

## Modelos matemáticos en biología: un viaje de ida y vuelta

R. ÁLVAREZ-NODARSE

Departamento de Análisis Matemático, Universidad de Sevilla

ran@us.es    <http://euler.us.es/~renato/>

*A Renato Alejandro y a Niurka, por el tiempo que les he robado*

### Resumen

El objetivo de este trabajo es mostrar la provechosa interacción entre la Biología y la Matemática. Para ello veremos cómo, por una parte, la Matemática es una herramienta sumamente interesante para entender distintos fenómenos biológicos como la dinámica del ADN, el crecimiento de tumores, dinámica de poblaciones, etc., y estos, a su vez, son una fuente de problemas matemáticos difíciles.

**Palabras clave:** *modelos matemáticos en biología, dinámica de poblaciones, modelos matemáticos del ADN, modelos de crecimiento de tumores*

**Clasificación por materias AMS:** *92B05, 92C05, 92D20, 92D25*

## Introducción

El objetivo de este *breve* trabajo es mostrar la provechosa interacción entre las matemáticas y las ciencias biológicas, ese camino de ida y vuelta: ida pues las matemáticas pueden ayudarnos a entender (e intentar predecir) muchos fenómenos biológicos, y vuelta, pues recorriendo ese camino los matemáticos tienen una inagotable fuente de problemas difíciles –a ese respecto véase el magnífico artículo [20]–. No en balde muchos científicos afirman que si bien el siglo XX fue el siglo de la Física, este siglo, el XXI, será el de la Biología. El trabajo está dividido en cuatro apartados. En el primero describiremos algunos modelos de poblaciones. En el segundo apartado daremos una breve introducción a la biología celular y molecular necesarios para entender mejor los modelos discutidos en los apartados 3 y 4.

## 1 Modelos de poblaciones

Los primeros modelos matemáticos aplicados en Biología han sido quizá los modelos que intentan describir la dinámica de poblaciones. Vamos a discutir aquí brevemente algunos de ellos. Para más detalles ver [10, 23]. Por simplicidad vamos a centrarnos en modelos para una única especie.

---

Fecha de recepción: 10/11/05

### 1.1 El modelo de Fibonacci

Quizá el modelo más antiguo de crecimiento de poblaciones es el modelo que Leonardo de Pisa (o Fibonacci, como se le conoce desde el siglo XVIII) utilizó para describir el crecimiento de una población de conejos y que describió en su famoso libro sobre la Aritmética, *Liber abaci*, de 1202. El problema es el siguiente: Partiendo de una pareja de conejos (macho y hembra) ¿cuántas parejas habrá al principio de cada temporada? ¿qué cantidad hay después de  $n$  temporadas?

Para resolverlo Fibonacci supuso ciertas *reglas*:

1. Comenzamos con una única pareja de conejos (macho y hembra).
2. Cada pareja de conejos (macho y hembra) madura (puede reproducirse) pasado cierto tiempo  $T$  (una temporada de crianza).
3. Cada pareja madura de conejos *produce* una única nueva pareja de conejos (macho y hembra) cada temporada de crianza (o sea pasado el tiempo  $T$ ).
4. Los conejos son inmortales.

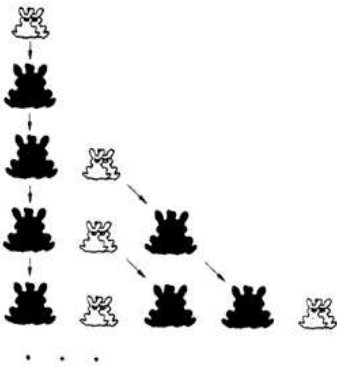


Figura 1: Los conejos de Fibonacci. En cada fila se representan las parejas de conejos por temporada. Las parejas maduras son las de color negro.

Si denotamos por  $N_t$  el número de parejas (macho y hembra) de conejos al principio de cada temporada y por  $t$  la correspondiente temporada, entonces la población de conejos se describe por la *ecuación en diferencias* (recurrencia)

$$N_{t+1} = N_t + N_{t-1}, \quad t = 1, 2, 3, \dots$$

Si empezamos por  $t = 1$  con las condiciones iniciales  $N_0 = N_1 = 1$ , la fórmula anterior nos genera la famosa sucesión de Fibonacci:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

La solución general de dicha ecuación es muy sencilla pues es una ecuación en diferencias lineal y homogénea. Si buscamos la solución en forma  $N_t = \lambda^t$ , sustituyendo en la ecuación anterior obtenemos

$$\lambda^2 - \lambda - 1 = 0 \Rightarrow \lambda_{1,2} = \frac{1}{2}(1 \pm \sqrt{5});$$

así, puesto que es una ecuación lineal de orden dos, su solución general es

$$N_t = A\lambda_1^t + B\lambda_2^t,$$

que usando las condiciones iniciales  $N_0 = N_1 = 1$  nos da:

$$N_t = \frac{1}{\sqrt{5}}(\lambda_1^{t+1} - \lambda_2^{t+1}).$$

Nótese que si  $t$  es suficientemente grande  $N_t \approx \lambda_1^{t+1}/\sqrt{5}$ . Un hecho curioso es que  $N_{t+1}/N_t \approx (1 + \sqrt{5})/2$ , que es la famosa razón áurea.

No vamos a detenernos aquí en el análisis de cuán real es el modelo de Fibonacci (especialmente si tenemos en cuenta la regla 4) y vamos a ver otros modelos *más realistas*.

## 1.2 El modelo malthusiano o exponencial

Imaginemos que tenemos una población de cierta especie (consideraremos que tenemos un número bastante alto de individuos) y sea  $p(t)$  el número de individuos de dicha especie en el momento  $t$  (evidentemente  $p(t) \in \mathbb{N}$ ). Supongamos que la población está aislada (o sea, no hay emigración ni inmigración). Sea  $r(t, p)$  la diferencia entre la razón de natalidad y mortalidad de la población. Entonces la variación  $p(t+h) - p(t)$  es proporcional a  $p(t)h$  siendo el coeficiente de proporcionalidad  $r(t, p)$ . Luego

$$p(t+h) - p(t) = r(t, p)p(t)h,$$

de donde, tomando límites cuando  $h \rightarrow 0$  queda  $p'(t) = r(t, p)p(t)$ .

La ecuación más sencilla posible se obtiene si consideramos  $r(t, p) = r$ , constante. Así, la población de individuos de la especie puede ser modelizada mediante el problema de valores iniciales (PVI)

$$p'(t) = r p(t), \quad p(t_0) = p_0, \quad r > 0, \quad (1)$$

cuya solución  $p(t) = p_0 e^{r(t-t_0)}$ . El modelo anterior se conoce como modelo de Malthus o modelo malthusiano pues fue propuesto por el economista inglés Thomas R. Malthus (1766–1834). Si  $r < 0$ , la especie está condenada a la extinción y, si  $r > 0$ , ésta crece en proporción geométrica.

Veamos un ejemplo. Según estimaciones del Departamento de Comercio de los EEUU, la Tierra estaba habitada en 1965 por  $3,34 \cdot 10^9$  personas. Según estudios realizados durante varios años, se cree que el índice de crecimiento es de un 2% anual, o sea  $r = 0,02$ . Entonces la evolución de la población mundial es

$$p(t) = 3,34 \cdot 10^9 e^{0,02(t-1965)}. \quad (2)$$

Por ejemplo, la ecuación nos predice para el año 1989 la cantidad  $5,39 \cdot 10^9$  que se acerca bastante a la estimada de  $5,18 \cdot 10^9$ , y para el año 2004 de  $7,29 \cdot 10^9$  (la cifra “real” es  $6,45 \cdot 10^9$ ). Además, la fórmula anterior nos dice que la población se duplica cada

$$2N = N e^{0,02T}, \quad T = 50 \log 2 = 34,6 \text{ años.}$$

¿Es realista este modelo? Para responder a esta pregunta vamos a ver qué ocurrirá en el año 2515, por ejemplo. Usando la fórmula (2) tenemos  $p(2515) = 3,34 \cdot 10^9 \exp(0,02 \cdot 535) = 2 \cdot 10^{14}$ , que es un número bastante grande. Veamos cómo estará repartida la población. La superficie de la tierra es de aproximadamente  $5,1 \cdot 10^8 \text{ Km}^2 = 5,1 \cdot 10^{14} \text{ m}^2$ , así que la densidad de población será de un habitante por cada  $0,4 \text{ m}^2$ , o sea un cuadrado de menos de 65 cm de lado, y eso sin contar que el 80 % de la superficie terrestre es agua, que nos da apenas un cuadrado de 28 cm de lado. ¿Qué ocurrirá en el 2550?<sup>1</sup>

El modelo discreto correspondiente es  $p(t+h) - p(t) = rp(t)$ , que reescribimos convenientemente en la forma

$$p(t+1) = \lambda p(t) \quad \implies \quad p(t) = p_0 \lambda^t.$$

Este modelo es, por tanto, muy similar al modelo continuo que hemos discutido antes.

### 1.3 El modelo logístico

Aunque hemos visto que el modelo funciona razonablemente bien para poblaciones grandes, hay que hacer varias correcciones pues si  $p(t)$  empieza a crecer demasiado habrá muchos otros factores como la falta de espacio o de alimentos que frenará el crecimiento. Así pues, varios años después, en 1837, un matemático y biólogo holandés, P. F. Verhulst, propuso un modelo algo más realista conocido como el modelo logístico (lo publicó 10 años más tarde). Verhulst razonó que, como estadísticamente el encuentro de dos individuos es proporcional a  $p^2$  (¿por qué?) entonces al término  $rp$  de la ecuación (1) habría que sustraerle el término  $cp^2$ . Así, que el PVI que modeliza el crecimiento de la población será

$$p'(t) = rp(t) - cp^2(t), \quad p(t_0) = p_0, \quad r, c > 0. \quad (3)$$

En general  $c$  ha de ser mucho más pequeño que  $r$  ya que si  $r$  no es muy grande la ecuación (1) es una aproximación bastante buena, pero si  $p$  comienza a crecer demasiado entonces el término  $-cp^2$  no se puede obviar y termina frenando el crecimiento exponencial. Usualmente  $c$  se escribe como el cociente  $r/K$  donde  $K$  es la capacidad de carga que está ligada a los recursos del hábitat.

---

<sup>1</sup>Malthus, en su ensayo de 1798 “*Acerca de los fundamentos de la población*”, abusando de su ecuación, predijo un colapso socioeconómico pues proclamó que la capacidad procreadora de los humanos es tan grande que siempre nacen más niños que los que pueden sobrevivir, lo cual se agrava por el hecho de que la cantidad de alimentos no crece exponencialmente (Malthus creía que dicho crecimiento era aritmético). Así Malthus concluyó que un gran número de humanos está predestinado a sucumbir en la lucha por la existencia. Aunque hoy día no se consideran ciertas las teorías malthusianas, en su momento tuvieron gran influencia; de hecho fue la lectura del ensayo de Malthus lo que inspiró a Darwin el mecanismo de selección natural universalmente aceptado hoy día.

La EDO (3) es una ecuación separable luego

$$\frac{dp}{p(r - cp)} = dt \Rightarrow \int \frac{dp}{p(r - cp)} = \frac{1}{r} \log \left| \frac{p}{r - cp} \right| = t + C \Rightarrow$$

$p/(r - cp) = Ce^{rt}$  y, por tanto, la solución del PVI (3) es

$$p(t) = \frac{rp_0 e^{r(t-t_0)}}{r - cp_0 + cp_0 e^{r(t-t_0)}} = \frac{rp_0}{cp_0 + (r - cp_0)e^{-r(t-t_0)}}.$$

Nótese que  $\lim_{t \rightarrow \infty} p(t) = r/c = K$  independientemente del valor inicial  $p_0$ . En el caso  $0 < p_0 < r/c$ , la evolución de la población está representada en la gráfica 2. Un ejercicio sencillo resulta comprobar que  $p = 0$  es un punto de equilibrio inestable de (3), mientras que  $p(t) = K$  es estable. También varía la forma de la curva si  $p_0 > K$  o  $p_0 < K$ ; además, en este último caso hay que discernir dos posibilidades  $K/2 \leq p_0 < K$  y  $0 < p_0 < K/2$  (ver para más detalles ver [23]).

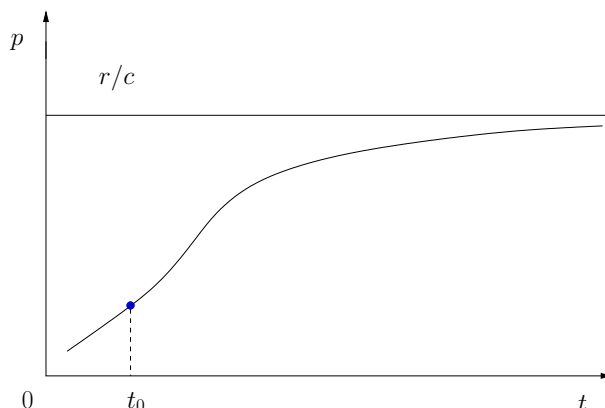


Figura 2: Evolución de la población según el modelo logístico (3) cuando  $p_0 < r/c$ .

Este modelo se ha comprobado con varias especies y ha dado muy buenos resultados. En particular, el hecho de que la población se estabilice ha sido comprobado en distintos experimentos con paramecios, obteniéndose una gran concordancia entre los resultados teóricos y experimentales.

Vamos a aplicar el modelo logístico a la población mundial. Según los ecólogos, el valor “natural” de  $r$  es de 0,029, así que si tomamos los datos de 1965 tenemos que  $p'/p = r - cp$  y, como hemos dicho antes, el incremento anual en 1965 era  $p'/p \approx 0,02$ , luego  $0,02 = 0,029 - c(3,34 \cdot 10^9)$  y por tanto  $c = 2,695 \cdot 10^{-12}$ . Entonces el modelo logístico predice que la población mundial se estabilizará en torno a la cantidad  $p_\infty = r/c = 1,07 \cdot 10^{10}$ , lo cual ocurrirá según la ley malthusiana en el 2024.

Obviamente este modelo sigue siendo muy simple ya que no tiene en cuenta ni las guerras (habituales desde hace cientos de años) ni las epidemias.

Un ejercicio interesante es aplicar el modelo logístico en España. Si usamos los datos de los censos de 1940 a 1990 nos da que  $r = 0,008$ . Si tomamos  $p(1940) = 2,62 \cdot 10^7$  podemos encontrar el número de habitantes en 1998 y compararlo con el número censado  $p = 3,98 \cdot 10^7$  personas. ¿Cuál será la población en España en el 2500 si sigue la misma tendencia? Como ejercicio al lector, proponemos que haga una estimación del número de habitantes que han de estabilizarse en España (suponiendo que la razón real representa el mismo porcentaje que en el caso de la población mundial y que no haya emigración ni inmigración).

El modelo logístico tiene una versión discreta muy interesante determinada por la ecuación (la hemos escrito convenientemente)

$$u(t+1) = ru(t)(1-u(t)), \quad r > 0, \quad (4)$$

donde se asume que  $u_0 \in (0, 1)$ . Este modelo, aparentemente sencillo, esconde bifurcaciones y caos [23]. En la figura 3 podemos apreciar varias soluciones de la ecuación (4) para diferentes valores de  $r$ . Para  $r = 2$  vemos que la solución tiende muy rápidamente al valor  $u^* \approx 1/2$ . Al aumentar  $r$  la solución comienza a oscilar, primero entre dos valores, cuatro, etc, hasta convertirse en caótica a partir  $r \approx 3,9$ .

#### 1.4 Modelos generales

En general, los modelos discretos de poblaciones más usados tienen la forma

$$u(t+1) = f(u(t)) = u(t)F(u(t)),$$

donde  $f$  es una función no lineal. Las soluciones estacionarias  $u^*$  del modelo corresponden a  $u^* = 0$  o  $F(u^*) = 1$ . Un análisis general de este tipo de ecuaciones está fuera del propósito de este trabajo. Una magnífica discusión, en particular de las condiciones de estabilidad de las soluciones estacionarias, se puede consultar en [23, §2.2]. Los correspondientes modelos continuos son del tipo  $p'(t) = f(p)$  donde  $f$  es generalmente una función no lineal.

Para terminar mencionaremos cómo un, aparentemente inocente, experimento de Biología ha llevado a desarrollar toda una rama de investigación matemática dentro de las ecuaciones en diferencias: las ecuaciones en diferencias no lineales forzadas periódicamente. En [16] Jillson describe un experimento muy interesante. El experimento consistía en estudiar cómo evolucionan dos poblaciones de escarabajos en dos medios muy diferentes. En un medio se mantenía una cantidad fija de flores (que constituían el alimento de los escarabajos): 20 gramos; mientras que en el otro medio la cantidad de flores alternaba entre 32

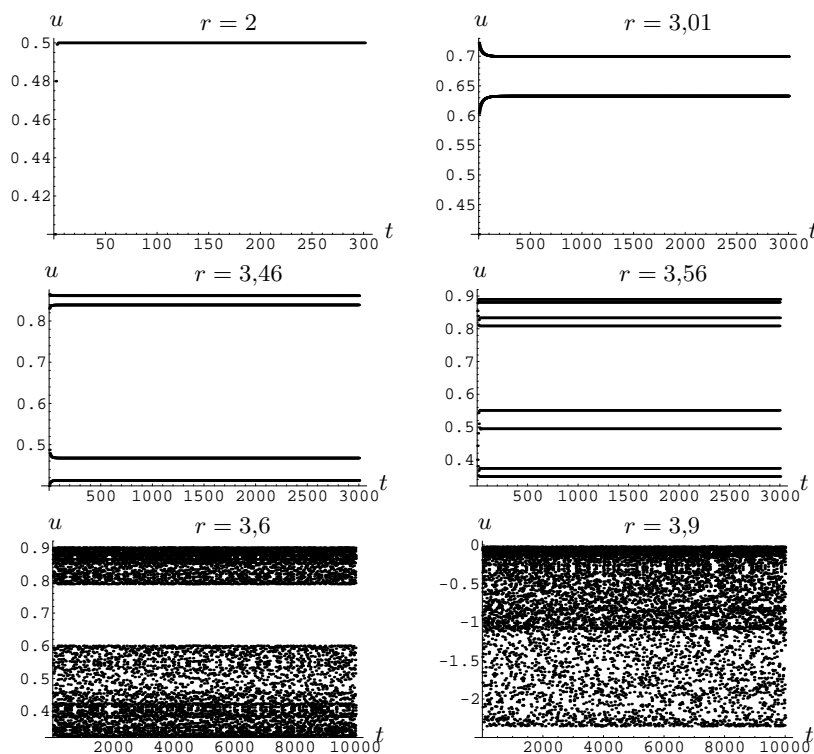


Figura 3: Caos en el modelo logístico discreto (4).

y 8 gramos cada dos semanas. El resultado fue que el número total de los escarabajos del medio oscilante era mucho mayor (más de dos veces) que el número de individuos del medio de control constante.

Entre las ecuaciones candidatas para explicar el fenómeno de Jillson estaban las ecuaciones en diferencias del tipo

$$u(t+1) = ru(t) \left( 1 - \frac{u(t)}{K(t)} \right), \quad r > 0, \quad K(t) > 0.$$

Nótese que, si  $K$  es constante, recuperamos la ecuación (4). El primer intento fallido de encontrar una ecuación que modelara el fenómeno de Jillson fue usar

$$K(t) = K(1 - a(-1)^t), \quad a \in (0, 1),$$

aunque no se obtuvieron los resultados deseados. Otra opción también fallida fue usar la ecuación de Beverton-Holt

$$u(t+1) = \frac{\mu K(t)u(t)}{K(t) + (\mu - 1)u(t)}, \quad K(t) > 0,$$

donde  $\mu$  es la razón de crecimiento *per capita*.

Otro modelo de poblaciones, objeto de un intensivo estudio en la actualidad, es el *modelo LPA* (ver e.g. [9]) basado en ecuaciones en diferencias de la forma

$$u(t+1) = [F(t, u(t)) + T(t, u(t))]u(t),$$

donde  $T$  y  $F$  son términos correspondientes a la *transición* y *fertilidad*, respectivamente. Para terminar debemos decir que aunque el fenómeno de Jillson ha generado una *gran cantidad de matemáticas* (e.g., las ecuaciones forzadas periódicamente antes mencionadas), todavía no se encontrado una ecuación que lo explique convenientemente.

## 2 Breve introducción a la biología celular y molecular

Antes de pasar a discutir algunos modelos matemáticos utilizados para describir la dinámica interna del ADN y el crecimiento y control tumoral, debemos hacer una breve incursión por la biología celular y molecular.

### 2.1 La célula

Célula [Diccionario RAE] (Del lat. *cellŭla*, dim. de *cella*, hueco). *f. Biol. Unidad fundamental de los organismos vivos, generalmente de tamaño microscópico, capaz de reproducción independiente y formada por un citoplasma y un núcleo rodeados por una membrana.*

Célula [Alberts et al. [1]]. *Es el vehículo a través del cual se transmite la información hereditaria que define cada especie y que además contiene la maquinaria necesaria para obtener materiales del ambiente y generar una nueva célula a su imagen, que contendrá una nueva copia de la información hereditaria.*

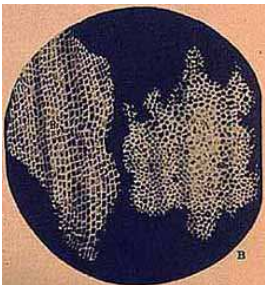


Figura 4: Las células del corcho. Foto del texto *Micrografía* de R. Hooke.

Todos los organismos vivos están compuestos por células. De hecho estas estructuras son los ladrillos sobre los cuales se construyen el resto de los organismos vivos que conocemos en la naturaleza, ya sean uni o pluricelulares. Por tanto es de esperar que el estudio de sus componentes químicos y la interacción entre los mismos pueda arrojar alguna luz sobre su funcionamiento.

El primero en observar (descubrir) las células (aunque muertas) fue Robert Hooke, quien estudió con un microscopio un delgado corte de corcho. Hooke notó que el material era poroso. Esos poros, en su conjunto, formaban cavidades poco profundas a modo de cajas a las que llamó células (del latín “*cella*”: cámara, espacio



vacío). Sus observaciones las publica en 1665 en el libro *Micrografía* (ver figura 4). Unos años más tarde, Marcelo Malpighi, anatomista y biólogo italiano, observó células vivas. Pasarían todavía más de 100 años hasta que en 1838, y gracias al perfeccionamiento de los microscopios, el biólogo Mathias Jakob Schleiden y el médico Theodor Schwann sentaron las bases de la moderna teoría celular que establece que todos los organismos vivos están constituidos por células.

Esencialmente existen dos tipos fundamentales de células: las procariotas y las eucariotas. Las primeras son mucho más simples y carecen de un núcleo definido y su organización interna es muy sencilla, mientras que las segundas poseen una estructura interna más complicada y poseen un núcleo bien definido y limitado por una membrana (ver figura 5).

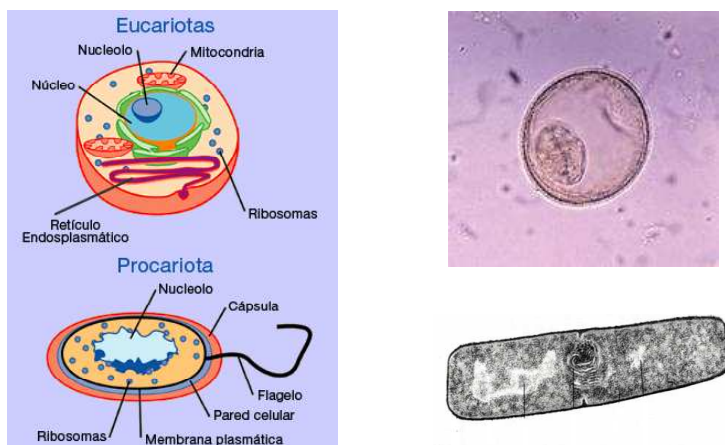


Figura 5: Clasificación de las células: Esquema (izquierda) y foto (derecha). Eucariotas (arriba) y procariotas (abajo).

Debido a que las células viven en un medio acuoso y están constituidas casi en su totalidad de agua, es preciso que haya una frontera bien definida que limite a la célula: la membrana plasmática (constituida por lípidos) dentro de la cual se encuentra el citosol –fase acuosa sin estructura del citoplasma.

Las procariotas tienen su ADN condensado en la región central y poseen varios ribosomas –partículas sintetizadoras de las proteínas en la célula–. Las eucariotas son mucho más complejas ya que contienen un sinnúmero de organelas: el núcleo, las mitocondrias –donde ocurre casi todo el metabolismo energético–, el aparato de Golgi, los lisosomas –degradan los componentes celulares desgastados y los innecesarios–, etc.

Todas las células guardan la información de la misma forma: **codificada en las moléculas de ADN**. Es por ello que nuestra atención se centrará en el estudio de dichas moléculas.

## 2.2 Estructura del ADN

En este apartado vamos a describir brevemente qué es el ADN y qué lo hace tan especial. Una información más completa se puede encontrar en [1, 19].

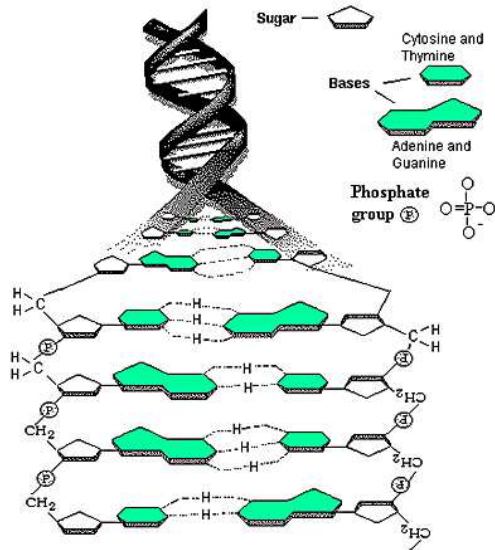


Figura 6: Estructura del ADN.

La molécula de ADN (ácido desoxirribonucleico) está formada por (ver figura 6) dos cadenas de nucleótidos o monómeros compuestos por un azúcar (la desoxirribosa) unida a un grupo fosfato  $PO_4$  que conforman el denominado esqueleto de azúcar-fosfato. A este esqueleto van unidas las bases que pueden ser la adenina (A), la citosina (C), la guanina (G) o la timina (T). Cada azúcar de la cadena está unida a la siguiente mediante el grupo fosfato y de ella sobresale la base.

La molécula de ADN no existe como una cadena libre sino que se junta con otra cadena de ADN formando una de doble hélice tridimensional con las bases colocadas hacia el interior (ver figura 6). Formalmente cada cadena simple puede estar constituida por una secuencia aleatoria de bases. No obstante, en la célula no hay semejantes cadenas simples y el ADN no se sintetiza como una cadena libre aleatoria, sino que lo hace a partir de la cadena original que es usada como molde o patrón. La razón fundamental se debe a que, debido a la estructura química de las bases, sólo se juntan las bases A con las T (mediante un enlace doble de puente de hidrógeno), y las G con las C (a través de un enlace triple).

Así pues, la doble cadena original se separa en dos cadenas simples a partir de las cuales se sintetizan dos nuevas cadenas dobles. Además, el procedimiento es muy sencillo: las bases de la nueva cadena (cadena **complementaria**) se van uniendo a las bases de la cadena patrón siguiendo el esquema preestablecido: A-T y C-G.

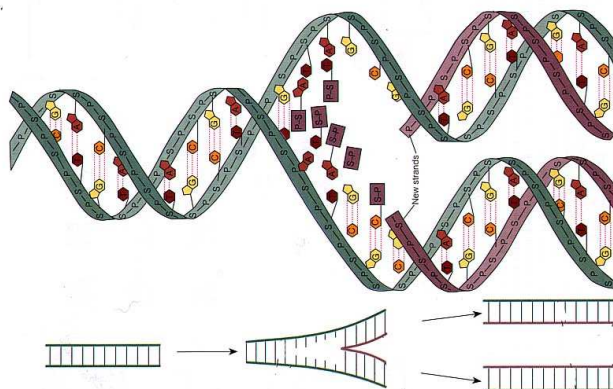


Figura 7: Autorreplicación del ADN.

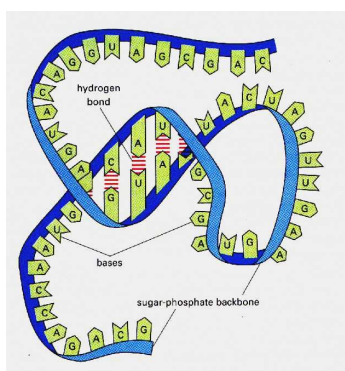


Figura 8: Estructura del ARN.

Aquí hay que destacar un punto de vital importancia: Resulta que los enlaces entre las bases (puentes de hidrógeno) son mucho más débiles que los enlaces azúcar- $\text{PO}_4$ . Esto precisamente permite que se rompa la doble cadena y queden las dos cadenas simples. Además, por la forma de unirse las bases es evidente que ambas cadenas no son independientes sino complementarias (ver la figura 7), de forma que al final quedan dos cadenas de ADN idénticas a la cadena original. El proceso anterior se conoce como la **replicación del ADN** y justifica el porqué la base de la *vida* –tal y como la conocemos hoy día– radica precisamente en el ADN.

Aparte de autorreplicarse, el ADN debe expresar la información que contiene utilizándola para dirigir la síntesis de otras moléculas de la célula. Este proceso es bastante complejo pero se puede resumir en dos fases:

1. La **transcripción**: El ADN sintetiza una molécula de ARN (ácido ribonucleico) muy similar en estructura al propio ADN (ver figura 8), pero más corta, constituida por otro azúcar (ribosa) y cuatro bases, A, C, G y U (uracilo) reemplazando, este último, a la timina T.

2. La **traducción**: El ARN se utiliza para sintetizar proteínas.

El proceso físico de la transcripción es bastante complicado ya que se requiere la interacción de distintos agentes químicos (moléculas) que pueden variar de célula a célula, pero su esencia es la misma. Nosotros nos limitaremos a describir brevemente cómo ocurre la traducción.

### 2.3 Las proteínas

Las proteínas son otras moléculas esenciales para la vida. Están constituidas por una larga cadena de nucleótidos, distintos de los del ADN y el ARN, denominados **aminoácidos**. Todos estos aminoácidos (hay 20 tipos distintos) tienen una estructura central por la que pueden unirse unos con otros (estructura primaria). Unida a dicha estructura central, el aminoácido tiene una estructura lateral (secundaria), que es la que les confiere las características químicas propias ( $\alpha$ -hélices y  $\beta$ -láminas).

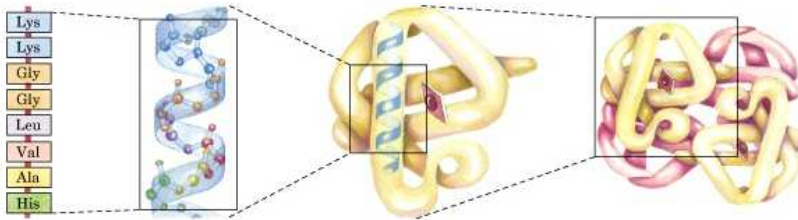


Figura 9: Estructuras primaria, secundaria ( $\alpha$ -hélice), terciaria y cuaternaria de una proteína.

Además, las proteínas se pliegan sobre sí mismas en una estructura tridimensional compacta (estructura terciaria) que en su superficie tiene varios grupos reactivos. Son estos grupos los que precisamente se juntan a otras moléculas y actúan como **enzimas** catalizando las reacciones químicas en las que se rompen o se crean **enlaces covalentes** –enlaces formados cuando varios átomos comparten electrones–. Finalmente, varias proteínas se pueden juntar (estructura cuaternaria) formando moléculas compactas (ver figura 9).

De esta forma las proteínas *dirigen* casi todos los procesos químicos de la célula, en particular la propia síntesis del ADN, la del ARN y la de ellas mismas, es decir tenemos un esquema de retroalimentación. En ello se basa el dogma central de la biología (ver figura 10) que consiste en la creencia de que el ADN

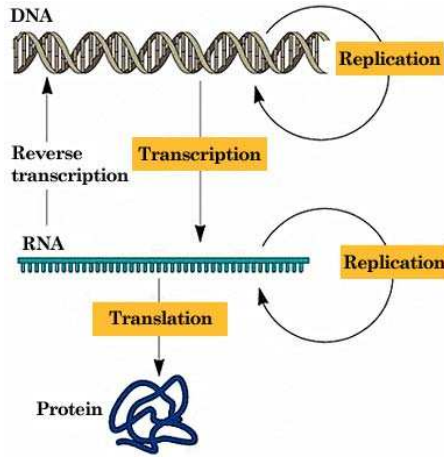


Figura 10: El dogma central de la biología molecular.

se transcribe en ARN y éste se traduce en proteínas, encargadas de la mayoría de las funciones biológicas, pero no a la inversa.

#### 2.4 ¿Cómo se traduce el ARN en proteínas?

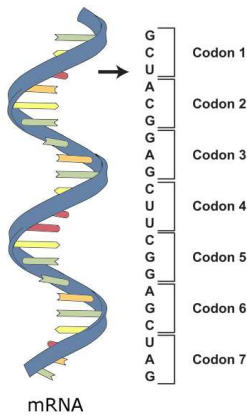


Figura 11: Los codones del mRNA

El proceso de traducción del alfabeto de 4 letras del ADN al de 20 de las proteínas es sumamente complicado por lo que sólo daremos un breve bosquejo del mismo.

Lo primero que hace el ADN es transcribir una parte de él en una cadena simple de ARN denominado ARN mensajero o mRNA que contiene la información de la proteína a sintetizar. La información contenida en el mRNA se lee en grupos de tres nucleótidos: el triplete o codón (ver figura 11). Cada codón codifica un aminoácido. Como tenemos  $4 \times 4 \times 4 = 64$  posibles codones y sólo

20 aminoácidos, tendremos varios codones distintos que codifican al mismo aminoácido.

El segundo paso comienza cuando al mRNA se le une un ARN muy corto conocido como ARN de transferencia o tARN. La estructura del tARN es tal

que en un extremo (físico) tiene un aminoácido, y, en el otro, un anticodón, de forma que el codón del mRNA se junta con el anticodón del tARN y se va formando la secuencia de proteínas. En la figura 12 se muestra un esquema del proceso. Resumiendo, una molécula de ADN contiene la codificación de miles

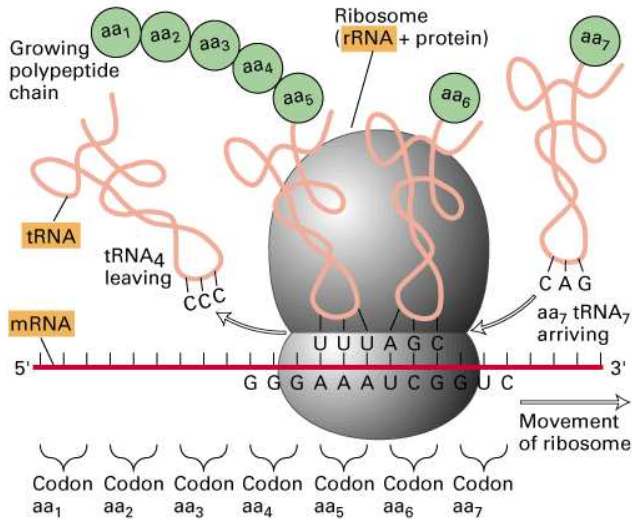


Figura 12: Síntesis de proteínas: el mRNA y el tARN

de proteínas. Esta información se traslada a un mRNA que es quien codifica la proteína al juntarse de una única forma con los tARN en cuyos extremos *vian* los nucleótidos (aminoácidos) de la correspondiente proteína.

## 2.5 ¿Qué es un gen?

Ya estamos en condiciones de definir el **gen**. Un **gen** es un fragmento o secuencia del ADN que corresponde a una determinada proteína, o a una molécula de ARN (ver figura 13). Además, se dice que un gen se *expresa* o está *expresado* si la proteína o ARN que dicho gen codifica está presente en la célula en un determinado instante de tiempo.

Lo sorprendente es que el ADN no sólo sintetiza las proteínas, sino que también regula el momento de tiempo cuando dicha proteína debe ser sintetizada. Es decir, la célula, mediante su ADN, ajusta la velocidad de transcripción y traducción de los distintos genes de forma distinta en función de sus necesidades. Lo anterior implica que, aparte de los genes, en el ADN tienen que haber secuencias no codificantes. Estas regiones se denominan ADN regulador –secuencias que regulan la velocidad de transcripción–, y otras regiones que funcionan como signos de puntuación: indican el inicio y final de la *información* para la síntesis de una proteína dada.

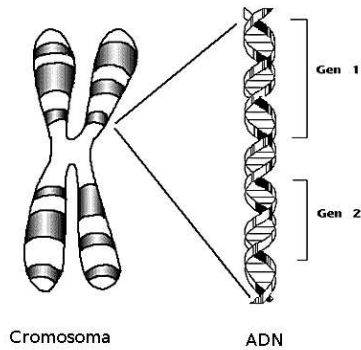


Figura 13: Un gen es una secuencia del ADN que codifica una proteína o ARN.

Así, la totalidad de la información genética incluida en la secuencia completa de ADN, denominada **genoma** de una célula, dicta no sólo qué moléculas han de sintetizarse, sino dónde y cuándo.

## 2.6 El ciclo celular

La célula de cualquier organismo adulto es un sistema en un perfecto estado de equilibrio. El proceso de transcripción del ADN en mRNA y de estos en proteínas está extremadamente bien regulado de forma que la célula no cambia ni su tamaño ni su función. Esta visión aparentemente estática de la célula es, sin embargo, errónea. De hecho, las células “nacen”, se “desarrollan” y “mueren” como cualquier organismo vivo. Una célula nueva se origina de dos maneras: por fusión –como el espermatozoide y el ovocito– o por división celular. En ambos casos se inicia un complejo programa de replicación celular codificado por el ADN y ejecutado por las correspondientes proteínas. Aquí nos centraremos en el segundo proceso: la división celular.

El crecimiento y división celular es un proceso muy bien regulado que es consecuencia de una necesidad del organismo. Por ejemplo, el crecimiento de las células musculares como consecuencia del ejercicio físico o el aumento de los eritrocitos debido a las grandes alturas. Si este proceso falla pueden producirse alteraciones muy peligrosas. Así, si las células empiezan a dividirse y crecer sin necesidad, pueden terminar formando tumores malignos o **cánceres**.

Antes de discutir las bases de esta temida enfermedad, veamos brevemente cómo está constituido el ciclo celular (ver figura 14).

El ciclo celular eucariota consta de dos fases fundamentales. La fase *S* (de síntesis), que dura unas 12 horas en un mamífero, y la *M* de mitosis, con una duración de menos de una hora. A estas dos fases se le intercalan dos periodos de *descanso* que le permiten a la célula crecer: La fase  $G_1$  entre *M* y *S* y la  $G_2$  entre *S* y *M*. Usualmente a las fases  $G_1$ , *S*, y  $G_2$  se les denomina **interfase**. La

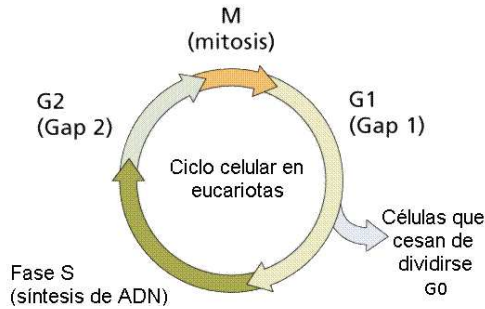


Figura 14: El ciclo celular eucariota.

interfase tiene una duración total de 23 horas en un mamífero. Además de estas fases, muchas células, después de la mitosis, entran en un estado latente o de reposo llamado  $G_0$ , que puede durar días e incluso años.

## 2.7 Cáncer

El cuerpo de un organismo vivo pluricelular, e.g. el hombre, se comporta como una sociedad o ecosistema cuyos miembros son células de distintos tipos y funciones que se reproducen por división celular y que están organizadas en tejidos. Lo anterior indica que, para su descripción, podemos aplicar las técnicas de dinámica de poblaciones que engloban tanto los nacimientos y muertes celulares, como los hábitats, límites territoriales, tamaño de las poblaciones, etc. También, como cualquier sistema vivo, en estas poblaciones tiene lugar el proceso de selección natural.

En general, todas las células del organismo están comprometidas en un equilibrio en el cual se  *cuentan*  unas a otras –mediante distintas señales químicas que emiten e interpretan– lo que tienen que hacer en cada momento: quedarse en estado latente  $G_0$ , dividirse  $M$ , diferenciarse, o morir (*apoptosis*). Las alteraciones moleculares que rompen esta armonía pueden ser extremadamente perjudiciales para el organismo. Por ejemplo, en el cuerpo humano hay unas  $10^{14}$  células y ocurren  $10^9$  mutaciones diarias. Normalmente la célula tiene un mecanismo de control y reparación de los daños de su ADN; el problema aparece cuando algunas de esas mutaciones proporcionan a la célula una ventaja evolutiva frente a sus compañeras del tejido permitiendo, por ejemplo, que se dividan más deprisa, y creando un clúster de células *anómalas* en crecimiento. Si además estos clústers empiezan a proliferar a expensas de las células vecinas pueden terminar por destruir toda la *sociedad celular*. Cuando esto ocurre estamos en presencia de un **cáncer**.

Así pues, el **cáncer** consiste en el crecimiento descontrolado y la diseminación de células anómalas en el organismo, que invaden y dañan tejidos y órganos.

Debemos hacer hincapié aquí en que el **cáncer** **no** es una única enfermedad, sino un grupo de al menos 100 enfermedades distintas aunque relacionadas,



a menudo con causas diferentes. La aparición de un cáncer no se debe a un único factor sino a la combinación de varios factores que se engloban en dos grupos: la herencia genética y el ambiente. Se ha demostrado que la herencia de versiones anormales de algunos genes es responsable de la predisposición a padecer algunos tipos de cáncer.

Las células cancerosas se caracterizan fundamentalmente por dos propiedades hereditarias:

1. Se reproducen a pesar de las restricciones normales (e.g. ignoran las señales celulares para que no ocurra la división celular).
2. Invaden y colonizan territorios reservados a otras células.

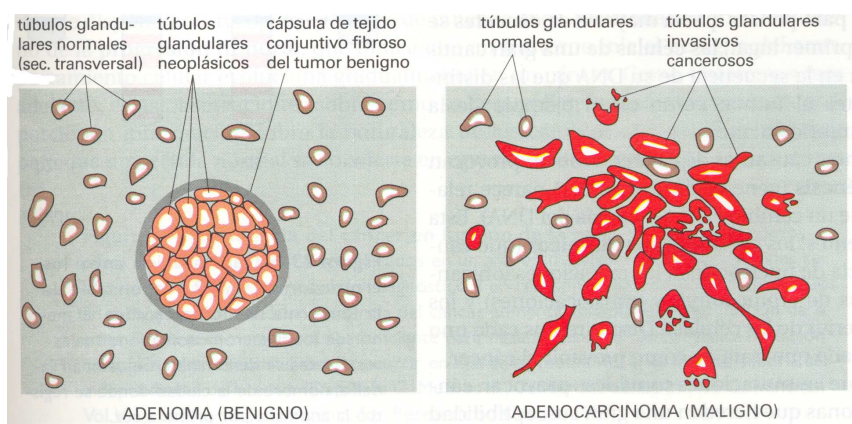


Figura 15: Tipos de tumor.

Si una célula normal no prolifera más que sus vecinas no se produce ningún daño sustancial, pero si su proliferación está fuera de control entonces se produce un tumor o neoplasma –masa de células anómalas que crece de forma inexorable–. Si estas células anormales se mantienen agrupadas en una única masa, usualmente limitada por una membrana de tejido conjuntivo fibroso, se dice que el tumor es benigno (se suele extirpar y problema resuelto). El verdadero problema surge cuando las células del tumor invaden el tejido circundante, en particular, si son capaces de liberarse y viajar por los vasos sanguíneos o linfáticos y formar tumores secundarios en otras partes del organismo (ver figura 15).

Así, la carcinogénesis o aparición de un cáncer es el resultado de dos procesos sucesivos:

1. El aumento descontrolado de la proliferación de un grupo de células que da lugar a un tumor o neoplasia.

2. La posterior adquisición por estas células de capacidad invasiva, que les permite diseminarse desde su sitio natural en el organismo y colonizar y proliferar en otros tejidos u órganos (proceso conocido como **metástasis**).

Este último es precisamente el más peligroso y el causante de que un cáncer sea mortal.

Cuando un tumor es detectable ha adquirido ya un tamaño considerable (alrededor de 0,5 cm de diámetro para su detección por rayos X, y de 1 cm por palpación) y está constituido por un elevado número de células (de 100 a 1.000 millones). Existen muchas pruebas científicas que indican que la mayoría de los tumores, benignos o malignos, derivan de una sola célula. Para que finalmente aparezca el tumor, los descendientes de esa célula han de continuar cambiando de forma que, para que alguna de ellas se vuelva cancerígena, han de ocurrir una enorme cantidad de alteraciones y mutaciones adicionales. Todo esto suele ocurrir a lo largo de muchos años. Hoy día se piensa que desde el inicio del proceso hasta que un cáncer puede ser diagnosticado transcurren una o más décadas. Así, el proceso de progresión tumoral se compone de cambios genéticos (mutaciones) y selección progresiva de células cada vez más anormales en su crecimiento y comportamiento, adquiriendo la capacidad de invadir el tejido circundante y, posteriormente, de originar metástasis.

## 2.8 Bioinformática

Como hemos visto a lo largo de esta sección, el ADN contiene toda la información necesaria para el correcto funcionamiento de la célula. Es por ello que no es de extrañar que se haya invertido tanto esfuerzo y dinero en su secuenciación, es decir, en descubrir el orden en el que se encuentran sus bases. Este es el punto de partida para luego identificar los correspondientes genes, lo cual tendría repercusiones inimaginables.

A partir de 1955 los científicos comienzan a disponer de las primeras secuencias completas de algunas proteínas: la hormona peptídica *insulina* [28], o la enzima *ribonucleasa* [15]. En 1965 ya se conocían las secuencias de unas 20 proteínas; en 1980, de unas 1500, en 1999 de 300000; en el 2006 unas 610000 (según la base de datos NCBI<sup>2</sup>). La situación se complica si queremos secuenciar el ADN, pues las cadenas de ADN son mucho más largas –la del hombre, por ejemplo, tiene entre  $5,5 \times 10^7$  y  $25 \times 10^7$  pares de bases (una proteína tiene muchas menos bases  $\sim 2000$ )–. A lo anterior hay que añadir que es técnicamente imposible secuenciar (al menos hoy día) una cadena tan larga, por lo que en la práctica hay que romperla en cadenas más cortas (usualmente de unos 500 pares de bases) para luego secuenciarlas y finalmente recomponer la cadena original.

Todo este proceso genera una ingente cantidad de datos (las secuencias de bases, o cadenas de letras, como la que se muestra en la figura 16) que hay que

---

<sup>2</sup>Véase la página web <http://www.ncbi.nlm.nih.gov/gquery/>.

analizar, cotejar, almacenar, etc., por lo que se hace imprescindible el uso de potentes ordenadores y avanzadas técnicas informáticas.

```

1   cctcgacgtc gacgttaacg gtaccgagct tgtggcagtt taaggcgggc gtcctgcccc
61  ccaccctccg ggccgttgct tcgcaacggt caaatccgct cccggcggat ttgtcctact
121 caggagagcg ttcaccgaca aacaacagat aaaacgaaag gcccagcttt tcgactgagc
181 ctttcgtttt atttgatgct tcaagctcgg tacctcgagg gaattccgga tccggcatca
241 tggttcgacc attgaactgc atcgtcgcg  tgtcccaaaa tatggggatt ggcaagaacg
301 gagacctacc ctggcctccg ctcaggaacg agttcaagta cttccaaaga atgaccacaa
361 cctcttcagt ggaaggtaaa cagaatctgg tgattatggg taggaaaacc tggttctcca
421 ttcctgagaa gaatcgacct ttaaaggaca gaattaatat agttctcagt agagaactca
481 aagaaccacc acgaggagct cttttcttg  ccaaaagttt ggatgatgcc ttaagactta
541 ttgaacaacc ggaattggca agtaaagtag acatggtttg gatagtcgga ggcagttctg
601 ttaccagga agccatgaat caaccaggcc accttagact ctttgtagaca aggatcatgc
661 aggaatttga aagtgcacg  tttttccag aaattgattt ggggaaatat aaacttctcc
721 cagaataccc aggcgtcctc tctgaggtcc aggaggaaaa aggcatcaag tataagtttg
781 aagtctacga gaagaagac taacaggaag atgctttcaa gttctctgct cccctctaa
841 agctatgcat ttttataaga ccatgggact tttgctggct ttagatccgg ccaagcttgg
901 actcctgttg atagatccag taatgacctc agaactccat ctggatttgt tcagaacgct
961 cggttgccgc cgggcgtttt ttattggtga gaatccaaag cttggcgaga ttttcaggag

```

Figura 16: Fragmento de la secuencia del plásmido pDS3 (primeras 1020 bases)

Esta necesidad ha llevado a la aparición de una nueva ciencia: la Bioinformática, término acuñado a mitad de los 80 (del siglo XX) para definir el conjunto de técnicas y aplicaciones informáticas (y ordenadores) usadas en las ciencias biológicas (ver figura 17).

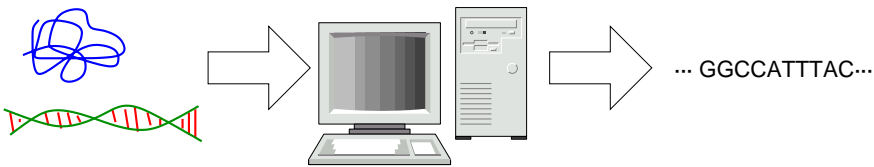


Figura 17: ¿Qué es la bioinformática?

En particular, existen en la WWW distintas bases de datos de libre acceso que contienen gran cantidad de información: secuencias completas de proteínas, genomas de algunas especies, etc. Algunas de ellas son:

1. <http://www.ncbi.nlm.nih.gov/> (The National Center for Biotechnology Information, Estados Unidos).
2. <http://www.ebi.ac.uk/Databases/> (The European Bioinformatics Institute, Reino Unido).

3. <http://au.expasy.org/> (The ExPASy –Expert Protein Analysis System–, Suiza).
4. <http://www.rcsb.org/pdb/> (The Research Collaboratory for Structural Bioinformatics, Estados Unidos).
5. <http://www.ensembl.org/> (The Ensembl Genome Browser).

Una magnífica introducción en castellano a la Bioinformática se puede encontrar en la WWW

[http://bvs.isciii.es/bib-gen/Actividades/curso\\_virtual/esquema.htm](http://bvs.isciii.es/bib-gen/Actividades/curso_virtual/esquema.htm)

Algunos textos interesantes sobre Bioinformática son [2, 4, 26].

### 3 Modelos matemáticos de la dinámica interna del ADN

De la discusión anterior es *fácil* adivinar que, modelos realistas que describan la dinámica interna del ADN y permitan entender los procesos de transcripción, etc. son prácticamente inviables. Por ejemplo, un modelo cuántico realista debería incluir las interacciones de miles de millones de partículas (los átomos), con el agravante de que no está claro el potencial de interacción para describir los enlaces por puentes de hidrógeno, por citar alguna de las muchas interacciones internas que tienen lugar en dicha molécula. Es por ello que vamos a considerar algunos modelos sencillos y los problemas matemáticos que se derivan de ellos. La referencia básica para este apartado es [31].

Antes de comenzar debemos recordar que la molécula de ADN está formada por dos cadenas muy largas enrolladas una sobre la otra en forma de doble hélice. La estructura de las cadenas es bastante regular y contiene cuatro bases colocadas con gran precisión. Esta estructura es muy similar a la de los *casicristales* unidimensionales [8], pero, a diferencia de éstos, el ADN es *bastante flexible*.

#### 3.1 Modelos lineales

##### 3.1.1 Una única cadena

Supongamos que tenemos una cadena simple de ADN (o ARN). Un modelo sencillo consiste en cambiar las bases por discos (cilindros) y suponer que la interacción de las mismas es una interacción armónica. Consideraremos únicamente los movimientos longitudinales y rotacionales de los discos.

Sea  $u_n(t)$  el desplazamiento longitudinal del  $n$ -ésimo disco de su posición de equilibrio en el instante de tiempo  $t$  y sea  $\varphi_n(t)$  el correspondiente desplazamiento angular (ver figura 18). Por simplicidad asumiremos que tanto las masas (y los momentos de inercia, respectivamente) como las constantes elásticas y de torsión son las mismas (es decir, una cadena homogénea). Entonces, si suponemos que las amplitudes de las oscilaciones internas son

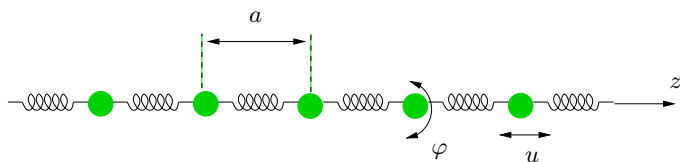


Figura 18: Modelo lineal de una cadena: movimientos longitudinal y angular.

pequeñas, podemos escribir el Hamiltoniano del sistema en la aproximación armónica, es decir, en la forma<sup>3</sup>  $H = H_l + H_r + H_{l-r}$ , donde

$$H_l = \sum_{n=1}^N \frac{1}{2} m \dot{u}_n^2 + \frac{1}{2} k (u_{n+1} - u_n)^2, \quad (5)$$

$$H_r = \sum_{n=1}^N \frac{1}{2} I \dot{\varphi}_n^2 + \frac{1}{2} \varkappa (\varphi_{n+1} - \varphi_n)^2,$$

siendo  $N$  el número total de monómeros de la cadena,  $m$  la masa del disco,  $I$  su momento de inercia,  $k$  la constante elástica,  $\varkappa$  la constante de torsión, y donde  $H_{l-r}$  representa el Hamiltoniano de la interacción entre ambos movimientos. Usualmente se imponen condiciones de frontera periódicas  $u_n = u_{n+N}$ ,  $\varphi_n = \varphi_{n+N}$ . En primera aproximación se puede suponer que  $H_{l-r}$  es constante.

Bajo estas restricciones, las ecuaciones de movimiento son

$$m \ddot{u}_n - k(u_{n+1} - 2u_n + u_{n-1}) = 0, \quad (6)$$

$$I \ddot{\varphi}_n - \varkappa(\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) = 0.$$

El sistema anterior es un sistema lineal de  $2N$ -ecuaciones diferenciales desacopladas (independientes). Las soluciones típicas del sistema anterior son las ondas viajeras planas

$$u_n(t) = u_{0n} e^{i(qna - \omega_l t)}, \quad \varphi_n(t) = \varphi_{0n} e^{i(qna - \omega_r t)}, \quad (7)$$

donde  $q$  es el número de onda y  $a$  es la distancia entre dos discos cualesquiera (que suponemos constante a lo largo de la cadena). Sustituyendo (7) en (6) obtenemos las relaciones de dispersión para las frecuencias longitudinales y rotacionales

$$\omega_l = \frac{2k}{m} (1 - \cos qa), \quad \omega_r = \frac{2\varkappa}{I} (1 - \cos qa).$$

Veamos el límite continuo de las ecuaciones anteriores, es decir pasemos del modelo discreto de discos a un modelo continuo tipo cuerda elástica (ver e.g. [14]). Comenzaremos reescribiendo la primera de las ecuaciones (6) en la forma

$$\frac{m}{a} \ddot{u}_n = (ka) \frac{u_{n+1} - u_n}{a^2} - (ka) \frac{u_n - u_{n-1}}{a^2}.$$

<sup>3</sup>En adelante, por  $\dot{f}$  denotaremos la derivada con respecto al tiempo de  $f$ .

Tomemos  $a \rightarrow 0$  y denotemos por  $\rho$  la densidad de masa por unidad de longitud; entonces, teniendo en cuenta que

$$u_n(t) = u(na, t) \rightarrow u(z, t),$$

tenemos

$$\frac{m}{a} \rightarrow \rho, \quad ka \rightarrow K, \quad \frac{u_{n+1} - u_n}{a^2} - \frac{u_n - u_{n-1}}{a^2} \rightarrow \frac{\partial^2}{\partial z^2} u(z, t) = u_{zz}(z, t),$$

luego obtenemos la ecuación de ondas

$$\rho u_{tt} = K u_{zz}, \text{ esde}$$

cuya solución es  $u(z, t) = u_0 e^{i(qz - \omega_l t)}$ , que sustituida en la ecuación anterior nos da

$$\omega_l = \sqrt{\frac{K}{\rho}} q.$$

Para la segunda ecuación obtenemos, de forma análoga,

$$\iota \varphi_{tt} = \mathcal{K} \varphi_{zz},$$

cuya solución es

$$\varphi(z, t) = \varphi_0 e^{i(qz - \omega_r t)}, \quad \omega_r = \sqrt{\frac{\mathcal{K}}{\iota}} q,$$

siendo  $\iota$  la densidad lineal del momento de inercia y  $\mathcal{K} = \lim_{a \rightarrow 0} \kappa a$ . En ambos casos el número de onda se encuentra en el intervalo  $[-\pi/a, \pi/a]$ .

De lo anterior se deduce que a través de la cadena se desplaza una onda viajera cuyas velocidades longitudinal y rotacional son

$$v_l = \frac{\partial \omega_l}{\partial q} = \sqrt{\frac{K}{\rho}}, \quad v_r = \frac{\partial \omega_r}{\partial q} = \sqrt{\frac{\mathcal{K}}{\iota}},$$

respectivamente.

El modelo anterior tiene una clara deficiencia: No tiene en cuenta el movimiento perpendicular al eje  $z$  (ver figura 19). Si tenemos en cuenta este

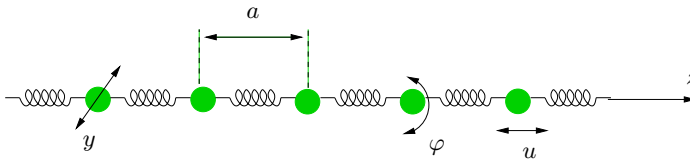


Figura 19: Modelo lineal de una cadena: movimientos longitudinal, angular y transversal.

movimiento se puede probar [31, §4.1.1.3] que el desplazamiento perpendicular

al eje  $z$ , que denotaremos por  $y$ , satisface la ecuación  $S\rho y_{tt} = -K\iota y_{zzzz}$ , donde  $S$  es el área de la base del cilindro (disco). Como ejercicio dejamos al lector que deduzca la expresión de la relación de dispersión y la velocidad de onda de las correspondientes ondas transversales.

Otra importante deficiencia del modelo es que la energía no está localizada ya que la solución onda plana corresponde a una energía uniformemente distribuida a lo largo del medio (en este caso la cadena de ADN). Por tanto, si pretendemos que el modelo explique el fenómeno de la transcripción, por ejemplo, es esencial que la onda viajera se propague de forma muy localizada para que la energía se localice en unos pocos pares de bases. A lo anterior hay que agregarle otra importante deficiencia y es que, para romper la doble cadena, se necesita además que la amplitud de la onda sea lo suficientemente grande, lo cual está en abierta contradicción con la suposición inicial de que podemos aproximar las interacciones mediante términos armónicos. Estas deficiencias se resuelven parcialmente en los modelos no lineales que discutiremos posteriormente.

### 3.1.2 Una cadena doble

Supongamos que tenemos una doble cadena de ADN. Un modelo sencillo consiste nuevamente en cambiar las bases por discos y suponer que tanto la interacción entre los nucleótidos como la de las bases de ambas cadenas son interacciones armónicas. Consideraremos los movimientos longitudinales  $u$ , transversales  $y$  y rotacionales  $\varphi$  de los discos. En este caso, el Hamiltoniano se

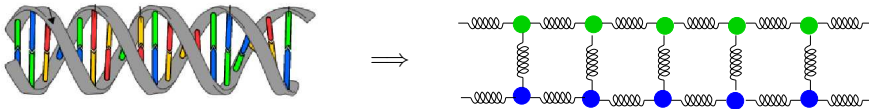


Figura 20: Modelo lineal de una doble cadena.

escribe como

$$H = H_l^{(1)} + H_l^{(2)} + H_r^{(1)} + H_r^{(2)} + H_t^{(1)} + H_t^{(2)} + H^{(1-2)},$$

donde  $H_l^{(j)}$ ,  $H_r^{(j)}$  y  $H_t^{(j)}$ ,  $j = 1, 2$ , representan los movimientos longitudinales, rotacionales y transversales de cada una de las dos cadenas, respectivamente, y  $H^{(1-2)} = H_l^{(1-2)} + H_r^{(1-2)} + H_t^{(1-2)}$  es la suma de las interacciones entre ambas cadenas asociadas a los correspondientes movimientos longitudinales, rotacionales y transversales.

Para los dos primeros Hamiltonianos tenemos una expresión análoga a (5),

de forma que

$$\begin{aligned} H_l^{(j)} &= \sum_{n=1}^N \frac{1}{2} m \dot{u}_{n,j}^2 + \frac{1}{2} k (u_{n+1,j} - u_{n,j})^2, \\ H_t^{(j)} &= \sum_{n=1}^N \frac{1}{2} m \dot{y}_{n,j}^2 + \frac{1}{2} b (y_{n+1,j} - y_{n,j})^2, \\ H_r^{(j)} &= \sum_{n=1}^N \frac{1}{2} I \dot{\varphi}_{n,j}^2 + \frac{1}{2} \kappa (\varphi_{n+1,j} - \varphi_{n,j})^2. \end{aligned}$$

La interacción  $H^{(1-2)}$  entre cadenas puede modelarse, en primera aproximación, mediante los términos

$$H^{(1-2)} = \sum_{n=1}^N \frac{\alpha}{2} (u_{n,1} - u_{n,2})^2 + \sum_{n=1}^N \frac{\beta}{2} (\varphi_{n,1} - \varphi_{n,2})^2 + \sum_{n=1}^N \frac{\gamma}{2} (y_{n,1} - y_{n,2})^2,$$

donde  $\alpha$ ,  $\beta$  y  $\gamma$  son ciertas constantes. Así, obtenemos un sistema de seis ecuaciones lineales acopladas dos a dos:

$$\begin{cases} m \ddot{u}_{n,1} - k(u_{n+1,1} - 2u_{n,1} + u_{n-1,1}) - \alpha(u_{n,2} - u_{n,1}) = 0, \\ m \ddot{u}_{n,2} - k(u_{n+1,2} - 2u_{n,2} + u_{n-1,2}) - \alpha(u_{n,1} - u_{n,2}) = 0, \\ I \ddot{\varphi}_{n,1} - \kappa(\varphi_{n+1,1} - 2\varphi_{n,1} + \varphi_{n-1,1}) - \beta(\varphi_{n,2} - \varphi_{n,1}) = 0, \\ I \ddot{\varphi}_{n,2} - \kappa(\varphi_{n+1,2} - 2\varphi_{n,2} + \varphi_{n-1,2}) - \beta(\varphi_{n,1} - \varphi_{n,2}) = 0, \\ m \ddot{y}_{n,1} - b(y_{n+1,1} - 2y_{n,1} + y_{n-1,1}) - \gamma(y_{n,2} - y_{n,1}) = 0, \\ m \ddot{y}_{n,2} - b(y_{n+1,2} - 2y_{n,2} + y_{n-1,2}) - \gamma(y_{n,1} - y_{n,2}) = 0. \end{cases} \quad (8)$$

Nuevamente buscamos las soluciones de tipo onda plana. Aquí restringiremos nuestro análisis al primero de los sistemas acoplados de (8),

$$u_{n,j}(t) = u_{0j} e^{i(qna - \omega_{1,j}t)}.$$

Sustituyéndolas en (8) obtenemos las relaciones de dispersión

$$\omega_{1,1} = \sqrt{\frac{4k}{m} \sin^2(qa/2)}, \quad \omega_{1,2} = \sqrt{\frac{4k \sin^2(qa/2) + 2\alpha}{m}}.$$

Si, como en el caso anterior, tomamos el límite cuando  $a \rightarrow 0$  (límite continuo) obtenemos las ecuaciones

$$\rho u_{1tt} = K u_{1zz} + \alpha(u_2 - u_1), \quad \rho u_{2tt} = K u_{2zz} + \alpha(u_1 - u_2).$$

Sus soluciones tipo onda plana son

$$u_1(z, t) = u_{01} e^{i(qz - \omega_{1,1}t)}, \quad u_2(z, t) = u_{02} e^{i(qz - \omega_{1,2}t)},$$



que nos conducen a las relaciones de dispersión:

$$\omega_{l,1} = \sqrt{\frac{ka^2q^2}{m}}, \quad \omega_{l,2} = \sqrt{\frac{ka^2q^2 + 2\alpha}{m}}.$$

Para los otros dos sistemas de ecuaciones los cálculos son similares y los omitiremos (se dejan como ejercicio al lector).

**Limitaciones.**

Esta claro que estos son los modelos más simples y por tanto tienen un sinnúmero de limitaciones. Esencialmente hemos considerado en la cadena de ADN los nucleótidos como una única partícula, cuando en realidad son mucho más complicados ya que constan de un azúcar, un grupo fosfato y una base heterocíclica de carbono (5-carbono). Por lo tanto, un modelo más realista debería tener en cuenta esta estructura. En la figura 21 se muestra un modelo donde se separa el esqueleto de carbón-fosfato (rectángulo), el azúcar (círculo intermedio –claro–) y la base (círculo exterior –oscuro–). Mediante líneas continuas se representan las interacciones internas de cada cadena y por la discontinua la interacción entre cadenas por puente de hidrógeno. Además, como ya mencionamos; la energía no está localizada, por lo que no es posible utilizar modelos lineales para dar una explicación del proceso de transcripción, por citar un ejemplo.

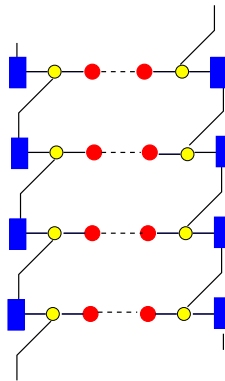


Figura 21: Modelo lineal de una doble cadena: segunda aproximación.

Denotemos, como antes, los movimientos longitudinales, transversales y rotacionales de las bases por  $u$ ,  $y$  y  $\varphi$ , respectivamente; y denotemos por  $v$  los transversales entre el azúcar y el grupo fosfato. Suponiendo que podemos usar la aproximación armónica podemos escribir el Hamiltoniano de la forma

$$H = \sum_{i=1,2} \left( H_l^{(j)} + H_t^{(j)} + H_r^{(j)} + H_{fa}^{(j)} \right) + H_l^{(1-2)} + H_t^{(1-2)} + H_r^{(1-2)} + H_{fa}^{(1-2)},$$

donde los términos son similares a los de antes pero ahora  $H_{fa}^{(j)}$  representa la nueva interacción fosfato-azúcar. Como ejercicio se deja al lector encontrar las ecuaciones de movimiento en este caso, así como las relaciones de dispersión. El modelo más exacto corresponde a la estructura atómica real de ADN (ver figura 22).

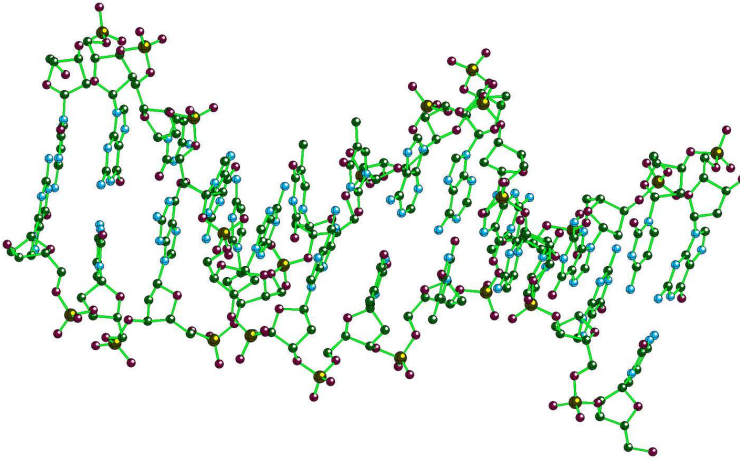


Figura 22: Modelo lineal más exacto de una doble cadena de ADN.

En todos estos casos también hay que tener en cuenta que el Hamiltoniano se está considerando en su aproximación armónica, es decir cuando las oscilaciones de las correspondientes partículas son suficientemente pequeñas. Nuevamente esto constituye una limitación para explicar la apertura o rotura de la doble cadena de ADN. Si estas amplitudes no son pequeñas, entonces entran a jugar un papel fundamental los efectos anarmónicos (o no lineales). Este tipo de modelos no lineales son de enorme interés como veremos en la próxima sección. Finalmente, debemos destacar que los Hamiltonianos de interacción aquí descritos no corresponden a los reales, los cuales son desconocidos.

### 3.2 Modelos no lineales

De entre los muchos modelos no lineales vamos a tratar aquí los más sencillos. Incluso en estos modelos veremos las dificultades matemáticas que aparecen cuando queremos hacerlos más *realistas*. Nuevamente la referencia básica será [31].

#### 3.2.1 Modelo de Englander et al. [13].

Vamos a comenzar por un modelo mecánico muy sencillo propuesto en 1980 por S.W. Englander et. al. para describir los estados abiertos del ADN (es decir, los estados cuando están rotos los enlaces de puente de hidrógeno entre

las bases de las cadenas del ADN). La idea de Englander et al. se basa en el modelo mecánico de Scott, que consiste en modelar el esqueleto azúcar-fosfato por una cadena de osciladores y las bases por péndulos (ver figura 23). Por simplicidad se considerarán sólo los movimientos giratorios, despreciando el resto. Además, supondremos que la aportación de la segunda cadena consiste en crear un potencial efectivo  $V_e$  sobre la primera.

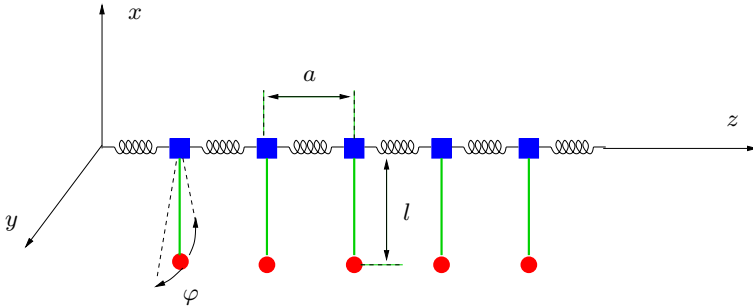


Figura 23: Modelo mecánico de una doble cadena de ADN.

Así, tenemos una cadena de péndulos que pueden rotar en el plano  $xy$  (ver la figura 24).

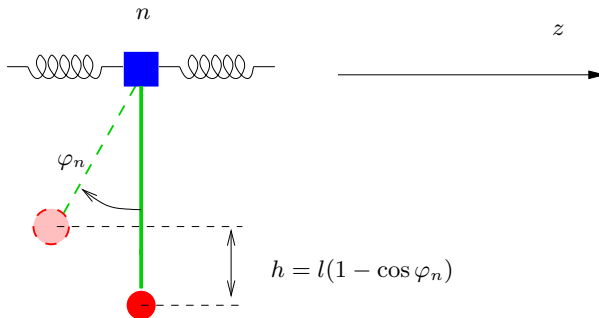


Figura 24: Movimiento de rotación en el modelo mecánico de una doble cadena de ADN.

Sea  $I$  el momento de inercia de las bases y  $\varkappa$  el coeficiente de torsión de los nucleótidos. Entonces el Hamiltoniano del sistema es

$$H = \sum_n \frac{I}{2} \dot{\varphi}_n^2 + \sum_n \frac{\varkappa l^2}{2} (\varphi_{n+1} - \varphi_n)^2 + V_e(\varphi_1, \dots, \varphi_N).$$

Las ecuaciones del movimiento son

$$I \ddot{\varphi}_n = \varkappa l^2 (\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) + \frac{\partial V_e}{\partial \varphi_n}. \tag{9}$$

Englander et. al. consideraron el caso más sencillo, en el que  $V_e = -mgh$ , i.e., un campo análogo al campo gravitatorio. Así, tenemos para  $\varphi_n$  la ecuación

$$I\ddot{\varphi}_n = \varkappa l^2(\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}) - mgl \sin \varphi_n.$$

Si ahora suponemos que la distancia  $a$  entre los nucleótidos es lo suficientemente pequeña, la ecuación difero-diferencial (9) se puede aproximar por la ecuación en derivadas parciales

$$I\varphi_{tt} = \varkappa a^2 l^2 \varphi_{zz} + \frac{\partial V_e}{\partial \varphi},$$

que, en variables *adimensionales*, tiene la forma

$$\varphi_{TT} = \varphi_{ZZ} + \frac{\partial V_e}{\partial \varphi}. \quad (10)$$

Así pues, escogiendo el potencial de Englander et al., y teniendo en cuenta que  $I = ml^2$ , obtenemos la ecuación en derivadas parciales

$$\varphi_{tt} = \frac{\varkappa a^2}{m} \varphi_{zz} - \frac{g}{l} \sin \varphi \quad \Rightarrow \quad \varphi_{tt} = \varpi^2 \varphi_{zz} - \omega^2 \sin \varphi, \quad \varpi^2 = \frac{\varkappa a^2}{m}, \quad \omega^2 = \frac{g}{l},$$

o, equivalentemente, en variables *adimensionales*

$$\varphi_{ZZ} - \varphi_{TT} = \sin \varphi,$$

que no es más que la conocida ecuación de sine-Gordon o sG [29, §3.2].

De entre las muchas soluciones de esta ecuación destacan las denominadas kink y antikink [29, pág. 73] (ver figura 25)

$$\varphi(Z, T) = 4 \operatorname{arctg} \left[ \exp \left( \pm \frac{Z - VT}{\sqrt{1 - V^2}} \right) \right], \quad |V| < 1.$$

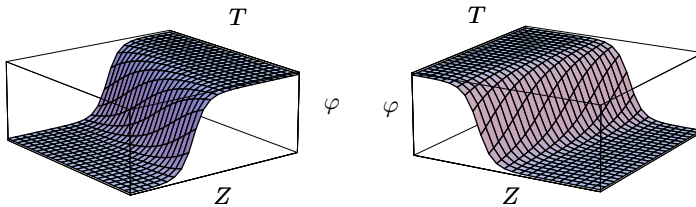


Figura 25: Soluciones kink y antikink de la ecuación sG.

Estas soluciones son ondas viajeras con una importante particularidad: su energía se localiza alrededor del centro del kink (antikink). Una interpretación de estos resultados es que las soluciones kink y antikink corresponden a la propagación de los estados abiertos en el ADN (ver figura 26).

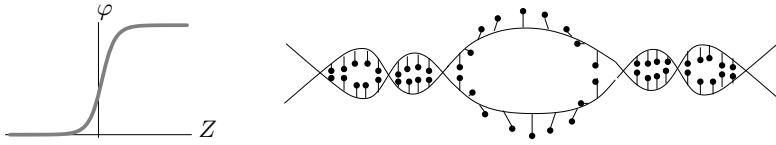


Figura 26: Interpretación del modelo mecánico de Englander et. al.

Imaginemos ahora que tenemos dos cadenas y supongamos que cada una de ellas genera un potencial efectivo  $V_e(\varphi_{n,j}) = mg_j l \cos \varphi_{n,j}$ ,  $j = 1, 2$ , sobre la otra. En este caso es fácil comprobar que las ecuaciones de movimiento son dos ecuaciones de sG desacopladas

$$\begin{cases} \varphi_{1tt} = \varpi_1^2 \varphi_{1zz} - \omega_1^2 \sin \varphi_1, \\ \varphi_{2tt} = \varpi_2^2 \varphi_{2zz} - \omega_2^2 \sin \varphi_2. \end{cases}$$

### 3.3 Otros modelos no lineales

Una limitación evidente de los dos modelos anteriores es que la interacción supuesta entre las dos cadenas sólo depende de una de las cadenas y no de ambas. Es por ello que se han introducido otras variantes del modelo anterior.

Por ejemplo, L. Yakushevich en [30] propone utilizar en (10) el potencial efectivo

$$V_e(\varphi_1, \varphi_2) = A(1 - \cos \varphi_1) + A(1 - \cos \varphi_2) - B[1 - \cos(\varphi_1 + \varphi_2)],$$

lo que le conduce a las siguientes dos ecuaciones acopladas cuya solución analítica en forma de onda solitaria (el kink y el antikink son casos particulares de estas ondas) es por el momento desconocida (algunos experimentos numéricos indican que semejante solución ha de existir):

$$\begin{cases} \varphi_{1TT} = \varphi_{1ZZ} - A \sin \varphi_1 + B \sin(\varphi_1 + \varphi_2), \\ \varphi_{2TT} = \varphi_{2ZZ} - A \sin \varphi_2 + B \sin(\varphi_1 + \varphi_2). \end{cases}$$

Por último, antes de terminar esta sección, debemos mencionar el modelo utilizado por Yomosa [32], quien propone el potencial efectivo

$$V_e(\varphi_1, \varphi_2) = A(1 - \cos \varphi_1) + A(1 - \cos \varphi_2) + B(1 - \cos \varphi_1 \cos \varphi_2).$$

Dejamos al lector como ejercicio que obtenga las correspondientes ecuaciones de movimiento.

Antes de terminar debemos hacer dos comentarios importantes. El primero es acerca de las limitaciones de estos modelos. En primer lugar, la afirmación de que el movimiento predominante en el proceso de transcripción es el rotatorio no es del todo correcta ya que las oscilaciones transversales (a lo largo del eje de

los enlaces puente de hidrógeno) son del mismo orden, por lo que un modelo más coherente tendría que tener en cuenta ambos movimientos. Para el caso de los desplazamientos transversales Peyrard y Bishop [24] introdujeron un modelo estadístico muy interesante que ha tenido bastante éxito (ver el interesante *survey* [25]).

El segundo comentario está relacionado con las interacciones consideradas. Modelos mecánicos más realistas (aun siendo simples) deben tener en cuenta todas las interacciones relevantes. Por tanto han de tener términos que modelen la interacción entre los movimientos longitudinales y transversales  $v_{l-t}$ , longitudinales y rotacionales  $v_{l-r}$ , e incluso términos que modelen los efectos térmicos (el ADN está en una disolución acuosa a cierta temperatura). Así, en general, tendremos ecuaciones no lineales de la forma

$$y_{TT} - y_{ZZ} = \left[ \begin{array}{c} \text{términos} \\ \text{no lineales} \end{array} \right] + v_{l-t}(y) + v_{l-r}(y) + \left[ \begin{array}{c} \text{efectos} \\ \text{térmicos} \end{array} \right].$$

El estudio de este tipo de ecuaciones constituye un reto para los matemáticos de hoy día.

## 4 Modelos de crecimiento tumoral

En esta última parte vamos a describir algunos modelos matemáticos de crecimiento tumoral con el objetivo de mostrar el tipo de problemas y técnicas matemáticas que suelen aparecer.

### 4.1 Modelo de un tumor cuerda estacionario

En la fase vascular de muchos tumores, las células cancerígenas rodean los capilares sanguíneos formando una estructura *cilíndrica* como la representada en la figura 27, donde el cilindro interior de radio  $r_0$  representa al vaso sanguíneo y el cilindro exterior de radio  $R$  al tumor.

Este tipo de tumores, conocidos como tumores cuerda, crece radialmente alejándose del centro de vaso. A medida que aumenta de tamaño, las células del borde se alejan del vaso y por tanto de su principal fuente de oxígeno y nutrientes, por lo que la mayoría entran en fase latente  $G_0$ . Esto conduce a un equilibrio entre el número de células activas (situadas en el interior del tumor, cerca del capilar) y las latentes (más alejadas del capilar); que conlleva a un equilibrio dinámico acorde con las observaciones clínicas para este tipo de tumores.

Vamos a describir un modelo matemático introducido por Dyson et. al. [11] que trata de explicar este tipo de comportamiento.

Supongamos que en el instante de tiempo  $t$  el tumor está compuesto por células activas dividiéndose y células en estado latente cuyos números denotaremos por  $p(x, r, t)$  y  $q(r, t)$ , respectivamente, donde  $r \in [r_0, R]$  es la distancia de la célula cancerígena al centro del vaso y  $x \in [0, 1]$  es la madurez de la célula ( $x = 0$  es una célula recién nacida y  $x = 1$  es una célula a punto

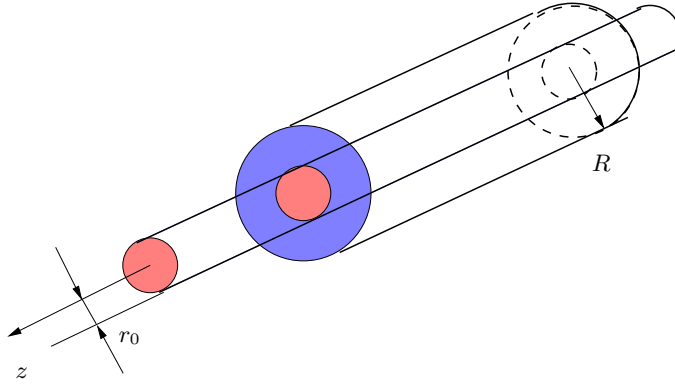


Figura 27: Modelo de un tumor cuerda alrededor de un vaso sanguíneo

de dividirse). Denotemos por  $\theta(r, t)$  la fracción de células que se dividen en la posición  $r$  (estamos asumiendo que el tumor tiene una simetría radial) en el momento  $t$  y por  $v(x)w(r)$  la fracción de células con madurez  $x$  en la posición  $r$  (el cociente de madurez). Como el cociente de madurez depende de  $r$ , la duración de los ciclos celulares a distintas distancias es, en principio, distinta. Un caso más sencillo consiste en considerar este cociente constante e independiente de  $r$  (todas las células tumorales tienen el mismo periodo de división celular). Sea  $u(r, t)$  el flujo radial de células tumorales a través de la superficie (que se asume independiente de la madurez).

Con todas estas suposiciones, las ecuaciones de evolución del tumor son las siguientes [11]: para  $t \geq 0$ ,

$$\begin{aligned} \frac{\partial}{\partial t} p(x, r, t) + \frac{\partial}{\partial x} (v(x)w(r)p(x, r, t)) + \frac{1}{r} \frac{\partial}{\partial r} (r u(r, t)p(x, r, t)) &= 0, \\ v(0)w(r)p(0, r, t) = 2\theta(r, t)v(1)w(r)p(1, r, t), \quad x \in [0, 1], \quad r \in [r_0, R], \\ p(x, r, 0) = p_0(x, r), \quad x \in [0, 1], \quad r \in [r_0, R], \\ \frac{\partial}{\partial t} q(r, t) + \frac{1}{r} \frac{\partial}{\partial r} (r u(r, t)q(r, t)) &= 2(1 - \theta(r, t))v(1)w(r)p(1, r, t), \\ q(r, 0) = \Psi(r), \quad r \in [r_0, R]. \end{aligned} \tag{11}$$

Supongamos que la densidad total de células en la posición  $r$  viene dada por

$$n(r, t) = \int_0^1 p(x, r, t) dx + q(r, t).$$

Entonces, de (11) se sigue que

$$\frac{\partial}{\partial t} n(r, t) + \frac{1}{r} \frac{\partial}{\partial r} (r u(r, t)n(r, t)) = v(1)w(r)p(1, r, t).$$

Una suposición habitual es que el flujo de células a través de la pared del capilar es despreciable (no existe), i.e.,  $u(r_0, t) = 0$  y  $n(r, t) \equiv N$ . Entonces  $\frac{\partial}{\partial t} n(r, t) = 0$

y la ecuación anterior se traduce en

$$r u(r, t) = \frac{1}{N} \int_{r_0}^r \zeta v(1) w(\zeta) p(1, \zeta, t) d\zeta.$$

Tal y como mencionamos, este tipo de tumores a partir de cierto tiempo  $t_e$  se encuentra en una fase de equilibrio (“steady state” o régimen permanente o estable) por lo que el problema de evolución (11) se transforma en el siguiente problema estacionario donde  $\nu(x, r) = p(x, r)/N$  representa la densidad de células activas en el estado estable del tumor:

$$\begin{aligned} \frac{\partial}{\partial x}(v(x)w(r)\nu(x, r)) + \frac{1}{r} \frac{\partial}{\partial r} \left( \nu(x, r) \int_{r_0}^r \zeta w(\zeta) v(1) \nu(1, \zeta) d\zeta \right) &= 0, \\ v(0)w(r)\nu(0, r) = 2\theta(r)v(1)w(r)\nu(1, r), \quad x \in [0, 1], \quad r \in [r_0, R]. \end{aligned} \quad (12)$$

Por comodidad introduciremos la función  $\varphi(x, r) = \nu(x, r)v(x)$ . De (12) se deduce que  $\varphi$  satisface la ecuación

$$\begin{aligned} \frac{\partial}{\partial x} \varphi(x, r) + \frac{1}{r w(r) v(x)} \frac{\partial}{\partial r} \varphi(x, r) \int_{r_0}^r \zeta w(\zeta) \varphi(1, \zeta) d\zeta &= -\frac{\varphi(x, r) \varphi(1, r)}{v(x)}, \\ \varphi(0, r) = 2\theta(r) \varphi(1, r), \quad x \in [0, 1], \quad r \in [r_0, R]. \end{aligned} \quad (13)$$

Para probar que la solución del problema anterior existe Dyson et al. transforman el problema en un problema de punto fijo para  $\varphi(1, r)$  y prueban que, bajo ciertas condiciones *razonables*,

1.  $w(r) > 0$  es continua y con derivada continua en  $[r_0, R]$  y tal que  $rw(r)$  es creciente,
2.  $v(x) > 0$  es continua y con derivada continua en  $[0, 1]$ ,
3.  $\theta(r) : [r_0, R] \mapsto [0, 1]$  es una función de Lipschitz y existe  $r_1 \in (r_0, R]$  tal que  $\theta(r) = \theta(r_0) = \theta_0$  en  $[r_0, r_1]$  con  $\theta_0 < 1/2$ ,

existe una única solución. Es decir, desde el punto de vista matemático la ecuación (13) tiene solución única, lo que indica que, efectivamente, el modelo es capaz de describir el estado estable del tumor.

Un problema abierto más complicado es el estudio de la evolución temporal. Es todo un reto encontrar la solución analítica del problema general (11) y probar que, asintóticamente, en efecto se tiene la solución predicha por el problema estacionario aquí descrito. Este problema ha sido parcialmente resuelto por los mismos autores en [12].

## 4.2 Modelo para protocolos de quimioterapia

Vamos a estudiar un modelo que intenta describir la acción de la quimioterapia sobre un tumor. En particular, de la solución del problema matemático se puede extraer un protocolo de actuación para las sesiones de quimioterapia usualmente usadas en medicina.



La idea biológica es la siguiente: Dentro del ciclo celular existen dos etapas bien diferenciadas: la interfase ( $G_1$ ,  $S$  y  $G_2$ ) y la mitosis  $M$  (ver apartado 2.6). Durante la fases  $G_2$  y  $M$  la célula resulta más vulnerable (en particular en  $M$  la célula es más porosa y su pared celular más fina), por lo que es un momento muy apropiado para aplicar dosis de medicamentos o agentes externos (quimioterapia) que eviten la división celular y que, a ser posible, aniquilen a las células cancerígenas. Este es un protocolo habitual en las terapias actuales contra el cáncer. El principal problema radica en que la quimioterapia no sólo mata a células tumorales, sino que también daña a muchas de las células sanas de los tejidos que rodean al tumor. Es por ello que es muy importante saber cuánto y cuándo hay que aplicar las dosis.

Comencemos describiendo un modelo para el crecimiento del tumor. Vamos a diferenciar dos estados dentro del ciclo celular de las células tumorales: el estado  $E_1$  y  $E_2$  correspondientes a las fases  $G_1, S$  y  $G_2, M$ , respectivamente. Sean  $N_1$  y  $N_2$  el número de células en cada estado  $E_1$  y  $E_2$ , respectivamente.

El modelo más sencillo de crecimiento tumoral corresponde al modelo exponencial o malthusiano que postula que el número de células cancerígenas salientes de cada fase es proporcional a  $N_i$ ,  $i = 1, 2$ . Así, para el estado  $E_2$  tenemos

$$\dot{N}_2(t) = -a_2 N_2(t) + a_1 N_1(t),$$

donde  $\dot{f}$  indica la derivada temporal de  $f$  y  $a_1$  y  $a_2$  ciertas constantes. Para el estado  $E_1$  hay que tener en cuenta que entran el doble de células procedentes de la división celular, luego

$$\dot{N}_1(t) = -a_1 N_1(t) + 2a_2 N_2(t).$$

En ambas expresiones  $a_1$  y  $a_2$  son dos parámetros que tienen que ver con los tiempos de duración de ambos estados. Usualmente el sistema se escribe en su forma matricial

$$\dot{N} = AN, \quad N(t_0) = N_0,$$

donde

$$A = \begin{pmatrix} -a_1 & 2a_2 \\ a_1 & -a_2 \end{pmatrix}.$$

La solución de esta ecuación es

$$N(t) = N(t_0) \exp((t - t_0)A),$$

donde  $\exp(zA)$  es la función exponencial de una matriz (ver e.g. [5]). Como ejercicio al lector se deja el estudio de la estabilidad del sistema anterior en función de los parámetros  $a_1$  y  $a_2$ .

Introduzcamos ahora la droga (medicamento) o, matemáticamente hablando, el control. Sea  $u(t)$  la dosis de droga introducida en el sistema en el momento  $t$ . Supondremos que  $u(t) = 0$  indica que no se suministra droga, y  $u(t) = 1$  que se aplica la dosis máxima. Se asume que la dosis de droga está en relación directa con la fracción de células que mueren en el estado  $E_2$ , así que en

la fase  $E_1$  sólo entrará la fracción  $2(1-u(t))a_2N_2$ . Esto nos conduce al problema [17]

$$\dot{N} = (A + uB)N, \quad u = (u(t), 0)^t, \quad N(t_0) = N_0,$$

donde  $A$  es la matriz de antes y  $B$  viene dada por

$$B = \begin{pmatrix} 0 & -2a_2 \\ 0 & 0 \end{pmatrix}.$$

El problema consiste en minimizar el funcional

$$J(u) = r_1N_1(T) + r_2N_2(T) + \int_{t_0}^T u(t)dt$$

sobre el conjunto de todas las funciones medibles  $u$  con valores en  $[0, 1]$  donde  $r_1$  y  $r_2$  son ciertos números que representan sendos pesos relativos para el número total de células en el instante final  $T$ . Además, se supone que el control modela también el número de células del tejido sano que mueren por la acción de la quimioterapia.

Nótese que al ser la norma de la matriz  $A + uB$  acotada para todo control admisible, entonces la ecuación diferencial matricial siempre tiene solución. En particular se tiene que si  $N_1(t_0) > 0$  y  $N_2(t_0) > 0$ , entonces  $N_1(t) > 0$  y  $N_2(t) > 0$  para todo  $t \geq t_0$ . La prueba es directa: basta acotar la matriz  $A + uB$  y usar la solución en forma de matriz exponencial.

Este problema se puede generalizar a  $n$  dimensiones (que corresponderían a más divisiones del ciclo celular) de la siguiente forma [18]:

Sea  $N = (N_1, N_2, \dots, N_n)^t$  el vector de estado cuyas coordenadas corresponden al número de células en la fase  $E_k$ ,  $k = 1, 2, \dots, m$ . Sea  $u$  el vector  $(u_1, u_2, \dots, u_m)^t$  siendo  $u_k$ , la correspondiente dosis de medicamento, que asumiremos no negativa, es decir, para todo  $k = 1, 2, \dots, m$ ,  $u_k(t) \in [0, 1]$ , donde como antes 0 implica que no hay medicación y 1 que se ha aplicado la dosis máxima. Sean  $r = (r_1, r_2, \dots, r_n)$  y  $s = (s_1, s_2, \dots, s_m)$  dos vectores de números positivos y no negativos, respectivamente. En este caso tenemos que minimizar el funcional

$$J(u) = rN(T) + \int_{t_0}^T s u(t)dt$$

sobre todas las funciones medibles  $u$  que cumplan la ecuación dinámica

$$\dot{N} = \left( A + \sum_{k=1}^m u_k B_k \right) N, \quad N(t_0) = N_0.$$

En el caso general hay que asumir que tanto las matrices  $A$ ,  $B_k$  como el control  $u$  son tales que la matriz  $A + \sum_{k=1}^m u_k B_k$  tiene elementos diagonales negativos y el resto no negativos. Cuando esto ocurre se dice que las matrices  $A$  y  $B_k$  satisfacen la condición **M**.

Una condición necesaria de optimalidad viene dada por el principio del máximo de Pontryagin [27], que establece que si  $u^* = (u_1^*, u_2^*, \dots, u_m^*)$  es un control óptimo, entonces existe una función (vector fila) continua  $\lambda$  que satisfice la ecuación adjunta

$$\dot{\lambda} = -\lambda \left( A + \sum_{k=1}^m u_k^* B_k \right), \quad \lambda(T) = r,$$

tal que el control óptimo minimiza el Hamiltoniano

$$H = \lambda AN + \sum_{k=1}^m u_k (s_k + \lambda B_k N).$$

Si suponemos que las matrices  $A$  y  $B_k$  satisfacen la condición **M**, entonces se tiene que tanto  $N_k(t)$  como  $\lambda_k(t)$ ,  $k = 1, 2, \dots, n$ , son positivas en  $[0, T]$  (probar como ejercicio).

Para resolver el problema se define la función de conmutación (switching function)

$$\Phi(t) = (\Phi_1(t), \dots, \Phi_m(t)), \quad \Phi_k = s_k + \lambda B_k N,$$

de forma que el control óptimo queda determinado por las expresiones

$$u_k^*(t) = \begin{cases} \alpha_k & \text{si } \Phi_k(t) > 0, \\ \beta_k & \text{si } \Phi_k(t) < 0. \end{cases}$$

Nótese que a priori los controles no se determinan por la condición del mínimo si  $\Phi_k(t) = 0$ . Ahora bien, si  $\Phi_k(t)$  se anula en todo un abierto  $(a_k, b_k)$ , entonces todas sus derivadas también deben anularse en dicho intervalo. Este tipo de controles se denominan controles singulares. Los controles constantes a trozos se conocen como controles de salto (*bang-bang*).

Recuperemos el caso  $n = 2$  de antes. Para este caso en [17] se prueba que los controles singulares no son nunca óptimos, por lo que la única opción es buscar controles de salto. Además, en el caso de dimensión 2 se tiene  $u = (u(t), 0)$ ,  $s = (1, 0)$ , y el control óptimo tiene la forma

$$u^*(t) = \begin{cases} 0 & \text{si } \Phi(t) > 0, \\ 1 & \text{si } \Phi(t) < 0, \end{cases} \quad \Phi = 1 + \lambda BN.$$

Esto es precisamente el protocolo usado en la actualidad en la quimioterapia: se alternan periodos de aplicación de máximas dosis de drogas con periodos de descanso sin aplicación de ningún agente químico. El problema es personalizar el tratamiento, es decir, saber cuántos periodos y por cuánto tiempo hay que aplicar los correspondientes controles para que el tratamiento sea óptimo. Es justo aquí donde este modelo podría ayudar a los médicos corrigiendo los correspondientes tiempos de quimioterapia y descansos para hacer el tratamiento lo más óptimo posible.

Para finalizar debemos mencionar que aunque el modelo predice justamente el mismo tipo de protocolo que el que se usa actualmente, todavía está lejos de poder ser aplicado. En primer lugar es un modelo, con todo lo que ello implica. En segundo lugar, hay que encontrar los parámetros correctos correspondientes a cada caso (paciente), lo cual sigue siendo algo bastante complicado y subjetivo hasta el momento.

### 4.3 Un modelo fractal para el crecimiento de tumores

Nuestro último ejemplo lo tomaremos de los trabajos de Brú et al. [6, 7].

La idea de Brú et al. consiste en asumir que el borde de un tumor crece siguiendo las pautas de la geometría fractal [21] y por tanto se puede usar el análisis de escala para medir la rugosidad de la frontera del tumor. A partir de dicho análisis se postula que la dinámica de crecimiento del tumor se puede describir usando la ecuación diferencial estocástica que modela el MBE (Molecular Beam Epitaxy). A partir de la ecuación MBE, que es una ecuación de difusión, se deduce que la dinámica de crecimiento tumoral posee tres características fundamentales:

1. Que la mayor parte de la actividad celular del tumor se concentra muy cerca del borde (ver figura 28).
2. El crecimiento del tumor ocurre por **difusión superficial** de las células tumorales. Es decir, que las nuevas células se mueven por la frontera del tumor hasta que encuentran una posición cóncava donde estén rodeadas por una mayor cantidad de células tumorales (ver figura 29).
3. El crecimiento del tumor, pasado algún tiempo, deja de ser exponencial y se convierte en lineal, es decir, el radio del tumor crece linealmente con el tiempo.

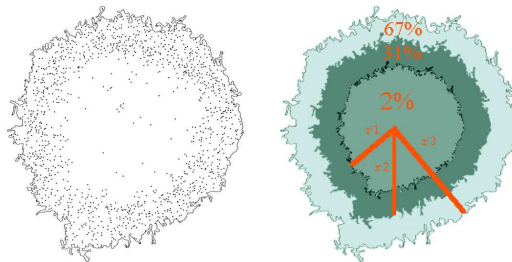


Figura 28: Actividad celular de un tumor. Con puntos negros se localizan las células marcadas que tienen gran actividad celular. Los anillos concentran un 2%, 31% y 67% de la actividad celular, respectivamente (de dentro a fuera). (Esta figura está tomada de [6]).

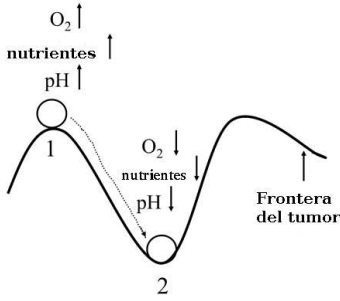


Figura 29: Difusión de células tumorales según la dinámica MBE. La célula nace en la posición 1 y se difunde hasta la posición 2.

Veamos una forma sencilla de explicar que el crecimiento del tumor es radial si la actividad celular se concentra casi en su totalidad en el borde del mismo. Por sencillez asumiremos que el tumor es esférico. Sea  $R$  el radio del tumor y  $r_0$  el de la célula (que también aproximaremos por una esfera). Si  $r_0 \ll R$ , entonces el número total de células del tumor  $N_T$  es del orden

$$N_T \sim \frac{4/3\pi R^3}{4/3\pi r_0^3} = \left(\frac{R}{r_0}\right)^3.$$

Si asumimos que las células se multiplican sólo en el borde, entonces la velocidad de crecimiento  $dN_T/dt$  ha de ser proporcional al número de células en la frontera  $N_F \sim 4\pi R^2/(\pi r_0^2) = 4(R/r_0)^2$ . Luego

$$\frac{dN_T}{dt} = \frac{d}{dt} \left(\frac{R}{r_0}\right)^3 = 3\frac{R^2}{r_0^3} \frac{dR}{dt} = K \frac{R^2}{r_0^2} \Rightarrow$$

$$\frac{dR}{dt} = \kappa r_0 \Rightarrow R(t) = R(t_0) + \kappa r_0(t - t_0).$$

Discutamos brevemente las bases matemáticas del modelo de Brú et al. (una magnífica introducción a la ecuación MBE se puede encontrar en [3, 22]). La ecuación MBE multidimensional tiene la forma

$$\frac{\partial}{\partial t} h(\vec{r}, t) = -K\Delta^2 h(\vec{r}, t) + F + \eta(\vec{r}, t), \tag{14}$$

donde  $\vec{r}$  es el radio vector de la posición del borde de la superficie en crecimiento y  $\Delta$  es el laplaciano multidimensional. Para dimensión 1 la ecuación anterior se reduce a

$$\frac{\partial}{\partial t} h(x, t) = -K \frac{\partial^4}{\partial x^4} h(x, t) + F + \eta(x, t), \tag{15}$$

donde  $h(x, t)$  representa la altura de la superficie,  $F$  es una constante (que en la teoría de Brú et al. representa el coeficiente de división celular) y  $\eta(x, t)$  es un ruido descorrelacionado en el espacio y en el tiempo, es decir es un ruido aleatorio de media cero  $\langle \eta(x, t) \rangle = 0$ , y función de correlación  $\langle \eta(x, t), \eta(x', t') \rangle = 2\nu\delta(x - x')\delta(t - t')$ . Esquemáticamente el crecimiento de la superficie según la ecuación (15) está representado en la figura 30.

Este tipo de ecuaciones es bien conocido en Física por los modelos de crecimiento de cristales. Además, la geometría fractal de estos objetos viene determinada por los exponentes de rugosidad  $\alpha$  y de crecimiento  $\beta$ . Dichos coeficientes son calculados en [7] para el borde de un tumor (en realidad para la proyección bidimensional de un tumor in vitro) y se obtienen los valores

$\alpha = 3/2$  y  $\beta = 3/8$ . Estos coeficientes de rugosidad corresponden a la ecuación unidimensional (15) y son la clave de la teoría de Brú et al. para explicar la dinámica de los tumores antes descrita.

El problema principal radica en que la ecuación (15) es unidimensional (describe el crecimiento de una superficie en una única dirección tal y como se muestra en la figura 30), por lo que no es aplicable a un caso bi o tridimensional como es el de un tumor. De hecho, para un caso de dimensión 2 (que correspondería a la proyección bidimensional del tumor), por ejemplo, a la ecuación MBE bidimensional le corresponden los coeficientes de rugosidad y de crecimiento  $\alpha = 1$  y  $\beta = 1/4$ , respectivamente, que no coinciden con los valores  $\alpha = 3/2$  y  $\beta = 3/8$  encontrados en [7]. En general, para la ecuación MBE  $d$ -dimensional (14) se tiene [3]  $\alpha = 2 - d/2$  y  $\beta = 1/2 - d/8$ , por lo que en el caso tridimensional los valores teóricos de los coeficientes  $\alpha$  y  $\beta$  tampoco coinciden con los valores obtenidos experimentalmente en [7].

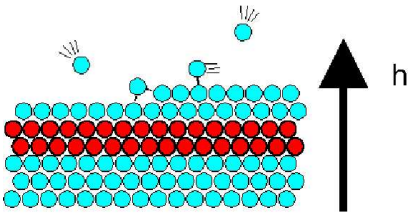


Figura 30: Esquema del crecimiento de una superficie según la ecuación MBE (15).

Para terminar notemos que este modelo difiere radicalmente del que discutimos anteriormente, donde el crecimiento es claramente exponencial y no se restringe sólo al borde. Lo más llamativo de la dinámica propuesta es que las células cancerígenas no atacan directamente a las células sanas intentando conseguir oxígeno y nutrientes, sino que se difunden buscando espacio vital, es decir, el tumor primero destruye el tejido huésped y luego lo invade, que es justo lo inverso de lo que es aceptado actualmente por la mayoría de los especialistas. Una consecuencia directa, en caso de que esta nueva dinámica fuese confirmada experimentalmente, sería un cambio radical en las terapias actuales para el tratamiento de la enfermedad.

**Agradecimientos.** Estas notas están basadas en el curso homónimo que tuvo lugar en la Universidad Internacional de Andalucía en el verano de 2005. Quiero expresar mi agradecimiento a la UNIA por la invitación a co-organizar dicha escuela y en particular a Antonio Durán por su apoyo incondicional. Quiero agradecer también a N.R. Quintero por sus comentarios sobre el apartado §3, a A. Sánchez y E. Moro por su ayuda para entender el modelo descrito en el §4.3 y a J. C. García, M. Pérez y N. R. Quintero por sus sugerencias. A Juan Luis Varona le agradezco su cuidadosa lectura y la infinidad de comentarios y correcciones que me han ayudado a que este trabajo sea legible. Agradezco a Enrique Fernández Cara su cordial invitación a publicar estas notas en el Boletín de SEMA. Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia de España (proyecto BFM-2003-6335-C03-01) y la Junta de Andalucía (proyecto FQM-262).

## Referencias

- [1] Alberts B., Johnson A., Lewis J., Raff M., Roberts K., Walter P., *Biología Molecular de la Célula*. Omega, 2004.
- [2] Attwood T.K., Parry-Smith D.J., *Introducción a la Bioinformática*. Prentice Hall, 2002.
- [3] Albert-Laszls Barabási A., Stanley H. E., *Fractal Concepts in Surface Growth*. Cambridge University Press, 1995.
- [4] Baxevanis A.D., Francis Ouellette B.F., *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. 3<sup>a</sup> ed., Wiley, 2004.
- [5] Braun M., *Differential Ecuations and their Applications*. 4<sup>a</sup> ed., Springer Verlag, 1993.
- [6] Brú A., Albertos S., Subiza J.L., López García-Asenjo J., Brú I., The Universal Dynamics of Tumor Growth, *Biophysical Journal* **85** (2003), 2948-2961.
- [7] Brú A., Pastor J.M., Fernaud I., Melle S., Brú I., Super-rough dynamics on tumour growth. *Phys. Rev. Lett.* **81** (1998), 4008-4011.
- [8] Bunn Ch., *Crystals. Their role in Nature and Science*. Academic Press, 1964.
- [9] Cushing J.M., The LPA model. *Fields Institute Communications* **43** (2004), 29-55.
- [10] Cushing J.M., Population Dynamics. In *The Encyclopedia of Nonlinear Science*, (A. Scott, editor), Routledge, Taylor and Francis Group, 2004.
- [11] Dyson J., Vilella-Bressan, R., Webb G.F., The steady state of a maturity structured tumor cord cell population. *Discr. Cont. Dyn. Sys. B* **4** (2004), 115-134.
- [12] Dyson J., Vilella-Bressan R., Webb G.F., The evolution of a tumor cord cell population, *Comm. Pure Appl. Anal.* **3** (2004), 331-352.
- [13] Englander S.W., Kallenbach N.R., Heeger A.J., Krumhans, A.J., Litwin A., Nature of the open state in long polynucleotide double helices: possibility of soliton excitations. *Proc. Natl. Acad. Sci. USA* **77** (1980), 7222-7226.
- [14] Kovalev A. S., Continuum Approximations. In *The Encyclopedia of Nonlinear Science*, (A. Scott, editor), Routledge, Taylor and Francis Group, 2004.
- [15] Hirs C.H.W., Moore S., Stein W.H., The Sequence of the Amino Acid Residues in Performic Acid-oxidized Ribonuclease. *Journal of Biological Chemistry* **235** (1960), 633-647.

- [16] Jillson D., Insect populations respond to fluctuating environments. *Nature* **288** (1980), 699-700.
- [17] Ledzewicz U., Schättler H., Optimal bang-bang controls for a 2-compartment model in cancer chemotherapy. *Journal of Optimization Theory and Applications* **114** (2002), 609-637.
- [18] Ledzewicz U., Schättler H., Swierniak A., Optimal control for a class of compartmental models in cancer chemotherapy. *International Journal of Applied Mathematics and Computer Science* **13** (2003), 357-368.
- [19] Lodish H., Berk A., Zipursky S.L., Matsudaira P., Baltimore V., Darnell J., *Biología Celular y Molecular*. Editorial panamericana, 2003.
- [20] Mackey M.C., Santillán M., Mathematics, Biology, and Physics: Interactions and Interdependence. *Notices of the AMS* **52** (2005), 832-840.
- [21] Mandelbrot B., *The Fractal Geometry of Nature*. W. H. Freeman and Company, 1982.
- [22] Moro E., *Estudio analítico y numérico de ecuaciones diferenciales estocásticas: Aplicación a la Mecánica Estadística*. Tesis doctoral. Universidad Carlos III de Madrid, 1999 (ver <http://gisc.uc3m.es/~moro/profesional.html>).
- [23] Murray J.D., *Mathematical Biology. I*. 3ª ed., Interdisciplinary Applied Mathematics **17**. Springer-Verlag, 2002.
- [24] Peyrard M., Bishop A.R., Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.* **62** (1989), 2755-2758.
- [25] Peyrard M., Nonlinear dynamics and statistical physics of DNA. *Nonlinearity* **17** (2004), R1-R40.
- [26] Pevsner J., *Bioinformatics and Functional Genomics*. Wiley, 2003.
- [27] Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V., Mishchenko E.F., *The Mathematical Theory of Optimal Processes*. MacMillan, 1964.
- [28] Ryle A.P., Sanger F., Smith L.F., Kitai R., The disulphide bonds of insulin. *Biochemical Journal* **60** (1955), 541-556.
- [29] Scott A., *Nonlinear Science*. 2ª ed., Oxford University Press, 2003.
- [30] Yakushevich L.V., Nonlinear DNA dynamics: a new model. *Phys. Letters A* **136** (1989), 413-417.
- [31] Yakushevich L.V., *Nonlinear Physics of DNA*. Wiley, 2004.
- [32] Yomosa S., Solitary excitations in deoxyribonucleic acid (DNA) double helices. *Phys. Rev. A* **30** (1984), 474-480.