



Facultad de Ciencias

**TOPOLOGÍA DE LOS PROCESOS CELULARES EN EL
ADN USANDO LA TEORÍA DE NUDOS**

(TOPOLOGY OF CELLULAR PROCESSES IN DNA USING
KNOT THEORY)

Trabajo de Fin de Grado
para acceder al

GRADO EN MATEMÁTICAS

Autor: Bruno Incera González

Director: Mario Fioravanti Villanueva

Febrero, 2023

AGRADECIMIENTOS

En primer lugar, me gustaría dar las gracias a Mario por su comprensión, su paciencia y guía en este trabajo. A mi familia por ayudarme en todo lo que estaba en sus manos y más. A Alba por su apoyo incondicional y por estar a mi lado en los momentos más difíciles.

*”Last but not least, I wanna thank me
I wanna thank me for believing in me
I wanna thank me for doing all this hard work
I wanna thank me for having no days off
I wanna thank me for, for never quitting
I wanna thank me for always being a giver
And tryna give more than I recieve
I wanna thank me for tryna do more right than wrong
I wanna thank me for just being me at all times...”*

Resumen

La doble hélice del ADN puede ser vista como una cinta que se enrosca y retuerce sobre sí misma. Mediante la teoría de nudos podemos medir el grado de enroscamiento y retorcimiento y cómo los procesos celulares los modifican aumentando en muchos casos la tensión producida. Para solucionar este problema están unas enzimas llamadas topoisomerasas que se encargan de modificar la topología del ADN liberando tensión excesiva. El enroscamiento y el retorcimiento sumados dan lugar a una propiedad invariante entre dos curvas llamada número de enlace. A parte de todo lo mencionado anteriormente en este trabajo se estudia como estos parámetros se comportan sobre el ADN y sobre el ADN que se encuentra envolviendo superficies de proteínas.

Palabras clave: ADN, Topología, Teoría de nudos, Teorema de Călugăreanu, número de enlace, Twist number, Writhe number, topoisomerasas.

Abstract

The DNA double helix can be seen as a ribbon that coils and twists around itself. Knot theory allows to measure the degree of twisting and coiling and how cellular processes modify them by increasing in many cases the tension produced. The enzymes called topoisomerases, in effect, exist to solve this kind of problems by modifying the DNA topology, releasing excessive tension. The twist and the coil degree give rise to an invariant property between two curves called the linking number. Apart from everything mentioned above, in this project we study how these parameters behave on DNA and when DNA wraps on protein surfaces.

Key words: DNA, Topology, Knot theory, Călugăreanu's Theorem, linking number, twist number, writhe number, topoisomerases.

Índice general

1. Introducción	4
1.1. ADN	5
1.2. Topología	7
2. Número de enlace	16
3. Twist, Writhe y el teorema de Călugăreanu	23
3.1. Twist y Writhe	24
3.2. Teorema de Călugăreanu	31
4. ADN sobre la superficie de una proteína	34
5. Topoisomerasas	47
5.1. Topoisomerasas de tipo I.	48
5.2. Topoisomerasas de tipo II.	49
Bibliografía	51

Capítulo 1

Introducción

A lo largo de la historia de la humanidad el conocimiento y el entendimiento del cuerpo humano siempre ha levantado interés entre las grandes mentes. A medida que la sociedad ha ido evolucionando y modernizándose se han ido desarrollando avances científicos que permitirían el estudio de nuestro cuerpo. Sin embargo, no fue hasta 1869 que se descubrió por accidente la molécula de ADN, llamada en un primera instancia *nucleína*, por su procedencia del núcleo de una célula. El descubrimiento fue logrado por el médico suizo Johann Friedrich Miescher. No fue hasta 1930 que se conoció el contenido del ADN y hasta 1953 que se propuso el modelo de la doble hélice del ADN. Como vemos, no hace más de un siglo que se conoce esta interesante molécula con su composición y estructura; y es por esto que el margen de investigación en este campo es tan amplio a la par que esperanzador por las consecuencias positivas que puede traer en el campo de la medicina, como la cura del cáncer o la predicción de enfermedades con inteligencia artificial y la lectura genética.

Uno de los campos que puede ayudar al progreso en la investigación del ADN es la topología pues mediante la teoría de nudos puede servir para identificar las distintas moléculas, así como definir las mediante propiedades geométricas y poder saber cuáles son equivalentes topológicamente. El ADN al estar compuesto por dos cadenas en forma de hélice se puede pensar en ella como dos curvas enrolladas entre sí dependiendo de cierta manera una de la otra.

En este capítulo se introduce el ADN para tener un conocimiento básico sobre él y poder comprender con mayor facilidad la aplicación de toda la matemática del trabajo al campo de la bioquímica. El campo principal que se trata en este trabajo es la teoría de nudos, rama de la topología.

Algunos de los parámetros geométricos de las curvas con más importancia en la topología del ADN son el número de enlace, del cual se habla en extensión en el capítulo 2, el índice de retorcimiento y el índice de enroscamiento que simplemente por sus nombres te pueden dar una pista sobre lo que miden. Estos dos índices junto al Teorema de Călugăreanu, que los relaciona con el número de enlace, serán explicados en el capítulo 3. En la gran mayoría de casos el ADN no se encuentra libre en la célula sino sobre la superficie de una proteína, esto da cabida a la geometría de curvas y superficies y la relación entre ambas, dando lugar a conceptos nuevos.

Por último, en el capítulo 5, hablamos sobre unas enzimas llamadas topoisomerasas encargadas de la ruptura y sutura del ADN para la facilitación de procesos celulares. Lo interesante de estas enzimas es que modifican la topología del ADN, y por tanto los conceptos topológicos vistos en este trabajo ayudan a entender la labor de estas enzimas.

Mencionar que en muchas situaciones los casos teóricos estudiados serán muy similares a los casos prácticos pero no exactamente iguales ya que reproducir estos con todo detalle sería muy dificultoso a la vez que poco práctico pues la diferencia en el resultado obtenido sería irrelevante.

1.1. ADN

El ADN, siglas de ácido desoxirribonucleico, es una molécula que se encuentra en el interior de las células y contiene la información genética responsable del desarrollo y el funcionamiento de todos los organismos vivos conocidos [1]. La función principal de este ácido nucleico es el almacenamiento de información a largo plazo. El ADN se organiza estructuralmente en cromosomas, estructuras altamente organizadas que recogen la mayor parte de la información genética de los seres vivos. A nivel funcional se organiza en genes, que son piezas de ADN que generan características físicas específicas. Estas características no vienen directamente del propio ADN, sino de una molécula llamada ARN, formada a

partir del ADN.

El ADN en una célula se encuentra, mayormente, en el núcleo de ésta y en menor medida en las mitocondrias, los orgánulos encargados de la respiración celular. El ADN, como bien es sabido a día de hoy, tiene una estructura de escalera de caracol, lo que se conoce como la doble hélice del ADN. Los azúcares y los fosfatos alternados forman las dos hebras del ADN. Éstas, a su vez, están unidas entre sí por lo que serían "los peldaños de la escalera", conocidos como pares de bases o de nucleótidos [2]. Hay cuatro tipos diferentes de nucleótidos o bases: A=adenina, T=tiamina, C=citosina y G=guanina. Estos se unen mediante enlaces de hidrógeno, formando siempre los dos mismos tipos de bases: adenina con tiamina y citosina con guanina. Podemos observarlo en la Figura 1.1. El ADN que se encuentra en el núcleo de una célula puede contener millones de pares de nucleótidos. Nos podemos hacer una idea del desorden de hebras de ADN que podemos encontrarnos dentro de un núcleo.

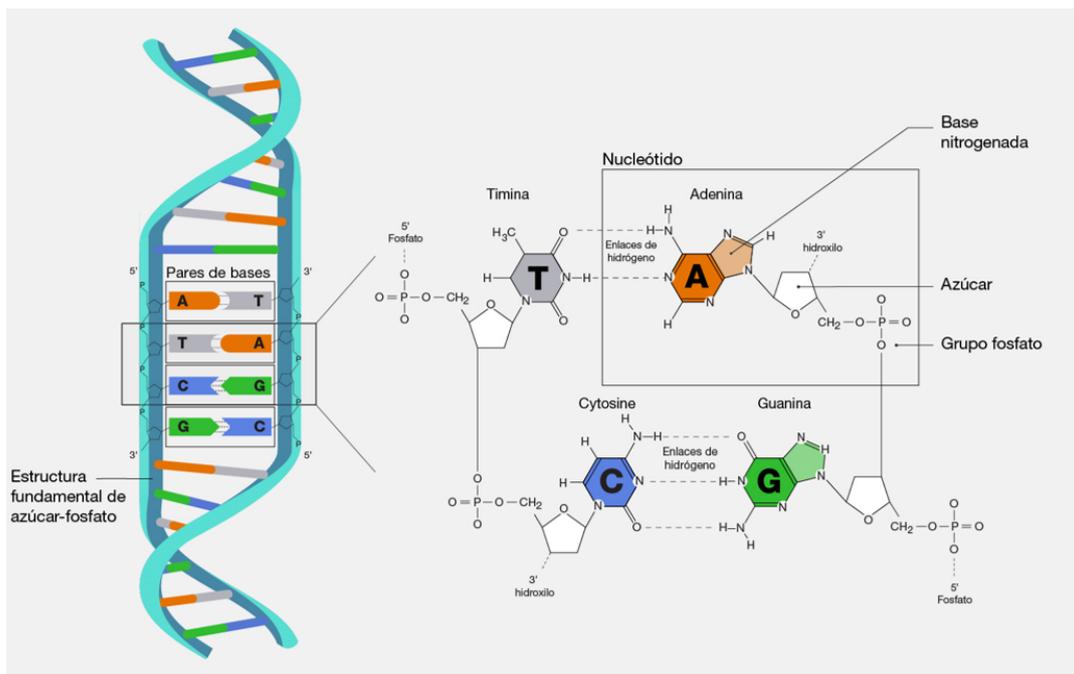


Figura 1.1: Ácido desoxirribonucleico (ADN). [1]

Existen una serie de procesos biológicos que se llevan a cabo en el ADN que son imprescindibles para la vida. Los procesos son los siguientes:

- **Recombinación:** a partir de dos secuencias de ADN se forma un nuevo ADN que tendrá información genética de las dos moléculas.

- Replicación: consiste en duplicar una molécula de ADN. Desenrollando la doble hélice y cada una de estas dos sirve de plantilla para sintetizar nuevas cadenas de ADN.
- Transcripción: es el proceso de generar una cadena de ARN a partir de una secuencia de ADN, por lo que ese ARN pasaría a tener la información genética que contenía esa secuencia de ADN.

Como mencionamos antes, al haber tantas cadenas de ADN y tan largas en tan poco espacio, es un completo desorden. Es por esto que en la práctica, la gran mayoría de veces, el ADN se encuentra enredado y anudado, dificultando la realización de estos procesos. De alguna manera tenía la naturaleza que intentar solucionar estos problemas, y precisamente esta es la función de las enzimas topoisomerasas. Estas se encargan de modificar la topología de la cadena del ADN mediante la rotura de una o dos de las hebras de la hélice para poder desenrollar y desenroscar con facilidad la propia hélice del ADN.

En 1963, se descubrió por primera vez, en el virus del Polyomaviridae, ADN con una estructura circular, es decir, siendo un dominio cerrado [3]. Al estudiar este virus se observaron dos formas de la molécula circular, con diferentes velocidades de sedimentación, pero ¿por qué sucedía esto? Se descubrió que la versión más compacta de las dos se transformaría en la otra al romper una de las hebras de la doble hélice. Posteriormente, gracias al estudio de J. Vinograd y colaboradores [4], se supo que la versión más compacta de las dos era porque estaba intacta y presentaba superenrollamiento. Este fue el descubrimiento del ADN superenrollado, lo que condujo a intentar entenderlo. Con la intención de medir el grado de enrollamiento y estudiar la geometría y topología de un cadena de ADN cerrada se introdujeron algunos primeros conceptos matemáticos, los cuáles veremos más adelante en profundidad.

1.2. Topología

Para poder adentrarnos más en profundidad en este tema primero tenemos que conocer términos matemáticos como topología, nudo, número de enlace, número de torsión, número de enrollamiento...

Para explicar sencillamente el concepto de topología conviene irnos a sus orígenes. Gottfried Leibniz (1646-1716) fue el primero en preocuparse por la parte de la geometría que no trata sobre medidas y cálculos, sino de su posición y las propiedades que son independientes de las medidas [5]. A este concepto lo llamó *geometriam situs* (geometría de la posición). Formalmente, la topología se describe como la rama de las matemáticas que se ocupa de estudiar las propiedades de los cuerpos geométricos que no son alteradas al aplicar a dicho cuerpo transformaciones continuas como, por ejemplo, estirar, torcer, doblar, arrugar... sin tapar ni hacer agujeros, ni romper, ni traspasar la superficie.

Todas las preguntas alrededor de la geometría del ADN empezaron a surgir a partir de 1953, a raíz del descubrimiento de la estructura de doble hélice del ADN, por Francis Crick y James Watson [6]. Y gracias a la topología podremos comprender por qué son geoméricamente posibles los cambios estructurales del ADN.

1.2.1. Teoría de Nudos

En matemáticas un nudo es una curva cerrada sin autointersecciones. Es decir, una aplicación continua inyectiva $g : [0, 1] \rightarrow \mathbb{R}^3$ donde $g(0) = g(1)$. También podemos expresar los nudos como un embebimiento f de la circunferencia en el espacio euclídeo tridimensional, expresando g como composición de dos aplicaciones, $g = f \circ p$:

$$\begin{cases} f : \mathbb{S}^1 \rightarrow \mathbb{R}^3 \\ p : [0, 1] \rightarrow \mathbb{S}^1, p(t) = (\cos(2\pi t), \sin(2\pi t)) \end{cases}$$

Siendo f continua e inyectiva por construcción [7].

Es fundamental tener claro el concepto de *isotopía del ambiente* para poder entender la teoría de nudos, a continuación se expone su definición.

Definición 1.2.1 (Isotopía del ambiente [8]). *Una isotopía del ambiente entre dos aplicaciones continuas $f : [a, b] \rightarrow \mathbb{R}^3$ y $g : [a, b] \rightarrow \mathbb{R}^3$ es otra aplicación continua $F : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ que cumpla $F(x, 0) = f(x)$, $F(x, 1) = g(x) \forall x \in \mathbb{R}^3$ y tal que si $t_0 \in [0, 1]$ entonces $F(x, t_0)$ es un homeomorfismo.*

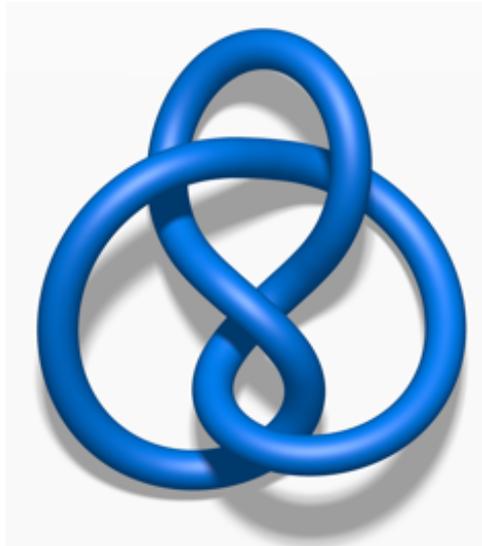


Figura 1.2: Nudo en 3 dimensiones.

Mediante la teoría de nudos estudiamos las variaciones a las que podemos someter a los nudos sin “romperlos”. Cuando un nudo se puede deformar en otro sin romperlo se dirá que son *nudos equivalentes*. Es decir, si podemos transformar uno en otro mediante isotopías del ambiente, siendo todos los nudos intermedios también nudos equivalentes. Todos los nudos equivalentes entre ellos forman una clase de equivalencia.

Esto es interesante porque geoméricamente es muy similar a cómo se encuentra gran parte del ADN en los organismos. Pensar en el ADN, ya sea circular o en un dominio cerrado, como un nudo nos simplificará el trabajo. Realmente no es tan importante cómo se encuentre el ADN en la naturaleza, sino su estructura topológica, ya que de esto dependerán muchas de sus propiedades.

Sabiendo esto, nos conviene trabajar con los nudos en su representación mas sencilla, deshaciendo sus cruces en la medida de lo posible. Pero, ¿cómo hacemos esto matemáticamente? No podemos representar con una cuerda cada nudo con el que vayamos a trabajar, menos aún, ir deshaciendo cada nudo sin ningún tipo de procedimiento riguroso y matemático, ya que estaríamos dependiendo de la habilidad o intuición individual para deshacer nudos. Además, llegaría un punto en que la complejidad de los nudos sería tal que se convertiría en una tarea imposible deshacer un nudo mediante la destreza humana.

Por todos estos inconvenientes y para poder simplificar su presentación, los nudos del espacio euclídeo son representados mediante una proyección en el plano, dónde indica-

remos en cada cruce de cuerda qué trozo pasa por arriba y cuál lo hace por debajo. A esto se le llama diagrama plano del nudo. En la Figura 1.3 podemos ver un ejemplo de diagrama de nudo. Pero, seguimos sin tener una metodología para deshacer nudos y qué nudos son o no equivalentes.

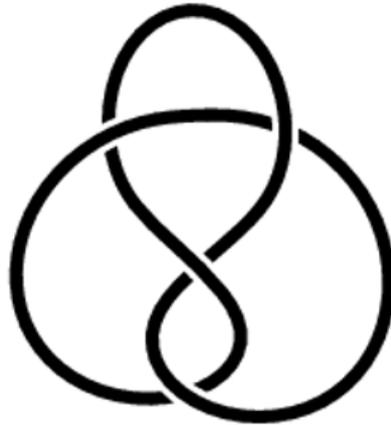


Figura 1.3: Diagrama del nudo de Figura 1.2.

A principios del siglo XX, K. W. F. Reidemeister observó una técnica para comprobar que dos nudos son equivalentes [9]. Esta técnica consiste en solventar la parte problemática de los diagramas, los cruces de cuerda, mediante una serie de isotopías ambientales locales, llamadas *movimientos de Reidemeister*. Este matemático demostró que dos nudos son equivalentes si y sólo si cualquiera de los dos diagramas puede ser transformado en el otro mediante una combinación finita de movimientos de Reidemeister.

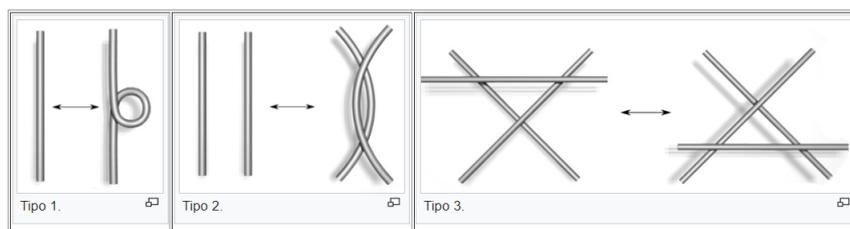


Figura 1.4: Movimientos de Reidemeister [10].

Gran parte del interés en el estudio de nudos surgió para intentar entender mejor la química de la materia [11]. A finales del siglo XIX William Thomson tenía una teoría sobre la materia y su composición. Creía que los diferentes tipos de materia no eran más que nudos de éter. El éter era un supuesto fluido invisible que se consideraba que llenaba todo el espacio. Muy pocos años después de que la química despertara interés sobre los

nudos a los matemáticos se comprobó que dicha teoría no era correcta, ya que surgió un modelo atómico más preciso. Sin embargo, los matemáticos siguieron investigando el campo de los nudos.

Sorprendentemente, alrededor de 100 años más tarde, gracias a avances en la bioquímica se descubrió que existen nudos en las moléculas de ADN. Gracias al modelo atómico equivocado de William Thomson, surgió un campo de las matemáticas que resulta tener varias aplicaciones a la química y a la biología, en este trabajo nos centraremos en la aplicación sobre el estudio de la doble hélice del ADN. La teoría de nudos también tiene otras aplicaciones en la teoría de cuerdas, en la gravedad cuántica y en mecánica estadística.

1.2.2. Topología del ADN

El ADN al estar formado por dos cadenas no es una única curva sino un par de curvas o interpretándolo como superficie, una cinta. Vamos a ver qué es una cinta desde el punto de vista matemático [11]. Cuando hablamos del concepto cinta (o banda), estamos haciendo referencia a la suma de una curva suave en el espacio de tres o más dimensiones (nosotros trabajaremos en \mathbb{R}^3) y un vector normal a la curva, cualquier vector que sea perpendicular al vector tangente. Si nos imaginamos un punto y un vector que se desplazan en el espacio, la estela que dejan forma una banda, siempre y cuando no se tome la dirección del vector. Otra forma de construir una cinta es con el eje central, una curva que se sitúa en la superficie de la cinta y es equidistante a los dos bordes, y un vector normal al eje que apunte a cualquiera de los bordes. Esta construcción de la cinta será la más usada en este trabajo más porque aporta más información geométrica y facilita muchos cálculos.

Cuando la curva directora de la banda es cerrada y al completar una vuelta los extremos del vector normal director de la banda vuelven a su origen diremos que estamos ante una banda cerrada con dos bordes. Si el extremo final del vector se sitúa donde el comienzo del vector y viceversa, entonces el resultado será una banda de con un solo borde. Las bandas cerradas de dos bordes tienen dos caras y las bandas cerradas de un solo borde tienen una cara, el ejemplo más conocido de estas últimas es la *Banda de Möbius*.

Como resultado, cuando tenemos una cinta con dos caras, cada borde forma una curva cerrada y estas podrían estar anudadas entre ellas. Todo esto es interesante porque la doble hélice del ADN tiene precisamente esta forma, una cinta cerrada de dos caras.

Vamos a ver porque sabemos que cada hélice del ADN circular es una curva independiente, es decir, que no es una banda de Möbius. Como vimos anteriormente, cada uno de los fosfatos de la hélice está unido a dos azúcares alternados, las pentosas, un tipo de azúcar que como su nombre indica tiene 5 átomos de carbono. En biología molecular, los ácidos nucleicos tienen asignados una orientación de entre los átomos, lo que es conocido como direccionalidad. Particularmente, esta convención hace que las pentosas tengan asignados el tercer y quinto átomo de carbono como 3' (“extremo cinco prima”) y 5' (“extremo cinco prima”). Estos son los dos extremos importantes de dicho azúcar ya que los grupos fosfatos conectan dos moléculas mediante dichos extremos, uniendo el extremo 3' de un molécula de azúcar con el extremo 5' de la otra molécula.

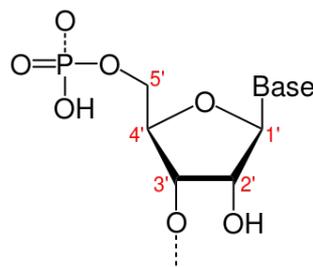


Figura 1.5: Átomos de carbono de una molécula de azúcar orientados según la direccionalidad estándar.
[12]

La doble hélice del ADN es una molécula antiparalela, las dos hebras tienen orientación opuesta. Se puede apreciar en la figura, por cuestiones de simetría esto ha de ser así. Esta orientación tiene sentido horario y podemos observar que las bases nitrogenadas se unen a los azúcares en los extremos 1'.

La asignación de una dirección para nombrar los átomos se debe a que los ácidos nucleicos sólo pueden sintetizarse *in vivo* en una dirección concreta, de 5' a 3'. Por esto mismo una hebra de ADN, tiene orientación propia, siguiendo el orden siguiente. Empieza en el extremo 5' del azúcar, y sigue hacia el extremo 3' del mismo, donde se une al extremo 5' de la siguiente molécula de azúcar mediante el enlace fosfodiéster. Es por este motivo que cada una de las hebras comienza en un extremo 5' y termina en un extremo 3' de la última pentosa, como podemos apreciar en la Figura 1.6.

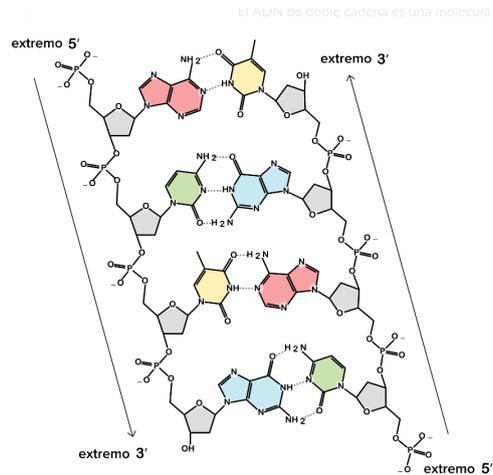


Figura 1.6: Cadenas antiparalelas del ADN. [13]

Las moléculas del ADN pueden presentarse de forma lineal o con los extremos de la doble hélice unidos, de forma circular. En este último caso nos podemos encontrar al ADN en estado de superenrollamiento, es decir, la doble hélice se retuerce sobre sí misma, de esta forma el eje de la hélice pasa de describir una curva plana a otra hélice pero en este caso de sentido opuesto al retorcimiento original, lo que se conoce como una superhélice. En el cromosoma, el ADN no es circular cerrado, pero se organiza en lazos cuya base está unida a un esqueleto proteico y se comporta como tal, pudiendo presetar también fenómenos de superenrollamiento. A las moléculas circulares de ADN o con comportamiento topológico cerrado en cierto dominio, se las conoce como moléculas CCC (Close Covalent Circles).

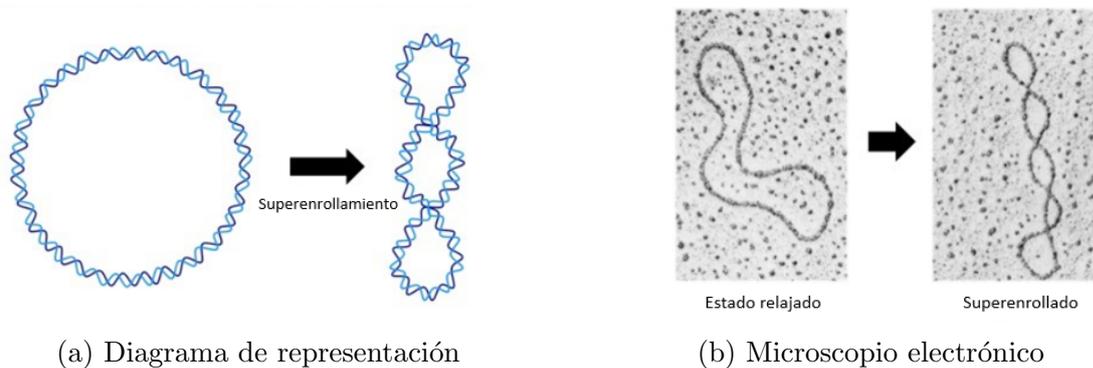


Figura 1.7: Molécula de ADN superenrollada.

A la hora de estudiar el efecto de las enzimas topoisomerasas sobre el ADN, los científicos utilizan moléculas circulares de ADN. Esto tiene una sencilla explicación, a la hora de observar el efecto de una de estas enzimas sobre el ADN, si el ADN no es circular, al tener los dos extremos de la cadena libre, si se formase un nudo o algún tipo de enrollamiento

podría deshacerse porque se liberaría la tensión producida. Para entender esto podemos imaginarnos que retorremos o anudamos una cuerda sin apretar el nudo, entonces lo que estaría aguantando la cuerda anudada sería la fricción de la cuerda con ella misma. Seguramente, al soltar cualquiera de los extremos, perderíamos el retorcimiento o nudo que hemos creado. Esto es lo que sucede si queremos analizar la aplicación de una enzima topoisomerasa a una cadena de ADN libre, al retorcer localmente un segmento de la cadena la tensión se escaparía por alguno de los extremos.

Una molécula de ADN con exactamente la misma secuencia puede encontrarse en estado relajado o en diferentes estados de enrollamiento. A cada una de estas posibles isoformas en las que podemos encontrarnos a la molécula se le llama topoisómero. Dado que la topología del ADN se ve modificada en los procesos celulares de replicación, transcripción y recombinación, su estudio es importante para comprender estos cambios.

Existen dos tipos de superenrollamiento, si el giro alrededor del eje es en el sentido de las agujas del reloj se define como superenrollamiento positivo y diremos que tenemos una doble hélice levógira. En el caso contrario, al girar en el sentido opuesto a las agujas del reloj, tendremos superenrollamiento negativo, formando una doble hélice dextrógira. Existe otro posible estado del ADN, cuando no existe superenrollamiento, llamado estado relajado. En la naturaleza, en la gran mayoría de ocasiones nos encontramos las moléculas de ADN superenrolladas negativamente. Esto se debe a que cuando la hélice se encuentra en superenrollamiento negativo es más fácil la separación de las dos hélices para que accedan las enzimas encargadas de los procesos de replicación y transcripción, las topoisomerasas.

El modelo de cuerdas anterior simula con bastante precisión una molécula circular de ADN. En la Figura 1.8a vemos dos hebras enroscadas entre sí y posteriormente unidas cada una con ella misma. Vemos como a medida que separamos las dos hebras con superenrollamiento negativo éste va disminuyendo hasta que las hebras se encuentran en estado relajado (Figura 1.8b), ya que la fuerza que realizamos para separarlas hace liberar tensión de superenrollamiento. Si seguimos separando las cuerdas empieza a producirse superenrollamiento positivo (Figura 1.8c).

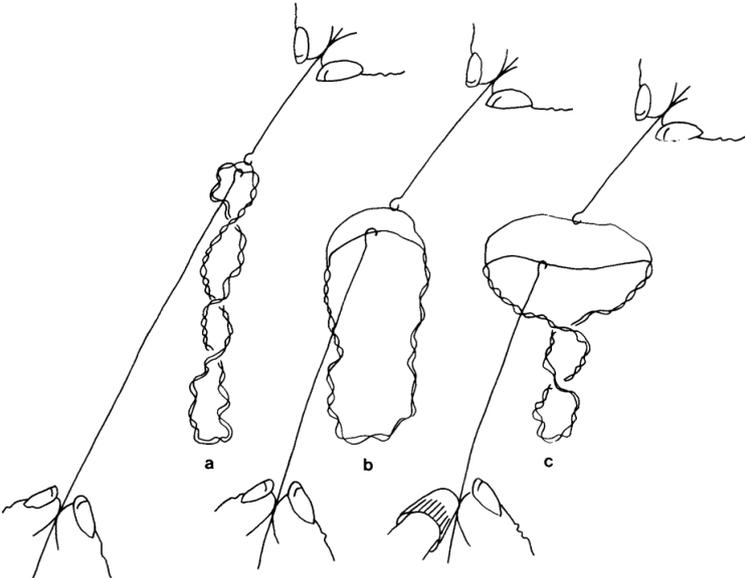


Figura 1.8: Posibles tipos de enrollamiento. [14]

Capítulo 2

Número de enlace

El número de enlace, *Linking number* (Lk) en inglés, también conocido como índice de ligazón, es el número de vueltas que un nudo da alrededor de otro [11]. A priori parece un concepto muy sencillo y fácil de medir, pero más adelante veremos que esto no es así. Como es de esperar, este valor siempre que las curvas sean cerradas será un número entero, y según la orientación de las dos curvas este valor puede ser negativo o positivo, información trascendental al estudiar la topología del ADN. Esta propiedad topológica asociada a un par de nudos no depende de la geometría de las curvas y es invariante por homeomorfismos, su valor no se verá modificado siempre y cuando no rompamos una de las hebras. Dadas dos curvas cerradas K y F , indicamos el número de enlace entre ambos como $Lk(K, F)$.

Cuando hablamos del número de enlace de una cinta cerrada (R), necesariamente con dos bordes, nos referimos al número de enlace entre las dos componentes del borde (R_1 y R_2):

$$Lk(R) = Lk(R_1, R_2)$$

Cálculo de Lk

Para entender con facilidad este concepto, sin duda, lo mejor es verlo en las imágenes de la Figura 2.1.

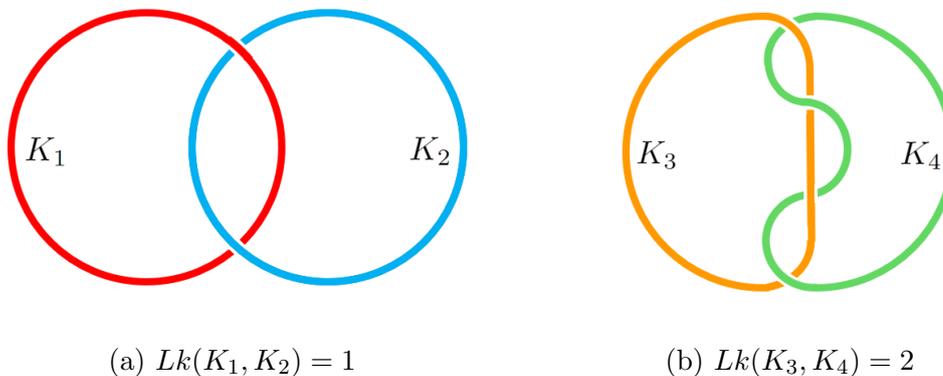


Figura 2.1: Nudos enlazados entre sí.

En el diagrama de la Figura 2.1a podemos observar como cada uno de los nudos pasa una vez por encima y otra por debajo del otro, quedando K_1 y K_2 unidos por un sólo enlace. Es decir, estos dos nudos no pueden ser separados sin romper uno de los dos. Con enlaces tan simples se puede apreciar a simple vista, incluso con K_3 y K_4 podemos intuir que el número de enlace en este caso será 2, ya que K_4 se enrolla dos veces sobre la parte recta de la hebra de K_3 . Estos son ejemplos muy simples para entender el concepto, es por eso que resulta bastante intuitivo. Sin embargo para cualquier otro par de nudos calcular el número de enlace es una tarea mucho más enredosa.

Existe un algoritmo para computar el número de enlace de dos nudos a partir de su diagrama [21].

Este algoritmo consiste en asignar a cada cruce entre los dos nudos uno de los dos valores $+1$ o -1 , el siguiente convenio de la Figura 2.2.

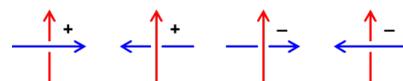


Figura 2.2: Regla para asignar valores a los diferentes cruces.

Para entenderlo mejor, hemos de asignar una orientación a cada una de las curvas cerradas y comprobar en cada cruce el giro menor de 180° que ha de realizar la hebra que

crucza por encima para deshacer el cruce, quedando ambas hebras con la misma orientación. Si el giro es en el sentido de las agujas del reloj asignamos a dicho cruce el valor -1 , y en el sentido contrario a las agujas del reloj asignaremos $+1$.

Una vez que todos los cruces tienen asignado su respectivo valor, el número de enlace será la mitad de la suma de dichos valores. La orientación de las curvas es irrelevante, cambiaría el signo de todos los cruces, modificando por tanto el signo del número de enlace.

Definición 2.0.1 (Número de enlace). *Sea J el conjunto de los cruces entre dos curvas cerradas K_1 y K_2 , n_j el valor asignado a un cruce j y $\{n_j : j \in J, \}$ el conjunto que recoge los valores de los cruces; entonces el número de enlace de K_1 y K_2 será:*

$$Lk(K_1, K_2) = \frac{1}{2} \sum_{i \in J} n_i.$$

Como mencionamos anteriormente, para la gran mayoría de casos no podemos pretender utilizar este algoritmo para calcular el número de enlace. Pero podemos estar tranquilos dado que existen otros métodos para calcular dicho índice. La fórmula original, descubierta por Carl Friedrich Gauss en 1833, no fue publicada hasta más de 30 años después de su muerte porque esta expresión se encontraba en una de sus notas privadas.

Se cree que Gauss estudió este concepto intentando buscar una expresión para el número de vueltas que da una γ_1 a γ_2 , siendo γ_1 y γ_2 curvas diferenciales cerradas, dando lugar a la próxima definición de número de enlace (Lk). Sin embargo, se cree que su interés no era tanto la topología entre curvas sino la aplicación al estudio del electromagnetismo.

Definición 2.0.2 (Número de enlace). *Siendo $r_1 = (x, y, z)$ y $r_2 = (x', y', z')$ las variables de γ_1 y γ_2 respectivamente, Gauss definió el número de enlace entre dos curvas como:*

$$\frac{1}{4\pi} \oint_{\gamma_1} \oint_{\gamma_2} \frac{(x' - x)(dydz' - dzdy') + (y' - y)(dzdx' - dxdz') + (z' - z)(dxdy' - dydx')}{[(x' - x)^2 + (y' - y)^2 + (z' - z)^2]^{\frac{3}{2}}}$$

Es decir,

$$Lk(\gamma_1, \gamma_2) = \frac{1}{4\pi} \oint_{\gamma_1} \oint_{\gamma_2} \frac{r_1 - r_2}{|r_1 - r_2|^3} \cdot (dr_1 \times dr_2)$$

El número de enlace es un invariante

Una vez introducido el concepto de número de enlace vamos a comprobar que es una característica invariable, es decir, que dos nudos equivalentes tienen mismo Lk . La forma de transformar nudos en nudos equivalentes es mediante los movimientos de Reidemeister. Comprobando que cada uno de estos movimientos, a pesar de modificar el número de cruces, no modifican el número de enlace, sabremos por tanto que una combinación finita de ellos tampoco lo hará y por tanto Lk será constante entre nudos equivalentes y por tanto un invariante. Veamos como para todas las posibles combinaciones de los 3 tipos de movimientos de Reidemeister cambia el valor de los cruces pero se compensan y no afecta a Lk .

Para el movimiento de Reidemeister tipo 1, ver Figura 1.4, no hace falta ninguna demostración más allá de aclarar que el número de enlace se calcula en los cruces creados entre curvas diferentes. En este movimiento el único cruce sobre el que se actúa es de una hebra con ella misma.

En el movimiento de Reidemeister tipo 2 aparecen dos hebras de curvas, Suponemos que son de dos curvas diferentes cada hebra, de no ser así el número de enlace sería cero por el mismo motivo que para el movimiento tipo 1. La Figura 2.3 demuestra que para todos los casos posibles el número de enlace se mantiene en cero. Cada una de las dos hebras puede tener dos sentidos (2^2) y a su vez cada hebra puede pasar por encima o por debajo, por lo que el número de combinaciones es 2^3 . No haría falta ver cada caso por separado si entendemos que cada cruce va a tener signo opuesto al otro cruce, luego $Lk = +1 - 1 = 0$, porque cada hebra se cruza con la otra dos veces, cada una en una dirección. El resultado de aplicar el movimiento son las dos hebras separadas sin ningún cruce y por tanto el número de enlace en esta situación también es cero.

Ahora vamos a ver todas las posibilidades dentro del movimiento de Reidemeister tipo 3. De las 3 hebras que lo forman, cada una puede tener dos sentidos, resultando en 2^3 posibilidades; y luego cada una de las hebras diagonales puede estar por encima de la otra, pero nunca por encima de la hebra horizontal, 2 posibilidades más. Teniendo, entonces, un total de $2^4 = 16$.

El número de enlace es una propiedad estudiada para un par de curvas cerradas, por

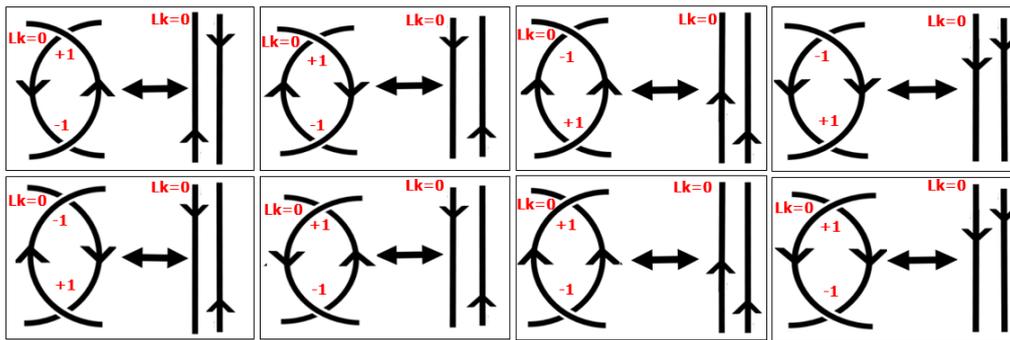


Figura 2.3: Movimiento de Reidemeister tipo 2 aplicado a todas las combinaciones de cruces posibles.

lo tanto de las tres hebras involucradas en los movimientos de Reidemeister tipo 3 dos pertenecerán a la misma curva. No tiene sentido estudiar el caso en que las tres hebras perteneciesen al mismo cruce pues el número de enlace sería 0 para todas las posibilidades, antes y después de aplicar el movimiento.

Tenemos dos tipos de cruces, cuando las dos diagonales pertenecen a la misma curva y la hebra horizontal a la otra curva, y cuando cada hebra diagonal pertenece a una curva diferente y la hebra horizontal comparte curva con alguna de las diagonales.

La invariancia para el primer caso queda demostrada en la Figura 2.4. Otra forma de demostrar que el número de enlace no varía para esta situación resulta de observar que al mover la hebra horizontal el cruce con cada hebra diagonal se mantiene intacto. Las direcciones de las hebras son las mismas, la hebra horizontal permanece encima de las diagonales y el ángulo que forma cada hebra diagonal con la horizontal es el mismo antes y después de la transformación. Esta observación nos basta para asegurar que el valor asociado a cada par (hebra diagonal, hebra horizontal) se mantiene intacto después del movimiento de Reidemeister tipo 3. Manteniéndose así constante el recuento de los valores de los cruces, que no es más que el número de enlace.

Para demostrar el segundo caso se puede aplicar la misma lógica que para el caso anterior. El cruce entre la hebra horizontal y la hebra diagonal de la otra curva mantendrá ángulo, orientación de las hebras y posición relativa entre ambas después de aplicar el movimiento. La única diferencia es dónde se cruzan pero eso no afecta a la hora de asignar el valor a un cruce, es por esto que se mantiene intacto. Como el cruce entre las hebras diagonales no se ve absolutamente modificado tras el movimiento, por supuesto,

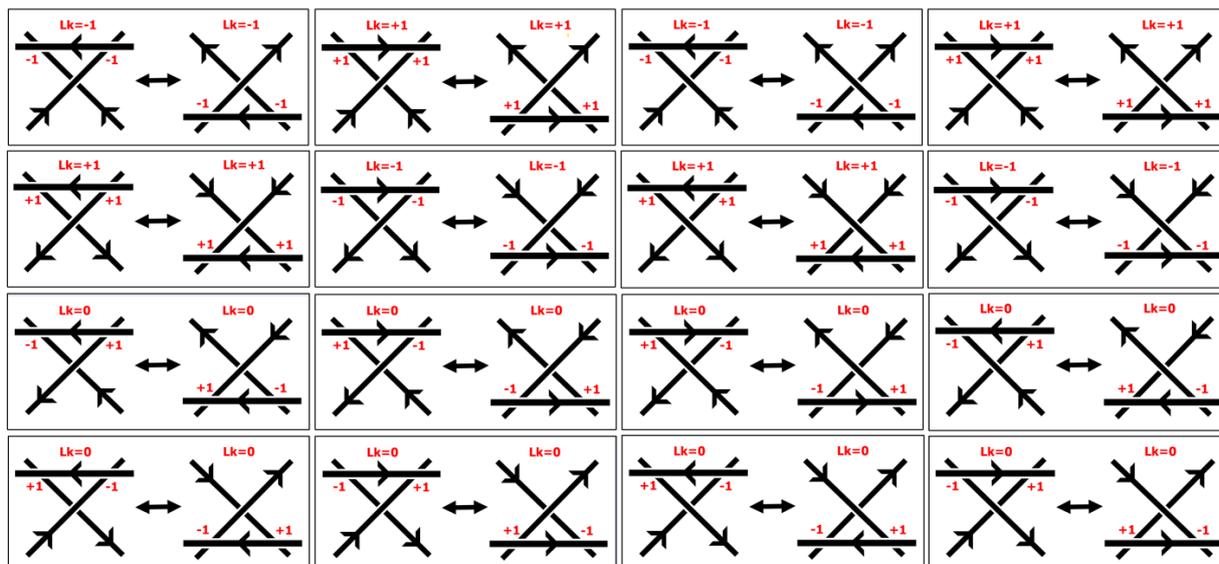


Figura 2.4: Movimiento de Reidemeister tipo 3 aplicado a todas las combinaciones de cruces posibles.

el valor del cruce se mantendrá. Resultando en que el número de enlace no varíe.

Con esto queda demostrado que para los tres tipos de movimientos de Reidemeister Lk se mantiene constante y por tanto es una propiedad invariante.

Número de enlace del ADN

En la naturaleza siempre nos encontramos al ADN formando una espiral, lo que se conoce como la doble hélice. En el caso de que dicha molécula se encuentre en estado relajado los cálculos de Lk se pueden simplificar ya que el número de enlace no será más que el número de vueltas que forme la doble hélice. Esto se explica con facilidad viendo la proyección de una molécula de ADN en un plano, ver Figura 1.1. Por cada vuelta las cadenas se cruzan dos veces. Con el primer cruce la cadena que se sitúa inicialmente a la derecha pasa a estar a la izquierda, y viceversa; y después del siguiente cruce cada cadena vuelve al lado de la proyección inicial. La Figura 2.2 muestra que el valor de los dos cruces de la proyección, ambos iguales, es +1. Recordemos que al computar el número de enlace de forma algorítmica se suma el valor de todos los cruces y se divide entre 2. Por este motivo, cada vuelta que da la doble hélice estará aportando una unidad a Lk .

Con este razonamiento se concluye que el número de enlace del ADN en estado relaja-

do, lo que llamaremos Lk_0 , no es más que el número de vueltas. Como vimos al principio del trabajo, el ADN está formado por pares de bases (pb), los peldaños que unen ambas cadenas, estos se usan como unidad de medida para calcular la longitud de la molécula. Entonces, el valor de Lk_0 es simplemente el cociente entre la longitud de la molécula de ADN (L) y el número de pares de bases por vuelta de la doble hélice (h):

$$Lk_0 = \frac{L}{h}$$

Al valor h en inglés se le conoce como *helical repeat*, que se podría traducir como *repetición helicoidal*. Es el número de pares de base que necesita una cadena de ADN para realizar una vuelta helicoidal alrededor de la otra. Este parámetro se ve afectado por distintos factores del medio en el que se encuentre dicha molécula de ADN, como son la temperatura, el pH, el tipo de sales, la concentración de estas... Además de esto también pueden afectar otras moléculas que se encuentren en el mismo medio actuando como ligandos.

Ahora que conocemos el número de enlace de una molécula en estado relajado, podemos medir la diferencia entre Lk y Lk_0 , que mide el grado de superenrollamiento del ADN, según el signo de esta diferencia indicará el sentido de giro del enrollamiento. Pero esta medida no aporta mucha información porque dependerá del tamaño de la molécula, cuanto más larga sea ésta, más probable es que tenga un alto grado de superenrollamiento. Si lo que queremos es medir de alguna forma la tensión, este valor tiene que depender de alguna manera de la longitud de la molécula. Este parámetro, σ , se conoce como densidad superhelicoidal y se obtiene de dividir la diferencia de número de enlace entre el número de enlace en estado relajado.

$$\sigma = \frac{\Delta Lk}{Lk_0}.$$

Capítulo 3

Twist, Writhe y el teorema de Călugăreanu

Este capítulo sigue las referencias [15], [16], [17], [18] y [19], así como el apoyo en el resto de referencias del trabajo.

Este teorema, también conocido como teorema de Călugăreanu-White-Fuller o fórmula de White, es un resultado fundamental en la geometría de bandas cerradas con retorcimiento. Relaciona el número de enlace, un invariante, con dos parámetros geométricos no invariantes de las bandas. Para entender con mejor claridad este teorema y los nuevos índices geométricos que relacionarán dos curvas cerradas enlazadas entre sí; podemos ver dichas curvas como los bordes de una banda, siempre y cuando estas dos curvas no tengan puntos en común. Será el caso de cuando trabajemos con moléculas del ADN, que veremos la doble cadena como una banda. Lo mismo sucede al revés, podemos estudiar los bordes de una banda como dos curvas, siempre y cuando la banda tenga dos caras y por tanto dos componentes del borde. Pero no podemos profundizar en el teorema de Călugăreanu sin antes introducir con anterioridad las propiedades que trata.

3.1. Twist y Writhe

El número de enlace es una propiedad invariante que se estudia sobre la estructura de la doble hélice. Ahora vamos a ver otras propiedades que no son invariantes de las cintas cerradas pero que están muy relacionadas con el número de enlace, el *Twist* (Tw) y el *Writhe* (Wr). No hay apenas literatura sobre estos conceptos en español por lo que he decidido traducir *Twist* como enroscamiento, y *Writhe* como retorcimiento. Sin embargo, a lo largo del trabajo haré referencia a estos términos en su nombre inglés.

Aunque con otra definición diferente, Tw y Lk_0 calculan el mismo parámetro ya que cuando el ADN se encuentra en estado relajado el número de enlace coincide con el *Twist number*. Más adelante cuando veamos el Teorema de Călugăreanu y estudiemos estos parámetros en las moléculas de ADN sobre superficies de proteínas entenderemos porque tiene sentido introducir estos conceptos.

El *Twisting number* es el número de giros que da la cinta sobre su propio eje, siendo el eje la curva que se encuentra en la mitad de la cinta, entre las dos curvas que forman los bordes de la cinta. Este número puede ser bien positivo o negativo, al igual que el número de enlace. Depende de la orientación de las curvas, será positivo cuando los giros se realicen en el sentido horario y negativo cuando gire en sentido antihorario. Cuando es positivo se le llama hélice dextrógira y se dice que gira hacia la derecha, y cuando es negativo diremos que es una hélice levógira, que gira hacia la izquierda.

La forma en la que se enrolla la doble cadena del ADN varía según el modelo de ADN. Las hebras de una molécula ADN-B, la más predominante en las células, dan una vuelta cada 10.5 pares de bases. Podríamos decir que el ADN-B es una escalera de caracol que da un giro completo cada 10.5 peldaños. Otros modelos como, por ejemplo, el ADN-A gira cada 11 pares de bases, o el ADN-Z que al girar en sentido antihorario lo hace cada -12 pares de bases.

$$Tw = \frac{n^{\circ} \text{ pares de bases}}{n^{\circ} \text{ pares de bases por vuelta}}$$

Como el modelo más frecuente con mucha diferencia es el ADN-B, generalizaremos que:

$$Tw = \frac{n^{\circ} \text{ pares de bases}}{10,5}$$

Para poder trabajar con la cinta primero vamos a dar una definición y a explicar su construcción. La cinta será una superficie reglada con borde formada por el eje de la cinta (\mathcal{A}) formado por los puntos $a(s)$, y por un segmento de longitud variable que se desplaza por la curva. La ecuación de la cinta será:

$$C(s) = a(s) + r(s)v_N(s)$$

donde $a(s)$ será una curva simple cerrada parametrizada por su longitud de arco, $v_N(s)$ un vector unitario normal a \mathcal{A} que apunte en la dirección de uno de los bordes de la cinta y $r(s)$ es el parámetro que recorre todo el grosor de la cinta, de un borde a otro, deslizándose en la dirección de $v_N(s)$. El grosor puede variar según el punto $a(s) \in \mathcal{A}$ en el que nos encontremos. $r(s) \in [-r_{max}(s), r_{max}(s)]$ $v_N(s)$ se encuentra en el plano normal y se puede descomponer en función del sistema de referencia móvil de Frenet de la siguiente manera:

$$v_N(s) = \cos(\alpha(s))N(s) + \sin(\alpha(s))B(s) .$$

Donde $N(s)$ es el vector normal del Triedro de Frenet, $N(s) = \dot{T}(s)$, vector ortonormal al vector tangente $T(s)$, $T(s) \cdot N(s) = 0$ para todo $s \in a(s)$. $B(s)$ es el vector binormal, $B(s) = T(s) \times N(s)$, que será también unitario, $\|B(s)\| = \|T(s) \times N(s)\| = \|T(s)\| \cdot \|N(s)\| \cdot \sin 90^\circ = 1 \cdot 1 \cdot 1 = 1$. Como $a(s)$ está parametrizada por su longitud de arco, el vector tangente $T(s)$ tiene medida 1, $\|T(s)\| = \|\dot{a}(s)\| = 1$.

Con esto podemos también dar las ecuaciones pertinentes de cada uno de los bordes de la cinta:

$$\gamma_1(s) = a(s) + r_{max}v_N(s)$$

$$\gamma_2(s) = a(s) + r_{max}v_N(s)$$

Twist

Ahora ya podemos definir formalmente el Twist, que con los términos descritos no es más que las vueltas que da el vector $N(s)$ a $v_N(s)$ a medida que nos desplazamos por \mathcal{A} :

$$Tw = \frac{1}{2\pi} \int_{\mathcal{A}} (T(s) \times v_N(s)) \cdot v_N(s) ds$$

Esta es la fórmula general utilizando el vector director que construye la cinta. Sin embargo, podemos usar cualquier otro vector normal a la curva porque realizará el mismo número de giros tras recorrer la curva \mathcal{A} . Para simplificar la notación abreviaremos $\alpha(s)$ como α y $T(s)$, $N(s)$ y $B(s)$ como T , N y B .

$$\begin{aligned} & (T \times v_N) \cdot v_N \\ &= (T \times (\cos(\alpha)N + \sin(\alpha)B)) \cdot v_N \\ &= (\cos(\alpha)B - \sin(\alpha)N) \cdot (-\dot{\alpha} \sin(\alpha)N + \cos(\alpha)\dot{N} + \dot{\alpha} \cos(\alpha)B) + \sin(\alpha)\dot{B} \\ &= -\cos(\alpha) \sin(\alpha)B \cdot N + \cos^2(\alpha)B \cdot \dot{N} + \dot{\alpha} \cos^2(\alpha)B \cdot B + \cos(\alpha) \sin(\alpha)B \cdot \dot{B} + \dot{\alpha} \sin^2(\alpha)N \cdot N \\ & \quad - \cos(\alpha) \sin(\alpha)N \cdot \dot{N} - \dot{\alpha} \cos(\alpha) \sin(\alpha)N \cdot B - \sin^2(\alpha)N \cdot \dot{B} \\ &= -\cos(\alpha) \sin(\alpha)B \cdot N + \cos^2(\alpha)B \cdot \dot{N} + \dot{\alpha} \cos^2(\alpha) + \cos(\alpha) \sin(\alpha)B \cdot \dot{B} + \dot{\alpha} \sin^2(\alpha) \\ & \quad - \cos(\alpha) \sin(\alpha)N \cdot \dot{N} - \alpha \cos(\alpha) \sin(\alpha)N \cdot B - \sin^2(\alpha)N \cdot \dot{B} \\ &= -2\alpha \cos(\alpha) \sin(\alpha)B \cdot N + \cos^2(\alpha)B \cdot \dot{N} - \sin^2(\alpha)N \cdot \dot{B} + \dot{\alpha} . \end{aligned}$$

Este sería el integrando desarrollando v_N con su definición. En el caso de utilizar el vector normal del Triedro de Frenet (N) estaríamos simplificando los cálculos en gran medida. N es un caso específico de $v_N(s) = \cos(\alpha(s))N(s) + \sin(\alpha(s))B(s)$, cuando $\alpha(s) = 0$. En ese caso el integrando del cálculo del Twist se reduce a :

$$-2\alpha \cos(\alpha) \sin(\alpha)BN + \cos^2(\alpha)B\dot{N} - \sin^2(\alpha)N\dot{B} + \dot{\alpha} = B \cdot \dot{N}$$

y por tanto la fórmula del Twist usando el Triedro de Frenet es:

$$Tw = \frac{1}{2\pi} \int_{\mathcal{A}} B \cdot \dot{N} ds .$$

También podemos escribir Tw en función de la torsión (τ) del eje central de la cinta (\mathcal{A}), utilizando las fórmulas de Frenet-Serret. Estas fórmulas describen propiedades geométricas de las curvas diferenciables continuas en el espacio euclidiano tridimensional utilizando los conocidos vectores unitarios tangente, normal y binormal, y sus derivadas.

Ahora tenemos una base de vectores normales sobre la que utilizar las fórmulas de Frenet-Serret.

$$\begin{bmatrix} \dot{T} \\ \dot{N} \\ \dot{B} \end{bmatrix} = \begin{bmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix} .$$

En la fórmula matricial anterior, $B = T \times N$ es el vector binormal y κ es la curvatura. Tenemos entonces que $\dot{B} = -\tau N \implies \tau = -N \cdot \dot{B}$. Ahora al ser N y B ortogonales $N \cdot B = 0$ y si derivamos esta última ecuación: $\dot{N} \cdot B + N \cdot \dot{B} = 0 \implies \dot{N} \cdot B = -N \cdot \dot{B} \implies \tau = \dot{N} \cdot B$. Con esto y sabiendo que $B = T \times N$ podemos afirmar que

$$Tw = \frac{1}{2\pi} \int_{\mathcal{A}} B \cdot \dot{N} \, ds = \frac{1}{2\pi} \int_{\mathcal{A}} \tau \, ds .$$

Vamos a ver otra forma alternativa de formular el Twist number y ver que ambas son equivalentes. La demostración que veremos a continuación es la más compleja, pero la considero necesaria porque es la original de Călugăreanu y nos arroja una idea de la línea que siguen las demostraciones de este teorema, aunque el resto ya se nos escapen por extensión y dificultad.

Empezamos describiendo una dirección de observación que llamaremos χ , diremos que tenemos un cruce local en $\gamma_1(s)$ cuando la cinta se aprecia de canto, que será cuando χ sea linealmente dependiente de $T(s)$ y de $N(s)$, y por tanto estará contenido en el plano que llamaremos $\Pi(s)$ con $T(s)$ y $N(s)$ como vectores directores, a este plano en geometría diferencial de curvas se le llama plano osculador. Esto podemos observarlo en la Figura 3.1. Entonces, existirá un cruce local para la dirección de observación χ cuando podamos definir ésta como:

$$\chi = T \cdot \cos\theta + N \cdot \sin\theta, \quad \theta \in [0, \pi] .$$

Podemos entonces crear el vector de todas las posibles direcciones de observación,

dependiente de los valores s y θ :

$$v(s, \theta) = T(s) \cdot \cos\theta + N(s) \cdot \sin\theta, \quad \theta \in [0, \pi].$$

Con esto, Călugăreanu afirmó que el doble de Tw es el promedio del número de cruces locales desde todas las direcciones de observación χ , lo que será calculado mediante una integral. Por lo tanto, para cada χ recorreremos $v(s, \theta)$ y habrá un cruce local cuando coincida v con χ o $-\chi$, es decir, cuando $\chi = \pm v(s, \theta)$, para algún s tal que exista $\gamma_1(s)$ y para algún θ entre 0 y π . Para saber si cada cruce contabilizará como positivo o como negativo hay que fijarse hacia qué lado está girando el plano tangente de la cinta a medida que nos acercamos a dicho cruce local, es decir, qué cara de la cinta es visible justo después del cruce, donde recordemos veíamos la cinta de canto. En la práctica lo veremos mediante el vector $(\partial_\theta v \times \partial_s v) \cdot v$, el signo del cruce local será el opuesto al vector anterior. Esto se debe a que θ aumenta en la dirección opuesta a s en la cinta.

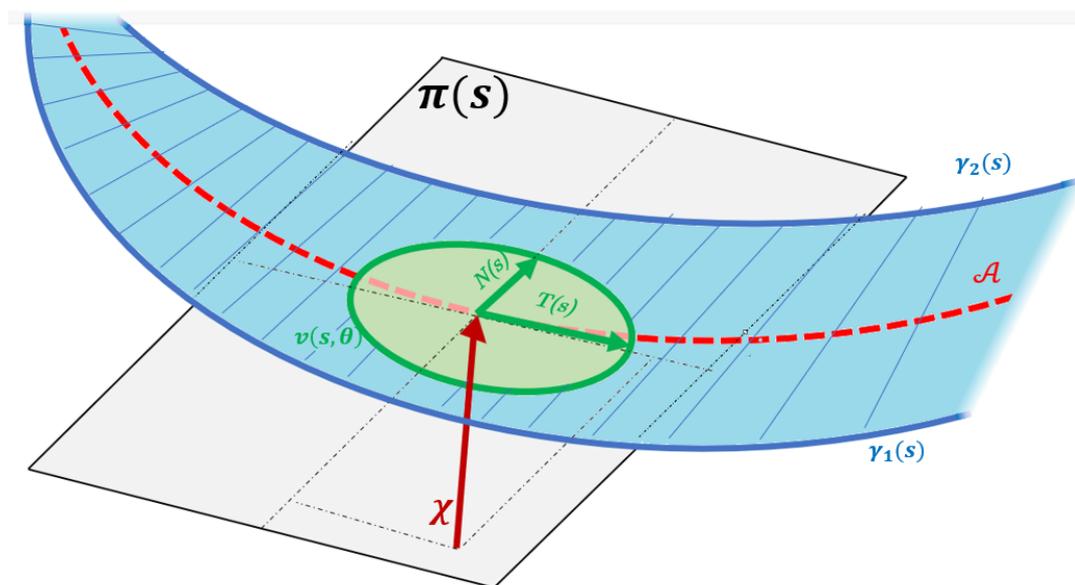


Figura 3.1: Representación de la estrategia seguida para calcular Tw .

Para obtener Tw mediante el promedio de todas las proyecciones posibles χ , calcularemos la integral sobre s y sobre θ del vector $(\partial_\theta v \times \partial_s v) \cdot v$. Como las únicas direcciones de observación que nos interesan son las que coinciden con v , dependiente de s y θ , estas son las variables de integración. Esto nos lleva a que, entonces, Tw será el promedio esférico del conteo de cruces locales:

$$Tw = \frac{1}{4\pi} \int_{\mathcal{A}} \int_0^\pi (\partial_\theta v \times \partial_s v) \cdot v \, d\theta \, ds .$$

Vamos ahora a comprobar que esta definición de Tw coincide con la primera que dimos. Para ello, vamos a intentar, partiendo de esta última expresión, llegar a la otra.

Lo primero de todo, empezamos derivando el vector v respecto de θ y de s , para posteriormente sustituir los términos que ya conocemos. Por el momento no vamos a tener en cuenta la integral respecto de s , más adelante entenderemos el porqué.

$$\frac{1}{4\pi} \int_0^\pi (\partial_\theta v \times \partial_s v) \cdot v \, d\theta \quad (3.1)$$

$$= \frac{1}{4\pi} \int_0^\pi ((-T \sin \theta + N \cos \theta) \times (\dot{T} \cos \theta + \dot{N} \sin \theta))(T \cos \theta + N \sin \theta) \, d\theta = \quad (3.2)$$

$$= \frac{1}{4\pi} \int_0^\pi ((T \cos \theta + N \sin \theta) \times (-T \sin \theta + N \cos \theta))(\dot{T} \cos \theta + \dot{N} \sin \theta) \, d\theta \quad (3.3)$$

$$= \frac{1}{4\pi} \int_0^\pi (\cos^2 \theta + \sin^2 \theta)(T \times N)(\dot{T} \cos \theta + \dot{N} \sin \theta) \, d\theta \quad (3.4)$$

$$= \frac{1}{4\pi} (T \times N) \int_0^\pi (\dot{T} \cos \theta + \dot{N} \sin \theta) \, d\theta \quad (3.5)$$

$$= \frac{1}{2\pi} \cdot (T \times N) \cdot \dot{N} = \frac{1}{2\pi} \cdot B \cdot \dot{N} . \quad (3.6)$$

Para pasar de la expresión (3.2) a la (3.3), como lo que tenemos es un producto escalar triple y sabemos que $[A, B, C] = -[B, A, C]$, es decir, que al hacer cualquier permutación entre los elementos de un producto triple escalar cambia el signo. Por lo cual, al hacer dos permutaciones cambiamos dos veces el signo y queda igual: $[A, B, C] = [C, A, B]$.

En el paso de (3.3) a la (3.4), simplemente calculamos el producto vectorial teniendo en cuenta que $T \cdot T = N \cdot N = 1$ por ser estos vectores unitarios.

Para pasar de (3.5) a (3.6) hemos calculado la integral, sabiendo que $\int_0^\pi \cos \theta \, d\theta = 0$ y que $\int_0^\pi \sin \theta \, d\theta = 2$.

Finalmente, veamos que los cálculos anteriores no afectan de ninguna manera a la integral respecto de s sobre γ_1 , por esto mismo hemos omitido dicha integral en las expresiones

anteriores, para facilitar su comprensión. Reescribiendo el resultado obtenido en (3.6) vemos que obtenemos la primera definición que vimos del Twist number: $Tw = \frac{1}{2\pi} \int_{\mathcal{A}} B \cdot \dot{N} ds$.

Writhe

El Writhe number mide cuánto se retuerce la banda, sucede cuando el propio eje de la banda se cruce con él mismo. Para el caso concreto del ADN, el Wr se refiere al número de hélices que forma la propia doble hélice del ADN. Para calcularlo el método es muy similar al del número de enlace, pero esta vez con una sola curva cerrada. Proyectamos el eje en cualquier plano y asignamos a cada cruce $+1$ o -1 con el mismo criterio que anteriormente para el Lk . Lo único diferente es que en este caso el número de Writhe (Wr) será la suma del valor de estos cruces, sin tener que dividir entre dos. A diferencia del Linking Number, Tw y Wr toman cualquier valor real, no solo números enteros.

Para comprender la forma de demostrar estos conceptos desde una perspectiva geométrica vamos a tener que distinguir dos tipos de cruces en las cintas y sus diagramas, los cruces locales y no locales. Los cruces locales se producen cuando desde la dirección escogida desde la que visualizar la cinta vemos a ésta de canto, y por tanto desde dicha perspectiva se produce un cruce entre las dos componentes del borde de la cinta. Por otro lado, los cruces no locales simplemente se producen cuando la propia cinta se cruza con ella misma.

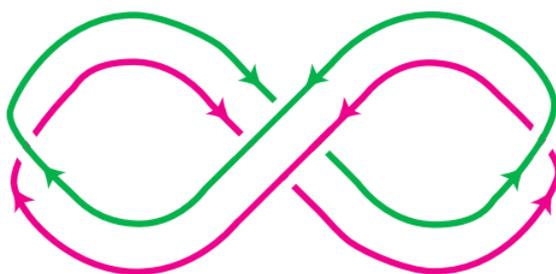


Figura 3.2: Cadenas antiparalelas del ADN [17].

En la figura 3.2 podemos observar los diferentes tipos de cruce. En los lados izquierdo y derecho de la figura tenemos un cruce local en cada uno. Observando la regla explicada en la Figura 2.2 para asignar el signo a cada cruce, vemos que el cruce de la izquierda es negativo y el de la derecha positivo. Además, en el centro de la figura se producen cuatro

cruces, correspondientes a cruces no locales, formados por el cruce de la propia cinta con ella misma. En este caso, de los cuatro cruces, solo contabilizarían para el número de enlace dos de ellos, dado que los otros dos no son cruces entre las diferentes curvas.

Para poder calcular el Writhe number vamos a construir el conjunto de todas las cuerdas de la cinta, a este concepto se le llama *writhe mesh* ($c_{\mathcal{A}}$), lo que se podría traducir como malla de retorcimiento.

$$c_{\mathcal{A}}(s, s') = \frac{a(s) - a(s')}{|a(s) - a(s')|}.$$

La integral que calcula el Writhe será:

$$\begin{aligned} Wr &= \frac{1}{4\pi} \int_{\mathcal{A}} \int_{\mathcal{A}} (\partial_s c_{\mathcal{A}} + \partial_{s'} c_{\mathcal{A}}) \cdot c_{\mathcal{A}} \, ds \, ds' \\ &= \frac{1}{4\pi} \int_{\mathcal{A}} \int_{\mathcal{A}} \frac{(\dot{a}(s) \times \dot{a}(s')) \cdot (a(s) - a(s'))}{|a(s) - a(s')|^3} \, ds \, ds' \end{aligned}$$

3.2. Teorema de Călugăreanu

La historia sobre este teorema comienza con un paper de Gheorghe Călugăreanu (1961) que a partir de la integral de Gauss definió un invariante que describía la medida en que una curva suave y simple en el espacio euclídeo de 3 dimensiones se enrolla a sí misma. Lo que en 1971 F. Brock Fuller llamó *writhing number*. En 1969 James White demostró esta fórmula sin asumir que la curvatura no se anulaba cómo había hecho Călugăreanu, logrando una demostración más completa.

Teorema 3.2.1 (Teorema de Călugăreanu). *Relaciona el número de enlace (Lk), el Twist number (Tw) y el Writhe number (Wr) de la siguiente manera:*

$$Lk = Tw + Wr$$

Este resultado es muy interesante porque nos ayuda a entender cómo afectan todo tipo de modificaciones y movimientos de la cinta sobre los índices que estamos estudiando. Aplicando esta fórmula también podemos obtener cualquiera de ellos conociendo los otros

dos, esto será útil ya que dependiendo de la situación alguno será más difícil de calcular que otro.

Podemos asegurar que la fórmula de Călugăreanu se cumple aplicada a las variaciones de cualquiera de los argumentos de la fórmula : $\Delta Lk = \Delta Tw + \Delta Wr$. La demostración de esto se reduce a la resta de la ecuación de Călugăreanu aplicada al estado inicial menos el estado modificado: $\Delta Lk = Lk - Lk_0 = Tw + Wr - Tw_0 - Wr_0 = \Delta Tw + \Delta Wr$. Lo que nos indica esta otra fórmula es que cualquier cambio en cualquiera de los tres términos será compensada con una variación en cualquiera de los otros dos. Tiene una muy sencilla demostración, basta aplicar la fórmula a una banda antes y después de las variaciones y restar ambas fórmulas.

Como el número de enlace no varía si no deshacemos el nudo o cortamos alguna de las hebras, Lk será fijo en estos casos y por tanto $\Delta Lk = 0$. Y por tanto cualquier variación en Tw o en Wr tendrá el mismo cambio de distinto signo en el otro término. Vamos a verlo con un ejemplo para entenderlo.

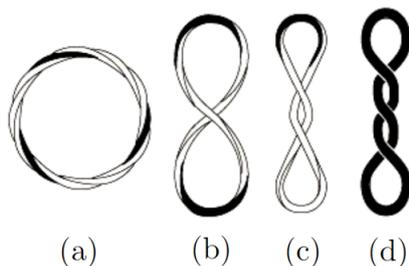


Figura 3.3: Representación de la misma cinta con $Lk = 3$. [20]

En la Figura 3.3 observamos como manipulando una cinta, por ejemplo haciendo o deshaciendo cruces del eje con él mismo, podemos modificar Tw y Wr .

En el caso de la Figura 3.3a vemos que el eje de la banda no se cruza en ningún momento, es por esto que $Wr_a = 0$. Y como el número de enlace de dicha banda es 3, automáticamente aplicando la fórmula de Călugăreanu $Tw = Lk - Wr = 3 - 0$, $Tw_a = 3$. Es decir, cada borde de la cinta da 3 vueltas sobre el otro borde.

Al pasar de (a) a (b) podemos apreciar que se ha retorcido la banda que antes formaba una circunferencia para pasar a tener un cruce del eje con él mismo, formando un ocho. Con este cambio aumentamos Wr , $\Delta Wr = +1$. Como Lk se mantiene constante en 3, $\Delta Lk = 0$. Luego, $\Delta Tw = \Delta Lk - \Delta Wr = 0 - 1 = -1$. La Figura 3.3b tiene los siguientes

Twist number y Writhe number:

$$Tw_b = Tw_a + \Delta Tw = 3 - 1 = 2$$

$$Wr_b = Wr_a + \Delta Wr = 0 + 1 = 1$$

Para las otras dos figuras, (c) y (d), observamos que cada una tiene un cruce de su propio eje más que el anterior. Entonces la Figura 3.3c cuenta con dos cruces del eje, y por tanto $Wr_c = 2$ y $Tw_c = 1$. En la Figura 3.3d apreciamos tres cruces del eje de la cinta con él mismo, entonces: $Wr_d = 3$ y $Tw_d = 0$.

Capítulo 4

ADN sobre la superficie de una proteína

Hasta ahora nos hemos centrado en las dobles cadenas de ADN libres ignorando la superficie sobre la que se encuentra la molécula como si ésta se encontrase flotando. Nos podemos encontrar una molécula de ADN envolviendo la superficie de una proteína, lo cual altera sus propiedades geométricas. Entran en juego conceptos nuevos como el número de enlace de la superficie (SLk) o el *Winding Number* (Φ), que lo podríamos traducir como número de serpenteo. Veremos cómo varían las propiedades geométricas del ADN cuando su eje no sigue una curva comprendida en un plano, por ejemplo el ADN cuando se encuentra envolviendo la superficie de una proteína como un cilindro o un toro. El eje del ADN superenrollado en la práctica tampoco se encuentra en un plano, por este motivo es importante introducir los nuevos conceptos que describen el modelo que se amolda más a la biología. Cuando hablamos de que el ADN se encuentra sobre la superficie la parte que se mantiene todo el rato sobre ella es el eje central, lo que anteriormente hemos llamado \mathcal{A} . Para este capítulo se han seguido las referencias [3], [23] y [24].

En muchas ocasiones, incluso en ausencia de una superficie envuelta por el ADN, será interesante el estudio con una superficie virtual adecuada que simule con exactitud el cuerpo al que envuelve la molécula de ADN, de esta manera el estudio será mucho más preciso.

A diferencia del ADN sobre un plano, la repetición helicoidal (h) de una molécula que

envuelve una proteína no depende de factores ambientales como la temperatura ambiente, concentración de sales... En estos casos h depende de la geometría de la superficie, como veremos más adelante. La demostración deriva de la nueva definición que tendremos de número de enlace al introducir los nuevos parámetros SLk y Φ . Vamos a aplicar este estudio geométrico a tres casos importantes biológicamente:

- ADN superenrollado plectonómico: este tipo de superenrollamiento es el que hemos visto hasta ahora, la doble hélice del ADN se enrolla alrededor de sí misma formando con su propio eje otra hélice. Este tipo de enrollamiento es característico de las células procariotas, aquellas que no tienen núcleo y el ADN se encuentra disperso en el citoplasma.
- ADN superenrollado toroidal: en esta conformación la molécula de ADN se enrolla sobre la superficie de un toro o aro imaginario formando una hélice de grado superior (ver Figura 4.3) Suele encontrarse en el núcleo de las células eucariotas.
- ADN en el minicromosoma del virus SV-40. Este caso tan específico lo que resulta interesante es la disposición geométrica del ADN y no tanto la utilidad de su estudio trasladado a la biología molecular.

Cuando hablamos de una molécula de ADN sobre una superficie realmente lo que se encuentra sobre dicha superficie es el eje de la doble cadena, las propias cadenas se encontrarán alternativamente encima y debajo de la superficie. En la Figura 4.1 se encuentra representada esta idea, junto a una serie de construcciones necesarias para el desarrollo de las fórmulas de este apartado. La curva C representa cualquiera de las cadenas del ADN, que va girando alrededor de su eje (A). El eje se encuentra sobre la superficie de proteínas M . P es un plano imaginario perpendicular a la curva A en el punto a y el punto c es la intersección de la curva C con el plano P . Con esto construimos el vector unitario v_{ac} con la misma dirección que la recta que une los puntos a y c . El vector v es el vector normal a la superficie M en el punto a . Esta representación no es más que una fotografía de un momento concreto del recorrido sobre la curva C , para cada otro instante tendremos otro plano, otros puntos y otros vectores.

El Winding Number (Φ) de C sobre A se define formalmente como el número de giros de v_{ac} sobre v a medida que P avanza sobre el eje A , recorriendo la curva A exactamente una vez. Como tanto a como c se encuentran en el plano P , perpendicular a la curva A ,

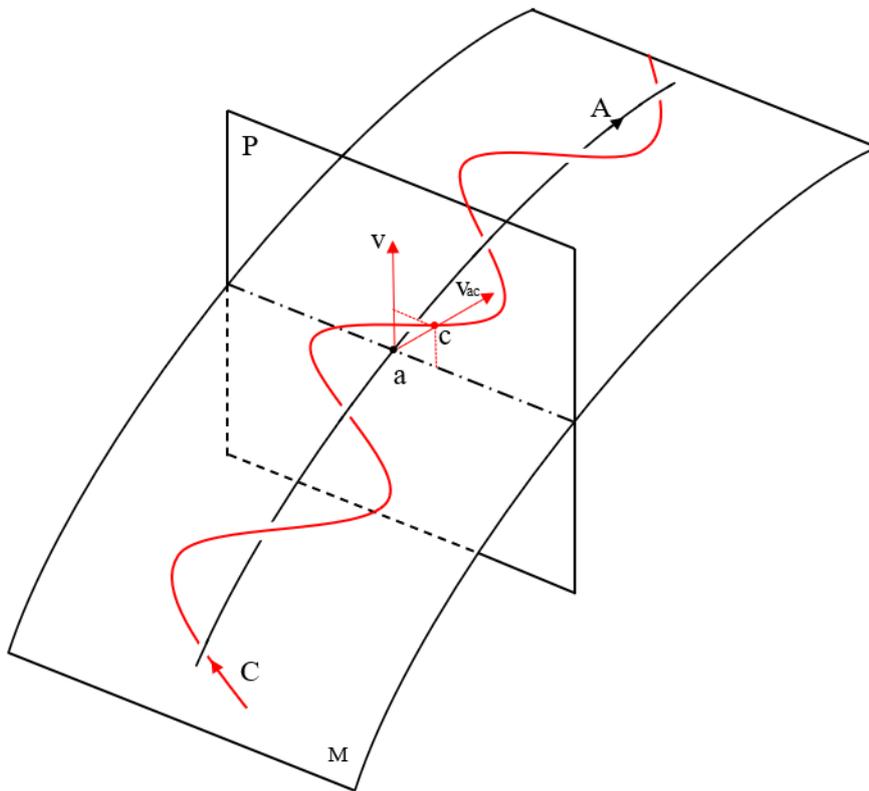


Figura 4.1: Gráfico de una de las hebras del ADN sobre la superficie de una proteína.

v_{ac} también lo será y va girando sobre ella. Para medir las rotaciones sobre A se usará v como referencia dado que al ser el vector normal de M en a , es perpendicular a A en el punto a . Φ será positivo o negativo siguiendo la regla de la mano derecha.

El Winding number es inversamente proporcional a la repetición helicoidal (h), lo cual es bastante intuitivo, cuantos más pares de base se necesiten para que una cadena complete una vuelta sobre la otra, menos vueltas se realizarán en total. Definiremos, entonces, el promedio de la repetición helicoidal como el número total de pares de bases de una molécula de ADN entre el número de vueltas de dicha molécula:

$$h = \frac{N}{\Phi} \quad (4.1)$$

Hablamos de promedio, porque al ser una propiedad biológica no se comporta con exactitud plena. Podría darse el caso, por ejemplo, en el que en una molécula con $\Phi = 2$ la primera vuelta se dé al cabo de 10.49 pares de base y la siguiente vuelta al cabo de 10.51 pb. En esta situación estaríamos hablando de que el promedio de la repetición helicoidal

es 10.5, sin ser este el de cada vuelta. Como no podía ser de otra forma, observando la fórmula anterior, el signo de h sigue la regla de la mano derecha.

El Winding number y la repetición helicoidal de una molécula de ADN relajada se denotan como Φ_0 y h_0 respectivamente, y la relación entre ambos parámetros es la misma:

$$h_0 = \frac{N}{\Phi_0} . \quad (4.2)$$

El Winding number está directamente relacionado con el número de enlace. En el caso de una molécula de ADN relajada sobre un plano, $Wr = 0$, y por tanto el vector v , con el que calcular Φ , tiene siempre la misma dirección y será constante. Para este caso el número de vueltas de la curva C sobre A es igual al número de enlace, $\Phi = Lk$. Sin embargo, para los casos en que el ADN se encuentre sobre una superficie que no sea plana de una proteína, la dirección del vector v generalmente no será constante. Por este mismo motivo el Writhe number no tendrá que ser 0, lo que conlleva que el número de enlace y el Winding number no sean necesariamente iguales, aunque Lk sí dependerá de este último como veremos a continuación.

Número de enlace de una superficie envuelta por una molécula de ADN

A diferencia de lo visto en el resto del trabajo, en este tema entra en juego el eje de la cinta (\mathcal{A}), y esta curva influirá en la mayoría de cálculos.

$$Lk \equiv Lk(C, \mathcal{A}) = Tw(C, \mathcal{A}) + Wr(\mathcal{A}) . \quad (4.3)$$

Como vemos en esta expresión el número de enlace no depende de uno de los bordes de la cinta.

James H. White y William R. Bauer demuestran en su trabajo que para los casos en los que nos encontramos en una superficie, $Tw(C, \mathcal{A})$ se puede dividir en dos parámetros: el Winding number (Φ) y el Twist de la propia superficie (STw) [25]. Este último término

mide como el vector de referencia v cambia.

$$Tw(C, \mathcal{A}) = STw + \Phi . \quad (4.4)$$

Juntando las dos ecuaciones anteriores tenemos que:

$$Lk = STw + \Phi + Wr . \quad (4.5)$$

El Twist de una superficie (STw) es igual al Twist number de una curva que se extiende a lo largo de una superficie sobre esta misma desplazada mínimamente. A continuación vamos a introducir los conceptos necesarios para definir formalmente lo anterior.

Anteriormente vimos el Triedro de Frenet, para este capítulo nos conviene saber un poco acerca del Triedro de Darboux, pero tampoco profundizaremos en él porque tampoco vamos a necesitar expresar todas sus cualidades. El triedro de Darboux se diferencia con el de Frenet en que el vector normal que se utiliza es el normal a la superficie, recordemos que en el de Frenet era el vector derivada del vector tangente.

Por lo tanto, para la curva $a(s)$ en la superficie orientada M :

$$\begin{aligned} T(s) &= a'(s), \\ N(s) &= N_M(s) . \end{aligned}$$

El vector binormal se construirá de igual manera:

$$B(s) = T(s) \times N(s)$$

Resultando $T(s), N(s), B(s)$ una base ortonormal diferenciable y positivamente orientada llamada *triedro de Darboux*.

Las ecuaciones resultantes de derivar los vectores de la base son:

$$\begin{bmatrix} \dot{T} \\ \dot{N} \\ \dot{B} \end{bmatrix} = \begin{bmatrix} 0 & \kappa_n & \kappa_g \\ -\kappa_n & 0 & \tau_g \\ -\kappa_g & -\tau_g & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix} .$$

y se las llama ecuaciones de Darboux. κ_g es la curvatura geodésica y τ_g es la torsión geodésica.

Vamos a asumir que la superficie M es regular en todos sus puntos y por tanto se puede definir plano tangente a M en todo punto y este varía de forma suave entre los puntos de M . Además M es orientable y por tanto tendrá sentido la definición de su vector normal (v) y este será continuo. Usaremos el vector normal unitario. Sea \mathcal{A} una curva suave sobre M , diremos que \mathcal{A}_ϵ es la curva que resulta de desplazar la curva \mathcal{A} una distancia ϵ ($\epsilon \neq 0$) en la dirección del vector normal v . Para evitar conflictos como intersección de las curvas o de la curva desplazada con la superficie escogeremos un ϵ suficientemente pequeño. El valor ϵ puede variar para cada punto de \mathcal{A} pero cambia de una manera suave, por tanto no es una constante sino una función. De esta forma la nueva curva obtenida \mathcal{A}_ϵ es suave.

Definition 4.0.1 (Twist number de una superficie). *Siendo T el vector tangente a \mathcal{A} , entonces el Twist de \mathcal{A}_ϵ sobre \mathcal{A} es:*

$$STw = Tw(\mathcal{A}_\epsilon, \mathcal{A}) = \frac{1}{2\pi} \int_{\mathcal{A}} T \times v \, dv . \quad (4.6)$$

Al argumento de esta integral, $T \times v$ se le llama torsión geodésica de \mathcal{A} en la superficie M . La función ϵ será escogida de siempre positiva o siempre negativa, de cualquier otra forma al ser una función continua existiría una intersección entre \mathcal{A} y \mathcal{A}_ϵ .

EL Twist de \mathcal{A}_ϵ sobre \mathcal{A} es la manera en que queda reflejada la geometría de la superficie en el Twist total. Y como pasa con todas las curvas cerradas, el número de enlace de \mathcal{A}_ϵ con \mathcal{A} es la suma del Twist number de \mathcal{A}_ϵ sobre \mathcal{A} y el Writhe number de \mathcal{A} . A este término se le conoce como número de enlace de la superficie:

$$SLk = Lk(\mathcal{A}_\epsilon, \mathcal{A}) = Tw(\mathcal{A}_\epsilon, \mathcal{A}) + Wr(\mathcal{A}) . \quad (4.7)$$

Escribiendolo en términos de la superficie obtenemos:

$$SLk = STw + Wr . \quad (4.8)$$

En la Figura 4.2 podemos observar de dónde surge la definición de este concepto en tres superficies diferentes y cuál es su valor, esto hará más sencilla su comprensión.

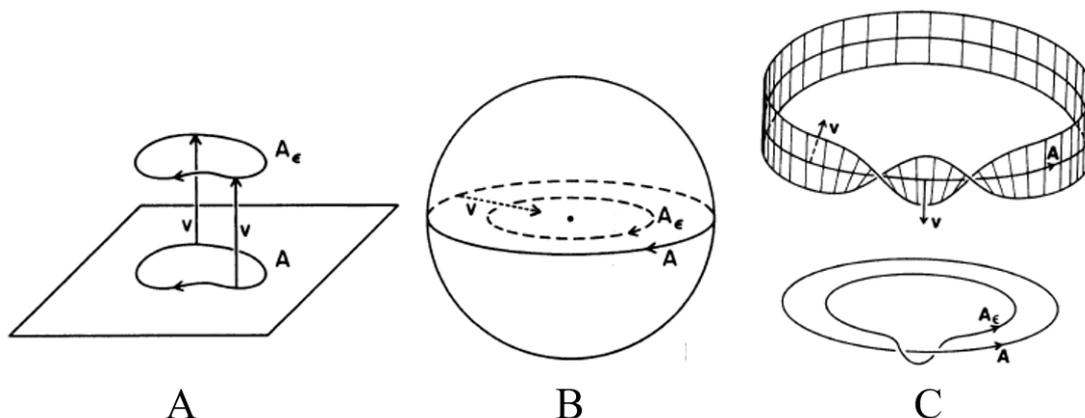


Figura 4.2: Representación sobre un plano, una esfera y una cinta de Möbius respectivamente, de una curva A y su curva desplazada A_ϵ en la dirección del vector normal de la superficie.

Teniendo claro el concepto de número de enlace podemos apreciar que tanto para las curvas en el plano como las curvas en la esfera $Lk(\mathcal{A}_\epsilon, \mathcal{A}) = 0$, pues ninguna curva da alguna vuelta sobre la otra. Por tanto, en ambos casos $SLk = 0$. Por otro lado, vemos que la curva \mathcal{A}_ϵ de la cinta de Möbius da una vuelta sobre \mathcal{A} y es por esto que en este caso $SLk = +1$, siguiendo la regla de la mano derecha para la elección del signo.

Combinando las ecuaciones anteriores podemos expresar el número de enlace de una molécula de ADN sobre una proteína mediante la suma de dos términos que se pueden calcular de manera independiente:

$$Lk = SLk + \Phi . \quad (4.9)$$

SLk se puede calcular mediante la geometría de la superficie y Φ , el Winding Number, puede calcularse mediante sondeos químicos y enzimáticos. Otra manera más precisa de calcular Φ es contando las veces que la curva C , la cual gira alrededor de la curva eje \mathcal{A} , interseca con la superficie M . Este conteo multiplicado por $\frac{1}{2}$ da como resultado Φ . Esto es así porque para cada vuelta que C da sobre \mathcal{A} , C interseca la superficie dos veces. En los casos en que la molécula de ADN se encuentre relajada el número de enlace de la superficie será nulo ($SLk = 0$) y por tanto $Lk_0 = \Phi_0$.

Poder expresar el número de enlace como la suma de dos parámetros geométricos independientes es muy útil principalmente para calcularlo en moléculas de ADN que envuelven una superficie matemática cerrada como son los esferoides y los toroides. Superficies de

revolución que se obtienen, el esferoide al girar una elipse alrededor de uno de sus ejes principales; y el toroide al girar una curva plana cerrada simple sobre un eje de rotación coplanar a la curva que no interseca con ésta. Muchas enzimas donde encontramos ADN son proteínas globulares que quedan muy bien representadas geoméricamente por esferoides. El caso de los toroides es interesante porque simulan la estructura de colecciones de nucleosomas, que no es más que la forma de organización del ADN en las células eucariotas.

Algo que sucede en muchas ocasiones al intentar describir cuerpos de la naturaleza como cuerpos geoméricos de las matemáticas es que no se puede lograr la precisión de detalle de la naturaleza pues sus cuerpos no son perfectos. Sin embargo, esto no es un problema en este ámbito dado que los valores Winding Number y número de enlace de una superficie no sufren modificaciones al aplicar un homeomorfismo a la superficie. Por esto, no será necesario una representación exacta de la superficie sobre la que se encuentra el ADN. Un claro ejemplo se puede observar en la Figura 4.3. Si tenemos una molécula de ADN sobre una superficie cualquiera (virtual o no) de las de la imagen podremos calcular SLk y Φ para dicha molécula sobre cualquier toro con el fin de simplificar los cálculos. Dado que SLk no cambia bajo estas transformaciones y Lk es una propiedad invariante como vimos en este trabajo con anterioridad, de la fórmula $Lk = SLk + \Phi$ se deduce que Φ se mantendrá también constante en estas situaciones.

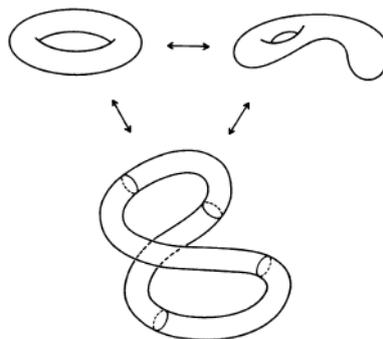


Figura 4.3: Representación de un toro y dos superficies cerradas homeomorfas a este.

Relación entre la repetición helicoidal y la densidad superhelicoidal

Como vimos en secciones anteriores la variación del número de enlace de una molécula circular y cerrada de ADN entre su estado con superenrollamiento y su estado relajado viene dada por $\Delta Lk = Lk - Lk_0$. Combinando esto con el número de enlace calculado desde la perspectiva de la superficie tratada en esta sección obtenemos la expresión:

$$\Delta Lk = SLk + \Phi - Lk_0 . \quad (4.10)$$

Recordemos que la densidad superhelicoidal que actúa como medidor del grado de superenrollamiento relativo viene dada por la fórmula $\sigma = \Delta Lk/Lk_0$. Igual que hemos hecho antes con la variación del número de enlace, podemos reescribir σ en función de los nuevos parámetros SLk y Φ . $\Delta\Phi = \Phi - \Phi_0$ será la variación del Winding Number. Recordando que $Lk_0 = \Phi_0 = N/h_0$ podemos expresar la densidad superhelicoidal de la siguiente forma

$$\sigma = \frac{SLk}{Lk_0} + \frac{\Phi - Lk_0}{Lk_0} = \frac{SLk}{Lk_0} + \frac{\Delta\Phi}{\Phi_0} . \quad (4.11)$$

Vemos aquí que σ es resultado de la suma de dos términos, uno que refleja la contribución del número de enlace de la superficie y el otro que mide la variación relativa del Winding number. Pero también podemos escribir la repetición helicoidal (h) en función del resto de parámetros, ya que próximamente veremos como cambia h según la superficie en la que nos encontremos. Vamos a utilizar que $\Phi = N/h$ y $\Phi_0 = N_0/h_0$ para relacionar la repetición helicoidal con la densidad superhelicoidal:

$$\sigma = \frac{SLk}{Lk_0} + \frac{h_0}{h} - 1 \quad \implies \quad h = \frac{h_0}{\sigma - \frac{SLk}{Lk_0} + -1} . \quad (4.12)$$

Gracias a esta última expresión podemos calcular de manera independiente la repetición helicoidal para cualquier molécula cerrada de ADN que se encuentre en cualquier superficie. Dado que h depende de valores que fluctúan al aplicar homeomorfismos sobre la superficie, dicho parámetro tampoco lo hará

Aplicaciones del análisis del Winding Number del ADN sobre diferentes superficies

Esferoides

El primer caso que vamos a tratar es el de los esferoides. Para cualquier curva cerrada simple \mathcal{A} sobre la superficie de un esferoide el vector normal a la superficie en \mathcal{A} , v , apuntará, por convención, al interior de la superficie y por tanto \mathcal{A}_ϵ se encontrará en el interior. Con esto podemos asegurar que $SLk = Lk(\mathcal{A}_\epsilon, \mathcal{A}) = 0$ en el caso de los esferoides. Con lo que de las ecuaciones (4,9) y (4,12) deducimos que para este tipo de superficies:

$$Lk = \Phi .$$

$$h = \frac{h_0}{\sigma + 1} . \quad (4.13)$$

Es decir, el número de enlace del ADN sobre la superficie de un esferoide será igual al Winding Number, y la repetición helicoidal es una función simple que depende exclusivamente de la repetición helicoidal en estado relajado y de la densidad superhelicoidal.

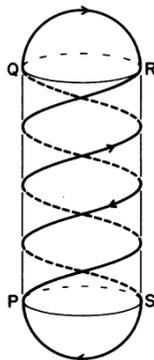


Figura 4.4: Molécula de ADN se enrolla sobre la superficie de un esferoide.

En la Figura 4.4 vemos un caso de ADN sobre un cilindro con casquetes en los extremos, superficie homeomorfa a un esferoide. Sobre esta superficie una molécula levógira de ADN se enrolla siguiendo la regla de la mano derecha y por tanto una orientación positiva, y al llegar al casquete superior yace sobre este hasta llegar del punto Q al punto R y continuar enrollándose positivamente hacia el casquete opuesto. Al llegar abajo se une el punto S con el P para acabar cerrando la molécula de ADN.

De la fórmula (4,13) deducimos que para este tipo de superficies, $h > h_0$ cuando la densidad helicoidal sea negativa ($\sigma < 0$) y viceversa, $h < h_0$ cuando $\sigma > 0$. Dado que para estos casos $SLk = 0$, acudiendo a la ecuación (14), vemos que $Wr = -STw$. Esto se traduce en que cualquier alteración en el Writhe será compensada completamente por el Twist de la superficie. Permaneciendo la repetición helicoidal inalterada pues depende de Φ , y h_0 que son constantes, ya que $Lk = \Phi$ y el número de enlace es un parámetro invariante.

Toroides

El siguiente tipo de superficies que vamos a estudiar son los toros y las superficies homeomorfas a estas. En la Figura 4.3 podemos ver tres ejemplos. Sea \mathcal{A} la curva que define el eje de la cinta de ADN, diremos que se enrolla sobre el toro n veces cuando envuelve el cuerpo del toro n veces. Existe la posibilidad de que la curva no se enrolle sobre el toro, será en los casos en que la curva sea homeomorfa al ecuador o cuando la curva sea homeomorfa a un punto. Este último caso no es de interés en este apartado pues la molécula no estaría envolviendo la superficie del toro. En estos casos, el resultado de definir la curva \mathcal{A}_ϵ será semejante al caso de los esferoides o de un plano, ver Figura 4.2, luego $SLk = 0$. Para los casos en que $n > 0$, si escogemos la orientación de la curva de forma que el vector normal v apunte al interior del toro, escogiendo ϵ adecuadamente, la curva \mathcal{A}_ϵ será la circunferencia directriz o eje central del toro. Por tanto, el número de veces que estas curvas se enrollan sobre \mathcal{A}_ϵ es el mismo que lo hacen sobre el propio toro, $SLk = Lk(\mathcal{A}_\epsilon, \mathcal{A}) = \pm n$, dependiendo el signo de si es una curva levógira (+) o dextrógira (-).

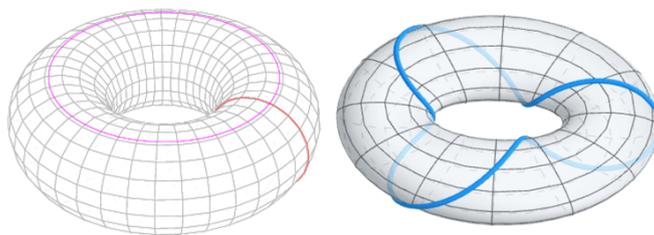


Figura 4.5: Dos toros con tres curvas diferentes sobre su superficie.

En la Figura 4.5, sobre el toro de la izquierda yacen dos curvas, una con $n = 1$, que pasa por el agujero; y la otra que se encuentra en la parte superior de la superficie con

$n = 0$. En el toro de la izquierda podemos observar una curva dextrógira que se enrolla un total de 3 veces sobre la superficie del toro, $n = -3$. Combinando las ecuaciones (15) y (18) obtenemos que para las curvas sobre toroides el número de enlace y la repetición helicoidal son los siguientes:

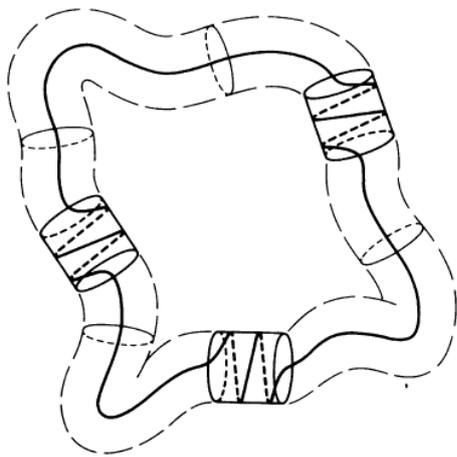
$$Lk = \pm n + \Phi , \quad (4.14)$$

$$h = \frac{h_0}{\sigma - \frac{\pm n}{Lk_0} + 1} . \quad (4.15)$$

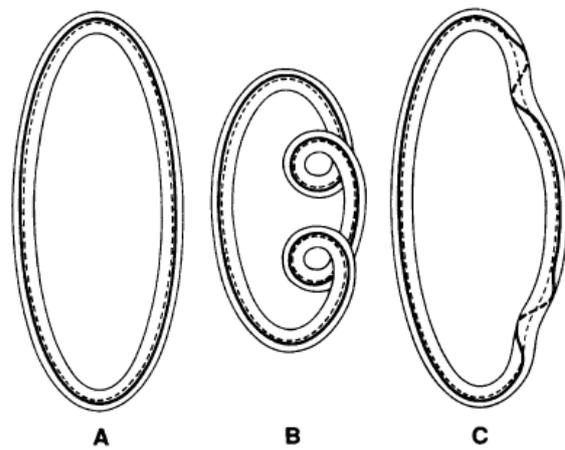
Conjunto de nucleosomas

Este es un caso especial que se puede encontrar en algunos virus, como por ejemplo en los minicromosomas del Papovirus SV-40. Cada nucleosoma se puede representar con un cilindro, igual que las superficies que conectan el ADN. Esto resulta en una superficie representada por una sucesión de cilindros, alternando un cilindro nucleosomático y un cilindro sobre el que descansa el ADN. Al cerrar esta superficie el resultado obtenido es homeomorfo a un toro, ver Figura 4.6a. Es importante que en los cilindros de descanso la molécula de ADN no se enrolle sobre la superficie, y así asegurar que las contribuciones al número de enlace de la superficie, SLk , vienen exclusivamente del ADN en los nucleosomas. De esta manera, el SLk podrá ser producido exclusivamente por retorcimiento del propio toro.

Un ejemplo del retorcimiento de la propia superficie lo podemos observar en la Figura 4.6b. La línea continua representa a la curva d , y la línea discontinua representa el eje central del toro. Sea M y d la terminología usada para las superficies y la curva sobre la superficie respectivamente. En la imagen A el writhe de la curva es nulo, $Wr(d_A) = 0$ y para todas las imagenes $STw = 0$. Entonces $SLk(M_A) = STw(M_A) + Wr(d_A) = 0$. Por otro lado, como $Wr(d_B) = -2$, entonces $SLk(M_B) = STw(M_B) + Wr(d_B) = -2$. Las curvas d_A y d_B no son dado que $SLk(M_A) \neq SLk(M_B)$. Al contrario sucede con B y C, que son equivalentes porque $SLk(M_C) = -2$ también. Esto es fácil de comprobar porque como vimos anteriormente para los toroides $SLk = \pm n$, y en la imagen se puede observar que se trata de un enroscamiento dextrógiro, orientación negativa, de grado 2.



(a) Minicromosoma formado por tres nucleosomas.



(b) Toroide en diferentes grados de retorcimiento.

Figura 4.6

Capítulo 5

Topoisomerasas

Este capítulo está apoyado en las referencias [3], [11], [22] y [26]. Las topoisomerasas son un grupo de enzimas esenciales que hay en los organismos vivos, presentes en bacterias, plantas, animales y humanos. El objetivo principal de estas enzimas es modular la estructura de la doble hélice del ADN, es decir, cambiar la forma en que el ADN se enrolla y se relaciona consigo mismo; a la vez que preservar la estabilidad y las funciones de los ácidos nucleicos, como son el ADN y el ARN. El trabajo de estas enzimas analizado desde un punto de vista matemático es modificar la topología del ADN alterando el valor del número de enlace de modo que el grado de superenrollamiento se mantiene en el adecuado para llevar a cabo procesos biológicos. El número de enlace es una propiedad invariante como ya hemos visto, es por esto que no puede alterarse sin romper de alguna manera la doble cadena. Esto es lo que hacen las topoisomerasas, una rotura temporal en alguna de las hebras del ADN.

Debido a su larga longitud, el ADN tiene que enrollarse de alguna manera para poder caber dentro de la célula. Si no estuviera enrollado, ocuparía demasiado espacio y no podría realizar sus funciones biológicas, ahí es donde entran en juego las topoisomerasas, ya que pueden modificar la topología de las cadenas sin necesidad de desanudarlas. Las topoisomerasas intervienen en los procesos de replicación, transcripción del ADN y en la reparación del ADN dañado. Las topoisomerasas se clasifican en dos tipos, las que cortan una hebra (Tipo I) y las que cortan las dos hebras de la doble cadena (Tipo II).

5.1. Topoisomerasas de tipo I.

Conociendo la importancia del superenrollamiento en los procesos celulares que se producen en el ADN, no es de extrañar que exista una enzima encargada de regular la cantidad de enrollamiento, las topoisomerasas de tipo I. Este tipo de enzimas es capaz de cortar una sola hebra del ADN, de esta forma consiguen liberar tensión producida por el exceso de enrollamiento, volviendo después a unir la hebra cortada consigo misma.

Después de procesos biológicos como la replicación, las dos cadenas del ADN se encuentran excesivamente enroscadas, es decir, el Twist entre ambas es muy alto. Por este motivo, la separación entre las hebras es pequeña, dificultando el acceso a los pares de bases que se encuentran entre las hebras, donde se encuentra la información genética. Aquí es donde entra en juego esta enzima que relajará la tensión helicoidal facilitando el acceso a los pares de bases.

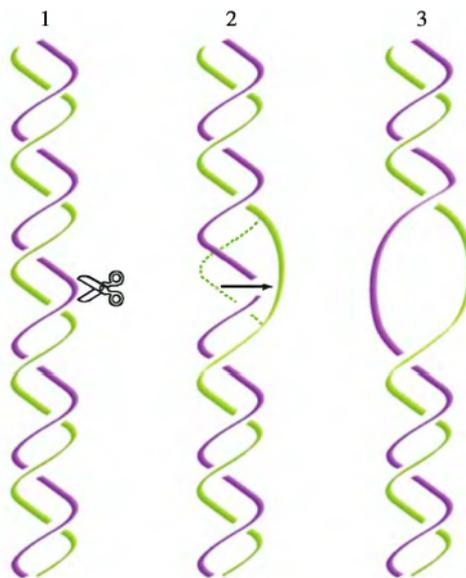


Figura 5.1: Doble cadena del ADN sobre la que ha actuado una enzima topoisomerasa tipo I.

Cuando la topoisomerasa tipo I actúa sobre una doble cadena está sucediendo lo que se aprecia en la Figura 5.1. Una de las hebras sufre un corte temporal, mientras la otra hebra aprovecha esta rotura para deslizarse a una posición con menos tensión helicoidal. Después de eso la enzima se encarga de suturar la hebra. Si analizamos este acontecimiento topológicamente lo que ha sucedido es que el número de giros de una de las hebras sobre la otra ha disminuido, es decir, ha disminuido el Twist number como valor absoluto en

una unidad. En el caso de una molécula de ADN-B, la más común, el sentido de giro es positivo (dextrógiro) y por tanto $\Delta Tw = -1$. Cuando el giro de las cadenas de una molécula sea levógiro, Twist number negativo, entonces $\Delta Tw = +1$.

Como adelantamos antes, el número de enlace también se ve afectado, exactamente en una unidad al igual que el Twist number. Cuando la enzima de tipo I actúa sobre la doble cadena no modifica el Writhe number porque el eje central de la doble cadena no se desplaza y por tanto no aumenta el número de cruces del propio eje con él mismo. El eje central conserva la posición y dirección original, luego $\Delta Wr = 0$. Con esta información podemos aplicar el ya conocido Teorema de Călugăreanu para comprobar la modificación sufrida por el número de enlace según el sentido de giro de la molécula.

Molécula de ADN dextrógira:

$$\Delta Lk = \Delta Tw + \Delta Wr = -1 + 0 = -1 .$$

Molécula de ADN levógira:

$$\Delta Lk = \Delta Tw + \Delta Wr = -1 + 0 = -1 .$$

5.2. Topoisomerasas de tipo II.

Estas enzimas son más complejas que las anteriores, su labor es la de la simplificación topológica mediante el *desanudamiento* o *desenlace* de moléculas. Para conseguir este objetivo las topoisomerasas de tipo II cortan las dos cadenas, quebrantando la molécula vista como una cinta. El hueco producido por la ruptura será aprovechado por la propia cinta para disminuir su tensión fruto del enrollamiento disminuyendo el número de enlace. También puede ser usada por otra molécula porque ambas se encuentren anudadas entre sí.

Cuando se produce el proceso celular de la replicación a partir de cada hebra de la doble cadena del ADN se crea otra doble cadena, resultando en dos moléculas con misma información genética. La enzima de tipo II es la encargada de desanudar las dos dobles cadenas, de no conseguir la división completa de estas dos se produce una muerte celular,

resultando el proceso de replicación en vano.

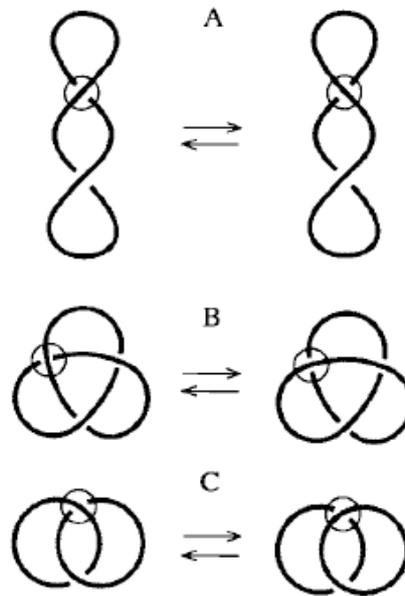


Figura 5.2: Representación de la actuación de una enzima topoisomerasa tipo II sobre el eje central del ADN circular cerrado en tres casos diferentes.

Existen dos motivos por los cuales la topoisomerasa tipo II actúa, reducir tensión procedente del grado de superenrollamiento de una molécula o separar dos moléculas anudadas entre sí, como en la replicación. Desde el punto de vista topológico, el primer caso lo que hace esta enzima es romper la cinta para deshacer un cruce del eje de la cinta con él mismo, resultando en una disminución del valor absoluto del Writhe number de dos unidades. Recordemos que el Writhe number se podía contabilizar con un algoritmo idéntico al del número de enlace, pero sin dividir entre dos, en el que asignamos $+1$ o -1 a cada cruce según lo indicado en la Figura 2.2 y posteriormente sumamos todos los valores. Por tanto, como se observa en la Figura 5.2, tras actuar la enzima de tipo II la parte del eje que se encuentra cruzando por arriba pasa a situarse debajo, lo que se traduce en que los cruces de valor $+1$ pasarán a valer -1 y viceversa, $\Delta Wr = \pm 2$. En este caso el Twist number no se ve modificado porque la enzima no afecta al retorcimiento de una hebra sobre la otra, luego $\Delta Tw = 0$. Como consecuencia el efecto de las topoisomerasas sobre el número de enlace es:

$$\Delta Lk = \Delta Tw + \Delta Wr = 0 \pm 2 = \pm 2 .$$

Bibliografía

- [1] National Institutes of Health. (9 de febrero 2023). Ácido desoxirribonucleico. National Human Genome Research Institute. URL: <https://www.genome.gov/es/genetics-glossary/%C3%81cido-desoxirribonucleico>
- [2] National Institutes of Health. (9 de febrero 2023). Nucleótido. National Human Genome Research Institute. URL: <https://www.genome.gov/es/genetics-glossary/Nucleotido>
- [3] Vologodskii, A. (2017). Topology and physics of circular DNA. CRC Press.
- [4] Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., y Laipis, P. (1965). The twisted circular form of polyoma viral DNA. *Proceedings of the National Academy of Sciences*, 53(5), 1104-1111.
- [5] Stewart, I. (1995). *Concepts of Modern Mathematics*. Dover Publications.
- [6] Watson, J. D., Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.
- [7] Llerena, I. (2021). *Nudos, enlaces, ADN y helicidad*.
- [8] Xiao, L. W. *Introducción a la teoría de nudos*. Universidad de Cantabria.
- [9] Reidemeister, K. (1927, December). Elementare begründung der knotentheorie. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* (Vol. 5, pp. 24-32). Springer-Verlag.
- [10] Wikipedia. (11 de septiembre 2021). Reidemeister move. URL: https://en.wikipedia.org/w/index.php?title=Reidemeister_move&oldid=1043695845
- [11] Adams, C. C. (1994). *The knot book*. American Mathematical Soc..

-
- [12] Wikipedia. (27 de mayo 2022). Direccionalidad. URL: <https://es.wikipedia.org/w/index.php?title=Direccionalidad&oldid=143823378>
- [13] Khan Academy. (s.f). Descubrimiento del ADN. URL: <https://cdn.kastatic.org/ka-perseus-images/111350b04ad2940d68a7e362875031c395eb5785.png>
- [14] Pohl, W. F. (1980). DNA and differential geometry. *The Mathematical Intelligencer*, 3(1), 20-27.
- [15] Buck, D. (2009). DNA topology. Applications of knot theory (Proc. Sympos. Appl. Math., 66, Amer. Math. Soc., 2009), 47-79.
- [16] Chirikjian, G. S. (2013). Framed curves and knotted DNA. *Biochemical Society Transactions*, 41(2), 635-638.
- [17] Dennis, M. R., Hannay, J. H. (2005). Geometry of Călugăreanu's theorem. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 461(2062), 3245-3254.
- [18] Klenin, K., Langowski, J. (2000). Computation of writhe in modeling of supercoiled DNA. *Biopolymers: Original Research on Biomolecules*, 54(5), 307-317.
- [19] Chirikjian, G. S. (2013). Framed curves and knotted DNA. *Biochemical Society Transactions*, 41(2), 635-638.
- [20] Inspiralis. (17 de enero 2023). The DNA Intercalation Assay. URL: <https://www.inspiralis.com/assets/TechnicalDocuments/Linking-Number-and-the-Intercalation-Assay2.pdf>
- [21] Qu, A., James, D. L. (2021). Fast linking numbers for topology verification of loopy structures. *ACM Trans. Graph.*, 40(4), 106-1.
- [22] Castán García, A. (2017). Dinámica de la topología del DNA durante la replicación.
- [23] White, J. H., Cozzarelli, N. R., Bauer, W. R. (1988). Helical repeat and linking number of surface-wrapped DNA. *Science*, 241(4863), 323-327.
- [24] White, J. H., Bauer, W. R. (1988). Applications of the twist difference to DNA structural analysis. *Proceedings of the National Academy of Sciences*, 85(3), 772-776.
- [25] White, J. H., Bauer, W. R. (1988). Applications of the twist difference to DNA structural analysis. *Proceedings of the National Academy of Sciences*, 85(3), 772-776.

-
- [26] Castán García, A. (2017). Dinámica de la topología del DNA durante la replicación.