## Benemérita Universidad Autónoma de Puebla

# FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS LICENCIATURA EN FÍSICA

BÚSQUEDA DE LAS ESTRUCTURAS TRIPLEX DEL ADN FORMADAS ENTRE LOS PARES DE WATSON Y CRICK Y LAS BASES CANONÍCAS

**TESIS** 

QUE PARA OBTENER EL TÍTULO DE LICENCIADO EN FÍSICA

PRESENTA ACEVEDO GONZÁLEZ JOSÉ FILIBERTO

DIRECTOR DE TESIS Dr. GONZÁLEZ JIMÉNEZ EDUARDO

PUEBLA, PUE.

11 de Noviembre del 2014

#### Dedicatoria

Le dedico la tesis presente a mi padre , el Dr. José Filiberto Acevedo Sanchez, por haberme apoyado y motivado a continuar mis estudios de licenciatura en física, también le dedico esta tesis a mi sobrina Irasema de León Acevedo cuyo ejemplo de tenacidad y sacrificio me hicieron seguir adelante aún en los tiempos mas difíciles de mi carrera.

## Agradecimientos

Agradezco a la FCFM-BUAP por haberme dado la oportunidad de realizar y concluir mis estudios de licenciatura en física.

Agredezco al Dr. Eduardo González Jiménez por haberme dado todo su apoyo y la oportunidad de ser parte de este interesante proyecto.

### Introducción

El ácido desoxirribonucleico (ADN) es altamente flexible y puede adoptar una gran variedad de estructuras helicoidales dependiendo de la secuencia, solvente, ambiente iónico, temperatura o la presencia de ligandos [1]. Algunas transcripciones entre las diferentes formas de ADN ocurren bajo condiciones fisiológicas y pueden influir significativamente en el funcionamiento del ADN. Algunas de las estructuras se presentan como multi-cadenas de oligonucleótidos, como las triples y tetra- hélices [2, 3]. Las triples hélices se forman cuando bases purinas o pirimidinas ocupan el surco mayor de doble hélice del ADN formando pares de Hoogsteen con las purinas de los pares de Watson- Crick. Éstas pueden formarse por cadenas de ADN o de ácido ribonucleico (ARN), o bien por combinación de ambas y se les denomina estructuras H.

Este tipo de estructuras inusuales del ADN son adoptadas bajo condiciones de superhelicidad. Experimentalmente se encontraron en regiones homopurinas-homopirimiodinas en plásmidos super-helicoidales [4].

Una descripción detallada de las estructura de este tipo de hélices se inicia con el estudio de las componentes individuales más simples que son la interacción de 3 bases de las cuales dos de ellas forman el par de Watson y Crick. El objetivo de este trabajo es el estudio de la estructura y las características energética de los posibles mínimos de interacción de triplex de bases de los ácidos nucleicos formados con los pares de Watson y Crick y otra base (A, T, G y C) del ADN, para ello se utilizará la metodología de la Mecánica Molecular que emplea las funciones de potencial de Poltev. La tesis se organiza primeramente haciendo en el capítulo 1 una revisión de los conceptos biológicos de los ácidos nucleicos, a continuación en el capítulo 2 se describe la metodología a utilizar, y posteriormente en el capítulo 3 se obtienen las estructuras de valores mínimos de energía de configuraciones: planas, apiladas y perpendiculares, en el par Adenina-Timina (A:T) y Guanina-Citosina (G:C) con todas las bases A, G, T y C del ADN. Por último se da la conclusión y la bibliografía del tema.

# Índice general

	Intr	roducción	I
1.	Esti	ructura y Función de los Ácidos Nucleicos	1
	1.1.	El ácido desoxirribonucleico ADN	1
	1.2.	Las bases nitrogenadas	3
	1.3.	Enlaces de hidrógeno y los pares de Watson y Crick	4
	1.4.	Pares incorrectos y mutaciones	9
	1.5.	Triple hélice del ADN	11
2.	Med	cánica molecular	17
	2.1.	Coeficientes de Van der Walls y tipos de átomos para el po-	
		tencial de Poltev	20
	2.2.	Enlaces de Hidrógeno	22
	2.3.		23
3.	Res	ultados de los cálculos de los pares de bases de Watson y	
	Crie	ck con las bases canónicas del ADN (triplex de bases)	<b>25</b>
	3.1.	Mínimos de interacción más profundos del par A-T de Watson-	
		Crick y las bases de los ácidos nucleicos	25
		3.1.1. Configuraciones Planas	26
		3.1.2. Configuraciones fuera del plano	28
	3.2.	Mínimos de interacción más profundos del par G-C de Watson-	
		Crick y las bases de los ácidos nucleicos	31
		3.2.1. Configuraciones planas	31
		3.2.2. Configuraciones fuera del plano	34
Co	onclu	sión	38

IV	ÍNDICE GENERAL
Bibliografía	39
Índice alfabético	42

## BÚSQUEDA DE LAS ESTRUCTURAS TRIPLEX DEL ADN FORMADAS ENTRE LOS PARES DE WATSON Y CRICK Y LAS BASES CANONÍCAS

Acevedo González José Filiberto

11 de Noviembre del 2014

## Capítulo 1

# Estructura y Función de los Ácidos Nucleicos

#### 1.1. El ácido desoxirribonucleico ADN

Los ácidos nucleicos (AN) son portadores de la información genética y se encuentran en todas las células. Hay dos tipos de ácidos nucleicos: el ácido desoxirribonucleico (ADN) y el ácido ribonucleico (ARN). En el ADN se encuentran las instrucciones génica usadas en el desarrollo y funcionamiento de todos los organismos vivos conocidos y algunos virus, y es responsable de su transmisión hereditaria. Los segmentos de ADN que llevan esta información genética son llamados genes, pero también las otras secuencias de ADN tienen propósitos estructurales o toman parte en la regulación del uso de esta información genética.

El ADN es un polímero de nucleótidos, es decir, un polinucleótido. Un polímero es un compuesto formado por muchas unidades simples conectadas entre sí, como si fuera un largo tren formado por vagones. En el ADN, cada vagón es un nucleótido, y cada nucleótido, a su vez, está formado por un azúcar (la desoxirribosa), una base nitrogenada que puede ser: adenina (A), timina (T), citosina (C) o guanina (G) y un grupo fosfato que actúa como enganche de cada vagón con el siguiente (ver Figura 1). Lo que distingue a un vagón (nucleótido) de otro es, entonces, la base nitrogenada, y por ello la secuencia del ADN se especifica nombrando sólo la secuencia de sus bases. La disposición secuencial de estas cuatro bases a lo largo de la cadena (el

ordenamiento de los cuatro tipos de vagones a lo largo de todo el tren) es la que codifica la información genética: por ejemplo una secuencia de ADN puede ser ATGCTAGATCGC.

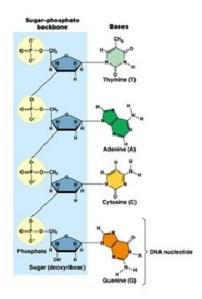


Figura 1. Estructura de la secuencia de nucleótidos en una cadena del ADN

Para que la información que contiene el ADN pueda ser utilizada por la maquinaria celular, debe copiarse en primer lugar en unos trenes de nucleótidos, más cortos y con unas unidades diferentes, llamadas ARN.

El ARN es el ácido ribonucleico (ARN o RNA en ingles) también formado por una cadena de ribonucleótidos. Está presente tanto en las células procariotas como en las eucariotas, y es el único material genético de ciertos virus (virus ARN). El ARN celular es lineal y de hebra sencilla, pero en el genoma de algunos virus es de doble hebra.

En los organismos celulares el ARN desempeña diversas funciones. Es la molécula que dirige las etapas intermedias entre la síntesis proteica y el ADN ya que este no puede actuar solo, y se vale del ARN para transferir esta información vital durante la síntesis de proteínas (producción de las proteínas que necesita la célula para sus actividades y su desarrollo). Varios tipos de ARN regulan la expresión génica, mientras que otros tienen actividad catalítica.

El ARN es, pues, mucho más versátil que el ADN.

#### 1.2. Las bases nitrogenadas

La estructura de las bases nitrogenadas que aparecen en los ácidos nucléicos está relacionada con la estructura del benceno, un hidrocarburo aromático que consta de un anillo de seis átomos. El anillo fundamental de la base nitrogenada es el anillo de pirimidina, en el cual el N reemplaza a dos C-H en las posiciones 1 y 3 del anillo bencénico

Los otros tipos de bases en los nucleótidos son las purinas, las cuales están muy relacionadas con las pirimidinas. Una molécula de purina consta de un anillo de pirimidina al que se le ha unido un anillo de 5 átomos, imidazol. La citosina y la timina son las dos pirimidinas que se encuentran usualmente en el ADN y Uracil en ARN (Figura 2), mientras que la Adenina y Guanina son purinas (Figura 3).

La citosina es una de las cinco bases nitrogenadas que forman parte de los ácidos nucleicos (ADN y ARN) y en el código genético se representa con la letra C. su fórmula semidesarrollada es:  $C_4H_5N_3O$ . Pertenece a la familia de bases nitrogenadas conocida como pirimidinas, que tienen un anillo simple. En el ADN y ARN la citosina se empareja con la guanina por medio de tres enlaces de hidrógeno. Forma los nucleósidos citidina (Cyd) y desoxicitidina (dCyd), y los nucleótidos citidilato (CMP) y desoxicitidilato (dCMP).

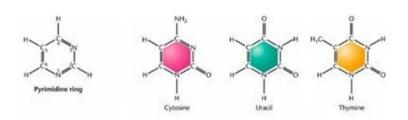


Figura 2. Estructura de la Pirimidina y sus derivados: la Citosina, Uracil y Timina

La Timina es un compuesto heterocíclico derivado de la pirimidina su

fórmula semidesarrollada es  $C_5H_6N_20_2$ . Forma el nucleósido Timidina (dThd) y el nucleótido Timidilato (dTMP). Es única en el ADN, no existe en el ARN. En el ADN, se empareja mediante dos enlaces o puentes de hidrógenos con su base complementaria Adenina.

La Guanina es una base nitrogenada púrica, en el código genético se representa con la letra G. su fórmula es  $C_5H_5N_5O$ . Esta forma los nucleósidos Guanosina (Guo) y Desoxiguanosina (dGuo) y los nucleótidos Guanilato (GMP) y Desoxiguanilato (dGMP). La guanina siempre se empareja en el ADN con la Citosina mediante tres puentes de hidrógeno. Además es una de las bases más importantes de los ácidos nucleicos.

La Adenina se encuentra en los dos ácidos nucleicos (ADN y ARN) y en el código genético se representa con la letra A. En el ADN la Adenina siempre se empareja con la Timina. Forma los nucleósidos Adenosina (Ado) y Aesoxiadenosina (dAdo), y los nucleótidos Adenilato (AMP) y Desoxiadenilato (dAMP). En la bibliografía antigua. Su fórmula es  $C_5H_5N_5$ . Es un derivado de la purina (es una base púrica) en la que un hidrógeno ha sido sustituido por un grupo amino (NH2). También forma parte de la molécula de trifosfato de adenosina, que constituye la fuente principal de energía a nivel celular, y está presente en muchas sustancias naturales como la remolacha, el té y la orina.



Figura 3. Estructura de la Purina y sus derivados: Guanina y Adenina

# 1.3. Enlaces de hidrógeno y los pares de Watson y Crick

El reconocimiento mutuo de A por T y de C por G se hace a través de enlaces de hidrógeno, estos determinan la fidelidad de la transcripción y translación de la información del ADN. Los grupos N-H de las bases son buenos donadores de enlaces de hidrógenos (D), mientras los pares de electrones híbridos sp2 sobre los grupos de los oxígenos de las bases C=0 y sobre los nitrógenos son mucho mejores aceptores de enlaces de hidrógenos (A) que los oxígenos de cualquiera de los fosfatos o las pentosas. Los enlaces de hidrógenos formados en A-D (aceptor-donador) tienen una carga de alrededor de +0.2e sobre los hidrógenos y alrededor de -0.2e sobre los oxígenos y nitrógenos y parecen tener una fuerza promedio de 6-10 kJ/mol [1].

Cuando Watson y Crick efectuaban la construcción los modelos del DNA en 1953 [2], reconocieron que la capacidad de los enlaces de hidrogeno de un par de bases de A:T usan complementariedad de ()-A-D a A-D-A, mientras un par de C:G usa la complementariedad de A-A-D a D-D-A. Este patrón del par de bases rápidamente se convierte como pares Watson-Crick.

En un par A:T hay dos enlaces de hidrógenos y tres en un par C:G (ver Figura 4). La geometría de estos pares ha sido completamente analizada en muchas estructuras desde fosfatos dinucleótidos hasta oligonucleótidos para especies de ARNt, ambos utilizando cristalografía de rayos X y, después, por espectroscopia NMR.

Figura 4.-Apareamiento de las bases de Watson-Crick para C:G (izquierda) y A:T (centro). Emparejamiento de base Hoogsteen para A:T (derecha)

En el plano de los pares de bases, los enlaces de hidrogeno se unen a los átomos de nitrógenos y oxígenos que están separados de  $2.8\dot{A}$  a  $3.4\dot{A}$ . Esta geometría da una distancia entre los átomos de carbono de C-1'-C-1'de  $10.79\dot{A}$  con ángulos de  $52.65^{\circ}$ y  $52.02^{\circ}$  entre los dos enlaces glicosilicos y la línea que une a los carbones C-1'y C-1'para los pares de bases A:T y para los pares C:G se tiene respectivamente  $10.80\dot{A}$  con  $55.23^{\circ}$  y  $52.03^{\circ}$  [3]

esto se muestra en las Figuras 5 y 6. A estos pares se les llama pares de bases canónicos, normales , naturales o de Watson y Crick, en la Figura 6 presentamos las características de estos pares.

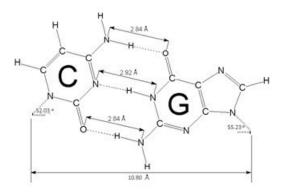


Figura 5. Geometría del par C:G de Watson y Crick

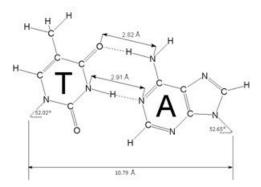


Figura 6. Geometría del par T:A de Watson y Crick

Como un resultado de esta geometría isomorfa, solo las combinaciones de las 4 pares de bases A:T, T:A, C:G y G:C de Watson y Crick pueden ser incluidos dentro de una cadena regular de ADN Figura 7. A esta propiedad de las bases se le llama complementaridad, siempre a una secuencia de bases le corresponde su complementaria. Ésta es la clave para proteger la información genética, para la replicación del ADN, la transferencia de la información y la biosíntesis de las proteínas como se esquematiza en la Figura 8.

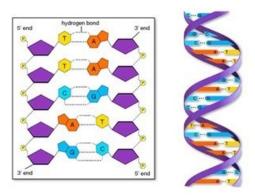


Figura 7.Esquema que muestra la estructura de la doble hélice del ADN

La mayoría de los procesos celulares son efectuados por las proteínas, entonces la célula debe cuidar que se mantenga la integridad de la secuencia de bases, ya que éstas codifican una proteína, esta es la base del Código Genético (Figura 8).

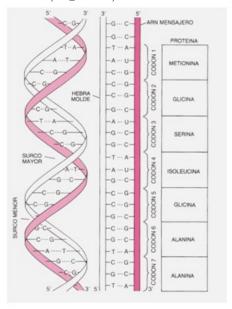


Figura 8.-Biosintesis del ADN y las proteínas

Por la estructura de la doble hélice, hay regiones de hendiduras llamadas surcos (mayor y menor) en los cuales se permiten el acceso a varios grupos

químicos de las bases en determinadas secuencias (genes), estos grupos son reconocidos por proteínas. Los grupos son mostrados en la Tabla 1,1 y en la Figura 9.

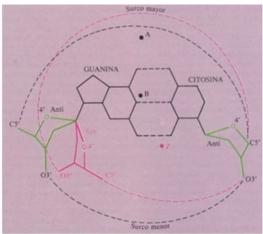


Figura 9 Estructura del par G:C que muestra la región de los surcos mayor y menor

Tabla 1,1 Grupos químicos de las bases de los AN								
que	que son expuestos en los surcos de la doble hélice							
Base Surco Mayor Surco Menor								
Adenina	N6 amino (H61), N7	N3						
Citosina	N4 amino (H41)	C2 carbonil (O2)						
Guanina	C6 carbonil (O61), N7	N2 amino (H21), N3						
Timina	C4 carbonil (O41), C5 metil	C2 carbonil (O2)						
Uracil	C4 carbonil (O4)	C2 carbonil (O2)						

Por ejemplo la trascripción de la secuencia de ADN a ARN es iniciada en la región promotora la cual contiene una secuencia específica (TATA) que es reconocida por un complejo de trascripción. Este complejo es una proteína que se une al ADN e inicia el auto ensamblaje, éste consiste de 10 o más proteínas denominado ARN-polimerasa II (RNApol II). Este complejo abre el ADN formando la horquilla de replicación que transcribe la secuencia codificada es decir el gen.

Un gen puede ser trascripto muchas veces y este puede producir muchas proteínas al pasar su RNA por varios ribosomas. Entonces éstos complejos también son factores reguladores, siendo este el mecanismo que tiene la célula para regular la expresión del gen.

Una distorsión física de la doble hélice (flexión o plegamiento) puede influen-

ciar en la interacción proteína-ADN, la distorsión puede obstruir los grupos químicos de reconocimiento de las bases. También estas secuencias específicas pueden ser removidas de la región codificadora y pueden no ser localizadas por las proteínas o bien algunas bases de la secuencia pueden ser sustituidas cambiando el patrón de reconocimiento o regulatorio. A esta sustitución de bases se llama una mutación. Si un gen que codifica a una proteína reguladora es mutado entonces esta mutación puede ser la diferencia entre una célula normal y una disfuncional.

El cáncer es una distorsión del comportamiento celular, las células cancerosas difieren de una normal en el mal funcionamiento del proceso que controla la proliferación celular, esto tiene que ver con el fallo de las proteínas reguladoras del ciclo celular que esta codificada en la secuencia del gen.

#### 1.4. Pares incorrectos y mutaciones

Mientras los pares-base Watson-Crick son el patrón dominante, otros pares han sido sugeridos de los cuales los más significativos son los pares Hoogsteen y los pares incorrectos "wobble". Entre estos se encuentran los pares Hoogsteen los cuales son ilustrados para A:T en la Figura 4, no sonisomorfos como los pares Watson-Crick porque ellos tienen un ángulo de  $80^{\circ}$  entre los enlaces glicosilicos y una separación de  $8,6\dot{A}$  entre los carbonos anoméricos. En el caso de los pares Hoogsteen invertidos y los pares Watson-Crick, una base es rotada a través de  $180^{\circ}$  relativo al otro.

Francis Crick propuso la existencia de los pares de bases "wobble" para explicar la degeneración de los códigos genéticos. Este fenómeno corresponde para una sola base en el posición 5′ anticodon del RNAt para ser habilitado para reconocer las pirimidinas o, alternativamente, las purinas como su base compañera 3′-codon. Un par "wobble" G:U tiene dos enlaces de hidrógeno entre los átomos (G)N1-H —O2(U) y O6 (G)—H-N3(U) y esto requiere un cambio lateral de una base relativa a su posición en la geometría regular de Watson-Crick (Figura 10). La pérdida resultante de un enlace de hidrógeno conduce a reducir la estabilidad, lo cual puede ser compensado en parte por el apilamiento mejorado de la base, que resulta del desplazamiento de la base.

La complementariedad expresada en el apareamiento específico de los pares

A:T y G:C, es esencial para la transferencia y la expresión de la información genética. Pero los pares incorrectos pueden ser introducidos en la doble hélice sin destruir ésta, incluso sin distorsionarla significativamente.

Los pares de bases incorrectos son significativos en tres situaciones estructurales. Primero, la estructura compacta de RNAt maximiza los pares de bases y la base apilada donde sea posible. Esto ha permitido la identificación de una considerable variedad de pares de Hoogsteen y wobble invertidos. Segundo, donde hay hélices triples unidas de ADN y ARN, como por ejemplo en la union (poly(dA)) y (poly(rG) 2poly(rC), la segunda cadena de pirimidina se une a la purina en el surco mayor con enlaces Hoogsteen y se dirige paralelo a la cadena de purina. Tercero, pares mal apareados son necesariamente identificados con anomalías de enlaces de hidrogeno y muchos de los cuales han sido encontrados por estudios de Rayos X en oligodeoxyribonucleotidos sintéticos.

Figura 10.-Emparejamiento "Wobble" para G:U (izquierda), I:U (centro) y I:A (derecha)

La formación de estos pares puede explicar los errores en la replicación, reparación, trascripción, traslación y regulación de la información genética. La inclusión de un par incorrecto dentro de la doble hélice puede ocurrir por diversos mecanismos, uno de ellos que es el más ampliamente aceptado, corresponde a la formación del par con dos enlaces de hidrógeno con las bases normales de los ácidos nucleicos, el otro mecanismo discutido en este trabajo es la formación del par donde una de las bases es un tautómero raro.

Se realizaron cálculos de la energía de interacción con el potencial empírico de Poltev-Malenkov [3] encontrándose que para cada par coplanar de bases existen varios mínimos locales de energía en los cuales la posición mutua del enlace glicosil difiere de los pares A:T y G:C por no más de  $3\dot{A}$ . Algunos de los pares incorrectos propuestos fueron descubiertos experimentalmente en hélices de ADN en cristales (por ejemplo en [4]).

#### 1.5. Triple hélice del ADN

Aunque, la forma B del ADN es la forma más estable de ADN, La forma dúplex de ADN puede asumir una forma de triple hélice debido a la tensión negativa superhelicoidal. Esta forma se estudió por primera vez en una secuencia de un espaciador entre los genes de las histonas de erizo de mar, y fue descrito como forma H de ADN, ya que se estabiliza por enlaces de hidrógeno (una estructura protonada).

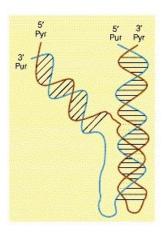


Figura 11. Representación de la formación de la estructura triple del ADN

Los nucleótidos que participan en un par de bases (Figura 11) pueden formar un enlace de hidrogeno adicional, principalmente con grupos funcionales dispuestos en el surco mayor; por ejemplo, una Timidina puede aparearse con la adenosina de un A:T.

Los átomos N7, O6, N6 de las purinas (ver Tabla 1,1) que participan en los enlaces de hidrógeno en los triplex del ADN son comúnmente referidos como posiciones Hoogsteen o apareamientos de no-Watson-Crick. Karst Hoogsteen en 1963 [5] fue el primero que reconoció el potencial de estos pares inusuales (Figura 4).

Figura 12. Estructuras de triplex bien caracterizadas en la triple hélice del ADN

Una estructura de ADN de triple cadena es una estructura de ADN en el que tres oligonucleótidos se dirigen una alrededor de la otra y forman una triple hélice. En esta estructura, una hebra se une a una doble hélice de ADN de forma B a través de enlaces de hidrogeno de Hoogsteen o inversos de Hoogsteen.

Por ejemplo, una base Timina se une a un apareamiento de bases de Watson-Crick de T:A por enlaces de hidrógeno de Hoogsteen entre un par A:T o bien una citosina protonada en N3, representado como C\*, también puede formar una base de triplete con un par C:G a través de la apareamiento de bases de Hoogsteen de un G:C\* (Figura 12). Por lo tanto, las triples hélices de ADN que utilizan estos apareamientos de Hoogsteen constan de dos cadenas de homopirimidinas y de homopurinas, y la tercera hebra de homopirimidina la cual es paralela a la cadenas de homopurinas.

Una tercera cadena de homopurina que se puede unir a una dúplex homopurina: homopirimidina usando pares inversos de Hoogsteen . En este triplex, una base nitrogenada A se une a un par de bases A:T y una G para un par C:G . Dado que las bases nitrogenadas en la tercera cadena se tienen que invertir, la tercera cadena de homopurina es antiparalela a la de homopurina del dúplex original.

Cuando la hebra central es rica en bases purínicas, la tercera hebra forma pares inversos de Hoogsteen y la orientación de esta tercera hebra es antiparalela (H\*-ADN); mientras que si la cadena central es rica en bases

pirimídicas, la tercera cadena forma pares de Hoogsteen y su orientación es paralela (H-ADN).

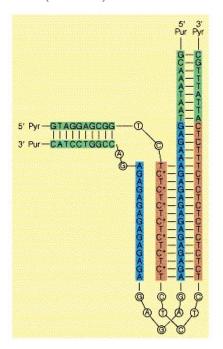


Figura 13. Ilustración de la formación del triple apareamiento de bases (Y\*UY)

Para que se pueda formar una triple hélice se requieren de ciertas condiciones que se deben cumplir, tales como:

- -Tener una región de homopurina-homopiridina en la hebra central.
- -Dicha región debe de tener una simetría de espejo de repetición.



Figura 14. Espejo de repetición

La estabilidad de estos tripletes depende de varios factores como:

- -La estabilidad intrínseca de los nucleótidos.
- -Los enlaces de hidrógeno que estén formando los pares de Hoogsteen.

- -La repulsión entre los grupos fosfato de las moléculas.
- -Las interacciones de apilamiento.

La triple hélice ha sido identificada in vivo en la molécula de ARN y se ha podido sintetizar la estructura de la triple hélice de ADN in vitro. Sin embargo también se puede formar una triple hélice combinando ambas moléculas (ADN y ARN).

Los tripletes se pueden formar en una molécula que contenga una sola hebra de polinucleótidos (intramolecular) o por la interacción de varias hebras de polinucleótidos (intermolecular). Ambos tipos de tripletes se pueden formar bajo ciertas condiciones.

Muchas secuencias en el genoma humano tienen estructuras que son potenciales a formar triples hélices. Estas secuencias están asociadas a regiones reguladoras de gen.

El superenrollamiento provee de la energía suficiente para desenrollar al ADN y permitir la formación de la triple hélice. Incluso se puede formar la estructura triplex cuando el superenrollamiento se lleva a cabo en un medio con pH neutro.

Estas estructuras triples se han encontrado in vivo en los telómeros de los cromosomas de bacterias Escherichia coli . Cada vez que se produce un enrollamiento negativo (en sentido contrario de las manecillas del reloj) y por cada 10,5 pb en una región de homopurina-homopitrimidina, se forma una triple hélice intramolecular .

Las triples hélices también han sido sintetizadas in vitro; su formación suele ser rápida en un medio con un rango de pH de 5 a neutro; el tiempo de vida media de esta triple hélice es menor de 2 minutos.

La transición de la estructura doble a la triple pude actuar como un interruptor que determina si algunas proteínas reguladoras se enlazan o no al ADN, ya que sobre el H-ADN, los factores de transcripción no pueden unirse al ADN, esto tiene como consecuencia que el gen no pueda ser expresado.

Por primera vez en 2012 se pudo conseguir por un grupo de españoles [6]

información estructural fidedigna de la triple hélice del ADN en fase gaseosa, es decir, en condiciones prácticamente en el vacío. Con esto se ha conseguido caracterizar esta estructura y demostrar que mantiene una sorprendente memoria de cómo era en su entorno biológico, en solución acuosa, donde normalmente es muy difícil caracterizar.

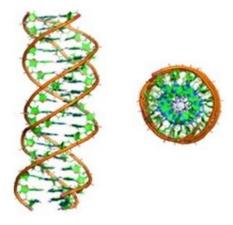


Figura 14a. Estructura de triple hélice reportada en la referencia [6]

## Capítulo 2

### Mecánica molecular

La Mecánica Molecular (MM) es una parte del modelado molecular que implica el uso de Mecánica Clásica/Mecánica Newtoniana para describir las bases físicas de los modelos. La mecánica molecular describe la energía de una molécula en términos de un conjunto de funciones de energía de potencial. Las funciones de energía potencial y los parámetros usados para su evaluación se conocen como campo de fuerza. Los estudios con campo de fuerza empíricos en macromoléculas biológicas se están convirtiendo en una herramienta común para investigar la relación estructura-actividad a un nivel atómico detallado. Tales estudios facilitan la interpretación de los datos experimentales y permiten acceder a información que no es fácilmente accesible con los métodos experimentales. Una gran parte del éxito de estos métodos se basa en la calidad de los campos de fuerza (un refinamiento continuo) combinados con los nuevos algoritmos que permiten la reproducción más precisa de las observables experimentales.

Los métodos de la mecánica molecular se basan en los siguientes principios:

- -El núcleo y los electrones están juntos y se tratan como una única partícula. Es decir los modelos moleculares describen normalmente a los átomos (núcleos y electrones en conjunto)
- -Estas partículas se tratan como si fueran esferas suaves.
- -Los enlaces entre partículas se comportan como osciladores armónicos.
- -Las interacciones no-enlazantes entre estas partículas se tratan utilizando funciones de potencial derivadas de la mecánica clásica.
- -Las funciones de energía potencial se relacionan con parámetros derivados

empíricamente (constantes de fuerza y valores de equilibrio) que describen las interacciones entre los átomos.

Se utilizan funciones de potencial individuales para describir las diferentes interacciones: tensión de enlaces, flexión de ángulos y energías torsionales, así como interacciones no enlazantes. Los átomos no-enlazantes, interaccionan a través de atracciones de van der Waals, repulsiones estéricas, y atracciones y repulsiones electrostáticas. La suma de todas las interacciones determina la distribución espacial (conformación) de un conjunto de átomos (molécula).

Las energías obtenidas en la mecánica molecular no son cantidades absolutas sino que sólo se pueden utilizar para comparar energías estéricas relativas entre dos o más conformaciones de la misma molécula.

Los cálculos de mecánica molecular contemplan las contribuciones de la energía potencial de las moléculas debido a la tensión del enlace  $(U_{ten})$ , deformación del ángulo de enlace  $(U_{def})$ , rotación interna alrededor del enlace, también llamada torsión  $(U_{tor})$ , interacciones de van der Waals que comprenden atracciones y repulsiones de entre los átomos no enlazados  $(U_{vdw})$  e interacciones electrostáticas entre los átomos  $(U_{elec})$ .

La suma de estas contribuciones produce la energía potencial  $U_{Total}$ , para una colectividad y conformación molecular determinada.

$$U_{Total} = U_{tension} + U_{deformacion} + U_{torsion} + U_{vdW} + U_{elec}$$

Este potencial depende exclusivamente de los parámetros estructurales internos de la cadena y comúnmente toma la siguiente forma:

con:

l: longitudes del enlace químico

 $\theta$ :ángulo de valencia.

 $\phi$ : ángulo de torsión

como se muestra en la Figura 15.

Las constantes  $K_i$  son constantes de Fuerza. El potencial electrostático y de Van der Waals son las interacciones no enlazantes  $U_{nb}$ . Las constantes de fuerzas de estrechamiento  $K_l$  y doblamiento  $K_{\theta}$ , así como los pesos de la

barrera de torsión  $K_{\phi}^{n}$ , son frecuentemente transferibles a especies relativamente similares. Es decir, se asume que los parámetros obtenidos por análisis espectroscópico y termodinámico de un compuesto-modelo, son directamente aplicables a la cadena polinucleótida. Desafortunadamente, las bases de datos experimentales sobre ácidos nucleicos relevantes es bastante limitada. Entonces los parámetros necesarios son estimados, si es posible, en base a las determinaciones de los métodos ab en mecánica cuántica.

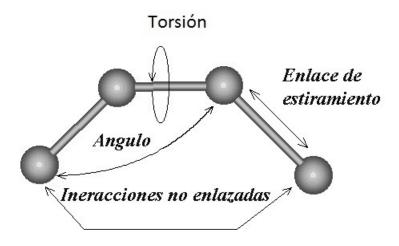


Figura 15. Diagrama que representa los términos del potencial usado en mecánica molecular.

Los sistemas moleculares que contienen la información genética, son complicados en su estudio. La construcción de campos de fuerza para estos sistemas es una de las tareas de los métodos de la simulación computacional aplicada a la Biofísica. En Mecánica molecular, la calidad de los campos de fuerzas se mide por la precisión de sus predicciones es decir, por su realismo. Los diferentes campos de fuerzas difieren comúnmente en sus parámetros y algunas veces en su forma matemática, pero sus resultados deben ser comparados con los resultados experimentales o de la mecánica cuántica. Existe un gran conjunto de funciones de potencial que describen diferentes sis-

Existe un gran conjunto de funciones de potencial que describen diferentes sistemas moleculares y diferentes características de estos sistemas. El conjunto de funciones de potencial para determinar las interacciones entre fragmentos de los ácidos nucleicos y, que empleamos en este trabajo fue elaborado en la década de los noventas por el grupo de investigación del Dr. V. I. Poltev [7], estos fueron recientemente optimizados para satisfacer los últimos resultados experimentales y de Mecánica Cuántica [3, 8].

# 2.1. Coeficientes de Van der Walls y tipos de átomos para el potencial de Poltev

El potencial con el cual trabajamos fue introducido por el Dr. Poltev en 1967 para pequeñas moléculas orgánicas. Como en este trabajo solo se utilizan los términos para las interacciones no-enlazantes solo se discute el término  $U_{nb}$ . Inicialmente el término atractivo tenia forma exponencial. En 1979 este fue cambiado por el termino de Lennard-Jones,  $r^{-6}$  y en 1984 el potencial fue presentado como el potencial de Poltev-Malenkov para la interacción de agua con fragmentos de ácidos nucleicos y se introdujo el termino  $r^{-10}$  para describir las interacciones entre átomos que forman los enlaces de hidrógeno [7]. El potencial desde entonces tiene la forma actual.

Ecuaciones (1)

$$V_{enlacedehidrogeno}(r_{ij}) = K \frac{q_i q_j}{\varepsilon r_{ij}} - A_{ij}^{(10)} r_{ij}^{-10} + B_{ij}^{(10)} r_{ij}^{-12}$$
$$V_{Vanderwaals}(r_{ij}) = K \frac{q_i q_j}{\varepsilon r_{ij}} - A_{ij} r_{ij}^{-6} + B_{ij} r_{ij}^{-12}$$

Donde  $r_{ij}$  son las distancias entre los átomos i y j;  $q_i$  y  $q_j$  son las cargas de los átomos i y j; K es la constante eléctrica y  $\varepsilon$  es la constante dieléctrica.  $A_{ij}$  y  $B_{ij}$ ,  $(A_{ij}^{(10)}$  y  $B_{ij}^{(10)}$  son los parámetros de parte atractiva y repulsiva de Van der Waals (de enlaces de hidrógeno) y dependen del tipo de átomos, de su estado de valencia, y de los átomos vecinos.

Para obtener las cargas electrostáticas de los átomos, estos se calculan por el método de Huckel para cargas  $\pi$  y para las cargas  $\sigma$  por el método Del Re, con los parámetros de Berthod y Pullman. Con tales métodos los momentos dipolares son reproducidos para una serie de componentes de los ácidos nucleicos respecto a los resultados experimentales.

En la Tabla 2 y 3 presentamos los parámetros para las funciones del potencial, estos parámetros se ajustaron en una base a series de cálculos para sistemas de diferente complejidad (desde pequeños cristales de moléculas orgánicas hasta fragmentos de ADN). La elección de estas funciones de potencial está basada en consideraciones teóricas y el ajuste de los parámetros se hizo en base a los últimos datos experimentales.

Los átomos en las moléculas sistema se clasifican en diferentes tipos, de acuerdo a su posición en la molécula o a sus vecinos cercanos, por ejemplo H1 es el átomo de hidrógeno unido covalentemente a N o O y que pueden formar enlaces de hidrógeno, H2, está unido a los átomos de carbono en los anillos y H3 es el hidrógeno del grupo metilo.

Los carbonos C4 y C5 son los carbonos aromáticos y alifáticos respectivamente, los tres nitrógenos corresponden a: N6 del grupo amino, N7 y N8 son nitrógenos del los anillos pirrol y piridin respectivamente. O9 es el oxígeno carbonilo mientras O11 es el oxígeno hidroxilo, aunque este último átomo no fue utilizado en este trabajo pero por completo se incluye ya que corresponde a los tautómeros enol de las bases, de importancia en mutagénesis.

En la Tabla 2 y 3 se muestran los coeficientes  $A_{ij}$  y  $B_{ij}$  que corresponden a las ecuaciones 1. Los átomos de las bases se muestran en la primera y última fila y columna respectivamente, por ejemplo el coeficiente a de la primera de las ecuaciones 4, es para la interacción entre los átomos N7 de una base y C5 de otra es igual a 400, pero si la interacción es entre N7 y H1 entonces el coeficiente A es de 126 pero para la segunda ecuación 4, ya que ésta interacción constituye un enlace de hidrógeno.

La buena concordancia con los resultados experimentales permite a los coeficientes ser recomendados para los cálculos en interacciones.

Ta	Tabla 2. Coeficientes $A_{ij}$ de las funciones de potencial								
Átomo	H1	H2	Н3	C1	C2	N1	N2	N3	О
H1	35	35	40	76	100	80	126	9073	10667
H2	35	40	40	100	126	105	126	146	121
Н3	40	40	40	100	126	105	126	146	121
C1	76	100	100	250	316	264	316	367	305
C2	100	126	126	316	400	477	400	464	385
N1	80	105	105	264	477	280	334	391	455
N2	126	126	126	316	400	334	400	440	616
N3	9073	146	146	367	464	391	440	550	465
0	10667	121	121	305	385	455	616	465	400

Tabla 3. Coeficientes $B_{ij}$ de las funciones de potencial									
Átomo	H1	H2	Н3	C1	C2	N1	N2	N3	О
H1	5914	5914	7740	53936	25686	29229	24151	27008	31084
H2	5914	7740	7740	70600	61700	53700	53700	58700	17800
Н3	7740	7740	7740	70600	81600	62000	71400	78300	42200
C1	533939	70600	70600	512000	601000	464000	538000	598000	349000
C2	25686	61700	81600	601000	704000	947869	630000	699000	406000
N1	29229	53700	62000	464000	947869	421000	488000	544000	663737
N2	24151	53700	71400	538000	630000	488000	565000	537323	815959
N3	27008	58700	78300	598000	699000	544000	537323	705000	413000
0	31084	17800	42200	349000	406000	663737	815959	413000	240000

### 2.2. Enlaces de Hidrógeno

Los enlaces de hidrógeno (hydrogen bonds HB) es un tema muy importante dentro del estudio de las conformaciones de los ácidos nucleicos. Como se sabe es el tipo de enlace más débil dentro de los enlaces químicos existentes. Pero a pesar de esto es esencial para mantener estables las conformaciones más importantes. Un HB se forma entre el grupo OH o NH, que es donador de electrones de una molécula y un átomo electronegativo N u O que es aceptor de electrones, ver Figura 16

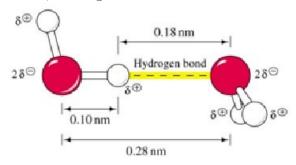


Figura 16. Características del enlace de hidrógeno en moléculas de agua

Existen ciertos criterios que permiten distinguir una interacción HB de un contacto de Van der Waals (VdW) dentro de cada potencial. Estos criterios pueden ser de carácter tanto geométricos como energéticos. Aunque los últimos suelen ser algo complicados y no son de uso generalizado, en cambio el criterio geométrico es más empleado.

Los HB se forman en agua, lo cual conlleva a la formación de cúmulos de moléculas de agua. Los enlaces HB son indispensables para la formación de los pares de bases nitrogenadas dentro del ADN. Los HB entre los átomos hidrofílicos de las bases determinan la estabilidad de los pares.

También los HB son importantes en el caso de las interacciones entre las bases de ADN y moléculas de agua, lo cual influye en la estructura de ADN y en algunos casos puede aumentar posibilidades de errores durante la transmisión del código genético.

La aproximación estrecha del donador y aceptor en un HB es favorecido por

las interacciones electrostáticas fuertes que ocurren cuando dos grupos son apropiadamente alineados. Las interacciones electrostáticas en estos casos son de dos a cuatro veces más grandes que los valores en las interacciones de VdW ordinarias, ya que la constante dieléctrica es considerada igual a uno. Para compensar la predominancia atractiva de las interacciones electrostáticas, el potencial de HB tiene que ser ajustado incrementando los factores repulsivos de escala  $A_{ij}$  en el potencial de Lennard Jones.

En muchos casos el ajuste necesario se logra cambiando el exponente de  $r_{ij}$ . Dentro de los parámetros también se está considerando implícitamente la dependencia del ángulo en un HB.

Por ejemplo en el potencial de Poltev-Malenkov [7] las interacciones HB se calculan con la expresión segunda de la ec 1. A veces La función  $U_{HB}$  es usada en vez de  $U_{VdW}$  para cada par de átomos capaces de formar HB entre ellos.

Un criterio para la consideración de un enlace de hidrógeno es cuando suponemos que dos moléculas (agua, base, etc.) se encuentran entre si, a una distancia O...O u O...N menor que 3,2 Å y la distancia O...H ó N...H es menor que 2,4 Å.

#### 2.3. Procedimiento de cálculo

Para calcular los mínimos de energía entre pares de bases y una base individual ocupamos el programa RANBAS6.exe que utiliza técnicas de minimización estándar. Este programa inicialmente fija un par de bases y mediante un procedimiento aleatorio coloca la tercera base, en torno al par, y se inicia el proceso de minimización.

El programa destinado para los cálculos entre las interacciones de los dos complejos de bases funciona en el sistema de coordenadas normales. Durante el cálculo se obtiene la energía potencial de interacción de las bases y sus derivadas a continuación se realiza el movimiento de las bases hasta que estas llegan a la posición de un mínimo local.

El movimiento de las bases se produce de la siguiente manera: primero, las bases se trasladan en las tres coordenada x, y y z, siendo los parámetros de salida  $\Delta x$ ,  $\Delta y$  y  $\Delta z$ , estos son los desplazamientos totales respecto a la posición inicial de la primera base, mientras la segunda base puede girar al-

rededor de estos tres ejes siendo  $\Delta \phi_{xy}$ ,  $\Delta \phi_{xz}$  y  $\Delta \phi_{yz}$  los valores finales de los ángulos de rotación alrededor de los ejes z, y y x respectivamente (en la Figura 17 se muestra un ejemplo de esto pero con la molécula de cafeína).

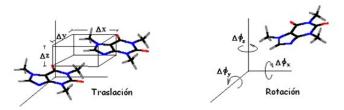


Figura 17. Movimientos de las moléculas durante el proceso de minimización

Originalmente las coordenadas iníciales de las bases serán leídas de un archivo de entrada, el cálculo se realiza con estas coordenadas iniciales. Durante este trabajo el centro de coordenadas del sistema se traslada al centro geométrico de la base que se fija y respecto a la cual rota la otra base. Esto se hace para que los valores de  $\Delta\phi_{xy}$ ,  $\Delta\phi_{xz}$  y  $\Delta\phi_{yz}$  sean independientes de los desplazamientos  $\Delta x$ ,  $\Delta y$  y  $\Delta z$ .

Por medio del programa HIPERCHEM vemos las figuras de las bases clasificando los complejos de un par con la base en bases en un plano, en par de bases y la otra base en posiciones apiladas y en posiciones casi perpendiculares.

Una vez localizadas las figuras planas, apiladas y perpendiculares se escogen empezando por los niveles de energía más altos de forma decreciente.

Los complejos que tengan el mismo valor de energía se eliminan y dejamos aquella que tenga la mayor energía en valor absoluto. Después editar nuestra imagen del par de bases

## Capítulo 3

# Resultados de los cálculos de los pares de bases de Watson y Crick con las bases canónicas del ADN (triplex de bases)

Utilizando la optimización de la energía por el método de Mecánica Molecular, se obtuvieron los mínimos locales de la interacción de los pares de Watson y Crick con las cuatro bases canónicas. En la sección 3,1 se buscaron los mínimos con el par A:T y en la sección 3,2 con el par G:C. Se cuantificaron los enlace de hidrogeno que se forman entre las bases. Para que pueda considerarse la formación de un enlace de hidrógeno, la distancia entre los átomos electronegativos involucrados debe ser menor que 3,2 Å y la distancia entre el átomo electronegativo y el átomo de hidrógeno debe ser menor que 2,4 Å [3].

### 3.1. Mínimos de interacción más profundos del par A-T de Watson-Crick y las bases de los ácidos nucleicos

Como ya se menciono los mínimos en esta sección se construyen teniendo al par de bases A-T (WC) fijo en un sistema de coordenadas cartesianas y una tercera base nitrogenada (A, T, G o C) moviéndose en torno al par mi-

nimizando la energía respecto a las coordenadas y a los ángulos alrededor de las coordenadas.

En las tablas siguientes muestran los valores de energía, enlaces de hidrógeno si los hay, así como la longitud de los mismos para todos los mínimos de interacciones entre bases nitrogenadas. Ordenamos los mínimos en dos grupos: mínimos en posiciones mutuas casi planas y en mínimos fuera del plano.

#### 3.1.1. Configuraciones Planas

En la Tabla 3,1 mostramos algunas de las configuraciones planas (o cuasiplanas) más energéticas del par A-T con las bases purinas: guanina y adenina.

Tabla 3,1 Resultados obtenidos para las configuraciones planas de los mínimos más profu	ndos
de energía de interacción entre el par A-T de WC y las bases purinas: A y G.	

A-T con Guanina						
Mínimo	E (kcal/mol)	Enlace	L(A)			
L956	-10.43	GN1AN7	3.03			
		AN7GH1	2.03			
		AN6GO6	2.82			
		GO6AH62	1.89			
L2662	-10.16	GN1AN7	2.99			
		AN7GH1	1.99			
L1877	-9.76	GN2AN7	2.9			
		AN7GH21	1.91			
L439	-9.30	GN2TO4	3.04			
		TO4GH21	2.12			
		GN1AN6	3.27			
		AN6GH1	2.49			

			A-T con Adenina
Minima	E (legal/mod)	Enlose	

Minimo	E (kcal/mol)	Enlace	L (A)
L3780	-11.53	AN6ÁN7	2.83
		ÁN7AH62	1.89
		ÁN6AN7	2.87
		AN7ÁH62	1.92
L6	-11.19	AN6ÁN1	2.85
		ÁN1 AH62	1.91
		ÁN6AN7	2.88
		AN7ÁH62	1.89
L2987	-10.11	AN6ÁN7	2.77
		ÁN7AH62	1.91
		ÁN6TO4	2.87
		TO4ÁH62	2.02
L2891	-10.10	AN6ÁN7	2.77
		ÁN7AH62	1.91
		ÁN6TO4	2.88
		TO4ÁH62	2.02
L2545	-9.21	ÁN6TO4	2.91
		TO4ÁH62	1.94
		AN6ÁN1	2.77
		ÁN1AH62	1.97
7 1	T 11 00		1 / 1 1 1 1 1

En la Tabla 3,2 se tienen los mínimos con las bases de las pirimidinas: citosina y timina. Una primera observación es que los mínimos más profun-

dos ocurren cuando la base se enlaza en el surco mayor, esto parece natural ya que en esta región es más amplia para interaccionar. Por otro lado comparando las dos Tablas se observa que los mínimos son más profundos (en aproximadamente más de 1 kcal/mol) en las configuraciones con las base A y T que con las G y C a pesar de que en estas últimas se tiene la base con más centros hidrofílicos, esto se debe a una repulsión entre los grupos de átomos con igual carga, como por ejemplo en los oxígenos de la configuración A.T-C, como se muestra en la Figura 3,1.

Tabla 3,2 Resultados obtenidos para las configuraciones planas de los mínimos más profundos						
de energía de interacción entre el par A-T de WC y las bases pirimidinicas C y T.						
A-T con Citosina						
Mínimo	E (kcal/mol)	Enlace	L(A)			
L1570	-10.88	CN4AN7	2.87			
		AN7CH41	1.88			
		AN6CN3	2.87			
		CN3AH62	1.91			
L1501	-9.55	CN4AN7	2.90			
		AN7CH41	1.90			
L3193	-9.21	CN4TO4	2.90			
		TO4CH41	2.91			
		AN6CO2	2.99			
		CO2AH62	2.00			
L1737	-9.17	CN4TO4	2.90			
		TO4CH41	1.91			
		AN6CO2	3.00			
		CO2AH62	2.01			
L2623	-8.04	CN4AN3	2.88			
		AN3CH42	1.93			
L1636	-7.79	CN4TO2	2.89			
		TO2CH41	1.9			
		A-7	Γ CON TIMINA			
L2962	-12.03	AN6T´O2	2.9			
		T'O2AH62	1.92			
		TÑ3AN7	2.86			
		AN7T′H3	1.86			
L62	-11.83	T'N3AN7	2.86			
		AN7T'H3	1.87			
		AN6T 'O4	2.89			
		T´O4AH62	1.92			
L2278	-7.55	T'N3TO4	3.09			
		TO4T´H3	2.1			
		AN6T 'O2	2.76			
		T'O2AH62	1.97			
L3785	-7.48	T´N3TO4	3.08			
		TO4T´H3	2.09			
		AN6T´O2	2.76			
		T´O2AH62	1.98			
L1952	-6.13	T'N3TO2	2.92			
11302	0.10	TO2T´H3	1.93			
L4407	-6.06	AN6T´O2	2.87			
D4401	-0.00	T'O2AH62	1.91			
L4845	-5.59	AN6T 'O2	2.87			
L4049	-5.59	ANU1 UZ	2.01			

En la Figura 3,1 se muestran los mínimos más profundos para las cuatro bases, los enlaces de hidrógeno se representan con las líneas punteadas, estos

mínimos corresponden con las Tabla 3.1 y 3.2. Se observa que en todas las estructuras que se obtienen, no contienen más de 2 enlaces de hidrogeno. Mínimos con un solo enlace también pueden llegar a ser energéticos (en valor absoluto) como es el caso del mínimo L2662 que es solo 0.26 kcal/mol menor que el mínimo global con 2 enlaces de hidrogeno.

Estas estructuras son importantes ya que puede existir esta intercalación de la bases en una cadena de ADN, como ocurre en las estructuras H de triples hélices.

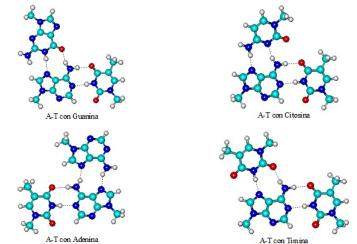


Figura 3.1 Configuraciones planos de mínima energía de los pare A-T de Watson y Crick y las bases de los ácidos nucleicos. Las estructuras corresponden con las Tabla 3.1 y 3.2.

#### 3.1.2. Configuraciones fuera del plano

La gran mayoría de los mínimos encontrados tienen configuraciones de este tipo, siendo estos los más energéticos. Estos mínimos los ordenamos en dos grupos, primero los mínimos apilados (A) donde los planos de las bases son paralelos y generalmente se traslapan la mayoría de sus superficies, la naturaleza de esta interacción es de tipo p. El segundo grupo corresponde a los mínimos cuyos planos forman hasta un ángulo cercano a 90°, estos mínimos nombrados perpendiculares (P) se unen generalmente con un enlace hidrogeno es amplio el espectro de este tipo de estructuras. En la Tabla 3,3

Tabla 3,3 Configuraciones de complejos en sus mínimos más profundos de energía de interacción de los pares de bases A-T con las bases purinas Guanina y Adenina, en posiciones apiladas y perpendiculares

	A-T con Guanina				
Mínimo	E (kcal/mol)	Tipo	Enlace (A)		
P2529	-13.81	P	GN6AN3 2.86		
			AN3GH62 1.91		
			GN2AN7 2.88		
			AN7GH22 1.88		
P503	-13.44	Р	GN1AN7 2.83		
			AN7GH1 1.93		
P109	-13.43	Р	GN2AN7 2.83		
			AN7GH21 1.93		
			GN1AN7 2.83		
			AN7GH1 1.93		
A2862	-12.56	A	-		
P509	-12.86	Р	AN6GO6 2.89		
			GO6AH62 1.94		
			GN1AN7 2.87		
			AN7GH1 1.88		
A3073	-12.49	A	-		
A-T con Adenina					
P1899	-11.61	Р	AN6ÁN7 2.85		
			AN7AH62 1.89		
			AN6AN7 2.86		
			AN7AH62 1.91		
A3962	-10.95	A	-		

Tabla 3,4 Configuraciones de complejos en sus mínimos más profundos de energía de interacción de los pares de bases A-T con las bases pirimidinas:

Citosina y Timina ,en posiciones apiladas y perpendiculares

A990

A794

P23

-10.92

-10.77

-10 13

Α

	Citosina y Tinina ;en posiciones apinadas y perpendiculares					
•	A-T con Citosina					
Mínimo	E (kcal/mol)	Tipo	Enlace (A)			
A4497	-12.59	A	-			
A1115	-12.58	A	-			
A4175	-12.44	A	-			
A2196	-12.33	A	-			
P1406	-11.60	Р	CN4TO2 2.92			
		Р	TO2CH41 1.97			
P4791	-11.51	Р	CN4TO2 2.92			
		Р	TO2CH41 1.97			
	A:T con Timina					
A4164	A4164 -10.64 A -					

	A4164	-10.64	A	-
	A3351	-10.61	A	-
ĺ	A1506	-10.54	A	-
	P3607	-9.12	P	T´N3TO2 2.89
İ			Р	TO2T'H3 1.96
ĺ	P5000	-9.10	Р	T´N3TO2 2.9
			P	TO2T´H3 1.96

Los valores de la energía de los mínimos globales fuera del plano para los mínimos apilados y los perpendiculares son muy cercanos entre sí. Las diferencias entre mínimos planos y apilados son de 1,25, 0,66, 0,99 y 1,52 para G, A, C y T respectivamente.

Se encontraron tres conjuntos de valores de mínimos locales de interacción

de configuraciones: Planas, Apiladas y Perpendiculares.

Las configuraciones planas de los mínimos de energía (en valor absoluto) entre A:T de Watson y Crick con las bases, se presentan con energías en Kcal/mol de 10,43 (guanina), 10,88 (citosina), 11,53 (adenina), 12,03 (timina).

Las configuraciones planas más estables se forman con las bases T y A en el surco mayor del par y en el mínimo más profundo para formar dos enlaces en cada caso.

En las configuraciones apiladas como sucede en los pares de bases [3], las distancias entre los planos son de 3.4 Å aproximadamente. Los mínimos más profundos tienen a las bases arriba (abajo) de la timina del par A-T.

Los mínimos que corresponden a la configuración perpendicular, con mayor energía en (valor absoluto) se forman en el surco mayor con dos enlaces para la guanina y adenina, y un enlace para citosina y timina. Una cantidad considerable de configuraciones sitúan a la tercera base sobre el par, pero estas configuraciones no son posibles en la doble hélice.

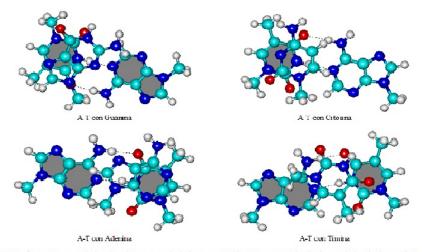


Figura 3.2 Configuraciones apilados de mínima energía de los pare A-T de Watson y Crick y las bases de los ácidos nucleicos. Las estructuras corresponden con las Tabla 3.3 y 3.4.

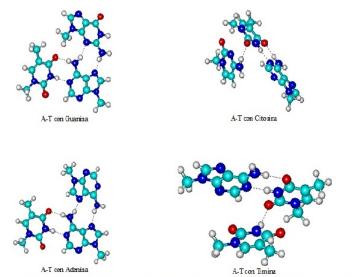


Figura 3.3 Configuraciones perpendiculares de mínima energía de los pare A-T de Watson y Crick y las bases de los ácidos nucleicos. Las estructuras corresponden con las Tabla 3.3 y 3.4.

### 3.2. Mínimos de interacción más profundos del par G-C de Watson-Crick y las bases de los ácidos nucleicos

En un estudio anterior de Mecánica Molecula, se realizó una investigación detallada de la interacción entre bases normales de los ácidos nucleicos. Se obtuvo un conjunto de los posibles pares de bases planos de mínima energía. Entre estos pares se encuentran los pares G:C y de A:T de WC, y de Hoogsteen así como sus inversos [3]. De estos se analizó y describió su geometría. Estos pares de bases se utilizaron para nuestros cálculos para los triplex de bases. En una tesis anterior también se realizo un estudio de triplex de bases solo con el par G-C [13], este estudio es la base para esta parte del trabajo.

#### 3.2.1. Configuraciones planas

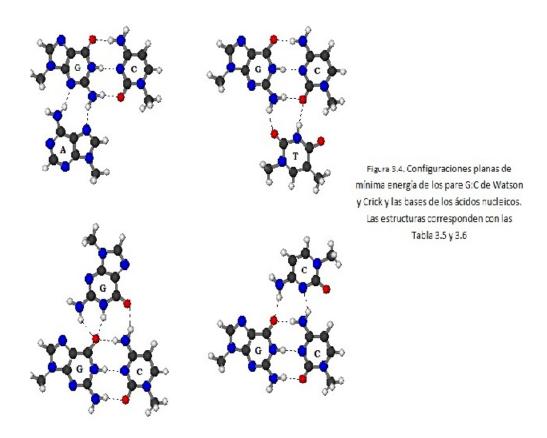
En la Tabla 3,5 mostramos algunas de las configuraciones planas (o cuasiplanas) más energéticas del par G-C con las bases purinas: guanina y adenina. En estos pares con la base guanina se forma el par más profundo (mínimo global) con -22,4 kcal/mol, los mínimos mas energéticos se muestran en la Figura 3,4.

A diferencia de los pares A-T, en los pares G:C, las bases A y T en los mínimos se situaron en el surco menor mientras que para las configuraciones de G Y C (que fueron las estructuras más energéticas) la tercera base se situó del lado del surco mayor, lo que le da una mayor estabilidad a la molécula debido a que esta tercera base tiene un impedimento estérico menor que el que pudiera tener si se ubicara del lado del surco menor. De hecho, en la bibliografía [9,10] se menciona que al formarse el H-ADN la tercera hebra de polinucleótidos se inserta en el surco mayor de la doble hélice.

Tabla 3,5 Configuraciones planas de los mínimos más profundos de energía de interacción entre el par G-C de WC y las bases purinas: A y G

G-C con Guanina						
3.57 :						
Mínimo	E (kcal/mol)	Enlace	L (A)			
GPLA1	-22.40	GO6GH1	1.62			
		GO6GH21	1.91			
		CH62GO6	1.98			
		GN7GH21	2.36			
GPLA6	-18.44	GN7H1	1.90			
		GN7GH21	2.11			
		GO6GH21	2.21			
GPLA3	-18.40	GN7GH1	1.86			
		GN7GH21	2.69			
		GO6GH21	1.98			
GPLA1209	-14.66	CH5GN7	2.79			
		CH4GN7	2.62			
		CH4GO6	2.35			
		CH31GO6	2.68			
GPLA1526	-13.96	GH91GO6	2.94			
GPLA803	-12.63	CO2GH21	1.96			
GPLA2	-11.62	CH62GN7	1.88			
		CH5GN7	2.96			
		CH5GO6	2.37			
		G-C con Aden	ina			

Mínimo	E (kcal/mol)	Enlace	L (A)
APLA1864	-13.39	GO6AH91 2.73	
		CH62AN3	1.93
APLA3385	-13.22	GO6H91	2.65
		CH62AN3	1.93
APLA4871	-12.14	GN7AH61	1.89
APLA839	-11.21	GH22AN3	1.88
APLA3208	-8.99	GO6AH8	2.44
		CH62AN7	1.94
APLA2423	-8.94	GN3AH91	2.86
		GH22AN3 1.90	
		CO2AH2	3.03
APLA3773	-8.83	GO6AH61	1.94
		CH62AN7	1.94



También se obtuvo una configuración en la cual uno de los átomos formó un enlace bifurcado (G-C...G), lo que le adjudica una mayor estabilidad a la molécula y por lo tanto tiene una mayor energía (en valor absoluto).

Las configuraciones planas de los mínimos de energía entre G-C de Watson y Crick con las bases, se presentan con energías en Kcal/mol de -22,40 (guanina), -16,99 (citosina), -13,39 (adenina), -9,72 (timina).

Tabla 3,6 Configuraciones planas de los mínimos más profundos de energí	a
de interacción entre el par G-C de WC y las bases pirimidinicas C y T	

	G-C con Citosin	a	
Mínimo	E (kcal/mol)	Enlace	L (A)
CPLA81	-16.99	GO6CH61	1.92
		CH62CN7	1.92
		CH62C02	2.48
		CH5CO2	2.82
CPLA666	-13.63	GN3CH61	1.89
CPLA829	-13.27	GH22CN7	1.88
		GH22CO2	2.99
CPLA232	-11.50	GN7CH62	1.93
		CO6CH62	2.97
		CO6CH61	1.97
CPLA720	-11.13	GH22CO2	2.51
		GH22CN7	1.94
		CO2CH61	1.90
CPLA53	-10.75 6	GN7CH62	1.89
		GN7CH5	2.99
		GO6CH5	2.33
CPLA773	-10.73	GN7CH5	2.58
		GO6CH5	2.86
		GO6CH62	1.89
CPLA560	-9.78	CH62CO2	1.89
		CH5CO2	2.50
		CH5CN7	2.80
CPLA61	-9.61	GN3CH62	2.30
		GH22CN4	2.78
CPLA68	-9.25	GN7CH61	1.89
CPLA990 -8.90	CO2CH61	1.91	
		CH32CN7	2.61
G-C con Timina			,
TPLA275	-9.72	GH22TO2	1.93
		C02TH3	2.06
		CH32T04	2.42
TPLA178	-9.33	GH91TO4	2.60
		GN3TH3	2.44
		GH22TO2	1.87
TPLA48	-8.91	T02CH42	1.89
		T02CN4	2.88
TPLA73	-8.25	GO6TH51	2.84
		CH62TO4	1.89
		CH5TO4	3.13
TPLA187	-7.31	GH8TO4	2.38
		GN7TH3	1.89
TPLA94	-7.09	GN3TH12	2.8
		GH22TO2	1.89
TPLA235 -6.71 GN7TH12	2.64		
		GO6TH6	2.68

#### 3.2.2. Configuraciones fuera del plano

Al igual que en pares A-T en configuraciones con el par G-C (WC) se tiene que, en mínimos apilados no se formaron enlaces de hidrógeno del par con la tercera base nitrogenada. Este tipo de configuraciones no forma puentes de hidrógenos entre planos debido a que la distancia entre los mismos es de aproximadamente 3.4 Å y, como se menciona en la referencia, se puede formar

un enlace de hidrógeno cuando la distancia entre átomos electronegativos es menor que 3.2 Å, además de que el enlace de hidrogeno es direccional. En la tabla 4,3 se tienen los mínimos fuera del plano, apilados y perpendiculares en orden descendiente de energía en valor absoluto.

Tabla 3,7 Configuraciones de complejos en sus mínimos más profundos de energía de interacción de los pares de bases G-C con las bases purinas Guanina y Adenina,

en posiciones apiladas y perpendiculares							
	G-C con Guanina			G-C con Adenina			
Mínimo	E (kcal/mol)	Tipo	Mínimo	E (kcal/mol)	Tipo		
GPER5	-18.22	P	AAPI1259	-13.98	A		
GAPI7	-18.20	A	AAPI1301	-13.70	A		
GAPI9	-18.08	A	APER2298	-13.57	P		
GAPI2	-16.08	A	APER2373	-13.41	Р		
GAPI3	-16.02	A	AAPI1099	-13.41	A		
GPER12	-15.73	P	APER1924	-13.37	P		
GPER13	-15.15	P	AAPI188	-13.29	A		
GPER15	-14.91	P	APER1322	-13.05	Р		
GAPI4	-14.67	A	APER357	-12.38	Р		
GPER11	-14.33	P	AAPI1717	-12.14	A		
GPER2	-14.26	P	AAPI1558	-11.76	A		
GAPI2355	-14.23	A	APER1375	-10.03	Р		
GPER4	-13.63	P	APER1995	-8.98	Р		
GAPI1	-12.98	A	APER2004	-8.77	P		
GPER3	-12.65	P					
GAPI8	-12.39	A					

En la Figura 3,5 se muestran solo los mínimos globales de las estructuras perpendiculares para todas las bases.

Los valores de la energía de los mínimos globales fuera del plano para los mínimos apilados y los perpendiculares son muy cercanos entre sí. Las diferencias entre mínimos planos y apilados son de 0.02, -0.41,0.61 para G, A y C, mientras que para T es de 2.72 kcla/mol, el valor más alto.

En las configuraciones apiladas la distancia entre planos de bases es en promedio 3,4 Å.

Aunque las configuraciones perpendiculares no se encuentren en la molécula de ADN, sí se tiene referencia de que han sido encontradas in vivo, lo cual nos hace suponer que se trata de estructuras intermediarias entre las apiladas a las planas y viceversa.

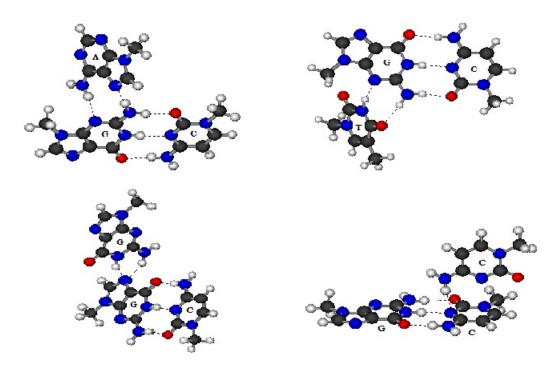


Figura 3.5 Configuraciones perpendiculares de mínima energía de los pare G:C de Watson y Crick y las bases de los ácidos nucleicos. Las estructuras corresponden con las Tabla 3.3 y 3.4.

En las Figuras 3,6 y 3,7 se muestran las estructuras de 2 triplex encontrados en el ARN de transferencia (Prolyl-tRNA Synthetase) de la bacteria Thermus thermophilus GGC imino-N7, amino-carbonil; Watson-Crick [11, 12]. Estos mínimos son comparados con los obtenidos por nosotros dando los mínimos con energías de -18,40 y 14,13 kcal/mol, una comparación detallada no es posible sin embargo en base a los enlaces de hidrógeno formados entre el par fijo G-C de WC y la tercera base nitrogenada podemos afirmar que las posibles configuraciones son planas. La formación se da en el surco mayor.

Tabla 3,8 Configuraciones de complejos en sus mínimos más profundos de energía de interacción de los pares de bases G-C con las bases pirimidinas: Citosina y Timina, en posiciones apiladas y perpendiculares

	G-C con Citosina			G-C con Timina	
Mínimo	E (kcal/mol)	Tipo	Mínimo	E (kcal/mol)	Tipo
CPER40	-14.64	Р	TAPI189	-16.52	A
CAPI56	-14.03	A	TPER752	-13.80	Р
CPER279	-13.79	P	TPER1374	-13.54	Р
CAPI28	-13.76	A	TPER896	-12.44	Р
CAPAI9	-13.71	A	TPER648	-12.41	Р
CPER186	-13.40	P	TPER573	-12.17	Р
CPER102	-13.37	P	TPER1114	-11.88	Р
CAPI20	-13.12	A	TPER1457	-11.24	Р
CAPI254	-13.05	A	TPER1088	-10.40	Р
CPER1	-12.90	P	TPER38	-10.01	Р
CPER263	-11.94	P	TPER1398	-9.09	Р
CPER134	-11.89	Р	TPER1086	-8.85	Р
CPER340	-11.63	P	TAPI387	-15.59	A
CAPI403	-10.75	A	TAPI302	-15.23	A
CAPI231	-10.22	A	TAPI190	-13.83	A
CAPI359	-9.98	A	TAPI980	-12.83	A
CAPI179	-9.85	A	TAPI527	-12.41	A
			TAPI575	-12.15	A
			TAPI508	-11.95	A
			TAPI1019	-11.17	A
			TAPI465	-10.96	A

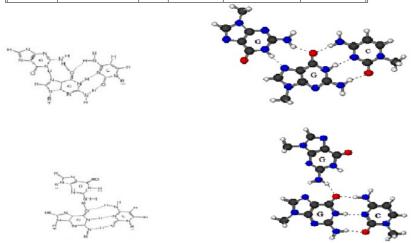


Figura 3.6. Comparación de Triplex de bases GGC reportadas en el banco de datos del Lab. de P.I. George E. Fox en la Universidad de Houston, Texas

### Conclusión

Se encontraron tres tipos de configuraciones entre las interacciones de las bases nitrogenadas: apiladas, perpendiculares y planas.

De estas configuraciones, las más energéticas (en valor absoluto) fueron aquellas en donde se formó más de un enlace de hidrógeno en un mismo átomo, es decir, en aquellas configuraciones con dos enlaces de hidrógeno y con enlaces bifurcados.

De los tres tipos de configuraciones se obtuvieron datos característicos de los mismos.

En el caso de las configuraciones planas, las más energéticas fueron aquellas estructuras en las cuales la tercera base nitrogenada se unió al par fijo G-C de WC por el lado del surco mayor. Y de manera contraria, las configuraciones planas menos energéticas fueron aquellas en donde la tercera base nitrogenada se unió del lado del surco menor. Para el caso de las configuraciones apiladas, no se forman enlaces de hidrógeno, las más energéticas resultaron ser, aquellas en las cuales la tercera base nitrogenada se situó más próxima a la citosina del par fijo G-C de WC.

En las configuraciones apiladas la distancia entre planos de bases es en promedio 3,4 Å.

Tanto en los pares A-T en configuraciones con el par G-C (WC) se tiene que, en mínimos apilados no se formaron enlaces de hidrógeno del par con la tercera base nitrogenada. Este tipo de configuraciones no forma puentes de hidrógeno entre planos debido a que la distancia entre los mismos es de aproximadamente  $3,4\ \text{Å}$ .

## Bibliografía

- 1. Hanvey J. C., Shimizu, M., Wells, R.D., Intramolecular DNA triplexes in supercoiled plasmids. II. Effect of base composition and noncentral interruptions on formation and stability. Biol Chem. 1989 Apr 5;264(10):5950-6.
- 2. Saenger W. Principles of nucleic acid structure. Ed. Springer-Verlag. New York. 1984.
- 3. Mathews C. van Holde K. Ahern K. Bioquímica. Ed. Addison Wesley. Tercera edición. España. 20034.
- 4. Van Holde K. Curtis W. Shing P. Principles of Physical Biochemistry. Ed. Prentice Hall. Primera edición. 1998.
- 5. Hernández L. Biología Molecular Integral. Ed. Noriega Limusa. Primera edición. México. 1990.
- $6.\ {\rm Ruiz}\ {\rm M}.$  Bioquímica estructural. Alfa<br/>omega. Primera edición. México. 1999.
- 7. Gosh A. Bansal M. A glossary of DNA structures from A to Z. Acta Crystallographica. 620-625. 2003.
- 8. Kamenetskii F. Biophysics of the DNA molecule. Physics reports.  $13-34.\ 1997.$
- 9. Van Dogen M. Doreleijers J. van der Marel G. van Boom J. Hilbers C. Wijmenga S. Structure and mechanism of formation of the H-y5 isomer of an intramolecular DNA triple helix. Nature America, Inc. 854-859. 1999.
- 10. Kamenetskii F. Mirkin S. Triplex DNA structures. Annu. Rev. Biochem. 65-95.1995.
- 11. Robles J. Grandas A. Pedroso E. Luque F. Eritja R. Orozco M. Nucleic Acid Triple Helices: Stability Effects of Nucleobase Modifications. Current Organic Chemistry. 1333-1368.2002.
- 12. Parel S. Marfurt J. Leumann C. DNA Triple-Helix Formation at

- Pyrimidine-Purine Inversion Sites. Marcel Dekker, Inc. 411 417,2001.
- 13. Cuevas G. Cortés F. Introducción a la química computacional. Fondo de Cultura Económica. Col. La Ciencia en el Nuevo Milenio. Primera edición. México. 2003.
- 14. Calladine C. Drew H. Understanding DNA. The molecule and how works. Academic Press. Segunda edición. New York. 1997.
- 15. Domenicano A. Hargittai I. Accurate Molecular Structures. Their Determination and Importance. International Union of Cristallography Oxford University Press. 1992.
- 16. Levine I. Química Cuántica. Ed. Pearson Educación. Quinta edición. Madrid.2001.
- 17. Sinden, R. DNA Structure and Function. Ed. Academic Press. 1994.
- 18. Poltev V.I. Grokhlina T.I. Malenkov G.C. Hydration of nucleic acid bases studied using novel atom-atom poential functions. J. Bomol. Struct. Dyn. Vol.2,413-429,1984.
- 19. Allinger N.L. Accurate Molecular Structures. Ed. Oxford Academic Press. 1992.
- 20. Potaman V. N. Oussatcheva Y.L. Lyubchenko L. S. Bidichandani. S. I. Ashizawa T. Sinden R. R. Length-dependent structure formation in Friedreich ataxia (GAA)n-(TTC)n repeats at nautral pH. Nucleic Acids Research. Vol. 32. No. 3,1224 1231,2004.
- 21. Deriabina A. González E. Groklina T. Poltev V. Biophysics. Vol. 47,972,2002.
- 22. Ban N. Nissen P. Hansen J. Moore P. Steitz T. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science. Vol. 289,905-920,11 August. 2000.
- 23. Silvian, L. Wang, J. Steitz T. Insghts into editing from an Ile-tRNA Synthetase Structure with tRNAIle and Mupirocin. Science. 13 August 1999.
- 24. Delagoutte, B. Moras, D. Cavarelli, J. tRNA aminoacylacion by arginyl-tRNA synthetase: iduced conformations during substrates binding. The EMBO journal. Vol. 19. No. 21,2000.
- 25. Sze'P, S. Wang, J. Moore, P. The crystal structure of a 26-nucleotide RNA containing a hook-turn. RNA. 9:44-51,2003.
- 26. Mao, H. Williamson, R. Assignment of the L30-mRNA complex using selective isotopic labeling and RNA mutants. Nucleics Acid Research. Vol. 27. No. 20. pp. 4059 4070,1999.
- 27. Krasilnikov, A. Yang, X. Pan, T. Mongragón A. Structure of the Spe-

BIBLIOGRAFÍA 41

cificity Domain of Bacterial RNase-P. Science Highlight. April. 2003.

- 28. González, E. Castro, I. López, E. Filippov, S. Teplukhin, A. Poltev, V. J. Mol. Struct. (Theochem), Vol. 493,301 308,1999.
- 29. Poltev, V. Sahulyupina, N. J. Biomol. Dtruc. Dyn. Vol. 3,739,1986.
- 30. Leontis, N. Stombaugh, J. Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Research. Vol. 30. No. 16,3497 3531,2002.
- 31. Yaremchuk A, Tukalo M, Grotli M, Cusack S A succession of substrate induced conformational changes ensures the amino acid specificity of Thermus thermophilus prolyl-tRNA synthetase: comparison with histidyl-tRNA synthetase J. Mol Biol vol.309, 4,989-1002,2001
- 32. Jiang F, Gorin A, Hu W, Majumdar A, Baskerville S, Xu W, Ellington A, Patel DJ Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing functional base triples Structure Fold Des 7(12):1461-72,1999

# Índice alfabético

```
superenrollamiento 14
Α
ácido desoxirribonucleico 1.
                                         surcos 7.
                                         Т
adenina 4
adn 1.
                                         timina 3.
                                         triplex 12.
arn 2
                                         W
В
benceno 3
                                         wobble 9.
\mathbf{C}
cáncer 9.
citosina 3.
complementaridad 6
coordenada 24
conformación 18.
\mathbf{E}
enlaces de hidrogeno 22.
estructuras H I
G
gen 8.
guanina 4.
Μ
mecánica molecular 17.
mutación 9
Ρ
polimero 1.
potencial de Poltev-Malenkov 20.
purinas 3.
\mathbf{R}
región promotora 8.
ribosomas 8.
S
```