilizando la regla de la mano derecha.

los péndulos trae consigo la deformación de los vecinos en la cadena. Es bonito ver cómo aparece de manera tan gráfica una *discretización efectiva* del modelo dependiendo de la intensidad relativa entre los términos segundo y tercero del Hamiltoniano (4.4), mientras que el primer término nos da la escala de tiempos en la que observar los fenómenos. Por eso, en la escala de tiempos adecuada, y de la misma manera que para el modelo de Frenkel-Kontorova, podemos adimensionalizar el Hamiltoniano (4.4) y escribirlo en función de un único parámetro, $d = \sqrt{mgL/K}$. Multiplicando, además, el Hamiltoniano (4.4) por la discretización, se obtiene

$$H(\{\phi_n\}) = d \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \dot{\phi}_n^2 + \frac{1}{2d^2} (\phi_{n+1} - \phi_n)^2 + (1 - \cos \phi_n) \right\}. \quad (4.5)$$

El Hamiltoniano (4.5) así obtenido es una versión discreta del Hamiltoniano continuo (2.7), y el parámetro d representa la discretización efectiva utilizada al realizar la aproximación $\phi_x \simeq (\phi_{n+1} - \phi_n)/d$. Debido a que este Hamiltoniano es la discretización de una integral, el parámetro d aparece multiplicando el sumatorio para normalizar el resultado. Así se consigue que, en el límite $d \rightarrow 0$, la energía del modelo discreto y del continuo sean iguales.

Las ecuaciones de movimiento asociadas al Hamiltoniano (4.4) son

$$\ddot{\phi}_n - \frac{1}{d^2} (\phi_{n+1} - 2\phi_n + \phi_{n-1}) + \sin \phi_n = 0, \quad (4.6)$$

las mismas que las del modelo de Frenkel-Kontorova, (4.3). De nuevo, dichas ecuaciones de movimiento son las mismas que (2.9) discretizadas. Debido a esta discretización, las soluciones exactas que presenta el modelo continuo dejan de serlo en el modelo discreto, aunque constituyen una muy buena aproximación para valores pequeños de d . Más adelante estudiaremos los efectos más importantes que produce la discretización del sistema, y que lo diferencian del modelo continuo.

Representación de solitones.

Pese a estas diferencias, con este modelo ahora es posible tener una imagen más intuitiva de las soluciones del sistema, a diferencia de la visión puramente matemática de que hablé en el capítulo anterior. Así, echando un vistazo a la expresión del *kink* (2.19) para $v = 0$ y discretizando de la forma $x = nd$, vemos que no es más que una vuelta de la goma sobre sí misma, estática y localizada, de manera que los péndulos, empezando por el $n = 0$ y terminando en el $n = N$, pasan de $\phi_0 = 0$ a $\phi_N = 2\pi$ poco a poco (la suavidad con que lo hacen depende de la discretización efectiva del sistema). Podemos utilizar la regla de la mano derecha para definir un sentido de giro en el ángulo ϕ_n , y así diremos que el *kink* gira en sentido horario o a derechas, mientras que el *antikink* lo hace en sentido antihorario o a izquierdas. En la figura 4.2 se representan ambas soluciones en el modelo mecánico. De esta manera, en el caso del *kink*, el péndulo N tiene una vuelta más que el péndulo 0 (o una menos para el *antikink*), y es por eso que se dice que el *kink* (*antikink*) tiene carga topológica⁶ igual a 1 (-1). Si fijamos los extremos de la goma, la carga topológica se tiene que conservar: esto hace de los *kinks* objetos muy estables, porque si hemos retorcido la goma con una vuelta ya no la podemos dejar como al principio por mucho que la giremos en otro sitio (siempre en el caso de que fijemos los extremos). Si deformamos poco la goma, al soltarla recuperará su forma de *kink*. Bajo deformaciones muy grandes, conseguiremos que haya más *kinks* y *antikinks* a lo largo de la goma, pero el número de vueltas *total* seguirá siendo el mismo. Las soluciones tipo *breather* en este modelo equivalen a oscilaciones acopladas de un grupo localizado de péndulos, siendo mayor la amplitud de oscilación en el centro del grupo, y decreciendo conforme nos alejamos de él.

⁶También se le llama número de *winding*, que significa algo así como *retorcimiento*, en español; es fácil entender por qué se le llama así imaginando este modelo de péndulos. También se entiende mejor el porqué del nombre de *kink* y *antikink*.

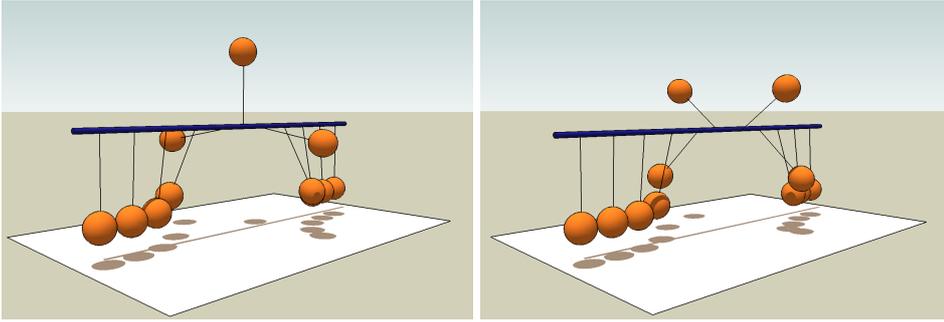


Figura 4.3: Diferentes configuraciones del sistema de péndulos con un *kink*, para un valor grande de discretización ($d = 1$). Se observa que las dos configuraciones no son equivalentes, siendo su única diferencia la posición de su centro, que en la de la izquierda está en un nodo de la red y en la segunda está en la mitad de dos nodos, y por tanto se pierde la simetría de traslación del sistema.

4.3. Efectos de la discretización.

El potencial de Peierls-Nabarro.

Como ya mencioné en el apartado anterior, las soluciones exactas del modelo continuo de sG dejan de serlo en el modelo discreto, aunque constituyen una buena aproximación para valores pequeños de d . El principal motivo por el que las soluciones dejan de ser exactas es porque se pierde la simetría de traslación de la ecuación (2.9). Supongamos la ecuación de sG (2.9) discretizada según el parámetro d , de manera que $x = nd$. Supongamos una solución en reposo localizada en x_0 , por ejemplo un *kink* (2.19) con $v = 0$. Si ahora trasladamos el *kink*, la configuración del sistema dejará de ser equivalente a la anterior, a menos que lo traslademos una distancia md , con m entero. Un ejemplo gráfico del efecto de la discretización se observa en la figura 4.3. Debido a este efecto, la energía de las posibles configuraciones entre x_0 y $x_0 + d$ no es constante, como muestra la gráfica 4.4a. En ella está dibujada la variación de la energía dependiendo de la posición x_0 para diferentes valores de la discretización, sin más que aplicar la expresión (4.5). Esta energía se denomina *potencial de Peierls-Nabarro* (PN), $V_{PN}(x_0)$, y fue introducida en el contexto de la teoría de dislocaciones en cristales por Peierls (Peierls 1940) y Nabarro (Nabarro 1947). A la diferencia de energía entre el mínimo y el máximo del potencial se le denomina *barrera de Peierls-Nabarro*, E_{PN} . El valor de E_{PN} en función de d aparece dibujado en la figura 4.4b.

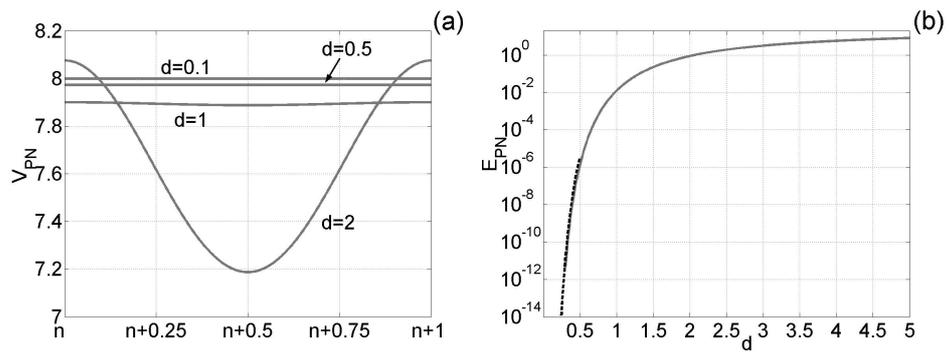


Figura 4.4: (a) Potencial de PN para diferentes valores de la discretización, d . El eje de abscisas representa la posición del centro del *kink*, relativa a dos puntos de la red, $x_0/d - n$, con n entero, de manera que el 0 se corresponde con una posición de la red y el 1 con la siguiente. La energía se ha calculado utilizando la expresión (4.5). (b) Línea continua: barrera de PN en función de la discretización del sistema, d ; el error producido en el cálculo numérico de E_{PN} es tan grande que los resultados calculados para valores menores de $d = 0.3$ no aparecen. Línea discontinua: aproximación (4.7) de E_{PN} para valores pequeños de d .

El efecto de la discretización se hace patente en el potencial de PN. La primera causa directa de este potencial es que no todas las soluciones de *kink* (2.19) en reposo y discretizadas son soluciones estacionarias de la ecuación discreta (4.6). Las posiciones de equilibrio inestable vendrán dadas por los máximos del potencial PN, y las de equilibrio estable por los mínimos. Como se desprende de la figura 4.4, dichas configuraciones se corresponden con los valores $x_{0,u} = md$ y $x_{0,s} = (m + 1/2)d$, respectivamente, y son precisamente los que aparecen representados en la figura 4.3. Por tanto, un sistema que se encuentre inicialmente en reposo en una configuración fuera del equilibrio, oscilará en torno a la posición de equilibrio estable. Debido a esta oscilación, el *kink* emitirá fonones hasta que alcance el estado estacionario correspondiente al mínimo del potencial de PN. Además, debido a este potencial, el *kink* quedará ligeramente “comprimido” y su anchura sufrirá un ligero cambio. Otra consecuencia del potencial de PN es que deja de haber soluciones exactas en movimiento. Al moverse en un potencial oscilatorio, la velocidad del *kink* ya no es constante, sino que también oscila, por lo que emite fonones y pierde energía. Con el tiempo, la energía cinética se hará tan pequeña que, cuando la velocidad sea menor que una cierta velocidad crítica v_{PN} , el *kink* quedará atrapado en el potencial PN y oscilará en torno a un mínimo hasta que se quede parado en él.

Dependencia con la discretización.

Ahora bien, aunque este efecto existe para cualquier discretización, su valor puede despreciarse para valores pequeños de d , como puede observarse en la figura 4.4b. En los estudios del modelo de sG realizados durante la elaboración de esta tesis, hemos elegido un valor de discretización $d \leq 0.1$. El valor de la barrera para dicha discretización no se ha podido calcular de forma numérica en la figura 4.4, ya que existe demasiado error numérico debido a la precisión del ordenador para $d \simeq 0.1$. Sin embargo, existen estudios analíticos para estimar esta barrera en el límite continuo (Pokrovsky 1981; Braun y Kivshar 2004), que dan como aproximación de la barrera de PN

$$E_{PN} \simeq \frac{64\pi^2}{d} e^{-\pi^2/d}. \quad (4.7)$$

Dicha curva aparece dibujada también en la figura 4.4b. El valor de la barrera para $d = 0.1$ utilizando esta aproximación es $E_{PN} \simeq 10^{-38}$, un error que en las simulaciones queda totalmente borrado por la precisión de la máquina, y por tanto no hemos tenido en cuenta el efecto de la discretización en los resultados de las simulaciones.

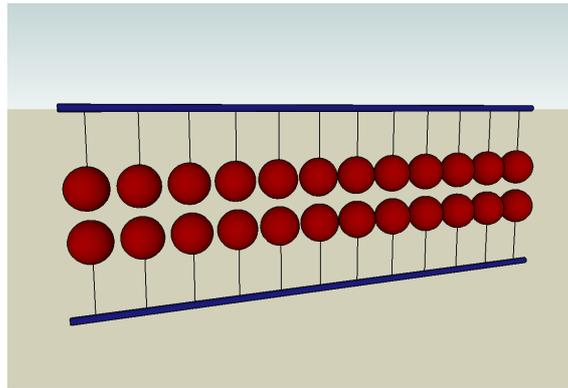


Figura 4.5: Aproximación del ADN como cadena de péndulos, tal y como fue propuesto en Englander et al. (1980).

4.4. El modelo de Englander del ADN.

Solitones en ADN.

El modelo de sG discretizado también ha sido utilizado como modelo del ADN. Dicho modelo fue introducido por Englander et al. (1980) a partir de los resultados obtenidos en un trabajo anterior (Mandal et al. 1979) en el que se estudian los parámetros que describen la cinética del intercambio de protones por bases abiertas de la doble cadena de ADN. En estos experimentos el ADN se coloca en una solución de agua pesada (agua con alto contenido en deuterio), y se observa que los protones imino, que son los protones que forman los enlaces de hidrógeno entre las dos bases de un par de bases, son intercambiados por deuterio proveniente del solvente. Puesto que estos protones están profundamente enterrados en la estructura del ADN, su intercambio significa que las bases pueden abrirse, al menos temporalmente, y exponer los protones imino al solvente. Englander et al. (1980), para explicar tiempos de vida de estados abiertos de hasta un segundo, introducen la hipótesis de que existe un estado abierto, a modo de burbuja, de una longitud de unos 10 pares de bases y que es capaz de difundir por la doble cadena de ADN. Como explicación de la existencia de dicha burbuja, proponen el modelo de sG discreto aplicado al ADN, donde las burbujas serían *kinks* que se propagan por la doble cadena.

Para utilizar el modelo de sG en el ADN, Englander y sus colaboradores tuvieron que hacer una serie de suposiciones que convirtieron la doble hélice de

ADN en el modelo mecánico de péndulos que describí anteriormente⁷. Para ello, aproximaron la doble hélice de ADN por una estructura plana en la que las dos cadenas de ADN están representadas por dos varillas paralelas, unidas por los pares de bases mediante enlaces de hidrógeno, como se muestra esquematizado en la figura 4.5. Dejando una de las cadenas de ADN fijas, para que un par de bases se rompiera era necesario que la base de la otra cadena girara, como si fuera un péndulo, en el plano perpendicular a la varilla. El análogo con el modelo de péndulos ya estaba prácticamente hecho. Sólo hacía falta suponer un acoplamiento armónico entre bases adyacentes y un “potencial gravitatorio” que cada base de enfrente ejerciera sobre su compañera y que favoreciera que éstas permanecieran unidas por el enlace de hidrógeno. La ecuación de movimiento del sistema sería entonces

$$I\ddot{\phi}_n - K(\phi_{n+1} - 2\phi_n + \phi_{n-1}) + E_H \sin \phi = 0, \quad (4.8)$$

donde I es el momento de inercia de las bases, K la constante de acoplamiento armónico entre bases adyacentes y E_H la energía de un enlace de hidrógeno. Dividiendo la ecuación por E_H y reescalando en el tiempo se vuelve a obtener la ecuación sG discreta adimensionalizada, (4.6).

Análogo al modelo mecánico.

Con estos ingredientes cocinaron un modelo mecánico para el ADN, análogo al modelo de péndulos. Tal y como Englander y sus colaboradores explicaron, aunque este modelo parece una simplificación excesiva del modelo real, el modelo de sG contiene la principal característica de que, al abrirse la cadena, el enlace se rompe alrededor de $\phi_n = 0$. Además, en cualquier caso, puntualizaron, se ha demostrado que la existencia de soluciones de tipo solitón es robusta bajo cambios en “detalles” del potencial. Con este modelo bajo el brazo, y suponiendo la formación de *kinks* mediante fluctuaciones térmicas, sus resultados encajaban con los resultados experimentales del intercambio de protones (Mandal et al. 1979). La era de los solitones en ADN había comenzado, a pesar de que, al final del artículo de Englander et al. (1980), se aclaraba explícitamente que no había ninguna evidencia experimental de estados abiertos ni de excitaciones de tipo solitón que se movieran por la doble cadena de ADN, sino que lo que se había obtenido era consistente con los datos experimentales y podría ser un posible estado excitado de la doble cadena.

Posteriormente a este trabajo, Guéron et al. (1987) mostraron que los resultados de Mandal et al. (1979) eran erróneos y que se debían a una malinterpretación

⁷Para una descripción más detallada de la estructura del ADN, véase el apéndice B.

de los datos del intercambio de protones imino, dando lugar a tiempos de vida más razonables, del orden de microsegundos.⁸ Pese a este nuevo resultado, que cuestionaba la consistencia de la investigación de Englander et al. (1980) con los resultados experimentales, aparecieron con posterioridad diferentes trabajos sobre solitones, tales como *kinks* o *breathers* que se mueven en ADN.

Existe otro problema fundamental del modelo, que proviene de no tener en cuenta los efectos de la viscosidad del agua sobre la dinámica del ADN. En la escala de fuerzas de la célula, del orden de piconewtons, el agua actúa como un fluido muy viscoso, donde el término inercial es despreciable frente al disipativo⁹. Un intento de describir la dinámica del ADN de forma realista que no incluya este efecto carece de validez. Sin embargo, se pueden estudiar algunas propiedades del modelo desde un punto de vista fenomenológico, como veremos en los capítulos siguientes.

⁸La física y la biología no siempre se han llevado bien; un ejemplo de ello es la controversia que ha habido en el tema de solitones que se propagan por cadenas de ADN. Así, Kamenetskii (1987) habla de intentos de “secuestrar” el ADN por parte de los físicos, mostrando claramente la opinión de algunos biólogos al respecto.

⁹Para una explicación de este fenómeno, véase el apéndice B.

5

Modelo de Salerno del ADN.

En este capítulo se describen de forma muy detallada algunos resultados basados en el modelo de Englander del ADN en los que se estudian secuencias inhomogéneas, tales como secuencias genómicas o secuencias aperiódicas, a partir de los resultados de la dinámica de *kinks* que se mueven por ellas. Estos trabajos relacionan de forma positiva la dinámica de los *kinks* con la posición de promotores o con el contenido de información de la misma. Posteriormente se introduce el potencial efectivo asociado a la dinámica de *kinks* en el modelo discreto. Al final del capítulo, como parte del trabajo original de esta tesis, se hace un estudio comparativo de la dinámica de *kinks* en secuencias inhomogéneas con la dinámica de una partícula que se mueve en el potencial efectivo asociado a dicha secuencia.

5.1. Secuencias de ADN.

Funcionalidad de la secuencia.

Como hemos visto, Englander et al. (1980) introdujeron el concepto de solitón propagándose por la doble cadena de ADN para explicar datos experimentales del intercambio de protones imino en las bases. La utilización de solitones para entender la funcionalidad del ADN fue sugerida unos años después por Polozov y

Yakushevich (1988). Más concretamente, en este estudio se relacionaba la posible implicación de cambios conformacionales en la estructura del ADN que se propagan por la doble cadena con la unión de proteínas específicas del ADN, como la ARN polimerasa en el proceso de transcripción de proteínas¹. Pero fue Mario Salerno (1991) el primero que introdujo la secuencia específica de bases de una determinada molécula de ADN para relacionarlo con el reconocimiento de la ARN polimerasa a la hora de transcribir un determinado gen y, posteriormente, descodificar la proteína.

El problema resultaba extremadamente interesante: por aquel entonces ya se conocía la existencia de *promotores*² en la secuencia genética que hacían que la ARN polimerasa transcribiese los genes de determinados organismos. Sin embargo, la distancia entre el promotor y el gen era en ocasiones de más de 1000 bases, así que Salerno se preguntó de dónde procede la energía de la excitación no lineal (los cambios conformacionales que mencioné antes) necesaria para viajar tales distancias. La pregunta no era trivial y, de hecho, si el comportamiento del modelo de Englander pudiera describir posiciones funcionalmente relevantes de la secuencia genética (tales como promotores), éste sería un método barato, rápido y eficiente para la genómica, que podría encontrar cómodamente dichas posiciones.

Implementación del modelo.

Puesto que había que introducir la secuencia en el modelo, Salerno (1991) tuvo en cuenta que, de los cuatro tipos de bases diferentes, adenina (A), timina (T), citosina (C) y guanina (G), los únicos pares de bases posibles al formarse la doble cadena de ADN son AT y CG. Además, el número de enlaces de hidrógeno en estos pares de bases difiere, siendo dos enlaces de hidrógeno en el par AT y tres enlaces de hidrógeno en CG. En la figura 5.1 está esquematizado el modelo. Con estos datos, Salerno introdujo una inhomogeneidad en la secuencia de bases del modelo de Englander, de manera que la ecuación de movimiento pasó a ser

$$\ddot{\phi}_n - \frac{1}{d^2}(\phi_{n+1} - 2\phi_n + \phi_{n-1}) + q_n \sin \phi_n = 0. \quad (5.1)$$

En esta ecuación, la inhomogeneidad aparece en el parámetro q_n , que antes era constante (e igual a uno)³. Dicho parámetro hace que la relación entre la intensi-

¹Para una explicación detallada del proceso de transcripción, véase el apéndice B.

²Para una definición detallada de promotor, véase el apéndice B. De forma poco precisa, un promotor es una secuencia fija de bases a las que se acopla la ARN polimerasa para comenzar el proceso de transcripción de un segmento de ADN que contiene uno o más genes.

³La expresión que Salerno utilizó no fue exactamente ésta, sino que colocó el parámetro de discretización en el término del potencial, multiplicando en lugar de dividiendo. Los resultados son

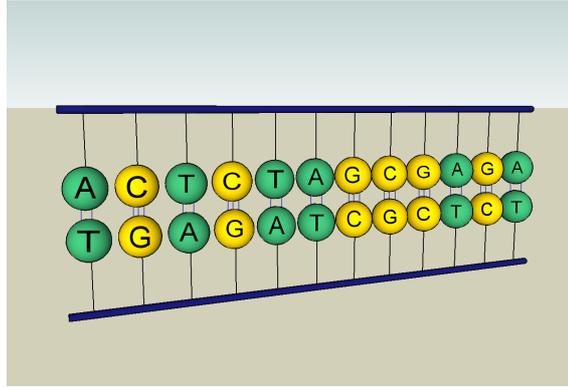


Figura 5.1: Modelo de Salerno del ADN que utiliza el esquema principal de Englander et al. (1980), pero que introduce la inhomogeneidad de la secuencia. Los dos colores diferentes en el modelo se corresponden con los dos tipos diferentes de pares de bases, los AT y los CG. Hay dos enlaces de hidrógeno en cada par AT y tres en cada par CG.

dad de los enlaces de hidrógeno (o la intensidad gravitatoria) y la intensidad del acoplamiento a próximos vecinos dependa del sitio n de cada base (o péndulo). Para relacionarlo con una secuencia de bases de un fragmento de ADN, el procedimiento es el siguiente: a la n -ésima base de la secuencia de bases se le asocia un parámetro q_n , que es q_{AT} en el caso de ser una A o una T, o q_{CG} en el caso de ser C o G. Estos valores son tales que $\frac{q_{AT}}{q_{CG}} = \frac{2}{3}$, que es la relación que existe entre los números de enlaces de hidrógeno de los dos pares de bases. Nótese que este parámetro influye en la discretización efectiva del problema, puesto que se puede dividir la ecuación (5.1) por $q_{avg} = (\sum_n q_n)/N$, donde N es el número de nucleótidos de la secuencia de ADN, y obtener una ecuación equivalente con una nueva discretización $dq_{avg}^{1/2}$ en otra escala de tiempos diferente. Esta discretización efectiva, $d_{eff} = dq_{avg}^{1/2}$, depende de la secuencia que se utilice.

Los valores que eligió para los parámetros fueron $q_{AT} = 2$, $q_{CG} = 3$ y $d \simeq 0.045$. Salerno eligió este último valor de un trabajo anterior (Fedyanin y Yakushevich 1984), donde se utilizaban valores de d^2 en torno a 10^{-3} y 10^{-4} . Con esos valores de q_{AT} y q_{CG} , la discretización efectiva puede variar entre

equivalentes, teniendo en cuenta la diferente escala de tiempos en un caso y en otro. Hemos elegido esta notación por consistencia con la notación anterior. Sin embargo, más adelante redefiniremos los parámetros utilizando una secuencia normalizada para evitar ambigüedades.

$d_{\text{eff}} = 0.063$, para una secuencia de sólo A y T, y $d_{\text{eff}} = 0.077$, para una secuencia de sólo C y G. En ambos casos, la discretización es tan pequeña que no influye el efecto del potencial de Peierls-Nabarro en los resultados obtenidos por integración numérica de (5.1).

Como se puede observar, la ecuación (5.1) no tiene en cuenta el efecto disipativo del agua que se mencionó en el capítulo anterior. Por el contrario, ni siquiera tiene término de rozamiento. Sin embargo, la fenomenología de este modelo, descrito por la ecuación (5.1), puede estar relacionada con el ADN de la forma descrita por Salerno; es decir, el estudio de la dinámica de *kinks* en secuencias genómicas puede relacionarse con la posición de elementos funcionales en dicha secuencia, aunque el modelo no describa de manera realista el comportamiento del ADN. Así es como entenderemos este modelo a partir de ahora a lo largo de esta memoria de tesis.

Metodología y resultados.

Para investigar los efectos sobre la dinámica debidos a la secuencia de bases, Salerno (1991) integró numéricamente la ecuación (5.1), con $\{q_n\}$ correspondiente a una secuencia S del promotor A_1 del fago T7⁴. La secuencia utilizada tenía sólo 168 pares de bases (PB)⁵ por lo que, para evitar efectos de borde, construyó una secuencia mayor, S' , de 1000 PB. Para ello utilizó la siguiente regla,

$$S'(1, 1000) = S(1, 5) + 8S(1, 50) + S(1, 168) + 15S(141, 168) + S(162, 168), \quad (5.2)$$

donde $kS(i, j)$ significa una secuencia que es k veces repetida la subsecuencia que va de la posición i a la j de la secuencia S . La forma en la que se construyó esta secuencia será clave para un posterior análisis de los resultados. Una vez conseguida la secuencia a estudiar, el trabajo de Salerno (1991) se centró en una región particular del promotor, la de la subsecuencia $S(51, 140)$, que se corresponde con la región $S'(456, 545)$. En dicha región se encuentra la secuencia de bases a la que la ARN polimerasa puede unirse en su proceso de transcripción, así como el sitio de comienzo de la transcripción (TSS)⁶, que se encuentra en la posición (o PB)

⁴Para conocer más detalles sobre el fago T7 y su genoma, véase el apéndice B.

⁵Realmente, la secuencia completa del fago T7 se conocía desde 1983 (Dunn y Studier 1983) y tiene 39936 pares de bases. Sin embargo, Salerno no conocía su existencia, probablemente debido a la dificultad que entrañaba encontrar información sobre resultados de investigación en biología celular para cualquiera que no fuera un experto en el tema, y probablemente debido a su vez a la falta de los buscadores de internet de que disponemos hoy en día para encontrar dicha información.

⁶Las siglas provienen del inglés, *transcription start site*. Para una mejor explicación del TSS y de las secuencia de bases de los promotores del fago T7, véase el apéndice B.

121 de la secuencia S y PB 526 de la secuencia S' . A continuación se realizaron simulaciones numéricas que consistieron en la integración de la ecuación (5.1), con condiciones de contorno fijas, partiendo de un *kink*, (2.19), discretizado y en reposo para diferentes posiciones iniciales en la secuencia central⁷.

En una secuencia homogénea, con esta discretización tan pequeña, un *kink* en reposo continúa en reposo al realizar la integración numérica, por ser solución estacionaria de la ecuación continua (2.9). Sin embargo, en una secuencia inhomogénea, el *kink* puede moverse espontáneamente a un lado o a otro, o permanecer en reposo, dependiendo del sitio en el que esté. La explicación a este comportamiento la veremos en detalle cuando introduzcamos el potencial efectivo en sistemas discretos aunque, en definitiva, lo que sucede es algo análogo a lo que sucede en el modelo continuo: la inhomogeneidad de las fuerzas que actúan a los dos lados del centro del *kink* hace que éste adquiera una aceleración y se mueva en un sentido y no en otro. La hipótesis introducida por Salerno (1991) es que este movimiento, provocado por la secuencia, se produce en regiones en la que la ARN polimerasa se une al ADN, en este caso en la región de S' (456, 545), mientras que en el resto de la secuencia este efecto es prácticamente nulo. Como veremos a continuación, los resultados que obtiene se ajustan bastante bien a esta hipótesis.

Los resultados mostrados por Salerno (1991) para diferentes posiciones iniciales son los siguientes: (i) partiendo de PB 415, el *kink* permanece en reposo; (ii) desde PB 415 a PB 505, no se aprecian diferencias significativas respecto al caso anterior, salvo pequeñas oscilaciones en torno a la posición inicial; (iii) para PB 510, el *kink* se mueve hacia la izquierda de la cadena con una velocidad $v = 0.18$, se refleja en el borde y luego vuelve de nuevo a ser reflejado en la región del promotor con velocidad $v = 0.18$; (iv) este comportamiento se acentúa (es decir, la velocidad que adquiere el *kink* es mayor) cuando la posición inicial del *kink* se incrementa de PB 510 a PB 535, siendo mayor en este último, donde alcanza $v = 0.3$; (v) para BP 540, el *kink* alcanza una velocidad $v \simeq 0.08$ hacia la derecha, y poco a poco dicha velocidad va disminuyendo conforme partimos de posiciones más a la derecha⁸, hasta que (vi) en PB 555 el *kink* permanece en reposo. En otra simulación, partiendo de PB 900, y en esta ocasión con velocidad inicial $v = 0.3$ hacia la izquierda, se obtuvo como resultado que el solitón se acelera en la región promotora cuando viaja hacia la izquierda, luego es reflejado y,

⁷La dinámica obtenida partiendo de un *antikink* es la misma, como veremos más adelante, tras introducir el potencial efectivo. A partir de ahora, cuando nos refiramos a un *kink* que se mueve de acuerdo con la ecuación (5.1), realmente podrá tratarse de un *kink* o un *antikink*, puesto que sus dinámicas son las mismas.

⁸Salerno (1991) no especifica lo que ocurre entre PB 536 y PB 539.

Promotor	Región PB	Respuesta
A_1	De 510 a 535 540	Propagación hacia la izquierda, más fuerte en 535. Pequeña velocidad hacia la derecha.
A_0 (o D)	De 530 a 540 De 543 a 555	Propagación hacia la izquierda, más fuerte en 540. Propagación hacia la derecha, más fuerte en 543.
A_3	De 435 a 460	Propagación hacia la derecha.

Cuadro 5.1: Resumen de resultados de Salerno (1992), donde se estudia la dinámica de *kinks* moviéndose en la secuencias sintéticas S' obtenidas a partir de secuencias de 168 PB donde se encuentran los promotores A_1 , A_0 (también llamado D) y A_3 .

en su vuelta hacia la derecha, el solitón es decelerado en la misma región.

Salerno (1991) concluye que estos resultados muestran la existencia de una región dinámicamente “activa” que va desde PB 510 hasta PB 540 dentro del promotor A_1 del fago T7. De esta forma, cuando un *kink* estático se encuentra dentro de esta región, recibe un impulso que le permite viajar a lo largo de la cadena de ADN, dependiendo su sentido de movimiento de la posición inicial en la cadena y siendo el efecto menor cuando el *kink* se coloca a la izquierda de PB 535. Además, este comportamiento dinámico desaparece conforme el *kink* se coloca en posiciones más alejadas de esta región. Finalmente, propone que esta región dinámicamente “activa” podría ser la explicación del funcionamiento de los promotores de ADN como activadores del proceso de transporte de la ARN polimerasa. Debido a la dificultad de encontrar secuencias de promotores en el ADN simplemente buscando en la secuencia de bases de un genoma bacterial, este resultado es muy importante.

Al año siguiente de la publicación de este estudio apareció otro análogo (Salerno 1992) donde se mostraban los resultados asociados a la secuencia del promotor A_1 y a otras dos secuencias, las de los promotores A_0 (también llamado D) y A_3 . Las nuevas secuencias vuelven a ser de 168 nucleótidos, y de nuevo se coloca la secuencia $S(51, 140)$ en el centro de la nueva secuencia S' que se construye con la misma regla (5.2) utilizada en Salerno (1991). En este estudio se utilizó la misma metodología seguida en Salerno (1991) para cada secuencia. En la tabla 5.1 aparece un resumen de los resultados obtenidos en las tres secuencias. En el artículo se especifica que, puesto que el *kink* inicialmente en reposo se encuentra en la secuencia central de 168 PB, las secuencias utilizadas para prolongar la secuencia original no juegan ningún papel en los resultados de la dinámica. De

nuevo la conclusión de este último trabajo es que, puesto que las únicas regiones dinámicamente “activas” en cada secuencia son las que se encuentran en las proximidades de los promotores, éstos actúan como activadores energéticos de la ARN polimerasa en el proceso de transcripción de ARN. En el próximo capítulo matizaremos dichas conclusiones después de estudiar en detalle las secuencias utilizadas en estos trabajos.

5.2. Secuencias aperiódicas.

La ecuación de sG forzada y amortiguada.

Movidos por los resultados de Salerno (1991) y Salerno (1992), en 1995 se realizaron otros estudios relacionados con la relevancia de la secuencia en la dinámica de *kinks* (Domínguez-Adame et al. 1995). En este caso, y puesto que carecían de la información necesaria utilizaron, en lugar de secuencias de ADN, secuencias binarias aperiódicas. Estas secuencias se caracterizan por: (i) ser secuencias cuyos elementos pueden tomar dos posibles valores, A o B ; (ii) son deterministas, construidas a partir de una regla que construye la secuencia S_{n+1} , de longitud $|F_{n+1}|$, a partir de la secuencia S_n , de longitud $|F_n|$, con $|F_n| < |F_{n+1}|$; (iii) son no periódicas para cualquier $n \in \mathbb{N}$. De esta manera, estas secuencias se parecen a las secuencias obtenidas a partir de secuencias de ADN descritas en la sección anterior, en el sentido (i) y (iii) que he mencionado. Además, la propiedad (ii) asocia a estas secuencias aperiódicas una especie de “información”, ya que cuanto más larga es la secuencia más “complicada” se vuelve. Esta información puede asociarse a la información que contiene el ADN en forma de genes y que lo convierte en un objeto funcionalmente muy complejo.

Las dos secuencias aperiódicas utilizadas fueron la secuencia de Fibonacci y la secuencia de Thue-Morse. La secuencia de Fibonacci de orden $n + 1$, F_{n+1} , se define con la regla generadora que actúa sobre F_n de la forma $A \rightarrow AB, B \rightarrow A$, mientras que la secuencia de Thue-Morse de orden $n + 1$, T_{n+1} , se define con la regla generadora que actúa sobre T_n de la forma $A \rightarrow AB, B \rightarrow BA$. Con estas reglas generadoras, y utilizando como secuencia inicial $F_0 = T_0 = A$, el número de elementos de una secuencia de Fibonacci F_{n+1} es $|F_{n+1}| = |F_n| + |F_{n-1}|$, y para una secuencia de Thue-Morse T_n es $|T_n| = 2^n$. La concentración de A en la secuencia de Fibonacci es $c_F \simeq 0.618\dots$ en el límite de la secuencia infinita⁹, pero la aproximación es bastante buena ya a partir de F_5 , con una concentración

⁹Dicha concentración es la inversa de la razón áurea, $(\sqrt{5} + 1)/2$, que se obtiene a partir de la secuencia de Fibonacci numérica, $\{x_n\} = \{1, 2, 3, 5, 8, 13, 21, \dots\}$, como $\lim_{n \rightarrow \infty} (x_{n+1}/x_n)$.

de $8/13 \simeq 0.615\dots$. En el caso de Thue-Morse, la concentración de A es siempre $c_T = 0.5$.

La ecuación que se utilizó en Domínguez-Adame et al. (1995) fue la ecuación de sG discretizada, forzada y amortiguada,

$$\ddot{\phi}_n - \frac{1}{d^2}(\phi_{n+1} - 2\phi_n + \phi_{n-1}) + q_n \sin \phi_n + \alpha \dot{\phi}_n + F = 0, \quad (5.3)$$

donde $\alpha > 0$ nos da la disipación del sistema y F la fuerza externa, global y constante, que se aplica. El efecto de una fuerza F positiva (negativa) sobre un *kink* es empujarlo hacia la derecha (izquierda), y en sentido contrario en el caso de un *antikink*. La disipación siempre ejerce una fuerza que se opone al sentido de movimiento del *kink* (o *antikink*)¹⁰.

La introducción de la fuerza externa en el modelo puede entenderse en el contexto del ADN de varias maneras. Cuando se utilizó la ecuación de sG para modelar el ADN por primera vez (Englander et al. 1980), ya se tuvo en cuenta que la imagen real de un *kink* que hace que las bases, cerradas, den un giro completo hasta volver a cerrarse no era realista, y se hizo énfasis en la descripción por parte del modelo del movimiento de rotación inicial que deben realizar las bases para desligarse de su compañera en la formación de un estado abierto de la doble cadena de ADN. Posteriormente, el modelo fue utilizado como “activador energético” de la ARN polimerasa en el proceso de transcripción (Salerno 1991), aunque de nuevo la topología del *kink* hacía imposible la comparación literal del modelo mecánico que describe la ecuación con la molécula de ADN. Con este espíritu, la introducción de la fuerza externa en el modelo puede representar la fuerza que la propia ARN polimerasa ejerce sobre la molécula para avanzar por ella. Pero la sencillez del modelo también permite jugar con él en el terreno de la replicación del ADN¹¹, donde el pozo de potencial en el que caen las bases a un lado del *kink* puede entenderse como la unión con la nueva hebra de ADN que se está formando. La fuerza externa en esta nueva imagen del modelo sería la fuerza que el “tenedor” de replicación de la ADN polimerasa ejercería para moverse por la doble cadena de ADN y producir las copias. Por último, otra imagen que puede encajar en este sencillo modelo proviene de los experimentos de desnaturalización mecánica de ADN, en los que el ADN es manipulado de manera se separan las dos hebras que lo forman al tirar del extremo unido de dos cadenas complemen-

¹⁰Por tanto, las dinámicas de un *kink* y de un *antikink* que se mueven de acuerdo a la ecuación de movimiento (5.3) ya no son equivalentes, salvo que se cambie F por $-F$ en el caso del *antikink*.

¹¹En la replicación del ADN, cada hebra de la doble cadena de ADN se usa como plantilla para obtener la hebra complementaria. De esta forma, el resultado final son dos dobles cadenas de ADN, cada una de las cuales formada por una hebra de la molécula original y una hebra recién sintetizada.

tarias¹². En este caso, la fuerza con que se tira del extremo de las hebras sería la fuerza externa del modelo, y el mínimo de potencial en el que caerían los pares de bases rotos sería la formación de nuevos enlaces de hidrógeno con el medio acuoso que las rodea.

Sin embargo, y como ya mencionamos al final del capítulo anterior, para que el efecto disipativo del agua sea tenido en cuenta de forma realista, el término de rozamiento de la ecuación (5.3) debe ser mucho mayor que el inercial, aunque los parámetros elegidos en este estudio muestran lo contrario. Pese a ello, y al igual que en el caso de Salerno (1991), esta investigación sigue siendo válida si se considera que relaciona la dinámica de *kinks* en secuencias inhomogéneas que se mueven de acuerdo a la ecuación (5.3) con determinadas propiedades de la secuencia estudiada de manera fenomenológica. No es necesario llevar el modelo a niveles más realistas, aunque los nuevos parámetros introducidos en él estén motivados por procesos del ADN conocidos, así como por experimentos reales.

La velocidad asintótica.

En el caso homogéneo, con $q_n = q_0$, y tomando el límite continuo de esta ecuación, se puede calcular la velocidad asintótica de un *kink* (o un *antikink*) en función de la intensidad de rozamiento α y la fuerza externa F (McLaughlin y Scott 1978). Esta relación debe existir ya que, si se aproxima la dinámica de un (*anti*)*kink* por la de una partícula y se le aplica fuerza externa y rozamiento, entonces las fuerzas que actúan sobre ella son la fuerza externa, constante, y el rozamiento, proporcional a la velocidad de la partícula. El término de fuerza externa hará que la partícula se acelere y aumente su velocidad; el término de rozamiento se hará cada vez más grande conforme aumenta ésta, hasta que llegue un punto en que toda la fuerza externa aplicada se contrarreste por el rozamiento, y la aceleración de la partícula se anule. El cálculo de la velocidad asintótica, v_∞ , se realiza de la siguiente manera. Sea la ecuación de sG continua, homogénea, forzada y amortiguada,

$$\phi_{tt} - \phi_{xx} + \sin \phi + \alpha \phi_t + F = 0, \quad (5.4)$$

y supongamos que un *kink* (2.19) es solución de dicha ecuación, siendo el único efecto de los términos de disipación y fuerza externa sobre el *kink* un cambio en su velocidad. A continuación, si introducimos esta expresión en el Hamiltoniano (2.7) asociado a la ecuación sin perturbar, es decir, a la ecuación de sG continua y homogénea (2.9) (o, lo que es lo mismo, a (5.4) con $\alpha = F = 0$), se obtiene la

¹²En el capítulo 7 se describe más detalladamente el experimento.

energía del (*anti*)*kink* en el sistema sin perturbar,

$$E(\phi) = \frac{8}{\sqrt{1-v^2}}. \quad (5.5)$$

La pérdida o ganancia de energía del sistema con el tiempo se obtiene derivando (5.5) con respecto al tiempo,

$$\frac{dE(\phi)}{dt} = \frac{8v}{(1-v^2)^{3/2}} \frac{dv}{dt}. \quad (5.6)$$

Por otro lado, si ϕ es solución de (5.4) entonces, derivando el Hamiltoniano (2.7) respecto al tiempo se obtiene que

$$\frac{d}{dt}H[\phi] = - \int_{-\infty}^{\infty} dx(\alpha\phi_t^2 + F\phi_t). \quad (5.7)$$

Sustituyendo la expresión (2.19) en (5.7), se obtiene que la energía del (*anti*)*kink* varía de la forma

$$\frac{dE(\phi)}{dt} = 2\pi Fv - 8\alpha v^2(1-v^2)^{-1/2}. \quad (5.8)$$

Igualando las expresiones (5.6) y (5.8) se obtiene una EDO para $v(t)$,

$$\frac{dv}{dt} = \frac{1}{4}\pi F(1-v^2)^{3/2} - \alpha v(1-v^2). \quad (5.9)$$

En el límite asintótico, cuando $dv/dt = 0$, se puede despejar la velocidad asintótica, v_∞ , y se obtiene

$$v_\infty = \pm \left[1 + \left(\frac{4\alpha}{\pi F} \right)^2 \right]^{-1/2}, \quad (5.10)$$

donde el signo de v_∞ depende del signo de la fuerza aplicada, F . El estudio de la velocidad del *antikink* es análogo, cambiando F por $-F$.

Metodología y resultados.

Domínguez-Adame et al. (1995) utilizaron este resultado para comparar la velocidad asintótica de un *kink* que se mueve en una secuencia inhomogénea aperiódica $\{q_n\}$ con la que tendría al moverse en el modelo homogéneo con $q_n = q_{\text{avg}} = c_X q_A + (1 - c_X)q_B$ (c_X es la concentración de A asociada a cada secuencia). Observaron una propiedad interesante: la velocidad asintótica de un

kink forzado y amortiguado en estas secuencias aperiódicas es nula para valores de la fuerza F por debajo de una cierta fuerza umbral, F_u , y por encima de ella se ajusta más o menos a la fórmula (5.10)¹³. A continuación describiré en detalle la metodología seguida y los resultados del artículo.

En todo el artículo se realizó la integración numérica de la ecuación (5.3) para diferentes secuencias, utilizando como parámetros de la simulación $\alpha = 0.1$, un valor de amortiguamiento habitual cuyo único efecto es fijar la escala de fuerzas, y $d = 0.1$, un valor de discretización lo suficientemente pequeño como para despreciar efectos del potencial de PN. Los valores de la fuerza F se tomaron en el rango $[0, 0.30]$ para dibujar la gráfica de (F, v_∞) y analizar los resultados. Respecto a los dos posibles valores de q_n asociados a la secuencia, q_A y q_B , fijaron $q_A = 1$, y de esta manera estudiaron el comportamiento de v_∞ para diferentes valores de q_B ; el mayor valor utilizado fue $q_B = 10$, que en ninguna secuencia da lugar a discretizaciones efectivas tan grandes como para tener en cuenta el potencial de PN ($d_{\text{eff}} \simeq 0.21$ para el caso de Fibonacci y $d_{\text{eff}} = 0.23$ para Thue-Morse). La condición inicial en la integración numérica fue en todos los casos un *kink* (2.19) en reposo con su centro situado en la mitad de la cadena.

En la primera parte del trabajo estudiaron las diferentes gráficas (F, v_∞) para diferentes secuencias usando un mismo q_B , y también para diferentes valores de q_B usando la misma secuencia. Para ello generaron cinco secuencias diferentes de más de 4000 elementos: F_{18} , de 4181 elementos; T_{12} , de 4096 elementos; una secuencia periódica $P_{AB} = ABABAB\dots$ de 4182 elementos; y, por último, dos secuencias aleatorias generadas con las concentraciones de A de la secuencia de Fibonacci y Thue-Morse, c_F y c_T . En todas estas secuencias representaron la gráfica (F, v_∞) para $q_B = 10$. Por otro lado, dibujaron la misma gráfica asociada a la secuencia de Fibonacci F_{18} para valores de $q_B = 2, 5$ y 10 . El resultado que obtuvieron se muestra en la figura 5.2 (izquierda). En ella se observa la existencia de una fuerza umbral, F_u , por debajo de la cual los *kinks* no se mueven, que aparece en todas las secuencias excepto en la secuencia periódica P_{AB} . Para un mismo valor de $q_B = 10$, las secuencias aleatorias tienen una fuerza umbral F_u tres veces mayor que las secuencias aperiódicas con su misma concentración de A ; además, por encima de este valor la velocidad v_∞ depende de la realización de la secuencia aleatoria que se haya escogido. Por otro lado, fijada la secuencia F_{18} , y para diferentes valores de q_B , se observa que el valor de F_u aumenta conforme

¹³Realmente, la fórmula utilizada por Domínguez-Adame et al. (1995) no es exactamente (5.10) sino que incluye el parámetro q_{avg} , ya que en dicho trabajo partieron de la ecuación (5.3) para obtenerla. Ambos resultados son equivalentes si se tiene en cuenta la discretización efectiva del problema, $d_{\text{eff}} = dq_{\text{avg}}^{1/2}$. Por ello, los resultados son diferentes para valores de q_{avg} diferentes.

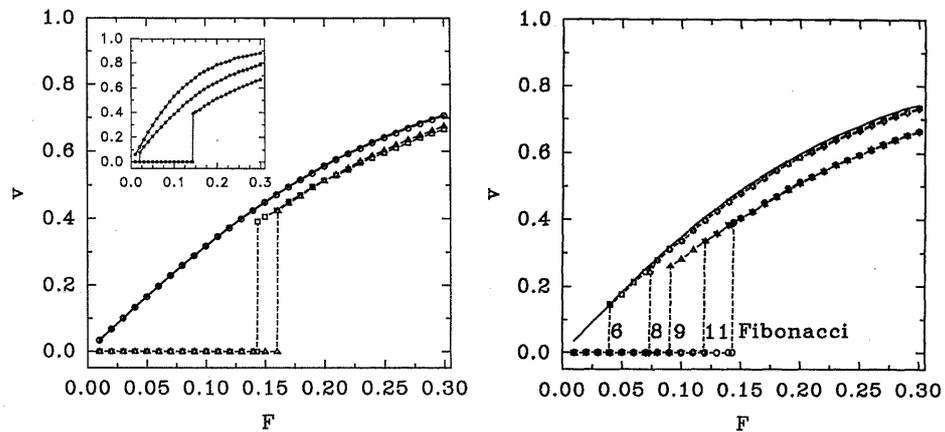


Figura 5.2: Figuras obtenidas de Domínguez-Adame et al. (1995). Velocidad asintótica frente a la fuerza aplicada para diferentes secuencias. Los puntos corresponden a integraciones numéricas, y la línea sólida a la predicción (5.10). Las líneas discontinuas son una guía para los ojos. Izquierda: Secuencias con $q_A = 1$ y $q_B = 10$; velocidad asintótica para q_{avg} (véase la nota al pie 13) correspondiente a $c = 0.5$; (○), secuencia periódica $ABABAB \dots$ con 4182 elementos; (□), secuencia de Fibonacci F_{18} con 4181 elementos; (△), secuencia Thue-Morse T_{12} con $2^{12} = 4096$ elementos. El valor F_u para la secuencia aleatoria está alrededor de 0.5, fuera del rango. Detalle: lo mismo, pero para secuencias de Fibonacci con $q_B = 2$, $q_B = 5$ y $q_B = 10$. Derecha: velocidad asintótica para q_{avg} (véase la nota al pie 13) correspondiente a $c = c_F$, con $q_A = 1$ y $q_B = 10$; secuencias periódicas construidas cuya celda unidad es una secuencia de Fibonacci, F_k . (□), $k = 6$ (13 elementos); (◇), $k = 8$ (34 elementos); (△), $k = 9$ (55 elementos); (▽), $k = 11$ (144 elementos); (○), secuencia de Fibonacci completa.

aumenta q_B .

En la segunda parte de su trabajo, estudiaron la dependencia de la fuerza umbral F_u con el orden de la secuencia de Fibonacci utilizado. Para ello construyeron secuencias periódicas cuya celda unidad era una secuencia de Fibonacci de orden k , F_k , de la forma $F_k F_k F_k \dots$. La longitud de las secuencias rondaba las 4180 posiciones en todos los casos. Los resultados que aparecen en la figura 5.2 (derecha) son para $k = 6$ ($|F_6| = 13$), $k = 8$ ($|F_8| = 34$), $k = 9$ ($|F_9| = 55$), $k = 11$ ($|F_{11}| = 144$), y una secuencia sin repetir con $k = 18$ ($|F_{18}| = 4181$) que ocupa toda la cadena. En dicha figura se observa una dependencia monótona de F_u con el orden k de la secuencia de Fibonacci utilizada al construir la secuencia periódica, siendo la fuerza umbral máxima para el caso de la secuencia completa aperiódica F_{18} . Al repetir el proceso con la secuencia de Thue-Morse, los autores explican que los resultados son muy parecidos, pero que son necesarias secuencias más cortas para alcanzar la F_u de la secuencia completa, y la dependencia de F_u con k no es monótona. También señalan que el valor de F_u depende de la posición inicial del *kink* en la cadena, cambiando hasta en un factor 2, pero que siempre se mantiene debajo del valor de la secuencia F_{18} .

Existe una tercera parte en Domínguez-Adame et al. (1995) en la que hacen un estudio de los resultados anteriores utilizando el potencial efectivo para sistemas discretos inhomogéneos. Dicho potencial fue introducido en Salerno y Kivshar (1994), y se estudiará más detenidamente en la próxima sección, donde veremos cómo se obtiene y las aplicaciones que aparecen en Salerno y Kivshar (1994) para el estudio de secuencias de ADN (Salerno 1991; Salerno 1992) y en Domínguez-Adame et al. (1995) para el estudio de secuencias aperiódicas.

Una de las conclusiones del artículo es que aparecen efectos de largo alcance cuando los *kinks* se propagan por la cadena, y que se muestran en un descenso de la velocidad asintótica v_∞ con respecto a la de la cadena homogénea. Estos efectos pueden relacionarse en el contexto del ADN, que hacen que una onda no lineal se propague más fácilmente por la cadena cuando ésta no es aleatoria, como muestran las figuras. De esta manera, es como si la “información” que contiene la secuencia sobre la que se mueve el *kink* permitiera a éste moverse. La relación con el contenido de genes o no en secuencias de ADN parece inmediata. Sin embargo, en el artículo también se dejan preguntas sin contestar, como una explicación analítica de v_∞ en secuencias aperiódicas, o por qué las aproximaciones periódicas convergen a la secuencia completa aperiódica. En el próximo capítulo aclararé estas cuestiones, así como otras relacionadas con Domínguez-Adame et al. (1995), describiendo en detalle algunos resultados originales de esta tesis que iluminan gran parte de los puntos oscuros de los trabajos sobre el modelo de sG

con secuencias inhomogéneas (Salerno 1991; Salerno 1992; Domínguez-Adame et al. 1995) que he descrito en este capítulo.

5.3. El potencial efectivo, V_{eff} .

Renormalización de la secuencia.

Antes de introducir el potencial efectivo en el modelo de sG discreto con secuencias inhomogéneas es necesario solucionar algunos problemas relacionados con la introducción de una secuencia inhomogénea en el modelo. Dichos problemas provienen del hecho de que, en la ecuación de sG (5.1), la discretización efectiva del problema, $d_{\text{eff}} = dq_{\text{avg}}^{1/2}$, deja de ser el parámetro de discretización de la red, d . Sin embargo, la discretización intrínseca del problema, y del *kink* en particular, sí es la discretización efectiva. Esto significa que la expresión del *kink* en reposo con centro en X es

$$\tilde{\phi}(z_n) = 4 \arctan(e^{z_n}), \quad (5.11)$$

con $z_n = d_{\text{eff}}n - X$. Si se usase el parámetro de la discretización, $z_n = dn - X$, daría lugar a un *kink* con una anchura diferente de la anchura en reposo, y por tanto con más energía que en el estado fundamental que tendría que ser relajada del *kink* en forma de fonones.

Para evitar este problema de nomenclatura es necesario renormalizar la ecuación, dividiéndola por q_{avg} , y reescalar el tiempo $t \rightarrow q_{\text{avg}}^{1/2}t$. De esta manera se llega a una expresión de la misma forma que (5.1),

$$\ddot{\phi}_n - \frac{1}{d_{\text{eff}}^2}(\phi_{n+1} - 2\phi_n + \phi_{n-1}) + \tilde{q}_n \sin \phi_n = 0, \quad (5.12)$$

pero donde $\tilde{q}_n = q_n/q_{\text{avg}}$ está normalizada de manera que su valor medio sea la unidad, y por tanto el parámetro de discretización de la ecuación ahora sí coincide con la discretización efectiva del problema, d_{eff} . De esta forma se resuelven algunas ambigüedades y posibles confusiones que puedan surgir en un futuro, con la única diferencia en la escala temporal de ambas ecuaciones, la normalizada y la no normalizada.

Ciertamente, establecer d_{eff} como escala intrínseca del *kink* en reposo no es más que una aproximación. Cuando veamos la expresión del potencial efectivo, entenderemos cómo el *kink* pesa de diferente manera los q_n que están cerca de su centro de los que están alejados de él. Por tanto, parece lógico pensar que la anchura intrínseca del *kink* dependa de un entorno de q_n alrededor de su centro.

Sin embargo, utilizar el valor medio de la secuencia completa es una aproximación bastante buena, como veremos a continuación, cuando comparemos el potencial efectivo del *kink* con la dinámica de una partícula.

En los casos en los que hay rozamiento y fuerza externa, como en los estudiados en (Domínguez-Adame et al. 1995), al dividir la ecuación (5.3) por q_{avg} los parámetros de fuerza y rozamiento quedan modificados, de forma que la ecuación renormalizada pasa a ser, una vez reescalado el tiempo,

$$\ddot{\phi}_n - \frac{1}{d_{\text{eff}}^2}(\phi_{n+1} - 2\phi_n + \phi_{n-1}) + \tilde{q}_n \sin \phi_n + \alpha' \dot{\phi}_n + F' = 0, \quad (5.13)$$

con $\alpha' = \alpha/q_{\text{avg}}^{1/2}$ y $F' = F/q_{\text{avg}}$. Esta variación en los parámetros hay que tenerla en cuenta a la hora de comparar con resultados numéricos en los que no se haya efectuado la renormalización de la secuencia. Por ejemplo, esta transformación en los parámetros hace que la ecuación de la velocidad asintótica utilizada por Domínguez-Adame et al. (1995) no sea exactamente (5.10), que proviene de la aproximación al continuo, y que es válida para la ecuación renormalizada. Esto es así porque, por un lado, $(\alpha'/F')^2 = q_{\text{avg}}(\alpha/F)^2$, y por otro el espacio y el tiempo están ambos reescalados con $q_{\text{avg}}^{1/2}$, de manera que la velocidad del centro del *kink* coincide en ambas visualizaciones.

La obtención del potencial efectivo puede hacerse tanto para la ecuación normalizada como la no normalizada, siempre que se tenga en cuenta que la anchura intrínseca del *kink* en ambos casos es la misma.

Coordenadas colectivas en el modelo discreto.

El espíritu del potencial efectivo en sistemas discretos es el mismo que el del modelo continuo: describir con una única variable, el centro del *kink*, el comportamiento dinámico de éste debido a la inhomogeneidad del potencial. El potencial efectivo en sistemas discretos inhomogéneos fue introducido por Salerno y Kivshar (1994) para explicar los resultados anteriores (Salerno 1991; Salerno 1992). En su artículo, los autores utilizan como *Ansatz* el *kink* discretizado dado por (5.11), donde $z_n = d_{\text{eff}}n - X(t)$, y $X(t)$ es la posición del centro del *kink*, que depende del tiempo. Dicho *Ansatz* se sustituye en la expresión del Hamiltoniano del sistema inhomogéneo sin renormalizar, que es

$$H(\{\phi_n\}) = d \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \dot{\phi}_n^2 + \frac{1}{2d^2} (\phi_{n+1} - \phi_n)^2 + q_n (1 - \cos \phi_n) \right\}, \quad (5.14)$$

y se obtiene la expresión de la energía de una partícula¹⁴,

$$E = \frac{1}{2}M(X)\dot{X}^2 + U(X), \quad (5.15)$$

donde

$$M(X) = 4d \sum_n \frac{1}{\cosh^2(z_n)}, \quad U(X) = 2d \sum_n \frac{q_n + q_{\text{avg}}}{\cosh^2(z_n)}. \quad (5.16)$$

Nótese que, cuando $q_n = q_{\text{avg}} = 1$, se obtiene que $U(X) = M(X)$. Además, $M(X)$ es la aproximación discreta de la masa del *kink* en el modelo continuo, $M_0 l_0 = 8$, y es un valor constante para valores de X alejados de los extremos de la red¹⁵.

A continuación, Salerno y Kivshar (1994), y puesto que E es una cantidad conservada, consideran la energía de un *kink* inicialmente en reposo en la posición $X(0) = X_0$, y entonces reescriben la ecuación (5.15) de la forma

$$\frac{1}{2}\dot{X}^2 + W(X, X_0) = 0, \quad (5.17)$$

donde

$$W(X, X_0) = \frac{\sum_n (q_n + q_{\text{avg}}) [\text{sech}^2(z_n) - \text{sech}^2(z_n^{(0)})]}{2 \sum_n \text{sech}^2(z_n)}, \quad (5.18)$$

y $z_n^{(0)} \equiv d_{\text{eff}} n - X_0$. Así, según los autores, la ecuación (5.17) se puede considerar como la ecuación de una partícula de masa unidad que se mueve en el potencial efectivo $W(X, X_0)$, y dicho potencial puede ser utilizado para predecir la dirección en la que se moverá el *kink* debido a la inhomogeneidad de la secuencia.

Comparación del potencial efectivo con simulaciones.

Para comprobar la fiabilidad del potencial efectivo (5.18), Salerno y Kivshar (1994) eligieron la secuencia utilizada anteriormente por Salerno (1991) para estudiar la “activación” del *kink* en torno al promotor A_1 del fago T7. Sin embargo, en lugar de aplicar directamente la fórmula (5.18), y para tener en cuenta que el *kink* es un objeto extenso, promediaron el potencial $W(X, X_0)$ un distancia igual

¹⁴La única aproximación necesaria en este caso es la aproximación $\phi_{n+1} - \phi_n \simeq d_{\text{eff}} \partial_x \phi(x, t)$, donde $\phi(x, t)$ es la aproximación al continuo $\phi_n(t) \rightarrow \phi(x, t)$, válida para $d_{\text{eff}} \ll 1$.

¹⁵Esto es así porque, por un lado, el *kink* está contenido en la cadena y se pueden despreciar efectos de borde, y por otro, la discretización efectiva hemos supuesto que es lo suficientemente pequeña como para despreciar el potencial de PN.

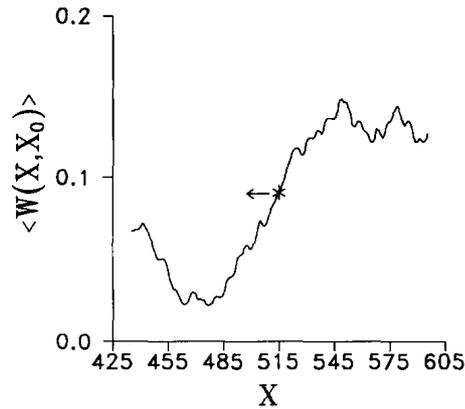


Figura 5.3: Figura obtenida de Salerno y Kivshar (1994). Potencial efectivo (5.18) dentro de la región del promotor T7 A_1 , en BP 515, de acuerdo con la secuencia S dada por (5.2). Claramente el potencial tiene una pendiente positiva que hace que el *kink* se mueva a la derecha.

a la anchura del *kink*, $l = d_{\text{eff}}^{-1}$ (según sus cálculos, $l = 30$ PB). El resultado que obtienen para X_0 en la posición 515 de la cadena se observa en la figura 5.3. En ella aparece una pendiente positiva del potencial en X_0 que hace que el *kink* claramente se vea impulsado hacia la izquierda¹⁶, tal y como revelaban las simulaciones realizadas por Salerno (1991) (ver tabla 5.1).

Posteriormente, Domínguez-Adame et al. (1995) también utilizaron el potencial efectivo (5.18) para explicar parte de sus resultados, más exactamente el porqué de la existencia de una fuerza umbral F_u por debajo de la cual los *kinks* no se mueven. La única diferencia con el potencial calculado por Salerno y Kivshar (1994) es que, puesto que en este caso había una fuerza externa, utilizaron el siguiente Hamiltoniano sin renormalizar,

$$H(\{\phi_n\}) = d \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \dot{\phi}_n^2 + \frac{1}{2d^2} (\phi_{n+1} - \phi_n)^2 + q_n (1 - \cos \phi_n) + F\phi \right\}, \quad (5.19)$$

que da lugar a las ecuaciones de movimiento con fuerza externa, pero sin disipación. El potencial efectivo que utilizaron, por tanto, tiene un término más corres-

¹⁶Posteriormente justificaban este resultado con la transcripción de la ARN polimerasa hacia la izquierda en ese promotor. Sin embargo, como se describe en el apéndice B, la dirección de propagación asociada al promotor A_1 es hacia la derecha.

pondiente a la fuerza externa, de manera que su potencial efectivo resultó ser de la forma $W(X, X_0) + W_F(X, X_0)$, con

$$W_F(X, X_0) = F \frac{\sum_n [\arctan(e^{z_n}) - \arctan(e^{z_n^{(0)}})]}{\sum_n \operatorname{sech}^2(z_n)}. \quad (5.20)$$

Además, tal y como hicieron Salerno y Kivshar (1994), y para incluir los efectos de la anchura del *kink*, realizaron un promedio del potencial en el intervalo de red dado por la anchura del *kink*. Una vez calculado el potencial efectivo observaron si, dependiendo de la fuerza aplicada, el *kink*, inicialmente en reposo, tendría que superar alguna barrera de potencial o no. En el primer caso se quedaría parado en algún lugar antes de la barrera, y en el segundo, podría seguir moviéndose, con lo que la existencia de la fuerza umbral quedaba explicada.

Significado del potencial efectivo.

Sin embargo, ni Salerno y Kivshar (1994) ni Domínguez-Adame et al. (1995) confiaron lo suficiente en las posibilidades que el potencial efectivo podía ofrecerles. A continuación voy a explicar una serie de propiedades del potencial efectivo que he ido observando a lo largo de estos años de estudio de esta tesis doctoral.

En primer lugar, no es necesario realizar ningún promedio del potencial efectivo en un número de sitios dado por la anchura del *kink*. La definición que se hace del potencial efectivo $W(X, X_0)$ por (Salerno y Kivshar 1994) es para reescribir la expresión de la energía de una partícula (5.15) de la forma (5.17). Y esta definición se hace a su vez para escribir un potencial que dependa explícitamente del punto inicial X_0 y así poder realizar el promedio. Pero veamos qué quiere decir el potencial efectivo $W(X, X_0)$. Para ello, antes observemos el potencial $U(X)$ del que proviene, tal y como se encuentra definido en (5.16). Como ya mencioné anteriormente, en el caso $q_n = q_{\text{avg}} = 1$ se obtiene $U(X) = M(X)$. Ahora, además, fijémonos en el sumatorio de $U(X)$. El término de la derecha del sumatorio se puede sumar y es proporcional a la masa, $M(X)$; dicho término no nos interesa. Si nos fijamos en el otro sumando, el que va con q_n , se encuentra multiplicado por la contribución a la masa del *kink*. Esto significa que el potencial $U(X)$ está compuesto por la suma de un término constante (para valores de X lo suficientemente alejados de los extremos de la cadena) y otro que es la media ponderada de la secuencia completa $\{q_n\}$. El peso en la posición n de dicha media ponderada, $w(n)$, es proporcional a la contribución a la masa total del *kink* en ese sitio,

$$w(n) \propto \operatorname{sech}^2 z_n. \quad (5.21)$$

Eso significa que el potencial $U(X)$ ya tiene en cuenta que el *kink* es un objeto extenso, y lo realiza de una manera lógica, puesto que no asigna el mismo peso a todos los sitios de la red, sino que le da más peso a los que están cerca del centro del *kink*, X , y menos a los que están alejados. Para dar una idea de la extensión de un *kink* (o su anchura), el valor del peso $w(n)$ para $z_n = 3$ es del orden de 10^{-2} , aunque el número de sitios que hay que recorrer hasta llegar a $z_n = 3$ depende del valor de la discretización efectiva, d_{eff} ¹⁷. Por ello, en el cálculo del potencial $U(X)$ para cada posición realmente no se tiene en cuenta *toda* la secuencia $\{q_n\}$ de la red, sino sólo los de los sitios más cercanos. Cuando, posteriormente, se promedia el potencial $W(X, X_0)$ para tener en cuenta que el *kink* es un objeto extenso, no se ha pensado que el potencial efectivo ya es un promedio en sí mismo.

Así que, puesto que no considero necesario realizar promedios posteriores, y puesto que resulta más intuitivo, prefiero utilizar la expresión del potencial $U(X)$ como potencial efectivo del *kink* en secuencias inhomogéneas en lugar de la expresión de $W(X, X_0)$, propuesta en Salerno y Kivshar (1994). La modificación que realizaré sobre $U(X)$ es dividirlo por el término de masa, $M(X)$, para asociarlo con el potencial de una partícula de masa unidad. De esta forma, el potencial efectivo que voy a considerar a partir de ahora es

$$V_{\text{eff}}(X) = \frac{\sum_n (q_n + q_{\text{avg}}) \text{sech}^2(z_n)}{2 \sum_n \text{sech}^2(z_n)}, \quad (5.22)$$

donde recordemos que $z_n = d_{\text{eff}}n - X$. Esta expresión puede utilizarse tanto para la ecuación sin renormalizar como para la ecuación renormalizada, teniendo en cuenta que, en este caso, $\tilde{q}_{\text{avg}} = 1$. La diferencia en la dinámica de dos partículas que se mueven en sendos potenciales efectivos será en la escala de tiempos, que se corresponderá con la escala de tiempos del *kink* que se mueve de acuerdo a la ecuación de sG de la proceden, según sea renormalizada o no.

En segundo lugar, tanto Salerno y Kivshar (1994) como Domínguez-Adame et al. (1995) subestimaron la capacidad de predicción del potencial efectivo¹⁸. En ambos casos sólo se utiliza el potencial efectivo para predecir el comportamiento inicial del *kink*, ya sea para estimar en qué sentido se mueve un *kink* colocado en reposo en un sitio de la red (Salerno y Kivshar 1994), o para estimar si un *kink* es capaz de sobrepasar la barrera de potencial más cercana aplicándole una fuerza externa (Domínguez-Adame et al. 1995). Además, no se dieron resultados cuantitativos, sino que la utilización del potencial efectivo fue para dar una explicación

¹⁷Esa distancia será la que hay que tener en cuenta para evitar efectos de borde al colocar el *kink* en la red.

¹⁸Quizá en parte porque se realizó esa media arbitraria del potencial efectivo.

cualitativa del proceso estudiado. Continuando el trabajo de Domínguez-Adame et al. (1995), unos años después, Lennholm y Hörnquist (1999), estudiaron de una manera más cuantitativa la relación entre la dinámica de *kinks* con rozamiento y fuerza externa y el potencial efectivo de diferentes secuencias aperiódicas. La principal conclusión del artículo es que, dada una secuencia aperiódica con un potencial efectivo no monótono, esto es, con máximos y mínimos, existe una fuerza umbral que depende del punto inicial en el que se coloca el *kink* al comienzo de la simulación. Si se aplica una fuerza menor que esa fuerza umbral, el *kink* se quedará parado siempre en un mínimo local del potencial efectivo.

Dinámica de *kink* y dinámica de partícula.

Resulta tentador realizar una simulación de la dinámica de una partícula que se mueve en un cierto potencial efectivo, para compararla posteriormente en una gráfica con la dinámica de un *kink* moviéndose en la correspondiente secuencia inhomogénea que ha generado dicho potencial. En este apartado compararemos ambos resultados en el caso de un sistema en ausencia de rozamiento y fuerza externa [como en el estudio realizado por Salerno (Salerno 1991; Salerno 1992)] y con ellas [como en el estudio realizado por Domínguez-Adame et al. (1995)].

En primer lugar consideremos el caso en el que la fuerza externa y el rozamiento son cero. En ese caso, la dinámica de una partícula depende exclusivamente del potencial en el que se encuentra, por tanto su ecuación de movimiento es

$$\ddot{x} = -\frac{dV_{\text{eff}}}{dx}, \quad (5.23)$$

y el sistema es conservativo. Eso significa que, si aproximamos el *kink* por una partícula, su energía total es la suma de la energía cinética, dada por su velocidad, y de la potencial, dada por la posición. El caso real del *kink* no es tan sencillo, puesto que su Hamiltoniano tiene tres términos: el cinético, asociado a su velocidad y anchura; el dispersivo, asociado a su anchura; y el potencial, asociado a la posición en la cadena y su anchura. Además, el *kink* puede emitir fonones, que le permiten perder energía, conservándose la energía del sistema. Por tanto, la comparación de la dinámica del *kink* y de la partícula asociada nos pueden dar una idea de lo que le está sucediendo al *kink*.

En la figura 5.4 se muestran la posición y la velocidad de un *kink* y de su aproximación de partícula, para diferentes posiciones iniciales del *kink* de las que habla Salerno (1991). En ellas se observa, en primer lugar, que el potencial efectivo V_{eff} describe muy bien la dinámica del *kink*. Los resultados obtenidos además se ajustan a los descritos por Salerno (1991). En los resultados se observa un li-

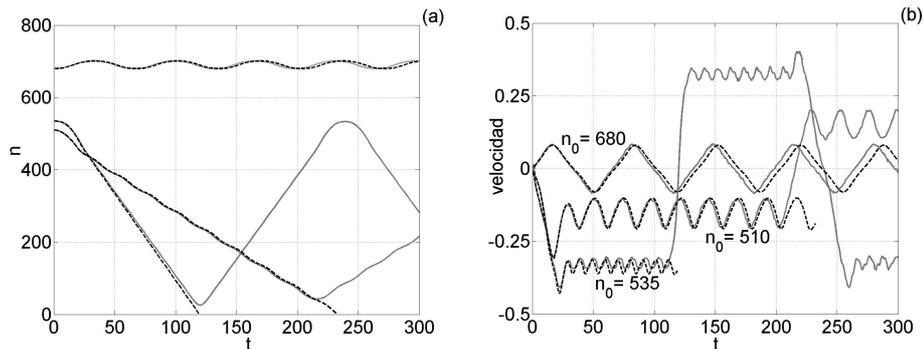


Figura 5.4: En (a) está dibujada la posición del centro y en (b) la velocidad de: un *kink* (línea continua, gris) y de una partícula (línea discontinua, negro). Las dinámicas de los *kinks* fueron obtenidas a partir de la integración numérica de la ecuación (5.12), renormalizada, utilizando la secuencia S dada por (5.2), creada por Salerno (1991), y con condiciones de contorno fijas; y la dinámica de las partículas, por la integración numérica de la ecuación (5.23), utilizando la misma secuencia S para construir el potencial efectivo, y terminando la integración cuando la partícula se encuentra muy cerca de los extremos de la red (por eso las velocidades se cortan en seco para un cierto t , que depende de cada simulación). Las posiciones iniciales utilizadas son $n_0 = 510$, $n_0 = 535$ y $n_0 = 680$. Véase el texto para un análisis de los resultados.

gero desfase de las dos velocidades, la del *kink* y la de la partícula. Dicho desfase es menor en el caso $n_0 = 510$, que da lugar a una velocidad pequeña pero no nula, oscilante, que se corresponde muy bien con la dinámica de la partícula hasta que el *kink* llega al final de la secuencia, donde el *kink* rebota y la integración de la partícula se para. En el caso $n_0 = 535$ se obtiene una mayor velocidad final (en valor absoluto), pero se observa que la velocidad del *kink* siempre tiende a ser ligeramente menor que la de la partícula, dando lugar por tanto también a un desfase entre ambos resultados. El peor caso se da en $n_0 = 680$, donde el *kink* y la partícula oscilan en torno a una cierta posición de la red y el desfase se acentúa al prolongarse ambas simulaciones en el tiempo. Este desfase puede deberse a una oscilación en la anchura del *kink* o a la emisión de fonones de éste, que hacen que la solución se aleje del *Ansatz* elegido. Una variación en la anchura del *kink* cambiaría la discretización efectiva del sistema, haciendo que el peso (5.21) utilizado en el cálculo del potencial efectivo se ensanchara o estrechara, y cambiando ligeramente con este efecto la forma del potencial. La emisión de fonones por parte del *kink* conllevaría una pérdida de energía de éste, que podría traducirse en una disminución en su velocidad, como si de un amortiguamiento se tratara. Puesto que, en general, no se observa que la velocidad del *kink* sea menor que la de la aproximación dada por la partícula, es probable que el mayor efecto que provoque la diferencia de velocidades en ambos casos esté producido por un cambio en la anchura del *kink*. Dicho efecto es mayor para velocidades altas y en oscilaciones en torno a $v = 0$.

En el caso de rozamiento y fuerza externa, sin embargo, es necesario realizar un pequeño estudio que mostrará que el problema es más complicado que en el caso puramente inercial. Fijémonos en la ecuación (5.9), que nos da la velocidad de un *kink* que se mueve en el modelo de sG continuo y homogéneo con rozamiento y fuerza externa¹⁹. Dicha ecuación es fácilmente integrable, y da lugar a la expresión de la velocidad

$$v(t) = \frac{r(t)}{\sqrt{1 + r^2(t)}}, \quad (5.24)$$

donde

$$r(t) = \frac{\pi F}{4\alpha} (1 - e^{-\alpha t}) + r(0)e^{-\alpha t}. \quad (5.25)$$

Es fácil comprobar que, en el límite cuando $t \rightarrow \infty$, la expresión (5.24) se convierte en (5.10).

¹⁹Dicha expresión proviene de suponer como *Ansatz* la solución tipo *kink* dada por (2.19).

Pensemos ahora en el caso de una partícula con velocidad $\nu(t)$, de masa unidad, inicialmente en reposo, que se mueve debido a una fuerza externa constante γ y con rozamiento $\beta\nu$,

$$\dot{\nu} = \gamma - \beta\nu. \quad (5.26)$$

Claramente, en el caso de la partícula, la velocidad asintótica asociada al balance del efecto producido por el rozamiento y la fuerza externa es $\nu_{\infty} = \gamma/\beta$, que no se parece mucho a la velocidad asintótica del *kink*, (5.10). Además, la integración de la ecuación (5.26) da lugar a

$$\nu(t) = \frac{\gamma}{\beta} \left(1 - e^{-\beta t}\right) + \nu(0)e^{-\beta t}. \quad (5.27)$$

Comparando ahora ambas expresiones, la de la velocidad del *kink*, (5.24), y la de la velocidad de la partícula, (5.27), no se parecen. Sin embargo, las expresiones (5.25) y (5.27) son iguales para $\beta = \alpha$ y $\gamma = \pi F/4$. Por tanto, se puede relacionar la velocidad del *kink*, $v(t)$, con la de la partícula, $\nu(t)$, a partir de las relaciones (5.24) y (5.25); es como si la velocidad de un *kink* en el modelo homogéneo, $v(t)$, que se mueve con rozamiento α y fuerza externa F fuera igual a la velocidad de una partícula, $r(t)$, que se mueve con rozamiento α y fuerza externa $\pi F/4$, transformada de acuerdo a la relación (5.24). En la figura 5.5 aparecen dibujadas las dos velocidades para $v(0) = r(0) = 0$. Para valores pequeños de la velocidad, se tiene que $v(t) \simeq r(t)$. Claro está que todo esto depende de la validez de la expresión (5.24) para describir la dinámica del *kink* forzado y amortiguado en un sistema homogéneo; por ello en la figura 5.6 está dibujada la diferencia entre la velocidad aproximada, $v(t)$, dada por (5.24), y el resultado de la simulación dinámica del *kink* para dos valores diferentes de F . En dicha figura se observa que la expresión $v(t)$ se ajusta muy bien a los resultados numéricos, y que el error aumenta conforme aumenta la fuerza aplicada, F , siendo siempre el resultado numérico ligeramente inferior al resultado teórico. Esta diferencia puede provenir de una pérdida de energía del *kink* en forma de fonones en el proceso de aceleración, mayor cuanto mayor es la fuerza, y que no se ha tenido en cuenta en el cálculo de (5.24). Además, puesto que dicha emisión de fonones es menor para valores de la aceleración pequeños (por tanto, F pequeña), la aproximación es mejor para velocidades asintóticas pequeñas, en las que la velocidad del *kink*, dado por $v(t)$, se puede aproximar muy bien por la velocidad de la partícula, $r(t)$.

Este análisis lo he realizado para sistemas homogéneos, con rozamiento y fuerza externa. Pero supongamos que esta relación entre la velocidad del *kink* y la de la partícula se sigue manteniendo en sistemas inhomogéneos, al menos de manera aproximada, y que la ecuación de movimiento de la partícula que aproxima

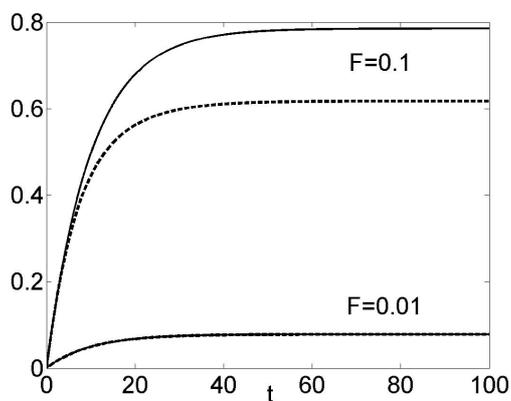


Figura 5.5: Dependencia temporal de $v(t)$, dada por (5.24), y $r(t)$, dada por (5.25), para $v(0) = r(0) = 0$. Las líneas continuas corresponden a $v(t)$ y las discontinuas a $r(t)$. Están dibujadas cuatro gráficas, dos para $F = 0.01$, en la parte de abajo de la figura, prácticamente indistinguibles, y dos para $F = 0.1$, las dos que se ven más arriba, con una clara diferencia entre ellas.

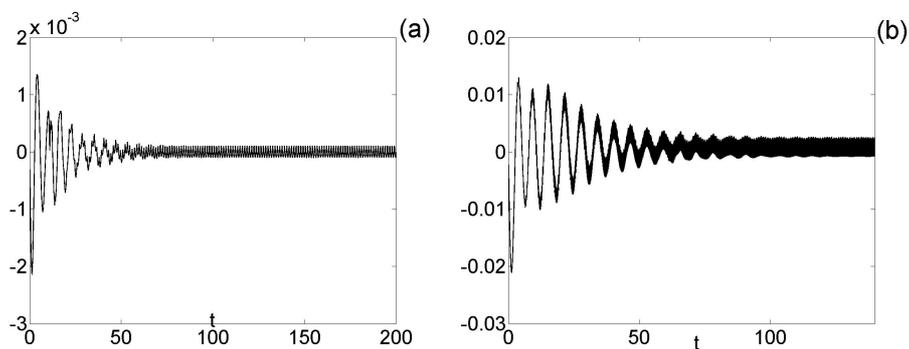


Figura 5.6: Diferencia entre $v(t)$ y la velocidad del centro del *kink* que se obtiene de la simulación numérica de (5.4), con $d = \alpha = 0.1$, y (a) $F = 0.01$ y (b) $F = 0.1$. Se observa que el error absoluto de la aproximación de $v(t)$ disminuye para valores pequeños de F .

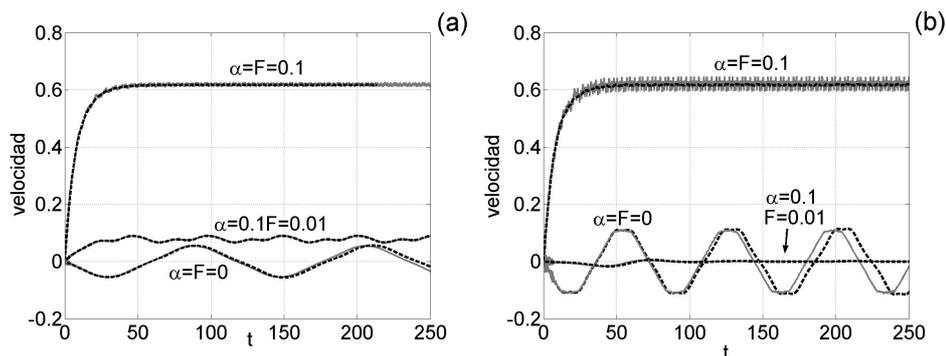


Figura 5.7: Velocidad del centro del *kink* y velocidad de la aproximación de partícula, transformada según (5.24), para el caso amortiguado y con fuerza externa. Resultados para diferentes valores del amortiguamiento y fuerza externa, en el caso de una secuencia con (a) $q_A = 1$, $q_B = 2$, y (b) $q_A = 1$, $q_B = 5$, en una secuencia periódica cuya celda unidad es F_8 (tamaño de celda 34), y el punto inicial es $n_0 = 341$. La dinámica de los *kinks* fue obtenida mediante la integración numérica de (5.13), normalizada, y la dinámica de la partícula mediante la integración numérica de (5.28). Para el análisis de resultados, véase el texto.

el movimiento del *kink* pasa a ser

$$\ddot{x} = -\frac{dV_{\text{eff}}}{dx} + \frac{\pi F}{4} - \alpha \dot{x}, \quad (5.28)$$

donde ya he utilizado la relación entre la fuerza asociada a la partícula y al *kink*. En la figura 5.7 están representados los resultados de la integración numérica de un *kink* y una partícula para un caso particular de secuencia periódica construida a partir de la secuencia de Fibonacci F_8 partiendo de una cierta condición inicial, común para todos (ver el pie de la figura para la explicación de los parámetros utilizados). En la figura de la izquierda se dibujan los resultados para la partícula en una secuencia con $q_A = 1$ y $q_B = 2$, y en la de la derecha con $q_A = 1$ y $q_B = 5$. Salvo en el caso de la partícula inercial, con $F = 0$ y $\alpha = 0$, en el que hay un desfase como el que se describe en la figura 5.4, en el resto de las simulaciones la dinámica de la partícula es muy parecida a la dinámica del *kink*, incluso en el caso en el que el *kink* se queda parado. En el caso en el que se aplica la fuerza mayor, la velocidad del *kink* tiene asociado demasiado ruido, pero aún así se ve que, en media, la aproximación de la velocidad de la partícula se ajusta muy bien a la del *kink*. Quizá esta mejor aproximación se deba a la existencia de rozamiento en el sistema, que hace que el *kink* disipe mucha menos energía en forma de fonones, y que la mayor pérdida de energía venga dada por el término disipativo de la ecuación (5.3), tal y como está supuesto en esta aproximación.

Con estos resultados uno puede hacerse una idea muy precisa de las ventajas y limitaciones que tiene el potencial efectivo V_{eff} (5.22) aplicado a secuencias inhomogéneas. Hasta ahora he mostrado que, salvo un ligero desfase en algunas ocasiones, el potencial efectivo V_{eff} aproxima muy bien la dinámica de un *kink*. Esto permitirá predecir el comportamiento de *kinks* en secuencias inhomogéneas estudiando únicamente la forma del V_{eff} asociado a ellas.

6

Aplicabilidad del modelo de Salerno del ADN

Este capítulo es íntegramente contenido original de la tesis. En él se revisan algunos de los resultados de los trabajos descritos en el capítulo anterior, primero para el caso de secuencias aperiódicas y luego para secuencias que contienen promotores de ADN. Posteriormente, a raíz de otros estudios en los que se utiliza el mismo modelo para estudiar los promotores del fago T7, se analiza el genoma del fago para encontrar alguna característica que permita identificar estas secuencias de entre el resto del genoma. Por último, se estudia la posibilidad de utilizar un potencial efectivo que aproxime la dinámica de *breathers* para identificar algunos elementos funcionales del genoma, como ha sido sugerido recientemente.

6.1. Secuencias aperiódicas.

Efecto de la posición inicial.

Las conclusiones obtenidas por Domínguez-Adame et al. (1995) sobre secuencias aperiódicas nos llevaron a realizar un estudio similar sobre secuencias de ADN, para así ver si la fuerza umbral dependía de si la secuencia contenía información o no. Dicho trabajo culminó con la publicación de dos artículos (Cuenda y Sánchez 2004a; Cuenda y Sánchez 2004b), que describiré en esta sección.

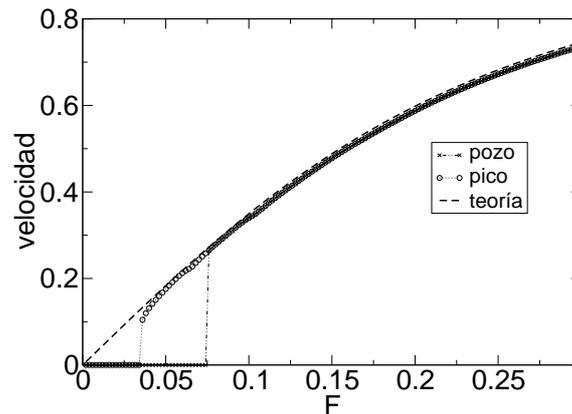


Figura 6.1: Resultados numéricos de la velocidad estacionaria frente a la fuerza aplicada para una secuencia periódica cuya celda unidad consiste en la secuencia de Fibonacci F_9 , para $q_A = 1$ y $q_B = 10$, pero empezando en un pico (\circ) y en un pozo (\times). La fuerza umbral empezando desde un pozo es mayor que desde un pico.

Para estudiar la posible dependencia de la dinámica con la secuencia de ADN, elegimos secuencias que pertenecían a algún gen y secuencias que no pertenecían a ninguno, y comparamos resultados¹. Puesto que Salerno (1991) muestra la diferencia en la dinámica de un *kink* dependiendo de la posición inicial de éste (aunque sin rozamiento ni fuerza externa), decidimos estudiar la influencia de la posición inicial del *kink* en la fuerza umbral. Parece lógico que, si la posición inicial del *kink* está en un pico de V_{eff} , costará menos esfuerzo (o requerirá menos fuerza externa, F) sobrepasar la siguiente barrera que si comienza desde un pozo del potencial, tal y como se muestra en la figura 6.1.

Teniendo esto en cuenta, uno puede esperar que, para secuencias periódicas formadas por celdas unidad de diferentes tamaños, pero con idéntico potencial efectivo en las proximidades del punto de inicio, y con la misma estructura de máximos y mínimos, la fuerza umbral sea la misma en ambos casos. Esto se explica mejor en la figura 6.2. En 6.2a se ve que, para tamaños pequeños de celda unidad ($|F_5| = 8$, $|F_6| = 13$), la estructura principal de picos y pozos de la secuencia de Fibonacci no se ha formado todavía. Dicha estructura se observa mejor

¹Realmente, el hecho de que una secuencia no codifique genes no significa que la secuencia no contenga información. Volveremos sobre este tema al final de esta sección, cuando analice los resultados obtenidos.

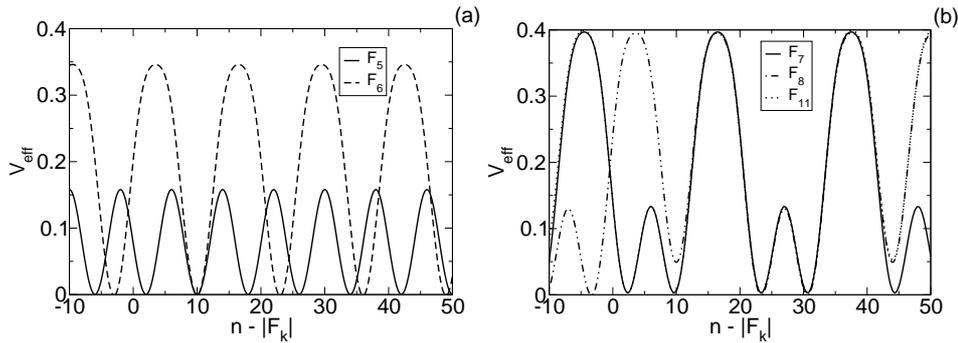


Figura 6.2: Potencial efectivo V_{eff} para secuencias periódicas cuya celda unidad es una secuencia de Fibonacci, F_k , para (a) $k = 5$ y $k = 6$; y (b) $k = 7$, $k = 8$ y $k = 11$. En (a) se observa que la estructura completa de los picos y los pozos no se ha formado todavía, como lo ha hecho en (b) a partir de $k = 7$.

en 6.2b, donde se dibuja V_{eff} para F_7 (de tamaño 21), F_8 y F_{11} , y se observa que en todos los casos el potencial efectivo es igual en un entorno de la posición inicial (el comienzo de una celda unidad), y la estructura de picos y pozos también es muy parecida. Así pues, escogiendo una posición inicial que esté cerca del comienzo de una celda unidad, la fuerza necesaria para superar la primera barrera grande será la misma que en otra secuencia con diferente celda unidad, siempre que las posiciones iniciales (relativas al comienzo de una celda unidad) sean las mismas. Para celdas unidad de tamaño menor el potencial efectivo es muy diferente y, por tanto, también lo es la fuerza umbral, tal y como se ve en el ejemplo de la figura 6.3.

Efecto de la información de la secuencia.

Volvamos ahora a los posibles efectos de la inhomogeneidad de la secuencia genética sobre la propagación del *kink*. Para ello integramos numéricamente la ecuación (5.3) utilizando secuencias del genoma humano². Los parámetros que utilizamos fueron, por un lado, y puesto que estamos aplicando este modelo sobre el ADN, $q_A = 2$ y $q_B = 3$, donde se ha utilizado q_A para sitios con bases A o T y q_B para sitios de la secuencia con bases C o G [de la misma forma que hizo Salerno (1991)]. Por otro lado, hemos usado un parámetro de discretización

²Las secuencias fueron obtenidas de la base de datos del NCBI. Una explicación de como bajarse secuencias se detalla en el apéndice B.

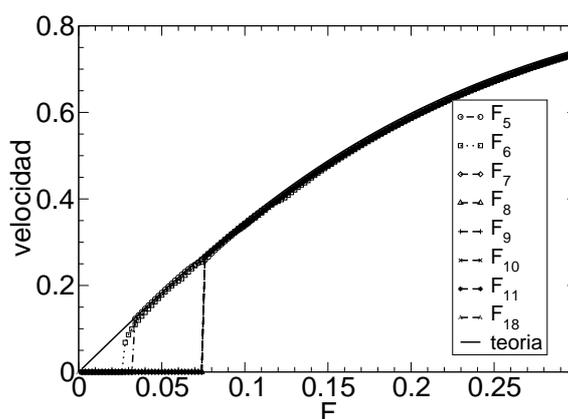


Figura 6.3: Velocidad asintótica para diferentes secuencias periódicas formadas por secuencias de Fibonacci, de la forma $F_k F_k F_k \dots$. Para F_5 y F_6 los valores de la fuerza umbral son distintas que para el resto de F_k debido a la diferente estructura de su potencial efectivo V_{eff} (como se explica en la figura 6.2).

$d = 0.1$, que permite ignorar cualquier efecto de la discretización del modelo, y un parámetro de rozamiento $\alpha = 0.1$. Las simulaciones numéricas fueron llevadas a cabo con un método estándar Runge-Kutta de cuarto orden, con condiciones de contorno periódicas o libres.

Resultados típicos de las simulaciones aparecen dibujados en la figura 6.4, en la que se observa la velocidad del centro del *kink* para dos valores diferentes de la fuerza aplicada en una misma secuencia de ADN del *contig* NT_028395.2 perteneciente al cromosoma humano 22, partiendo de la misma posición inicial. Para que sirva de referencia, hemos representado en la misma gráfica de la velocidad el potencial efectivo V_{eff} , debidamente escalado y desplazado para que sea fácilmente visible frente a la velocidad. Vemos que la dinámica es muy similar al caso de secuencias aperiódicas, en el sentido de que, para esa secuencia y esa posición inicial, hay una fuerza umbral por debajo de la cual el *kink* se queda parado en algún punto de la secuencia, y por encima de la cual el *kink* es capaz de alcanzar el final de la secuencia. En este caso, la fuerza umbral F_u se encuentra entre $F = 0.06$, para la cual el *kink* se queda parado, y $F = 0.07$, para la cual el *kink* llega al final con velocidad no nula.

La relevancia del contenido de información de una secuencia con la existencia y características de la fuerza umbral se pueden entender en la figura 6.5. En ella se encuentran recogidas diferentes observaciones de la dinámica referentes a

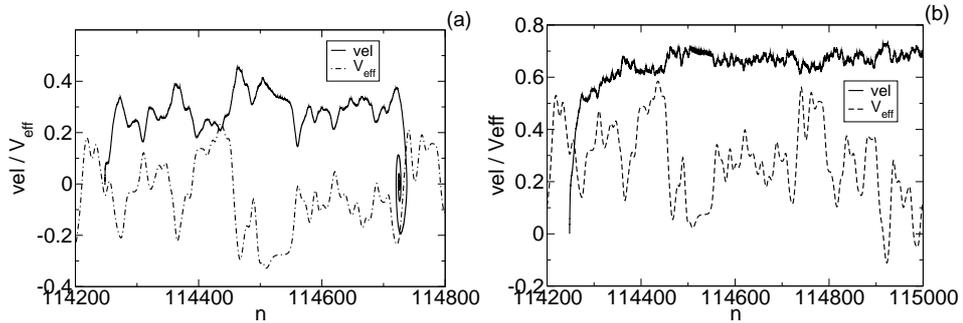


Figura 6.4: Simulaciones de la ecuación (5.3) con secuencias de ADN correspondientes al *contig* NT_028395.2 del cromosoma humano 22, entre las posiciones 114100 y 115100. Están dibujadas la velocidad (línea sólida) y el potencial efectivo (línea discontinua), para: (a) $F = 0.06$, donde se ve que el *kink* se queda parado en un pozo del potencial; y (b) $F = 0.07$, donde el *kink* alcanza el final de la secuencia. Nótese las escalas diferentes en ambas figuras y en ambos ejes.

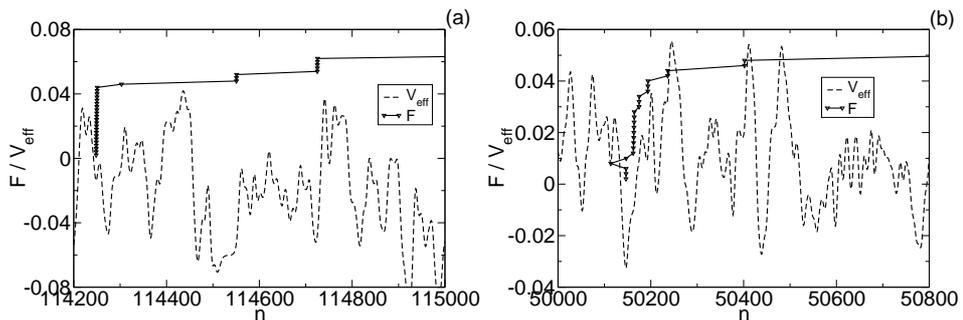


Figura 6.5: Simulaciones de la ecuación (5.3) con secuencias de ADN correspondientes al *contig* NT_028395.2 del cromosoma humano 22, entre las posiciones (a) 114100 y 115100 (región que codifica un gen) y (b) 50000 y 51000 (región que no codifica ningún gen). En línea discontinua está dibujado el potencial efectivo, reescalado y trasladado verticalmente. Los triángulos unidos por línea continua representan, en la coordenada horizontal, la posición a la que el *kink* llega cuando se le aplica la fuerza dada por la coordenada vertical.

dos ejemplos de secuencia: una región que codifica un gen y otra que no. De nuevo está dibujado el potencial efectivo V_{eff} , debidamente reescalado y desplazado verticalmente para que sirva de referencia; y la abscisa de los triángulos muestra el punto de parada del *kink* cuando se le aplica una fuerza dada por el valor de la ordenada. Como antes, se observa la existencia de fuerzas umbral, de acuerdo con la descripción en términos del potencial efectivo. Sin embargo, no se aprecia ninguna diferencia cualitativa relevante entre la dinámica de los *kinks* (o entre los potenciales efectivos) en ambas secuencias. Este es el caso de todas las secuencias que hemos analizado. Por tanto, la conjetura de Domínguez-Adame et al. (1995), basada en sus resultados con secuencias de Fibonacci, de que la información de la secuencia pudiera conllevar diferentes propiedades de la dinámica de *kinks* no está, al menos al nivel de este modelo sencillo, de acuerdo con los resultados de las simulaciones. Por supuesto, esto es sólo un ejemplo; aunque hemos obtenido diferentes resultados de simulaciones para secuencias que codifican genes o no, no podemos excluir que el comportamiento de la fuerza umbral sea estadísticamente diferente en los dos tipos de regiones. Sin embargo, de nuestras simulaciones consideramos esta posibilidad bastante poco probable.

Secuencias periódicas de ADN.

Ahora que hemos dejado de tener en cuenta el contenido de información de la secuencia para la determinación de la fuerza umbral, es interesante proceder a la comparación entre el trabajo previo sobre secuencias de Fibonacci (Domínguez-Adame et al. 1995) y los resultados que se presentan en la presente memoria, para así entender qué diferencias entre ellos llevan a conclusiones diferentes. Para ello, hemos simulado secuencias periódicas cuya celda unidad es un trozo de secuencia genética, de la misma forma que se construyeron secuencias periódicas utilizando secuencias aperiódicas como la de Fibonacci. Los tamaños de celda unidad imitan los tamaños de las sucesiones de Fibonacci de orden k , $|F_k|$, y de nuevo esta vez las posiciones iniciales se escogieron de manera que todas tuvieran la misma posición relativa con respecto al inicio de una celda unidad. Los resultados se recogen en la figura 6.6a, donde se observa que la mayor diferencia con el resultado de Fibonacci de la figura 6.3 es que, en la secuencia de ADN, la fuerza umbral varía mucho más dependiendo del tamaño de la celda unidad que en el caso de Fibonacci, y no hay un máximo de la fuerza umbral cuando se incrementa el tamaño de la celda unidad. Esto es así debido a la mayor libertad en la estructura del potencial efectivo que existe en el caso de la secuencia de ADN comparada con la secuencia de Fibonacci, como se puede deducir comparando la figura 6.6b con los resultados mostrados en 6.2a y 6.2b. Esta diferencia en el potencial efec-

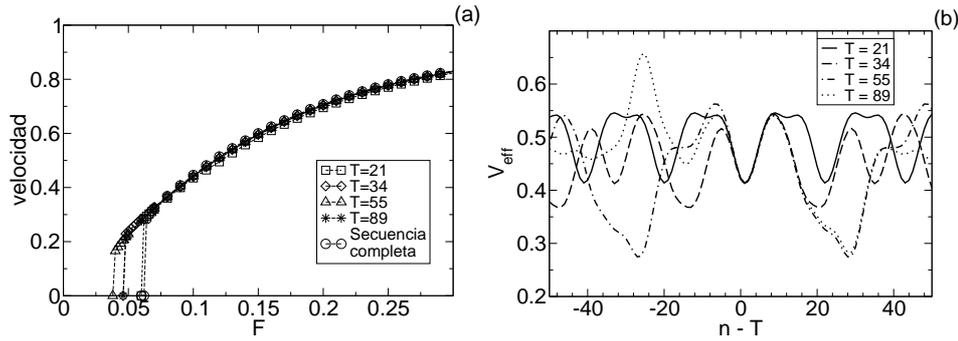


Figura 6.6: (a) Lo mismo que la figura 6.3, pero con secuencias periódicas cuya celda unidad es una secuencia de ADN de tamaño T . La secuencia de ADN es la del mismo *contig* que la figura 6.4 partiendo del sitio 114241; la secuencia completa se refiere a la secuencia de 114100 a 115100. (b) Potencial efectivo V_{eff} para dichas secuencias, desplazado en el eje horizontal para que coincidan los comienzos de las celdas unidad en el origen, sin efectos de borde.

tivo se debe a su vez a la restricción que existe en la secuencia de Fibonacci: la regla de creación de dicha secuencia provoca que en ella no se puedan encontrar nunca dos sitios consecutivos de tipo B , ni más de dos de tipo A ; además, existen siempre patrones que se repiten hasta un cierto orden en diferentes puntos de la secuencia y que hacen que el potencial efectivo tenga pocos “tipos” diferentes de máximos y mínimos y, por tanto, su fuerza umbral sea siempre la misma a partir de un cierto tamaño³. En el caso del ADN existe mucha más libertad en la colocación de ATs y CGs, lo que hace que su potencial efectivo varíe más dependiendo del tamaño de la secuencia que se use como celda unidad. Son estos picos altos y pozos profundos, provocados por una alta concentración localizada de CGs o ATs, respectivamente, los que más marcan la dinámica de los *kinks* al propagarse por la cadena haciendo que se paren o que lleguen a velocidades críticas, mientras que el resto del paisaje del potencial sólo influye en detalles menos importantes de la velocidad.

³Este tamaño está dado por el tamaño de celda unidad para el que aparecen en su V_{eff} todos los “tipos” diferentes de máximos y mínimos que se verán en órdenes superiores.

Diferencia entre ADN que no codifica gen y ADN que no codifica nada.

En Cuenda y Sánchez (2004b) mostramos las primeras conclusiones de nuestro trabajo sobre los efectos de la secuencia genética en la propagación de *kinks* en cadenas de ADN. Dicho estudio fue motivado por resultados previos sobre cadenas dadas por secuencias aperiódicas (Domínguez-Adame et al. 1995) que llevaron a la hipótesis de que el contenido de información de las inhomogeneidades podría ser relevante en el comportamiento dinámico del sistema. Aunque este problema no ha sido resuelto, al menos desde un punto de vista estadístico, como he señalado anteriormente, creemos que para el modelo de Salerno del ADN las simulaciones que realizamos no confirman tal hipótesis. Más exactamente, no encontramos diferencias cualitativas en la dinámica de *kinks* que se propagan en secuencias que codifican genes o no, lo que en principio hace que este sencillo modelo no se pueda usar como herramienta genómica para descubrir regiones en las que se encuentran genes del resto.

Sin embargo, la relación entre la dinámica de *kinks* en este modelo y el contenido de información de la secuencia puede no haber sido discutida correctamente. Esto es así porque, a la hora de decidir si una secuencia contiene o no información, sólo nos hemos fijado en si codifica un gen o no. Sin embargo, el ADN contiene secuencias que no codifican genes pero que funcionalmente son muy activas, principalmente porque codifican moléculas de ARN que no se utilizan en la transcripción de proteínas pero que tienen una alta funcionalidad en la regulación de genes⁴. Es en los últimos años cuando los biólogos están empezando a darse cuenta de que las regiones que no codifican proteínas tienen implicaciones cruciales en la funcionalidad del ADN, y por tanto no pueden ser consideradas como regiones que no valen para nada. Así pues, nuestra conclusión debe ser formulada sólo en relación a las regiones que codifican genes o no, sin hablar, en general, de información. Las regiones que codifican ARN que no va destinado a la síntesis de proteínas forman parte de investigaciones punteras en biología, y no están por ahora en bases de datos fácilmente accesibles (como lo están los genes). Por otro lado, el estudio de posiciones relevantes en la secuencia que no son genes, como promotores en virus, es estudiado de forma exhaustiva en el contexto del modelo de sG en la sección 6.3 de esta memoria.

⁴Un artículo de divulgación muy accesible que explica algunos de estos fenómenos es Gibbs (2003).

6.2. Introducción de ruido.

Influencia de la temperatura.

Hasta ahora, en esta memoria, aparecen resultados puramente deterministas, donde el único desorden del modelo ha aparecido en la inhomogeneidad de la intensidad del término potencial de la ecuación de sG, tanto en el caso continuo como en el discreto. Sin embargo, este enfoque no resulta muy realista a la hora de utilizar el modelo de sG inhomogéneo y discreto para describir algunas características del ADN, ya que la vida tiene lugar a temperatura no nula, y lo mismo ocurre con los experimentos que se realizan con ADN, normalmente a temperatura ambiente. Es lícito pensar, por tanto, cuál puede ser el efecto que la temperatura introduce en el modelo; en ese sentido, en esta sección estudiamos si los resultados anteriores siguen siendo válidos en presencia de ruido térmico, así como la formación de excitaciones localizadas a partir de condiciones iniciales ruidosas (Cuenda y Sánchez 2004a).

Robustez de resultados.

En primer lugar estudiaremos la validez de la descripción de la dinámica de *kinks* en términos del potencial efectivo en presencia de ruido térmico utilizando dinámica de Langevin. Para ello, integramos numéricamente la ecuación

$$\phi_{tt} - \phi_{xx} + q_{\text{avg}} \sin \phi + \alpha \phi_t + F + \sigma \xi(x, t) = 0, \quad (6.1)$$

donde $\xi(x, t)$ es ruido blanco gaussiano normalizado (de media cero y varianza unidad). Entonces, por el teorema de fluctuación disipación, se tiene que

$$\sigma^2 = \frac{2\alpha K_B T}{E_H}, \quad (6.2)$$

donde K_B es la constante de Boltzmann, T es la temperatura del sistema⁵, y E_H es la energía de un enlace de hidrógeno⁶, por la que tuvimos que dividir la ecuación original de sG (4.8) para adimensionalizarla.

Los resultados así obtenidos se ilustran con un ejemplo en la figura 6.7. En ella se representa la trayectoria del centro de un *kink* para diferentes valores de

⁵Ciertamente, ésta es una de las partes de la memoria donde más física estadística aparece. Para una explicación muy buena de los métodos estocásticos utilizados, véase Gardiner (1985).

⁶En los cálculos donde intervenga esta cantidad tomaremos $E_H \simeq K_B T_a = 0.025$ eV, la energía térmica a temperatura ambiente. Para una explicación de por qué se toma este valor, véase el apéndice B.

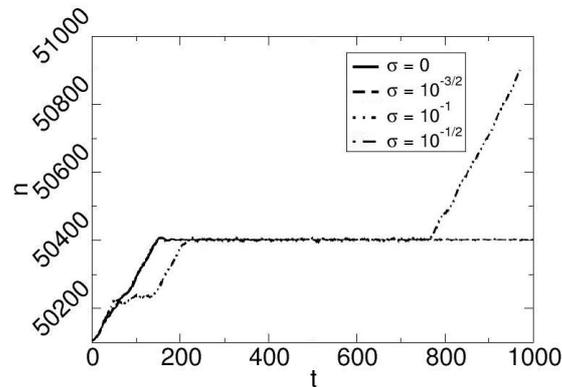


Figura 6.7: Propagación de *kinks* para diferentes intensidades de ruido, integrando numéricamente la ecuación (6.1). La secuencia es la misma que la de la figura 6.5b. La fuerza aplicada es $F = 0.048$, cerca de la fuerza umbral para esa secuencia y esa posición inicial, que es $F_u = 0.05$.

la intensidad del ruido, σ . La fuerza aplicada, F , está por debajo pero cerca de la fuerza umbral F_u , por lo que para $\sigma = 0$ el *kink* se queda atrapado alrededor de la posición 50400. Como se puede observar, es necesario aplicar una intensidad de ruido mucho mayor que la de la fuerza umbral para conseguir sacar el *kink* del pozo en el que se encuentra parado, y sólo se consigue después de llevar bastante tiempo en él. La temperatura asociada a dicha intensidad es $T \simeq 50T_a$ (T_a es un valor habitual de la temperatura ambiente, $295K$ o $22^\circ C$), una temperatura fuera de cualquier rango de medición experimental de procesos biológicos *in vivo* o *in vitro*. Ésta y otras simulaciones parecidas nos llevaron a la conclusión de que el modelo determinista que estamos usando se comporta aproximadamente igual que en presencia de ruido térmico.

Formación de espontánea de excitaciones no lineales.

El modelo discreto de sG aplicado al ADN introducido por Englander et al. (1980) presupone la existencia de excitaciones no lineales (*kinks*) que se propagan por la doble hélice. En este apartado tratamos la formación de dichas excitaciones a partir de configuraciones iniciales ruidosas. Para este estudio utilizamos la secuencia del virus adenoasociado 1 (AAV1), que consta de 4718 pares de bases, y

que contiene el promotor P_5 , estudiado anteriormente en otros modelos más realistas (Kalosakas et al. 2004; Choi et al. 2004). En esos trabajos se realizaron simulaciones de dinámica Langevin en el modelo de Peyrard-Bishop⁷ para secuencias de ADN cortas que contienen el promotor, y encontraron que la secuencia se abre preferentemente en el sitio de comienzo de la transcripción. De sus resultados, concluyeron que el ADN dirige dinámicamente su propia transcripción. Mirando la misma secuencia que en los estudios mencionados, nosotros hemos abordado el problema del origen de las excitaciones no lineales, encontrando resultados que encajan con los encontrados por ellos.

En nuestras simulaciones hemos considerado de nuevo la ecuación determinista (5.3), pero integrada a partir de condiciones iniciales dadas por datos aleatorios de una distribución gaussiana, $\phi_n(0) = \xi_n$, con media ϕ_0 y varianza σ . La fuerza externa F es cero para ver dónde se forman de manera espontánea las excitaciones no lineales en ausencia de forzamiento externo, y el rozamiento también lo hemos anulado porque, de otro modo, las excitaciones no topológicas desaparecerían con el tiempo debido a la pérdida de energía por disipación. Tomamos condiciones de contorno libres, lo que permite que se vaya energía en forma radiación por los extremos del sistema y no interactúe con las excitaciones localizadas que se forman. Esto, a su vez, hace que el sistema no sea conservativo, y no reproduzca la colectividad microcanónica, aunque esto no represente un problema para nosotros puesto que sólo estamos interesados en los lugares más característicos en los que se forman las excitaciones localizadas. Por último, la secuencia utilizada va del PB 200 al PB 400⁸.

La figura 6.8 representa un ejemplo de lo que ocurre al dejar evolucionar condiciones iniciales aleatorias. Para $\phi_0 = 1.0$ se observa que se empiezan a formar excitaciones localizadas, pero la energía de estas excitaciones se escapa por los bordes antes de llegar a localizarse más y “sentir” el efecto de la inhomogeneidad. Para $\phi_0 = 1.5$, aunque se escapa energía, se mantiene la suficiente dentro del sistema para que, con el tiempo, se forme un *breather* que oscila entre dos puntos. Para $\phi_0 = 2.5$ y $\phi_0 = 3.0$, el sistema es capaz de formar *kinks*, que oscilan en torno a dos sitios diferentes.

En cuanto a las posiciones en las que se colocan los centros de las excitaciones, claramente no aparecen al azar: los lugares en los que se focalizan las excitaciones de la figura 6.8 son los pozos del potencial efectivo de la secuencia estudiada⁹. El

⁷Hablaré en detalle de este modelo en el capítulo 7.

⁸El promotor P_5 va de la posición 256 a la 324, y el TSS está en la posición 296.

⁹Parece normal pensar que los *kinks* se sitúen en torno a mínimos de V_{eff} . En la próxima sección veremos que se puede calcular un potencial efectivo similar para los *breathers*, que resulta muy parecido al de los *kinks*, y es por esto por lo que los *breathers* también se colocan en torno a

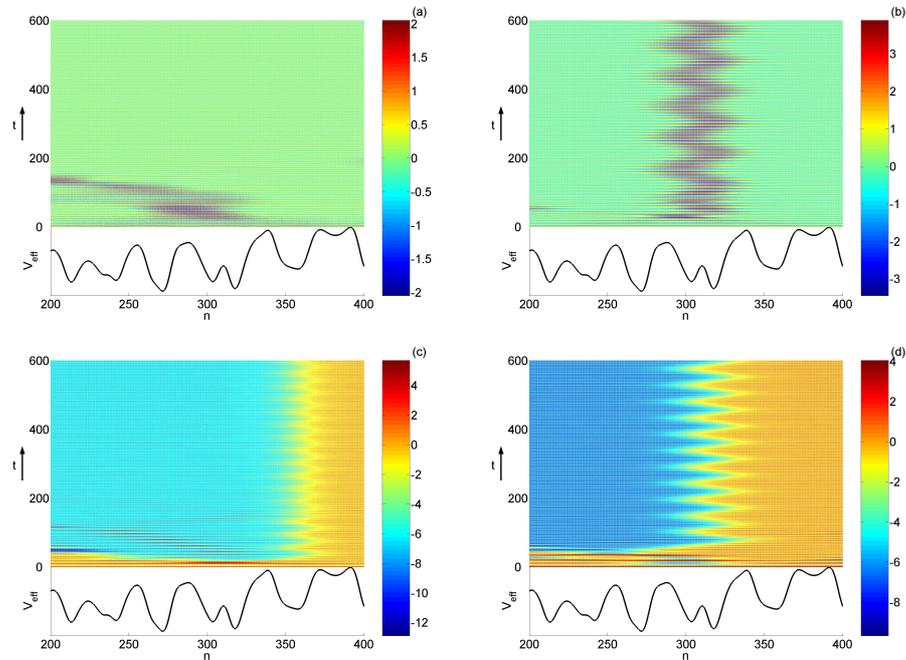


Figura 6.8: En la parte de arriba de cada dibujo está representada, en escala de colores, la evolución temporal de la ecuación de sG a partir de condiciones iniciales aleatorias, con el tiempo de abajo a arriba; las condiciones iniciales son $\sigma = 0.1$ y (a) $\phi_0 = 1.0$, (b) $\phi_0 = 1.5$, (c) $\phi_0 = 2.5$ y (d) $\phi_0 = 3.0$. En la parte de abajo de cada dibujo se representa el potencial efectivo de la secuencia, la del virus AAV1 de la posición 200 a la 400 (ver el texto para más detalles). Se observa una muy buena correspondencia entre la posición de los *kinks* y *breathers* que aparecen con los pozos del potencial.

estudio de la posible relación entre la localización de excitaciones y la posición del TSS, como en otros estudios realizados en el modelo de Peyrard-Bishop (Choi et al. 2004; Kalosakas et al. 2004), muestra que la apertura encontrada en el dibujo (b) de la figura 6.8 se encuentra muy cerca de él, ya que se mueve dentro del pozo que está entre las posiciones 300 y 320, aproximadamente, y el TSS está en PB 296. Sin embargo, existen otros pozos más alejados del TSS, en los que también se forman excitaciones de tipo *kink*, como en (d). Por otro lado, en este estudio no hemos realizado una estadística de la localización de las aperturas en la secuencia de ADN y de su posible relación con los sitios funcionales de la molécula. Por tanto, no es posible dar una conclusión al respecto. En la próxima sección se estudia la relación entre los pozos del potencial efectivo y los sitios funcionales de una secuencia de ADN, que, como se ve en la figura 6.8, pueden tener que ver con los sitios de formación de excitaciones.

6.3. Secuencias de ADN.

La secuencia real del fago T7.

Como se vio en la sección 6.1, no parece existir relación entre la dinámica de *kinks* y el hecho de que una secuencia de ADN codifique genes o no (Cuenda y Sánchez 2004b). Sin embargo, también se han descrito otros trabajos que indican una clara correspondencia entre posiciones relevantes de la secuencia de ADN (promotores) y dinámica de *kinks* (Salerno 1991; Salerno 1992). En esta sección se describe el estudio, realizado durante esta tesis doctoral, para intentar encontrar una explicación a esta aparente contradicción (Cuenda et al. 2006).

Para ello volvamos a la figura 5.3. En ella se muestra el potencial efectivo [con el promedio extra, tal y como propusieron Salerno y Kivshar (1994)] de parte de la secuencia S' que utilizó Salerno (1991) para concluir la relación entre los promotores y las regiones dinámicamente “activas” de la secuencia. Pero, en lugar de limitarnos al rango desde la posición 425 a la 605, veamos cuál es el aspecto de V_{eff} en la secuencia completa. El resultado aparece en la figura 6.9a en línea continua. En la misma figura, en línea discontinua, se muestra el potencial efectivo V_{eff} correspondiente a la secuencia real de ADN alrededor del promotor A_1 del genoma del fago T7. Se observa, por un lado, que las posiciones y alturas de los picos y pozos del potencial efectivo de la secuencia S' explican muy bien los resultados de la dinámica obtenida por Salerno (1991) y mencionados anteriormente en términos de una partícula puntual que se mueve en ese potencial efectivo; por otro

mínimos de este V_{eff} .

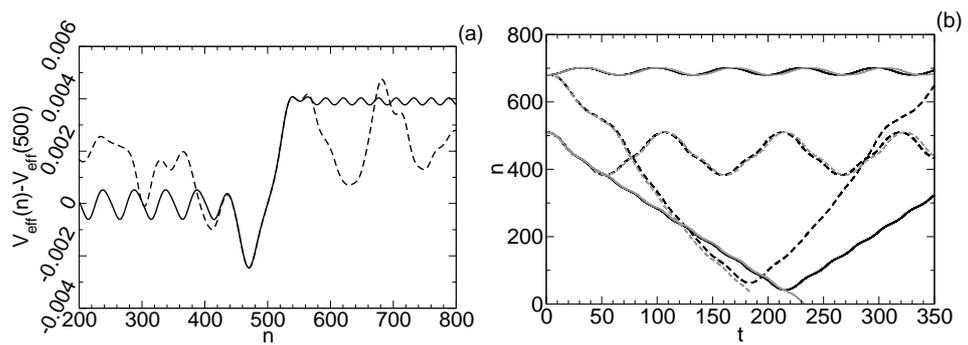


Figura 6.9: (a) Potencial efectivo (5.22) calculado para: la secuencia completa S' [definida en (5.2)], en línea continua; y para la secuencia real de ADN en los alrededores del promotor A_1 , en línea discontinua. A ambas se les ha restado el valor en el centro de la secuencia para que ambos potenciales se hagan cero en $n = 500$ y no influyan efectos de constantes en la comparación de ambas curvas. (b) Dinámicas correspondientes a *kinks* moviéndose por la secuencia S' (líneas continuas) y la secuencia real del fago T7 (línea discontinua), partiendo del reposo desde PB 510 y PB 680. Ambas dinámicas se comparan con las dadas por el potencial efectivo (líneas grises próximas).

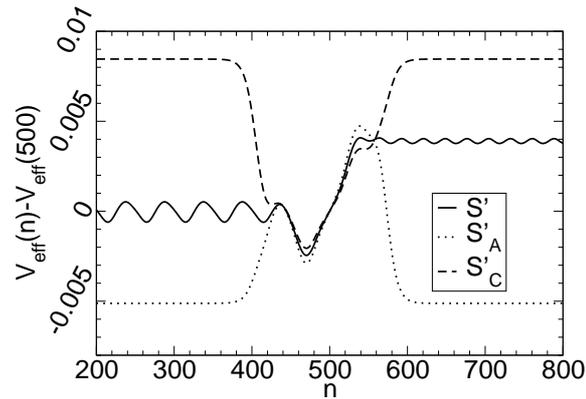


Figura 6.10: Importancia de las secuencias adyacentes: potencial efectivo (5.22) calculado para las secuencias S' , S'_A y S'_C (ver el texto para las definiciones).

lado, en la figura se aprecia la diferencia entre ambos potenciales fuera de la zona central de la secuencia, lo que hace que la dinámica de *kinks* también sea diferente cuando éstos se encuentran en las regiones adyacentes a la secuencia central S . En la figura 6.9b se muestra la dinámica de dos *kinks* en cada secuencia, S' y la secuencia original del fago T7. Claramente, las dinámicas en una y en otra secuencia son diferentes: por ejemplo, en el caso del *kink* que parte de PB 510, en la secuencia real no debería considerarse un sitio “activo”, mientras que el *kink* que parte de PB 680 sí debería serlo (y en la secuencia S' no lo es). En la figura también se muestra el resultado de la dinámica dado por la aproximación del potencial V_{eff} , para demostrar una vez más la validez de la predicción en términos del potencial efectivo.

Efecto de las secuencias adyacentes.

La diferencia entre el potencial de la secuencia S' y la del fago T7 es debida a las dos secuencias periódicas, $8S(1, 50)$ y $15S(141, 168)$, que se colocaron adyacentes a la secuencia S (perteneciente a la secuencia real de ADN del fago T7) en la definición de S' dada por la expresión (5.2). La relación del contenido de AT/CG de estas secuencias adyacentes con respecto al de la secuencia completa S' hace que su potencial efectivo oscile en torno a un cierto valor. Para ver la importancia de este efecto supongamos que, en lugar de estas secuencias que dieron lugar a la secuencia S' , se hubiesen elegido dos secuencias de la forma $AAA\dots$ o dos secuencias de la forma $CCC\dots$, de manera que las nuevas secuencias así creadas

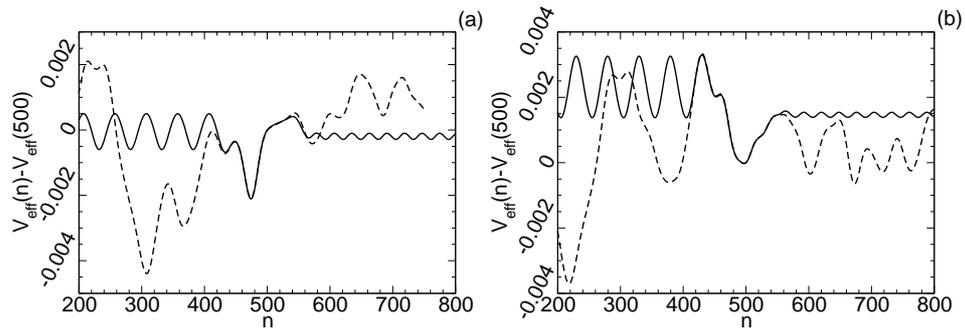


Figura 6.11: Potenciales efectivos V_{eff} de las secuencias sintéticas de los promotores (a) A_0 y (b) A_3 , y de las secuencias reales del fago T7. En (a), la secuencia real del fago T7 aparece truncada porque se encuentra muy cerca del extremo final de la cadena “plantilla” de ADN, que en este caso es la cadena contraria a la del resto de promotores del fago.

fueran S'_A y S'_C , respectivamente. Los potenciales efectivos V_{eff} correspondientes a las tres secuencias se muestran en la figura 6.10. Claramente, las dinámicas correspondientes a los potenciales V_{eff} de las secuencias S'_A y S'_C serían muy diferentes a las descritas por Salerno (1991) y explicadas en el capítulo anterior, aunque las tres tienen la misma secuencia central de 168 nucleótidos.

La misma comparación que en la figura 6.9a entre el potencial efectivo de secuencias sintéticas y secuencias reales puede hacerse para los otros dos promotores, A_0 (también llamado D) y A_3 , estudiados por Salerno (1992). Los resultados aparecen en la figura 6.11, donde de nuevo se observa que el potencial efectivo de las secuencias sintéticas describe muy bien la dinámica de *kinks* descrita en la tabla 5.1 [obtenida del trabajo de Salerno (1992)]. En ellas también aparece una clara diferencia entre el potencial efectivo en la región de las secuencias adyacentes y la secuencia de ADN del genoma del fago T7, lo que implica dinámicas diferentes por parte de los *kinks*. Por ejemplo, un *kink* que partiera de PB 245 en la secuencia real correspondiente al promotor A_0 alcanzaría el final de la secuencia, en lugar de quedarse oscilando, como ocurriría si partiera de la misma posición de la secuencia sintética S' asociada al promotor.

Por tanto, aunque el *kink* parta inicialmente de un sitio que se encuentre dentro de la secuencia original de 168 nucleótidos, las secuencias adyacentes influyen en su dinámica en el momento en que el *kink* llegue a estas secuencias y se tenga que mover por ellas. Este hecho contradice el argumento dado por Salerno (1992), que

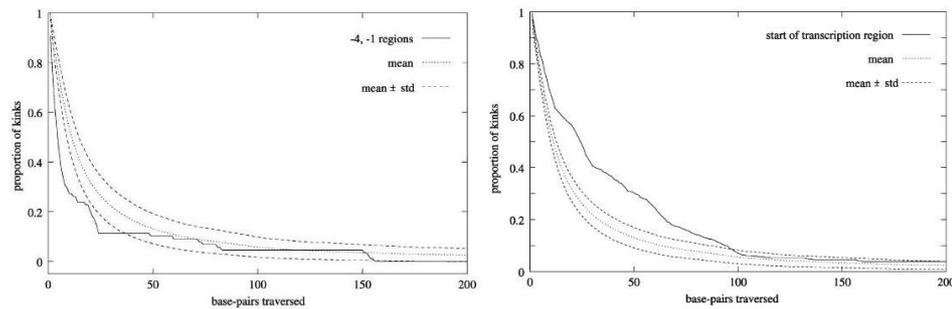


Figura 6.12: Figuras obtenidas de (Lennholm y Hörnquist 2003). Izquierda: en línea continua, la fracción de *kinks* que parten de la región de -4 a -1 y que se mueven al menos el número de PB dado por el eje x ; en línea discontinua, la misma curva para posiciones aleatorias en la secuencia, más menos una desviación típica. Derecha: en línea continua, lo mismo que en (a), pero para *kinks* que parten de los primeras 20 posiciones de la secuencia a partir del TSS; las otras líneas significan lo mismo que en (a).

asumía que estas regiones adyacentes no jugaban ningún papel en la dinámica del *kink*. Por tanto, estamos forzados a concluir que los resultados de la dinámica de *kinks* en promotores descritos en el capítulo anterior (Salerno 1991; Salerno 1992) dependen en gran medida de la construcción de las secuencias S' , y que cuando se utilizan las secuencias de ADN reales del fago T7, las regiones promotoras ya no pueden considerarse “activas” o “especiales”. Como hemos visto, existen otras regiones de la secuencia que pasan a ser más “activas” (en el sentido de poder inducir movimiento al *kink*); y al contrario, existen zonas “activas” en la secuencia sintética que dejan de serlo en el genoma real del fago.

6.4. Trabajos posteriores sobre el modelo de Salerno.

Estadística de dinámica sobre el fago T7.

Existen otros trabajos que, siguiendo la interesante propuesta de Salerno, es decir, la relación entre promotores del fago T7 y la dinámica de solitones moviéndose en secuencias genómicas inhomogéneas, intentaron arrojar un poco más de luz al problema. Así, Lennholm y Hörnquist (2003) midieron la distancia máxima (en cualquier dirección) a la que podía llegar un *kink* inicialmente en reposo que partía de cada uno de los sitios de la secuencia completa del genoma del fa-

go T7. También tomaron los 24 promotores del genoma del fago T7 (excepto el primero y el último para evitar efectos de borde), estudiaron los resultados obtenidos para las posiciones entre -4 y -1 (donde $+1$ es la posición del TSS de cada promotor) y compararon estos resultados con los resultados del genoma completo. El objetivo de este análisis era encontrar si la región de “fusión” de la ARN polimerasa¹⁰, la que va de -4 a -1 en cada promotor, actúa como una región dinámicamente “activa” en el sentido de Salerno, o por el contrario se comporta de la misma forma que el resto de nucleótidos del genoma. En este sentido no encontraron diferencias relevantes¹¹ [ver la figura 6.12a, obtenida del trabajo de Lennholm y Hörnquist (2003)]. En su estudio también investigaron la “actividad” de los primeros n pares de bases después del TSS de cada promotor y encontraron que, para $n = 20$, las regiones estudiadas eran más “activas” que el promedio con una diferencia de más de 5 desviaciones típicas (ver la figura 6.12b, obtenida de dicho trabajo). En el estudio no dieron ninguna interpretación biológica sencilla pero, en las conclusiones, sugirieron que era necesaria una relación más cuantitativa entre la dinámica de *kinks* y el potencial efectivo.

Los promotores en T7.

Como ya se ha discutido a lo largo de este capítulo y del capítulo anterior, el potencial efectivo V_{eff} ha probado ser capaz de predecir la dinámica de *kinks* en secuencias inhomogéneas (al menos en los rangos de discretización efectiva en que nos estamos moviendo) con bastante nivel de precisión, lo suficiente como para resolver el problema del alcance de un *kink* inicialmente en reposo. Por tanto, podemos estudiar el genoma completo del fago T7 utilizando esta herramienta, buscando algún tipo de patrón en el potencial efectivo o algún conjunto de propiedades que puedan ser aplicadas a todos los promotores en el genoma del fago¹² y que permitan diferenciarlos del resto del genoma.

Para ello, en la figura 6.13 se muestra el potencial efectivo para los 18 promotores mayores¹³ del fago T7. Claramente no se observa ningún patrón que se

¹⁰En inglés se le llama *melting region* porque la ARN polimerasa separa las dos hebras de ADN en esa región como si de una desnaturalización térmica se tratara. Para más información sobre la actuación de la ARN polimerasa en el proceso de transcripción, véase el apéndice B.

¹¹Los parámetros que utilizaron en su modelo fueron los mismos que los que utilizó Salerno en los trabajos mencionados en esta memoria (Salerno 1991; Salerno 1992).

¹²Esta generalización de propiedades podemos aplicarla a dos conjuntos de promotores, los denominados *tempranos*, objetivos de la ARN polimerasa de la *E. coli*, y los llamados *tardíos*, que son transcritos por la ARN polimerasa del fago. Para un mejor conocimiento de los tipos de promotores del fago T7, sus características y sus diferentes rendimientos, véase el apéndice B.

¹³Existen promotores mayores y menores, dependiendo de si se transcriben *in vivo* o sólo *in vitro*,

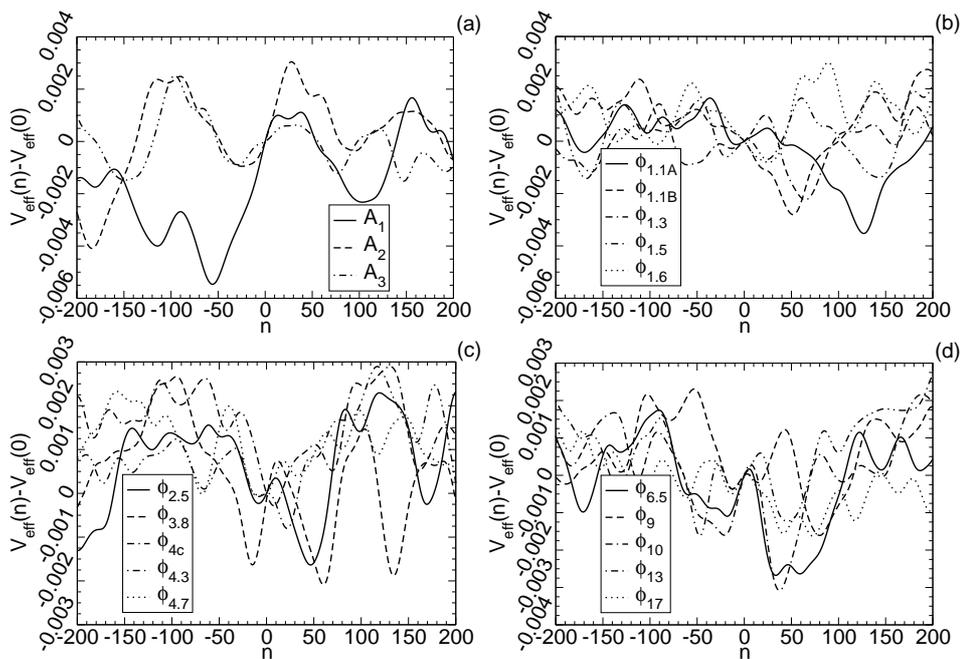


Figura 6.13: Potenciales efectivos V_{eff} de las secuencias de promotores mayores del genoma del fago T7. (a) promotores tempranos A_1 , A_2 y A_3 ; (b) y (c) promotores tardíos de clase II; (d) promotores tardíos de clase III. El origen en el eje horizontal se refiere al TSS de cada promotor.

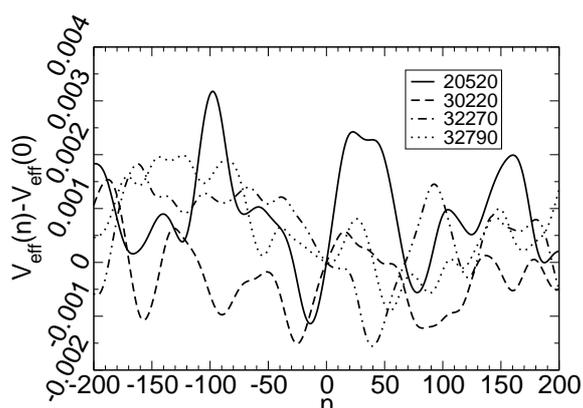


Figura 6.14: Potenciales efectivos V_{eff} en diferentes lugares del genoma del fago T7 que pueden parecerse al V_{eff} de los promotores de la figura 6.13, pero que no lo son. El origen de cada secuencia se corresponde con el número de la leyenda.

repita en todos ellos y que les permita diferenciarse del resto del genoma. Para comprobarlo, en la figura 6.14 se muestra como ejemplo el potencial efectivo V_{eff} de algunas partes del genoma fago que no contienen promotores, y se puede ver su parecido con los de la figura 6.13. Por tanto, debemos concluir que el potencial efectivo, y por tanto la dinámica de *kinks* en el modelo de sG inhomogéneo, no pueden explicar el proceso de iniciación de la transcripción de la ARN polimerasa en los promotores del fago T7 (ni, probablemente, en ningún otro organismo), o al menos no para el nivel de discretización utilizado hasta ahora.

La resolución espacial.

Cabe la posibilidad de que las conclusiones del párrafo anterior se deban a una elección equivocada de la discretización del sistema. Si así fuera, quizá para otra elección de la discretización se pudieran encontrar unas características únicas que diferenciaran el potencial efectivo en torno a promotores del resto del genoma. Pero debemos recordar que el potencial efectivo, V_{eff} , dado por la expresión (5.22), es sólo una media ponderada de la secuencia de q_n alrededor de cada sitio (como ya mencioné en la sección 5.3), con peso dado por la expresión (5.21). Podemos hacer una estimación de la resolución espacial del potencial efectivo si pensamos que, al truncar la suma de la media ponderada (5.22) en $\pm 1.5d_{\text{eff}}^{-1}$ alrededor de

respectivamente. Para más información sobre promotores del fago T7, véase el apéndice B.

cada X , el error que se introduce es aproximadamente del 10%. Si consideramos que, para sitios más alejados, la contribución de q_n a la media ponderada es despreciable, entonces el número de sitios promediados al calcular V_{eff} para cada X va como $\Delta n = 3d_{\text{eff}}^{-1}$. Esto significa que, para $d_{\text{eff}} \simeq 0.07$ [este es el valor de la discretización efectiva utilizada por Salerno (1991) y otros trabajos basados en él (Salerno 1992; Lennholm y Hörnquist 2003; Cuenda et al. 2006)], Δn es aproximadamente 40, una resolución mucho menor que la necesaria para reconocer promotores con una secuencia consenso de 23 PB correspondientes a la ARN polimerasa del fago T7, por no hablar de los hexámeros de la ARN polimerasa de la *E. coli*¹⁴. Por tanto, podemos decir que el *kink* es demasiado ancho como para poder identificar los promotores a través del potencial efectivo asociado a él. Por otro lado, veamos qué ocurre cuando cambiamos la resolución espacial del potencial efectivo aumentando la discretización efectiva hasta encontrar $\Delta n = 1$: en ese caso podríamos encontrar la secuencia consenso de los promotores repetida en el potencial efectivo de cada uno de ellos; sin embargo, aparte de que el potencial efectivo dejaría ya de ser útil por efectos de la discretización del sistema (la barrera de PN ya no es despreciable), de todas formas no tendríamos más información de la que nos ofrece el análisis directo de la secuencia de nucleótidos¹⁵.

El trabajo realizado por Lennholm y Hörnquist (2003) resulta muy interesante, ya que es la primera vez que se estudia la secuencia completa del fago T7 para encontrar diferencias entre los promotores y el resto del genoma que permitan utilizar el modelo de Salerno del ADN como herramienta genómica para la localización de lugares de la secuencia funcionalmente relevantes. Sin embargo, creemos que la metodología utilizada no es la adecuada al problema que se proponían resolver, ya que el análisis estadístico de la dinámica de *kinks* no dio resultados concluyentes. Por el contrario, una gráfica de la distancia máxima recorrida (con signo, para distinguir movimientos del *kink* en un sentido y en otro) en función de la posición inicial de partida hubiese dado lugar a conclusiones diferentes de las obtenidas por Salerno, y el trabajo hubiese sido más concluyente. Además, al no tener en cuenta la dirección de propagación de los *kinks*, no se puede relacionar en el estudio la dirección de transcripción, tan importante en el proceso real. Por tanto, pensamos que es necesario un estudio individual de cada promotor si se quieren encontrar propiedades de estos lugares funcionalmente relevantes. Este estudio individual se ha realizado durante esta tesis doctoral (Cuenda et al. 2006), y la conclusión es desilusionantemente clara: el potencial efectivo de *kinks*, V_{eff} ,

¹⁴Para saber qué es una secuencia consenso, véase el apéndice B.

¹⁵De hecho, tendríamos menos debido a la equivalencia de A y T por un lado y C y G por otro que se introduce en el modelo.

no puede diferenciar los promotores del resto de la secuencia genómica.

Potencial efectivo para breathers.

En este último apartado estudiamos la posibilidad de que los lugares funcionalmente relevantes puedan ser estudiados con este modelo, pero utilizando *breathers* en lugar de *kinks*. Esta posibilidad fue explorada por Bashford (2006), utilizando para ello un potencial efectivo para los *breathers*. En su trabajo, el autor estudia el reconocimiento de determinadas secuencias de ADN por parte de la ARN polimerasa; para ello compara los *breathers* con deformaciones locales de la doble cadena de ADN producidas por el deslizamiento de la ARN polimerasa sobre los surcos del ADN¹⁶. Para ello construye un potencial para los *breathers* siguiendo los pasos utilizados por Salerno y Kivshar (1994), pero utilizando como *Ansatz* el *breather* discreto con centro en X , dado por

$$\phi_{br,n}(X, t) = 4 \arctan \left(\tan \mu \frac{\sin(d_{\text{eff}} n \cos \mu)}{\cosh[\sin \mu (d_{\text{eff}} n - X)]} \right), \quad (6.3)$$

donde μ está relacionado con la frecuencia intrínseca del *breather*, y de nuevo $d_{\text{eff}} = dq_{\text{avg}}^{1/2}$. El potencial obtenido por Bashford (2006) es el siguiente,

$$V_{br}(X) = \sum_n (q_n + q_{\text{avg}}) \text{sech}^2 z_n (1 + \tan^2 \mu \text{sech}^2 z_n)^{-3/2}, \quad (6.4)$$

donde $z_n = \sin \mu (d_{\text{eff}} n - X)$ y, por tanto, aparte de d_{eff} existe otro parámetro que influye en la anchura del *kink*, $\sin \mu$. La principal diferencia entre este potencial V_{br} y el V_{eff} para *kinks* de la expresión (5.22) es que los *breathers* (6.3) no son soluciones estáticas de la ecuación de sG (2.9). Esto significa que el autor, al derivar con respecto al tiempo $\phi_{br,n}(X(t), t)$, obtiene un término que no depende de \dot{X} ; posteriormente, al elevar esta derivada al cuadrado en el término cinético del Hamiltoniano del modelo, (5.14), obtiene de más, por un lado, un término lineal en \dot{X} , y por otro, un término que depende explícitamente de t . Para solucionarlo, los dos términos que dependen de \dot{X} los deja en el término cinético, y en el término potencial hace un promedio temporal después de haber introducido este nuevo término dependiente del tiempo. Para ello debe suponer que la posición del *breather* no varía en un periodo de oscilación interna. Otra diferencia importante con respecto a V_{eff} es la estabilidad: como hemos visto hasta ahora, los *kinks* son objetos muy robustos que se comportan muy bien en secuencias inhomogéneas, y

¹⁶De nuevo supondremos que este estudio, en el que no aparece el efecto disipativo del agua, es fenomenológico.

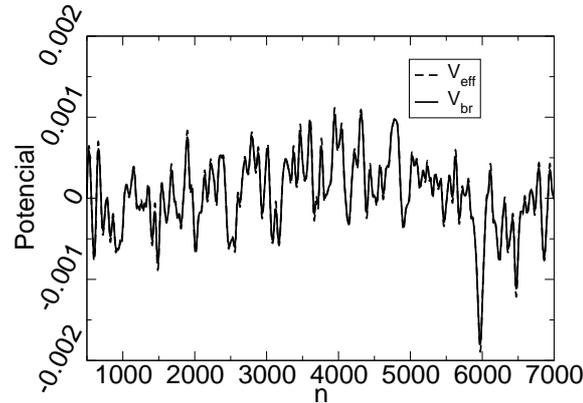


Figura 6.15: Potenciales V_{br} (6.4) para *breathers* (en línea continua) y V_{eff} (5.22) para *kinks* (en línea discontinua) de la secuencia del genoma del fago T7 que va de PB 500 a PB 7000. Ver texto para los parámetros utilizados. El potencial V_{br} ha sido escalado en el eje y para que coincida con la escala de V_{eff} , pero no ha sido necesaria ninguna traslación vertical. El parecido entre ambos potenciales es tan grande que apenas se distinguen ambas curvas. Esta región del genoma contiene siete promotores.

por eso pueden ser expresados en función de su centro utilizando el *Ansatz* (5.11); los *breathers*, por el contrario, son muy inestables en el modelo de sG discreto y homogéneo, por lo que se espera que lo sean mucho más en secuencias inhomogéneas. Por tanto, es probable que el potencial efectivo para *breathers* (6.4) no pueda describir de forma precisa y/o a largos tiempos la dinámica de un *breather* inicialmente en reposo. La construcción de un potencial para *breathers* en el modelo discreto de sG no es, por tanto, tan sencillo como en el caso de los *kinks*, y esto debe ser tenido en cuenta a la hora de analizar los resultados. Sin embargo, también debemos decir que no se han realizado estudios en esta tesis doctoral que relacionen de manera cuantitativa la dinámica de *breathers* en secuencias inhomogéneas con la dinámica de una partícula que se mueve de acuerdo al potencial (6.4), y cabe la posibilidad de que, a la vista de las simulaciones que aparecen en la figura 6.8, dicho potencial también describa con la precisión necesaria la dinámica de *breathers*.

Después de construir el potencial (6.4), el autor lo utiliza para analizar la región de promotores tempranos del genoma del fago T7, y una región determinada del fago T5. Entre otras cosas, en el trabajo se sugiere la correlación entre pozos

profundos en el potencial y promotores en la región de promotores tempranos y de clase III¹⁷. Se sugiere también otra relación entre pozos profundos y secuencias terminadoras¹⁸. En la figura 6.15 se recoge el potencial para *breathers* (6.4) para el genoma del fago T7 en la región que va desde PB 500 hasta PB 7000, para un valor de $d_{\text{eff}} \simeq 0.07$, $q_{AT} = 2$, $q_{CG} = 3$, similares a los de los trabajos de Salerno y otros relacionados con él y descritos en esta memoria, y $\sin \mu = 0.5$. En la misma figura se presenta el potencial efectivo para *kinks*, V_{eff} (con $d_{\text{eff}} = 0.07 \times 0.5$ para ajustarse a la anchura del *breather*), aunque apenas se observa diferencia entre uno y otro resultado. Esto es así porque la función peso del potencial para *breathers*, dada por

$$w_{br}(n) \propto \text{sech}^2 z_n (1 + \tan^2 \mu \text{sech}^2 z_n)^{-3/2}, \quad (6.5)$$

donde $z_n = \sin \mu (d_{\text{eff}} n - X)$, tiene una forma muy parecida a la función peso asociada a *kinks*, (5.21), para $\tan \mu < 1$ [que es el caso, de acuerdo con Bashford (2006)], tal y como se puede ver en la figura 6.16.

Por tanto, en el caso de que el potencial V_{br} (6.4) describa bien la dinámica de *breathers* en secuencias inhomogéneas, podemos aplicarle las mismas conclusiones que en el caso del potencial efectivo para *kinks* V_{eff} (5.22). El potencial V_{br} del modelo de sG no puede explicar el inicio del proceso de transcripción de la ARN polimerasa en los promotores. Así, tal y como aprecia Bashford (2006), existen pozos profundos en el potencial V_{br} cerca de algunos promotores del genoma del fago T7. Sin embargo, también existen promotores (los de la clase II) que no están cerca de pozos profundos, así como pozos profundos que no están cerca de ningún promotor, como los encontrados por el propio autor cerca de terminadores. Cuando se eliminan restricciones para tener en cuenta pozos no tan profundos, entonces otros pozos lejos de promotores deberían ser también tenidos en cuenta. Por tanto, la conclusión es que no existe una característica especial en el potencial para los *breathers* que permita la identificación de promotores de una secuencia del genoma, tal y como ocurría con el potencial efectivo para *kinks*.

¹⁷Los promotores del fago T7 tardíos se subclasifican, a su vez, en promotores de clase II y III. Los tempranos también pueden llamarse de clase I. Para más información sobre promotores del fago T7, véase el apéndice B.

¹⁸Un terminador es una secuencia que hace que la ARN polimerasa, después de leerla, se desacople de la doble cadena de ADN que estaba leyendo. Para más información sobre terminadores, véase el apéndice B.

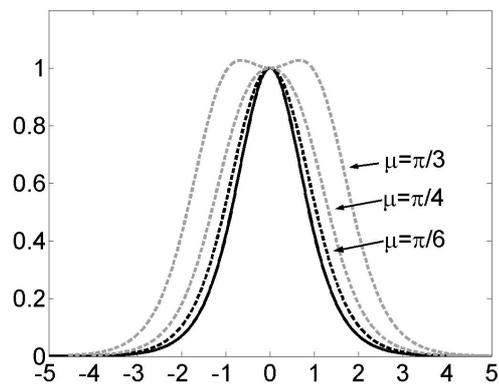
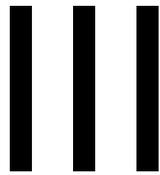


Figura 6.16: Función peso para *kinks*, (5.21), en línea continua. En línea discontinua, funciones peso para *breathers*, (6.5), con el eje x reescalado para que no tenga la contribución de $\sin \mu$, y con el eje y multiplicado por $(1 + \tan^2 \mu)^{3/2}$ para normalizar su altura; en negro, peso para $\mu = \pi/6$, el mismo valor al utilizado en la figura 6.15, y en gris, $\mu = \pi/4$ ($\tan \mu = 1$) y $\mu = \pi/3$ ($\tan \mu > 1$). La línea discontinua apenas se ve debido al gran parecido entre ambas curvas.



Modelo de PB

7

El modelo de Peyrard-Bishop del ADN

En la primera parte del capítulo se introduce el modelo de Peyrard-Bishop del ADN, un modelo de Klein-Gordon nacido para explicar la desnaturalización térmica del ADN, así como la modificación del modelo en el término armónico, introducida por Dauxois, que permite describir mejor la transición de fase del ADN. También se describen las excitaciones dinámicas estudiadas en el modelo, como los *breathers* en el modelo semidiscreto y la solución pared de dominio en la aproximación al continuo. En la segunda parte se explica la analogía existente entre los experimentos de desnaturalización mecánica de ADN y la solución de tipo pared de dominio. Por último, como parte del trabajo original de esta tesis, se presenta el estudio analítico de soluciones de tipo pared de dominio en sistemas finitos con condiciones de contorno fijas como primer paso del estudio de un posible potencial efectivo que permita describir la dinámica de estas excitaciones.

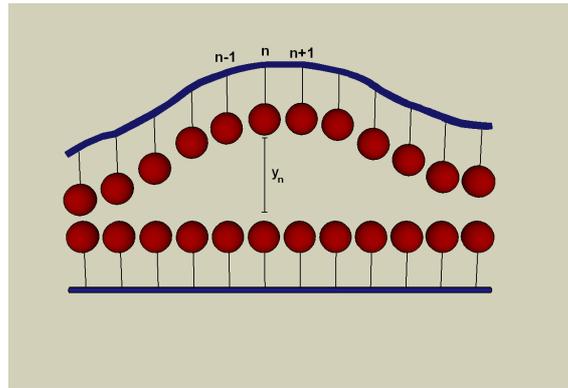


Figura 7.1: Esquema del modelo de PB. y_n representa la separación de la distancia equilibrio del n -ésimo par de bases.

7.1. Introducción del modelo¹.

La transición de fase del ADN.

El modelo de Peyrard-Bishop (PB) del ADN (Peyrard y Bishop 1989) es un modelo de Klein-Gordon no lineal, discreto, que fue introducido a finales de los 80 para describir la desnaturalización térmica del ADN (Inman y Baldwin 1964). Dicha desnaturalización consiste en la separación de las dos hebras de ADN aplicándoles calor, y es fácil de medir el porcentaje de separación de las dos hebras porque la rotura de los pares de bases va acompañada de un aumento de la absorción de la luz ultravioleta en el rango de los 260 nm. En el caso de secuencias homogéneas, esto es, formadas sólo por un tipo de pares de bases, se detecta una transición abrupta entre los estados de doble cadena y dos cadenas separadas, como una transición de fase genuina. Sin embargo, las moléculas de ADN de organismos vivos son altamente inhomogéneas; en ese caso, el proceso de desnaturalización ocurre en múltiples pasos, y depende en gran medida de los detalles de la secuencia (Wartell y Benight 1985).

Peyrard y Bishop (1989) aplican por primera vez al ADN un modelo no lineal y altamente discreto con el que utilizan herramientas de mecánica estadística para determinar la separación de las dos hebras de ADN en función de la temperatura.

¹La mayor parte de estas notas provienen de una revisión muy exhaustiva del modelo que ha publicado recientemente el propio Michel Peyrard (Peyrard 2004).

El hamiltoniano utilizado en el modelo es

$$H = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} m \dot{y}_n^2 + \frac{1}{2} K (y_{n+1} - y_n)^2 + V(y_n) \right\}, \quad (7.1)$$

donde y_n representa la separación de la posición de equilibrio de las bases del n -ésimo par de bases, medida a lo largo de la dirección de los enlaces de hidrógeno que las conectan, tal y como se esquematiza en la figura 7.1. La energía de los enlaces de hidrógeno en función de la distancia entre las bases viene dada por el potencial de Morse,

$$V(y) = D(1 - e^{-ay})^2, \quad (7.2)$$

donde D es la energía de disociación del par y a es un parámetro que da la escala espacial del potencial. Este potencial es típico en enlaces químicos², pero confiere al modelo una diferencia fundamental frente a sG: sólo posee un mínimo, situado en $y = 0$, es decir, a la distancia de equilibrio entre las dos bases que forman un par. Esta característica hace que no posea soluciones de tipo onda solitaria, como sG, y que la única solución estática en un modelo finito con condiciones de contorno libres sea la cadena cerrada. Más adelante, cuando discuta algunas soluciones de este modelo, volveré sobre este tema.

Para el estudio de la desnaturalización térmica del ADN, Peyrard y Bishop (1989) utilizan técnicas de mecánica estadística: calculan la función de partición del modelo en la colectividad canónica, el valor medio de la apertura de las bases, $\langle y \rangle$, y la fracción de pares de bases abiertas, definiendo “estar abierto” como superar una abertura umbral $y_0 = 2\text{Å}$, todo ello en función de la temperatura. Este sencillo modelo da lugar a una temperatura crítica, T_c , por encima de la cual $\langle y \rangle$ diverge. Sin embargo, la fracción de bases abiertas en función de la temperatura, $P(y > y_0, T)$, da una transición muy suave en comparación con los datos experimentales. Aunque se puedan elegir los parámetros adecuados en el modelo de manera que T_c se ajuste al valor real, el modelo necesita ser mejorado. Para ello, Dauxois et al. (1993) modificaron el hamiltoniano (7.1) añadiendo un término anarmónico. El hamiltoniano de este modelo, que llamaré Peyrard-Bishop-Dauxois (PBD), toma la forma

$$H = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} m \dot{y}_n^2 + \frac{1}{2} K \left(1 + \rho e^{-\alpha(y_{n+1} + y_n)} \right) (y_{n+1} - y_n)^2 + V(y_n) \right\}, \quad (7.3)$$

²Realmente, el potencial V es un potencial efectivo en el sentido de que también se incluye el efecto de repulsión de las bases debido a los grupos fosfato, lo que hace muy difícil elegir los parámetros.

donde ρ y α son constantes positivas. Con esta modificación en el término dispersivo, el acoplamiento entre bases vecinas depende de si están abiertas o no. Cuando ambas están cerradas, $y_n \simeq y_{n+1} \simeq 0$, el acoplamiento entre las dos bases es $K' \simeq (1 + \rho)$, mientras que, en cuanto una de ellas se abre (o las dos), el acoplamiento cae a $K' \simeq K$. Con la modificación introducida en la expresión (7.3), cambia de forma drástica el carácter de la transición de desnaturalización en el modelo, volviéndose mucho más abrupto, como en los resultados experimentales.

Dinámica de excitaciones.

Aunque el modelo de PB nació para reproducir la transición de fase del ADN, también es interesante analizar las soluciones dinámicas del modelo (Dauxois et al. 2002; Theodorakopoulos et al. 2004; Cuesta-López et al. 2005). Para revisar brevemente la dinámica del modelo de PB utilizaré el Hamiltoniano (7.1) adimensionalizado,

$$H = \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \dot{Y}_n^2 + \frac{1}{2d^2} (Y_{n+1} - Y_n)^2 + v(Y_n) \right\}, \quad (7.4)$$

donde $Y_n = ay_n$, $d^2 = Da^2/K$, $v(Y) = (e^{-Y} - 1)^2$ y el tiempo está reescalado de la forma $t \rightarrow \sqrt{Da^2/mt}$. Las ecuaciones de movimiento asociadas a dicho Hamiltoniano son

$$\ddot{Y}_n - \frac{1}{d^2} (Y_{n+1} - 2Y_n + Y_{n-1}) + v'(Y_n) = 0, \quad (7.5)$$

donde $v'(Y_n) = 2e^{-Y_n} (e^{-Y_n} - 1)$. Estas ecuaciones no pueden ser resueltas de manera exacta, sino que es necesario introducir algunas aproximaciones.

Una de ellas es el desarrollo en amplitudes pequeñas, es decir, suponiendo $Y_n = \epsilon \phi_n$, con ϵ pequeño, en el límite *semidiscreto* (Remoissenet 1986). Así se obtiene la ecuación de Schrödinger no lineal, y la solución Y_n da lugar a oscilaciones localizadas de gran amplitud, denominadas *breathers*³. Las simulaciones numéricas muestran que estas soluciones, aunque aproximadas, son soluciones estables de (7.5) para determinadas frecuencias de oscilación. Esta solución puede considerarse una buena explicación de la aparición experimental de “burbujas”

³Estas soluciones están relacionadas con los *breathers* de sG, ya que en el límite de pequeña amplitud de estas soluciones en sG se obtiene la ecuación de Schrödinger no lineal perturbada (Newell 1985).

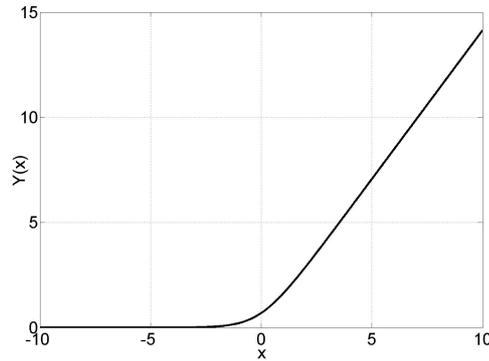


Figura 7.2: Pared de dominio, dada por la expresión (7.9), para $x_0 = 0$.

en la doble hélice de ADN (Guéron et al. 1987), que también aparecen en simulaciones de la dinámica del modelo de PB con temperatura no nula cercana a la temperatura crítica.

Otra aproximación que da lugar a otro tipo diferente de soluciones es la del continuo, tomando $Y_n(t) \rightarrow Y(x, t)$ en la ecuación (7.5),

$$Y_{tt} - Y_{xx} + v'(Y) = 0, \quad (7.6)$$

donde $x = nd$, válida por tanto cuando $d \ll 1$. Esta ecuación posee soluciones estáticas exactas, totalmente diferentes de las soluciones tipo *breather* que vimos anteriormente. Dichas soluciones estáticas deben verificar

$$\frac{d^2 Y}{dx^2} = v'(Y), \quad (7.7)$$

fácilmente integrable por cuadraturas. Multiplicando (7.7) por dY/dx e integrando, se obtiene

$$\frac{1}{2} \left(\frac{dY}{dx} \right)^2 = v(Y) + C, \quad (7.8)$$

donde C es una constante determinada por las condiciones de contorno. Suponiendo un sistema infinito con $Y = Y' = 0$ para $x \rightarrow -\infty$, entonces se obtiene $C = 0$, y por tanto la ecuación (7.8) se integra directamente, obteniéndose

$$e^{Y(x)} = 1 + e^{\sqrt{2}(x-x_0)}, \quad (7.9)$$

donde x_0 es una constante de integración que determina la posición de la solución. Esta solución describe una configuración que está cerrada para $x < x_0$, mientras que para $x > x_0$ está abierta, con una separación entre las bases que crece casi linealmente con x . La ecuación (7.9) representa, por tanto, una pared de dominio entre el estado abierto y cerrado de la molécula de ADN, como se representa en la figura 7.2. En cuanto a la estabilidad de esta solución, un simple análisis energético nos muestra su inestabilidad. La energía de una solución de este tipo en una cadena finita de N elementos viene dada, por un lado, por la contribución de la parte de la cadena cerrada ($Y_n \simeq 0$), que es cero en el término potencial y cero en el de acoplamiento; por otro, la parte de la cadena totalmente abierta (suficientemente alejada de x_0) contribuye a la energía con una unidad en el término potencial y, para el término de acoplamiento, $(\sqrt{2})^2/2 = 1$; finalmente, el entorno de x_0 contribuye con una cantidad constante e_0 , independiente del tamaño del sistema (salvo efectos de borde). Por tanto, la energía total de una pared de dominio centrada en $x_0 = n_0 d$ en una cadena de tamaño N es

$$E_{pd} \simeq 2(N - n_0) + e_0. \quad (7.10)$$

Esto significa que, para N finito, la energía de la pared de dominio disminuye conforme se desplaza su centro a la derecha (es decir, conforme la cadena se cierra) de manera continua, siendo mínima cuando la cadena está completamente cerrada, por lo que la solución es inestable. En el límite $N \rightarrow \infty$ la expresión (7.10) hace que la energía diverja. Esta diferencia es fundamental con respecto a los otros modelos de Klein-Gordon no lineales que hemos visto hasta ahora, como sG, con más de un mínimo equipotencial. A pesar de estos problemas de inestabilidad, cuando se introducen fluctuaciones térmicas en el modelo el análisis de estabilidad de la pared de dominio puede utilizarse para investigar la desnaturalización térmica del ADN, tanto en el límite continuo (Dauxois et al. 2002) como en el discreto (Theodorakopoulos et al. 2004); este tipo de soluciones también se han utilizado en el estudio de la desnaturalización mecánica del ADN (Cuesta-López et al. 2005).

7.2. Estudio analítico de la pared de dominio.

Los experimentos de desnaturalización mecánica.

Los experimentos de desnaturalización mecánica de moléculas de ADN fueron realizados por primera vez por Essevaz-Roulet et al. (1997). En ellos, uno de los extremos de una de las hebras de una molécula de ADN se une a otra molécula

de ADN que, a su vez, tiene su otro extremo fijo en un plato experimental. El extremo libre de la otra hebra de ADN se sujeta a una cuenta de cristal de la que se tira. En un principio se tiraba con una microaguja, aunque más tarde se consiguieron resultados más precisos mediante pinzas ópticas (Bockelmann et al. 2002).

Estos experimentos (Essevaz-Roulet et al. 1997; Bockelmann et al. 2002; Liphardt et al. 2001; Harlepp et al. 2003; Danilowicz et al. 2003) pueden ofrecer información relevante en la dinámica de replicación del ADN, y ya han sido abordados teóricamente para predecir los resultados experimentales (Lubensky y Nelson 2002; Cocco et al. 2002; Marenduzzo et al. 2002; Cuesta-López et al. 2005), tanto en secuencias homogéneas como inhomogéneas. También existen trabajos en los que se estudia la inferencia de la secuencia inhomogénea a partir de los resultados experimentales (Baldazzi et al. 2006; Baldazzi et al. 2007). En estos experimentos de desnaturalización mecánica de una sola molécula de ADN se observa la rotura de los pares de bases en grupo, no de uno en uno, seguidos de pausas en las que no se rompe ningún par de bases hasta el próximo ciclo de roturas, en un mecanismo denominado *stick-slip*⁴. Existen estudios teóricos de estos experimentos que utilizan el modelo de PB (Cuesta-López et al. 2005). En una parte del estudio añaden una fuerza armónica sobre el primer elemento de la cadena del modelo de PBD y estudian numéricamente la desnaturalización mecánica para una secuencia inhomogénea; posteriormente aproximan la molécula abierta por una solución de tipo pared de dominio (7.9) del modelo de PB y lo utilizan como *Ansatz* para calcular la variación de energía de la molécula conforme se mueve por la secuencia; con este resultado apuntan a un posible método para obtener el potencial efectivo de secuencias inhomogéneas.

Estabilidad en modelos discretos de Klein-Gordon no lineales.

Motivados por la búsqueda de una herramienta que permitiera predecir el comportamiento de soluciones de pared de dominio en secuencias inhomogéneas, y así poder comparar fácilmente resultados con los experimentos de desnaturalización mecánica, en el desarrollo de esta tesis estudiamos de forma analítica las soluciones de tipo pared de dominio del modelo de PB (Cuenda y Sánchez 2006).

La primera parte del estudio se centra en la estabilidad de las soluciones de tipo pared de dominio. Dichas soluciones, como he mencionado anteriormente, son inestables para N finito, y su energía diverge en el caso infinito. Por tanto, nosotros nos centraremos en el caso de condiciones de contorno fijas, aplicable a los experimentos de desnaturalización mecánica, en los que el extremo abierto de la

⁴Se pega-desliza.

cadena está sujeto, ya sea con pinzas ópticas o con otro mecanismo, y evita que se cierre. Para trasladar esto a las ecuaciones de movimiento (7.5), las reescribimos como una la relación de recurrencia para las soluciones estáticas,

$$Y_{n+1} = 2Y_n - Y_{n-1} + d^2 v'(Y_n), \quad (7.11)$$

para $n = 1, 2, \dots, N$. Si impusiéramos condiciones de contorno libres, tendríamos una última ecuación $Y_{N-1} = Y_N$ que daría lugar a la solución de cadena cerrada, pero para describir paredes de dominio utilizaremos las condiciones *iniciales* $\{Y_0, Y_1\}$. Nos centraremos en el caso con $Y_0 = 0$, que es el que nos interesa, lo que nos deja un único parámetro, $Y_1 = y$, en función del cual se puede obtener la apertura del n -ésimo elemento de la cadena; por tanto, siempre que me refiera a Y_n será como función de y , $Y_n(y)$, y discutiré el comportamiento de estas soluciones en función de y . Siempre que $y > 0$ tendremos soluciones distintas de la cadena cerrada.

La ecuación (7.11) describe las soluciones estáticas, tanto estables como inestables, del sistema definido por el Hamiltoniano (7.4). Una forma de estudiar la estabilidad de dichas soluciones es utilizando el Hessiano del sistema,

$$\mathcal{H}_N(y) = \begin{pmatrix} a_1(y) & -1 & 0 & \dots & 0 \\ -1 & a_2(y) & -1 & \dots & 0 \\ 0 & -1 & a_3(y) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_N(y) \end{pmatrix}, \quad (7.12)$$

donde $a_n(y) = 2 + d^2 v''(Y_n(y))$. Sea $\Delta_n(y)$ el menor principal de orden n del hessiano \mathcal{H}_N , es decir, $\Delta_n(y) = \det(\mathcal{H}_n)$; entonces una solución estable debe satisfacer $\Delta_n(y) > 0$ para todo $n = 1, 2, \dots, N$. Puesto que el hessiano (7.12) es una matriz tridiagonal, hay una relación recursiva para Δ_n ,

$$\Delta_{n+1}(y) = a_{n+1}(y)\Delta_n(y) - \Delta_{n-1}(y), \quad (7.13)$$

con $\Delta_1 = a_1$ y $\Delta_2 = a_1 a_2 - 1$.

La expresión (7.13) puede reescribirse en términos de $Y'_n(y)$. Derivando la expresión (7.11) respecto a y , se tiene que

$$Y'_{n+1}(y) = a_n(y)Y'_n(y) - Y'_{n-1}(y), \quad (7.14)$$

con $Y'_2(y) = \Delta_1(y)$ y $Y'_3(y) = \Delta_2(y)$. Por tanto, debe cumplirse que $\Delta_n(y) = Y'_{n+1}(y)$ para $n = 1, 2, \dots, N$, con lo que las regiones de estabilidad de las soluciones (7.11) son los puntos que satisfacen la condición $Y'_n(y) > 0$ para todo

$n = 2, 3, \dots, N + 1$. Este resultado es válido para cualquier valor de d y para cualquier potencial $v(Y)$. En los siguientes apartados nos restringiremos al caso del potencial de Morse, y estudiaremos por separado la forma de las soluciones y su estabilidad en los límites $d \ll 1$ y $d \gg 1$.

Límite continuo del modelo de PB.

Este límite se corresponde al caso $d \ll 1$, lo cual significa que se puede tomar la aproximación $Y_n(y) \rightarrow Y(x, y)$, donde $x = nd$, en la ecuación (7.11). El resultado es la ecuación (7.7), como ya mostré anteriormente. Sin embargo, la diferencia con el estudio anterior es que, en este caso, suponemos una cadena con un origen, de manera que sólo consideramos $x \geq 0$, en contraposición a la cadena infinita de antes. Además, las condiciones iniciales son, para que se ajusten a las de la ecuación (7.11), $Y(0, y) = 0$ y $\partial Y(0, y)/\partial x = y/d$. La integración de (7.7) en este caso da lugar a la solución

$$e^{Y(x,y)} = \frac{\sqrt{2}\frac{y}{d} \sinh \left[\sqrt{2 + \frac{y^2}{d^2}} x + \sinh^{-1} \left(\frac{y}{d\sqrt{2}} \right) \right] + 2}{2 + \frac{y^2}{d^2}}, \quad (7.15)$$

válida para $x \geq 0$. Nótese la diferencia con la expresión (7.9), que sólo es válida en el límite continuo de una cadena infinita. En el límite continuo de sistemas finitos, sin embargo, la condición inicial (7.9) no permanece estática al hacerla evolucionar en la ecuación (7.5). Por tanto, no podemos decir que (7.9) sea una solución estacionaria de la ecuación (7.5) en el límite continuo para sistemas finitos.

Una característica importante de la aproximación al continuo es que, en la ecuación (7.15), Y puede escribirse como función de x y ξ , donde $\xi = y/d$. Esto implica una relación de escala entre estos dos parámetros, una relación que desaparece en las soluciones en el límite discreto. Este comportamiento se puede observar en la figura 7.3, donde se representa la aproximación al continuo Y frente a la posición en la cadena, x , para dos valores de ξ , comparando con el resultado exacto $Y_n(y)$ (recordemos que $x = nd$) para diferentes valores de la discretización del sistema, d , siendo $y = \xi d$. Claramente se observa que para valores grandes de d ($d = 1$ y, sobretudo, $d = \sqrt{10}$) la relación de escala no se verifica, lo que indica el paso al régimen discreto.

Anteriormente vimos que, para analizar la estabilidad de una solución estática en un sistema discreto de tamaño N , era necesario estudiar el signo de $Y'_n(y)$ para $n = 2, 3, \dots, N + 1$. En la aproximación del límite continuo se puede generalizar el resultado al estudio del signo de $\partial Y(x, y)/\partial y$ para $x > 0$. Puesto

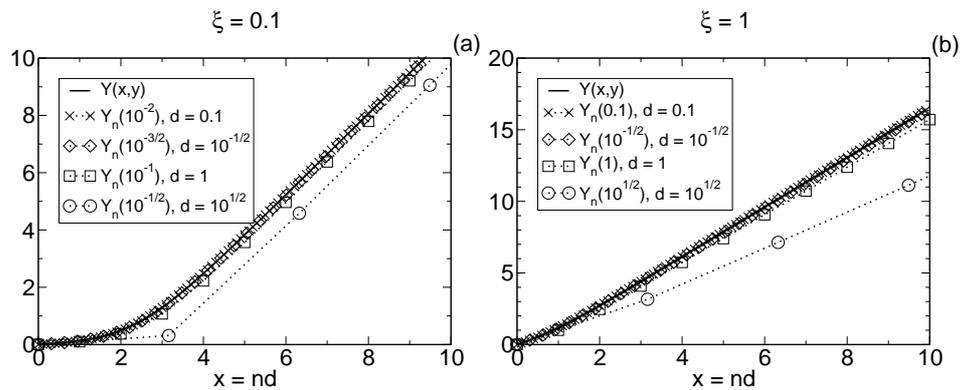


Figura 7.3: Gráficas de $Y(x, y)$ obtenidas a partir de (7.15) para (a) $\xi = 0.1$ y (b) $\xi = 1$, con $\xi = y/d$, comparadas con la solución exacta $Y_n(y)$, obtenida a partir de la relación de recurrencia (7.11), calculada para el mismo cociente $\xi = y/d$ pero para diferentes valores de d (y, por tanto, diferentes valores de y). En ambas gráficas se observa que la aproximación es buena para valores pequeños de d , siendo peor para $d = \sqrt{10}$.

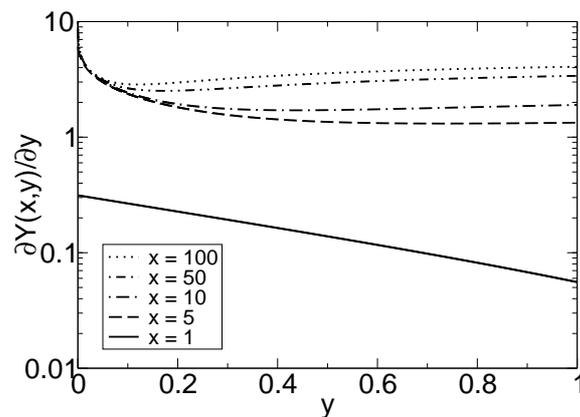


Figura 7.4: Gráfica de $\partial Y(x, y)/\partial y$ para (de abajo a arriba) $x = 1$, $x = 5$, $x = 10$, $x = 50$ y $x = 100$. Todos ellos son positivos para cualquier $y > 0$, de manera que todas las soluciones de la forma (7.15) son estables para cualquier $y > 0$.

que la derivada de la expresión (7.15) es bastante complicada, en la figura 7.4 se muestran las gráficas del resultado para diferentes valores de x . En ellas se observa que $\partial Y(x, y)/\partial y > 0$ para cualquier $x > 0$, y por tanto la expresión (7.15) es una solución estable de (7.11). Para comprobar estos resultados, integramos las ecuaciones de movimiento (7.5) a partir de las condiciones iniciales dadas por la expresión (7.15) con pequeñas perturbaciones. Utilizando condiciones de contorno fijas para evitar el cierre espontáneo de la cadena, las simulaciones confirman la estabilidad de las soluciones. Hay que insistir en que dichas soluciones son las relevantes en el problema de desnaturalización mecánica, donde la fuerza ejercida sobre el extremo de la cadena impide su cierre espontáneo.

Límite discreto del modelo de PB.

El límite discreto del modelo de PB corresponde al caso $d \gg 1$, que es el caso relevante al aplicar el modelo al ADN debido a que se trata de un sistema discreto. En el modelo de PB el parámetro de discretización del sistema puede variar desde $d^2 = 10.1$ (Peyrard 2004; Theodorakopoulos et al. 2004) hasta $d^2 \simeq 75$ (Campa y Giansanti 1998), o incluso $d^2 = 100$ (Qasmi et al. 2004), dependiendo de los parámetros que se usen para describir el sistema. En todos ellos, los valores de d corresponden a sistemas que se encuentran lejos del límite continuo. Por tanto, entender el comportamiento de las soluciones de tipo pared de dominio en el límite discreto es fundamental para el estudio de la desnaturalización mecánica del ADN.

La ecuación en diferencias de partida, (7.11), puede reescribirse utilizando la suma telescópica de $Y_{n+1} - Y_n$, teniendo en cuenta las condiciones iniciales $Y_0 = 0$ e $Y_1 = y$, obteniéndose

$$Y_{n+1}(y) = (n+1)y + d^2 \sum_{k=1}^n (n+1-k)v'(Y_k(y)). \quad (7.16)$$

Esta expresión no da de forma explícita la dependencia de Y_n en función de y ; sigue siendo una relación de recurrencia, que ahora tiene en cuenta todas las $Y_k(y)$, con $k < n$, a la hora de calcular $Y_n(y)$. Definamos

$$f_k(y) \equiv v'(Y_k(y)) = 2e^{-Y_k(y)}(1 - e^{-Y_k(y)}). \quad (7.17)$$

Puesto que $Y_1 = y$, es claro que se cumple $f_k(y) = f_1(Y_k(y))$ para todo $k \geq 1$. Nótese que, puesto que no tenemos una expresión explícita de Y_k en función de y para $k > 1$, la única forma de calcular $f_k(y)$ es la numérica, ya sea utilizando la fórmula de recurrencia (7.16) o con la expresión (7.11). En la figura 7.5 se muestran estas funciones para valores grandes de d . Como se puede observar, las

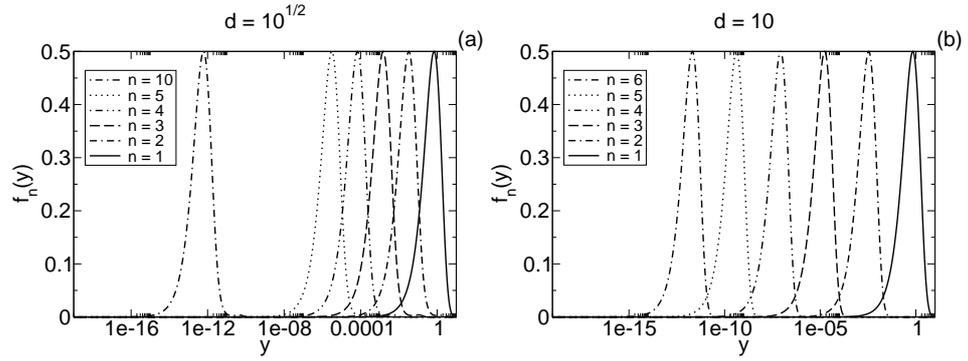


Figura 7.5: Funciones $f_n(y)$ para diferentes valores de n , y (a) $d = \sqrt{10}$ y (b) $d = 10$.

funciones f_k están muy localizadas, y el solapamiento de su soporte depende de d , siendo menor el solapamiento cuanto mayor es la discretización. De hecho, en el límite discreto, lo que implica un bajo solapamiento de las curvas, podemos aproximar

$$f_k(y) \simeq f_k^{(1)} \equiv f_1(b_k y), \quad (7.18)$$

siendo b_n una dilatación que traslada la posición del máximo, como aparece trasladado en la figura 7.5. Para calcularlo hay que tener en cuenta que $y_1 = \log 2$ es la posición del máximo de $f_1(y)$; aproximando $f_1(y) \simeq 2y$ para $y \ll y_1$ en la expresión de $f_2(y)$ obtenida a partir de (7.17), se tiene que $b_2 = 2d^2 + 2$. Aunque aparentemente la traslación de las jorbas de la figura 7.5 es constante, lo que daría lugar a potencias de b_2 para el resto de las b_n , existe una pequeña corrección. Para encontrarla basta repetir el mismo proceso que para f_2 con el resto de las f_n , ahora en función de $f_{n-1}, f_{n-2}, \dots, f_1$, y suponiendo siempre que $y \ll y_{n-1}$, donde y_{n-1} es la posición del máximo de $f_{n-1}(y)$. De esta manera se obtiene la relación de recurrencia lineal

$$b_{n+1} = (2d^2 + 2)b_n - b_{n-1}, \quad (7.19)$$

que puede resolverse utilizando como condiciones iniciales $b_1 = 1$ y $b_2 = 2d^2 + 2$:

$$b_n = \frac{1}{2d\sqrt{d^2 + 2}} \left\{ \left[d^2 + 1 + d\sqrt{d^2 + 2} \right]^n - \left[d^2 + 1 - d\sqrt{d^2 + 2} \right]^n \right\}. \quad (7.20)$$

Con estos valores de b_n podemos comparar la solución f_n , que proviene del cálculo numérico de la ecuación (7.17), y la solución aproximada $f_n^{(1)}$ para dife-

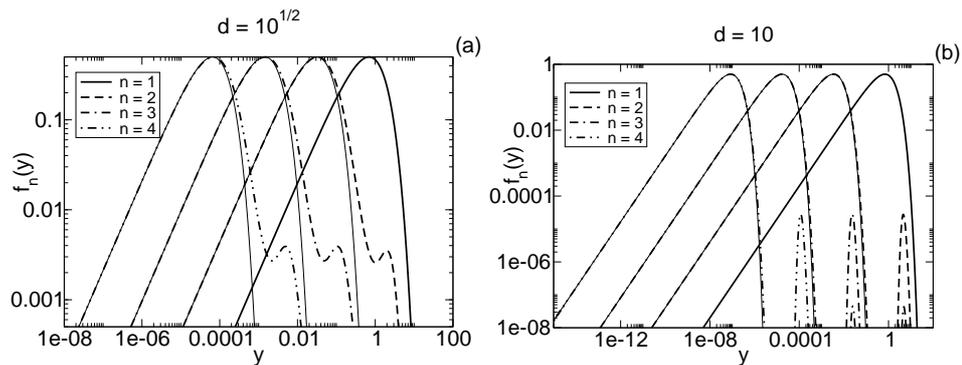


Figura 7.6: Aproximación $f_n^{(1)}(y)$ dada por la expresión (7.18) frente a la $f_n(y)$ exacta, (7.17), para diferentes valores de n , y (a) $d = \sqrt{10}$ y (b) $d = 10$. Las soluciones exactas están dibujadas con líneas gruesas, mientras que la correspondiente aproximación se dibuja con líneas continuas finas. El eje vertical está en escala logarítmica para acentuar las diferencias entre cada par de curvas. Es claro que la aproximación es mejor para valores grandes de d (nótese la diferencia de escala en ambas gráficas). Las gráficas también muestran la diferencia principal entre f_n y f_{n-1} , aparte de su traslación (dilatación), que es la aparición de una nueva joroba, mucho menor que la anterior, a la derecha del máximo de f_{n-1} . De esta forma, f_1 sólo muestra una joroba, f_2 muestra una joroba secundaria, f_3 tres, y así.

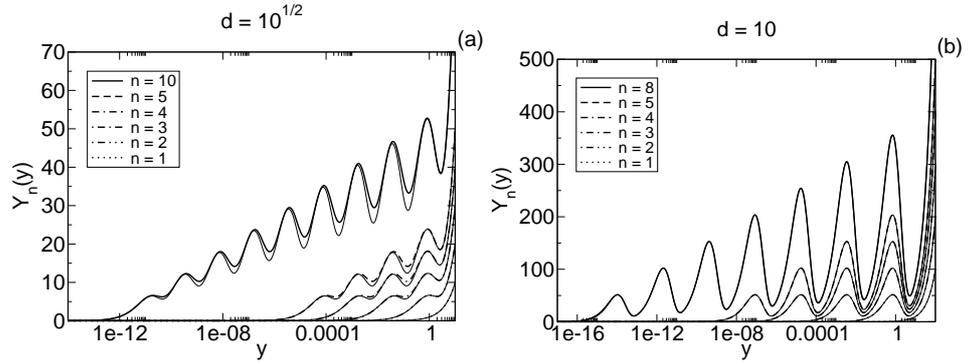


Figura 7.7: Aproximación $Y_n^{(1)}(y)$ dada por la expresión (7.21) frente a la solución exacta obtenida numéricamente a partir de (7.11), para (a) $d = \sqrt{10}$ y (b) $d = 10$ y diferentes valores de n . Las soluciones exactas están dibujadas con líneas gruesas, mientras que la correspondiente aproximación se dibuja con líneas continuas finas. La aproximación es mejor cuanto mayor es d .

rentes valores de n y d . La figura 7.6 muestra la validez de la aproximación, sobre todo para valores grandes de d (véase el pie de la figura para más detalles).

Este resultado nos permite obtener una expresión analítica aproximada de las soluciones estáticas en el límite discreto en función de y . Así, sustituyendo (7.18) en (7.16), obtenemos que

$$Y_{n+1}^{(1)}(y) = (n+1)y + d^2 \sum_{k=1}^n (n+1-k) f_k^{(1)}(y) \quad (7.21)$$

es una buena aproximación de la solución exacta $Y_n(y)$ [calculada numéricamente con la ecuación de recurrencia (7.16)] para valores grandes de d . La figura 7.7 muestra la precisión de la aproximación: para $d \gtrsim \sqrt{10}$, la aproximación es muy buena, mientras que la aproximación se va haciendo peor conforme disminuye d (consideramos valores de $d \geq \sqrt{10}$, lo que incluye todas las constantes utilizadas en la literatura y que mencionamos anteriormente). Para valores menores de d , la aproximación anterior puede mejorarse utilizando la función $f_2(y)$ en lugar de $f_1(y)$ como aproximación del resto de los $f_k(y)$, definiendo

$$f_k(y) \simeq f_k^{(2)} \equiv f_2 \left(y \frac{b_k}{b_2} \right) \quad (7.22)$$

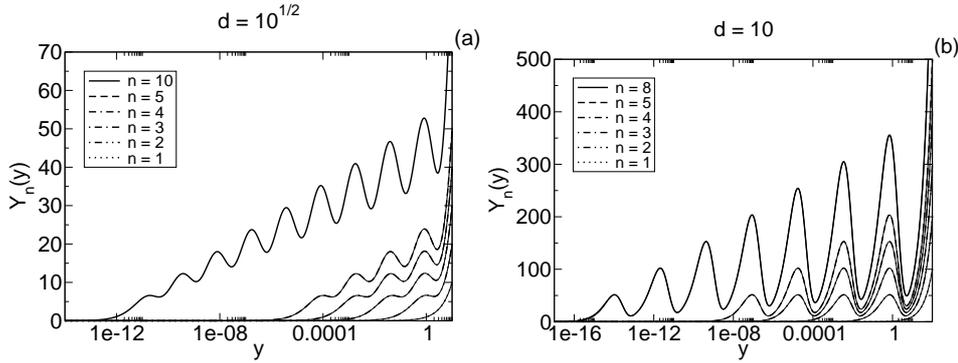


Figura 7.8: Aproximación $Y_n^{(2)}(y)$ dada por la expresión (7.23) frente a la solución exacta obtenida numéricamente a partir de (7.11), para (a) $d = \sqrt{10}$ y (b) $d = 10$ y diferentes valores de n . Las soluciones exactas están dibujadas con líneas gruesas, mientras que la correspondiente aproximación se dibuja con líneas continuas finas. En esta escala apenas se observa diferencia entre las dos soluciones.

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

$$Y_{n+1}^{(2)}(y) = (n+1)y + nd^2 f_1(y) + d^2 \sum_{k=2}^n (n+1-k) f_k^{(2)}(y). \quad (7.23)$$

En este caso la aproximación es incluso mejor que $Y_n^{(1)}(y)$, en incluso para $d = \sqrt{10}$ se obtienen resultados muy parecidos a los obtenidos numéricamente a partir de (7.16), como se puede observar en la figura 7.8.

Los errores cometidos en estas aproximaciones dependen de los valores de d y n . Así, por ejemplo, $Y_n^{(1)}(y)$ es exacto para $n = 1$ y $n = 2$, para cualquier valor de d , mientras que $Y_n^{(2)}(y)$ es exacto hasta $n = 3$ para cualquier d . Para valores pequeños de n , la principal diferencia entre el $f_n(y)$ exacto, obtenido numéricamente, y $f_n^{(1)}$, aproximado, se encuentra en la segunda joroba de $f_n(y)$, con un error máximo $E_{\text{máx}}^{(1)} \simeq 0.06$ para $d = \sqrt{10}$ y $E_{\text{máx}}^{(1)} \simeq 0.006$ para $d = 10$. La aproximación $f_n^{(2)}(y)$ cubre esa segunda joroba, y el error proviene de la tercera, con un error máximo $E_{\text{máx}}^{(2)} \simeq 2.7 \times 10^{-3}$ para $d = \sqrt{10}$ y $E_{\text{máx}}^{(2)} \simeq 3.2 \times 10^{-5}$ para $d = 10$. Hay errores mucho menores alrededor de las posiciones de las siguientes jorobas. Se puede realizar este mismo cálculo para órdenes mayores de la aproximación, $f_n^{(k)}(y)$, utilizando $f_k(y)$, y puede verse que el error disminuye al menos un orden de magnitud al pasar de k a $k+1$.

Existe un límite computacional, cerca de la precisión de la máquina, que nos impide comprobar la validez de estas aproximaciones más allá de unos ciertos valores de k y n , que dependen del valor de d . Sin embargo, por lo que hemos visto, parece razonable esperar el mismo comportamiento para valores mayores de k y n . Por tanto, nuestra conjetura es que las funciones $f_k(y)$ pueden usarse como aproximación de $f_n(y)$, de la forma $f_n(y) \simeq f_n^{(k)}(y)$, con

$$f_n^{(k)}(y) = f_k \left(y \frac{b_n}{b_k} \right) \quad (7.24)$$

para $n > k$, para obtener mejores aproximaciones $Y_n^{(k)}(y)$ de la solución exacta $Y_n(y)$. El error de la aproximación de orden k , $E_n^{(k)}(y) = |Y_n(y) - Y_n^{(k)}(y)|$, puede estimarse como la diferencia

$$E_n^{(k)}(y) \simeq |Y_n^{(k+1)}(y) - Y_n^{(k)}(y)| + \mathcal{O} \left(E_n^{(k+1)}(y) \right), \quad (7.25)$$

con $E_n^{(k+1)}(y) \ll E_n^{(k)}(y)$.

Las aproximaciones definidas en las expresiones (7.21) y (7.23), así como la generalización a órdenes superiores de k utilizando la función (7.24), nos permiten calcular de manera muy precisa la solución $Y_n(y)$ para prácticamente cualquier valor de n e y . Esto es importante, ya que en el cálculo numérico de las expresiones exactas (7.11) y (7.16) surgen problemas para valores de y cercanos a cero debido a la precisión del ordenador, como veremos más adelante en un ejemplo concreto. En estos casos resulta más conveniente utilizar la aproximación $Y_n^{(k)}(y)$ hasta el orden de aproximación que nos permita la precisión del ordenador. De esta manera podemos trabajar con sistemas de mayor tamaño y para valores de y más cercanos a cero de lo que se podría hacer al resolver directamente de manera numérica la relación original de recurrencia (7.11) o (7.16).

Las soluciones estables en el límite discreto de un sistema de tamaño N deben verificar la condición $Y_n'(y) > 0$ para todo $n = 2, 3, \dots, N + 1$. En la figura 7.9 se muestra $dY_n(y)/d \log_{10}(y)$ en función de la condición inicial y , en escala logarítmica, para diferentes valores de n . Hemos elegido $dY_n(y)/d \log_{10}(y)$ en lugar de $Y_n'(y)$ para dibujar una curva suave, ya que la gráfica de $Y_n'(y)$ hubiese hecho muy difícil observar los intervalos con $Y_n'(y) > 0$, y el signo de ambas gráficas, que es en lo que estamos interesados, es el mismo. La figura 7.9 revela algo parecido a una sinusoide modulada para cada n , con $n - 1$ máximos y mínimos alrededor de $Y_n'(y) = 0$. En la figura se observa la aparición de un nuevo intervalo de inestabilidad para valores pequeños de y en un sistema de tamaño n

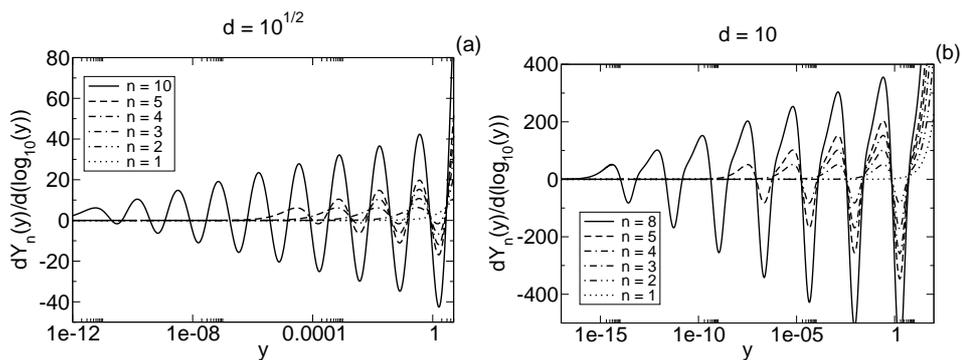


Figura 7.9: Derivada de $Y_n(y)$ respecto a $\log_{10}(y)$ para diferentes valores de n y para (a) $d = \sqrt{10}$ y (b) $d = 10$ en escala logarítmica. La región de estabilidad de un sistema de tamaño N es la intersección de todos los puntos que satisfacen $dY_n/dy > 0$ para $n = 1, 2, \dots, N + 1$. De las figuras, se observa que la región de estabilidad se corresponde con los puntos que satisfacen la condición para $n = N + 1$, ya que la región de estabilidad de un sistema de tamaño N parece estar contenida en la región de estabilidad de un sistema de tamaño $N - 1$ (véase el texto para una explicación más detallada).

comparado con el sistema de tamaño $n - 1$. La figura también sugiere que el conjunto de puntos inestables del sistema de tamaño n contiene el conjunto de puntos inestables del sistema de tamaño $n - 1$. Un posible argumento en favor de esta conjetura es el siguiente: fijémonos en los puntos que satisfacen $Y'_n(y_0) = 0$, en los extremos de un intervalo de inestabilidad; entonces, si $Y'_{n-1}(y_0) > 0$, de acuerdo a la ecuación (7.14), se debe cumplir que $Y'_{n+1}(y_0) < 0$, y por tanto el intervalo de inestabilidad de un sistema de tamaño n será mayor que el de un sistema de tamaño $n - 1$. Esta condición se satisface para todos los nuevos intervalos que aparecen para cada $Y_n(y)$, comenzando desde $Y_2(y)$, y por tanto, por inducción, puede aplicarse a cualquier tamaño del sistema, n . Siguiendo este razonamiento, y teniendo en cuenta la condición de estabilidad, los puntos estables de un sistema de tamaño N son aquellos que satisfacen

$$Y'_{N+1}(y) > 0. \quad (7.26)$$

Comparación de resultados en el límite discreto.

Para comprobar la validez del estudio realizado en el límite discreto, comparamos nuestros resultados con los presentados por Theodorakopoulos et al. (2004). Como parte de su trabajo estudiaron las soluciones estáticas con condiciones de contorno fijas en el límite discreto, con $d^2 = 10.1$, en un sistema con $N = 28$ elementos. Las condiciones de contorno que utilizaron fueron $Y_0 = 0$ e $Y_N = 80$, encontrando ocho soluciones estables y siete inestables. Para ello generaron numéricamente diferentes secuencias utilizando la relación (7.11) para una gran número de condiciones iniciales Y_1 , y aceptaron como soluciones las secuencias que verificaban $Y_{28} \simeq 80$ para un cierto error relativo; posteriormente, rechazaron los casos que eran “vecinos próximos” de la solución, quedándose finalmente con las quince soluciones; la estabilidad de las mismas la establecieron estudiando el signo de los autovalores del Hessiano (7.12) para cada una de ellas, obteniendo las ocho estables y siete inestables mencionadas anteriormente.

En la figura 7.10 se muestra la solución $Y_N(y)$ calculada numéricamente a partir de la relación de recurrencia (7.11), así como la solución aproximada $Y_N^{(2)}(y)$ del sistema descrito en el párrafo anteriormente. En la gráfica se observa claramente el problema de la precisión del ordenador que aparece cuando se intenta calcular la solución exacta de forma numérica para valores pequeños de y . Por otro lado, la aproximación $Y_N^{(2)}(y)$ fue calculada sin ningún problema en un amplio rango de y . La intersección de la curva $Y_N(y)$ (o, en su defecto, de la curva $Y_N^{(2)}(y)$) con la abscisa 80 da las soluciones del sistema con condiciones de contorno fijas. El estudio de la estabilidad se realiza estudiando el signo de $Y_{29}(y)$

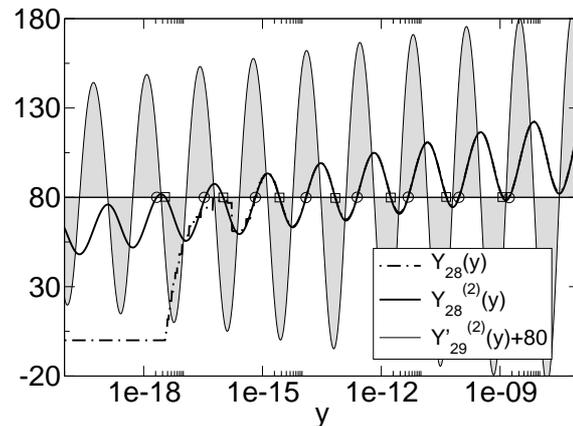


Figura 7.10: Soluciones estables (\circ) e inestables (\square) dados por la aproximación $Y_N^{(2)}$ para un sistema con $N = 28$ elementos, $Y_0 = 0$ e $Y_N = 80$. Las soluciones son las intersecciones de $Y_{28}(y)$ (en línea discontinua) con la línea horizontal $Y_{28} = 80$. Debido a problemas con la precisión del ordenador, hemos utilizado la aproximación $Y_{28}^{(2)}$ (en línea gruesa continua). Para estudiar la estabilidad de las soluciones es necesario conocer el signo de Y_{29}' (línea delgada continua). De nuevo, por claridad, hemos dibujado en su lugar $dY_{29}(y)/d\log_{10}(y)$, y el valor de dicha derivada ha sido trasladado hasta la abscisa 80 para estudio más sencillo de su signo. Las soluciones estables (\circ) son aquellas con $Y_{29}' > 0$, mientras que las soluciones inestables (\square) tienen $Y_{29}' \leq 0$. Nuestra aproximación da ocho soluciones estables y siete inestables, exactamente igual que en (Theodorakopoulos et al. 2004).

(o, en su defecto, de su aproximación $Y_{29}^{(2)}(y)$), que también aparece dibujada en la figura, de manera que las soluciones que se encuentren en un intervalo con la derivada positiva son estables, mientras que el resto son inestables (véase el pie de la figura para más detalles). El resultado del estudio da ocho soluciones estables y siete inestables, como en el resultado anterior de Theodorakopoulos et al. (2004). Este ejemplo muestra la utilidad de nuestro método para calcular soluciones estables e inestables, utilizando las aproximaciones propuestas cuando la resolución numérica directa da problemas debido a la precisión de la máquina. Por este motivo, el método propuesto permite obtener las soluciones estables e inestables para valores mayores de N utilizando las aproximaciones propuestas. Es, además, mucho más económico que el método de disparo de Theodorakopoulos et al. (2004), que tiene que calcular por integración numérica muchas soluciones que luego hay que descartar, frente a la evaluación de una función que facilita nuestro método.

IV

Conclusión

8

Conclusiones y problemas abiertos.

El trabajo original realizado durante esta tesis es de dos tipos: estudios de la validez del modelo de sG como aproximación de la dinámica del ADN, y trabajos relativos a aproximaciones y propiedades del modelo matemático (principalmente de sG, aunque también de PB), sin aplicación práctica directa a casos reales. Por este motivo, al final de esta tesis doctoral se pueden extraer varias conclusiones en diferentes contextos, pero relacionadas entre sí.

La relevancia de la anchura del *kink*.

Una de ellas está relacionada con el modelo de sG continuo. En el capítulo 2 vimos que el *kink* es un objeto extenso con propiedades de partícula relativista. En el capítulo 3, además, vimos que es un objeto muy estable y rígido, y por lo general su dinámica en sistemas inhomogéneos está bien descrita en términos de su centro. Para ello se utiliza la técnica de coordenadas colectivas, con la que se obtiene un potencial efectivo que define el movimiento del centro del *kink* (Fogel et al. 1976). Sin embargo, mostramos un ejemplo concreto de inhomogeneidad sinusoidal en el que, para determinados valores de la longitud de onda, dicha aproximación no es válida: mientras que la aproximación dada por el potencial efectivo resulta útil en rangos de longitud de onda pequeña o grande en relación

a la anchura del *kink*, no es válida en un rango intermedio de valores (Sánchez y Bishop 1994); en ese intervalo, el *kink* se queda atrapado en algún mínimo del potencial efectivo. En nuestro estudio (Cuenda y Sánchez 2005) mostramos que una descripción del *kink* en términos de dos coordenadas colectivas, el *kink* y la anchura, es suficiente para reflejar el mismo atrapamiento del *kink* en un rango intermedio de longitudes de onda. El hecho de que esta descripción tan sencilla tenga éxito nos muestra la relevancia de la anchura del *kink* en el problema. La comparación de los resultados de la aproximación con el modelo completo, sin embargo, nos dice que el nivel de descripción es sólo cualitativo. Esto se debe a que hay otros factores en el atrapamiento del *kink* que en esta aproximación no hemos tenido en cuenta. Así, el *kink* emite radiación al moverse por la inhomogeneidad, de manera que va perdiendo energía poco a poco hasta que, finalmente, no tiene la energía suficiente para sobrepasar la siguiente barrera; en estos casos, el *kink* superará varias longitudes de onda hasta que, finalmente, se pare. Sin embargo, de los resultados numéricos del modelo completo se observa que estos casos se diferencian bastante bien de otros en los que el *kink* se queda atrapado en uno de los primeros pozos. Éstos son aquellos en los que la anchura del *kink* puede tener un papel relevante.

Sin embargo, todavía no hay una explicación clara de la forma en la que la anchura del *kink* interviene en el problema. La integración numérica de las ecuaciones que se obtienen en la aproximación no permite descubrir por qué, para ciertos valores de la longitud de onda de la perturbación, el *kink* decide pararse en un mínimo de potencial. Aunque hay algunos trabajos al respecto realizados durante la elaboración de esta tesis, esta pregunta todavía no tiene una respuesta clara.

La utilidad del V_{eff} .

El problema del potencial efectivo descrito en el apartado anterior no es habitual. Por lo general, el potencial efectivo describe bastante bien la dinámica de *kinks* en modelos de sG inhomogéneos. En el capítulo 5 se introduce el potencial efectivo para el modelo de sG discreto en secuencias inhomogéneas (Salerno y Kivshar 1994), y se hace un estudio comparativo de la dinámica de *kinks* y de la dinámica de una partícula que se mueve en el potencial efectivo, tanto con rozamiento y fuerza externa como sin ellas. El estudio se realiza para valores de la discretización efectiva en torno a $d_{\text{eff}} \simeq 0.1$, con una barrera de PN despreciable; quizá para otras discretizaciones aparezcan efectos en los que la anchura del *kink* sea relevante, como el descrito en el apartado anterior o, para discretizaciones demasiado grandes, el potencial de PN deje de ser despreciable. Sin embargo, para

el rango en el que nos hemos movido la similitud entre las dinámicas del *kink* y de la partícula es bastante grande. La mayor diferencia se observa en el caso sin rozamiento ni fuerza externa, en el que aparece un desfase entre ambas dinámicas cuando la velocidad del *kink* oscila en un cierto rango (aunque en algunos casos la diferencia se acentúa más que en otros); dicha diferencia se debe a que en ese proceso están interviniendo otros mecanismos que no han sido tenidos en cuenta en la aproximación de potencial efectivo, como la emisión de radiación por parte del *kink* o la diferencia en su anchura. Cuando se incluyen rozamiento y fuerza externa en el modelo, sin embargo, la aproximación mejora, probablemente porque el rozamiento anula la radiación que pueda emitir el *kink*, así como la oscilación de su centro. La comparación realizada en ambos casos nos permite afirmar que, en las condiciones estudiadas, la dinámica de la partícula en el potencial efectivo nos sirve para predecir resultados en la dinámica de *kinks*.

Un trabajo interesante como continuación de éste sería una comparación más exhaustiva de las dos dinámicas para diferentes tipos de secuencias (y, por tanto, de potenciales efectivos) y con diferentes valores de los parámetros. Por ejemplo, construir secuencias periódicas para que den lugar a potenciales efectivos de tipo sinusoidal y estudiar los efectos de la anchura en ellos; estudiar el grado en que el rozamiento y la fuerza externa afectan a la similitud entre una dinámica y otra; o también, comprobar los efectos que la introducción de ruido térmico induce en el parecido entre ambas dinámicas.

Aplicabilidad de sG al ADN.

En el capítulo 6 se revisan algunos resultados referentes al modelo de sG aplicado al ADN (Salerno 1991; Salerno 1992; Domínguez-Adame et al. 1995). Para ello se usan, entre otras cosas, los resultados descritos anteriormente referentes a la utilidad del potencial efectivo para describir la dinámica de *kinks* en secuencias inhomogéneas. Gracias a ellos, resulta mucho más intuitivo entender (y a veces imaginar) la dinámica de un *kink* en una secuencia inhomogénea en términos de su potencial efectivo. Así, se deduce que, una vez fijados el resto de parámetros (rozamiento, fuerza externa y discretización efectiva), la dinámica de un *kink* depende en gran medida de la posición inicial de su centro y de la secuencia inhomogénea por la que se mueva. Este resultado, aparentemente trivial en términos de una partícula, resulta de gran ayuda para entender algunos trabajos anteriores referentes al modelo, aplicado a secuencias aperiódicas (Domínguez-Adame et al. 1995) y a secuencias en las que aparecen promotores de ADN (Salerno 1991; Salerno 1992). Dichos resultados sugieren la posibilidad de que la dinámica de *kinks* esté relacionada con el contenido de información de la secuencia, en el caso de se-

cuencias aperiódicas, o con las posiciones de elementos funcionales de secuencias de ADN, como promotores. Sin embargo, en nuestro estudio (Cuenda y Sánchez 2004a; Cuenda y Sánchez 2004b) se explican los resultados obtenidos en trabajos anteriores en términos de la posición inicial del *kink* y la secuencia utilizada. Teniendo en cuenta esto, nuestra conclusión es que no existen dichas relaciones en el modelo.

Posteriormente, y basándonos en otros trabajos que estudian la misma relación (Lennholm y Hörnquist 2003; Bashford 2006), se realiza un estudio exhaustivo del genoma del fago T7 utilizando para ello el potencial efectivo para *kinks*. Tras un análisis de todos los promotores del fago, así como de otras secuencias que no contienen promotores, no encontramos rasgos característicos que identifiquen estos elementos de entre el resto del genoma. Además, tras un análisis del peso del potencial efectivo concluimos que la resolución espacial de éste, para los parámetros utilizados en la literatura estudiada, no es lo suficientemente alta como para encontrar las secuencias de promotores del fago y que, para otras discretizaciones que aumentarían la resolución del potencial efectivo, el potencial efectivo da información sesgada de la secuencia por no poder diferenciar las bases A y T y las C y G. Otro análisis de un potencial efectivo para *breathers* que ha aparecido recientemente aplicado al ADN (Bashford 2006) indica, por un lado, que el potencial efectivo para estos objetos es más complicado que en el caso de los *kinks* y, por otro, que su expresión da lugar a un potencial efectivo muy parecido al de éstos. La conclusión de este estudio es que los promotores del fago T7, y probablemente en cualquier otro organismo, no pueden ser diferenciados del resto del genoma estudiando la dinámica de *kinks*.

Comparación con ADN real.

La conclusión anterior es probablemente generalizable a otros elementos funcionales de la secuencia de ADN, no sólo a promotores. Un estudio de la literatura que trata la transcripción del ADN, así como del reconocimiento de secuencias específicas por parte de algunas proteínas (Alberts et al. 2002), es suficiente para entender que la complejidad del problema no se pueda explicar en términos de este sencillo modelo.

Otra limitación muy importante que ha sido tenida en cuenta en esta tesis es el efecto real que produce la viscosidad del agua en el ADN (Nelson 2004). En el rango de fuerzas en el que se mueve el ADN, del orden de los piconewtons, la fricción del agua sobre éste es tan grande que su dinámica deja de ser inercial y es prácticamente sólo disipativa. Sin embargo, los modelos aplicados al ADN mostrados a lo largo de esta memoria son inerciales, ya que, en sus ecuaciones

de movimiento, su término de rozamiento es cero o mucho menor que el término inercial. Cuando Englander et al. (1980) introdujeron el modelo de sG aplicado al ADN, lo hicieron para describir ciertos aspectos de la dinámica de éste, al igual que hizo Salerno (1991) cuando estudió los promotores en secuencias inhomogéneas. Sin embargo, el caso de Salerno, así como otras investigaciones posteriores (Domínguez-Adame et al. 1995; Lennholm y Hörnquist 2003; Bashford 2006) y las investigaciones realizadas sobre el modelo durante esta tesis (Cuenda y Sánchez 2004b; Cuenda y Sánchez 2004a; Cuenda et al. 2006), pueden entenderse de forma fenomenológica. De esta manera, aunque realmente no puede haber *kinks* o *breathers* que se propagan por la cadena de la forma descrita, el estudio de la relación entre la dinámica de estos solitones en secuencias genómicas y la existencia o no de promotores u otros elementos funcionales en dichas secuencias es igualmente válido.

El modelo de PB.

En el capítulo 7 se estudian las soluciones de tipo pared de dominio del modelo de PB (Peyrard y Bishop 1989). Estas soluciones pueden modelar el estado del ADN en los experimentos de desnaturalización mecánica, por lo que puede resultar útil estudiar este tipo de soluciones para explicar los resultados experimentales (Cuesta-López et al. 2005). Esto, unido a los resultados obtenidos con el potencial efectivo de *kinks* en el modelo de sG, nos llevaron a estudiar estas soluciones para intentar construir una herramienta análoga en este modelo. Para ello realizamos un análisis de las soluciones estáticas en este modelo en sistemas finitos con condiciones de contorno fijas (que son las representativas en los experimentos de desnaturalización mecánica), y estudiamos su estabilidad. En este trabajo encontramos expresiones analíticas aproximadas de las soluciones, tanto en el límite continuo como en el límite discreto, aunque las soluciones relevantes son estas últimas. Para finalizar, comparamos con un trabajo reciente en el que se aborda un caso particular del problema de forma numérica, y obtenemos el mismo número de soluciones estables e inestables. Las conclusiones de esta investigación son que, por un lado, hemos obtenido expresiones analíticas aproximadas, válidas para sistemas finitos, que no habían sido obtenidas anteriormente, y que describen correctamente las soluciones de un sistema con condiciones de contorno fijas, como vimos al comparar nuestros resultados con otros obtenidos por otro grupo de forma diferente.

Este trabajo es el primer paso en la búsqueda de un potencial efectivo para este tipo de soluciones que pueda explicar las observaciones experimentales de desnaturalización mecánica (Essevaz-Roulet et al. 1997). Un problema que hay

que resolver es que, para aplicar la técnica de coordenadas colectivas, la expresión analítica aproximada que tenemos para estas soluciones, en términos de la apertura del segundo elemento de la cadena, hay que pasarla a otra variable que represente el centro de la pared de dominio. Otra barrera difícil de superar en el modelo es la dependencia lineal de la energía con el tamaño del sistema. Esto hace, por un lado, que el sistema, con condiciones de contorno libres, tienda a estar cerrado. Por otro, a la hora de calcular un posible potencial efectivo para estas soluciones en el modelo, el término de masa depende de la posición (lo cual es consistente con el problema anterior). Así, aunque la aplicación de la fuerza externa en el último elemento puede prevenir el cierre de la misma, no está claro cómo se puede superar la dependencia de la masa con la posición.

Además de lo anteriormente mencionado, es necesario tener en cuenta el efecto disipativo del agua si queremos realizar un estudio realista de este tipo de experimentos con este modelo. Las soluciones estáticas encontradas son igualmente válidas en el caso disipativo y, aunque es necesario realizar un estudio sobre cómo esta disipación afecta a la dinámica de estas soluciones de tipo pared de dominio, es posible que el efecto de la disipación y el ruido térmico explique el mecanismo de *stick-slip* que se produce en los experimentos de desnaturalización mecánica. Sin embargo, es necesario un estudio detallado del problema para obtener resultados concluyentes.

V

Apéndices

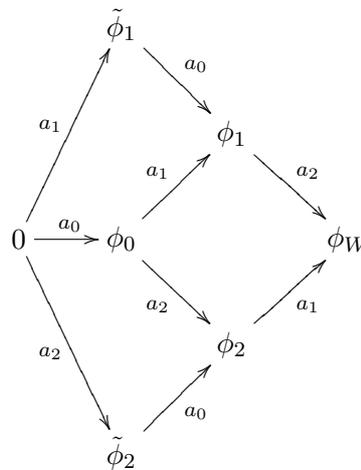
A

Wobbles en el modelo de sG.

La TEI fue aplicada a la ecuación de sG (2.1) por primera vez por Ablowitz et al. (1973) para la búsqueda de soluciones exactas de la misma, y las expresiones principales de tipo *kink* y *breather* fueron encontradas con esta técnica. En 1983, Harvey Segur consiguió la expresión de una solución que él llamó *wobbling kink*: un objeto que satisface las condiciones de contorno del *kink*, pero con un grado de libertad interno. El propio Segur advirtió en su trabajo que estos objetos debían ser probablemente inestables, aunque con una inestabilidad que crece de manera algebraica en lugar de exponencial. Recientemente han aparecido estudios que vuelven a recuperar la expresión de este tipo de soluciones (Kalbermann 2006), intentando hacer un análisis numérico de la estabilidad del objeto. En estos estudios, en los que utilizan la TEI como medio para encontrar la expresión de la solución, consideran el objeto así obtenido como un *kink* con una oscilación, modo interno o “quasimodo”. Dicho modo interno fue sugerido por Rice (1983) e introducido por Boesch y Willis (1990). Debido a las discrepancias que este trabajo ha suscitado recientemente (Quintero et al. 2000; Willis 2006; Quintero et al. 2006) sobre la existencia o no de estos quasimodos, en este apéndice se recoge la evidencia de que, en el caso descrito por Segur y Kälbermann referente a los *wobbles*, esta solución no se trata de un *kink* con un grado de libertad interno. Por el contrario, son dos *kinks* y un *kink* superpuestos de forma no lineal y de manera

que un par *kink-antikink* forma un *breather*. Este *breather* es el que puede confundirse con la oscilación interna si no se tiene el cuidado de estudiar el significado de los parámetros de la solución final. Para demostrarlo ha bastado utilizar la TB, conocida desde finales del siglo XIX (Bäcklund 1883), para así obtener una expresión donde todos los parámetros tienen un significado conocido en términos de posiciones y frecuencias.

Para encontrar la expresión del *wobble* como el referido por Kalbermann (2006) basta con construir la solución de tres *kinks* a partir del siguiente diagrama de Bianchi:



La única dificultad radica en conocer la forma que deben tener los coeficientes a_i , con $i = 0, 1, 2$, para obtener las soluciones parciales $\tilde{\phi}_1$, $\tilde{\phi}_2$, ϕ_0 , ϕ_1 y ϕ_2 con las que construir la solución final del *wobble* ϕ_W . Puesto que postulamos que el *wobble* no es más que un *kink* y un *breather*, los parámetros a_i que se deben usar en la integración son aquellos que se utilizan en la construcción de un *kink* y un *breather*. Nosotros consideraremos que ϕ_0 es un *kink* en reposo, que $\tilde{\phi}_1$ y $\tilde{\phi}_2$ son las soluciones parciales que dan lugar a un *breather* a partir del diagrama (o de la identidad) de Bianchi apropiada; ϕ_1 y ϕ_2 son las soluciones que se obtienen como superposición no lineal de ϕ_0 y $\tilde{\phi}_1$, y ϕ_0 y $\tilde{\phi}_2$, respectivamente, y con las que, finalmente, se construye la solución de tipo *wobble* ϕ_W .

A continuación se muestran las transformaciones básicas que son necesarias para obtener las soluciones parciales ϕ_0 , $\tilde{\phi}_1$ y $\tilde{\phi}_2$ y así conocer el valor de los parámetros a_0 , a_1 y a_2 necesarios para construirlas.

A.1. Integración de TB a partir de $\phi = 0$.

En este caso no se puede utilizar el diagrama de Bianchi, sino que es necesario integrar directamente la TB (2.2) a partir de la solución trivial $\phi = 0$ utilizando el parámetro a . La ecuación que se obtiene es, como ya se vio en el primer capítulo,

$$\begin{aligned}\left(\frac{\phi}{2}\right)_{\xi} &= a \sin\left(\frac{\phi}{2}\right), \\ \left(\frac{\phi}{2}\right)_{\tau} &= \frac{1}{a} \sin\left(\frac{\phi}{2}\right).\end{aligned}$$

Multiplicando la ecuación de arriba por $d\xi$, la de abajo por $d\tau$ y sumando ambas ecuaciones se obtiene

$$\left(\frac{\phi}{2}\right)_{\xi} d\xi + \left(\frac{\phi}{2}\right)_{\tau} d\tau = \sin\frac{\phi}{2} \left(ad\xi + \frac{d\tau}{a}\right), \quad (\text{A.1})$$

donde el término de la izquierda de la expresión anterior es la diferencial exacta de $\phi/2$. Se puede, por tanto, escribir la expresión (A.1) como

$$\frac{d\frac{\phi}{2}}{\sin\frac{\phi}{2}} = ad\xi + \frac{d\tau}{a}, \quad (\text{A.2})$$

que al integrarse da la expresión

$$\tan\left(\frac{\phi}{4}\right) = \exp\left[a\xi + \frac{\tau}{a} + \delta\right], \quad (\text{A.3})$$

o en coordenadas estándar,

$$\tan\left(\frac{\phi}{4}\right) = \exp\left[\left(a + \frac{1}{a}\right)x + \left(a - \frac{1}{a}\right)t + \delta\right], \quad (\text{A.4})$$

donde δ es una constante de integración asociada al origen de coordenadas en x y en t .

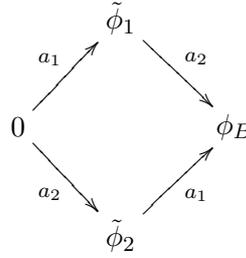
La expresión (A.4) la utilizaremos en las expresiones de ϕ_0 , $\tilde{\phi}_1$ y $\tilde{\phi}_2$. Por ejemplo, en el caso de ϕ_0 estamos interesados en que sea un *kink* en reposo de la forma

$$\tan\left(\frac{\phi_0}{4}\right) = e^{x-x_k}, \quad (\text{A.5})$$

por lo que, para ello, el parámetro de la transformación debe ser $a_0 = 1$, y la constante $\delta = -x_k$, real, nos da la posición del *kink*.

A.2. Construcción de un *breather*.

Tal y como se ha mencionado, $\tilde{\phi}_1$ y $\tilde{\phi}_2$ deben ser soluciones de la forma (A.4) definidas en este caso con el parámetro a_i y la constante δ_i para $i = 1, 2$. Estos a_i y δ_i deben ser tales que la combinación no lineal de ambas expresiones de lugar a un *breather*, ϕ_B , de acuerdo al diagrama de Bianchi



La expresión del *breather* viene dada por la identidad de Bianchi (2.3), que en este caso particular es

$$\tan\left(\frac{\phi_B}{4}\right) = \frac{a_1 + a_2}{a_1 - a_2} \tan\left(\frac{\tilde{\phi}_1 - \tilde{\phi}_2}{4}\right). \quad (\text{A.6})$$

Utilizando la expresión de la tangente de la suma, y operando con las exponenciales de $\tilde{\phi}_i$, se llega a

$$\tan\left(\frac{\phi_B}{4}\right) = \frac{a_1 + a_2}{a_1 - a_2} \frac{\text{sh}\left[\left(a_1^+ - a_2^+\right)\frac{x}{4} + \left(a_1^- - a_2^-\right)\frac{t}{4} + \frac{\delta_1}{2} - \frac{\delta_2}{2}\right]}{\text{ch}\left[\left(a_1^+ + a_2^+\right)\frac{x}{4} + \left(a_1^- + a_2^-\right)\frac{t}{4} + \frac{\delta_1}{2} + \frac{\delta_2}{2}\right]} \quad (\text{A.7})$$

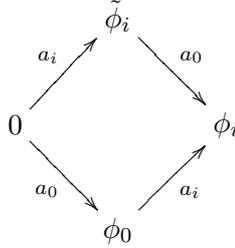
donde, para una escritura más reducida, se ha definido $a_i^+ = a_i + 1/a_i$, y $a_i^- = a_i - 1/a_i$. Para que esta expresión se corresponda con la expresión de un *breather* en reposo, (2.30), con una cierta traslación espacial y temporal, es necesario que los parámetros de la transformación verifiquen $a_1 = \beta + i\alpha$, con β y α reales tales que $\beta^2 + \alpha^2 = 1$, y que $a_2 = a_1^*$ (el asterisco significa complejo conjugado). Las constantes de integración también están relacionadas, de la forma $\delta_1 = \beta x_b + i\alpha t_b$, con x_b y t_b reales, y con $\delta_2 = \delta_1^*$. De esta manera, la expresión final del *breather* es

$$\tan\left(\frac{\phi_B}{4}\right) = \frac{\beta \sin(\alpha(t - t_b))}{\alpha \text{ch}(\beta(x - x_b))}, \quad (\text{A.8})$$

y quedan definidos todos los parámetros que son necesarios para construir el *wobble*.

A.3. Superposición no lineal de un *kink* y medio *breather*.

Esta superposición se refiere a la penúltima que es necesario hacer para conseguir el resultado parcial ϕ_i , con $i = 1, 2$, a partir de las cuales ya se construye directamente el *wobble* con la última superposición no lineal. La identidad de Bianchi para conseguir ϕ_i es



El título de esta sección se debe a que, en ella, se calcula la superposición de un *kink* y sólo uno de los dos $\tilde{\phi}_i$ necesarios para construir el *breather* completo. Las expresiones que se necesita usar para aplicar la identidad de Bianchi,

$$\tan\left(\frac{\phi_i}{4}\right) = \frac{a_0 + a_i}{a_0 - a_i} \tan\left(\frac{\phi_0 - \tilde{\phi}_i}{4}\right), \quad (\text{A.9})$$

vienen dadas por las ecuaciones (A.5) para ϕ_0 , con $a_0 = 1$, y (A.4) para $\tilde{\phi}_i$, con $a_1 = \beta + i\alpha$ ($\beta^2 + \alpha^2 = 1$), $a_2 = a_1^*$, $\delta_1 = \beta x_b + i\alpha t_b$ y $\delta_2 = \delta_1^*$. Una forma de escribir de forma reducida las tres expresiones es

$$\tan\left(\frac{\phi_0}{4}\right) = e^{z_0}, \quad (\text{A.10})$$

$$\tan\left(\frac{\tilde{\phi}_1}{4}\right) = e^{z_1 + iz_2}, \quad (\text{A.11})$$

$$\tan\left(\frac{\tilde{\phi}_2}{4}\right) = e^{z_1 - iz_2}, \quad (\text{A.12})$$

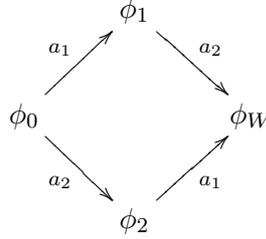
con $z_0 = x - x_k$, $z_1 = \beta(x - x_b)$ y $z_2 = \alpha(t - t_b)$. Con esta notación se obtienen las expresiones de ϕ_i , que son

$$\tan\left(\frac{\phi_1}{4}\right) = \frac{i\alpha}{1 - \beta} \frac{e^{z_0} - e^{z_1 + iz_2}}{1 + e^{z_0 + z_1 + iz_2}}, \quad (\text{A.13})$$

y $\phi_2 = \phi_1^*$.

A.4. Obtención del *wobble*.

Para obtener el *wobble* ϕ_W es necesario superponer las tres soluciones solitónicas, ϕ_0 , ϕ_1 y ϕ_2 , utilizando el siguiente diagrama de Bianchi:



que equivale a la siguiente expresión,

$$\tan\left(\frac{\phi_W - \phi_0}{4}\right) = \frac{a_1 + a_2}{a_1 - a_2} \tan\frac{\phi_1 - \phi_2}{4}. \quad (\text{A.14})$$

Operando con la expresión (A.13) y teniendo en cuenta que $\phi_2 = \phi_1^*$ y $a_1 = a_2^* = \beta + i\alpha$, se llega a la siguiente expresión,

$$\begin{aligned} \frac{\phi_W - \phi_0}{4} = \arg & \left[(1 - \beta) (1 + e^{2z_0 + 2z_1} + 2e^{z_0 + z_1} \cos z_2) \right. \\ & + (1 + \beta) (e^{2z_0} + e^{2z_1} - 2e^{z_0 + z_1} \cos z_2) \\ & \left. + i2\beta (e^{z_0} - e^{z_0 + 2z_1} + e^{2z_0 + z_1} \cos z_2 - e^{z_1} \cos z_2) \right], \quad (\text{A.15}) \end{aligned}$$

donde hemos usado que

$$\arctan\left(\frac{y}{x}\right) = \arg(x + iy), \quad (\text{A.16})$$

para $x > 0$. Utilizando la relación (A.16) también es posible escribir ϕ_0 , definido por (A.10), como

$$\frac{\phi_0}{4} = \arg(1 + ie^{z_0}). \quad (\text{A.17})$$

Esta expresión nos permite sumar los argumentos (A.15) y (A.17) sin más que multiplicar los números complejos correspondientes, dando lugar a

$$\begin{aligned} \frac{\phi_W}{4} = \arg & \left[(1 - \beta)(1 + e^{2z_0}) + (1 + \beta)(1 + e^{2z_0})e^{2z_1} \right. \\ & - 2\beta(1 + e^{2z_0})e^{z_0 + z_1} \cos z_2 \\ & + i(1 + \beta)(1 + e^{2z_0})e^{z_0} + i(1 - \beta)(1 + e^{2z_0})e^{z_0 + 2z_1} \\ & \left. - i2\beta(1 + e^{2z_0})e^{z_1} \cos z_2 \right]. \quad (\text{A.18}) \end{aligned}$$

Esta expresión es muy farragosa, pero si recordamos que $z_0 = x - x_k$, $z_1 = \beta(x - x_b)$ y $z_2 = \alpha(t - t_b)$, así como el significado de x_k como posición del centro del *kink*, x_b la posición del centro del *breather*, t_b el origen temporal en la oscilación del *breather* y α la frecuencia del oscilación del *breather*, uno puede imaginar el aspecto que tendrá este objeto para unos valores dados de estos parámetros.

Comparemos ahora la expresión (A.18) y la obtenida por (Kalbermann 2006),

$$\frac{\phi_W}{4} = \arg(U + iV), \quad (\text{A.19})$$

con

$$U = 1 + \frac{|m_2|^2 \left(\frac{1}{4} - \omega^2\right)}{\omega^2} e^{4\omega x} - \frac{2m_1|m_2| \left(\frac{1}{2} - \omega\right)}{\frac{1}{2} + \omega} e^{x+2\omega x} \cos(2\sigma(t + t_0)), \quad (\text{A.20})$$

$$V = -\frac{m_1|m_2|^2 \left(\frac{1}{2} - \omega\right)^3}{\omega^2 \left(\frac{1}{2} + \omega\right)} e^{x+4\omega x} - m_1 e^x + 2|m_2| e^{2\omega x} \cos(2\sigma(t + t_0)), \quad (\text{A.21})$$

y donde $\omega^2 + \sigma^2 = 1/4$, $m_1 \in \mathbb{R}$ y $m_2 \in \mathbb{C}$, y t_0 se define mediante la expresión

$$\tan(2\sigma t_0) = \frac{\text{Im}(m_2)\omega + \text{Re}(m_2)\sigma}{\text{Im}(m_2)\sigma - \text{Re}(m_2)\omega} \quad (\text{A.22})$$

A primera vista, ambas expresiones son totalmente diferentes. Sin embargo, se puede obtener (A.19) a partir de (A.18) utilizando, en primer lugar, la propiedad $\arg[a(x + iy)] = \arg(x + iy)$, para $a > 0$. De esta forma, sacando factor común $(1 - \beta)(1 + e^{2z_0})$ (válido para $0 < \beta < 1$) del argumento de la derecha de (A.18), dicha ecuación se puede escribir como

$$\frac{\phi_W}{4} = \arg \left[1 + \frac{1 + \beta}{1 - \beta} e^{2z_1} - \frac{2\beta}{1 - \beta} e^{z_0+z_1} \cos z_2 + i \frac{1 + \beta}{1 - \beta} e^{z_0} + i e^{z_0+2z_1} - i \frac{2\beta}{1 - \beta} e^{z_1} \cos z_2 \right]. \quad (\text{A.23})$$

Ahora sólo queda encontrar la relación entre los parámetros de Kalbermann (2006) y los nuestros. Dicha relación se obtiene identificando término a término los exponentes de las dos expresiones, y el resultado es el siguiente:

$$\begin{aligned} m_1 &= -\frac{1 + \beta}{1 - \beta} e^{-x_k} & t_0 &= -t_b \\ & & \sigma &= 2\alpha \\ |m_2| &= -\frac{\beta}{1 - \beta} e^{-\beta x_b} & \omega &= 2\beta \end{aligned} \quad (\text{A.24})$$

Nótese que en la expresión anterior se iguala el módulo de m_2 a un valor negativo, lo que carece de sentido. Sin embargo, se puede entender que $|m_2|$ es un parámetro real más al que llamaremos m'_2 , con $m'_2 < 0$, definido a partir de m_2 como $m'_2 = -\sqrt{(\text{Re}(m_2))^2 + (\text{Im}(m_2))^2}$ para que no exista este conflicto. De esta manera, puesto que las dos expresiones son equivalentes y se conoce el significado de la solución ϕ_W en términos de la posición de un *kink* y un *breather*, la naturaleza del *wobble* queda clara. No se trata de un *kink* con un modo interno que oscila, sino tan sólo de una superposición no lineal de un *kink* y un *breather*. Este resultado confirma una vez más la inexistencia de modos internos en sG (Quintero et al. 2000).

B

Nociones de biología y biofísica.

B.1. Estructura del ADN¹.

Composición del ADN y estructura básica.

El ADN es una macromolécula muy compleja con una estructura muy rica que le permite tener la funcionalidad que tiene. Sin entrar mucho en detalles, a continuación se describe su estructura. El ADN es una macromolécula formada a su vez por dos cadenas de polímeros complementarias, unidas entre sí; cada una de esas cadenas de polímeros se denominarán *cadena* de ADN, y la unión de ambas, *doble cadena* de ADN, *doble hélice* o simplemente ADN. Los monómeros que conforman estas cadenas de ADN son los *nucleótidos*, que están compuestos a su vez por tres moléculas, que son: un grupo fosfato, unido a un azúcar (en el caso del ADN es la desoxirribosa), unido a su vez a una base, que puede ser de cuatro tipos diferentes: adenina (A), citosina (C), guanina (G) y timina (T); estos símbolos (A, C, G, T) también se usan para referirse a los diferentes nucleótidos

¹La mayor parte de la información contenida en esta sección me fue impartida en la escuela de verano de Les Houches, *Múltiples aspectos del ADN y el ARN: de la biofísica a la bioinformática*, de la que existe un libro impreso, realizado por los propios profesores de la escuela (Duke et al. 2004)

formados con estas bases. Los nucleótidos de una cadena de ADN están unidos entre sí a través del azúcar de un nucleótido y el grupo fosfato de otro por enlaces covalentes, formando la “columna vertebral” de la cadena de ADN. Esta unión no es simétrica, sino que existe una orientación en la cadena de ADN; al extremo donde está el grupo fosfato se le llama 5' y al extremo del azúcar se le llama 3'. La unión entre dos cadenas de ADN para formar la doble hélice se realiza emparejando cada una de las bases de una cadena con una base de la otra mediante *enlaces* o *puentes de hidrógeno*. Para que esto sea posible es necesario que las dos cadenas de ADN sean *antiparalelas* (esto es, que el extremo 5' de una esté junto al extremo 3' de la otra y viceversa) y *complementarias*. Que sean complementarias significa que las bases que se tienen que unir son complementarias, es decir, que los únicos enlaces posibles² entre bases son A con T y C con G. Es precisamente la secuencia específica de bases de la doble cadena de ADN la que contiene la información genética de la célula. Cada base A, T, C ó G puede considerarse como una letra en un alfabeto de cuatro letras que codifica los diferentes mensajes (proteínas) que necesita un organismo para vivir. La unión mediante los pares de bases de las dos cadenas de ADN hacen que, en primera aproximación, se pueda visualizar la doble cadena de ADN como una escalera cuyos lados serían las cadenas de azúcar-fosfato y cuyos peldaños serían los pares de bases, como se ve esquematizado en el dibujo izquierdo de la figura B.1. Pero realmente, como propusieron Watson y Crick (1953), la estructura de la doble cadena de ADN es una doble hélice, tal y como aparece representada en la figura B.1, derecha, donde la información genética está “apilada” como bloques de ladrillos unos encima de otros, y además se encuentra protegida por las dos hélices de azúcar-fosfato. Como puede observarse en el dibujo, la doble hélice no es simétrica, sino que la separación en alturas entre una hélice y otra da lugar a dos surcos de tamaño diferente, llamados surco menor y surco mayor, muy importantes en la funcionalidad de la molécula.

Estructura primaria, secundaria y terciaria.

En biología, al referirse a la estructura de macromoléculas definidas por una secuencia de monómeros, tales como proteínas, moléculas de ARN o la doble cadena de ADN, se habla de ellas en diferentes niveles de complejidad, que son la estructura primaria, la secundaria y la terciaria. En la estructura primaria se defi-

²Realmente, hay más enlaces posibles que hacen del ADN un objeto muy rico, versátil y complicado. Aquí solamente voy a referirme a las más conocidas, que son las llamadas uniones Watson-Crick, llamadas así porque son las que propusieron Watson y Crick (1953) cuando descubrieron la estructura básica del ADN.

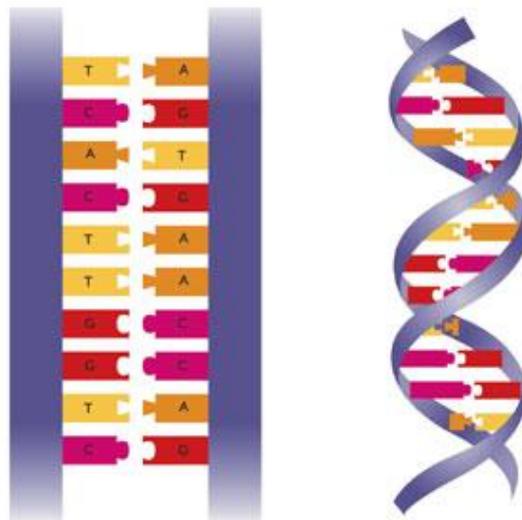


Figura B.1: A la izquierda, doble cadena de ADN representada como una escalera. A la derecha, doble cadena de ADN representada como doble hélice.

ne la secuencia ordenada concreta de la macromolécula. En el caso de proteínas, la secuencia de aminoácidos que la forman; en el caso del ARN y del ADN, la secuencia de bases, con la diferencia de que el ARN sólo está formado por una cadena y el ADN por dos. En la estructura secundaria se tienen en cuenta, además, las uniones físicas de monómeros no consecutivos de la cadena, dando lugar a plegamientos en la macromolécula. En las proteínas, la unión de aminoácidos; en ARN, la formación de horquillas por bases complementarias y de otras uniones entre nucleótidos alejados³; en ADN, la unión de las dos cadenas de ADN para formar la doble cadena. Por último, en la estructura terciaria se tiene en cuenta la configuración tridimensional final de la molécula, dados por los enlaces de la estructura secundaria y la polaridad de la molécula. Esta estructura es la más complicada de predecir, y hay muchos grupos en biofísica que tratan el plegamiento de proteínas y de ARN; el ADN, aparte del enrollamiento en forma de doble hélice, que se puede entender como una primera aproximación a su estructura, puede

³En ARN no existe timina (T), sino uracilo (U), que es la base complementaria a la adenina (A). Esta nueva base hace que el ARN sea un objeto funcionalmente más complejo que el ADN, puesto que le permite una gran variedad de uniones de los nucleótidos entre sí, aparte de la unión de Watson-Crick típica del ADN, dándole una gran variedad de formas y funciones dentro de la célula.

doblarse ligeramente en algunos puntos de la secuencia (fenómeno conocido como *bending*), y también resulta muy importante la interacción de los surcos de la doble hélice con proteínas que actúan sobre ella, como la ARN polimerasa o factores de transcripción, para el reconocimiento de lugares funcionales.

En el caso de organismos eucariotas, además, existen ordenamientos superiores que permiten el empaquetamiento de la información genética en configuraciones que ocupen menos volumen. Así, la doble hélice de ADN se encuentra “enrollada” alrededor de un octámero de unas proteínas denominados *histonas*, dando lugar a los *nucleosomas*. Estos nucleosomas, a su vez, se apilan y se enrollan entre sí en una hélice para formar la *cromatina*. Por último, en un enrollamiento más de la cromatina, se forman los *cromosomas*, compuestos por una (en células *haploides*) o dos (en células *diploides*) *cromátidas*, que contienen distintos *alelos* de los mismos genes. Realmente, los cromosomas sólo se hacen visibles en la célula en los procesos de *mitosis* (división celular en la que de una célula se obtienen dos células hijas, genéticamente idénticas a la madre) y *meiosis* (división celular en la que una célula diploide da lugar a dos haploides, con el mismo número de cromosomas pero cada uno de ellos formado por una única cromátida). El estado normal del ADN celular, en el que se producen todos los procesos de transcripción y regulación génica, es en forma de cromatina. Por tanto, la maquinaria necesaria para controlar estos procesos en células eucariotas es mucho más compleja que en células procariotas.

B.2. Transcripción del ADN⁴.

La ARN polimerasa.

La producción de las proteínas necesarias para el funcionamiento de la célula se realiza mediante el proceso de transcripción de una porción de ADN que contiene uno o varios genes en ARN, el posterior procesamiento del ARN para eliminar la información irrelevante a la hora de construir la proteína y la posterior traducción del ARN maduro en una secuencia de aminoácidos que se pliega para formar una proteína. Éste es el llamado dogma central de la biología molecular. En esta sección nos ocuparemos de describir el mecanismo del proceso de transcripción de una porción de ADN por la ARN polimerasa. Cada elemento transcrito de ADN se llama *unidad de transcripción*. En organismos eucariotas, una unidad de transcripción normalmente contiene la información relativa a un gen; en bacterias, por el contrario, se suele traducir un conjunto de genes adyacentes en una única

⁴Información obtenida casi íntegramente del gran libro de biología molecular de Alberts et al. (2002).

unidad de transcripción. De hecho, el mecanismo de transcripción en eucariotas y en procariotas es muy diferente: mientras que en bacterias sólo existe un tipo de ARN polimerasa que se encarga del proceso de transcripción, en el núcleo de las eucariotas existen tres tipos de ARN polimerasa (llamadas ARN polimerasa I, II y III), que transcriben diferentes tipos de secuencias de ADN. Las estructuras de todas ellas (incluida la bacteriana) son muy parecidas entre sí, pero transcriben diferentes tipos de genes. La diferencia fundamental entre las ARN polimerasas eucariotas y la ARN polimerasa bacteriana es que, mientras que la ARN polimerasa bacteriana (unida a una proteína denominada *factor* σ) es capaz de iniciar la transcripción de un fragmento de ADN *in vitro* sin la ayuda de proteínas adicionales, las ARN polimerasas eucariotas no pueden. Además, debido a la compleja estructura del ADN eucariota, la transcripción en estos organismos debe tener en cuenta el enrollamiento del ADN en los nucleosomas y el enrollamiento de éste para formar la cromatina, problemas totalmente ausentes en el caso de los cromosomas procariotas. En esta memoria de tesis bastará con que nos ocupemos del caso de la transcripción del ADN por la ARN polimerasa bacteriana, un mecanismo lo suficientemente complicado como para intentar describirlo matemáticamente con un modelo sencillo.

La ARN polimerasa bacteriana (en adelante, ARNP) es un complejo proteínico formado por varias subunidades. Una de ellas, el factor σ , es una subunidad capaz de acoplarse y desacoplarse de la ARNP que puede leer las señales en el ADN que le dicen dónde empezar a transcribir. Por lo general, la ARNP se adhiere de forma muy débil al ADN bacteriano cuando choca con él, de manera que se desliza rápidamente por el ADN antes de separarse de nuevo de él. Sin embargo, cuando la ARNP (junto con el factor σ) se desliza hasta una región del ADN donde se encuentra un *promotor*, se une firmemente a él. La ARNP utiliza el factor σ para reconocer esta secuencia de ADN mediante contactos específicos con las bases a través de los surcos de la doble hélice.

Después de unirse fuertemente al promotor de ADN, la ARNP abre la doble hélice para exponer una secuencia corta de nucleótidos en cada hebra (a este proceso se le llama *melting*, o fusión) para formar la burbuja de transcripción. Para ello no requiere energía química en forma de ATP, sino que tanto la molécula de ADN como la ARNP sufren una serie de cambios estructurales reversibles que llevan a un estado energéticamente más estable. Una vez el ADN está abierto, la ARNP utiliza una de las dos hebras de ADN como plantilla para crear la cadena de ARN complementaria, y empieza a transcribir en el “sitio de comienzo de la transcripción” (TSS) del promotor. Esta síntesis de ARN (a veces llamada *etapa abortiva*) es relativamente ineficiente. Sin embargo, una vez que la ARNP ha

sintetizado una cadena de unos diez ribonucleótidos, el factor σ se desacopla del complejo y la ARNP sufre una serie de cambios conformacionales para pasar a la *etapa de elongación*, en la que la ARNP se mueve rápidamente por la cadena (a una velocidad de unos 50 nucleótidos por segundo) mientras sintetiza el ARN. Este proceso continúa hasta que la ARNP encuentra una segunda señal en el ADN, el *terminador*, momento en el cual la ARNP libera tanto la cadena de ADN como la recién sintetizada de ARN. Una vez terminado el proceso de transcripción, esta ARNP puede unirse de nuevo a un factor σ para iniciar otra transcripción de ARN.

Promotores y terminadores.

Como hemos visto, el proceso de inicio y fin de la transcripción están regulados por secuencias específicas de nucleótidos que los puntos en los que la ARNP debe empezar y terminar la transcripción. No es de extrañar que los promotores y terminadores sean difíciles de reconocer por los investigadores, ya que esos estadios de la transcripción conllevan una serie de cambios conformacionales en las moléculas de ADN, ARN y en la ARNP. De hecho, una comparación de diferentes promotores bacterianos revela que son heterogéneos en la secuencia de ADN⁵. Sin embargo, todos ellos contienen secuencias relacionadas que reflejan en parte aspectos del ADN que son reconocidos directamente por el factor σ . Estas características comunes a menudo se resumen en forma de *secuencias consenso*. En general, una secuencia de nucleótidos consenso se deriva comparando muchas secuencias con la misma función básica y eligiendo el nucleótido más común que se encuentra en cada posición. Sirve, por tanto, como resumen o promedio de un gran número de secuencias de nucleótidos individuales (véase la figura B.2).

Una razón por la que los promotores bacterianos poseen diferentes secuencias de manera individual es que la secuencia particular de cada promotor determina su *fuerza* (el número de eventos de iniciación por unidad de tiempo). Por tanto, la evolución ha refinado cada promotor para que sea iniciado con la frecuencia necesaria, creando para ello un amplio espectro de promotores. De esta manera, los promotores asociados a genes que codifican una proteína muy abundante son mucho más fuertes que aquellos asociados a genes que codifican proteínas escasas, y su secuencia de nucleótidos es responsable de esa diferencia.

Otra característica de los promotores es que son asimétricos, como se observa en la figura B.2. En principio, puesto que el ADN está formado por dos cadenas, dos ARNPs podrían transcribir dos cadenas diferentes de ARN a partir de un mismo promotor. Sin embargo, la asimetría de éstos, unida al hecho de que la ARNP

⁵Esta sencilla afirmación parece echar por tierra cualquier modelo matemático que intente explicar la búsqueda de promotores en los términos explicados en los capítulos 5, 6 y 7.

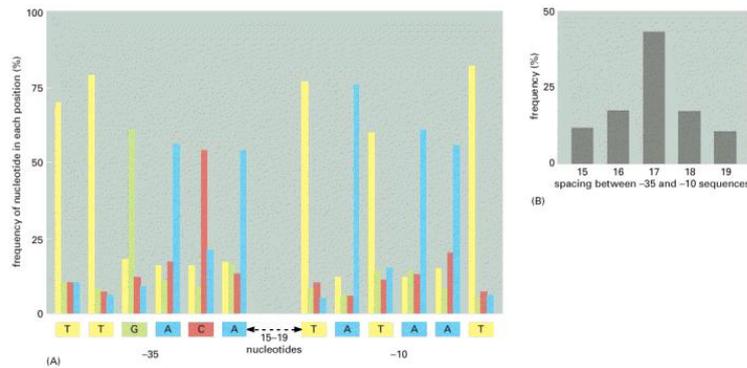


Figura B.2: Secuencia consenso para el promotor de clase mayor de la *E. coli*. (a) Los promotores se caracterizan por dos secuencias hexámeras (formadas por seis elementos) de ADN, la -35 y la -10 , llamadas así por su localización aproximada relativa al TSS (designado por $+1$). Por conveniencia, se muestra la secuencia de nucleótidos de una hebra de ADN, aunque en realidad la ARNP reconoce el promotor como una doble cadena de ADN. Tras estudiar 300 promotores, se muestran las frecuencias de los cuatro nucleótidos en cada posición de los hexámeros. La secuencia consenso se muestra bajo el gráfico. Las distintas secuencias de nucleótidos de los promotores estudiados no muestran parecidos significativos entre los hexámeros -35 y -10 . (b) Distribución del espaciado entre los hexámeros -35 y -10 encontrados en los promotores de la *E. coli*. La información de estas dos gráficas se aplica sólo a los promotores de la *E. coli* que son reconocidos por la ARNP unida al factor σ mayor (llamado σ^{70}). Sin embargo, las bacterias también poseen factores σ menores, cada uno de los cuales reconoce un promotor con secuencia diferente. Algunos promotores particularmente fuertes, reconocidos por la ARNP y el factor σ^{70} tienen una secuencia adicional, localizada más a la izquierda (si suponemos que la ARNP transcribe de izquierda a derecha) del hexámero -35 , y que es reconocida por otra subunidad de la ARNP, denominada α . (Figura obtenida de (Alberts et al. 2002).)

sólo puede transcribir en la dirección $5' - 3'$, hace que la ARNP sólo se pueda unir de manera que el proceso de transcripción sólo utilice una de las cadenas de ADN como plantilla. La elección de la plantilla para cada gen, por tanto, queda definida por la localización de un promotor y su orientación. Las secuencias genómicas revelan que la cadena de ADN utilizada como plantilla varía de un gen a otro.

Como ya hemos visto, los terminadores son secuencias de ADN que indican a la ARNP el final del proceso de elongación de la cadena de ARN sintetizada. La mayoría de los terminadores bacterianos consisten en una secuencia de pares $A - T$ precedida por otras dos secuencias, complementarias entre sí, de manera que, cuando el terminador es transcrito en ARN, se dobla sobre sí misma como una horquilla para unir las bases complementarias a la manera de la doble cadena de ADN. Al mismo tiempo, la doble cadena híbrida de ADN-ARN, formada predominantemente por pares $U - A$ (menos estables que los $C - G$ por tener dos enlaces de hidrógeno en lugar de tres), no es lo suficientemente fuerte para mantener el ARN en su sitio, y se disocia, causando la liberación de la ARNP del ADN.

A pesar de todo lo dicho anteriormente, aunque se conocen muchos datos de promotores y terminadores bacterianos que permiten desarrollar secuencias consenso que resumen las características más importantes de estos elementos, su variación en nucleótidos hace difícil a los investigadores (incluso con la ayuda de ordenadores) localizarlos mediante la simple inspección de la secuencia de nucleótidos de un genoma. A menudo es necesaria información adicional, proveniente a menudo de la experimentación directa, para localizar con precisión estas cortas señales de ADN contenidas en los genomas.

B.3. El genoma del fago T7⁶.

Regiones del genoma y función.

La secuencia completa del fago T7 se consiguió en uno de los primeros proyectos de secuenciación genómica durante los 80 (Dunn y Studier 1983), debido a la pequeña longitud de su genoma de 39937 nucleótidos. Desde entonces pocos cambios ha sufrido la secuencia, aunque se han encontrado nuevos genes. El genoma del fago T7 está compuesto por una doble cadena de ADN lineal que se caracteriza por un contenido de GC del 48 %, y tiene 60 genes que ocupan el 92 % de su genoma. Los genes de T7 se expresan ordenadamente en tres grupos: los

⁶Información obtenida en su mayor parte del artículo original de Dunn y Studier (1983), contrastada en la propia página del NCBI, y de otras publicaciones que se indican en el texto.

tempranos, o de clase I, son transcritos por la ARNP de la *E. coli*, y se encargan, en la primera etapa de infección, de superar las restricciones del huésped y de convertir el metabolismo de la célula huésped a la producción de proteínas de T7. Los genes de clase II son los siguientes en ser expresados, e incluyen funciones del metabolismo del ADN. Por último, son expresados los genes de clase III, con funciones involucradas en la maduración y empaquetamiento del ADN. A las regiones de clase I y II se las denomina tardías.

El ADN del T7 se transcribe enteramente de izquierda a derecha, primero por la RNAP de la *E. coli*, que transcribe la región temprana, y luego por la recién creada RNAP del T7, que transcribe una porción de la región temprana y toda la región tardía. Existen tres promotores mayores en la región temprana, A_1 , A_2 y A_3 , que se encuentran en la región que no codifica cerca del extremo izquierdo del ADN del fago. Además, existen varios promotores menores, incluido el promotor A_0 (también llamado D), orientado a la izquierda, que transcriben *in vitro* pero no tienen función conocida *in vivo*. El ARN producido en la transcripción de la región tardía se debe quince promotores de la T7 ARNP distribuidos a lo largo del 85 % de la porción derecha del ADN. De estos promotores, cinco de ellos son de clase III (es decir, dirigen la transcripción de los genes de clase III), y resultan ser los promotores más fuertes, tanto *in vivo* como *in vitro*; todos ellos tienen la misma secuencia exacta de 23 nucleótidos. Los diez promotores de la clase II (asociados a los genes de esta clase) son más débiles que los de la clase III, y se diferencian de éstos en tener de dos a siete nucleótidos diferentes de la secuencia conservada. Además, se cree que la existencia de una secuencia rica en A o T justo antes de la secuencia consenso puede influir en el grado de agarre de la ARNP del fago a la doble cadena de ADN; en este sentido, los promotores de la clase III tienen una secuencia continua en A o T en las posiciones de -17 a -22 , mientras que los de la clase II, más débiles, la interrumpen con una o más bases G o C. Existe un único promotor de la clase II con una secuencia ininterrumpida de A o T, desde -13 a -22 , y su fuerza es comparable a la de clase III (Tang et al. 2005).

Tres de estos promotores de clase II, realmente se encuentran en la región temprana (clase I), pero dirigen la transcripción dentro de la región de clase II. Además, todos los promotores de la clase II continúan dentro de la región de clase III.

Promotores y terminadores.

Por estar dirigidos a la RNAP de la *E. coli*, la estructura de los promotores tempranos del fago es de dos hexámeros, en -35 y -10 , para que sean reconocidos por el factor mayor σ^{70} de la ARNP bacteriana. Además, la ARNP de la *E.*

coli es capaz de reconocer, con la ayuda de las subunidades α , secuencias en la región de -35 a -60 , denominadas el elemento *UP*, y que consiste en una región de cerca de veinte nucleótidos en la que básicamente se alternan A y T (Gourse et al. 2000). Los promotores mayores de la región temprana del genoma del fago T7 poseen una alta concentración en nucleótidos A o T, mucho mayor que la de cualquier promotor menor de esa misma región.

La secuencia consenso de los promotores tardíos, por el contrario, consta de 23 nucleótidos, que van de -17 a $+6$, donde $+1$ es la posición del TSS, y siendo las posiciones de -4 a -1 las primeras que se funden en la burbuja de transcripción. De hecho, la estructura de la ARNP del T7 es bastante diferente. Aunque, al igual que otras ARNP, es capaz de reconocer una secuencia de un promotor específica, iniciar una nueva cadena de ARN, tener ciclos abortivos de síntesis, ser regulada por un inhibidor de la transcripción, y terminar la transcripción con un terminador, su principal característica consiste en conseguir realizar todo lo anterior estando formada por una única cadena polipeptídica (Cheetham et al. 1999). De esta manera, la ARNP del T7 consta de una única unidad, a diferencia de la ARNP bacteriana, formada por diferentes subunidades, como el factor σ o las subunidades α que vimos anteriormente.

La transcripción de la región temprana del ADN del T7 por la ARNP bacteriana termina en una señal que produce una horquilla, utilizada a su vez para desestabilizar el complejo de transcripción y hacer que el proceso de transcripción termine. La horquilla así formada consta de ocho pares de bases unidas entre sí y un bucle de cuatro bases para unir las dos cadenas de ARN. La transcripción de la ARNP del fago no termina ahí, sino que continúa en la región tardía, hasta que encuentra un terminador más fuerte, formado por una horquilla de catorce o quince pares de bases y un bucle de seis a ocho bases. Este terminador no es completamente eficiente, tanto *in vitro* como *in vivo*. De hecho, un terminador perfectamente eficiente sería letal para el fago, puesto que existen dos genes que se encuentran entre este terminador y el siguiente promotor, específicos de proteínas estructurales del T7, por lo que estos genes sólo son transcritos cuando una ARNP pasa de largo el terminador sin afectarle.

B.4. Algunas cosillas de biofísica⁷.

Los enlaces de hidrógeno.

Recordemos que, en la sección 6.2, supusimos que la energía de un enlace de hidrógeno, E_H , era aproximadamente la energía térmica a temperatura ambiente, $k_B T_a$. En este apartado se aportan datos del enlace de hidrógeno para entender dicha elección.

La molécula de agua está formada por dos átomos de hidrógeno unidos a uno de oxígeno. Los electrones no están repartidos uniformemente, sino que casi siempre se encuentran alrededor del átomo de oxígeno. Los átomos de hidrógeno, sin su electrón, son esencialmente un protón, con un tamaño mucho menor que el del hidrógeno neutro. Pero, además, las uniones de los dos hidrógenos al oxígeno forman un ángulo de 104° entre sí, lo que le da forma asimétrica. Esta forma asimétrica, unida a la distribución de su carga, hacen que el agua tienda a formar una red tetraédrica, donde los hidrógenos de una molécula apuntan hacia la parte más negativa de las moléculas adyacentes. Esta red le confiere al agua propiedades características, como su tensión superficial o su temperatura de evaporación, entre otras.

La unión entre un oxígeno de una molécula de agua y un hidrógeno de la molécula adyacente se realiza mediante un *enlace de hidrógeno*. La energía de este enlace en dos moléculas de agua aisladas (en vapor) es aproximadamente $8 k_B T_a$, donde T_a es la temperatura ambiente ($295 K$). Para comparar con otros enlaces químicos, la energía del enlace covalente sencillo (por ejemplo, $C - C$) es de $140 k_B T_a$, mientras que el enlace van der Waals es típicamente de 0.6 a $1.6 k_B T_a$.

Sin embargo, los enlaces de hidrógeno también ocurren entre moléculas que contienen hidrógenos unidos covalentemente a cualquier átomo electronegativo (oxígeno, nitrógeno o flúor). Por ejemplo, como hemos visto a lo largo de esta memoria de tesis, la doble cadena de ADN se mantiene unida mediante los enlaces de hidrógeno de los pares de bases. Dichos enlaces de hidrógeno se producen entre el hidrógeno unido covalentemente a un nitrógeno de una base y un oxígeno o nitrógeno de la base complementaria. Sin embargo, no se puede aplicar directamente la estimación de la energía del enlace de hidrógeno en moléculas de agua dada anteriormente. Cuando las dos bases complementarias de un par se separan en un medio acuoso, los enlaces de hidrógeno que los mantenían unidos se rompen pero, inmediatamente, formarán enlaces de hidrógeno nuevos con las moléculas de agua que las rodean, compensando en parte la pérdida anterior. De hecho, el

⁷Información obtenida del libro de Nelson (2004)

coste de energía neta necesario para romper un único enlace de hidrógeno en el agua es sólo de una o dos veces $k_B T_a$.

Viscosidad en la célula.

Cuando un objeto se mueve en un fluido, puede hacerlo en dos regímenes muy diferentes, dependiendo de la fuerza que el fluido ejerza sobre él. Esta fuerza se opone siempre al movimiento del objeto en el fluido, y su intensidad depende de la geometría del problema y del movimiento del objeto en el fluido, así como del propio fluido. Todo el efecto del fluido se incluye en una constante, η , que es su viscosidad. Cuando esta viscosidad es pequeña, el régimen de movimiento se dice que es *turbulento*, y en él los efectos inerciales son importantes (por ejemplo, al remover el café con una cucharilla, éste sigue dando vueltas aunque pare de removerlo); por el contrario, cuando la viscosidad es grande, el régimen de movimiento se dice que es *laminar*, y el efecto predominante es la disipación [remover un vaso con glicerina no siempre resulta como uno esperaría, como se explica de manera muy gráfica en el capítulo 5 de Nelson (2004)]. Pero, ¿qué se entiende por “grande” y “pequeño”? ¿Relativo a qué? La viscosidad no es adimensional, así que no hay una manera absoluta de decidirlo. La mayoría de los fluidos (los fluidos newtonianos isotrópicos) quedan completamente caracterizados por su viscosidad, η , y su densidad, ρ . Sin embargo, tampoco se puede construir ninguna cantidad adimensional combinando la viscosidad (que tiene dimensiones de $ML^{-1}T^{-1}$) con la densidad (de dimensiones ML^{-3}). Lo que sí se puede construir es una cantidad característica con las dimensiones de fuerza,

$$f_{\text{crit}} = \frac{\eta^2}{\rho}, \quad (\text{B.1})$$

la *fuerza crítica viscosa*. Esta fuerza nos dice en qué régimen de movimiento nos encontramos, dependiendo de la fuerza aplicada, f . Así, si el cociente f/f_{crit} es pequeño, significa que el efecto que produce la viscosidad sobre el objeto amortigua cualquier efecto inercial, y el flujo es laminar. Por el contrario, cuando f/f_{crit} es grande, los efectos inerciales dominarán sobre la fricción del fluido, y es de esperar que exista flujo turbulento.

La fuerza crítica viscosa asociada al sirope de caramelo es $f_{\text{crit}} = 0.03 \text{ N}$, y la del agua $f_{\text{crit}} = 8 \times 10^{-10} \text{ N}$. Esto significa que, en nuestra escala macroscópica de fuerzas, si removemos un tarro lleno de sirope con una fuerza mucho menor que 0.3 N , el movimiento estará dominado por la fricción; los efectos inerciales no existen y, de hecho, la cuchara y el sirope se quedarán quietos en el momento en el que dejemos de hacer fuerza. En el caso del agua, por el contrario, incluso

un milinewton de fuerza aplicado a un objeto en el agua nos garantiza estar en el régimen turbulento, dominado por la inercia (como se observa, por ejemplo, al remover suavemente agua con una cuchara). Por el contrario, para una célula, que se mueve en el rango de los piconewtons, el agua le parecerá un fluido tan viscoso como el sirope lo es para nosotros. Los efectos inerciales en el mundo de la célula son despreciables, y el rozamiento gobierna la dinámica celular.

B.5. La base de datos del NCBI.

En esta tesis se han utilizado a menudo secuencias genómicas reales con las que realizar simulaciones numéricas. Dichas secuencias fueron obtenidas utilizando la aplicación telemática que tiene el NCBI (*National Center for Biotechnology Information*), y que permite gestionar la ingente base de datos de genómica que existe en estos momentos y que sigue creciendo. De hecho, la página web del NCBI, www.ncbi.nlm.nih.gov, ha cambiado mucho en los últimos años y, aunque lógicamente todavía mantiene la utilidad básica de bajar secuencias, ahora contiene mucha más información y utilidades de las que poder sacar partido con tiempo y dedicación. Pese a ello, en este apartado voy a dar un esbozo de los pasos a seguir para buscar y bajarse secuencias genómicas de diferentes organismos. Puesto que no es necesaria su lectura para la comprensión del resto de la memoria, a los que no les interese el tema pueden omitirla con total impunidad.

Una vez en la página de inicio del NCBI (mencionada en el párrafo anterior), y puesto que estamos interesados en conseguir secuencias de genomas, pinchamos en el apartado *Genomic biology*, en la columna de la izquierda. Una vez en esa página, de nuevo pinchamos en la columna de la izquierda, esta vez en el enlace *Genome, complete genome sequences*, para obtener las secuencias genómicas que necesitamos. Y de nuevo en la columna de la izquierda aparecen, entre otras cosas, los diferentes tipos de organismos vivos divididos taxonómicamente. Pinchando en cualquiera de ellos, salvo en *Chromosome* de *Eukaryota*, que describiré después, aparece una lista con los nombres científicos, por orden alfabético, de todos los organismos que se ajustan a esa clasificación, y a la derecha, un enlace con una etiqueta que suele empezar por “NC_” o “NT_” seguido de un número de seis cifras, a veces seguido de un punto y otro número, por ejemplo, “NT_028395.2”. Éso es lo que se denomina un *contig*, y es el objetivo de nuestra búsqueda.

Un *contig* es una secuencia específica de nucleótidos que ha sido obtenida mediante el solapamiento de clones de ADN de una región de un cromosoma en particular. A la derecha del *contig* aparece el número de nucleótidos que contiene, y más a la derecha la última fecha de revisión. Una vez que hemos pinchado en el

contig en el que estamos interesados, en una página se nos muestra la información principal del *contig* seleccionado: longitud, %GC, porción de ADN que codifica, tipo de molécula, genes que contiene, proteínas que codifica, etc. Para conseguir la información de la secuencia de nucleótidos es necesario ir a *GenBank*, ya sea pinchando en el enlace que aparece en la tabla resumen o cambiando el *Display* (en la parte superior izquierda) de *Overview* a *GenBank*. De esta forma la pantalla cambia a una pantalla sin gráficos, en modo texto, que da los datos técnicos de la secuencia seleccionada, como el *Locus*, el organismo al que pertenece, los artículos de los que se ha obtenido la información, y lo más importante, en el apartado de características (*features*), la descripción de cada intervalo de la secuencia: si es un intervalo que codifica proteína, si es un promotor, si no es nada... Al final aparece la secuencia de nucleótidos del *contig* completa, separado en columnas de diez elementos. Sin embargo, para obtener la secuencia completa, limpia, sin información extra, es necesario cambiar el formato en el que aparece el *Display*, de *GenBank* a *FASTA* (que era un formato que no aparecía antes). De esta forma podemos cortar y pegar el texto a un fichero o, si es demasiado grande, mandar la información (utilizando el cuadro de texto *send to*) a un fichero, texto o portapapeles. Existen otras formas de visualización del *Display*, como *Graph*, que da una imagen esquemática de la porción de genoma que se está viendo, que puede ser interesante (e incluso útil, nunca se sabe).

Si, en la organización taxonómica, pinchamos en la opción de *Eukaryota, chromosome* o *Eukaryota, genome*, entonces llegamos a una página en la que se observan los organismos eucariotas para los que se está estudiando el genoma junto con sus cromosomas. Pinchando en los cromosomas (o en el organismo y luego en los cromosomas, da igual), nos muestra una página en la que se esquematiza el cromosoma, los *contigs* que contiene junto con su posición en la figura, y mucha más información. La página es una aplicación que nos permite alejarnos o acercarnos para ver más detalle, elegir la región de cromosoma que queremos ver, bajarnos las secuencias de los *contigs* que se muestran, elegir una vista en modo texto, y posiblemente muchas más cosas que no he podido comprender. Al pinchar en los *contigs* de nuevo se llega a una página de la base de datos de *GenBank* en la que se encuentran los datos técnicos de la secuencia y de la que podemos cambiar al formato *FASTA* para bajarnos la secuencia que deseemos, de la misma forma que hicimos antes.

El formato más útil para utilizar secuencias en simulaciones es el *FASTA*, y que permite escribir un código muy sencillo en cualquier lenguaje de programación que vaya leyendo letra por letra la secuencia genética y asigne un valor a cada nucleótido. Sin embargo, en el formato *GenBank* podemos encontrar información

acerca de las posiciones de sitios relevantes de la secuencia que hayan sido encontrados, como genes y, en ocasiones, promotores. Esta información resulta sumamente útil en el análisis de resultados numéricos en regiones determinadas de la secuencia.

Publicaciones.

Durante la elaboración de esta tesis se han publicado los siguientes artículos:

- Cuenda S. y Sánchez A. (2004a). Disorder and fluctuations in nonlinear excitations in DNA. *Fluct. Noise Lett.* **4**, L491–L504.
- Cuenda S. y Sánchez A. (2004b). Nonlinear excitations in DNA: Aperiodic models versus actual genome sequences. *Phys. Rev. E* **70**, 51903.
- Cuenda S. y Sánchez A. (2005). Length scale competition in nonlinear Klein- Gordon models: a collective coordinate approach. *Chaos* **15**, 23502.
- Cuenda S. y Sánchez A. (2006). On the discrete Peyrard-Bishop model of DNA: stationary solutions and stability. *Chaos* **16**, 23123.
- Cuenda S., Sánchez A. y Quintero N. R. (2006). Does the dynamics of sine-Gordon solitons predict active regions of DNA? *Physica D* **223**, 214–221.
- González J. A., Cuenda S. y Sánchez A. (2007). Kink dynamics in spatially inhomogeneous media: The role of internal modes *Phys. Rev. E* **75**, 36611.

Referencias

- Ablowitz M. J., Kaup D. J., Newell A. C. y Segur H. (1973). Method for solving the sine-Gordon equation. *Phys. Rev. Lett.* **30**, 1262–1264.
- Alberts B., Johnson A., Lewis J., Raff M., Roberts K. y Walter P. (2002). *Molecular biology of the cell*. Garland Science.
- Arecchi F. T., Masserini G. L. y Schwendimann P. (1969). Electromagnetic propagation in a resonant medium. *Riv. Nuovo Cimento* **64**, 740–742.
- Baldazzi V., Brade S., Cocco S., Marinari E. y Monasson R. (2007). Inferring DNA sequences from mechanical unzipping data: the large-bandwidth case. *Phys. Rev. E* **75**, 11904.
- Baldazzi V., Cocco S., Marinari E. y Monasson R. (2006). Inference of DNA sequences from mechanical unzipping: An ideal-case study. *Phys. Rev. Lett.* **96**, 128102.
- Bashford J. D. (2006). Salerno's model of DNA re-analysed: could breather solitons have biological significance? *J. Biol. Phys.* **32**, 27–47.
- Bäcklund A. V. (1883). Om ytor med konstant negativ krökning. *Lunds Universitets Årsskrift Avd.* **19**, 1–48.
- Bianchi L. (1879). Ricerche sulle superficie a curvatura costante e sulle elicoidi. Tesi di Abilitazione. *Ann. Scuola Norm. Sup. Pisa (1)* **2**, 285–304.
- Bianchi L. (1892). Sulla trasformazione di Bäcklund per le superficie pseudosferiche. *Rend. Lincei* **5**, 3–12.
- Bockelmann U., Thomen P., V B., Essevaz-Roulet B., y Heslot F. (2002). Unzipping DNA with optical tweezers: high sequence sensitivity and force flips. *Biophys. J.* **82**, 1537–1553.
- Boesch R. y Willis C. R. (1990). Existence of an internal quasimode for a sine-Gordon soliton. *Phys. Rev. B* **42**, 2290.

- Bour E. (1862). Théorie de la déformation des surfaces. *J. l'École Imperiale Polytech* **19**, 1–148.
- Braun O. M. y Kivshar Y. S. (2004). *The Frenkel-Kontorova model: Concepts, methods and applications*. Springer.
- Campa A. y Giansanti A. (1998). Experimental tests of the Peyrard-Bishop model applied to the melting of very short DNACHAINS. *Phys. Rev. E* **58**, 3585–3588.
- Cheetham G. M. T., Jeruzalmi D. y Steitz T. A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83.
- Choi C. H., Kalosakas G., Rasmussen K. Ø., Hiromura M., Bishop A. R. y Usheva A. (2004). DNA dynamically directs its own transcription initiation. *Nuc. Acids. Res.* **32**, 1584–1590.
- Cocco S., Monasson R. y Marko J. F. (2002). Force and kinetic barriers to initiation of DNA unzipping. *Phys. Rev. E* **65**, 41907.
- Cuenda S. y Sánchez A. (2004a). Disorder and fluctuations in nonlinear excitations in DNA. *Fluct. Noise Lett.* **4**, L491–L504.
- Cuenda S. y Sánchez A. (2004b). Nonlinear excitations in DNA: Aperiodic models versus actual genome sequences. *Phys. Rev. E* **70**, 51903.
- Cuenda S. y Sánchez A. (2005). Length scale competition in nonlinear Klein-Gordon models: a collective coordinate approach. *Chaos* **15**, 23502.
- Cuenda S. y Sánchez A. (2006). On the discrete Peyrard-Bishop model of DNA: stationary solutions and stability. *Chaos* **16**, 23123.
- Cuenda S., Sánchez A. y Quintero N. R. (2006). Does the dynamics of sine-Gordon solitons predict active regions of DNA? *Physica D* **223**, 214–221.
- Cuesta-López S., Errami J., Falo F. y Peyrard M. (2005). Can we model DNA at mesoscale? *J. Biol. Phys.* **31**, 273–301.
- Danilowicz C., Coljee V. W., Bouzigues C., Lubensky D. K., Nelson D. R. y Prentiss M. (2003). DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc. Natl. Acad. Sci.* **100**, 1694–1699.
- Dauxois T., Peyrard M. y Bishop A. R. (1993). Entropy-driven DNA denaturation. *Phys. Rev. E* **47**, R44–R47.
- Dauxois T., Theodorakopoulos N. y Peyrard M. (2002). Thermodynamic instabilities in one dimension: correlations, scaling and solitons. *J. Stat. Phys.* **107**, 869–891.

- Debnath L. (1997). *Nonlinear partial differential equations for scientists and engineers*. Birkhäuser, Boston.
- Dehlinger U. (1929). Zur Theorie der Rekristallisation reiner Metalle. *Ann. Phys. (Leipzig)* **394**, 749–793.
- Domínguez-Adame F., Sánchez A. y Kivshar Y. S. (1995). Soliton pinning by long-range order in aperiodic systems. *Phys. Rev. E* **52**, R2183–R2186.
- Duke T., van Helden J., de Jong H., Kadonaga J., Khokhlov A., Lavery R., Marko J., Samsonov A., Samsonova M., Strick T., Thieffry D. y Westhof E. (2004). *Multiple aspects of DNA and RNA: from biophysics to bioinformatics*. Editado por D. Chatenay, S. Cocco, R. Monasson, D. Thieffry y J. Dalibard, Elsevier.
- Dunn J. J. y Studier F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Bio.* **166**, 477–535.
- Englander S. W., Kallenbach N. R., Heeger A. J., Krumhansl J. A. y Litwin S. (1980). Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. *Proc. Natl. Acad. Sci.* **77**, 7222–7226.
- Enneper A. (1870). Ueber eine Erweiterung des Begriffs von Parallelfächen. *Nachrichten von der Königl. Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen aus dem Jahre 1870*, 70–82.
- Enz U. (1964). Die Dynamik der blochschen Wand. *Helv. Phys. Acta* **37**, 245.
- Essevaz-Roulet B., Bockelmann U. y Heslot F. (1997). Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci.* **94**, 11935–11940.
- Fedyanin V. K. y Yakushevich L. V. (1984). Scattering of neutrons and light by DNA solitons. *Stud. Biophys.* **103**, 171–178.
- Fei Z., Konotop V., Peyrard M. y Vázquez L. (1993). Kink dynamics in the periodically modulated ϕ^4 model. *Phys. Rev. E* **48**, 548–554.
- Fogel M. B., Trullinger S. E., Bishop A. R. y Krumhansl J. A. (1976). Classical particlelike behaviour of sine-Gordon solitons in scattering potentials and applied fields. *Phys. Rev. Lett.* **36**, 1411–1414.
- Frenkel J. y Kontorova T. (1939). On the theory of plastic deformation and twinning. *J. Phys. (Moscow)* **1**, 137–149.
- Gardiner C. W. (1985). *Handbook of stochastic methods*. Springer.

- Gardner C. S., Greene J. M., Kruskal M. D. y Miura R. M. (1967). Method for solving the Korteweg-deVries equation. *Phys. Rev. Lett.* **19**, 1095–1097.
- Gardner C. S., Greene J. M., Kruskal M. D. y Miura R. M. (1974). Korteweg-de Vries equation and generalizations, VI, Methods for exact solution. *Comm. Pure Appl. Math.* **27**, 97–133.
- Gibbs W. W. (2003). The unseen genome: Gems among the junk. *Sci. Am.* **289**, 47–53.
- González J., Cuenda S. y Sánchez A. (2007). Kink dynamics in spatially inhomogeneous media. The role of internal modes. *Phys. Rev. E*.
- Gourse R. L., Ross W. y Gaal T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* **37**, 687–695.
- Guéron M., Kochoyan M. y Leroy J.-L. (1987). A single mode of DNA base-pair opening drives imino proton exchange. *Nature* **328**, 89–92.
- Hammack J. L. y Segur H. (1974). The Korteweg-de Vries equation in water waves, Part 2, Comparison with experiments. *J. Fluid Mech.* **65**, 289–314.
- Harlepp S., Marchal T., Robert J., Leger J. F., Xayaphoummine A., Isambert H. y Chatenay D. (2003). Probing complex RNA structures by mechanical force. *Eur. Phys. J. E* **12**, 605–615.
- Hirota R. (1971). Exact solution of the Korteweg-de Vries equation for multiple collisions of solitons. *Phys. Rev. Lett.* **27**, 1192–1194.
- Hirota R. (1973a). Exact envelope-soliton solutions of a nonlinear wave equation. *J. Math. Phys.* **14**, 805–809.
- Hirota R. (1973b). Exact N-solutions of the wave equation of long waves in shallow water and in nonlinear lattices. *J. Math. Phys.* **14**, 810–814.
- Inman R. B. y Baldwin R. L. (1964). Helix-random coil transitions in DNA homopolymer pairs. *J. Mol. Biol.* **8**, 452–469.
- Josephson B. D. (1965). Supercurrents through barriers. *Adv. Phys.* **14**, 419–451.
- Kalbermann G. (2006). The sine-gordon wobble. *J. Phys. A: Math. Gen.* **37**, 11603–11612.
- Kalosakas G., Rasmussen K. Ø., Bishop A. R., Choi C. H. y Usheva A. (2004). Sequence-specific thermal fluctuations identify start sites for DNA transcription. *Europhys. Lett.* **68**, 127–133.

- Kamenetskii M. F. (1987). Physicists retreat again. *Nature* **328**, 108.
- Korteweg D. J. y de Vries G. (1895). On the change of form of long waves advancing in a rectangular channel, and on a new type of long stationary waves. *Phil. Mag.* **39**, 422–443.
- Lamb G. L. (1980). *Elements of soliton theory*. John Wiley, New York.
- Lax P. D. (1968). Integrals of nonlinear equations of evolution and solitary waves. *Comm. Pure Appl. Math.* **21**, 467–490.
- Lax P. D. (1978). *A Hamiltonian approach to the KdV and other equations*, 207-244; in *Nonlinear evolution equations*, M. G. Crandall. Academic Press, New York.
- Lennholm E. y Hörnquist M. (1999). Role of aperiodic order for fluxon dynamics in Josephson junction arrays. *Phys. Rev. E* **59**, 381–389.
- Lennholm E. y Hörnquist M. (2003). Revisiting Salenos's sine-Gordon model of DNA active regions and robustness. *Physica D* **177**, 233–241.
- Liphardt J., Onoa B., Smith S. B., Tinoco I. y Bustamante C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science* **297**, 733–737.
- Lubensky D. K. y Nelson D. R. (2002). Single molecule statistics and the polynucleotide unzipping transition. *Phys. Rev. E* **65**, 31917.
- Malomed B. A. (1990). Motion of a kink in a spatially modulated sine-Gordon system. *Phys. Lett. A* **144**, 351–356.
- Malomed B. A. y Tribelsky M. I. (1990). Emission of waves by moving kinks in a spatially modulated sine-Gordon system. *Phys. Rev. B* **41**, 11271–11281.
- Mandal C., Kallenbach N. R. y Englander S. W. (1979). Base-pair opening and closing reactions in the double helix. A stopped-flow hydrogen exchange study in poly(rA).poly(rU). *J. Mol. Biol.* **135**, 391–411.
- Marenduzzo D., Bhattacharjee S., Maritan A., Orlandini E. y Seno F. (2002). Dynamical scaling of the DNA unzipping transition. *Phys. Rev. Lett.* **88**, 28102.
- McLaughlin D. W. y Scott A. C. (1978). Perturbation analysis of fluxon dynamics. *Phys. Rev. A* **18**, 1652–1680.
- Mkrtychyan G. S. y Shmidt V. V. (1979). On the radiation from inhomogeneous Josephson junction. *Solid State Commun.* **30**, 783–784.

- Nabarro F. R.Ñ. (1947). Dislocations in a simple cubic lattice. *Proc. Phys. Soc.* **59**, 256–272.
- Nelson P. (2004). *Biological physics: energy, information, life*. W. H. Freeman and Co., New York.
- Newell A. C. (1985). *Solitons in mathematics and physics*. SIAM, Philadelphia.
- Peierls R. (1940). The size of a dislocation. *Proc. Phys. Soc.* **52**, 34–37.
- Peyrard M. (2004). Nonlinear dynamics and statistical physics of DNA. *Nonlinearity* **17**, R1–R40.
- Peyrard M. y Bishop A. R. (1989). Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.* **62**, 2755–2758.
- Pokrovsky V. L. (1981). Splitting of commensurate-incommensurate phase-transition. *J. Phys.* **42**, 761–766.
- Polozov V. y Yakushevich L. V. (1988). Nonlinear waves in DNA and regulation of transcription. *J. Theor. Biol.* **130**, 423.
- Prandtl L. (1928). Ein Gedankenmodell zur kinetischen Theorie der festen Körper. *Z. Angew. Math. Mech.* **8**, 85–106.
- Press W. H., Teulosky S. A., Vetterling W. T. y Flannery B. P. (1992). *Numerical recipes in C*. Cambridge University Press.
- Qasmi H., Barré J. y Dauxois T. (2004). Links between nonlinear dynamics and statistical mechanics in a simple one-dimensional model. *arXiv:cond-mat/0407662*, 9.
- Quintero N. R., Sánchez A. y Mertens F. (2000). Existence of internal modes of sine-Gordon kinks. *Phys. Rev. E* **62**, R60–R63.
- Quintero N. R., Sánchez A. y Mertens F. (2006). Reply to “comment on ‘Existence of internal modes of sine-Gordon kinks’ ”. *Phys. Rev. E* **73**, 068602.
- Quintero N. R. y Zamora-Sillero E. (2004). Lagrangian formalism in perturbed nonlinear Klein-Gordon equations. *Physica D* **197**, 63–68.
- Remoissenet M. (1986). Low-amplitude breather and envelope solitons in quasi-one-dimensional physical models. *Phys. Rev. B* **33**, 2386–2392.
- Remoissenet M. (1994). *Waves called solitons: concepts and experiments*. Springer-Verlag.
- Rice M. J. (1983). Physical dynamics of solitons. *Phys. Rev. E* **28**, 3587–3589.

- Rogers C. y Schief W. K. (2002). *Bäcklund and Darboux transformations: Geometry and modern applications in soliton theory*. Cambridge University Press.
- Rubinstein J. (1970). Sine-Gordon equation. *J. Math. Phys* **11**, 258–266.
- Russell J. S. (1844). *Report on Waves*. 14th meeting of the British Association for the Advancement of Science (BAAS).
- Salerno M. (1991). Discrete model for DNA-promoter dynamics. *Phys. Rev. A* **44**, 5292–5297.
- Salerno M. (1992). Dynamical properties of DNAPromoters. *Phys. Lett. A* **167**, 49–53.
- Salerno M. y Kivshar Y. S. (1994). DNA promoters and nonlinear dynamics. *Phys. Lett. A* **193**, 263–266.
- Scott A. (1999). *Nonlinear science: Emergence & dynamics of coherent structures*. Oxford University Press.
- Scott A. C. (1969). A nonlinear Klein-Gordon equation. *Phys. Rev. Lett.* **37**, 52–61.
- Scott A. C., Chu F. Y. F. y McLaughlin D. W. (1973). The soliton - A new concept in applied science. *Proc. IEEE* **61**, 1443–1483.
- Seeger A. K. (1980). The beginnings of solid state physics. *Proc. Roy. Soc. London* **371**, 173–177.
- Segur H. (1983). Wobbling kinks in the sine-Gordon theory. *J. Math. Phys.* **24**, 1439–1443.
- Skyrme T. H. R. (1958). A nonlinear theory of strong interactions. *Proc. Roy. Soc. London* **A262**, 237–145.
- Sánchez A. y Bishop A. R. (1994). Kink stability, propagation, and length-scale competition in the periodically modulated sine-Gordon equation. *Phys. Rev. E* **49**, 4603–4615.
- Sánchez A., Scharf R., Bishop A. R. y Vázquez L. (1992). Sine-Gordon breathers on spatially periodic potentials. *Phys. Rev. A* **45**, 6031–3067.
- Tang G.-Q., Bandwar R. P. y Patel S. S. (2005). Extended upstream A-T sequence increases T7 promoter strength. *J. Biol. Chem.* **280**, 40707–40713.
- Theodorakopoulos N., Peyrard M. y MacKay R. S. (2004). Nonlinear structures and thermodynamics instabilities in a one-dimensional lattice system. *Phys. Rev. Lett.* **93**, 258101.

- Wartell R. M. y Benight A. S. (1985). Thermal denaturation of DNAmolecules: a comparison of theory with experiments. *Phys. Rep.* **126**, 67.
- Watson J. D. y Crick F. H. C. (1953). Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.
- Weidman P. D. y Maxworthy T. (1978). Experiments on strong interactions between solitary waves. *J. Fluid Mech.* **85**, 417–431.
- Willis C. R. (1997). Spontaneous emission of continuum sine-Gordon kink in the presence of spatially periodic potential. *Phys. Rev. E* **55**, 6097–6100.
- Willis C. R. (2006). Comment on “Existence of internal modes of sine-Gordon kinks”. *Phys. Rev. E* **73**, 068601.
- Yoon B. (1976). Infinite sequence of conserved currents in the sine-Gordon theory. *Phys. Rev. D* **13**, 3440–3445.
- Zabusky N. J. (1967). A synergetic approach to problems of nonlinear dispersive wave propagation and interaction, *Proc. Symp. on Nonlinear Partial Differential Equations*. W. F. Ames, Academic Press, pp. 223-258.
- Zabusky N. J. y Galvin C. J. (1971). Shallow water waves, the Korteweg-de Vries equation and solitons. *J. Fluid Mech.* **47**, 811–824.
- Zabusky N. J. y Kruskal M. D. (1965). Interaction of solitons in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett.* **15**, 240–243.
- Zakharov V. E. y Shabat A. B. (1972). Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet Phys. JETP* **34**, 62–69.