



# Universidad Nacional Autónoma de México

Maestría en Ciencias (Física)

Simulando la acción enzimática de la helicasa en ds-DNA:  
un estudio de primeros principios

## T E S I S

que para optar por el grado de:

Maestro en Ciencias

PRESENTA:

Angel Ivan Rodríguez León

Tutor Principal:

Dr. Rubén Santamaría Ortíz, Instituto de Física

Miembros del comité tutor:

Dr. Jaques Andre Claude Soullard Saintrais, Instituto de Física

Dra. Angélica Estrella Ramos Peña, Instituto de Investigaciones en Materiales

Ciudad Universitaria, CD. MX., Junio 2023



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Esta pagina se dejo en blanco intencionalmente

*Para:*

*Mima, Cube, Yuri y Poncho*





# Agradecimientos

Principalmente, me gustaría agradecer al Dr. Rubén Santamaría Ortiz por su apoyo, tiempo y paciencia que dedicó a ayudarme con la culminación de este trabajo, así como sus consejos y conocimientos para poder afrontar este gran reto en mi vida académica.

Gracias a los miembros del comité tutor. Al Dr. Jaques Soullard Saintrains y la Dra. Angélica Estrella Ramos Peña por sus consejos y comentarios. Al Instituto de Física y a la Universidad Nacional Autónoma de México de la cual me llena de orgullo pertenecer.

Me gustaría agradecer también a Cristian Ordoñez, por su amistad y su paciencia, para ayudarme en todas y cada una de mis dudas, inclusive a larga distancia.

En particular quiero agradecer a mi familia, Kuka y Kike mis segundos padres. Le agradezco a mis hermanos, Chucho y Suricata, que siempre me motivan a ser el mejor ejemplo para ellos y que indirectamente guían mi vida para lograr ser mi mejor versión. A Eva, por ser mi mejor amiga y mi principal compañera de antojos y comida.

Una dedicatoria especial a Adrián, me mostraste lo que es sonreírle a la vida, y te recordaré con mucho cariño cada que toque una guitarra, o vea un acertijo, y no tengo duda que estarías muy feliz compartiendo este momento conmigo.

A Fryda, por ayudarme a hacer las imágenes y el diseño de mis trabajos como muchas otras veces. Por compartir todos mis planes de vida y lograrlos juntos. Por todos esos días en los que nos cuesta trabajo seguir adelante y nos alentamos el uno al otro. Te amo hoy y siempre. *Ich liebe dich Honey.*

A mis abuelos que fueron mis primeros y mis mejores maestros y que sin darme cuenta me dieron ese cariño hacia la Ciencia.

Finalmente, a mis padres, Poncho y Yuri. Me han enseñado a trabajar y luchar por mis sueños, sueños que poco a poco estoy logrando alcanzar y que no lo habría logrado sin su apoyo incondicional. Estaré eternamente agradecido con ustedes. Quiero que sepan que significa el mundo para mí, todos esos desayunos ruidosos en familia, todas esas tardes de café donde sus consejos y opiniones me ayudaron a tomar las decisiones correctas, esos momentos que individualmente parecen sin importancia, pero que en conjunto son parte importante de mí. Este es un logro más que también es de ustedes y espero los pueda hacer sentir orgullosos de mí, así como yo me siento orgulloso de que sean mis padres. Los amo.

# Simulando la acción enzimática de la helicasa en ds-DNA: un estudio de primeros principios

Angel Ivan Rodríguez León

tesis presentada para obtener el grado de:  
Maestro en Ciencias

## Resumen

En este trabajo de tesis se presenta un estudio de dinámica molecular a primeros principios sobre estructuras de ADN pequeñas con diferentes secuencias de bases nucleicas, con el objetivo de analizar las propiedades mecánicas del ADN bajo fuerzas externas.

Las moléculas se encuentran inmersas en un solvente acuoso implícito y bajo la influencia de fuerzas externas de tipo resorte. El tratamiento electrónico se hace de manera cuántica con teoría de funcionales de la densidad (DFT), a un nivel B3LYP/6-31g\*. Además, se realizó un tratamiento clásico sobre los núcleos y el reservorio térmico es del tipo Langevin, el cual emplea términos estocásticos.

La dinámica molecular se realizó con un código computacional desarrollado por el lenguaje Fortran, el cual permite la construcción de una subrutina de disociación de cadenas de ADN. Esto se logró al introducir resortes, detalladamente colocados sobre los distintos nucleótidos y que se mueven secuencialmente entre pares de base, simulando la acción enzimática que realizan las helicadas en los procesos celulares.

Se presentan cálculos de termalización de las distintas secuencias nucleicas a una temperatura de equilibrio específica. De igual manera, se obtuvieron valores para fuerzas de disociación de par de bases, hasta lograr la separación completa de las hebras. Además, se reportan valores de transferencia de carga entre pares de base y entre hebras, variaciones energéticas y fluctuaciones térmicas durante el proceso de disociación. Se realizaron análisis detallados sobre dependencias de secuencias en el proceso de separación y se calcularon las diferencias de fuerzas de disociación. Los resultados muestran la importancia de considerar tratamientos cuánticos que permiten el estudio de rompimientos de puentes de hidrógeno, así como tomar en cuenta la influencia de apilamiento entre bases para el cálculo de fuerzas de disociación de ADN. Este tipo de trabajos ayudan a expandir el conocimiento sobre el comportamiento del ADN durante procesos de replicación celular, a una escala cuántica.

# Simulating the helicase enzymatic action on ds-DNA: a first principle molecular dynamic simulation

Angel Ivan Rodríguez León

a thesis submitted in partial fulfillment of the requirements for the degree of  
Master of Science in Physics

## Abstract

In this thesis, we performed a first principle molecular dynamics simulations on short *ds-DNA* molecules with different nucleic sequences, with the aim of analyzing the mechanical properties of *DNA* under external forces.

The structures are immersed in an aqueous solvent and under the influence of harmonic external forces. We employed a quantum mechanical treatment for the electronic motion with density functional theory (*DFT*) on level B3LYP/6-31g\*. A thermal reservoir of the Langevin type was also used, which has shown good results in previous research.

The molecular dynamics was implemented on a computational code in the language Fortran, which allows the construction of a subroutine for helix dissociation. This was achieved with the use of springs attached to specific sites on the nucleotides. The springs move sequentially along the *DNA* strands, simulating the enzymatic behavior of helicases in cellular processes.

Thermal equilibration for each sequence and base pair force dissociation values until strand separation is presented. In addition, we report charge transfers, energy variations, and thermal fluctuations during the process. Furthermore, a detailed analysis of sequence dependence is carried out to elucidate force variations.

These results expose the importance of considering a quantum mechanical treatment on *DNA* simulations, which are useful to study hydrogen bond breakage and charge distributions. Likewise, the inclusion of several nucleic bases showed the influence of base stacking on the dissociation forces. This type of research provides deeper knowledge of *DNA* behavior during replication processes on a first principle theory level.

# Índice general

<b>Agradecimientos</b>	<b>IV</b>
<b>Resumen</b>	<b>V</b>
<b>Abstract</b>	<b>VI</b>
<b>Índice</b>	<b>VI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes y alcance de la tesis . . . . .	1
1.2. El gen y las proteínas . . . . .	5
1.2.1. Dogma Central de la Biología Molecular . . . . .	7
1.2.2. Enfermedades genéticas y el origen de la vida . . . . .	8
1.3. Ácidos nucleicos (ADN,ARN) . . . . .	10
1.3.1. Estructura biológica . . . . .	10
1.4. ADN . . . . .	11
1.4.1. Bases nucleicas . . . . .	13
1.4.2. Nucleósidos . . . . .	14
1.4.3. Nucleótidos . . . . .	14
1.4.4. Puentes de Hidrógeno . . . . .	14
1.4.5. Apilamiento . . . . .	15
1.5. Enzimas (Helicasas) . . . . .	16
1.6. Métodos experimentales sobre ADN . . . . .	18
1.7. Simulación Molecular . . . . .	20
<b>2. Marco Teórico</b>	<b>22</b>
2.1. Ecuación de Schrödinger . . . . .	23
2.2. Problema molecular . . . . .	23
2.2.1. Aproximación Born-Oppenheimer: movimiento electrónico y nuclear . . . . .	24
2.3. Métodos de solución para la estructura electrónica molecular . . . . .	27
2.3.1. Conjunto de funciones base . . . . .	28
2.3.2. Correlación electrónica y métodos post-HF . . . . .	29
2.3.3. Teoría del Funcional de la Densidad (DFT) . . . . .	30

2.3.4.	Primer teorema de Hohenberg y Kohn . . . . .	31
2.3.5.	Segundo teorema Hohenberg-Kohn . . . . .	32
2.3.6.	Esquema Kohn y Sham . . . . .	33
2.3.7.	Funcionales de correlación e intercambio . . . . .	36
2.4.	Superficie de energía potencial . . . . .	38
2.5.	Análisis poblacional (Mulliken) . . . . .	40
2.6.	Método de solución movimiento nuclear . . . . .	42
2.6.1.	Límite clásico . . . . .	42
2.6.2.	Integrador de Verlet . . . . .	44
2.7.	Reservorio térmico de Langevin . . . . .	46
2.7.1.	Ecuación Langevin-Zwanzig . . . . .	47
2.7.2.	Integrador de Verlet-Langevin . . . . .	51
2.8.	Fuerzas externas sobre ADN (steering forces) . . . . .	55
2.9.	Dinámica Molecular . . . . .	58
<b>3.</b>	<b>Resultados</b>	<b>61</b>
3.1.	Estructuras de ADN y detalles computacionales . . . . .	62
3.1.1.	Estructura (T-T-G-C-G) . . . . .	63
3.1.2.	Estructura (T-C-G-C-G) . . . . .	64
3.2.	Equilibración térmica y energética de las estructuras de ADN . . . . .	66
3.3.	Separación de ds-DNA . . . . .	69
3.3.1.	Fuerzas disociativas . . . . .	78
3.3.2.	Distribución de cargas eléctricas . . . . .	83
3.3.3.	Dependencia de la secuencia de bases nucleicas . . . . .	92
<b>4.</b>	<b>Conclusiones</b>	<b>94</b>
<b>A.</b>	<b>Hartree-Fock</b>	<b>97</b>
A.1.	Método de Hartree-Fock . . . . .	97
A.2.	Método de Roothaan . . . . .	99
<b>B.</b>	<b>Derivadas de la energía y matriz Hessiana</b>	<b>101</b>
<b>C.</b>	<b>Publicaciones</b>	<b>103</b>

# Índice de figuras

1.1.	Imagen de células bajo un microscopio (a). Representación gráfica de la célula, cromosoma y ADN (b).	5
1.2.	Dogma Central de la Biología. Usada con permiso del "National Human Genome Research Institute" [34].	7
1.3.	Proceso de replicación según la hipótesis semi-conservativa, se sintetizan dos hebras hijas de una hebra madre [35].	8
1.4.	Fotografía 51 [40].	11
1.5.	Replica del modelo original del ADN por Watson-Crick, tomado de la referencia [41].	12
1.7.	Estructura de ADN (grupo fosfato, ribosa y bases). Figura tomada de "National Human Genome Research Institute" [34].	14
1.8.	Puentes de hidrógeno entre pares de bases CG y TA [43].	15
1.9.	Fuerzas de apilamiento entre pares de bases (perpendiculares a los puentes de hidrógeno).	16
1.10.	Replicación del ADN con ayuda de la helicasa [45]. Donde: ds-DNA (ADN de cadena doble), ss-DNA (ADN de cadena sencilla) y SSBP (proteínas ligantes de ADN de cadena sencilla que se encargan de la estabilización de la apertura del ADN).	17
1.11.	Pinzas ópticas para estudio de interacción ADN-proteína [50].	19
1.12.	Arreglo común AFM [54].	19
2.1.	Diagrama de flujo para el cálculo de la energía electrónica bajo la metodología de DFT-Kohn/Sham.	35
2.2.	Niveles de teoría utilizados en este trabajo.	54
2.3.	Posicionamiento de los resortes virtuales sobre cada una de las bases nucleicas.	55
2.4.	Diagrama de flujo para la implementación de la dinámica molecular cuántica, con uso de la teoría de Langevin.	60
3.1.	Primera estructura de ADN cuenta con 315 átomos y 8 grupos fosfatos en total.	63
3.2.	Segunda estructura de ADN cuenta con 314 átomos y 8 grupos fosfatos en total.	65
3.3.		67
3.4.	Equilibración energética del sistema.	67
3.5.		68
3.6.	Equilibración energética del sistema.	68

3.7. Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-T-G-C-G). . . . .	70
3.8. Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-T-G-C-G) junto con intervalos temporales de apertura de cada par de bases. Se muestran imágenes de la estructura con 2 (izquierda) y 5 (derecha) pares de base disociados. . . . .	71
3.9. Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-T-G-C-G, durante toda la dinámica (3 etapas). La parte sombreada se refiere a la sección durante la cual actúan los resortes. . . . .	73
3.10. Fluctuaciones de temperatura (a) y de energía (b), para la secuencia T-T-G-C-G, durante el proceso de disociación (13 ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base. . . . .	73
3.11. Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-C-G-C-G). . . . .	74
3.12. Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-C-G-C-G) junto con intervalos temporales de apertura de cada par de bases. Se muestran imágenes de la estructura con 2 (izquierda) y 5 (derecha) pares de base disociados. . . . .	75
3.13. Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-C-G-C-G, durante toda la dinámica (3 etapas). La parte sombreada se refiere a la sección durante la cual actúan los resortes. . . . .	77
3.14. Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-C-G-C-G, durante el proceso de disociación (23 ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base. . . . .	77
3.15. Distancia de separación entre el centro de masa para cada par de base como función de las fuerzas de disociación. . . . .	78
3.16. Fuerzas de disociación como función de la distancia de separación entre el centro de masa para cada par de bases. . . . .	79
3.17. Distancia de separación entre el centro de masa para cada par de bases en función de las fuerzas de disociación. . . . .	80
3.18. Fuerzas de disociación de función de la distancia de separación entre el centro de masa para cada par de bases. . . . .	81
3.19. Distribuciones de carga para las hebras de la estructura de ADN durante el proceso de disociación. La parte sombreada se refiere al momento en el que los resortes se apagan. . . . .	83
3.20. Cargas eléctricas de átomos involucrados en puentes de hidrógeno en el par de bases central (GC), para antes (lado izquierdo) y después (lado derecho) de la separación. . . . .	84
3.21. Variaciones de carga para bases apiladas (segundo y tercer par de base) para el primer ADN (secuencia: T-T-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente. . . . .	85

3.22. Variaciones de carga para bases apiladas (tercer y cuarto par de base) para el primer ADN (secuencia: T-T-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente. . . . .	85
3.23. Distribuciones de carga para bases nucleicas del par de base central (GC). . . .	86
3.24. Distribuciones de carga para las hebras de la estructura de ADN durante el proceso de disociación. La parte sombreada se refiere al momento en el que los resortes se apagan. . . . .	87
3.25. Cargas eléctricas de átomos involucrados en puentes de hidrógeno en el par de bases central (GC), para antes (lado izquierdo) y después (lado derecho) de la separación. . . . .	88
3.26. Variaciones de carga para bases apiladas (segundo y tercer par de base) para el segundo ADN (secuencia: T-C-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente. . . . .	89
3.27. Variaciones de carga para bases apiladas (tercer y cuarto par de base) para el segundo ADN (secuencia: T-C-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente. . . . .	90
3.28. Distribuciones de carga para bases nucleicas del par de base central (GC). . . .	90
C.1. Página principal del artículo próximo a enviar para su publicación. . . . .	103



# Capítulo 1

## Introducción

*"The advantage of the emotions  
is that they lead us astray, the  
advantage of science is that is  
not emotional"*

---

*Oscar Wilde*

### 1.1. Antecedentes y alcance de la tesis

El origen de la vida es una de las incógnitas más antiguas para el ser humano. El planeta Tierra forma parte de la inmensidad de un universo que hasta la fecha no conocemos por completo, con miles de millones de estrellas y planetas con condiciones distintas. Esto lleva a preguntarnos, ¿Hay vida en otros planetas?, y de no ser así, ¿Qué condiciones especiales tiene la Tierra para haber producido vida por primera vez?

Se tienen registros de vida en este planeta que datan de hace más de 3 millones de años. La abiogénesis (surgimiento de vida, a partir de materia inerte) sigue siendo un tema discutido por distintas áreas científicas y existen aún varias teorías vigentes, que no han logrado explicar el proceso de manera completa [1].

A pesar de esto, algo que sí es identificable de cualquier sistema “vivo”, es la característica de poder crear más organismos similares a sí mismo. Hasta el momento, sabemos que todos los organismos del planeta Tierra hacen uso de los ácidos nucleicos para la replicación celular [2]. Por lo que podemos suponer que de encontrar vida en otros rincones del universo, seguramente su funcionamiento recaerá en este tipo de moléculas. Citando a *Francis Crick* inmediatamente después del descubrimiento de la estructura del ADN, “we have discovered the secret of life” [3]. Si tomamos al ADN y el ARN como las moléculas primordiales que desencadenaron en el origen de la vida, y además consideramos que esto ocurrió por medio de un proceso de autorreplicación, estudiar este fenómeno podría esclarecer varias incógnitas [4].

Como veremos más adelante, la maquinaria celular para la creación de nuevos organismos, depende de ciertas enzimas que separan las cadenas del ADN para poder extraer la información genética resguardada dentro de su doble hélice. La manera en que la helicasa logra disociar los puentes de hidrógeno entre bases nucleicas en procesos de replicación, es similar a como la

colagenasa (otro tipo de enzima) corta los enlaces peptídicos entre aminoácidos en la molécula de colágeno. Debido a esto, entender de manera general cómo es que ciertas moléculas controlan átomos y grupos funcionales específicos para romper enlaces químicos, puede ayudar a elucidar muchos procesos biológicos generales.

El interés por comprender el funcionamiento del copiado y movimiento de la información genética, no solo es para elucidar el origen de la vida<sup>1</sup>, sino que, nos ayuda a caracterizar las bases de las enfermedades genéticas. Se conocen actualmente más de 5000 desórdenes genéticos que afectan a millones de personas en todo el mundo; enfermedades como la fibrosis quística, Parkinson, síndrome de Down, entre otras [5]. La causa de estas enfermedades se asocia a errores en el proceso de replicación, así como mutaciones puntuales y mutaciones en proteínas que ayudan a la síntesis de las helicasas (ver sección enfermedades genéticas 1.2.2).

Una posible estrategia para comprender la dinámica ADN-helicasa, es estudiar la respuesta del ADN bajo fuerzas externas. Recordemos que en la replicación, las enzimas ejercen estas fuerzas para mantener abierta la doble hélice [6, 7].

Se ha progresado bastante en el desarrollo de técnicas experimentales para medir el desenrollamiento del ADN bajo la acción de la helicasa y las fuerzas que genera sobre las estructuras [5]. Fabian Kilchherr et al. lograron medir fuerzas de apilamiento para distintas secuencias al nivel de moléculas individuales. El arreglo experimental hace uso de pinzas ópticas en combinación con ADN origami para permitir cierto posicionamiento y aislar dichas fuerzas [8]. Por otro lado, Lu Bai et al. se enfocaron directamente en una helicasa particular, llamada helicasa T7; midieron el tiempo de desamarre de ADN en función del número de pares de base nucleicas usando pinzas ópticas. Curiosamente, notaron una dependencia de este tiempo con la secuencia nucleica, rápida en zonas AT y más lenta en zonas GC [9]. A pesar del refinamiento continuo de los experimentos de nano-manipulación, las demandas económicas para llevarlos a cabo, además de las dificultades experimentales debido al poco control de las condiciones físicas, conllevan a buscar alternativas o complementos para esta área de estudio [10]. Una alternativa para abordar el problema es con ayuda de las simulaciones moleculares.

El sustancial avance en el poder computacional, desde los años 80, permitió extender el estudio de los fenómenos biomoleculares desde el punto de vista de la química y física computacional [11]. La primera simulación de dinámica molecular sobre el ADN fue realizada por Michael Levitt<sup>2</sup> en 1983. En los últimos 25 años las simulaciones a nivel atómico han logrado entrar al régimen de microsegundos, una mejora de 5 órdenes de magnitud [13].

En el caso de estudios analizando el comportamiento de fuerzas externas sobre el ADN, Amar Singh et al. aplicaron una fuerza en distintas posiciones y analizaron cómo varía con la longitud de la cadena; mostraron una característica conocida, cuando la fuerza se aplica perpendicularmente al eje helicoidal, las bases se estiran y se rompen secuencialmente. Por otro lado, cuando la fuerza se aplica a lo largo del eje helicoidal, esta se distribuye sobre muchos pares de base [14].

En la literatura se pueden encontrar otros trabajos que utilizan dinámica molecular clásica junto con fuerzas de *steering* sobre estructuras de ADN [15, 16]. Estas investigaciones dependen

---

<sup>1</sup>Existen 123 definiciones distintas para el concepto de la vida en publicaciones formales [1]

<sup>2</sup>Michael Levitt, Premio Nobel de Química 2013 [12].

de la buena elección y adaptación del campo de fuerzas específico para el sistema bajo estudio. De igual manera, fallan al no incluir efectos cuánticos y la imposibilidad de hacer análisis de rompimientos y formación de enlaces químicos [17].

Justamente con la implementación de dinámicas moleculares cuánticas, podemos estudiar fenómenos biológicos en términos de arreglos electrónicos y caracterizar propiedades químicas. En lugar de tratar la energía potencial del sistema en términos de campos de fuerzas (FF, por sus siglas en inglés), se usa el campo debido a las interacciones electrónicas. Existen distintos métodos para el cálculo de la contribución electrónica. En particular, en este trabajo se utiliza teoría del funcional de la densidad (DFT, por sus siglas en inglés).

Lo que se busca encontrar al usar esta metodología, es la estructura y análisis energético del ADN basados en su estructura electrónica [18].

Debido a que en esta tesis se utilizó una estructura de ADN con consideración de la columna vertebral de fosfatos, desoxirribosas y bases apiladas, es importante considerar las contribuciones debido a las fuerzas de apilamiento entre pares de base y hacer uso de los mejores métodos en términos de eficacia y costo computacional.

Varios estudios han encontrado que, en casos donde se presenta el fenómeno de apilamiento, métodos cuánticos como MP2 sobreestiman este tipo de energía, mientras que métodos de DFT, con correcciones de dispersión, resultan una buena opción [19].

Existen diversos datos teóricos sobre las bases nucleicas de manera individual y bajo metodologías cuánticas. Santamaría et al; obtuvieron estructuras de las bases nucleicas en sus estados de mínima energía [20]. DFT se ha utilizado también para estudiar el comportamiento de las bases en presencia de moléculas contaminantes [21].

Por otro lado, Cristian Ordoñez et al. realizaron simulaciones de primeros principios sobre sistemas A-T y G-C en vacío y en solvatación, bajo fuerzas tipo resorte. Utilizaron un termostato de Langevin y obtuvieron energías de disociación, transferencias de carga y valores para las fuerzas de estiramiento de bases [22]. El posicionamiento de los resortes fue detalladamente realizado, para evitar efectos de torsión, y es un arreglo similar al utilizado en el presente trabajo.

Santosh Mogurampelly et al. trabajaron con distintos protocolos de estiramiento sobre DNA a lo largo del eje helicoidal, así como de manera perpendicular. La fuerza armónica fue aplicada en distintos puntos específicos, pero no de manera secuencial. Utilizaron estructuras conocidas que contienen la secuencia palíndroma “GAATTC” y una razón de estiramiento de  $0.0001 pN/fs$ . Cabe señalar que, restringieron la dinámica de los puentes de hidrógeno y usaron un campo de fuerzas tipo ff10 [23].

Por otro lado, A. M. Naserian-Nik et al; estudiaron las propiedades mecánicas de cadenas cortas de ds-DNA (ADN de doble cadena) bajo estrés. Además, incluyeron efectos de solvatación por medio de dinámica de Langevin. Las fuerzas externas aplicadas fueron de tipo resorte y estudiaron la dependencia del ángulo de aplicación y de la razón con la que se jalaba la estructura. Analizaron la fuerza requerida para estirar al ADN sobre el eje helicoidal, en función del tiempo de simulación y una cierta coordenada reactiva. Además, emplearon un criterio geométrico para observar la existencia de puentes de hidrógeno. Átomos aceptores y donado-

res deben estar a menos de 3.5 Å. Similarmente, para el caso de existencia de apilamiento, el criterio es una distancia menor a 5 Å para los centros de los anillos de las bases nucleicas [24, 25].

Parte crucial de la realización de un cálculo cuántico con DFT, recae en la buena elección de los funcionales de correlación e intercambio, así como la base que se utiliza.

Este trabajo en particular utiliza el funcional *B3LYP*, el cual ha demostrado modelar de buena manera sistemas de ácidos nucleicos, tanto para los valores energéticos de puentes de hidrógeno, como para las bases individuales en general [26, 27]. Además, se hace uso del conjunto de funciones base 6-31g\*, las cuales funcionan de manera efectiva para moléculas biológicas [28, 29]

En este trabajo se presentan simulaciones de dinámica molecular cuántica, empleando teoría de funcionales de la densidad a un nivel B3LYP/6-31g\* sobre distintas estructuras de ADN bajo fuerzas armónicas. Se aplican las fuerzas externas sobre cada par de bases para abrir de manera secuencial la doble hélice y poder obtener datos estructurales, energéticos y de transferencia de carga. Además, se emplea un termostato de Langevin para simular contacto con un reservorio térmico. La razón para emplear estructuras pequeñas (no más de 5 pares de bases) es para disminuir el costo computacional y usar métodos químico-cuánticos *ab-initio*. El ADN es una molécula flexible y como se mencionó previamente, el análisis de estructuras cortas, modela bien el comportamiento de secuencias más largas. El efecto de fuerzas de estiramiento depende de la longitud de la cadena, incrementa de manera lineal hasta llegar a un punto de saturación para cadenas de más de 12 bases [30].

Los cálculos de la energía electrónica se realizaron con el software Terachem [31] que realiza las operaciones a nivel de GPUs <sup>3</sup>. La dinámica molecular y la implementación de las fuerzas externas que abren a las cadenas nucleicas, se realizaron por medio de un código de Fortran, el cual se mencionará más adelante.

Generalmente, los trabajos que simulan estructuras de ADN considerando la cadena de fosfatos y varios pares de bases, hacen uso de metodologías clásicas. Por otro lado, cuando se analizan fuerzas externas sobre estas moléculas, suelen hacerlo simulando experimentos tipo AFM, lo cual limita el posicionamiento de las fuerzas. Se acercan en replicar el experimento y no tanto el fenómeno natural. La ventaja de utilizar métodos cuánticos en esta tesis, es poder analizar la estructura electrónica de la molécula de ADN, así como caracterizar la disociación de los puentes de hidrógeno entre pares de base bajo fuerzas externas. Además, se emplean fuerzas que rompen los puentes de hidrógeno base por base, de manera secuencial, lo cual es más fiel a como lo hacen ciertas enzimas en procesos biológicos.

Este tipo de trabajos ayudan a aportar mayor conocimiento sobre los efectos que generan las fuerzas realizadas por helicasas sobre estructuras de ADN, importantes en replicación de ADN. Más aún, dan oportunidad a crear estrategias médicas para prevención y erradicación de enfermedades conocidas por mal funcionamiento de estos procesos, como inestabilidad genómica o predisposición cancerígena.

---

<sup>3</sup>Una tarjeta gráfica Tesla V100 (NVIDIA) genera el rendimiento equivalente a 100 CPU's, ver ([32]), para información de eficiencia de TeraChem con tarjetas NVIDIA.

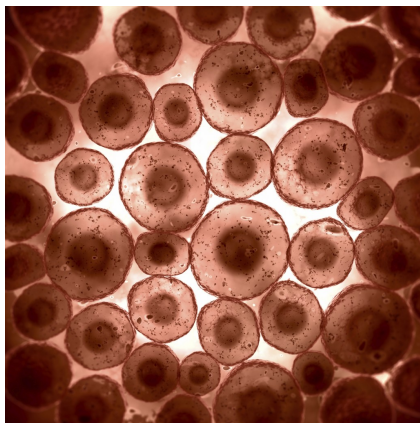
## 1.2. El gen y las proteínas

En el año 1590 Zacharias Janssen, un fabricante de lentes neerlandés, haría uno de los aportes más importantes en la historia de la ciencia con la creación del microscopio, un aparato que permitiría observar más allá de las capacidades humanas y elucidar sistemas cuyas dimensiones físicas son increíblemente pequeñas, pero cuyas dimensiones conceptuales son, sin lugar a duda, inconmensurables.

Años más tarde, el prolífico científico inglés Robert Hooke (1635-1703), utilizaría el microscopio para analizar pedazos de corcho y observar por primera vez lo que hoy en día conocemos como células. De las palabras de Hooke *“Con los microscopios no hay nada tan pequeño como para escapar de nuestro análisis; por lo tanto, hay un nuevo mundo visible listo para ser descubierto”*.

Estos descubrimientos generarían una cascada de preguntas respecto al contenido celular y su funcionamiento que tardarían varios años en poder responderse con la ayuda de distintos trabajos en diferentes áreas científicas. Principalmente, con la investigación del monje austriaco Gregor Mendel (1822-1884) en Brno, al notar que existen ciertos rasgos de color y forma en la crianza de guisantes y arvejas, que cumplen con factores específicos que pueden ser transmitidos de generación en generación. Eventualmente, estos llamados factores de Mendel terminarían siendo bautizados como “genes”.

En 1902, el genetista estadounidense Walter Sulton mostraría que ciertas estructuras encontradas en las células, llamadas cromosomas, cumplen con las leyes de Mendel y propuso que los genes se encuentran ubicados dentro de estos cromosomas.



(a) Obtenida de la referencia [33].

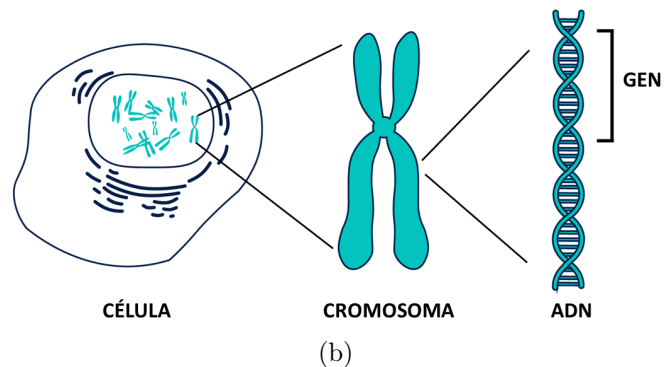


Figura 1.1: Imagen de células bajo un microscopio (a). Representación gráfica de la célula, cromosoma y ADN (b).

Finalmente, con el avance de las técnicas experimentales, así como el refinamiento de microscopios de mayor resolución, se encontró que los verdaderos responsables de las reglas hereditarias que siguen los sistemas biológicos son unos polímeros que sirven de estructura para los cromosomas, conocidos como ácidos nucleicos.

Los ácidos nucleicos son los guardianes de la información genética y sirven como una especie de instructivo que cada célula debe leer cuidadosamente para la construcción de proteínas.

Desde proteínas fibrosas responsables de la estructura de huesos, tejidos, músculos y tendones, hasta proteínas membranosas y enzimas. Todas las proteínas son creadas de esta manera por el sistema biológico.

Uno de los problemas más importantes en biología es entender cómo es que la célula logra traducir el código genético proporcionado por el ADN, a una secuencia de aminoácidos para la generación de moléculas más complejas, debido a que son lenguajes diferentes e involucra distintos fenómenos que hasta la fecha no se entienden por completo.

### 1.2.1. Dogma Central de la Biología Molecular

Hoy en día tenemos una idea general del proceso que debe seguir la célula para la lectura y divulgación de la información genética. Los ácidos nucleicos son los protagonistas de este proceso y con ayuda de ciertas moléculas que ayudan a acelerar reacciones químicas, cortar estructuras moleculares y leer información, se logra la transmisión, replicación y traducción de la información contenida en los genes.

Francis Crick (1916-2004), 5 años después del descubrimiento de la estructura del ADN (ver sección ADN 1.3), nombró al proceso del flujo de información genética; de ADN a ARN y finalmente a proteínas, como el *dogma central de la biología molecular*.

Este dogma, o mejor dicho, esta hipótesis central de la biología, dice que las células utilizan el código genético para traducir la secuencia de nucleótidos (bases constituyentes del ADN), en secuencias de aminoácidos. Si la célula necesita sintetizar alguna proteína, tiene que seguir las instrucciones dadas por el ADN y copiar el gen. La sutileza está en que estas secuencias (nucleótidos/aminoácidos) están escritas en lenguajes distintos, por lo que el proceso de traducción es de vital importancia para entender como ocurre este proceso celular.

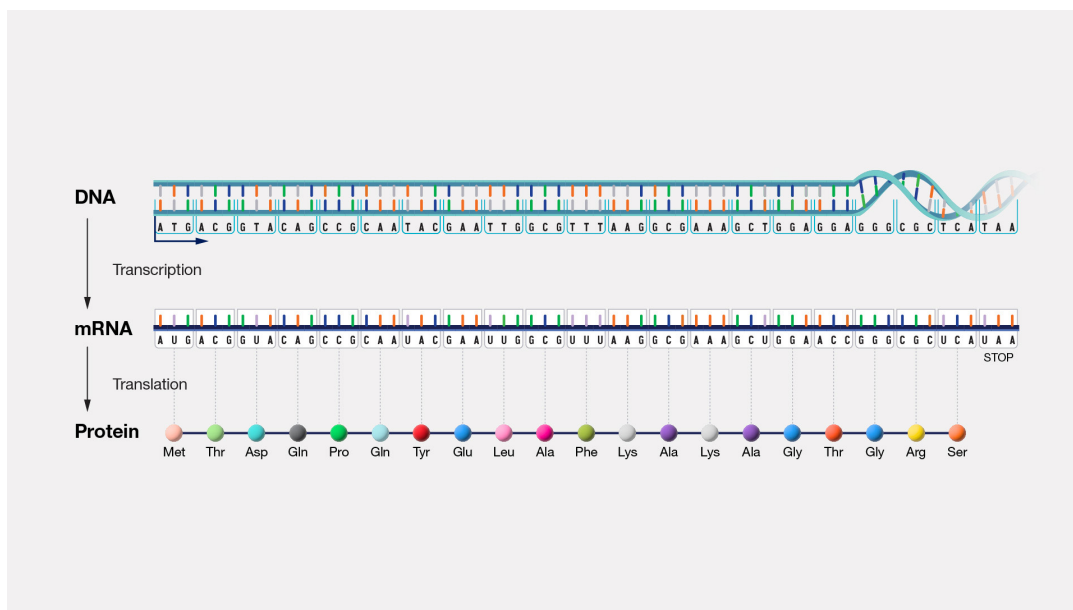


Figura 1.2: Dogma Central de la Biología.

Usada con permiso del "National Human Genome Research Institute"[34].

Es bien conocido que el ADN tiene una estructura de hélice doble y en su interior está codificada la información genética en forma de una secuencia de 4 bases nucleicas.

Debido a que la temperatura dentro de la célula es constante y la molécula de ADN es estable, para poder acceder a dicha secuencia el ADN debe desamarrar sus dos hélices (ver fig. 1.3<sup>4</sup>), lo cual ocurre únicamente con ayuda de moléculas catalizadoras, mejor conocidas como

<sup>4</sup>A esto se le conoce como la *hipótesis semi-conservativa*, en donde se estipula que el proceso de replicación involucra la separación de las hélices para posteriormente sintetizar dos nuevas hélices hijas (confirmado por el

enzimas (ver sección enzimas 1.5). Experimentalmente, podemos provocar esta separación al cambiar el pH para ionizar las bases, o al cambiar la temperatura y separar a la estructura.

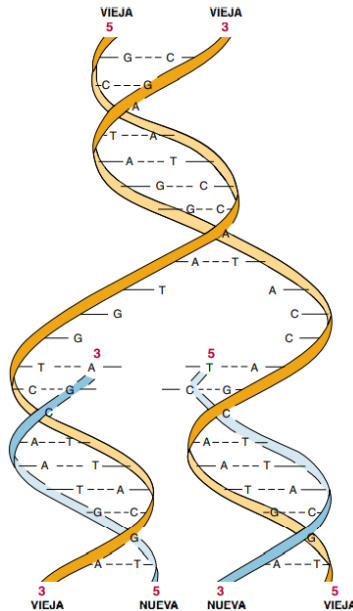


Figura 1.3: Proceso de replicación según la hipótesis semi-conservativa, se sintetizan dos hebras hijas de una hebra madre [35].

Una vez teniendo acceso a la secuencia de los nucleótidos tenemos que poder hacer la traducción necesaria para que la célula entienda como sintetizar a los aminoácidos. Esto requiere dejar de lado al ADN y concentrarnos en el ARN. A diferencia del ADN, el ARN puede dejar la célula pasando por pequeños agujeros de la membrana, para así llegar al ribosoma, los cuales son los encargados de construir a las proteínas [4].

La traducción se realiza al considerar que el ADN solo tiene 4 bases nucleicas y solo se conocen 20 tipos de aminoácidos, por lo que la única manera, es leer la secuencia nucleica de 3 en 3 bases (conjunto llamado codón) que caracteriza a cada aminoácido.

### 1.2.2. Enfermedades genéticas y el origen de la vida

El buen funcionamiento de los órganos internos de los seres humanos recae en la efectividad de nuestras células para sintetizar proteínas, esto es, efectividad para realizar el proceso descrito por el dogma central de la biología. Cualquier alteración o falla puede desencadenar en distintos tipos de enfermedades, que incluso pueden ir transmitiéndose de generación en generación.

“*En general, del virus más simple, hasta la complejidad de los humanos, somos simplemente paquetes llenos de proteínas que logran replicar su material genético*” [4]. Si tuviéramos la capacidad de no sólo entender como es que los organismos leen el código genético, sino, aprender a escribirlo nosotros mismos, tendríamos en nuestras manos la cura a miles de enfermedades experimento de Matthew Meselson and Franklin Stahl lo confirmo en 1958).



que azotan a los seres humanos. Más aún, tendríamos una herramienta para la mejora de los seres vivos en general, con todo y los problemas éticos que eso conlleva.

En particular, las enfermedades genéticas se dan por mutaciones o alteraciones del código hereditario o fallas en el proceso de replicación, transcripción o traducción de esta información. Alteraciones tales como sustitución de nucleótidos individuales, desajustes entre pares de bases, errores de lectura por parte de las polimerasas o rompimientos de enlaces químicos por algún tipo de radiación [36].

Muchos tipos de cáncer se generan por mutaciones en genes que controlan el crecimiento de células, llevando a un crecimiento sin control y generación de tumores [37]. De igual manera, desordenes de genes individuales llevan al desarrollo de enfermedades como Tay Sachs (afectaciones cerebrales y del sistema nervioso) o Huntington<sup>5</sup> [4, 38]. De igual manera, muchas enfermedades son ocasionadas por mutaciones en proteínas que ayudan a la buena síntesis de las helicasas (enzimas que desamarran al ADN en el proceso de replicación). Síndrome de Bloom<sup>6</sup> y síndrome de Werner (envejecimiento prematuro) son algunos de los más comunes. De ahí la importancia de poder describir de manera precisa la naturaleza de todos estos procesos biológicos, con la finalidad de poder erradicar todas estas terribles enfermedades.

Descubrir los mecanismos que generan mutaciones celulares no es importante únicamente para la comprensión de enfermedades genéticas, sino también, para el entendimiento del origen de la vida misma. Las células son fábricas en constante movimiento, responsables de muchas actividades realizadas a la perfección en ambientes con ajeteos impredecibles. ¿Qué mantiene encendida a la fábrica? ¿De dónde viene el motor que mantiene su funcionamiento? Es interesante pensar, si el clima específico del planeta Tierra fue lo que provocó organismos basados en ácidos nucleicos, donde los puentes de hidrógeno tienen las energías exactas para la buena estabilidad y flexibilidad de estas moléculas.

Sabemos ahora que el dogma central requiere de distintas biomoléculas, ADN, ARN y proteínas. Por lo que podríamos preguntarnos, ¿qué vino primero, ¿las proteínas o los ácidos nucleicos?, un clásico enigma tipo “la gallina y el huevo” [4]. Crick planteó una hipótesis en la que propone que, al inicio de la historia, un ARN logró replicarse por sí mismo para comenzar por primera vez con este proceso, algo así como un ARN primordial<sup>7</sup>.

---

<sup>5</sup>El gen HTT contiene un número incrementado de repeticiones de nucleótidos CAG (26 o menos), un gen HTT dañado contiene más de 26 repeticiones, esto genera una proteína que es tóxica para las células.

<sup>6</sup>Mutaciones en el gen BLM, que codifica una proteína de la familia de las helicasas del ADN. Estas mutaciones pueden producir la inactividad de la proteína o la inhibición de su síntesis.

<sup>7</sup>En el famoso experimento de Miller-Urey (1950's) se aplicó una descarga eléctrica a una mezcla de compuestos químicos que simulaban la atmósfera terrestre temprana. Mostraron que se podían obtener aminoácidos de manera sencilla (aunque no se ha logrado mostrar experimentalmente la creación de ácidos nucleicos bajo condiciones de una sopa primordial).

### 1.3. Ácidos nucleicos (ADN,ARN)

Los ácidos nucleicos se encuentran dentro de la célula y son los guardianes y distribuidores de la información genética. Son polímeros biológicos compuestos por tiras largas de monómeros llamados nucleótidos, todos amarrados entre sí, por enlaces covalentes.

Los tipos de ácidos nucleicos que nos interesan son ADN (ácido desoxirribonucleico) y ARN (ácido ribonucleico).

El ARN entra en juego en el proceso de transcripción, al copiar ciertos genes (secuencia de bases nucleicas) del ADN. Los ácidos ribonucleicos se diferencian del ADN al tener una ribosa distinta y una base nucleica diferente; debido a esto es una estructura de una sola hebra, como veremos más adelante. Su función es transcribir la información guardada en el ADN a una forma que la célula pueda leer y sintetizar proteínas [19]. Además, puede actuar como agente catalítico por sí mismo, por lo que puede ser más fundamental que el mismo ADN en temas del origen de la vida, ya que no requiere de enzimas para su replicación.

#### 1.3.1. Estructura biológica

La estructura general de los ácidos nucleicos se conforma por moléculas de ribosa amarrando bases nucleicas por medio de enlaces glicosídicos<sup>8</sup> y una columna vertebral sostenida por grupos fosfodiéster. El grupo fosfodiéster<sup>9</sup> básicamente es un átomo de fósforo rodeado de 4 oxígenos, estos átomos se conectan entre ellos por medio de enlaces covalentes (O-C). El grupo OH en el tercer carbono de la ribosa<sup>10</sup> se conecta con el OH del quinto carbono de una ribosa adyacente gracias a este fosfodiéster, para dar la forma de la columna vertebral. Los grupos introducen una carga negativa, por lo tanto, en ambientes acuosos (como lo es la célula) estas regiones hidrofílicas interactúan con las moléculas polares para estabilizar la estructura de los ácidos nucleicos. Estas cargas negativas repelen a cualquier nucleófilo por repulsión electrostática, haciéndolos menos susceptibles a hidrólisis.

Por otro lado, a la ribosa en el ADN le hace falta un grupo hidroxilo en el segundo carbono (de ahí el nombre desoxirribosa). Esta ausencia es esencial para la estabilidad y flexibilidad del ADN, así como su forma de doble hélice, lo cual lo hace más resistente a la hidrólisis.

---

<sup>8</sup>Enlace glicosídico: enlaces covalentes que se dan entre azúcares.

<sup>9</sup>Grupo fosfodiéster: tipo de enlace covalente que se produce entre un grupo hidroxilo (OH-) en el carbono 3' y un grupo fosfato formándose así un doble enlace éster.

<sup>10</sup>Ribosa: es una pentosa (monosacárido de cinco átomos de carbono).

## 1.4. ADN

Las estructuras nucleicas analizadas en esta tesis son de ADN, por lo que resulta importante mencionar con más detalle sus propiedades básicas y la historia del descubrimiento de su forma helicoidal.

El cuerpo humano contiene dentro de cada una de sus células alrededor de 2 metros de ADN; tomando en cuenta todas las células, hay una cantidad suficiente de ADN como para ir de la Tierra a Plutón. Esta cantidad debe estar contenida a su vez en un núcleo de apenas 10 micrones (milésima parte de un milímetro) de diámetro [39].

Estas características hacen al ADN una molécula biológica con características especiales y un nivel de adaptación perfecto, que nos permitiría suponer que si existe vida fuera de este planeta, seguramente está conformada por ADN. Algunas veces se encuentra alrededor de proteínas, o compactado dentro de la célula, puede ser dañado, reparado, e incluso puede ser cortado y reconstruido.

En 1952, la química y cristalógrafa Rosalind Franklin (1920-1958), mostraría por primera vez el patrón de difracción de una molécula de ADN. La famosa “fotografía 51” (fig. 1.4), hace notar una especie de cruz reflectiva, la cual solo puede ser proyectada por una estructura helicoidal.

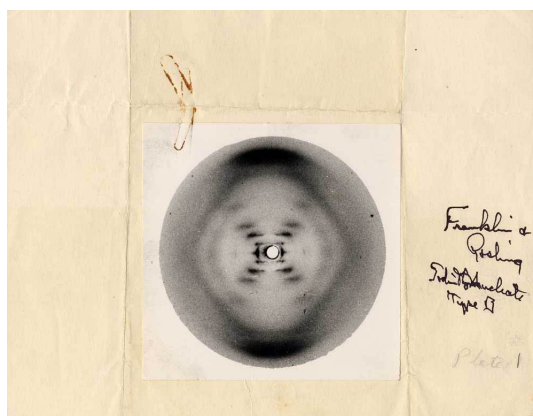


Figura 1.4: Fotografía 51 [40].

Posteriormente, el biólogo molecular James Watson y el físico británico Francis Crick deducirían, con base en los trabajos de Rosalind Franklin (sin permiso de ella<sup>11</sup>), la estructura del ADN. Su famoso artículo se publicaría en 1953 en la revista “*Nature*” [41], les valdría el Nobel de Medicina en 1962<sup>12</sup>.

<sup>11</sup>Por desgracia, Rosalind Franklin fallece 4 años antes que se otorgase el premio Nobel. Sufrió todo tipo de abuso por parte de sus colaboradores, e incluso comentarios sexistas de sus colegas, en especial de Watson y Crick.

<sup>12</sup>*The Double Helix*, libro escrito por Crick donde hace un recuento detallado del proceso del descubrimiento de la estructura del ADN, se convirtió en *bestseller*, y considerado uno de los 100 libros más importantes de la ciencia [42].

En el trabajo mencionado de la estructura del ADN, ellos proponen ciertas características estructurales:

- Conformada por dos clases de bases nucleicas, purinas (A,G) y pirimidinas (T,C)
- Las bases se encuentran dentro de una doble hélice
- La parte exterior del ADN esta conformada por grupos fosfatos y ribosas

El arreglo de estas bases a lo largo de la cadena de ADN simula una secuencia lineal de letras. Francis Crick fue parte de la milicia durante la Segunda Guerra Mundial, y su familiaridad con códigos militares lo pudo haber influenciado para poder identificar a esta secuencia de bases nucleicas como una especie de código molecular unidimensional [37].

Estos componentes del ADN pueden darnos una mejor idea del porqué toma la forma de una doble hélice. Los grupos fosfatos son solubles en agua, al igual que la ribosa (azúcar). Por otro lado, las bases nucleicas son insolubles en agua.

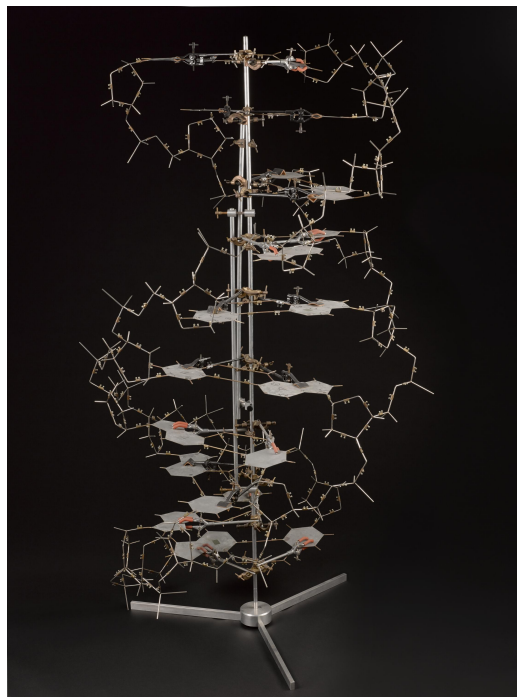


Figura 1.5: Replica del modelo original del ADN por Watson-Crick, tomado de la referencia [41].

Esta insolubilidad de las bases genera ciertas constricciones geométricas sobre la estructura que puede adoptar el ADN. Las bases deben estar “ocultas” del agua que llena a las células, por lo que se encuentran en la parte interior de la doble hélice. Al tomar en cuenta estas características de las componentes del ADN, la solución geométrica posible es una escalera “chueca”, o la famosa doble hélice que todos conocemos [39].

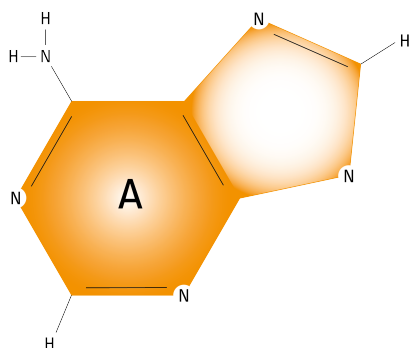
En resumen, la estructura del ADN es un “*cierre molecular*” torcido. Cada hebra del cierre consiste de una columna vertebral exterior (fosfatos y ribosas) y nucleótidos internos actuando como los dientes del cierre. Los puentes de hidrógeno entre las bases complementarias (solo pueden conectarse en pares bien específicos AT-GC) “cierran a las dos hebras”. Los enlaces covalentes son fuertes, ya que mantienen unidos los átomos de cada hélice, mientras que los puentes de hidrógeno son relativamente débiles, lo que permite posible separación de hebras. La estabilidad de este polímero biológico recae en distintas contribuciones sobre las bases tales como: puentes de hidrógeno, fuerzas de van der Waals y el efecto hidrofóbico.

### 1.4.1. Bases nucleicas

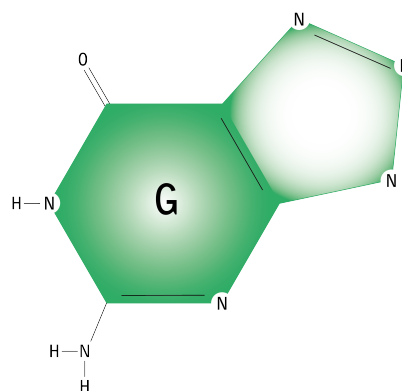
Como se mencionó antes, existen 2 clases de bases nucleicas. Las purinas (cuentan con 2 estructuras de anillos fusionadas A y G, fig. 1.6a y 1.6b) y las pirimidinas (una sola estructura de anillo T y C/U, fig. 1.6c y 1.6d).

Las bases nucleicas son de aproximadamente 0.34 nm de tamaño y por su naturaleza aromática absorben apreciablemente la luz UV. Esta peculiaridad de absorción resulta útil experimentalmente para saber si una estructura de ADN está abierta, debido a que los enlaces covalentes entre la ribosa y la base evitan la buena absorción. Por lo que, entre más abierta la estructura, mejor la absorción UV.

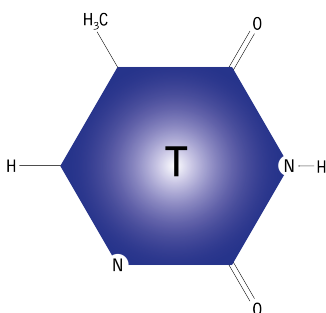
Las bases nucleicas en una cadena de ADN están apiladas entre sí e interactúan por medio de repulsiones de corto alcance y dispersión atractiva [19].



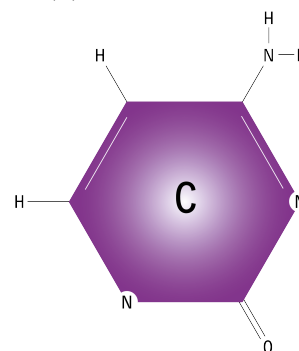
(a) Adenina



(b) Guanina



(c) Timina



(d) Citosina

### 1.4.2. Nucleósidos

Resulta útil conocer cierta nomenclatura a la hora de hablar de la estructura del ADN. En particular, un *nucleósido* consiste de la base nucleica considerando la ribosa unida a ella, por medio de un enlace covalente. Este enlace es de tipo carbono-nitrógeno y se le conoce como enlace glicosídico. También se puede notar que la base nucleica está en un plano por encima del de la ribosa, lo cual es importante a la hora de analizar la apertura de la doble hélice.

### 1.4.3. Nucleótidos

Otra terminología bastante utilizada es la de *nucleótido*. Estos consisten en la base nucleica, la ribosa y un grupo fosfato. Los fosfatos están enlazados al oxígeno unido al quinto (5') carbono de la ribosa. De aquí podemos entender de mejor manera la direccionalidad del ADN. La terminación de la cadena 5' tiene un fosfato, mientras que el extremo 3' tiene un grupo OH, por lo que la secuencia se lee de 5' a 3'.

### 1.4.4. Puentes de Hidrógeno

Las bases que conforman los ácidos nucleicos están bajo dos tipos de interacciones no covalentes. Los puentes de hidrógeno que sostienen a los pares de bases y las interacciones de apilamiento que se dan entre bases continuas de la cadena. Las bases nucleicas son estructural y electrónicamente complementarias, como piezas de rompecabezas. Poseen distribuciones de cargas que favorecen la interacción electrostática. Todos los átomos aceptores poseen carga negativa y los protones correspondientes carga positiva [18].

Los pares de bases complementarias (AT y GC), están enlazadas por medio de puentes de hidrógeno, por lo que entender este tipo de enlace es importante para el análisis de todo tipo de proceso que involucre al ADN.

La naturaleza de este tipo de enlace es debido a la interacción dipolo-dipolo entre el átomo de hidrógeno que se encuentra cerca de un átomo electronegativo (como nitrógeno, oxígeno o flúor) y un segundo átomo electronegativo en la vecindad. Se puede argumentar que el valor de electro-negatividad de cada átomo está ligado con la fuerza que tendrá el puente de hidrógeno. Es decir, el electrón no es compartido de manera equitativa entre el átomo de hidrógeno y el átomo electronegativo, lo que deja al hidrógeno con una carga positiva débil que le permite interactuar con otros átomos.

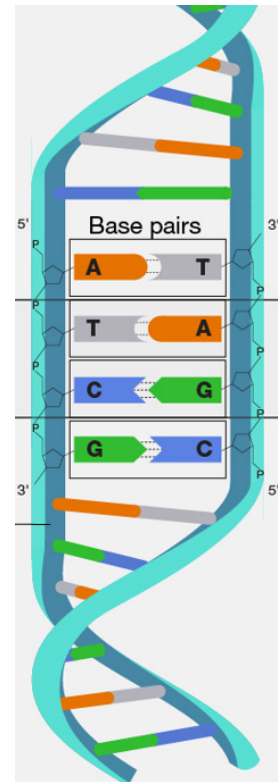


Figura 1.7: Estructura de ADN (grupo fosfato, ribosa y bases). Figura tomada de “National Human Genome Research Institute” [34].

Como se mencionó anteriormente, el emparejamiento de las bases ocurre únicamente entre A-T y G-C. El par de bases Adenina-Timina cuenta con dos enlaces de hidrógeno, mientras que el par Guanina-Citosina tiene 3 enlaces de hidrógeno, por lo que G-C es más estable y podría costar más romper su interacción.

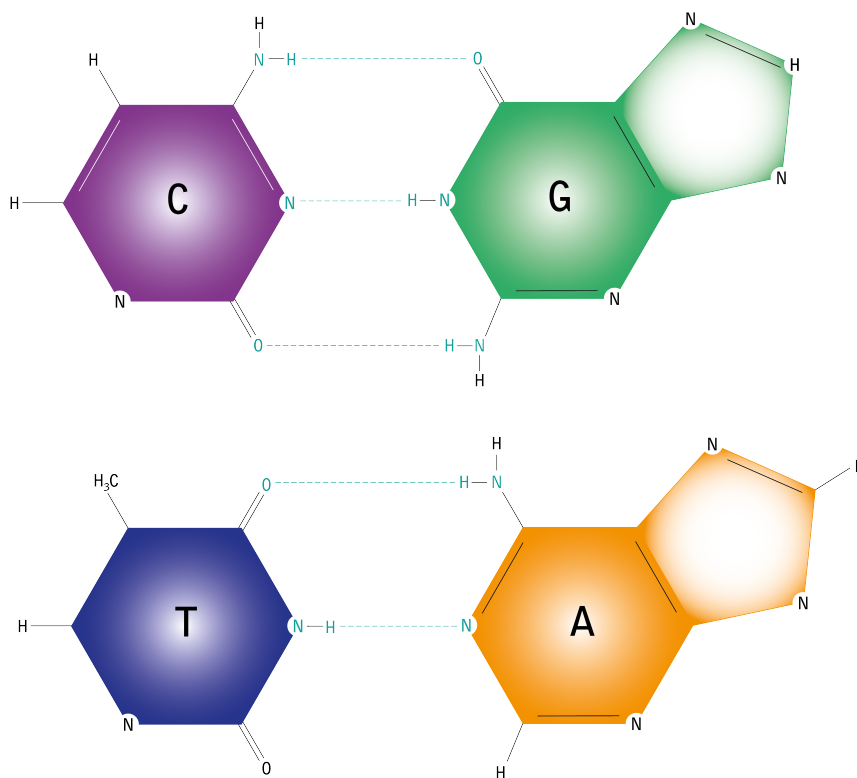


Figura 1.8: Puentes de hidrógeno entre pares de bases CG y TA [43].

#### 1.4.5. Apilamiento

El apilamiento o “stacking”, es una interacción no-covalente muy importante para la estabilidad de los ácidos nucleicos. Se da por apilamiento de las bases nucleicas a lo largo de la cadena del ADN o ARN. Es frecuente encontrarla al tratar con moléculas aromáticas y en particular para el caso de ácidos nucleicos, su contribución principal es por fuerzas de tipo van der Waals, o de dispersión. Aunque también las interacciones electrostáticas (momentos dipolares y cuadrupolares) y la interacción entre orbitales son parte de las contribuciones al apilamiento.

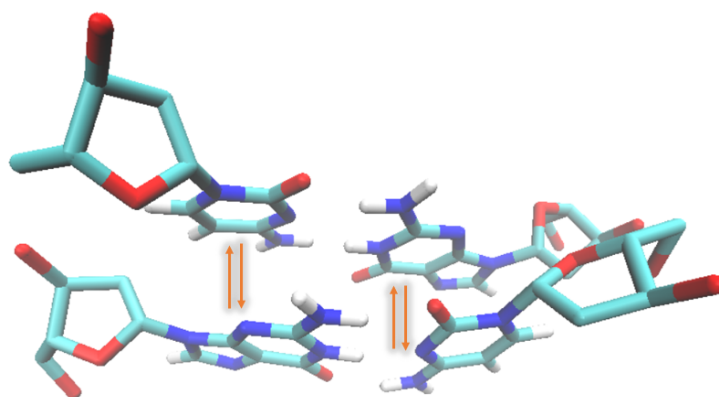


Figura 1.9: Fuerzas de apilamiento entre pares de bases (perpendiculares a los puentes de hidrógeno).

A través de estas interacciones aromáticas, la nube  $\pi$ <sup>13</sup> de una base nucleica, interactúa con la nube  $\pi$  de las bases adyacentes. Dando lugar a este enlace químico no covalente. Es decir, se considera como una interacción débil.

Para que la interacción sea atractiva, las moléculas deben estar desalineadas levemente de manera paralela, es decir, no deben estar completamente emparejadas. Uno de los anillos de las bases está desalineada de tal manera que los átomos de carbono (con cargas parciales negativas) se colocan justo arriba de los átomos de hidrógeno (con cargas parciales positivas) del anillo de la base apilada, dando así una interacción electrostática atractiva.

De manera general, las interacciones entre pares de bases G-C sucesivas son más fuertes en comparación con pares de bases A-T sucesivas. Por lo que, estructuras de ADN con contenido dominante de bases G-C tienden a ser más estables [35].

## 1.5. Enzimas (Helicasas)

La mayoría de las sustancias químicas y moléculas biológicas son estables y no sufren alteraciones en su estado natural. La forma para lograr que ocurra cierta reacción es con ayuda de moléculas catalizadoras, esto es, agentes que nos ayuden a acelerar estos procesos <sup>14</sup>. Estos catalizadores biológicos se les conoce como enzimas. Tienen la capacidad de convertir reactivos en productos al reducir la pendiente energética (energía de activación) que tienen que escalar para producir la reacción (ver teoría de estado de transición [37]).

La palabra enzima (levadura en griego) fue adoptada por el fisiólogo alemán Wilhelm Kühne en 1877 para describir el proceso de fermentación. En nuestro día a día utilizamos estas moléculas para lograr distintos objetivos. Enzimas como las proteasas en los detergentes para quitar manchas, peptidasas que se agregan a la leche para hacer queso o pectinasas en la fruta para hacer vino [37].

En particular, el ADN es una de estas moléculas que se mantiene estable, debido a que

<sup>13</sup>Enlace pi: ocurre cuando dos lóbulos del orbital p de un átomo se solapan con los dos lóbulos de un orbital p de otro átomo.

<sup>14</sup>Aunque no todas las reacciones químicas necesitan catalizadores.



protege la información genética y requieren del uso de enzimas para abrir la doble hélice y lograr copiar los genes en la replicación celular. A este tipo particular de enzimas se les conoce como *helicatas*. Las helicatas hacen uso de la energía proveniente de un proceso de hidrólisis<sup>15</sup> sobre unos nucleótidos especiales que cuentan con varios grupos fosfatos, siendo el más común, el ATP (adenosín trifosfato). El ATP es producido a nivel de respiración celular, donde intervienen varios procesos como glucólisis y el ciclo de Krebs. La energía guardada en los grupos fosfatos del ATP es liberada cuando se rompe el enlace por hidrólisis (generando ADP). Posteriormente, esta energía sirve como combustible para producir la fuerza mecánica que ayudará a la separación de las hebras del ADN [44].

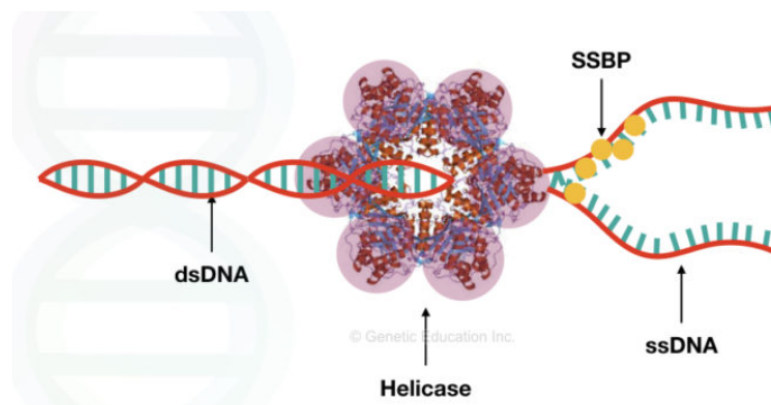


Figura 1.10: Replicación del ADN con ayuda de la helicasa [45]. Donde: ds-DNA (ADN de cadena doble), ss-DNA (ADN de cadena sencilla) y SSBP (proteínas ligantes de ADN de cadena sencilla que se encargan de la estabilización de la apertura del ADN).

La helicasa se enrolla sobre las cadenas y las separan al ir rompiendo (terminando la interacción) los puentes de hidrógeno que sostienen los pares de bases nucleicas, moviéndose secuencialmente de enlace a enlace. Podemos decir que lo que hace esta enzima, es una manipulación de átomos individuales, protones y electrones dentro de moléculas y entre moléculas [37]. Este juego entre partículas permite a la helicasa mantener los átomos del ADN en ciertas posiciones que permiten el rompimiento de enlaces y movimiento a lo largo de las hebras nucleicas. El funcionamiento de este proceso sigue siendo tema de debate y de ahí la importancia de entender como ocurre y en especial en la replicación del ADN. Distintas áreas de la física contribuyen a responder este fenómeno biológico<sup>16</sup>.

Debido a la sensibilidad para la medición directa del comportamiento de la helicasa, se requiere un grupo grande de moléculas enzimáticas para producir cantidades de productos medibles. Por lo que, se tiene datos que reflejan únicamente la actividad promedio de helicatas individuales. Actualmente, el avance de las técnicas experimentales permiten observar dinámicas catalíticas que involucran al ADN y la helicasa de manera individual. Cabe señalar que gracias a la fluorescencia de las bases nucleicas es posible observar el comportamiento del ADN bajo

<sup>15</sup>Una molécula de agua se separa y logra la separación de una macromolécula en sus componentes, liberando energía.

<sup>16</sup>Incluido análisis cuánticos de efecto túnel para analizar disociación de puentes de hidrógeno. Un área de investigación en desarrollo, llamada biología cuántica[46].

esta acción, pero hay dificultades para la observación directa de la helicasa. De aquí que las simulaciones moleculares son una herramienta para incrementar nuestra comprensión de la actividad ADN-helicasa, que experimentalmente sería imposible [35].

## 1.6. Métodos experimentales sobre ADN

Debido a las dimensiones de las moléculas de ADN no es posible visualizarlas de manera natural por el ojo humano. La única manera de hacerlo visible consiste en utilizar grandes cantidades de ADN, hasta volverlo una masa blanca viscosa [47].

Con la llegada de técnicas cristalográficas utilizando rayos X, fue posible elucidar la estructura del ADN y de muchos otros compuestos. La idea detrás de esta técnica experimental está en cristalizar la molécula que se quiere analizar (para tener un arreglo periódico) y bombardearla con rayos X. Los rayos se van a reflejar y difractar a ciertos ángulos debido a los electrones que componen a la sustancia y con esto se puede determinar su estructura. Algo así como analizar la sombra que produce al ser iluminada [48].

Recientemente, la microscopía electrónica nos ha permitido mejorar la resolución con la que se analizan estas moléculas, al aprovechar que la longitud de onda de los electrones es mucho más pequeña. Con esto, se pudo ver por primera vez una “imagen” del ADN de manera individual, confirmando su existencia.

Los avances experimentales de las últimas décadas se enfocaron en técnicas “single molecule”, para la nano-manipulación de moléculas individuales bajo fuerzas externas y condiciones fisiológicas variables [49]. Estas técnicas sostienen un extremo del ADN a una superficie, mientras que el otro se manipula con un tipo de sensor para cuantificar la fuerza aplicada [6].

En condiciones naturales, los procesos moleculares tienen contribuciones de varias moléculas pertenecientes a un ensamble debido a su medio ambiente y a lo mucho aspiramos a encontrar un valor promedio del comportamiento de cierto fenómeno [50]. Con estos experimentos podemos seguir la dinámica en tiempo real y considerando únicamente la molécula de interés. De este tipo de arreglos experimentales que permiten manipular moléculas individuales a alta resolución, los más usados son la microscopía de fuerza atómica (AFM, por sus siglas en inglés, (*atomic force microscopy*.) y pinzas ópticas/magnéticas<sup>17</sup>. Estas técnicas permiten medir y observar cambios estructurales de moléculas bajo distintos tipos de fuerzas aplicadas [6, 52].

En el caso de trampas ópticas, se hace uso de un rayo láser para sostener moléculas individuales<sup>18</sup>. Se han realizado estudios de estiramiento de ADN, con el llamado método “*Flow-stretch*”, en donde se usa un fluido laminar mientras la molécula está sujeta a una micro-esfera atrapada por pinzas ópticas (al igual que con pinzas magnéticas, cambiando el campo magnético para aplicar la fuerza deseada). El desamarre de la doble hélice es equivalente a la altura que se movió la micro-esfera respecto de la superficie. Estos cambios se monitorean con microscopía de luz y se convierten a cambios en distancias entre ds-DNA y ss-DNA (*double stranded* y *single stranded*, respectivamente) [11].

---

<sup>17</sup>Otros métodos utilizados son, grabación de canales iónicos, espectroscopia de fluctuación-correlación, *flow stretching*, entre otros [51]

<sup>18</sup>En 1970, Arthur Ashkin (premio Nobel de Física en 2018) explotó el hecho de que los fotones tienen momento para transportar micro-esferas suspendidas en agua usando rayos láser. Posteriormente, en 1986, se demostraría por primera vez que usando luz altamente enfocada podía servir para atrapar micro-partículas [50]

Por otro lado, AFM utiliza una punta afilada con un cabezal, como una especie de tornamesa. La información topográfica, magnitud de fuerzas aplicadas y manipulación, se obtiene al “tocar” la molécula con la punta afilada (microfabricada, generalmente de silicio). El cambio de forma del cabezal altera la cantidad de luz de láser que se refleja en un diodo, la cual puede ser cuantificada.

Para el análisis del proceso enzimático sobre ADN, se adhiere una micro-esfera a un extremo de la molécula y se mide su posición, mientras se mantiene una fuerza constante.

La principal diferencia entre estos dos tipos de experimentos es el valor de la constante de resorte. Las trampas ópticas tienen una constante de resorte del orden de 0.1 pN/nm, es decir, se consideran resortes blandos [53]. Por lo que, AFM resulta útil para fuerzas grandes (>100 pN), mientras que pinzas ópticas/magnéticas son más efectivas para aplicación de fuerzas pequeñas. Para una tabla con las resoluciones obtenidas por cada uno de estos métodos ver [50].

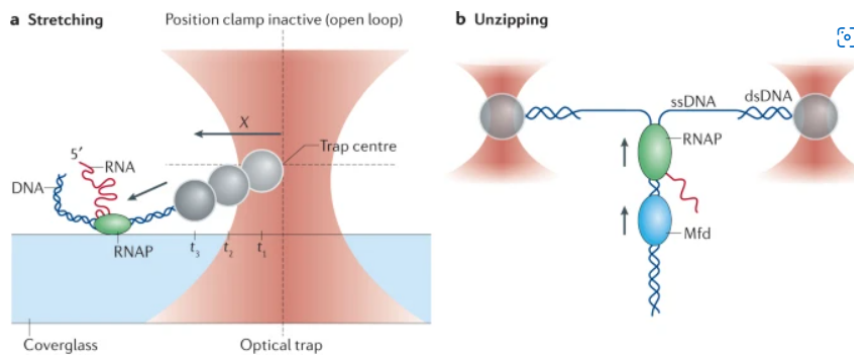


Figura 1.11: Pinzas ópticas para estudio de interacción ADN-proteína [50].

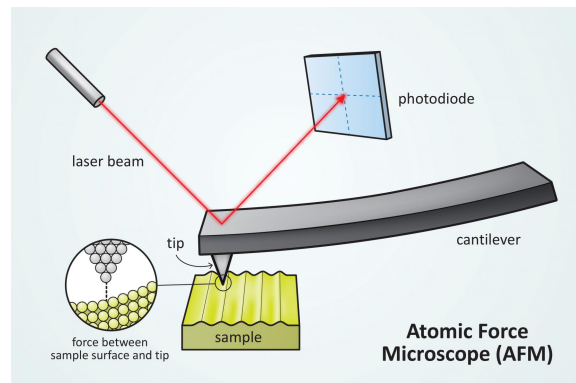


Figura 1.12: Arreglo común AFM [54].

## 1.7. Simulación Molecular

El emplear metodologías computacionales y teóricas para analizar sistemas moleculares nos ayuda a ganar conocimiento de muchos fenómenos interesantes, ya que intentamos imitar el mundo natural sin observación directa de él. Además, tenemos un control total de las características de este mundo y con esto podemos hacer inferencias importantes sobre ciertos comportamientos en las moléculas que están en estudio. Claramente, podemos ver que el alcance de este tipo de métodos depende fuertemente del poder computacional que se esté utilizando, es por eso que en el desarrollo o en la elección del marco teórico que se quiere utilizar para los cálculos, el costo computacional es un factor importante que se debe de analizar. Este es un ejemplo claro del gran rango que tiene este tipo de investigación, ya que no solo dependemos de la teoría química o física sino también de los avances que se dan en las ciencias computacionales, mientras más crezca esa área, de igual forma crecerá la investigación en temas de simulación molecular.

En particular, en el caso del ADN, estas simulaciones se enfocan en entender las propiedades mecánicas de la molécula, flexibilidad, dependencia de la secuencia y reconocimiento por otras moléculas [6].

Las simulaciones moleculares dependen fuertemente de la estructura inicial utilizada, esto es, de los buenos datos experimentales y cristalográficos. Posteriormente, se investigan los movimientos térmicos e interacciones de los átomos en escalas de tiempo cortas (en comparación con el proceso en estado natural). Todos estos cálculos se basan en ciertas metodologías, que se explicaran a detalle más adelante, en donde se puede utilizar dinámica molecular puramente clásica, cuántica o una combinación de ambas. El proceso general es tener una buena estructura inicial, calcular fuerzas en cada átomo, integrar ciertas ecuaciones de movimiento y dejar evolucionar el sistema en el tiempo [55].

El principal problema que se tiene con las simulaciones moleculares es la limitación computacional en el número de átomos que se pueden modelar de manera precisa. El cromosoma humano contiene alrededor de  $5 \times 10^7$  pares de bases, mientras que en simulaciones cuánticas podemos estudiar estructuras de menos de 15 pares de base [55, 56].

Afortunadamente, como se mencionó antes, el ADN es una molécula flexible y varios estudios han demostrado que el comportamiento de segmentos cortos describe de buena manera el comportamiento de estructuras completas. De igual manera, los distintos métodos experimentales trabajan con cadenas pequeñas de las moléculas, por lo que las simulaciones moleculares entregan buenos resultados [56].

Además, las dinámicas moleculares describen estos procesos en escalas temporales de femtosegundos ( $10^{-15}$  seg) o nanosegundos ( $10^{-9}$  seg), mientras que las escalas temporales naturales de los fenómenos biológicos que nos interesan se encuentran en los milisegundos. La razón de movimiento promedio de una helicasa requiere de 10 milisegundos por par de bases [6]. Estas circunstancias implican que existe un error implícito en las dinámicas por simular un proceso que ocurre a escalas distintas.

Hay que tomar en cuenta que el objetivo de las simulaciones moleculares no debe ser la reproducción exacta de datos experimentales, ya que no es posible tener una construcción idéntica entre un experimento real y uno “*in silico*”. Los métodos computacionales deben ser un

Cuadro 1.1: Escalas temporales de procesos biológicos de interés<sup>1</sup>

Escala temporal	Tipo de movimiento
Picosegundo	Vibraciones de enlaces, transiciones conformacionales locales (por ejemplo, en la columna vertebral de azúcar-fosfatos); y tiempos de residencia para moléculas de agua en sitios hidratados.
Nanosegundo	Formación de regiones desnaturalizadas en ADN bajo estrés y cambios conformacionales a grandes escalas (por ejemplo, transiciones de A a B)
Microsegundo	Respiración de bases, asociación de ligandos; y asociación de oligonucleótidos complementarios.
Milisegundo	Experimentos de estiramiento de ADN y reparación de proteínas y procesos de helicasas.
Segundo	Topoisomerización, disociación de ligandos; y remodelación de nucleosoma.

<sup>1</sup> Información obtenida de [6].

complemento, un puente entre teoría y experimento real [6].

El modelado molecular nos ayuda en la interpretación de datos experimentales provenientes de técnicas de nano-manipulación. Esto es debido a que, por medio de dinámicas computacionales, podemos visualizar cambios estructurales que ocurren a nivel atómico, que de otra manera serían imposibles de estudiar.

El objetivo a futuro de estos métodos teóricos-computacionales es lograr el análisis de estructuras más grandes a escalas temporales de micro/milisegundos. En particular, para el ADN, el estudio de distintas estructuras existentes (cruciformes, cuádruples, etc.). Así como el modelado cuántico de interacciones helicasa-ADN con gran precisión.

## Capítulo 2

# Marco Teórico

*"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."*

---

*Marie Curie*

La teoría de la mecánica cuántica es la culminación de una serie de trabajos y escritos que buscaron resolver una serie de problemas para los cuales no se tenía solución a finales del siglo XIX; problemas generados debido a las mejoras de las técnicas experimentales para estudiar sistemas a escala atómica. Sin duda, esta fue la mayor revolución dentro del mundo de la Física, ya que para su desarrollo fue necesario abandonar muchas ideas y principios bien arraigados en el pensamiento de los científicos clásicos, muchos de los cuales no aceptaban esta nueva teoría.

El éxito inminente de la mecánica cuántica para describir el comportamiento de la materia a una escala atómica, fue lo que hizo que la teoría venciera los prejuicios impuestos en aquella época para convertirse en una de las teorías más importantes en la ciencia como hoy en día la conocemos.

Históricamente, se conocen dos formulaciones independientes de la mecánica cuántica, pero equivalentes; la mecánica matricial de *W. Heisenberg* (1925) y la mecánica ondulatoria de *E. Schrödinger* (1926). Esta última se basa en una generalización del postulado de *de Broglie* y describe la dinámica de la materia microscópica mediante la ecuación de Schrödinger.

## 2.1. Ecuación de Schrödinger

Para la descripción de sistemas cuánticos no relativistas, tenemos la ecuación de Schrödinger dependiente del tiempo [57]:

$$i\hbar \frac{\partial \psi(\vec{r}, t)}{\partial t} = \hat{\mathcal{H}}\psi(\vec{r}, t) \quad (2.1)$$

donde  $\hat{\mathcal{H}}$  representa el operador asociado a la energía total del sistema y se le conoce como el operador Hamiltoniano del sistema, dado por la suma del operador de energía cinética y el operador de energía potencial.

$$\hat{\mathcal{H}} = \hat{K} + \hat{V}. \quad (2.2)$$

Si  $\hat{V}$  no depende explícitamente del tiempo, se puede hacer una separación de variables para resolver la ecuación de Schrödinger y obtener la ecuación independiente del tiempo:

$$\hat{\mathcal{H}}\psi = E\psi \quad (2.3)$$

La función de onda  $\psi$  contiene toda la información referente al sistema según la interpretación probabilística de Born (*Interpretación de Copenhague*<sup>1</sup>).

## 2.2. Problema molecular

El buen planteamiento del Hamiltoniano es esencial para construir la ecuación de Schrödinger y obtener la función de onda, la cual nos dará la descripción del sistema cuántico.

En este trabajo se estudiaron moléculas de muchos átomos y por ende de muchas partículas. El Hamiltoniano que describe una molécula con  $m$  núcleos y  $n$  electrones está dado por<sup>2</sup>:

$$\hat{\mathcal{H}} = - \sum_{A=1}^m \frac{\nabla_{\vec{R}_A}^2}{2m_A} - \sum_{i=1}^n \frac{\nabla_{\vec{r}_i}^2}{2} - \sum_{A=1}^m \sum_{i=1}^n \frac{Z_A}{|\vec{R}_A - \vec{r}_i|} + \sum_{A=1}^{m-1} \sum_{B>A}^m \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} + \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{1}{|\vec{r}_j - \vec{r}_i|} \quad (2.4)$$

Los dos primeros términos representan las energías cinéticas de los núcleos y de los electrones respectivamente, seguido de la energía potencial atractiva entre electrones y núcleos, el cuarto término representa la repulsión entre los núcleos y finalmente el término de la repulsión interelectrónica.  $Z_A$  y  $Z_B$  son las cargas nucleares y los vectores  $\vec{R}_A$  y  $\vec{r}_i$  representan las posiciones del A-ésimo núcleo y del i-ésimo electrón respectivamente, la ecuación está expresada en unidades atómicas [58].

<sup>1</sup>Existen otras interpretaciones de la cuántica. Entre las más conocidas, la interpretación de muchos universos y la interpretación de Bohm.

<sup>2</sup>Se ignoran efectos relativistas e interacciones magnéticas.

### 2.2.1. Aproximación Born-Oppenheimer: movimiento electrónico y nuclear

Si pudiéramos resolver la ecuación de Schrödinger con el Hamiltoniano molecular (2.4) de manera exacta, encontraríamos todas las propiedades exactas del sistema, en particular la energía. Lamentablemente, esta ecuación no es separable, por lo que se debe hacer uso de ciertas aproximaciones.

La aproximación de *Born-Oppenheimer* toma el hecho de la gran diferencia de masas entre los electrones y núcleos<sup>3</sup>. Debido a que los núcleos son más masivos que los electrones, los electrones se mueven más rápido y experimentan el campo electrostático de los núcleos como si estuvieran fijos [58]. Esta aproximación nos entrega dos ecuaciones para el movimiento de electrones y núcleos, respectivamente, y que además son más fáciles de resolver.

- Movimiento electrónico en el campo de los núcleos (lentos).
- Movimiento nuclear bajo la energía potencial determinada por la energía electrónica.

#### Ecuación de Schrödinger electrónica

En la aproximación BO, podemos omitir el término de la energía cinética de los núcleos en el Hamiltoniano molecular y además el término de repulsión nuclear  $V_{NN}$  será una constante una vez que se definan las coordenadas nucleares fijas. Con esto obtenemos el llamado Hamiltoniano electrónico:

$$\hat{H}_{elec} = -\sum_{i=1}^n \frac{\nabla_{\vec{r}_i}^2}{2} - \sum_{A=1}^{m-1} \sum_{i=1}^n \frac{Z_A}{|\vec{R}_A - \vec{r}_i|} + \sum_{i=1}^{n-1} \sum_{j>i}^n \frac{1}{|\vec{r}_j - \vec{r}_i|} \quad (2.5)$$

$$\hat{V}_{NN} = \sum_{A=1}^{m-1} \sum_{B>A}^m \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} \quad (2.6)$$

Formalmente, lo que estamos suponiendo es que podemos escribir la función de onda total molecular como un producto de función electrónica y nuclear:

$$\psi[(\vec{r}_i), (\vec{R}_A); t] = \sum_k \psi_{elec_k}[(\vec{r}_i), (\vec{R}_A)] \psi_{nucl_k}(\vec{R}_A) \quad (2.7)$$

Al sustituir en la ecuación de Schrödinger, usando la ortonormalidad de las funciones y despreciando términos de correlación electrón-núcleo (aproximación BO) se llega a la ecuación electrónica.

<sup>3</sup>Masa del protón es aprox. 2000 veces la masa del electrón.

$m_p = 1.67267 \times 10^{-27} \text{ kg}$

$m_e = 9.01939 \times 10^{-31} \text{ kg}$



Las funciones de onda y energías dependen paramétricamente de las coordenadas nucleares, ya no son variables, la ecuación de Schrödinger electrónica se escribe entonces como:

$$i\hbar \frac{\partial \psi_{elec}}{\partial t} = - \sum_{i=1}^n \frac{\hbar^2}{2} \nabla_i^2 \psi_{elec} + \left( \int \psi_{nuc}^* \hat{V} \psi_{nuc} dR \right) \psi_{elec} \quad (2.8)$$

Los electrones están en un campo promedio (sección verde de la ec. 2.10) debido a los núcleos, es decir, no ven a cada núcleo individual sino al comportamiento colectivo. Se puede notar que aún aparece la función de onda nuclear, por lo que la separación no está del todo completa.

Este potencial promedio está constituido por varias interacciones, de las cuales la interacción núcleo-núcleo complica la integración (involucra posiciones de dos núcleos). Para simplificar, tomamos en cuenta la característica de los núcleos de ser puntuales (en comparación con los electrones<sup>4</sup>) y calculamos la probabilidad como productos de deltas de Dirac:

$$|\psi_{nuc}(R_\alpha(t); t)|^2 = \Pi_\alpha \delta(R_\alpha - R_\alpha(t)) \quad (2.9)$$

Hacer esto nos permite tener el potencial promedio dependiendo paramétricamente de las posiciones nucleares al instante  $t$ :

$$i\hbar \frac{\partial \psi_{elec}}{\partial t} = - \sum_{i=1}^n \frac{\hbar^2}{2} \nabla_i^2 \psi_{elec} + \left( \hat{V}(r_i; R_\alpha(t)) \right) \psi_{elec} \quad (2.10)$$

Además, los electrones se adaptan instantáneamente al movimiento de los núcleos y es por esto que podemos hacer a la ecuación electrónica, una ecuación estacionaria y hacer uso de una ecuación de Schrödinger independiente del tiempo (como veremos en la sección del límite clásico).

$$\hat{\mathcal{H}}_{elec} \psi_{elec} = E_{elec} \psi_{elec} \quad (2.11)$$

La aproximación de *Born-Oppenheimer*<sup>5</sup> nos ayuda a pseudo-separar el problema molecular (ec. 2.4), “pseudo” porque aún existe una dependencia paramétrica de las coordenadas nucleares. Los electrones están inmersos en un campo promedio debido a los núcleos. Si fuera de otra forma no existiría la formación de átomos y moléculas, por lo que las ecuaciones deben estar acopladas.

<sup>4</sup>El diámetro del núcleo es del orden de  $10^{-15}m$  y el campo del electrón de  $10^{-10}m$ , la densidad de probabilidad de los núcleos los podemos aproximar como una delta de Dirac o gaussianas muy delgadas (es así como se resuelve la integral de campo promedio en la ec. (2.10)).

<sup>5</sup>La aproximación es efectiva para estados electrónicos base, puede llegar a fallar para estados excitados.

### Ecuación de Schrödinger nuclear

De manera similar, ahora consideramos el movimiento de los núcleos en un campo promedio debido a los electrones. Aplicando la aproximación BO, llegamos a la ecuación dinámica para los núcleos, donde el Hamiltoniano nuclear es:

$$\hat{\mathcal{H}}_{nuc} = - \sum_{A=1}^m \frac{\nabla_{\vec{R}_A}^2}{2} + \sum_{A=1}^{m-1} \sum_{B>A}^m \frac{Z_A Z_B}{|\vec{R}_A - \vec{R}_B|} + E_{elec}[\vec{R}_A] \quad (2.12)$$

Sustituimos en la ecuación de Schrödinger y aplicando un procedimiento similar al del caso electrónico, usando la ortonormalidad de las funciones y despreciando términos de correlación electrón-núcleo (aproximación BO) se llega a la ecuación nuclear dependiente del tiempo.

$$i\hbar \frac{\partial \psi_{nuc}}{\partial t} = - \sum_{A=1}^n \frac{\hbar^2}{2M_A} \nabla_A^2 \psi_{nuc} + \left( \int \psi_{elec}^* \hat{\mathcal{H}}_{elec} \psi_{elec} dr \right) \psi_{nuc} \quad (2.13)$$

Los núcleos no ven a cada electrón individual, sino, una nube electrónica que promedia los movimientos electrónicos que sentirán los núcleos.

Como veremos más adelante, al resolver la ecuación de Schrödinger electrónica (ec. 2.11) a diferentes configuraciones nucleares y encontrar  $E_{elec}$ , se puede utilizar un proceso de optimización para encontrar un mínimo de la función de varias variables  $U$  (PES) (ver ec. 2.14), la cual nos servirá como estructura de equilibrio de la molécula y con la que se calcularan las propiedades deseadas.

$$U = E_{elec} + V_{NN} \quad (2.14)$$

El conocimiento  $\psi_{nuc}$  permite obtener información de traslación, rotación y vibración de la molécula. Así como datos de movimiento nuclear y espectroscópicos.

Existen distintas maneras de resolver las ecuaciones electrónicas (2.11) y nuclear (2.13) respectivamente, dependiendo del nivel de teoría deseado. En particular, en esta tesis se hace un tratamiento clásico sobre los núcleos y la parte electrónica se resuelve con DFT, como veremos más adelante.

Hasta el momento tenemos las ecuaciones para el movimiento electrónico y nuclear, dadas respectivamente por:

$$\hat{\mathcal{H}}_{elec} \psi_{elec} = E_{elec} \psi_{elec} \quad (2.15)$$

$$i\hbar \frac{\partial \psi_{nuc}}{\partial t} = - \sum_{A=1}^m \frac{\hbar^2}{2M_A} \nabla_A^2 \psi_{nuc} + \left( \int \psi_{elec}^* \hat{\mathcal{H}}_{elec} \psi_{elec} dr \right) \psi_{nuc} \quad (2.16)$$

En la siguiente sección se presentan de manera sintetizada los diferentes métodos que existen para resolver la ecuación de Schrödinger electrónica (2.11).

## 2.3. Métodos de solución para la estructura electrónica molecular

No hay que olvidar que el punto de partida es la aproximación de *Born-Oppenheimer* y la meta es la resolución de la ecuación de Schrödinger electrónica (ec. 2.11) para un conjunto de coordenadas nucleares fijas. En general existen varios métodos para lograr esto, los métodos semi-empíricos, *ab initio*, y DFT.

- **Semi-empírico:** Hacen uso de un Hamiltoniano simplificado (se pueden ignorar electrones de core, y se omiten algunas integrales de dos electrones), además de parámetros cuyos valores han sido ajustados para coincidir con datos experimentales y resultados de cálculos *ab initio*.
- ***Ab initio*:** Proviene del latín *primeros principios* y utiliza el Hamiltoniano correcto y no hace uso de datos experimentales, únicamente valores de constantes físicas fundamentales.

**Perturbativos:** Inician con un cálculo de Hartree-Fock y corrigen la correlación electrónica con términos perturbativos.

- Møller Plesset: El hamiltoniano de orden cero, es una suma de operadores de Fock [59].

**Variacionales:** Las energías aproximadas serán igual o mayores a la energía exacta. Se propone una forma matemática para la función de onda  $\psi$  y se realiza un proceso de minimización de la energía al variar parámetros de esta función.

- **Hartree-Fock**
  - Usa la aproximación de campo medio y trata a las repulsiones electrónicas de manera promedio. La función de onda se escribe como producto de orbitales moleculares (determinante de Slater)<sup>6</sup>.
- **Interacción de Configuraciones (CI, por sus siglas en inglés)**
  - Inicia con una función de onda HF y crea nuevos determinantes al promover electrones a orbitales desocupados. Al usar todas las posibles excitaciones, logramos el conocido “*full CI*”.
- **Cúmulos Acoplados**
  - Similar a CI, en el sentido que la función de onda es una combinación lineal de determinantes. La manera de construir los determinantes es más complicada.
- **Teoría de funcional de la densidad:** Se puede determinar la energía de la molécula y todas las propiedades del sistema, haciendo uso de la densidad electrónica  $\rho$  en lugar de la función de onda. Debido a que en ciertas aproximaciones del funcional de intercambio y correlación se han utilizado algunos parámetros empíricos, hay cierto debate entre si considerar a DFT un método *ab initio*, ya que estos parámetros se pueden omitir para ciertos cálculos [58-60].

---

<sup>6</sup>Para más sobre teoría de Hartree-Fock ver el apéndice A

### 2.3.1. Conjunto de funciones base

Tanto para el método de Hartree-Fock-Roothaan, como para DFT, como veremos posteriormente, parte fundamental del cálculo de estructura electrónica depende de escoger una buena base con la cual expresar los orbitales moleculares.

En principio, se necesita un conjunto completo de funciones base para representar los espin-orbitales exactamente y el uso de un número infinito de funciones base resultan en la energía de HF igual a la dada por la expresión variacional (A.2). Una base infinita no es computacionalmente posible, por lo que se debe elegir la base de tal forma de minimizar el número de integrales de 2-electrones y además que el esfuerzo para resolverlas sea pequeño.

El tipo de funciones base más utilizadas son las de tipo exponencial,

$$g(r) = f(x, y, z)e^{-\xi r^n} \quad (2.17)$$

- $n = 1$  STO (Slater)
- $n = 2$  GTO (Gaussianas)

Un conjunto de bases de este tipo, son las conformadas por funciones tipo Slater (STO's, por sus siglas en inglés) [61] centradas en cada átomo de la molécula. La parte radial es más sencilla que en el caso de funciones hidrogenoides, ya que se sustituyen los polinomios de Laguerre, por simples  $r^k$ .

Sin embargo, las integrales con funciones de tipo Slater requieren el uso de integración numérica, la cuál se vuelve más complicada para moléculas no-lineales. Además, la dificultad se introduce con la presencia de más átomos, ya que causa dificultades para evaluar las integrales, para moléculas medianas se necesitan de 50 a 500 funciones base, lo que da un aproximado de  $10^{10}$  integrales de repulsión inter-electrónica [59].

Para ayudar a este problema, S. Boys [59] propuso en 1950 el uso de funciones tipo gaussianas (GTO's). El único problema de estas funciones es que no tienen el pico deseado para valores cercanos a  $r \rightarrow 0$  (cercaos al núcleo) y decae de manera más rápida para  $r \rightarrow \infty$ . Lo que nos lleva a malas descripciones para orbitales atómicos. La ventaja está en que podemos usar una combinación lineal de varias gaussianas centradas en el mismo núcleo atómico, o sea una base más grande. Además, la resolución de este tipo de integrales se puede realizar de manera analítica, a diferencia de las integrales de tipo Slater. Debido a la característica matemática importante de este tipo de funciones, el producto de dos gaussianas centradas en dos puntos distintos es equivalente a una sola gaussiana centrada en un tercer punto, por lo que el número de integrales de 2-electrones se reduce considerablemente, mejorando así el costo computacional. Al tomar distintos valores para  $\xi$  en la ecuación 2.17 podemos construir nuestro conjunto de gaussianas *primitivas*, a la combinación lineal se le conoce como gaussianas *contraídas*. En particular, en este trabajo se hace uso de una base de tipo 6-31g\*<sup>7</sup>, donde tenemos una gaussiana *contraída* compuesta de 6 primitivas para representar orbitales internos. Para los orbitales de valencia se usan dos funciones, una *contraída* de 3 primitivas y 1 sola primitiva difusa<sup>8</sup>. El asterisco, significa la inclusión de funciones de polarización.

<sup>7</sup>Bajo la notación de Pople. Esta base agrega funciones p y d a los átomos no hidrogenoides.

<sup>8</sup>Una función difusa se refiere a una función con exponentes pequeños, que representan adecuadamente la

### Base mínima

El conjunto base más simple es una base mínima en la cual una función es usada para representar cada orbital de capa interna y de valencia, según la teoría elemental de valencia. Por ejemplo:

- 1 función para H y He (orbital 1s)
- 5 funciones de Li hasta Ne (orbitales 1s, 2s y 2p)
- 9 funciones desde Na hasta Ar
- y así sucesivamente . . .

Una base doble-zeta (DZ) utiliza dos funciones base para cada orbital de valencia en lugar de una, similarmente para una base triple-zeta (TZ) [59].

#### 2.3.2. Correlación electrónica y métodos post-HF

El método de HF toma en cuenta las interacciones entre los electrones solo de manera promediada como vimos anteriormente, esto lleva a ciertos errores en los cálculos, debido a que los electrones se repelen entre ellos, es decir, tienden a mantenerse fuera del camino entre ellos. Podemos pensar mejor en esto al usar la famosa analogía de una ciudad con tráfico [62]. Los edificios están en posiciones fijas (núcleos) y los carros (electrones) se mueven alrededor de ellos, con movimientos que pueden ser complejos. HF toma a un solo carro (electrón) y considera que su movimiento es tal que evita las calles (lugares) donde hay más tráfico. Es decir, vemos un promedio del tráfico (un mapa de densidad de tráfico como en los GPS, dado por la nube de densidad de probabilidad), no a cada carro de manera individual. Debido a esto, el conductor tendrá una imagen borrosa de los otros automóviles, lo cual podría ser peligroso (errores en el cálculo por solo considerar promedios).

Por lo que se dice que el movimiento de los electrones está correlacionado. Mejorar esta correlación es el principal motivo para el desarrollo de teorías mejores a HF (DFT siendo una de estas teorías).

---

parte lejana del núcleo para los orbitales. Útiles para modelar sistemas con interacciones débiles.

### 2.3.3. Teoría del Funcional de la Densidad (DFT)

Uno de los grandes problemas que existen, más allá de calcular la función de onda molecular para algún sistema en particular, es el de almacenar computacionalmente la solución si es que la tuviéramos. Necesitaríamos mucho poder de memoria, lo cual sería extremadamente costoso. Por lo que, encontrar algún método que le dé la vuelta a calcular la función de onda molecular total de manera exacta y aun así encontrar las propiedades moleculares que nos interesan, sería algo muy útil.

DFT nace precisamente de esta necesidad de encontrar una función dependiente de menos variables que la función de onda electrónica y que además nos sirva para encontrar las propiedades moleculares antes mencionadas, junto con la inspiración otorgada por el método de Hartree-Fock para moléculas y los trabajos previos de Thomas y Fermi con la densidad electrónica. La idea general es escribir la energía electrónica del sistema en términos de la densidad electrónica. Se dice que  $E$  es un funcional de la densidad, es decir, para una  $\bar{\rho}(r)$  hay una única energía correspondiente.

El trabajo de Thomas y Fermi, los llevo al famoso modelo de Thomas-Fermi (1927). Ellos intentaron escribir cada contribución a la energía total de la forma siguiente:

$$E_{\alpha}[\rho] = \int \rho(\vec{r}) \epsilon_{\alpha}[\rho(\vec{r})] d^3\vec{r}, \quad (2.18)$$

pero este modelo no funcionaba bien, las energías totales no eran buenas en comparación con datos experimentales, la densidad no decae exponencialmente a largas distancias. La complicación más grande está en escribir la contribución de la energía cinética en términos de la densidad. Lo que este modelo nos dice es que la energía es un funcional de la densidad electrónica y que la densidad se podía relacionar de manera única con el potencial externo, aunque la prueba llegaría hasta 1964 con el primer teorema de Hohenberg y Kohn [63].

La formulación de DFT recae en la ayuda de los dos teoremas de H-K. El primer teorema nos dice que podemos utilizar la densidad electrónica del estado base, en lugar de la función de onda, para describir todas las propiedades del sistema en cualquier estado electrónico. El segundo teorema nos habla de la existencia de un funcional, el cual toma su valor mínimo con  $\rho_0$ . A continuación, se presentan de manera formal estos dos teoremas.

### 2.3.4. Primer teorema de Hohenberg y Kohn

Con este teorema se demostró que para una molécula con un estado base no degenerado, las propiedades electrónicas moleculares, como la energía molecular del estado base, la función de onda y demás propiedades, se pueden determinar de manera única con ayuda de la densidad de probabilidad electrónica del estado base  $\rho_0(x, y, z)$ , donde la densidad de probabilidad electrónica se define como la probabilidad de encontrar cualquiera de los  $N$  electrones con cualquier espín, dentro del elemento de volumen  $dr$  (ec. 2.33). De las palabras de Walter Kohn<sup>9</sup> “Finally it occurred to me that for a single particle there is an explicit elementary relation between the potential  $v(r)$  and the density”:

$$\rho(\vec{r}) = \sum_{i=1}^N |\phi_i(\vec{r})|^2 \quad (2.19)$$

En realidad, lo que probaron fue que  $\rho_0(r)$  determina de manera única el potencial externo (hasta una constante arbitraria) y el número de electrones, los cuales a su vez determinan el Hamiltoniano y las funciones de onda electrónicas. Por lo que podemos determinar todas las propiedades del sistema, en particular la energía.

Generalmente, la demostración se realiza al suponer que existen dos potenciales externos  $V_a$  y  $V_b$  que no generan la misma densidad electrónica del estado base hasta llegar a un resultado absurdo [63].

Al tomar los valores esperados de los términos del funcional  $E_0 = E_0[\rho_0]$  se puede notar que hay dos funcionales los cuales no tienen una forma conocida,

$$E_0 = E_v[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{ne}[\rho_0] + \bar{V}_{ee}[\rho_0] \quad (2.20)$$

El primer término representa el promedio de la energía cinética de los electrones, el segundo representa el promedio de las atracciones entre electrones y núcleos y por último las repulsiones inter-electrónicas. Generalmente, se designa a los términos desconocidos como  $F[\rho_0] = \bar{T}[\rho_0] + \bar{V}_{ee}[\rho_0]$  debido a esto, no podemos calcular  $E_0$ , por lo que se usa un segundo teorema.

<sup>9</sup>Galardonado con el premio Nobel en Química (1998) por el desarrollo de DFT, junto con John Pople.

### 2.3.5. Segundo teorema Hohenberg-Kohn

Hohenberg y Kohn probaron también que para toda función de densidad de prueba  $\rho_{tr}(\vec{r})$  que satisfice:

1.  $\int \rho_{tr}(\vec{r}) d\vec{r} = n$
2.  $\rho_{tr}(\vec{r}) \geq 0$

$E_0 \leq E_v[\rho_{tr}]$  donde  $E_0 = E_v[\rho_0]$  con  $\rho_0$  la densidad electrónica del estado base. Este teorema establece que la verdadera densidad electrónica del estado base minimiza al funcional de la energía, justo como la función de onda del estado base nos minimiza el valor esperado de un Hamiltoniano electrónico en el método de HF.

El teorema de Hohenberg y Kohn solo nos dice que es posible calcular todas las propiedades electrónicas del estado base molecular usando  $\rho_0$  y sin tener que calcular la función de onda, pero no nos dice como calcular  $E_0$ , solo nos dice que existe tal relación entre  $E_0$  y  $\rho_0$ .

En 1965 Walter Kohn y L. J. Sham propusieron un esquema práctico para encontrar  $\rho_0$  y por ende  $E_0$ , al relacionar el sistema interactuante a un problema auxiliar no interactuante [64].



### 2.3.6. Esquema Kohn y Sham

En la presentación del recibimiento del premio Nobel de Kohn en 1998, él mencionó que su objetivo era extraer ecuaciones tipo Hartree, pero para el funcional de la energía Hohenberg-Kohn, de aquí la inspiración antes mencionada en la que se basó DFT.

Kohn y Sham consideraron un sistema ficticio de  $n$ -electrones no interactuantes ( $S$ ), que experimentan un mismo potencial externo  $V_s(\vec{r}_i)$ , condicionado a que este potencial esta asociado a una densidad de probabilidad electrónica  $\rho_s(\vec{r})$ , y que la densidad de este sistema sea igual a la densidad de probabilidad electrónica real del estado base de la molécula de interés,  $\rho_s(\vec{r}) = \rho_0(\vec{r})$ . Es un potencial *mágico* que nos genera la densidad del estado base del sistema real.

Se hace esto debido a que para un sistema de partículas no interactuantes, la energía cinética se encuentra fácilmente y es exacta (conocemos los términos del Hamiltoniano en términos de orbitales, no de la densidad).

Como el sistema de referencia consiste de partículas no interactuantes, sabemos que la función del estado base se puede escribir como una combinación lineal antisimétrica de productos de orbitales (determinantes de Slater, de espin orbitales de Kohn-Sham). Entonces de la siguiente ecuación:

$$\begin{aligned} E_0 = E_v[\rho_0] &= \int \rho_0(\vec{r})v(\vec{r})dr + \bar{T}[\rho_0] + \bar{V}_{ee}[\rho_0] \\ &= \int \rho_0(\vec{r})v(\vec{r})dr + F[\rho_0] \end{aligned} \quad (2.21)$$

Kohn y Sham reescribieron a los funcionales desconocidos como:

$$\Delta\bar{T}[\rho] = \bar{T}[\rho] - \bar{T}_s[\rho] \quad (2.22)$$

$$\Delta\bar{V}_{ee}[\rho] = \bar{V}_{ee}[\rho] - \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} dr_1 dr_2 \quad (2.23)$$

donde  $\bar{T}[\rho]$  y  $\bar{T}_s[\rho]$  representan los funcionales de la energía cinética del sistema *real* y *ficticio*, respectivamente. Podemos notar que para el caso del término de interacción electrónica en el caso del sistema ficticio, usamos la expresión clásica, con una distribución de carga continua. Con esto podemos escribir entonces:

$$E_v[\rho] = \int \rho(\vec{r})v(r)dr + \bar{T}_s[\rho] + \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} \vec{dr}_1 \vec{dr}_2 + \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho] \quad (2.24)$$

De aquí obtenemos la famosa definición del funcional de correlación-intercambio:

$$E_{xc}[\rho] = \Delta\bar{T}[\rho] + \Delta\bar{V}_{ee}[\rho]$$

La motivación para hacer estas definiciones es intentar expresar  $E_v[\rho]$  en términos de tres cantidades fáciles de evaluar de  $\rho$ , más el término  $E_{xc}$ . Por lo que podemos ver que para lograr una descripción aceptable se necesita una buena aproximación del funcional  $E_{xc}$ .

Para encontrar la densidad se hace uso de la definición anteriormente vista (ec. 2.33) y de

igual manera de las reglas de Slater-Condor, para escribir  $E_0$  como:

$$E_0 = - \sum_{\alpha} Z_{\alpha} \int \frac{\rho(\vec{r}_1)}{r_{1\alpha}} dr_1 - \frac{1}{2} \sum_{i=1}^n \langle \theta_i^{KS}(1) / \nabla_1^2 / \theta_i^{KS}(1) \rangle + \frac{1}{2} \int \int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} dr_1 dr_2 + E_{xc}[\rho] \quad (2.25)$$

Por lo tanto, podemos encontrar  $E_0$  y  $\rho$  si podemos encontrar los orbitales KS ( $\theta_i^{KS}$ ) y si sabemos también la forma del funcional de correlación-intercambio.

Para encontrar los orbitales de Kohn y Sham hacemos uso del teorema variacional que nos dice que podemos encontrar  $E_0$  al variar la densidad sujeta a ciertas restricciones, de tal forma que nos minimice el funcional de la energía  $E_v[\rho]$ . De forma equivalente podemos variar a los orbitales KS y después utilizar la fórmula (ec. 2.33) para encontrar la densidad.

De la misma forma en que podemos ver que los orbitales ortonormales que minimizan la expresión de la energía de HF satisfacen las ecuaciones de HF, podemos mostrar que los orbitales KS que minimizan la expresión de la energía satisfacen:

$$\left[ -\frac{1}{2} \nabla_i^2 - \sum_{\alpha} \frac{Z_{\alpha}}{r_{1\alpha}} + \int \frac{\rho(\vec{r})}{r_{12}} dr_2 + V_{xc}(1) \right] \theta_i^{KS}(1) = \epsilon_i^{KS}(1) \theta_i^{KS}(1) \quad (2.26)$$

que también se puede escribir como:

$$\left[ -\frac{1}{2} \nabla_1^2 + V_{eff}(1) \right] \theta_i^{KS}(1) = \epsilon_i^{KS}(1) \theta_i^{KS}(1) \quad (2.27)$$

Para finalmente obtener, al mejorar la notación, las famosas ecuaciones de Kohn y Sham (ec. 2.28), donde podemos notar el parecido con las ecuaciones de Hartree-Fock.

$$\hat{h}^{KS}(1) \theta_i^{KS}(1) = \epsilon_i^{KS}(1) \theta_i^{KS}(1) \quad (2.28)$$

Se realizó una variación de  $E$  respecto a los orbitales. Matemáticamente, estamos haciendo un cambio en el funcional  $E[\rho]$  debido a  $\rho \rightarrow \delta\rho$ . En particular, para el potencial de correlación-intercambio tenemos un cambio  $\delta E_{xc}$ . Se puede probar que este cambio en el funcional esta definido a través de la derivada funcional de  $E_{xc}$ , denotada por  $\frac{\delta E_{xc}[\rho]}{\delta \rho(\vec{r})}$ .

El potencial de correlación-intercambio (ec. 2.29) se encuentra al obtener la derivada funcional de  $E_{xc}$  y depende de cuanto cambie  $E_{xc}[\rho]$  cuando  $\rho(x, y, z)$  cambia por una pequeña cantidad en una región pequeña centrada en  $(x, y, z)$

$$V_{xc}(\vec{r}) = \frac{\delta E_{xc}[\rho(\vec{r})]}{\delta \rho(\vec{r})} \quad (2.29)$$

Por lo que podemos decir que nuestro potencial *mágico*, propuesto en el esquema de Kohn-Sham tiene la forma:  $V_s = V + V_{coulomb} + V_{xc}$ .

La diferencia de las ecuaciones de KS con las de HF, es que  $V_{xc}$  si toma en cuenta el intercambio (antisimetría) y correlación electrónica, Hartree-Fock no incluye la correlación electrónica.

Como no sabemos la forma exacta del funcional  $E_{xc}[\rho]$  no sabemos tampoco la de  $V_{xc}[\rho]$ . Hay que recordar que los orbitales de Kohn y Sham son orbitales para el sistema ficticio no interactuante, por lo que estrictamente hablando carecen de significado físico, solo ayudan a

encontrar la densidad electrónica  $\rho$ . El método de Kohn y Sham usa estos orbitales simplemente como intermediarios para encontrar la densidad electrónica.

A partir de aquí, todo se centra en encontrar buenas aproximaciones para el funcional de correlación-intercambio como se verá posteriormente (gracias a que es un término pequeño en comparación con los demás).

El procedimiento para realizar un cálculo de KS-DFT es el siguiente:

1. Empezamos con una densidad inicial de prueba, la cual se encuentra usualmente al superponer densidades electrónicas calculadas de átomos individuales en la geometría molecular escogida.
2. Con esta prueba inicial encontramos el potencial  $V_{xc}[\rho]$  el cual se utiliza para resolver las ecuaciones de Kohn y Sham (ec. 2.28).
3. Al resolver las ecuaciones, generalmente se expanden los orbitales KS en términos de alguna base de funciones para encontrar ecuaciones parecidas a las de HF-Roothaan.
4. Los orbitales obtenidos en este primer cálculo se usan para encontrar una densidad mejorada, la cual nuevamente se usa para encontrar el potencial de correlación-intercambio y así sucesivamente.
5. Las iteraciones continúan hasta alcanzar algún criterio de convergencia previamente establecido.
6. Una vez que el método converge, podemos encontrar  $E_0$ .

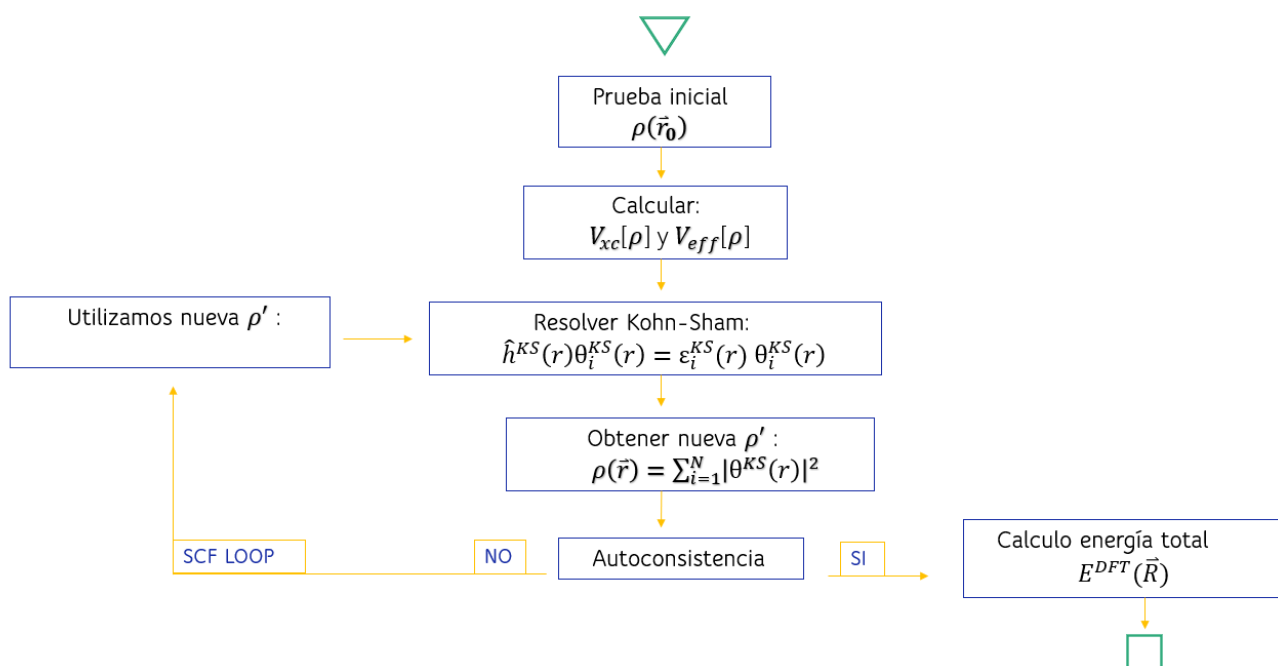


Figura 2.1: Diagrama de flujo para el cálculo de la energía electrónica bajo la metodología de DFT-Kohn/Sham.

### 2.3.7. Funcionales de correlación e intercambio

Como se mencionó antes, parte vital de un buen cálculo con DFT es usar una buena aproximación del funcional de correlación-intercambio, en este apartado solo se menciona brevemente las aproximaciones empleadas para los cálculos de esta tesis en particular. Una herramienta que se utiliza en las aproximaciones de los funcionales, es escribirlo como la contribución de dos cantidades  $E_{xc} = E_x + E_c$ , es decir, un funcional de intercambio y uno de correlación respectivamente.

#### Aproximación de densidad local (LDA)

Hohenberg y Kohn notaron que si  $\rho$  varía lentamente con la posición entonces  $E_{xc}[\rho]$  está descrito adecuadamente por:

$$E_{xc}^{LDA}[\rho] = \int \rho(\vec{r}) \epsilon_{xc}(\rho) d\vec{r} \quad (2.30)$$

donde  $\epsilon_{xc}$  es la energía de correlación-intercambio por electrón en un gas electrónico homogéneo con densidad electrónica  $\rho$ <sup>10</sup>. La idea es que sabemos que la densidad electrónica es inhomogénea, pero localmente (dentro de un volumen pequeño) podemos suponerla homogénea.

Se extiende más esta aproximación si permitimos que electrones con spín opuesto tengan orbitales espaciales KS diferentes, para la llamada *Local Spin Density Approximation* (LSDA) [58, 59].

#### Corrección de gradiente generalizado (GGA)

Este tipo de funcionales dejan de lado la suposición de un modelo de un gas de electrones uniforme incluyendo gradientes de la densidad para una mejor descripción de la variación de la densidad electrónica con la posición, es decir,  $E_{xc}$  depende de  $\rho$  en un cierto punto y de  $\rho$  en lugares cercanos [59, 62].

#### Funcionales PBE y B3LYP

Dos funcionales utilizados comúnmente en la literatura, para la realización de distintos cálculos, son *PBE* y *B3LYP*. El primero recibe ese nombre por *Perdew-Burke-Ernzerhof* y ocupan funcionales de intercambio y correlación que no usan parámetros empíricos, pero si ocupan algunos ajustes de datos obtenidos previamente con Hartree-Fock.

B3LYP es un funcional híbrido, el cual mezcla una fórmula para el funcional de intercambio y funcionales GGA de intercambio y correlación, fue propuesto por Becke y el 3 se refiere a los tres parámetros empíricos utilizados para que se ajusten a energías de atomización experimentales. La mezcla que toma este funcional es la siguiente:

- $E_{xc}$  toma el funcional del método LSDA y agrega 20 % de la diferencia entre  $E_{xc}^{HF}$  y  $E_{xc}^{LSDA}$ . Mezcla 72 % del funcional de intercambio de Becke ( $E_x^{B88}$ ) y un 81 % del potencial

<sup>10</sup>El Jellium es el nombre que se le dió a un sistema hipotético eléctricamente neutro, el cual consiste de un número infinito de electrones interactuando moviéndose en el espacio con carga positiva uniformemente distribuida [58]

de correlación de Lee-Yang-Park ( $E_c^{LYP}$ ) y finalmente un 19% del potencial Vosko-Wilk-Nusair.

$$E_{xc} = E_{xc}^{LSDA} + 0.20(E_{xc}^{HF} - E_{xc}^{LSDA}) + 0.72E_x^{B88} + 0.81E_c^{LYP} + 0.19E_c^{VNW} \quad (2.31)$$

Como se mencionó en la introducción es un funcional que produce resultados aceptables para cálculos sobre moléculas biológicas y en particular en ácidos nucleicos [59, 60, 65].

## 2.4. Superficie de energía potencial

Una vez que encontramos la energía electrónica  $E_{elec}$  de nuestro sistema, podemos calcular distintas propiedades importantes, como la geometría de equilibrio. Para esto, es necesario calcular las derivadas de la energía potencial con respecto a las coordenadas nucleares.

Las derivadas pueden ser obtenidas de manera numérica, al calcular la energía potencial a distintas geometrías y determinar el cambio en energía conforme las coordenadas nucleares varían. Sin embargo, métodos de gradientes nos permiten obtener las derivadas de manera analítica, más rápidamente y de manera eficaz.

### Optimización geométrica

Se conoce como *optimización geométrica* al proceso de encontrar la estructura geométrica de equilibrio de una molécula, esta llamada estructura de equilibrio corresponde al arreglo nuclear que minimiza la función de varias variables  $U$  (ec. 2.14), que es la energía electrónica con repulsión internuclear.

Esta función  $U$  da la superficie de energía potencial (PES) de la molécula; debido al gran número de variables  $U$  (ec. 2.14) es una superficie en un espacio abstracto de  $3M - 5$  dimensiones para moléculas lineales y  $3M - 6$  para moléculas no lineales (3 traslacionales, 3 rotacionales y los demás vibracionales). Estas superficies especifican como varía la energía del sistema cuando el arreglo nuclear cambia. Es la relación que existe entre la estructura molecular y su energía, cada punto sobre esta superficie representa una geometría particular.

Para encontrar  $U$  se debe resolver la ecuación electrónica de Schrödinger (2.11), con alguno de los métodos antes mencionados.

Una vez que encontremos la geometría de equilibrio, se realiza un cálculo *single point*, esto es, resolver la ecuación electrónica de Schrödinger nuevamente. De este último arreglo optimizado se obtienen posteriormente todas las propiedades moleculares del sistema, de ahí la importancia de conocer a grandes rasgos como se encuentra esta geometría al minimizar  $U$ .

Existen diversos métodos [59] para encontrar un mínimo (local por lo menos) de una función de varias variables, estos métodos se basan en la siguiente idea en general.

Empezamos con una geometría de prueba inicial basada en valores experimentales de distancias y ángulos de enlace. Posteriormente, buscamos el mínimo más cercano a la estructura inicial, se escoge una base y se ocupa algún método para resolver aproximadamente la ecuación de Schrödinger electrónica como los que se mencionaron anteriormente y se encuentra  $U$  y su gradiente  $\nabla U$ . El algoritmo cambia las  $3N - 6$  coordenadas nucleares a un nuevo conjunto que es probable que este más cerca de un mínimo que el conjunto inicial, esto se repite hasta alcanzar un criterio de convergencia.

Un método comúnmente usado es el del descenso más rápido<sup>11</sup>, cuyo algoritmo sería el siguiente:

1. Se da la estructura inicial  $(R_1^0, \dots, R_m^0)$
2. Se evalúa el gradiente  $\nabla U$  en ese punto inicial, si este da cero, entonces  $(R_1^0, \dots, R_m^0)$  es un punto estacionario
3. Encontraremos un  $(R_1', \dots, R_m')$  en la vecindad de  $(R_1^0, \dots, R_m^0)$  tal que  $\nabla U(R_1', \dots, R_m') < E((R_1^0, \dots, R_m^0))$
4. Calcular  $\nabla U$  en el nuevo punto, resolviendo el problema electrónico en  $R_1', \dots, R_m'$  se hace  $\vec{R}_i^0 = \vec{R}_i'$  y se regresa al paso 2

---

<sup>11</sup>Existen más de los llamados métodos de derivadas: descenso más rápido y gradiente conjugado, corresponden a métodos de primer orden. De segundo orden el más utilizado es el método de Newton-Raphson [66].

## 2.5. Análisis poblacional (Mulliken)

Algo que resulta muy útil en dinámica molecular, es utilizar algún método que permita segmentar o dividir la densidad electrónica para contar electrones asociados con un átomo que forma parte de una molécula entera. Las cargas atómicas no se pueden observar experimentalmente, ya que no corresponden a una propiedad física única<sup>12</sup>. Debido a esto los métodos pueden cambiar y resultan un tanto arbitrarios.

En 1955, *R. S. Mulliken* propuso un método para realizar esta partición de la densidad electrónica [67], al obtener fracciones de electrón correspondientes al porcentaje de tiempo que un electrón se encuentra en un cierto núcleo. Es decir, cada núcleo tiene un número (no necesariamente entero) de electrones asociado a él.

La suposición fundamental es que el traslape (overlap) entre dos orbitales se distribuye equitativamente. Como vimos anteriormente, el orbital molecular lo escribimos como una combinación lineal de funciones base [58, 66, 68]:

$$\phi_i = \sum_{s=1}^b C_{si} \xi_s \quad (2.32)$$

Esta expansión contiene varios términos, algunos son el cuadrado de funciones bases, algunos otros son productos de funciones base distintas, que al integrar nos dan el traslape:

$$\rho(r) = \sum_{i=1}^N |\phi_i(r)|^2 = C_{1i}^2 \xi_1^2 + C_{2i}^2 \xi_2^2 + \dots + 2C_{1i}C_{2i} \xi_1 \xi_2 + 2C_{1i}C_{3i} \xi_1 \xi_3 + \dots \quad (2.33)$$

Al integrar y hacer uso de la normalización de la base, tenemos:

$$1 = C_{1i}^2 + C_{2i}^2 + \dots + 2C_{1i}C_{2i}S_{12} + 2C_{1i}C_{3i}S_{13} + \dots \quad (2.34)$$

En el método de Mulliken, las integrales de un orbital no se suman. En lugar de eso, la contribución de una función base en todos los orbitales se suman para dar una *población neta* de esa función base. Por ejemplo, un electrón en el orbital  $\phi_i$  contribuye  $C_{1i}^2$  a la población neta en  $\xi_1$ ,  $C_{2i}^2$  en  $\xi_2$  y así sucesivamente.

Similarmente, los traslapes para un par de funciones base se suman para todos los orbitales y poder determinar la población de traslape (*overlap*) para ese par de funciones. Podemos ver que hay una contribución  $2C_{1i}C_{2i}S_{12}$  a la población de traslape entre  $\xi_1$  y  $\xi_2$ , y así sucesivamente<sup>13</sup>. Si tenemos  $n_i$  electrones en el orbital  $\phi_i$ , podemos definir la contribución neta de  $\xi_r$  debido a esos electrones ( $n_{r,i}$ ). Similarmente, la contribución a la población de traslape entre  $\xi_r$  y  $\xi_s$ :

$$n_{r,i} = n_i C_{ri}^2 \quad n_{r-s,i} = n_i (2C_{ri}C_{si}S_{rs}) \quad (2.35)$$

<sup>12</sup>Un átomo en realidad tiene muchos electrones orbitando, no cargas parciales. No existe un operador cuántico para la carga atómica.

<sup>13</sup>Las poblaciones de traslape pueden ser cero por simetría. Anti-enlazantes en el caso en que son negativas y cuando son positivas pueden ayudarnos a identificar enlaces químicos.



Al sumar sobre el número de orbitales ocupados, queda:

$$n_r = \sum_i n_{r,i} \qquad n_{r-s} = \sum_i n_{r-s,i} \qquad (2.36)$$

Podemos notar que el número total de electrones en la molécula es:

$$\sum_r n_r + \sum_{r>s} \sum_s n_{r-s} = n \qquad (2.37)$$

En algunos casos resulta conveniente repartir a los electrones sobre las funciones base únicamente, sin poblaciones de solapamiento. Es aquí donde se usa la suposición fundamental de dividir cada población de solapamiento ( $n_{r-s}$ ) equitativamente entre los dos orbitales. Tendremos una población bruta  $N_r$  que es igual a la población neta más la mitad de cada población de solapamiento a la que esa función base contribuye, es decir:

$$N_r = n_r + \frac{1}{2} \sum_{s \neq r} n_{r-s} \qquad (2.38)$$

La población bruta de todos los orbitales centrados en un cierto átomo, pueden sumarse de tal manera de calcular una población bruta atómica para ese átomo en particular. Por ejemplo, para todas las funciones base centradas en un átomo B, tendríamos la población atómica bruta  $N_B = \sum_{r \in B} N_r$ .

Esta población atómica bruta se puede restar de la carga nuclear para obtener finalmente la carga neta  $q_B$  del átomo B:

$$q_B = Z_B - N_B \qquad (2.39)$$

Este método es fuertemente criticado debido a que depende fuertemente del tamaño de la base y las cargas de Mulliken solo tienen significado si las funciones base pueden ser asociadas a un sitio atómico. Existe el problema general de orbitales tipo p, d y f, los cuales están propagados lejos del núcleo al cual están asociados. Debido a esto, los orbitales pueden estar más cerca de otros núcleos y aun así la ocupación electrónica de tales orbitales se asigna al átomo en el cual están centrados.

Por otro lado, es un método sencillo de implementar y computacionalmente barato<sup>14</sup>.

---

<sup>14</sup>Existen otros métodos para análisis de cargas, como análisis de población natural (NPA) o método de Bader (átomos en moléculas) [68].

## 2.6. Método de solución movimiento nuclear

Pensando en que los núcleos son mucho más masivos que los electrones y que además podemos tratarlos como partículas puntuales bien localizadas (a diferencia de las posiciones electrónicas) podemos llevar la ecuación cuántica para los núcleos a su límite clásico donde cumplen con la mecánica Newtoniana. Se tiene una ecuación de Newton con un potencial dado por el movimiento electrónico, la cual se puede resolver numéricamente.

### 2.6.1. Límite clásico

El famoso principio de correspondencia habla de la conexión entre la física clásica y la mecánica cuántica. Afirma que existe una reducción de una teoría general a una más particular en un cierto límite.

Históricamente, la primera versión de este principio fue propuesta por Max Planck en 1906, al observar que su ley de radiación para la densidad de energía se reduce a la ley de Rayleigh-Jeans clásica en el límite de  $\hbar \rightarrow 0$ . Posteriormente, Niels Bohr presenta su versión del principio de correspondencia en 1913, al notar que el espectro de frecuencia del átomo de hidrógeno se reduce a su espectro clásico en el límite de número cuántico  $n$  grande. Vale la pena mencionar la versión del teorema de Eherenfest. Este teorema muestra que al representar los observables con sus valores esperados, estos cumplen las ecuaciones clásicas. Incluso, entrega una forma de analizar leyes de conservación.

En 1926, G. Wentzel, H. Kramers y L. Brillouin encontraron una forma de obtener una ecuación de Hamilton-Jacobi partiendo de la ecuación de Schrödinger al usar el límite de Planck [69]. Me parece la manera más útil de mostrar la reducción clásica de la ecuación de los núcleos, ya que es la manera más clara y además hace contacto con una de las interpretaciones de la mecánica cuántica, la llamada *interpretación de Bohm*<sup>15</sup>.

La idea parte de reescribir la función de onda en su forma polar:

$$\psi(x, t) = D(x, t)e^{S(x, t)/\hbar} \quad (2.40)$$

Al hacer este cambio (inspirado en un cambio parecido que conecta óptica geométrica con ondulatoria), la ecuación de Schrödinger se separa en dos ecuaciones acopladas (al separar parte real e imaginaria), una para la amplitud  $D(x, t)$ , y otra para la fase  $S$ . Cuando se introduce la densidad de probabilidad como el módulo al cuadrado de esta función de onda, la densidad de flujo y haciendo las sustituciones necesarias, notamos que la ecuación para  $D$  es una ecuación de continuidad. Por otro lado, la ecuación para  $S(x, t)$  nos da:

$$\frac{\partial S}{\partial t} + \sum_{n=1}^N |\nabla S|^2 + V - \sum_{n=1}^N \frac{\hbar^2}{2M} \frac{\nabla^2 D}{D} = 0 \quad (2.41)$$

donde el último término del lado izquierdo de la ecuación anterior es el potencial de Bohm. Este potencial es el que guía el movimiento de las partículas cuánticas, visto desde la interpretación de Bohm de la mecánica cuántica. Es un potencial puramente cuántico y suponemos que

<sup>15</sup>La interpretación de Bohm, formulada por David Bohm, quien empezó a cuestionar la *interpretación de Copenhague* a finales de 1940. Esta interpretación considera a la mecánica cuántica como determinista, al considerar que los sistemas físicos tienen una función de onda y una posición real, la cual evoluciona de manera determinista por una ecuación guía.

se origina por una onda que acompaña o guía a las partículas cuánticas.

Si pensamos en el límite clásico como  $\hbar \rightarrow 0$  nos quitamos el potencial de Bohm y reconocemos que  $S(x, t)$  es la función de acción, la cual es solución de la ecuación de Hamilton-Jacobi. La teoría de Hamilton-Jacobi dice que en un ensamble de partículas que tiene trayectorias que son normales a una superficie de  $S$  constante,  $\nabla S(x, t)$  es el momento de la partícula bajo estudio.

Cabe señalar que  $\hbar$  es una constante y no se puede hacer tender a cero, por lo que es mejor decir, en el límite donde el potencial de Bohm se hace cero (cuando el término del potencial es despreciable en comparación con los otros términos). Al tomar el momento de la partícula como el gradiente de la función de acción, las partículas tienen un momento bien definido y a su vez sufren la acción de un potencial cuántico, junto con los potenciales clásicos específicos del sistema a tratar. Los electrones tienen posiciones y momentos bien definidos, esto es, trayectorias continuas y bien definidas, las cuales están controladas por el potencial cuántico de Bohm. Finalmente, usando el límite clásico (es decir, cuando el potencial de Bohm  $\rightarrow 0$ ) la ecuación de los núcleos es la ecuación de Hamilton-Jacobi (equivalente a una ecuación de Newton). Lo podemos ver al aplicar el gradiente a la ecuación anterior e identificando el momento de la partícula, se obtiene:

$$\frac{d\vec{P}_A}{dt} = -\nabla_A \int \psi_{elec}^*(r_i; t) \hat{\mathcal{H}}_{elec}(r_i; R_A) \psi_{elec}(r_i; t) dr \quad (2.42)$$

$$\frac{d\vec{P}_A}{dt} = -\nabla_A V_{elec}(R_A; t) = \vec{F}_A^{DFT}(R_A) \quad (2.43)$$

Se obtiene una ecuación clásica (Newton) para el movimiento nuclear. El potencial  $V_{elec}$  es el generado por el movimiento electrónico y se encuentra al resolver la ecuación de los electrones. La reducción de la ecuación de Schrödinger de muchos electrones a la descripción de un potencial clásico requiere de dos aproximaciones. La aproximación de Born-Oppenheimer, la cual dice que los electrones se mueven más rápido que los núcleos y podemos considerar que reaccionan de manera inmediata al movimiento de los núcleos. Esta condición permite tratar a los electrones de manera separada. La segunda aproximación toma en cuenta que los núcleos son mucho más pesados que los electrones y los trata como partículas puntuales bien localizadas que satisfacen las ecuaciones de Newton. En dinámica molecular el efecto de los electrones se aproxima como una superficie de energía potencial [69-71]<sup>16</sup>.

<sup>16</sup>Para más sobre el límite clásico de la mecánica cuántica en el contexto de la dinámica molecular, ver [72]

### 2.6.2. Integrador de Verlet

La integración de Verlet es un método numérico (hay muchos otros), usado para integrar las ecuaciones de movimiento de Newton, esto es, ayuda a encontrar soluciones aproximadas de estas integrales. Particularmente, fue utilizado en el ámbito de la dinámica molecular por *Loup Verlet* en 1960 [73], incluso se llegó a usar para describir la órbita del cometa Halley en 1909 [74].

El método de Verlet hace uso de las posiciones y aceleraciones al tiempo  $t$ , y las posiciones al tiempo anterior ( $t - \Delta t$ ), para calcular las nuevas posiciones al tiempo ( $t + \Delta t$ ).

La ecuación de Newton es una ecuación diferencial de segundo orden, la cual podemos escribir como:

$$\vec{F}_A = m_A \frac{d\vec{v}_A}{dt} \quad (2.44)$$

$$\vec{v}_A = \frac{d\vec{r}_A}{dt} \quad (2.45)$$

Podemos integrar estas ecuaciones en pequeños intervalos de tiempo  $\Delta t$  al hacer una expansión en serie de Taylor de la coordenada de la partícula alrededor del tiempo  $t$ , esto es:

$$r_A(t + \Delta t) = r_A(t) + v(t)\Delta t + \frac{f(t)}{2m}(\Delta t)^2 + \mathcal{O}^3 + \dots \quad (2.46)$$

$$r_A(t - \Delta t) = r_A(t) - v(t)\Delta t + \frac{f(t)}{2m}(\Delta t)^2 - \mathcal{O}^3 + \dots \quad (2.47)$$

donde despreciamos términos mayores de tercer orden. Al sumar las dos ecuaciones anteriores obtenemos:

$$r_A(t + \Delta t) \approx 2r_A(t) - r_A(t - \Delta t) + \frac{f(t)}{2m}(\Delta t)^2 \quad (2.48)$$

La aceleración ( $a(t)$ ) se calcula con la fuerza a la que está sometido el átomo  $A$  al tiempo  $t$ . Como vemos, el método es muy directo y requiere únicamente del almacenamiento de dos posiciones ( $r(t)$  y  $r(t - \Delta t)$ ) y la aceleración ( $a(t)$ ).

Una vez que se tienen las nuevas posiciones, se desechan las posiciones al tiempo ( $t - \Delta t$ ), las posiciones actuales se convierten en posiciones viejas y las posiciones nuevas en las posiciones actuales.

Una de las desventajas de este método numérico, es que las posiciones  $\vec{r}(t + \Delta t)$  se obtuvieron al añadir un término  $(a\Delta t)^2$  a la diferencia de dos términos mucho más grandes ( $2r(t)$  y  $r(t - \Delta t)$ ), lo cual introduce una pérdida de precisión. Se puede ver que el error en el cálculo de la nueva posición es del orden de  $(\Delta t)^4$ . Además, podemos ver que la velocidad no aparece de manera explícita y se debe calcular hasta que las posiciones del siguiente paso de tiempo sean conocidas. En distintas circunstancias, resulta conveniente conocer la velocidad en el mismo paso del tiempo que las posiciones. En nuestro caso es importante, ya que hacemos uso de la energía cinética del sistema, la cual depende de las velocidades.

**Velocity-Verlet:**

El método de *Velocity-Verlet* [75], nos entrega  $r(t)$ ,  $v(t)$  y  $a(t)$ , todo en el mismo paso del tiempo  $t$ :

$$\vec{r}_A(t + \Delta t) = \vec{r}_A(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)}{2}(\Delta t)^2 + \dots \quad (2.49)$$

$$\vec{v}_A(t + \Delta t) = \vec{v}_A(t) + \frac{1}{2}\Delta t[\vec{a}(t) + \vec{a}(t + \Delta t)] \quad (2.50)$$

podemos calcular las nuevas velocidades únicamente hasta haber calculado las nuevas posiciones y de ahí las fuerzas. Es decir, necesitamos la fuerza en el nuevo paso de tiempo [66, 68, 72, 76]. Se utilizará un integrador numérico similar a Velocity-Verlet para resolver las ecuaciones de las partículas del ADN dentro del reservorio térmico (ecuaciones de Langevin), como veremos más adelante.

## 2.7. Reservorio térmico de Langevin

La molécula de ADN se encuentra dentro de la célula, donde las condiciones de temperatura, presión y solvatación, afectan directamente el comportamiento biológico de la molécula.

El medio ambiente en el que se encuentra sumergido el ADN debe ser tomado en cuenta. La manera sencilla de hacerlo, sería introduciendo las partículas del solvente de manera explícita y optimizar el arreglo completo. Debido al elevado costo computacional que eso impondría, nos lleva a buscar alternativas para tratar al solvente.

De manera general, podemos notar que las partículas del medio ambiente agitan térmicamente a las partículas que constituyen a la estructura biológica a analizar y esto es similar a tomar al ADN como una partícula browniana con estructura atómica.<sup>17</sup>

Las moléculas dentro de la solución sufren colisiones con partículas del solvente. Además, experimentan fuerzas de fricción conforme se mueven a través de él. Como veremos más adelante, el termostato de Langevin nos permite incluir estos efectos, por medio de la inclusión de una fuerza aleatoria y un término de fricción que depende de la velocidad de la partícula, para poder representar un baño térmico simple [77-79].

### Movimiento Browniano

El movimiento Browniano es un término acuñado en honor al botánico Robert Brown, quien observó por primera vez este fenómeno en 1827. En sus investigaciones, Brown notó que ciertas partículas de polen inmersas en agua presentan movimientos irregulares debido a colisiones con las moléculas del solvente.

Posteriormente, en 1905 Albert Einstein publica un artículo describiendo el movimiento Browniano y obtiene una ecuación de difusión para estas partículas, donde relaciona el coeficiente de difusión con el desplazamiento cuadrático medio. La explicación de este fenómeno por parte de Einstein, sirvió como evidencia formal de la existencia de átomos y moléculas<sup>18</sup>.

Einstein notó que una partícula en un fluido sin fricción recibe una especie de golpe debido a colisiones con las moléculas, por lo que la velocidad de la partícula cambia. Sin embargo, si el fluido es muy viscoso, el cambio en velocidad se disipa rápidamente y el resultado neto de la colisión es simplemente cambiar el desplazamiento de la partícula. El efecto acumulado de las colisiones es producir saltos aleatorios de la partícula.

Los movimientos de la partícula son debido a un proceso estocástico gobernado por la función  $f(x, t)$ , la cual es solución de una ecuación de difusión. Representa el número de partículas cuya coordenada  $x$  (se puede generalizar a 3 dimensiones) incrementa entre  $t = 0$  y  $t = t'$  por una cantidad  $x$  y  $x + dx$ .

---

<sup>17</sup>No es trivial hacer esta suposición y la idea es pensar en que cada átomo que conforma al ADN, esta en contacto con su **propio** solvente. Debido a que cada partícula recibe golpeteos aleatorios por parte de su propio baño térmico y además las partículas pertenecen a la misma molécula (amarradas por fuerzas sistemáticas), después de un intervalo de tiempo, la molécula en general sufrirá un desplazamiento estocástico.

<sup>18</sup>En 1908, Jean Perrin calculó el número de Avogadro con la descripción del movimiento Browniano de Einstein. Se le otorgó el premio Nobel en 1926 por esta contribución.

Podemos decir, en manera de resumen, que esta descripción se basa en el concepto de la distribución de densidad de probabilidad de las partículas Brownianas y la ecuación para la evolución temporal de esta distribución.

$$\frac{\partial f(x, t)}{\partial t} = D \frac{\partial^2 f}{\partial x^2} \quad (2.51)$$

### Ecuación de Langevin

En 1908, Paul Langevin obtuvo el mismo resultado, pero con una ecuación de movimiento para la variable aleatoria misma. Está basada en la evolución temporal de la variable aleatoria describiendo el proceso y no en la evolución temporal de la distribución de probabilidad, como en el caso de la derivación de Einstein. Esto abrió el camino al campo de las llamadas ecuaciones diferenciales estocásticas.

Langevin escribió las ecuaciones de movimiento para la partícula Browniana de acuerdo a Newton y suponiendo que la partícula es afectada por una fuerza debido al baño térmico, dividida en dos componentes; una fuerza disipativa proporcional a la velocidad de la partícula, y una fuerza fluctuante aleatoria la cual se describe únicamente por medios estocásticos.

$$m \frac{d\vec{v}}{dt} = -\lambda\vec{v} + \eta(t). \quad (2.52)$$

A la ecuación de Langevin se le puede agregar el potencial específico de cada problema (fuerzas sistemáticas). Podemos pensar en estos dos términos anteriormente descritos, como un término que reduce la energía cinética (término disipativo) y un término estocástico que aumenta la energía cinética. Esto se describe por el *teorema de fluctuaciones disipativas*, que nos dice que en todo sistema en el que haya un proceso de disipación de energía y conversión a calor (fricción, por ejemplo), entonces existe un proceso revertido relacionado con fluctuaciones térmicas <sup>19</sup>. Históricamente, Langevin toma al término de fricción dado por la ley de Stokes  $\lambda v = 6\pi\eta a v$  (suponiendo una partícula esférica) [77-80].

#### 2.7.1. Ecuación Langevin-Zwanzig

Algo conocido para nuestro problema molecular es el Hamiltoniano del sistema. Por lo que, por medio de transformaciones de Legendre, es posible conocer también el Lagrangiano, dado por la suma de las energías cinéticas y de los potenciales interactuantes ( $T - V$ ).

El siguiente paso para el tratamiento correcto y específico del baño térmico en el que se pondrá a la estructura del ADN es la obtención de la ecuación de Langevin a partir de una función Lagrangiana. Esto fue realizado por Robert Zwanzig, al hacer la simplificación de tratar

<sup>19</sup>Por ejemplo, el caso de una corriente eléctrica circulando en un cable: La corriente decae debido a la resistencia, ya que disipa energía eléctrica y la transforma en calor (Efecto Joule). La fluctuación correspondiente es el ruido de Johnson.

Para el caso de movimiento Browniano: El objeto se mueve en el solvente y experimenta un arrastre que disipa energía cinética y la transforma en calor. La fluctuación correspondiente es el movimiento Browniano.

al baño térmico como un mar de osciladores armónicos. [77, 81, 82].

La función Lagrangiana que se utiliza es la de tipo Zwanzig-Langevin [83], la cual considera las partículas del baño como un conjunto de osciladores armónicos con frecuencias  $w_{i\alpha}$ :

$$\mathcal{L} = \sum_{\alpha=1}^m \frac{M_{\alpha}}{2} \dot{R}_{\alpha}^2 + \sum_{\alpha=1}^m \sum_{i=1}^s \frac{m_{i\alpha}}{2} \dot{q}_{i\alpha}^2 - v(q_{i\alpha}, R_{\alpha}) - U([R_{\alpha}]) \quad (2.53)$$

La idea a grandes rasgos, es expandir en serie de Taylor el término  $v(q_{i\alpha}, R_{\alpha})$  y considerar únicamente términos cuadráticos, es decir, una aproximación de osciladores armónicos.

$$\begin{aligned} \mathcal{L} = & \sum_{\alpha=1}^m \frac{M_{\alpha}}{2} \dot{R}_{\alpha}^2 + \sum_{\alpha=1}^m \sum_{i=1}^s \frac{m_{i\alpha}}{2} \dot{q}_{i\alpha}^2 \\ & - \sum_{\alpha=1}^m \sum_{i=1}^s \left[ \frac{m_{i\alpha}}{2} w_{i\alpha}^2 q_{i\alpha}^2 - c_{i\alpha} q_{i\alpha} R_{\alpha} + \frac{c_{i\alpha}^2}{2} \frac{R_{\alpha}^2}{m_{i\alpha} w_{i\alpha}^2} \right] - U([R_{\alpha}]) \quad (2.54) \end{aligned}$$

En este caso, tenemos 2 tipos diferentes de partículas:

- partículas que conforman a la estructura del ADN ( $R_{\alpha}, m_{\alpha}$ )
- partículas que conforman al baño térmico ( $q_{i\alpha}, m_{i\alpha}$ )

El primer y segundo término del lado derecho de la ecuación anterior corresponden a la energía cinética de los osciladores, y de las partículas dentro del baño térmico, respectivamente. Tenemos interacciones entre el  $\alpha$ -ésimo átomo del ADN y la  $i$ -ésima partícula del baño térmico. El segundo término de la sección entre corchetes representa las interacciones de la molécula con las partículas del baño térmico, y se introduce la suposición de que las interacciones con los átomos de la molécula son bilineales (acoplamientos bilineales). El tercer término en la sección con corchetes nos ayuda a quitar preferencias entre las posiciones de las partículas. El potencial  $U$  representa interacciones de los átomos del ADN mismo y con fuerzas externas (es la energía que se obtendrá con DFT).

Hay que notar que el doble índice que se presenta  $q_{i\alpha}$  significa que cada  $\alpha$ -ésima partícula del ADN interactúa con su propio baño térmico (con  $i$  partículas), el cual depende del coeficiente de viscosidad y de la temperatura.

Para encontrar las ecuaciones de movimiento se hace uso de las ecuaciones de Euler-Lagrange con el Lagrangiano anterior [77, 84]:

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = 0 \quad (2.55)$$

Tenemos  $3n - 1$  ecuaciones diferenciales de segundo orden, para las coordenadas generalizadas  $\dot{q}_j, q_j$ . Al introducir el Lagrangiano de Zwanzig-Langevin obtenemos:



$$M_\alpha \ddot{R}_\alpha - \sum_i \left( c_{i\alpha} q_{i\alpha} - \frac{c_{i\alpha}^2}{m_{i\alpha} w_{i\alpha}^2} \dot{R}_\alpha \right) + \frac{\partial U([R_\alpha])}{\partial R_\alpha} = 0 \quad (2.56)$$

$$m_{i\alpha} \ddot{q}_{i\alpha} + m_{i\alpha} w_{i\alpha}^2 q_{i\alpha} - c_{i\alpha} R_\alpha = 0 \quad (2.57)$$

La primera ecuación (2.56) corresponde al movimiento de los átomos del ADN y la segunda (2.57) para las partículas del baño térmico. Estas ecuaciones están acopladas, es decir, hay dependencia de movimiento entre los distintos tipos de partículas.

Para desacoplar las ecuaciones se resuelve la ecuación para  $q_{i\alpha}$ <sup>20</sup> y se sustituye en la segunda ecuación,

$$q_{i\alpha} = \left[ q_{i\alpha}(0) - \frac{c_{i\alpha}}{m_{i\alpha} w_{i\alpha}^2} R_\alpha(0) \right] \cos(w_{i\alpha} t) + \frac{\dot{q}_{i\alpha}(0)}{w_{i\alpha}} \sin(w_{i\alpha} t) - \frac{c_{i\alpha}}{m_{i\alpha} w_{i\alpha}^2} R_\alpha(t) - \frac{c_{i\alpha}}{m_{i\alpha} w_{i\alpha}^2} \int_0^t \cos[w_{i\alpha}(t-t')] \dot{R}_\alpha(t') dt' \quad (2.58)$$

Se pueden definir ciertas cantidades que ayudan a simplificar la relación obtenida del paso anterior. La idea es incluir el efecto completo de las partículas del baño en dos términos de naturaleza estadística. La razón para esta simplificación es debido a que no nos interesa describir cada partícula del baño térmico individualmente, sino el comportamiento global. Para esto definimos la fuerza estocástica, comúnmente conocida como “ruido”<sup>21</sup>  $G_\alpha$  y el kernel disipativo  $K_\alpha$ , que es función de los coeficientes de acoplamiento  $c_{i\alpha}$ :

$$G_\alpha(t) = \sum_i c_{i\alpha} \left[ \frac{\dot{q}_{i\alpha}(0)}{w_{i\alpha}} \sin(w_{i\alpha} t) + \left[ q_{i\alpha}(0) - \frac{c_{i\alpha}}{m_{i\alpha} w_{i\alpha}^2} R_\alpha(0) \right] \cos(w_{i\alpha} t) \right], \quad (2.59)$$

$$K_\alpha(t-t') = M_\alpha \sum_i \frac{c_{i\alpha}^2}{m_{i\alpha} w_{i\alpha}^2} \cos[w_{i\alpha}(t-t')] \quad (2.60)$$

La fuerza estocástica  $G_\alpha$  representa las interacciones aleatorias del fluido sobre cada partícula dentro del reservorio térmico (en este caso, las del ADN) y corresponde al movimiento térmico que tendrán estas partículas, gracias al intercambio energético.

La función memoria lleva ese nombre, ya que incluye una correlación entre el tiempo presente  $t$  y el valor de la función en un tiempo pasado  $t'$ , al meter la dependencia de la diferencia  $(t-t')$ , nos ayuda a introducir la fricción que genera el fluido. De igual manera, se le conoce como Kernel disipativo, debido a que se utiliza como el Kernel de una integral sobre el tiempo.

Sustituyendo estas expresiones, llegamos a lo que se conoce como la ecuación generalizada de Langevin:

$$M_\alpha \ddot{R}_\alpha - \int_0^t K_\alpha(t-t') \dot{R}_\alpha(t') dt' - G_\alpha(t) + \frac{\partial U([R_\alpha])}{\partial R_\alpha} = 0 \quad (2.61)$$

<sup>20</sup>Es una ecuación diferencial inhomogénea. Se puede resolver por el método de funciones de Green o variación de parámetros.

<sup>21</sup>En relación con el ruido blanco, el cual lleva ese nombre en analogía con la luz blanca, está conformado por todas las frecuencias y misma potencia.

Podemos notar que esta ecuación es no-Markoviana, debido a la función memoria que aparece en el término de fricción. Zwanzig resuelve ese problema, al considerar que estamos en un baño térmico formado por un mar de osciladores con una cierta distribución de frecuencias  $g(w_\alpha)$ , la cual se tomará como continua. Ya que se consideran osciladores acoplados, se utiliza una distribución de frecuencias de Debye.

El resultado de la integral de la función memoria<sup>22</sup>, se obtiene de hacer la simplificación de tomar a la función memoria como pulsos rectangulares que aparecen a intervalos temporales regulares<sup>23</sup>, es decir, el movimiento de la partícula se da a intervalos temporales cortos debido a las fluctuaciones térmicas (aproximación de Delta de Dirac o respuesta de impulso). Lo cual es consistente con el movimiento rápido y aleatorio de los golpes estocásticos del baño térmico. De esta manera, tendremos un término que es proporcional a la magnitud de la velocidad y que se opone al movimiento de la partícula, como una fuerza de arrastre con coeficiente de viscosidad  $\gamma$  [77],

$$M_\alpha \gamma \frac{dR_\alpha(t)}{dt} = \int_0^t K_\alpha(t-t') \dot{R}_\alpha(t') dt'. \quad (2.62)$$

El término de fricción introduce energía y la fuerza aleatoria la quita en forma de energía cinética. Utilizando estas simplificaciones (de naturaleza estadística) podemos escribir la ecuación para las partículas en términos de  $K$  y  $G$ :

$$M_\alpha \frac{d^2 R_\alpha(t)}{dt^2} - G_\alpha(t) + \gamma M_\alpha \frac{dR_\alpha(t)}{dt} + \frac{\partial U([R_\alpha])}{\partial R_\alpha} = 0 \quad (2.63)$$

Esta es la ecuación de Langevin (obtenida desde un Lagrangiano), la cual tomará a la estructura de ADN como una partícula Browniana con estructura atómica, es decir, la molécula de ADN está en contacto con un baño térmico de naturaleza estadística.

Hay que notar que hasta el momento no se ha dicho nada respecto a como obtener el término estocástico  $G_\alpha$ . A pesar de estar tratando con términos aleatorios, debe haber consistencia con el medio de solvatación que se está eligiendo. La idea es determinar a  $G_\alpha$  con ayuda de una función de distribución bivariada de Chandrasekhar<sup>24</sup>, que depende de la posición y velocidad del átomo  $\alpha$ . Usar esto nos permite argumentar de mejor manera la suposición de tratar a cada partícula del ADN como inmersa en su propio baño térmico, ya que la función de distribución puede ser la misma (bajo las características específicas del solvente) y obtener en términos generales un solo baño térmico homogéneo para la molécula de ADN entera. [85, 86]

<sup>22</sup>Este tipo de integrales aparecen en las conocidas series de Volterra (similar a expansiones de Taylor, pero se toma en cuenta tiempos pasados), con aplicaciones en neurociencias, donde se relacionan inputs y outputs a todo tiempo con integrales de convolución. En nuestro caso tendríamos el término de orden 1.

<sup>23</sup>Similar a la respuesta de un capacitor a un conjunto de pulsos eléctricos rectangulares.

<sup>24</sup>Subramanyan Chandrasekhar premio Nobel de Física 1983, por sus contribuciones en astrofísica.

### 2.7.2. Integrador de Verlet-Langevin

Debido a que consideramos a todas las partículas que conforman la estructura de ADN, como partículas Langevinas, es necesario implementar un método numérico para integrar la ecuación de Langevin (2.63).

El algoritmo comúnmente utilizado, es uno que tiene como límite (límite de fricción cero) las expresiones clásicas de *Velocity-Verlet*. La adaptación del algoritmo se debe a que estamos trabajando ahora con una ecuación diferencial estocástica. Se puede ver que la exactitud del método depende de las fuerzas sistemáticas (restringe el paso de tiempo). La complicación está en que al intentar integrar la ecuación de Langevin, aparece más de una integral con propiedades estocásticas. Por lo que es necesario caracterizar a las dos variables aleatorias correlacionadas<sup>25</sup>, por medio de una distribución Gaussiana bivariada, como se ha mencionado anteriormente.

La ecuación de Langevin queda escrita entonces de la siguiente manera:

$$M_\alpha \dot{\mathbf{v}}_\alpha(t) = \mathbf{G}_\alpha(t) - \gamma M_\alpha \mathbf{v}_\alpha(t) + \frac{\partial U([R_\alpha])}{\partial R_\alpha} \quad (2.64)$$

El término en azul es la llamada fuerza sistemática (puede ser cualquier tipo de campo externo). Lo escribiremos por el momento como  $\mathbf{f}_\alpha(t)$ .

Podemos despejar la velocidad y multiplicar por un factor  $e^{-\gamma t}$ , y al integrar nos queda:

$$\int_{t_n}^{t_n+\Delta t} \frac{d}{dt} e^{\gamma(t-t_n)} v dt = \int_{t_n}^{t_n+\Delta t} e^{\gamma(t-t_n)} \frac{f(t) + G(t)}{M} dt. \quad (2.65)$$

El lado izquierdo basta con ser evaluado en los límites de integración. Obtenemos para la velocidad:

$$v(t) = v(t_n) e^{-\gamma(t-t_n)} + e^{-\gamma(t-t_n)} \int_{t_n}^{t_n+\Delta t} e^{\gamma(t-t_n)} \frac{f(t) + G(t)}{M} dt. \quad (2.66)$$

Para  $f$  (fuerzas sistemáticas) podemos expandir en serie de potencias (suponemos intervalo de tiempo pequeño, de ahí que la fuerza sistemática nos restringe la dinámica):

$$f(t) = f(t_n) + \dot{f}(t_n)(t - t_n) + \mathcal{O}(t - t_n)^2. \quad (2.67)$$

Nos quedamos hasta segundo orden para la fuerza e integramos. Para la integral del término estocástico, definimos (contribución estocástica a la velocidad):

$$\mathcal{V}_\alpha \Delta t = \frac{1}{M} e^{-\gamma \Delta t} \int_{t_n}^{t_n+\Delta t} e^{\gamma(t-t_n)} G_\alpha(t) dt. \quad (2.68)$$

Se puede simplificar aún más, al utilizar los siguientes coeficientes [87]:

<sup>25</sup>Correlacionadas en el sentido que, para una velocidad grande las partículas se moverán a distancias mayores y viceversa, debe haber consistencia.

$$\begin{aligned}
 c_0 &= e^{-\gamma\Delta t} \\
 c_1 &= (\gamma\Delta t)^{-1}(1 - c_0) \\
 c_2 &= (\gamma\Delta t)^{-1}(1 - c_1) \\
 c_3 &= (\gamma\Delta t)^{-1}\left(\frac{1}{2} - c_2\right)
 \end{aligned}$$

Finalmente tenemos:

$$v(t_n + \Delta t) = c_0 v(t_n) + \frac{1}{M_\alpha} c_1 f(t_n) \Delta t + c_2 \frac{1}{M_\alpha} \dot{f}(t_n) (\Delta t)^2 + \mathcal{V}_\alpha \quad (2.69)$$

De manera similar, integrando una segunda vez sobre un intervalo  $\Delta t = t_{n+1} - t_n$  podemos obtener  $x(t_n + \Delta t)$ , sustituyendo lo encontrado para  $v(t)$  llegamos a la siguiente expresión<sup>26</sup>:

$$\begin{aligned}
 x(t_n + \Delta t) &= x(t_n)[1 + e^{-\gamma\Delta t}] - x(t_n - \Delta t)e^{-\gamma\Delta t} + \frac{1}{M} f(t_n) (\Delta t)^2 (\gamma\Delta t)^{-1} [1 - e^{-\gamma\Delta t}] \\
 &+ \frac{1}{M} \dot{f}(t_n) (\Delta t)^3 (\gamma\Delta t)^{-2} \left[ \frac{1}{2} \gamma\Delta t [1 + e^{-\gamma\Delta t}] - [1 - e^{-\gamma\Delta t}] \right] \\
 &+ X_n(\Delta t) + e^{-\gamma\Delta t} X_n(-\Delta t) + \mathcal{O}[(\Delta t)^4],
 \end{aligned} \quad (2.70)$$

donde nuevamente se define la contribución estocástica a la posición:

$$\mathcal{X}_\alpha(\Delta t) = \frac{1}{M_\alpha} \int_{t_n}^{t_n + \Delta t} e^{-\gamma(t-t_n)} \int_{t_n}^t e^{\gamma(t'-t_n)} G_\alpha(t') dt' dt. \quad (2.71)$$

Al escribir en términos de los coeficientes previamente definidos, finalmente obtenemos:

$$x(t_n + \Delta t) = x_\alpha(t_n) + c_1 v_\alpha(t_n) \Delta t + c_2 \frac{1}{M_\alpha} f(t) (\Delta t)^2 + \mathcal{X}_\alpha. \quad (2.72)$$

Entonces este método nos entrega las siguientes ecuaciones para las velocidades y posiciones de las partículas del ADN, dentro de un reservorio térmico de Langevin,

$$v(t_n + \Delta t) = c_0 v(t_n) + \frac{1}{M_\alpha} c_1 f(t_n) \Delta t + c_2 \frac{1}{M_\alpha} \dot{f}(t_n) (\Delta t)^2 + \mathcal{V}_\alpha, \quad (2.73)$$

$$x(t_n + \Delta t) = x_\alpha(t_n) + c_1 v_\alpha(t_n) \Delta t + c_2 \frac{1}{M_\alpha} f(t) (\Delta t)^2 + \mathcal{X}_\alpha. \quad (2.74)$$

Podemos notar que en el límite  $\gamma \rightarrow 0$  obtenemos las ecuaciones típicas del método de integración de Verlet. Las contribuciones estocásticas a las posiciones y velocidades son integrales del término estocástico, cuyo comportamiento desconocemos por definición. Lo único que si sabemos es que deben estar en sintonía con el medio que conforma el baño térmico, es decir, la ecuación de Langevin. Estos términos se obtendrán de una función de distribución de probabilidad bivariada [85]:

<sup>26</sup>La deducción completa se puede ver en el artículo de W.F. Van Gunsteren H.J.C. Berendsen [88].

$$\mathcal{W}(r, v, t; r_0, v_0, t_0) = \frac{1}{(2\pi)^3(ac - b^2)^{3/2}} e^{-(aR^2 - 2b\vec{R}\cdot\vec{S} + cS^2)/2(ac - b^2)} \quad (2.75)$$

donde los coeficientes están dados por:

$$a = \langle V_n^2(\Delta t) \rangle \quad (2.76)$$

$$b = \langle X_n(\Delta t + V_n(\Delta t)) \rangle \quad (2.77)$$

$$c = \langle X_n^2(\Delta t) \rangle \quad (2.78)$$

Haciendo el cambio de variable  $Y = V_n - \frac{b}{c}X_n$ , obtenemos:

$$\mathcal{W}(r, v, t; r_0, v_0, t_0) = \frac{1}{2\pi\sqrt{ac - b^2}} e^{-X_n^2/2c} e^{-Y_n^2/2(a - b^2/c)} \quad (2.79)$$

Como podemos ver, es como si tuviéramos dos gaussianas (en  $X$  y  $Y$ , respectivamente). Debido a que tenemos una distribución de probabilidad condicional (posiciones y velocidades correlacionadas) podemos integrar sobre todos los posibles valores de  $Y$  y tenemos:

$$\mathcal{W}(X_n(\Delta t)) = \frac{1}{2\pi\sqrt{ac - b^2}} e^{-X_n^2/2c} \int_{-\infty}^{\infty} e^{-Y_n^2/2(a - b^2/c)} dY, \quad (2.80)$$

la probabilidad condicional, dado que el valor de  $X_n$  sería  $\mathcal{W}(V_n(\Delta t)|X_n(\Delta t))$ .

Para poder implementar esta metodología de manera computacional, es necesario obtener los valores para la contribución estocástica a la posición y velocidad ( $X_n$  y  $V_n$ ). Esto se hace al utilizar una distribución de números aleatorios que cumplan reproducir una gaussiana de ancho 1 ( $\sigma = 1$ ) y centrada en el origen. El generador de números aleatorios nos entrega los valores para las contribuciones estocásticas y poder calcular la distribución de probabilidad  $\mathcal{W}$ , la cual tiene integrada, las fuerzas específicas de las partículas (de la ecuación de Langevin), la temperatura ( $T$ ), viscosidad ( $\gamma$ ) y la masa de las partículas.

Con esto se logra cambiar el sistema de un baño térmico de miles de partículas a la descripción del baño utilizando únicamente distribuciones de probabilidad para las posiciones y velocidades

La manera formal en la que se especifica la función de distribución desde una ecuación de Fokker-Plank<sup>27</sup> se puede encontrar en el artículo original de Chandrasekhar [85].

<sup>27</sup>Básicamente una ecuación generalizada de difusión.

El proceso que sigue este método es el siguiente:

- Conocemos los valores  $x(t_n)$ ,  $x(t_{n-1})$ ,  $\mathcal{X}_{n-1}(\Delta t)$  y  $f(t_{n-1})$
- Obtenemos  $f(t_n)$  a partir de los resultados de DFT
- Calculamos  $\dot{f}(t_n) = \frac{[f(t_n) - f(t_{n-1})]}{\Delta t}$
- Usando la distribución de probabilidad podemos calcular  $\mathcal{X}_n(\Delta t)$  y obtener las posiciones  $x(t_{n+1})$
- Similarmente obtenemos las velocidades  $v(t_n)$  y repetimos el procedimiento

Como se ha mencionado anteriormente, debido a la imposibilidad de resolver la ecuación de Schrödinger para muchos núcleos y muchos electrones, ciertas aproximaciones deben ser utilizadas para lograr obtener resultados útiles. Estas aproximaciones se deben a nuestra falta de herramientas tanto matemáticas como computacionales.

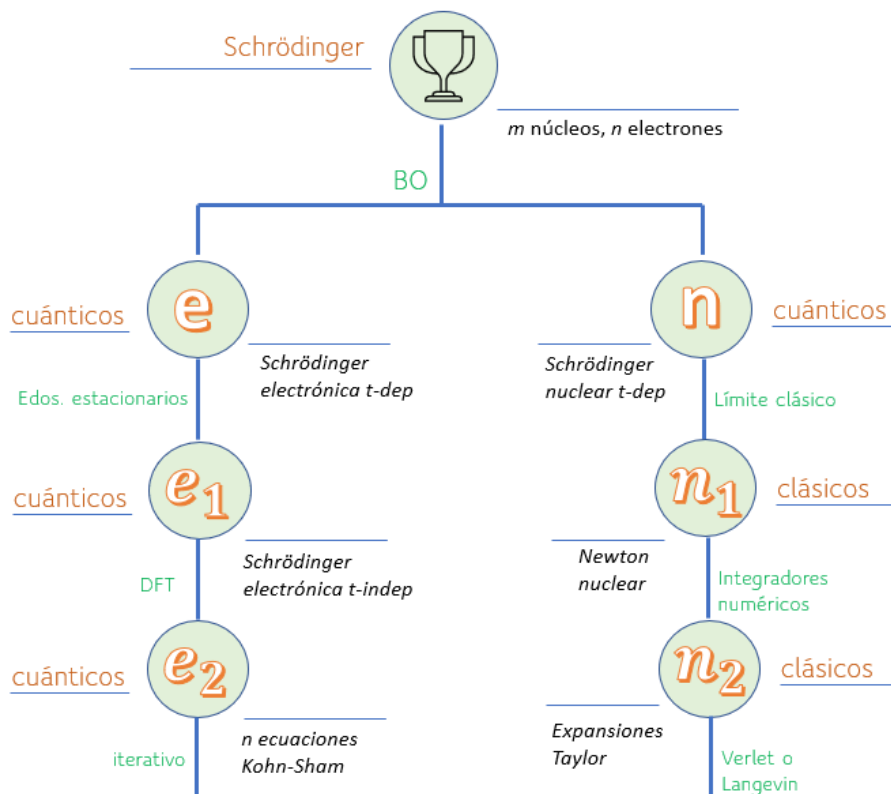


Figura 2.2: Niveles de teoría utilizados en este trabajo.

## 2.8. Fuerzas externas sobre ADN (steering forces)

Para estudiar los efectos de la separación de las hebras del ADN, simulando la acción de enzimas en procesos biológicos, se utilizaron fuerzas externas dirigidas de tipo armónicas. Esto se logró al colocar dos resortes en átomos específicos de cada par de bases y además, con un movimiento sobre una línea de acción bien definida. La elección de los átomos se hizo de tal manera que se estiraran las bases sobre una misma línea de acción, para dar uniformidad al movimiento y minimizar los efectos de torsión de las bases. Para el caso del par de bases AT, el cual cuenta con dos puentes de hidrógeno, se seleccionaron átomos de carbono de cada base, cuya línea de unión pasa aproximadamente por el centro de los dos enlaces (ver fig. 2.3). Similarmente, para el par de bases GC el cual tiene tres puentes de hidrógeno, la línea que une los átomos de carbono seleccionados está posicionada con el enlace central (ver fig. 2.3).

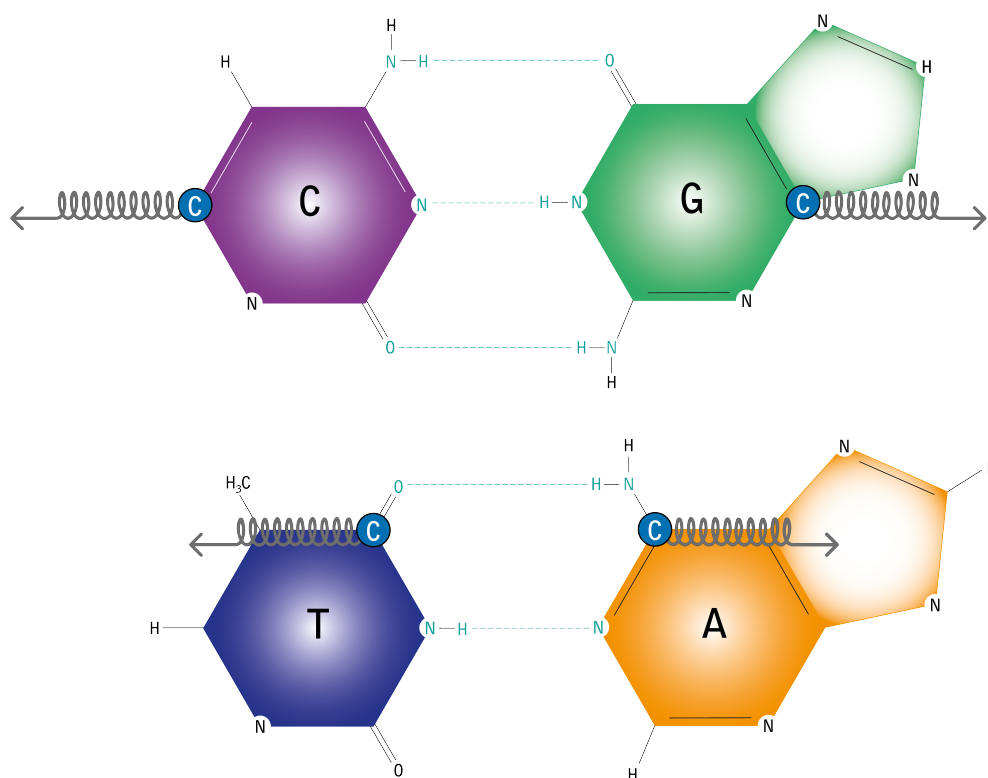


Figura 2.3: Posicionamiento de los resortes virtuales sobre cada una de las bases nucleicas.

El proceso que llamaremos "*disociación de ADN*", donde ocurre la generación de la línea de acción, colocación de resortes y estiramiento secuencial de par de bases, es el siguiente:

- **Línea de acción:** Una vez seleccionados los átomos de carbono (uno en cada base) que experimentarán las fuerzas armónicas, el código computacional genera la línea que los une. Conocemos las posiciones de los átomos para cada paso de tiempo  $x(C_{i,j})$ ,  $y(C_{i,j})$ ,  $z(C_{i,j})$ , entonces:

$$l = \sqrt{[x(C_i) - x(C_j)]^2 + [y(C_i) - y(C_j)]^2 + [z(C_i) - z(C_j)]^2} \quad (2.81)$$

Posteriormente, se obtiene el vector unitario de la línea, para cada una de sus componentes ( $l_x, l_y, l_z$ ):

$$\hat{l}_1 = \frac{l_x}{l}; \quad \hat{l}_2 = \frac{l_y}{l}; \quad \hat{l}_3 = \frac{l_z}{l} \quad (2.82)$$

Finalmente, dada la magnitud deseada para la velocidad con la que se moverá el resorte  $v$ , podemos obtener las componentes de la velocidad en la dirección de la línea calculada previamente, al multiplicar por el vector unitario:

$$v_x = v\hat{l}_1; \quad v_y = v\hat{l}_2; \quad v_z = v\hat{l}_3 \quad (2.83)$$

■ Orígenes de resortes:

Una vez que el código obtuvo las líneas de acción sobre las cuales se moverán los resortes, se procede a escribir los orígenes de los resortes en movimiento como:

$$O_x(C_i) = x(C_i) - v_x t \quad O_x(C_j) = x(C_j) + v_x t \quad (2.84)$$

$$O_y(C_i) = y(C_i) - v_y t \quad O_y(C_j) = y(C_j) + v_y t \quad (2.85)$$

$$O_z(C_i) = z(C_i) - v_z t \quad O_z(C_j) = z(C_j) + v_z t \quad (2.86)$$

Podemos notar que para el inicio de la dinámica ( $t = 0$ ), el origen del resorte está sobre cada carbono en su posición inicial, como debe ser. Además, los signos se escogieron con cuidado, de tal manera que los resortes jalen a las bases hacia la izquierda y derecha según la base nucleica a tratar. Los resortes van incrementando su longitud con el paso del tiempo ( $\Delta t$ ).

■ Potencial y fuerzas externas:

Las fuerzas armónicas actúan sobre el  $i$ -ésimo y  $j$ -ésimo carbono a lo largo de la línea que los conecta. Las fuerzas aplicadas son similares a las que se aplicarían al jalar una cuerda por los dos extremos, por lo que basta con tomar en cuenta una de las fuerzas (la positiva, por ejemplo). Las componentes para el potencial de resorte son (se utilizó la misma constante de resorte  $k$ ):

$$V_x(t; C_i) = k \frac{[x(t; C_i) - O_x(t; C_i)]^2}{2} \quad (2.87)$$

$$V_y(t; C_i) = k \frac{[y(t; C_i) - O_y(t; C_i)]^2}{2} \quad (2.88)$$

$$V_z(t; C_i) = k \frac{[z(t; C_i) - O_z(t; C_i)]^2}{2}. \quad (2.89)$$



Similarmente, las componentes para la fuerza aplicada a cada par de bases nucleicas son:

$$F_x(C_i) = -k[x(t; C_i) - O_x(t; C_i)] \quad (2.90)$$

$$F_y(C_i) = -k[y(t; C_i) - O_y(t; C_i)] \quad (2.91)$$

$$F_z(C_i) = -k[z(t; C_i) - O_z(t; C_i)], \quad (2.92)$$

donde la fuerza total externa por par de bases en cada paso del tiempo esta dada por:

$$F_T^{ext} = \sqrt{(F_x)^2 + (F_y)^2 + (F_z)^2} \quad (2.93)$$

Nuevamente podemos notar que en  $t = 0$ , las fuerzas externas se hacen cero, como debe ser.

■ Movimiento secuencial entre par de bases:

Hay que recordar que se busca abrir la doble hélice del ADN de manera secuencial, es decir, par de base por par de base (justo como lo hace la helicasa en procesos biológicos). Esto se realizó al tomar en cuenta un criterio geométrico, similar al utilizado en otros trabajos [22, 24].

1. Se identifica el número de par de bases de la estructura de ADN a estudiar. Discernimos entre par de bases AT y GC e iniciamos el proceso “*disociación de ADN*” para el primer par de bases.
2. El código calcula la posición del CM de cada base nucleica para la estructura de ADN completa. Si los CM de las bases que constituyen el primer par de bases llega a una distancia de separación límite, suponemos que se interrumpió la interacción de los puentes de hidrógeno.
3. Cuando el punto anterior ocurre, el código quita los resortes de los carbonos del primer par de bases y escoge 2 carbonos del segundo par de bases. El proceso se repite hasta llegar al último par de bases de la estructura.
4. Finalmente, cuando los CM para el último par de bases llega a la distancia de separación límite, el código elimina los resortes. Se obtiene una estructura de ADN donde todos los puentes de hidrógeno se rompieron y la doble hélice se separó.

Se puede notar que, al pasar los resortes al siguiente par de bases, se eliminan los resortes del par de bases anterior. Esto se puede hacer, ya que no se consideró un ambiente circundante de aguas explícitas, las bases nucleicas no se recombinan. El mismo argumento funciona si se consideran aguas explícitas, pero hay que garantizar que la distancia entre las bases nucleicas es lo suficiente para mantenerlas separadas.

Este método permite introducir el número de resortes que se deseen y aplicados a los átomos que se necesiten, lo cual hace que sea muy flexible y adaptable al experimento que sea y al estudio específico a realizar.

## 2.9. Dinámica Molecular

Ahora que ya tenemos las ecuaciones y métodos de solución para los distintos tipos de partículas, podemos englobar el tratamiento específico que se realizó en este trabajo. Tenemos estructuras de ADN inmersas en un reservorio térmico y bajo la influencia de fuerzas externas armónicas. El tratamiento electrónico se hace a un nivel cuántico con DFT, el nuclear a un nivel de mecánica analítica (clásico) y el termostato con ayuda de una ecuación de Zwanzig-Langevin, la cual emplea términos estocásticos. En nuestro caso, se toman a todos los átomos del sistema como átomos de Langevin, ya que el ADN completo está dentro del baño térmico y no estamos usando algún tipo de contenedor.

La ecuación de Langevin queda escrita entonces de la siguiente manera:

$$M_\alpha \frac{d^2 R_\alpha(t)}{dt^2} = G_\alpha(t) - \gamma M_\alpha \frac{dR_\alpha(t)}{dt} + \frac{\partial U([R_\alpha])}{\partial R_\alpha}, \quad (2.94)$$

donde el último término esta dado por dos contribuciones en nuestro caso:

$$\frac{\partial U([R_\alpha])}{\partial R_\alpha} = -\frac{\partial V_{DFT}(R_\alpha)}{\partial R_\alpha} + F_{steering}. \quad (2.95)$$

El primer término del lado derecho corresponde a la energía electrónica del estado base (se calcula con DFT) y representa una energía potencial que afecta a los núcleos. El segundo término es la fuerza externa armónica que actúa sobre cada par de bases nucleicas, con la cual se produce la separación de las cadenas del ADN y está dada por la ecuación (2.93).

La agitación térmica que experimentarán los núcleos se introduce por medio de  $G_\alpha$  y la viscosidad del fluido con ayuda de la función  $K_\alpha$  por medio del coeficiente  $\gamma$ . El intercambio de energía lleva eventualmente a la termalización del sistema entero (ya tenemos una temperatura constante<sup>28</sup>). Las velocidades iniciales que toman los átomos se inicializan por medio de una distribución de velocidades (Maxwell-Boltzman) compatible con las características macroscópicas del fluido [77, 85]. La forma de resolver este tipo de ecuaciones es con un método Velocity-Verlet, como se vio previamente, adaptado para la ecuación de Langevin (ver [77]).

---

<sup>28</sup>La temperatura estadística es igual al promedio de temperatura mecánica, dado por las energías cinéticas de los átomos. Cada observable que nos interesa se expresa en términos de posiciones y velocidades de las partículas.

El procedimiento general que sigue una dinámica molecular sobre estructuras de ADN (son las posiciones nucleares las que se dejan evolucionar con el tiempo), es el siguiente [66, 68, 76]:

1. Seleccionamos los parámetros iniciales de la dinámica:
  - Número de partículas
  - Temperatura inicial
  - Paso de tiempo  $\Delta t$ <sup>29</sup>
2. Inicializamos el sistema (Equilibración):
  - Posiciones iniciales de los átomos del ADN, comúnmente obtenidos de datos experimentales ( $R_\alpha$ ). No necesariamente una estructura optimizada.
  - Velocidades iniciales, con ayuda de una distribución de velocidades que asigne valores aleatorios, de acuerdo a una temperatura seleccionada. Se llega a la termalización en un cierto número de pasos de tiempo.
3. Se calculan las fuerzas sobre todos los átomos del sistema (en este caso, los átomos particulares de los cuales se estirarán las bases nucleicas, experimentarán una fuerza extra debido al resorte).

Recordemos que las fuerzas sobre los núcleos vienen de  $\nabla E_{elec}(R_\alpha)$  (donde  $E_{elec}$  esta dada por la energía cinética de los electrones y un funcional de la energía que incluye la correlación y el intercambio).
4. Se integran las ecuaciones de movimiento (numéricamente). En nuestro caso, estamos tratando con núcleos clásicos sumergidos en un baño térmico, por lo que se resuelven las ecuaciones de Langevin con ayuda de un integrador tipo Velocity-Verlet adaptado.
5. Se repite el proceso hasta alcanzar la evolución temporal deseada. Se construye la trayectoria o dinámica del sistema.

---

<sup>29</sup>Un paso de tiempo muy grande podría causar que los átomos se mueven muy lejos de la trayectoria y no simularía de buena manera el movimiento. Pasos de tiempo muy cortos nos llevaría a tener que realizar demasiadas iteraciones.

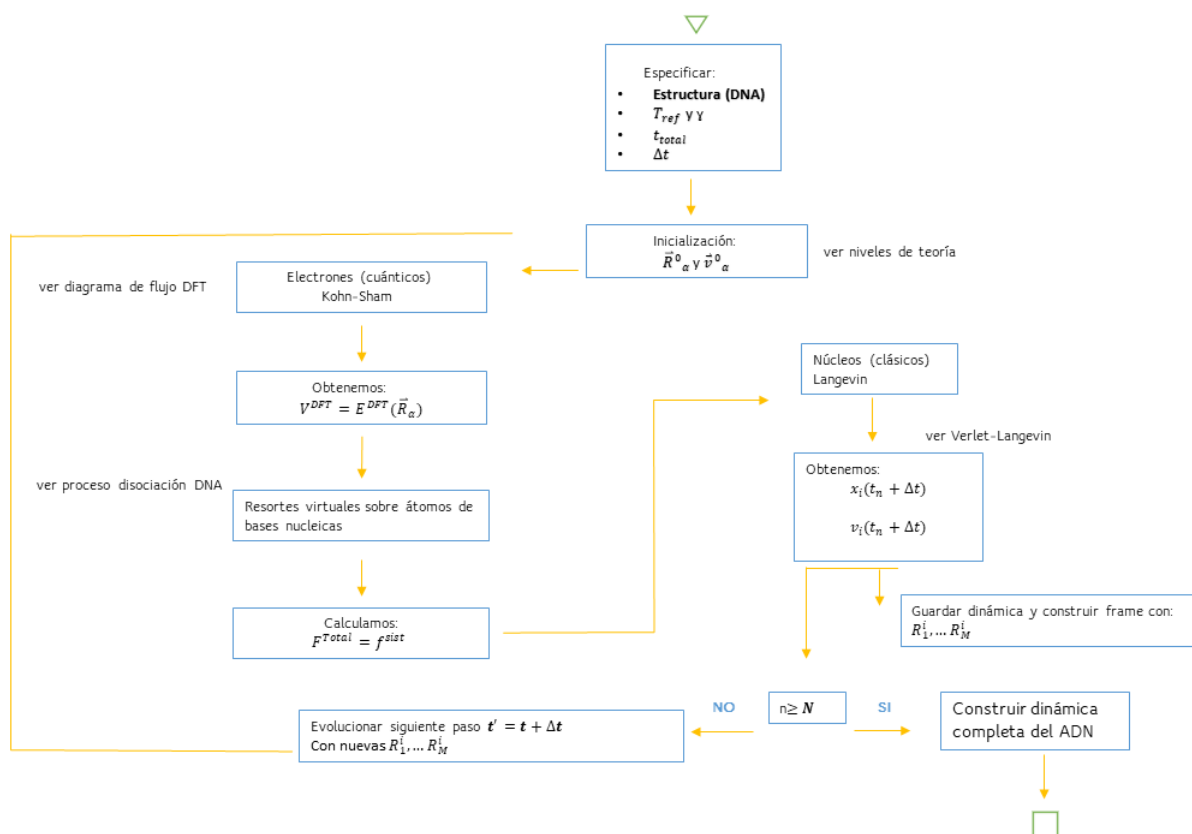


Figura 2.4: Diagrama de flujo para la implementación de la dinámica molecular cuántica, con uso de la teoría de Langevin.

## Capítulo 3

# Resultados

*"The expanding eyes of man  
behold the depths of wondrous  
worlds."*

---

*William Blake*

En este capítulo se presentan los resultados obtenidos al trabajar con estructuras de ADN con distintas secuencias nucleicas bajo fuerzas armónicas externas. La naturaleza secuencial de las fuerzas de separación trata de simular el comportamiento real de las helicasas en procesos de replicación celular.

Con este tipo de procedimientos es posible calcular fuerzas de disociación para la separación de pares de base, tomando en cuenta la columna vertebral nucleica e interacciones de apilamiento. Debido al elevado costo computacional que requieren las simulaciones de primeros principios a niveles cuánticos, el trabajo se enfoca en estructuras cortas de ADN, de 5 pares de base de largo. Esto no impone restricciones sobre los resultados, ya que múltiples estudios han mostrado la presencia de una longitud máxima a partir de la cual las biomoléculas dejan de experimentar la presencia de las fuerzas externas. Por lo que, los efectos sobre estructuras cortas, describen de buena manera el comportamiento de segmentos más grandes.

En la primera parte se encuentran las descripciones de las moléculas que se analizaron y su importancia en investigaciones biológicas. Posteriormente, se reportan equilibraciones térmicas y energéticas de las estructuras, así como las variaciones de energía y fluctuaciones térmicas durante el proceso de disociación. Además, se describen a detalle las razones de separación de la doble hélice, junto con las fuerzas disociativas y distribuciones de cargas. Finalmente, se analizan posibles dependencias en las secuencias para las variaciones de las fuerzas.

### 3.1. Estructuras de ADN y detalles computacionales

Las geometrías iniciales se obtuvieron de distintas fuentes bibliográficas basadas en datos experimentales y se realizó el proceso de optimización geométrica.

Los detalles computacionales de la dinámica son los mismos para las distintas estructuras, únicamente cambia el número de átomos debido a las diferencias en la secuencia.

Cuadro 3.1: Detalles computacionales

Paso del tiempo ( $\Delta t$ )	1 fs
Temperatura de equilibrio ( $T_{equ}$ )	300 K
Coefficiente de fricción ( $\gamma$ )	$4 ps^{-1}$
Número de átomos ( $N$ )	(315,314)
Constante de resorte ( $k$ )	0.01 N/cm
Velocidad de resorte ( $v_{spring}$ )	0.0020 Å/fs

<sup>1</sup> Tabla de detalles computacionales para la dinámica molecular sobre estructuras de ADN.

Debido a las condiciones celulares en las que se encuentra el ADN, consideramos una temperatura de equilibrio ambiente (300 K) y un coeficiente de fricción para un medio de solvatación de agua. Los valores para la constante y velocidad de los resortes atados a los pares de base se eligieron de tal forma que la separación se realiza de manera delicada para minimizar efectos de movimientos de torsión no deseados. Como se mencionó con anterioridad, la resolución numérica de las ecuaciones de Langevin dependen fuertemente de tener un paso de tiempo pequeño. Se empleó el software de química computacional *TeraChem* para los cálculos electrónicos cuánticos, a un nivel de teoría *B3LYP/6-31G\**. Este software permite la utilización de tarjetas gráficas (GPU's), lo cual ayuda a mejorar el tiempo y costo computacional [89, 90]. La dinámica molecular (con reservorio de Langevin) y aplicación de las fuerzas armónicas sobre los pares de bases nucleicas se realizó con un código computacional desarrollado en Fortran. Los cálculos se realizaron con la ayuda de la supercomputadora *MIZTLI* de la UNAM, la cual cuenta con dos GPUs *Tesla V100*<sup>1</sup> y la computadora *Holiday* del *Instituto de Física (UNAM)* que cuenta con 1 V100.

<sup>1</sup>Una tarjeta gráfica NVIDIA V100, tiene una equivalencia de aproximadamente 32 GPUs.

### 3.1.1. Estructura (T-T-G-C-G)

Para la primera estructura de ADN se utilizó la secuencia: T-T-G-C-G. Una característica importante a notar, es la presencia de la famosa “caja TATA”<sup>2</sup>. Es una secuencia de ADN ubicua, que se encuentra en sitios promotores de los genes. Es decir, se encuentra en sitios donde las proteínas se adhieren al ADN para comenzar el proceso de transcripción. Se sabe que distintas enfermedades son derivadas por defectos en este tipo de secuencia, lo que genera deficiencias en el proceso de transcripción [91].

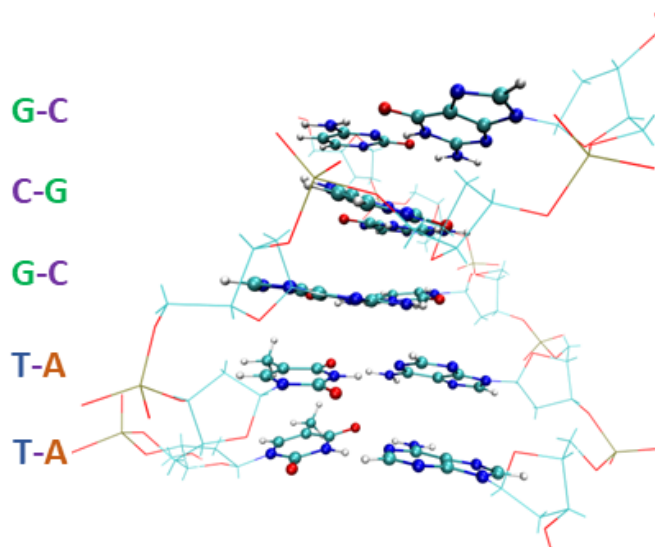


Figura 3.1: Primera estructura de ADN cuenta con 315 átomos y 8 grupos fosfatos en total.

<sup>2</sup>También conocida como caja de *Golberg-Hogness* debido a sus descubridores.

Cuadro 3.2: Valores iniciales de distancia entre CM<sup>1</sup>

Par de bases	d-CM (Å)	d-CM <sub>teo</sub> <sup>2</sup>
1bp (T-A)	5.585	6.0
2bp (T-A)	6.250	6.0
3bp (G-C)	5.639	5.6
4bp (C-G)	6.138	5.6
5bp (G-C)	5.377	5.6

<sup>1</sup> Tabla de valores iniciales de separación entre centros de masa (AT,GC) con la estructura termalizada.

<sup>2</sup> Valores teóricos comparativos obtenidos de [92]. Valores realizados a nivel de Hartree-Fock/6-31G\*\*.

Los valores iniciales para la separación entre centros de masa para cada una de las bases (AT,GC) están dentro de valores encontrados en trabajos previos, las pequeñas diferencias pueden ser debido a que en este trabajo se considera una estructura entera de ADN, los efectos de apilamiento afectan estas distancias iniciales.

La distancia de separación entre pares de base seleccionada fue de 8.2 Å, esta distancia nos permite decir con seguridad que la interacción de puentes de hidrógeno fue terminada (distancia máxima entre átomos de enlaces de hidrógeno es de 3.5Å [24, 25]).

### 3.1.2. Estructura (T-C-G-C-G)

En ciertas regiones del genoma humano resulta común encontrarse con las llamadas islas  $G_pC$ . Es decir, regiones donde nucleótidos de citosina son seguidos por nucleótidos de guanina y separados por un grupo fosfato. De ahí la importancia de analizar el proceso de separación bajo fuerzas externas sobre estructuras que contengan repetición de bases G-C.



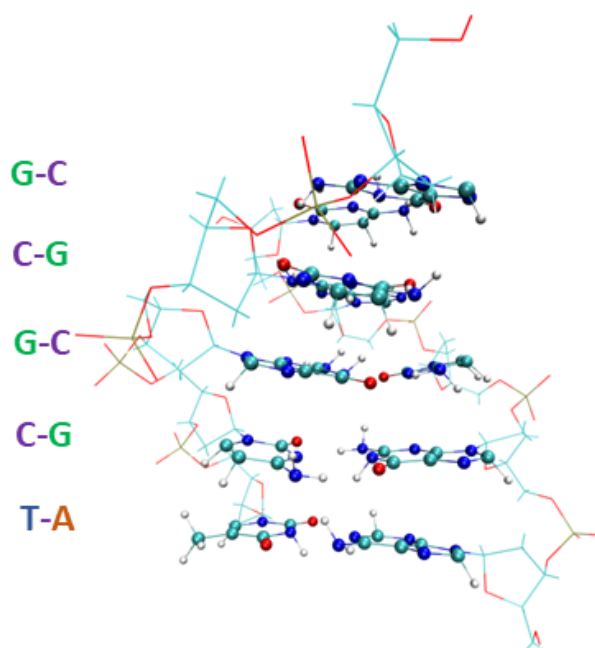


Figura 3.2: Segunda estructura de ADN cuenta con 314 átomos y 8 grupos fosfatos en total.

Los valores iniciales para la separación entre centros de masa (AT,GC) están dentro de valores encontrados en trabajos previos, las pequeñas diferencias pueden ser debido a que en este trabajo se considera una estructura entera de *ADN*, los efectos de apilamiento afectan estas distancias iniciales:

Cuadro 3.3: Valores iniciales de distancia entre CM<sup>1</sup>

Par de bases	d-CM (Å°)	d-CM <sub>teo</sub> <sup>2</sup>
1bp (T-A)	5.922	6.0
2bp (C-G)	5.472	5.6
3bp (G-C)	5.687	5.6
4bp (C-G)	5.473	5.6
5bp (G-C)	5.415	5.6

<sup>1</sup> Tabla de valores iniciales de separación entre centros de masa (AT,GC) con la estructura segunda termalizada.

<sup>2</sup> Valores teóricos comparativos obtenidos de [92]. Valores realizados a nivel de Hartree-Fock/6-31G\*\*.

### 3.2. Equilibración térmica y energética de las estructuras de ADN

Para garantizar que la dinámica se haga de manera correcta, es necesario realizar la separación de la doble hélice sobre la estructura en la temperatura de equilibrio. Se utilizó un reservorio térmico de Langevin, el cual está constituido por términos estocásticos, por lo que la equilibración se alcanza cuando la temperatura mecánica y la temperatura estadística son iguales:

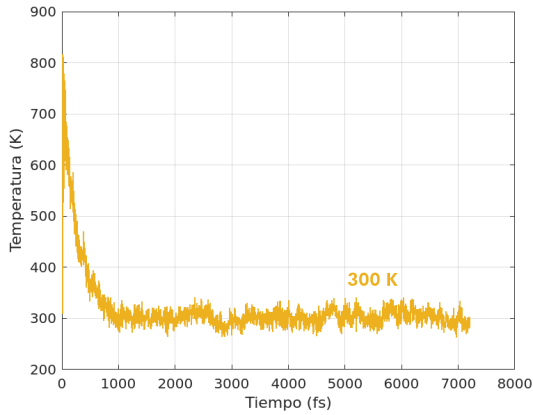
$$T_{mec} = T_{stat}, \quad (3.1)$$

donde la temperatura mecánica, es el promedio de la energía cinética de las partículas:

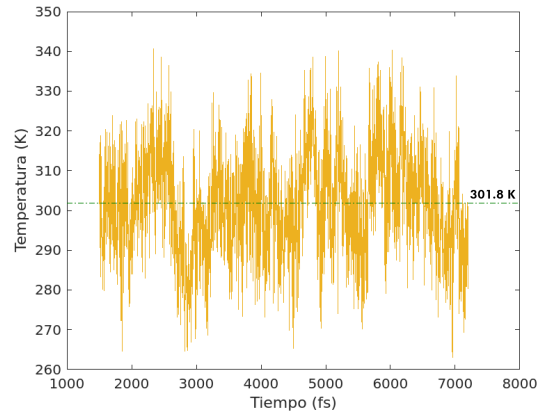
$$T_{mec} = \frac{2}{3Nk_B} \sum_{i=1}^{3N} \frac{1}{2} m_i \langle v_i^2 \rangle \quad (3.2)$$

**Estructura 1:**

Para la primera estructura de ADN (secuencia: T-T-G-C-G) la equilibración se hizo para 7000 fs, alcanzando un promedio de  $T = 301.8$  K.



(a) Termalización estructura 1 ADN (secuencia: T-T-G-C-G) y temperatura,  $T=300$  K.



(b) Promedio de temperatura alcanzado después de los 1500 fs (zona de equilibración).

Figura 3.3

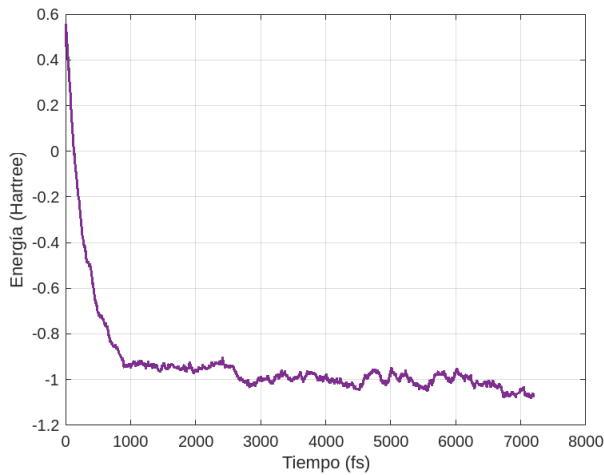


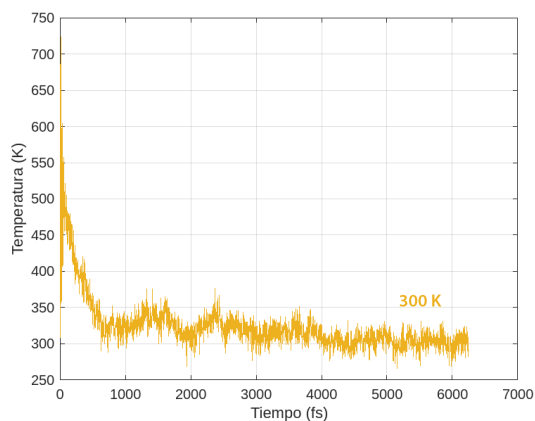
Figura 3.4: Equilibración energética del sistema.

Podemos notar que es hasta después de los 1000 fs que el sistema comienza a acercarse a la temperatura seleccionada.

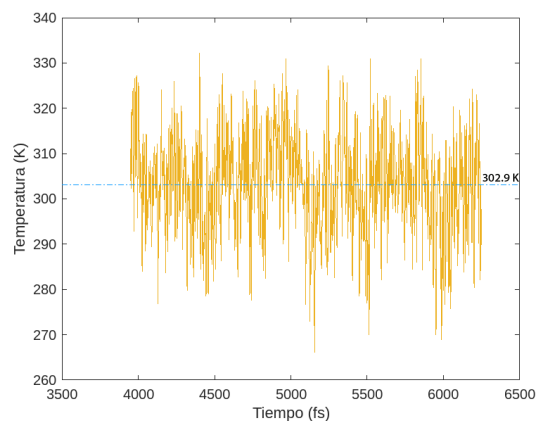
Similarmente, se muestra la gráfica de la equilibración de la energía, con un promedio de  $E_{prom} = -1.0121$  H.

**Estructura 2:**

Para la **segunda estructura** de ADN (secuencia: T-C-G-C-G) la equilibración se hizo para 6000 fs, alcanzando un promedio de  $T = 302.9$  K.



(a) Termalización estructura 2 ADN (secuencia: T-C-G-C-G) y temperatura,  $T=300$  K.



(b) Promedio de temperatura alcanzado después de los 3500 fs (zona de equilibración).

Figura 3.5

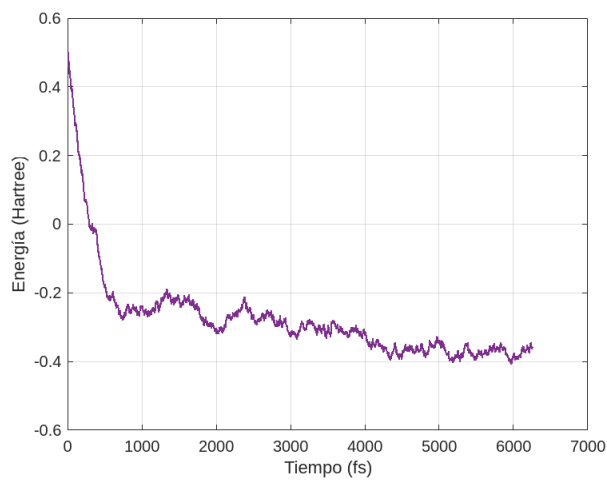


Figura 3.6: Equilibración energética del sistema.

Similarmente, se muestra la gráfica de la equilibración de la energía, con un promedio de  $E_{prom} = -0.3706$  H.

### 3.3. Separación de ds-DNA

Una de las características más importantes a analizar de la molécula de ADN es la flexibilidad de las bases nucleicas durante los procesos de disociación. La naturaleza de los puentes de hidrógeno que mantienen unidos los pares de base (AT,GC) permiten la apertura de las hélices por parte de las helicasas.

Las helicasas tiene como fuente de energía procesos de hidrólisis de ATP; sigue siendo tema de investigación si esta energía es suficiente para la apertura de un solo par de bases o varios pares de base a la vez [38].

Podemos ver a los pares de base como una especie de material dúctil, donde existe una fuerza límite en la cual ocurrirá la ruptura, es decir, se acabaría la interacción de los puentes de hidrógeno. En este caso, los resortes generan la fuerza que va incrementando con el paso de tiempo, hasta llegar al punto en que la fuerza es lo suficientemente grande para vencer a la fuerza de los puentes de hidrógeno que mantienen unido el par de bases. La *razón de separación* (definida como  $kv$ ) utilizada dado los valores de la constante de resorte y velocidad de separación ( $k = 0.01N/cm, v = 0.0020A/fs$ ) es de  $0.20pN/fs$ .

Las dinámicas involucran 3 etapas temporales descritas de la siguiente manera:

1. **Primera** parte de la dinámica: La estructura está equilibrada y no existen fuerzas externas de ningun tipo.
2. **Segunda** parte de la dinámica: Comienza el proceso de disociación de ADN. Resortes virtuales abren la estructura base por base de acuerdo a una distancia máxima.
3. **Tercera** parte de la dinámica: La estructura de ADN está completamente abierta y se deja equilibrar nuevamente.

#### Estructura 1: T-T-G-C-G

Las siguientes gráficas muestran los intervalos temporales de separación de bases nucleicas. El intervalo temporal para llegar a esta distancia máxima para el *primer par de bases* (TA) de la estructura, es de  $1860 fs$ . Podemos notar que el *segundo par de bases* es igualmente un par TA y el intervalo encontrado es de  $2090 fs$ . Esto tiene sentido al darnos cuenta de que la primera base cuesta menos romperla debido a que no sufre efectos de apilamiento por debajo. Para el caso de los siguientes pares de base, que son del tipo GC, se encontraron los intervalos de  $4800 fs$ ,  $990 fs$  y  $2860 fs$ . El *tercer par de base* es el que mayor tiempo tomó de separar, esto puede ser debido a que se encuentra justo en la parte media de la cadena y sufre más influencia del apilamiento. Por otro lado, el *cuarto par de base* mostró un tiempo relativamente corto en comparación con las demás. Esto podemos entenderlo al ver la gráfica (3.8), y darnos cuenta de que este par comenzó a separarse incluso antes de que los resortes actuaran directamente sobre él. Las fluctuaciones térmicas podrían ser responsables de esta separación. Existen ciertas helicasas que actúan de manera pasiva, es decir, esperan fluctuaciones térmicas que provoquen una apertura del par de bases para comenzar a avanzar sobre la cadena, el comportamiento visto por este *cuarto par de base* nos podría indicar justo este caso.

Para esta primera secuencia de ADN, la separación base por base tomo **12 ps** en llegar a la separación total de la doble hélice. Además, se le permitió un proceso de equilibración a la estructura de **3200 fs** una vez completada la separación, con la finalidad de analizar efectos de recombinación. Se muestra, en los resultados, que las bases continúan abriéndose y no existe una recombinación de los pares de base. Cabe señalar, que la distancia de separación que se utilizó puede ser demasiado grande y es la razón de no ver posibles recombinaciones. En procesos celulares, otras moléculas y el mismo cuerpo de las helicasas ayudan a evitar que las bases se vuelvan a juntar.

Es importante notar de las gráficas, que al inicio de la aplicación de los resortes, cada base se resiste a la separación.

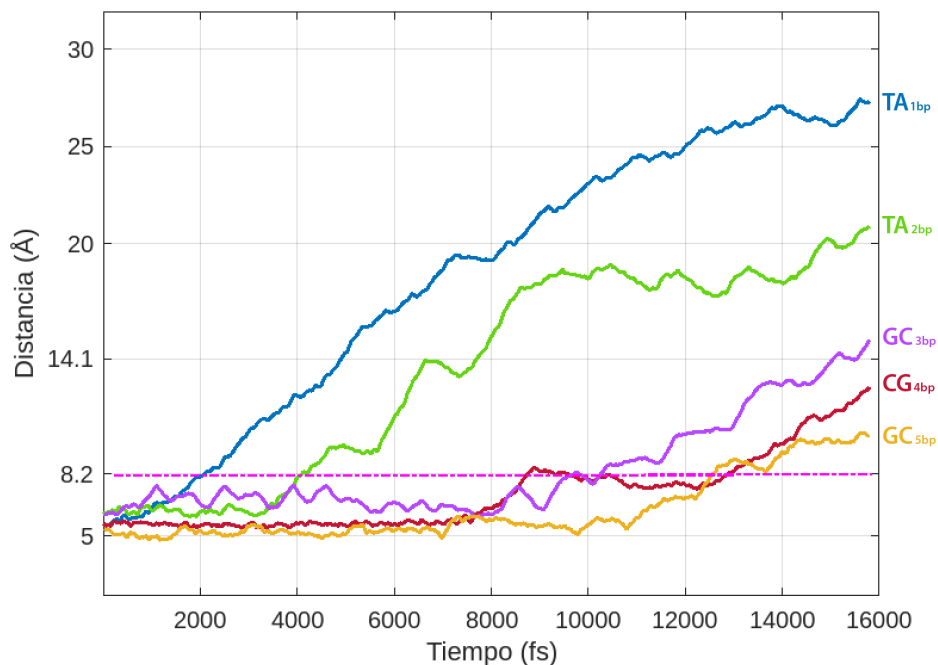


Figura 3.7: Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-T-G-C-G).

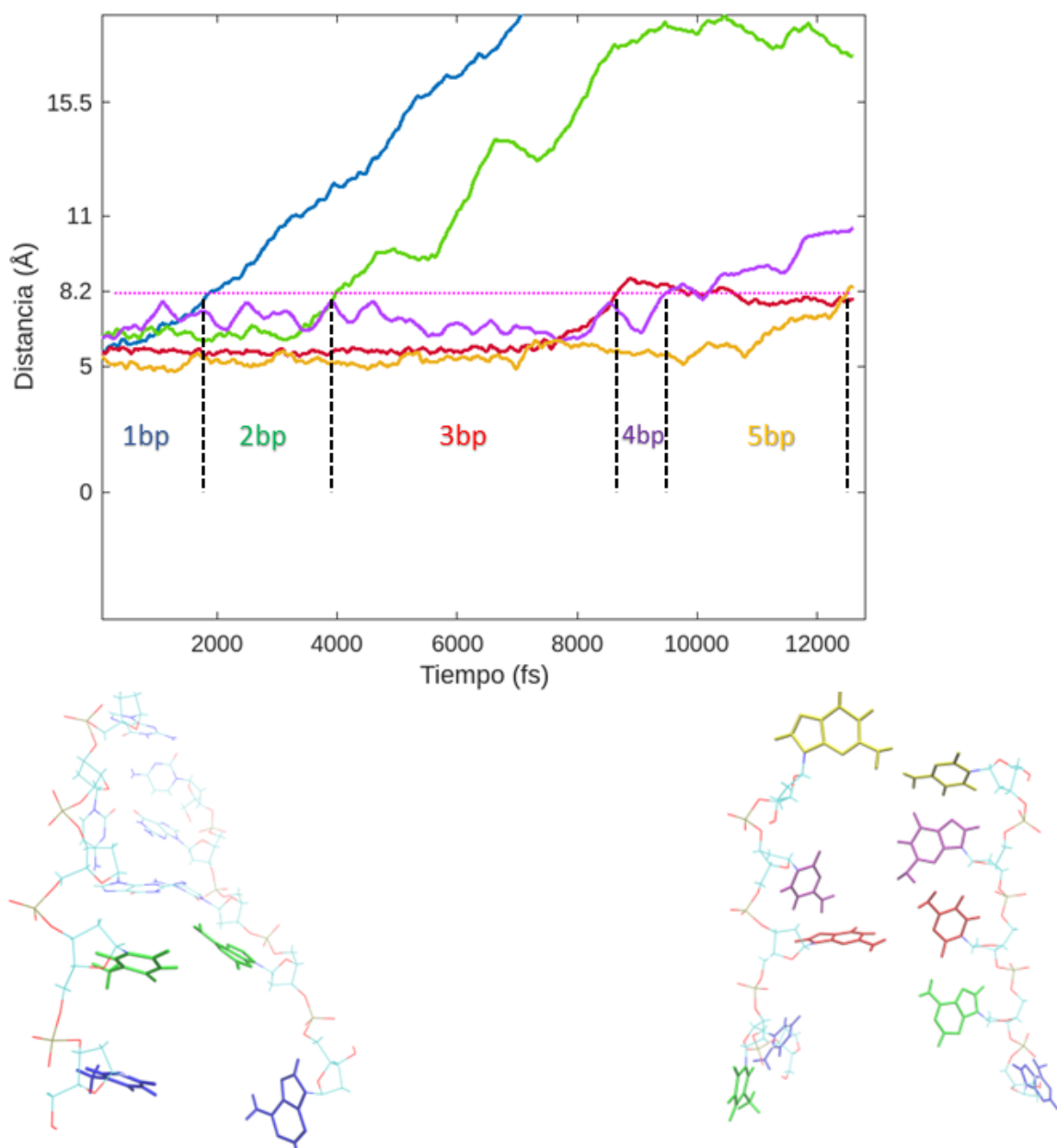


Figura 3.8: Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-T-G-C-G) junto con intervalos temporales de apertura de cada par de bases. Se muestran imágenes de la estructura con 2 (izquierda) y 5 (derecha) pares de base disociados.

Las fluctuaciones térmicas se analizaron en las 3 etapas de la dinámica, previamente mencionadas. Durante el proceso de disociación el sistema recibe energía por parte de los resortes virtuales, por lo que se espera que el sistema este fuera del equilibrio durante esta etapa.

- ◇ **Primera etapa (0-3000 fs):** El sistema está en equilibrio y no se ha comenzado el proceso de separación ( $T_{prom} = 302.92 \text{ K}$  ,  $E_{prom} = -1.012 \text{ H}$ )
- ◇ **Segunda etapa (3000-15600 fs):** Comienza el proceso de disociación base por base. Las fluctuaciones son más grandes y se llega a una  $T_{max} = 356.2 \text{ K}$  y  $E_{max} = -0.7658 \text{ H}$ . El pico máximo de temperatura se da justo en la separación del segundo par de base, y el máximo en energía exactamente en la separación del tercer par de bases. Hay fluctuaciones de  $\approx 18 \%$ .
- ◇ **Tercera etapa (15600-18800 fs):** La estructura está completamente separada y vemos que comienza de nuevo la equilibración.  
Se obtienen valores  $T_{prom} = 297.72 \text{ K}$  y  $E_{prom} = -1.062 \text{ H}$ .

Al inicio de la dinámica, el valor promedio de energía es  $E_i = -1.012 \text{ H}$  y el cambio de energía entre esta etapa inicial y la separación entera de ADN está dada por  $\Delta E = 31.37 \text{ Kcal/mol}$ . De la figura 3.10 podemos ver que la temperatura es relativamente indiferente a la apertura de los pares de base. En contraste, la energía es sensible al proceso de separación y podemos ver picos de energía cercanos a los momentos de separación de cada par de base de las secuencias (asteriscos en la figura 3.10).



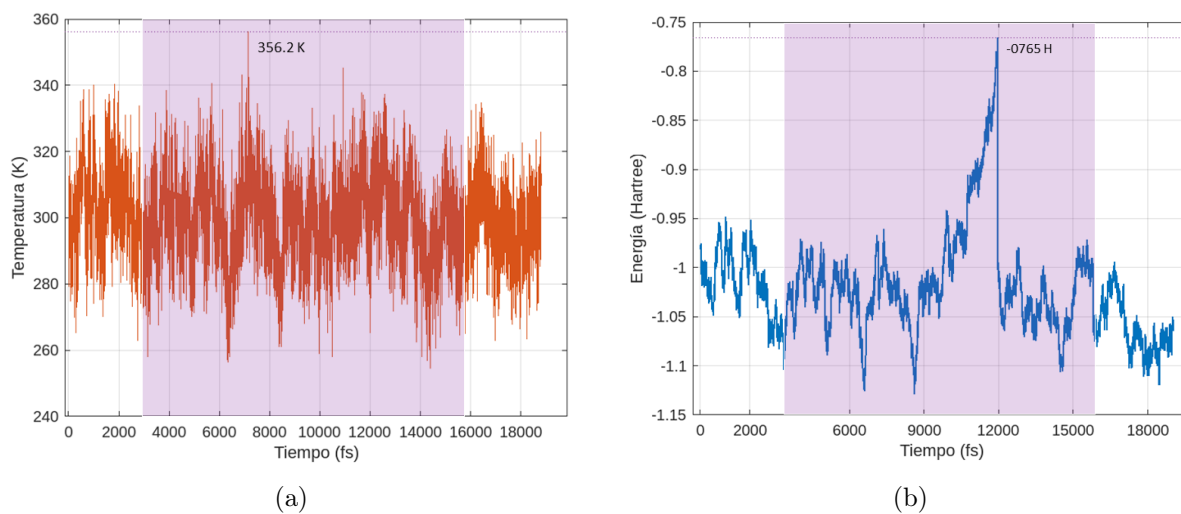


Figura 3.9: Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-T-G-C-G, durante toda la dinámica (3 etapas). La parte sombreada se refiere a la sección durante la cual actúan los resortes.

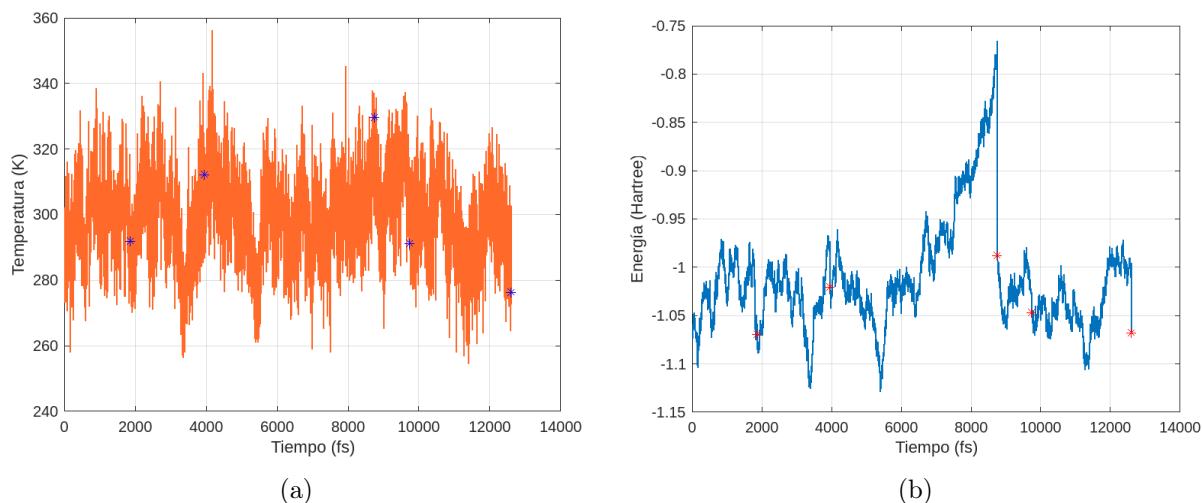


Figura 3.10: Fluctuaciones de temperatura (a) y de energía (b), para la secuencia T-T-G-C-G, durante el proceso de disociación (13 ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base.

### Estructura 2: T-C-G-C-G

Similarmente, para la segunda estructura se encontraron los siguientes resultados.

El intervalo temporal para llegar a esta distancia máxima para el *primer par de base* (TA) de la estructura, es de 2970 fs. Podemos notar que es el par de bases que cuesta menos romper debido a que no cuenta con una base apilada por debajo. Además, es la única base del tipo T-A para esta estructura de ADN en particular, y debido a que cuenta con solo dos puentes de hidrógeno, el resultado es justo el esperado.

Para el caso de los siguientes pares de base, que son del tipo GC, se encontraron los intervalos de 5700 fs, 4680 fs, 4170 fs y 5220 fs. El *segundo par de bases* es el que mayor tiempo tomó de separar, seguido del *quinto par de base* y es importante señalar que es el par de bases que más fluctuaciones de distancia sufrió durante la disociación. El valor encontrado de intervalo de separación para este último par de bases, va en contra de lo que se esperaría, ya que es la base que se encuentra en uno de los extremos y no sufre apilamiento por arriba. Parecería indicar que al tener la cadena prácticamente abierta, la última base cuesta más en abrir

Para esta estructura de ADN, la separación base por base tomo 22.7 ps en llegar a la separación total de la doble hélice. Además, se le permitió un proceso de equilibración a la estructura de 3900 fs una vez completada la separación, con la finalidad de analizar efectos de recombinación. Se muestra en los resultados, que las bases continúan abriéndose y no existe una recombinación de los pares de base (ver 3.11), tal y como fue el caso de la primera estructura de ADN previamente analizada.

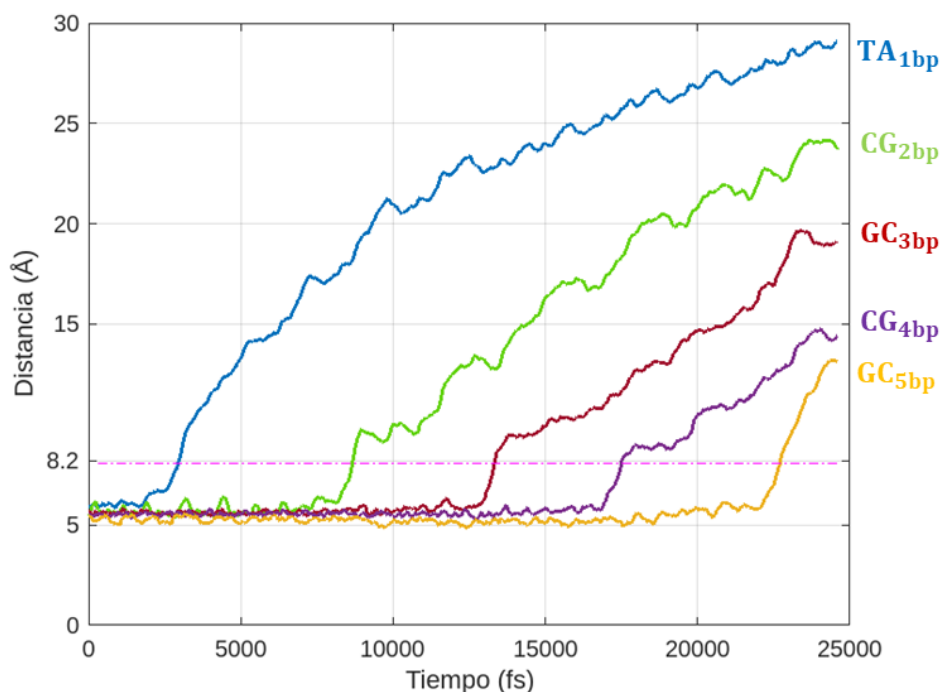


Figura 3.11: Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-C-G-C-G).

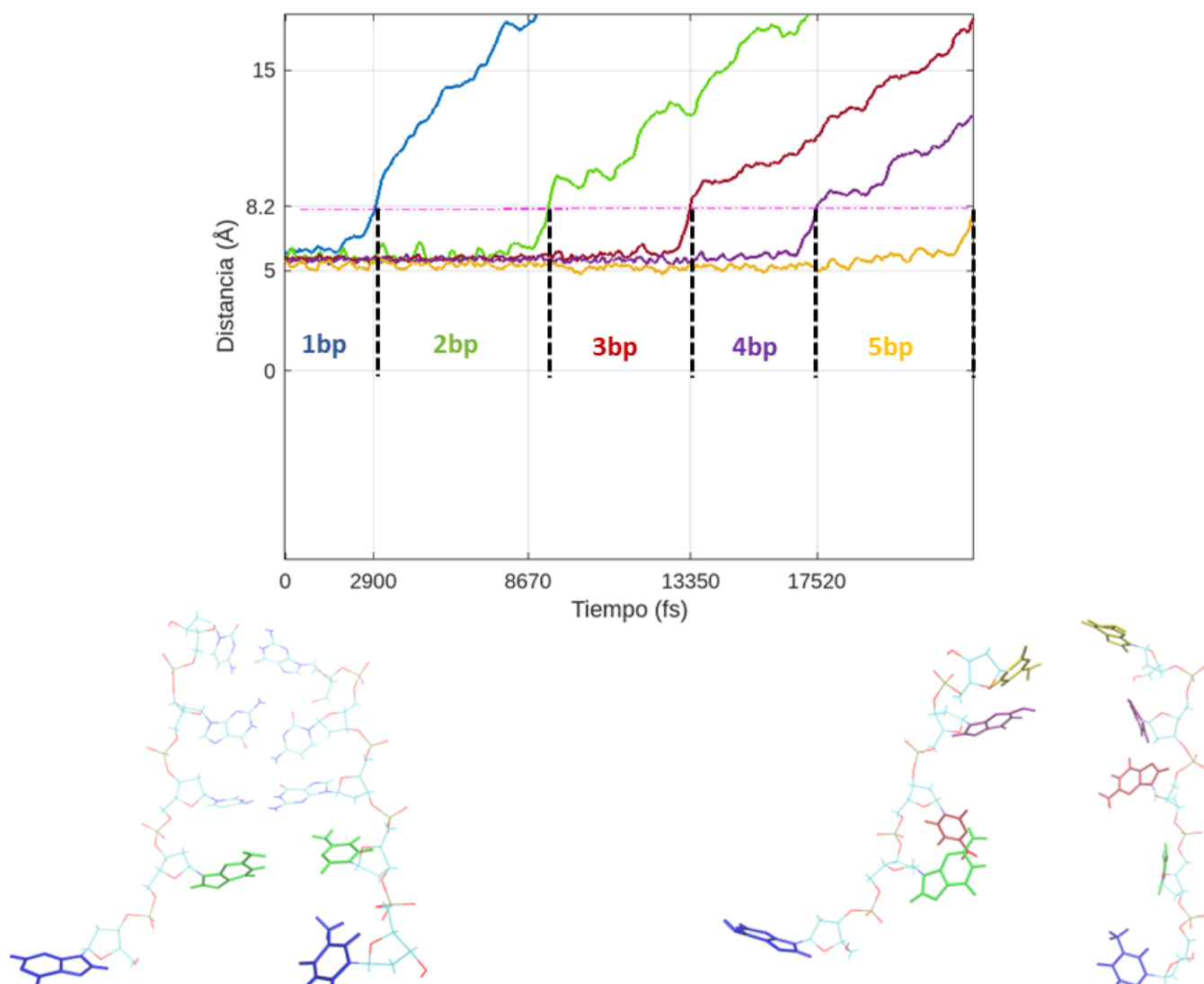


Figura 3.12: Distancias relativas entre el centro de masa de pares de base, durante el proceso de disociación de ADN (secuencia: T-C-G-C-G) junto con intervalos temporales de apertura de cada par de bases. Se muestran imágenes de la estructura con 2 (izquierda) y 5 (derecha) pares de base disociados.

Nuevamente se analizaron las fluctuaciones térmicas para esta segunda estructura en las 3 etapas de la dinámica. Durante el proceso de disociación el sistema estará recibiendo energía por parte de los resortes virtuales, por lo que se espera que el sistema este fuera del equilibrio durante esta etapa.

- ◇ **Primera etapa (0-2000 fs):** El sistema está en equilibrio y no ha comenzado el proceso de separación ( $T_{prom} = 302.82 \text{ K}$  ,  $E_{prom} = -0.3711 \text{ H}$ )
- ◇ **Segunda etapa (2000-24742 fs):** Comienza el proceso de disociación base por base. Las fluctuaciones son más grandes y se llega a una  $T_{max} = 372.4 \text{ K}$  y  $E_{max} = -0.0123 \text{ H}$ . El pico máximo de temperatura se da justo en la separación del primer par de bases, y el máximo en energía justo después de la separación del segundo par de bases.
- ◇ **Tercera etapa (24742-28600 fs):** La estructura está completamente separada y vemos que comienza de nuevo la equilibración.  
Se obtienen valores  $T_{prom} = 301.280 \text{ K}$  y  $E_{prom} = -0.456 \text{ H}$ .

Al inicio de la dinámica el valor promedio de energía es  $E_i = -0.3711 \text{ H}$  y el cambio de energía entre esta etapa inicial y la separación entera de ADN esta dada por  $\Delta E = 53.339 \text{ Kcal/mol}$ . Además, la segunda secuencia de ADN muestra un comportamiento similar respecto a las fluctuaciones de temperatura y energía. De la figura 3.14 podemos ver que la temperatura es relativamente indiferente a la apertura de los pares de base. En contraste, la energía es sensible al proceso de separación y podemos ver picos de energía cercanos a los momentos de separación de cada par de base de las secuencias (asteriscos en la figura 3.14).

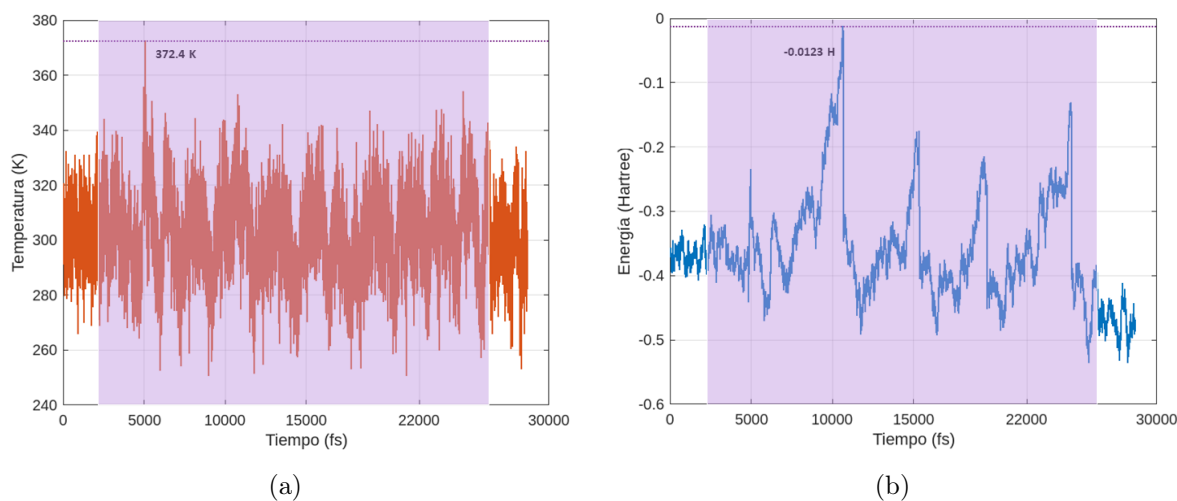
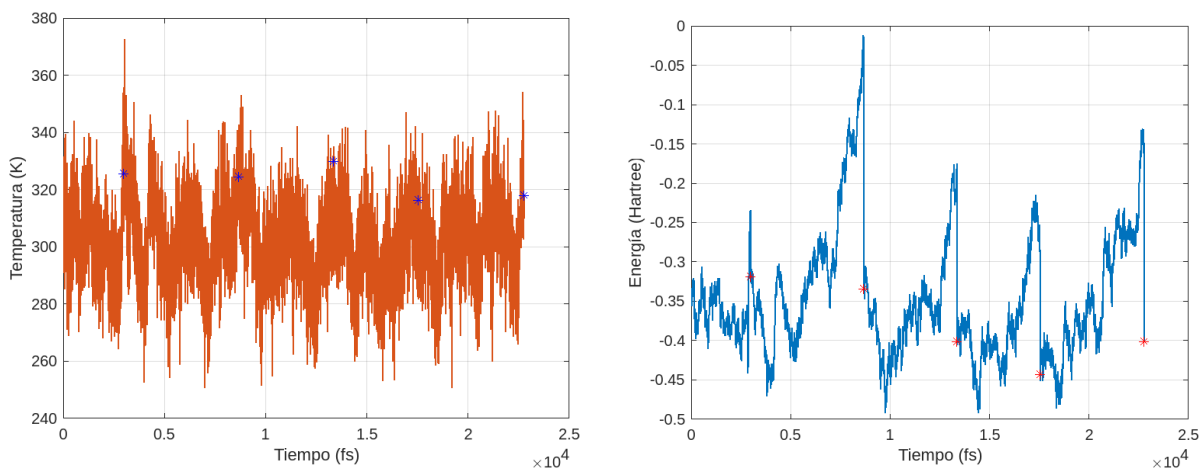


Figura 3.13: Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-C-G-C-G, durante toda la dinámica (3 etapas). La parte sombreada se refiere a la sección durante la cual actúan los resortes.



(a) Fluctuaciones térmicas para la estructura 2 (secuencia: T-C-G-C-G) durante el proceso de disociación (23ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base.

(b) Fluctuaciones energéticas para la estructura 2 (secuencia: T-C-G-C-G) durante el proceso de disociación (23ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base.

Figura 3.14: Fluctuaciones de temperatura (a) y energía (b) para la secuencia: T-C-G-C-G, durante el proceso de disociación (23 ps). Las marcas muestran el instante de separación de cada uno de los cinco pares de base.

De manera general, podemos notar de las figuras 3.14,3.9, que las fluctuaciones de temperatura son relativamente indiferentes al proceso de separación de pares de base. Por otro lado, la energía es más sensible a la separación de bases nucleicas. Por ejemplo, hay picos (ó máximos locales de energía) cerca de los momentos de separación total de cada par de base de la secuencia (asteriscos).

### 3.3.1. Fuerzas disociativas

Existen dos puntos interesantes que caracterizan las fuerzas de disociación. El primer punto es caracterizado por una fuerza que produce una separación rápida entre las bases de cada par de base (ver aumento pronunciado de la pendiente en las figuras 3.15 y 3.17). Llamamos a tal fuerza la *fuerza de ruptura*,  $F_{rup}$ . El segundo punto está caracterizado por la fuerza en la cual cada par de bases llega a la distancia máxima 8.2 Å, el cual etiquetamos como *fuerza máxima*,  $F_{max}$ . Podemos notar que  $F_{max}$  actúa como una cota superior de  $F_{rup}$ , ya que cada base se logra separar antes de llegar a  $F_{max}$ . Sin embargo, el valor de  $F_{rup}$  se determina por una *región de fuerza* y no puede ser asignada de manera adecuada. Para fines prácticos, se calcularon valores de  $F_{rup}$  para cada par de bases de las secuencias de ADN analizadas y se estableció como el último punto a partir del cual el valor de la distancia entre bases nucleicas de cada par de bases no decrece.

#### Estructura 1: T-T-G-C-G

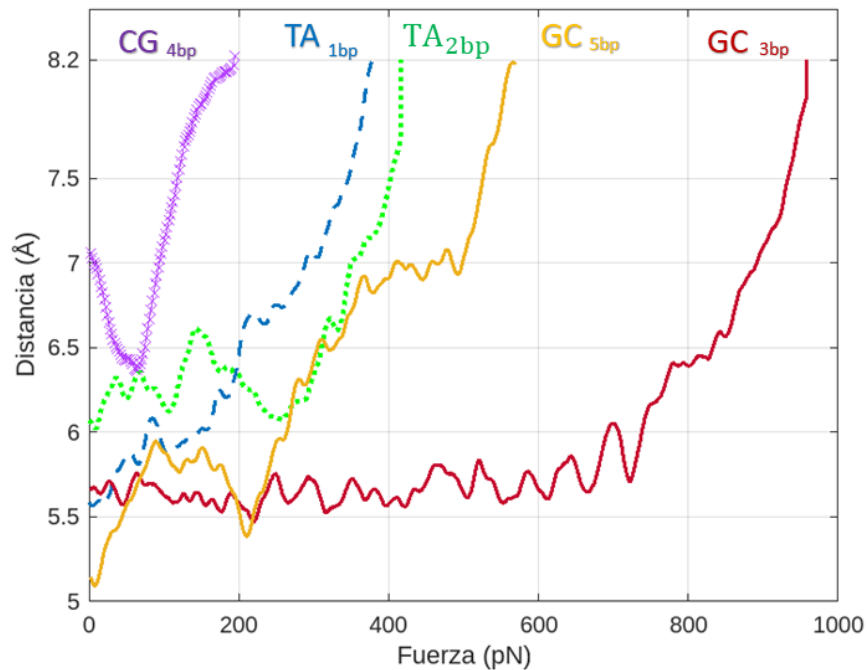


Figura 3.15: Distancia de separación entre el centro de masa para cada par de base como función de las fuerzas de disociación.

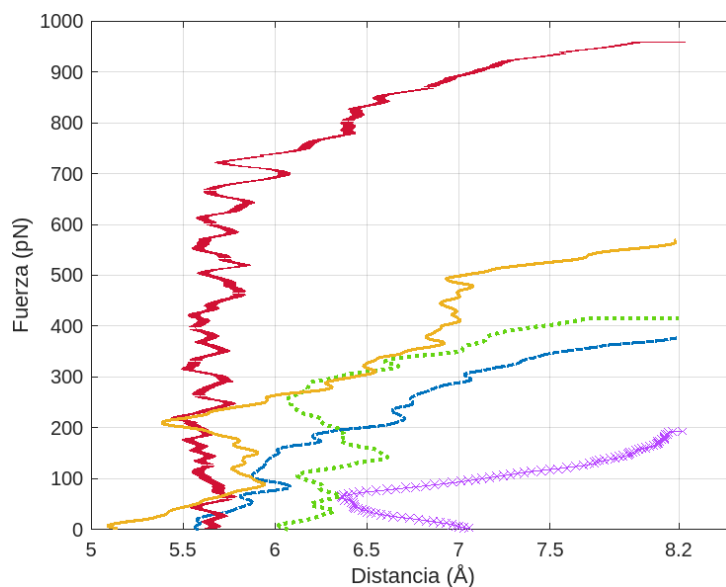


Figura 3.16: Fuerzas de disociación como función de la distancia de separación entre el centro de masa para cada par de bases.

Cuadro 3.4: Fuerzas de separación primer ADN<sup>1</sup>

Par de bases	$F_{max}$ (pN)	$F_{rup}$ (pN)	Tiempo de separación (fs)	$F_{exp}$ , <sup>2</sup> $F_{teo}$ <sup>3</sup> (pN)
1bp (T-A)	377.509	307.973	1860	(630,550)
2bp (T-A)	416.067	379.980	2090	(630,550)
3bp (G-C)	958.274	853.969	4800	(1080,860)
4bp (C-G)	194.027	64.016	990	(1080,860)
5bp (G-C)	569.971	494.008	2860	(1080,860)

<sup>1</sup> Valores de fuerzas disociativas para cada par de bases del primer ADN analizado.

<sup>2</sup> Valores experimentales comparativos obtenidos de [93]. Los valores se obtuvieron sobre bases nucleicas aisladas mediante microscopía de fuerza atómica (AFM, por sus siglas en inglés).

<sup>3</sup> Valores teóricos comparativos obtenidos de [94]. Los valores se obtuvieron sobre bases nucleicas aisladas con el software AMBER 4.1.

Podemos ver que la fuerza de disociación para los pares de base TA y GC de los extremos es ligeramente menor a las bases TA y GC que se encuentran con bases apiladas por arriba y por abajo (segundo y tercer par de bases). La excepción, como ya se mencionó antes, es con la cuarta base CG, la cual mostró una fuerza de disociación más pequeña de lo que se esperaba, y se le adjudica a fluctuaciones debido al baño térmico. Comparando directamente con valores tanto teóricos como experimentales, los valores obtenidos son más pequeños, esto tiene sentido

con el conocido efecto de cooperatividad de las bases. Al ir abriendo la estructura de ADN, el costo de disociación debería ir disminuyendo, que es justo lo que nos dicen los valores de las fuerzas de rompimiento. Adicionalmente, podemos ver que el par de base central (tercer par-GC) es el que más cuesta abrir con una fuerza de disociación de  $F_{rup} = 853.969$  pN. Las gráficas obtenidas (ver figura 3.16) tienen la forma típica de curvas de fuerza-extensión de estructuras de ADN inmersas en agua explícita [22, 24, 49], lo que nos muestra la efectividad del modelo de Langevin para modelar la solvatación del sistema. Esta forma es debido a que el solvente amortigua la separación, para tener una ruptura suave.

### Estructura 2: T-C-G-C-G

A continuación se presentan las gráficas y valores de fuerzas de disociación para la segunda estructura de ADN estudiada.

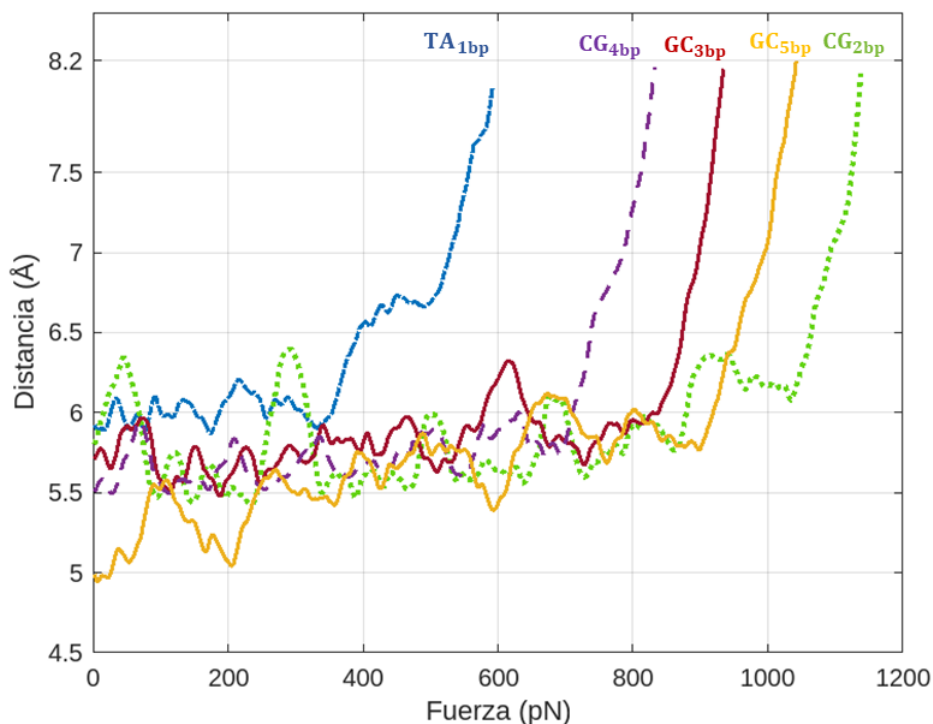


Figura 3.17: Distancia de separación entre el centro de masa para cada par de bases en función de las fuerzas de disociación.



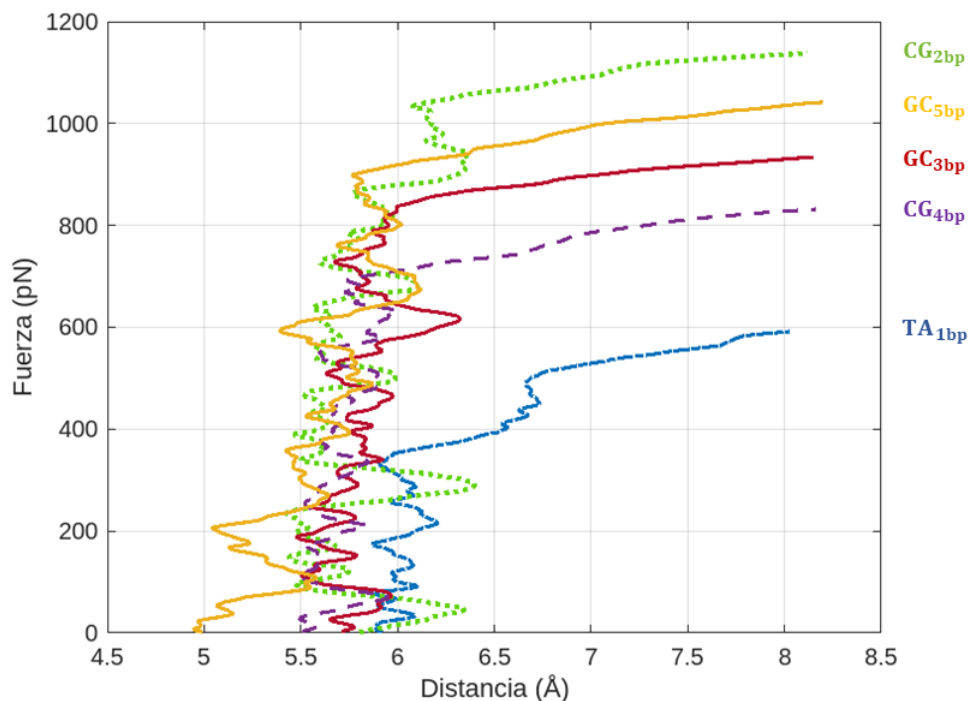


Figura 3.18: Fuerzas de disociación de función de la distancia de separación entre el centro de masa para cada par de bases.

Cuadro 3.5: Fuerzas de separación segundo *ADN*<sup>1</sup>

Par de bases	$F_{max}$ (pN)	$F_{rup}$ (pN)	Tiempo de separación (fs)	$F_{exp}$ , <sup>2</sup> $F_{teo}$ <sup>3</sup> (pN)
1bp (T-A)	591.969	492.031	2970	(630,550)
2bp (C-G)	1137.966	1033.990	5700	(1080,860)
3bp (G-C)	933.969	836.000	4680	(1080,860)
4bp (C-G)	831.970	691.99	4170	(1080,860)
5bp (G-C)	1041.982	897.960	5220	(1080,860)

<sup>1</sup> Valores de fuerzas disociativas para cada par de bases del segundo ADN.

<sup>2</sup> Valores experimentales comparativos obtenidos de [93]. Los valores se obtuvieron sobre bases nucleicas aisladas mediante microscopía de fuerza atómica (AFM, por sus siglas en inglés).

<sup>3</sup> Valores teóricos comparativos obtenidos de [94]. Los valores se obtuvieron sobre bases nucleicas aisladas con el software AMBER 4.1.

El valor más pequeño para las fuerzas de disociación lo presentó el primer par de bases, lo cual tiene sentido, ya que es una base T-A y se encuentra al inicio de la cadena. El *segundo par de la cadena* (CG), mostró una fuerza de disociación más alta con respecto a los otros pares de base GC similares que se encuentran en otras posiciones dentro del ADN. Este valor más alto se puede adjudicar a las fluctuaciones que sufrió este par durante la disociación. En particular, podemos ver en la figura 3.17, los picos pronunciados en la separación del segundo par durante la separación del primer par de bases. Este intento de separación, influenciada por la apertura de su base apilada por debajo, puede ser la razón de encontrar una fuerza de disociación más grande. Por otro lado, el par de bases final (*quinto par-GC*) es el que más cuesta abrir con una fuerza de disociación de  $F_{rup} = 897.96$  pN. Nuevamente, este valor es más alto de lo que se hubiera pensado, tomando en cuenta que es el último par de base y la mayor parte de la cadena ya está abierta.

### Resultados generales:

Los resultados referentes a fuerzas de separación de pares de base nucleicos de cada secuencia muestran comportamientos generales similares en cada caso. Se encontraron fuerzas relativamente pequeñas para bases nucleicas posicionadas al inicio de la cadena. Fuerzas para el mismo tipo de par de base ubicado en posiciones diferentes presentan fuerzas mayores y es atribuido a las influencias de bases apiladas (superiores e inferiores). Además, las gráficas de separación de pares de base durante la dinámica completa nos muestran que existen fluctuaciones de distancia para pares de base incluso antes de que los resortes actúen sobre ellos. Es decir, la acción de los resortes en un par de base específico afecta el comportamiento de pares de base cercanos. Esto tiene sentido al notar que la ribosa interactúa con los pares de base apilados. Al abrir un cierto par de base, la tensión sobre la ribosa y el grupo fosfato es transmitido al siguiente par de base. Finalmente, las gráficas obtenidas (figuras 3.16 y 3.18) tienen la forma típica de curvas de fuerza-extensión de estructuras de ADN inmersas en agua explícita [22, 24, 49], lo que nos muestra la efectividad del baño térmico de Langevin para modelar la solvatación del sistema.

### 3.3.2. Distribución de cargas eléctricas

#### Estructura 1: T-T-G-C-G

A continuación se presentan gráficas del comportamiento y caracterización de las cargas eléctricas de Mulliken de las dos hélices de la estructura de ADN durante el proceso de disociación de pares de base nucleicas. Cabe señalar que la estructura de ADN está cargada negativamente debido a los grupos fosfatos. Esta carga es de vital importancia en procesos de reconocimiento molecular. Distintas moléculas se ven atraídas por la fuerte interacción de los fosfatos, lo cuales también cumplen varias funciones de protección y estabilización.

En este caso, podemos notar el comportamiento oscilatorio de las cargas durante el proceso de separación de bases. Esto es debido a que los resortes virtuales alteran este comportamiento y generan las fluctuaciones. Una vez con la estructura completamente abierta (parte sombreada de la figura 3.19) las cargas se estabilizan nuevamente y tenemos las dos hebras completamente separadas, justo como ocurre en los procesos de replicación celular. Las cargas para la *hebra izquierda* fluctúan entre los valores  $-4.374 e$  y  $-4.079 e$ , es decir, fluctuaciones de  $\approx 7.3\%$ . Para la *hebra derecha* los valores están alrededor de  $-3.921 e$  y  $-3.626 e$ .

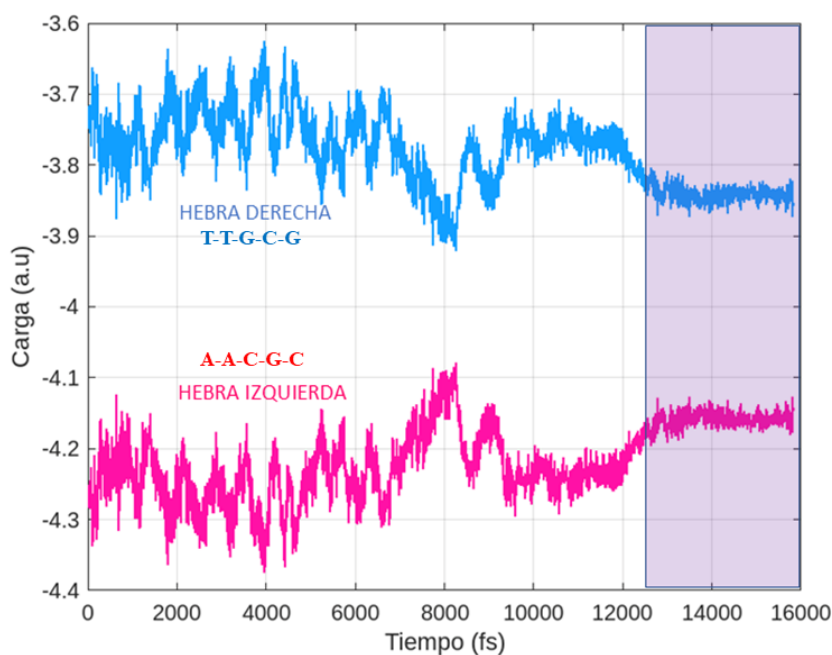


Figura 3.19: Distribuciones de carga para las hebras de la estructura de ADN durante el proceso de disociación. La parte sombreada se refiere al momento en el que los resortes se apagan.

Nos interesa analizar el par de base central de cada secuencia (en este caso es del tipo GC) para minimizar posibles efectos de borde. A continuación se presentan valores para cargas eléctricas entre átomos que conforman los puentes de hidrógeno entre bases, para el momento inicial y final de la separación de dicho par de bases.

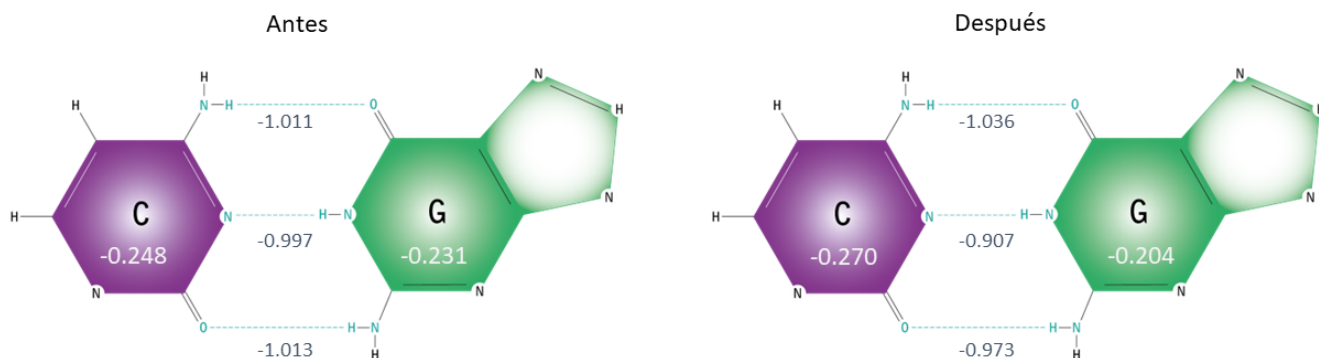


Figura 3.20: Cargas eléctricas de átomos involucrados en puentes de hidrógeno en el par de bases central (GC), para antes (lado izquierdo) y después (lado derecho) de la separación.

Se realizó un análisis del comportamiento de cargas eléctricas para bases apiladas. Nuevamente, tomamos de referencia el par de base central (tercer par de base de la secuencia, tipo GC) para cada estructura de ADN estudiada. Se caracterizaron apilamientos superiores (directamente arriba de la guanina y citosina del par GC) e inferiores (directamente abajo de la guanina y citosina del par GC). Para el apilamiento inferior entre el segundo (TA) y tercer (GC) pares de base (figura 3.21) de la *estructura 1* (secuencia: T-T-G-C-G), vemos los apilamientos inferiores directos entre G,C (morados) y T,A (verdes), respectivamente.

Se puede notar que durante la apertura de la base central (parte sombreada de la figura 3.21) las bases apiladas G-T (lado izquierdo) comienzan a mostrar una polarización pronunciada y al término de esta disociación del par de base central las cargas se neutralizan. Por otro lado, para el apilamiento entre las bases C-A (lado derecho), la mayor diferencia entre los valores de las cargas se dan durante la disociación del par de base central (GC) y al término de ésta, se estabilizan. Este comportamiento nos habla de la existencia de transferencia de carga entre bases apiladas durante el proceso de disociación de la estructura de ADN en su totalidad, seguido de una etapa de estabilización cuando se desactivan los resortes.

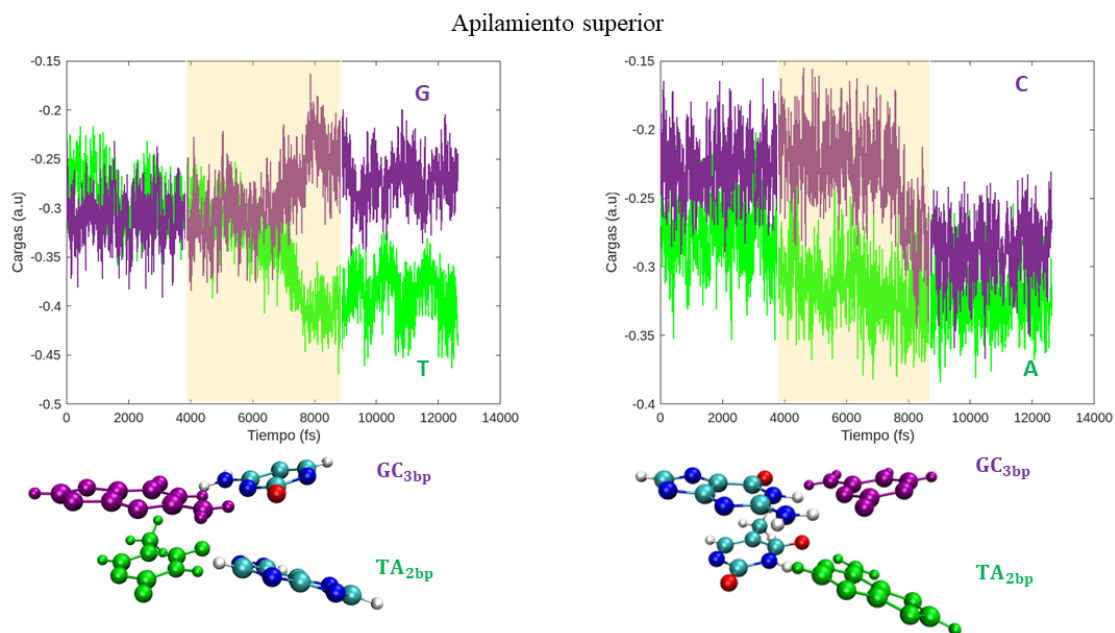


Figura 3.21: Variaciones de carga para bases apiladas (segundo y tercer par de base) para el primer ADN (secuencia: T-T-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente.

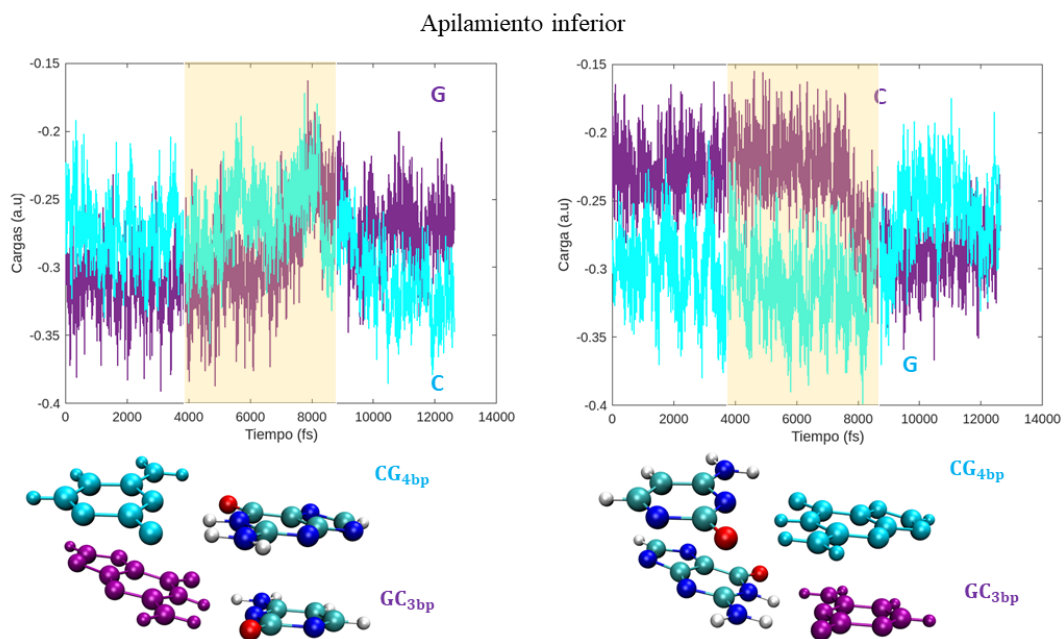


Figura 3.22: Variaciones de carga para bases apiladas (tercer y cuarto par de base) para el primer ADN (secuencia: T-T-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente.

Similarmente, para los apilamientos superiores entre el tercer (GC) y cuarto (CG) pares de base (figura 3.22) para la *estructura 1*. Vemos los apilamientos superiores directos entre G,C (morados) y C,G (azules), respectivamente.

Se puede notar que al igual que en el caso de apilamientos inferiores para el lado izquierdo, existe una polarización, aunque en este caso es más sutil. En contraste con lo visto para apilamiento inferior por derecha, para el apilamiento superior, notamos que existe un cambio en los valores de las cargas durante el proceso de disociación del par de base central (GC-morado), por lo que se tiene transferencia de carga.

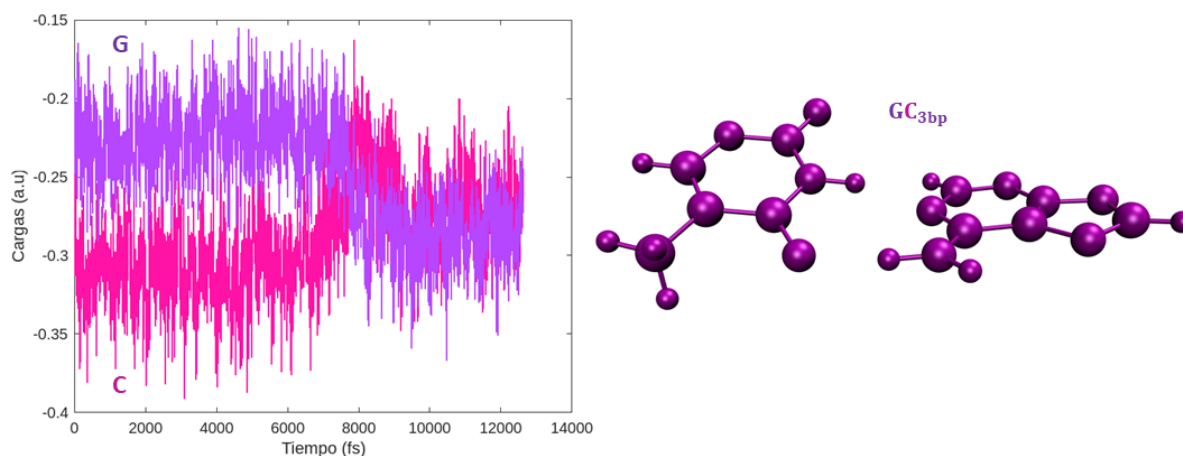


Figura 3.23: Distribuciones de carga para bases nucleicas del par de base central (GC).

Finalmente, se presentan los comportamientos de cargas eléctricas entre las bases involucradas en el par de base central (GC) de la primera secuencia T-T-G-C-G. Vemos que a partir del intervalo de disociación ( $\approx 3900 - 9000$  fs), existe un cambio en la polarización de las bases y los valores promedios de las cargas se vuelven más cercanos, seguido de la estabilización al termino del protocolo de apertura del ADN.

**Estructura 2: T-C-G-C-G**

A continuación se presenta un análisis de comportamiento de cargas para la segunda estructura de ADN estudiada (secuencia: T-C-G-C-G). Similarmente a lo encontrado para la primera estructura, notamos que una vez que se logra la disociación completamente (parte sombreada de la figura 3.19) las cargas se estabilizan nuevamente y tenemos las dos hebras completamente separadas, justo como ocurre en los procesos de replicación celular.

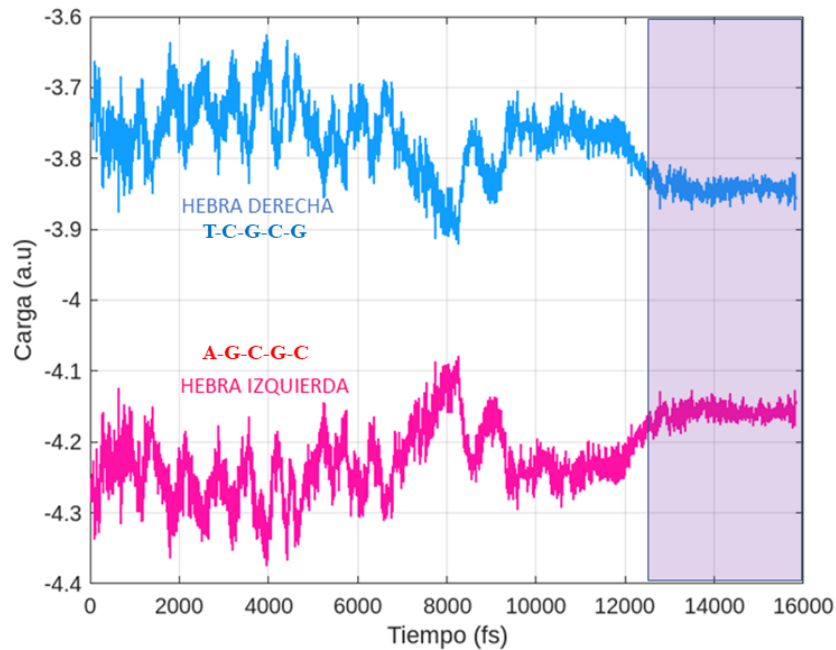


Figura 3.24: Distribuciones de carga para las hebras de la estructura de ADN durante el proceso de disociación. La parte sombreada se refiere al momento en el que los resortes se apagan.

Nos interesa analizar el par de base central de cada secuencia (en este caso es del tipo GC) para minimizar posibles efectos de borde. A continuación se presentan valores para cargas eléctricas entre átomos que conforman los puentes de hidrógeno entre bases, para el momento inicial y final de la separación de dicho par de bases.

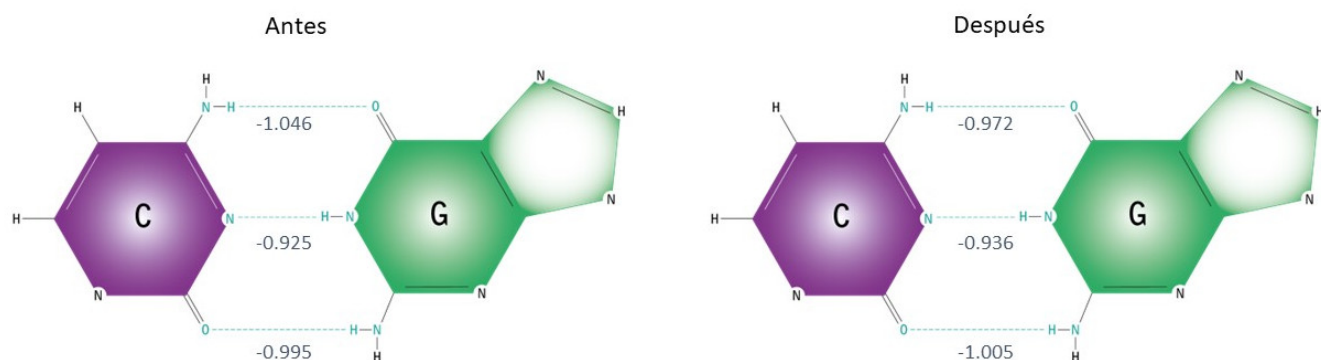


Figura 3.25: Cargas eléctricas de átomos involucrados en puentes de hidrógeno en el par de bases central (GC), para antes (lado izquierdo) y después (lado derecho) de la separación.

Además, se presenta un análisis de cargas para la *segunda estructura* (secuencia: T-C-G-C-G), similar al mostrado para la primera secuencia.

Para el apilamiento inferior entre el segundo (CG) y tercer (GC) pares de base (fig. 3.26) de la estructura 2 (secuencia: T-C-G-C-G), vemos los apilamientos inferiores directos entre G,C (morados) y C,G (verdes), respectivamente. Se puede notar que durante la apertura de la base central (parte sombreada de la figura) las bases apiladas G-C (lado izquierdo) muestran una polarización simétrica prácticamente constante durante la disociación. Por otro lado, para el apilamiento entre las bases C-G (lado derecho), notamos un cambio en la polarización inicial, lo que nos habla de posible transferencia de carga, seguido de una estabilización al final del proceso de separación de bases.



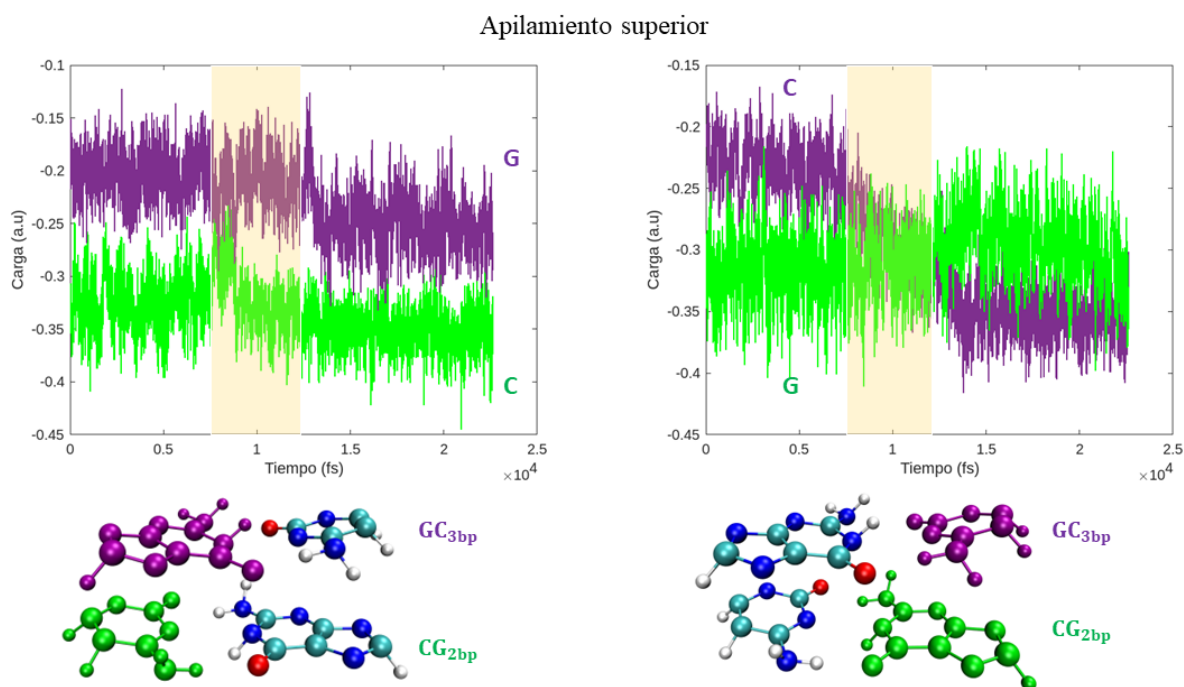


Figura 3.26: Variaciones de carga para bases apiladas (segundo y tercer par de base) para el segundo ADN (secuencia: T-C-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente.

Similarmente, para los apilamientos superiores entre el tercer (GC) y cuarto (CG) pares de base (fig. 3.27) para la estructura 2. Vemos los apilamientos superiores directos entre G,C (morados) y C,G (azules), respectivamente. Para las dos secuencias utilizadas, el apilamiento superior de la base central es del mismo tipo, es decir, GC(morado)-CG(azul). Algo interesante, es notar que el comportamiento de las cargas no es idéntico, lo que nos podría decir, que el resto de la estructura tiene influencia sobre este tipo de apilamiento.

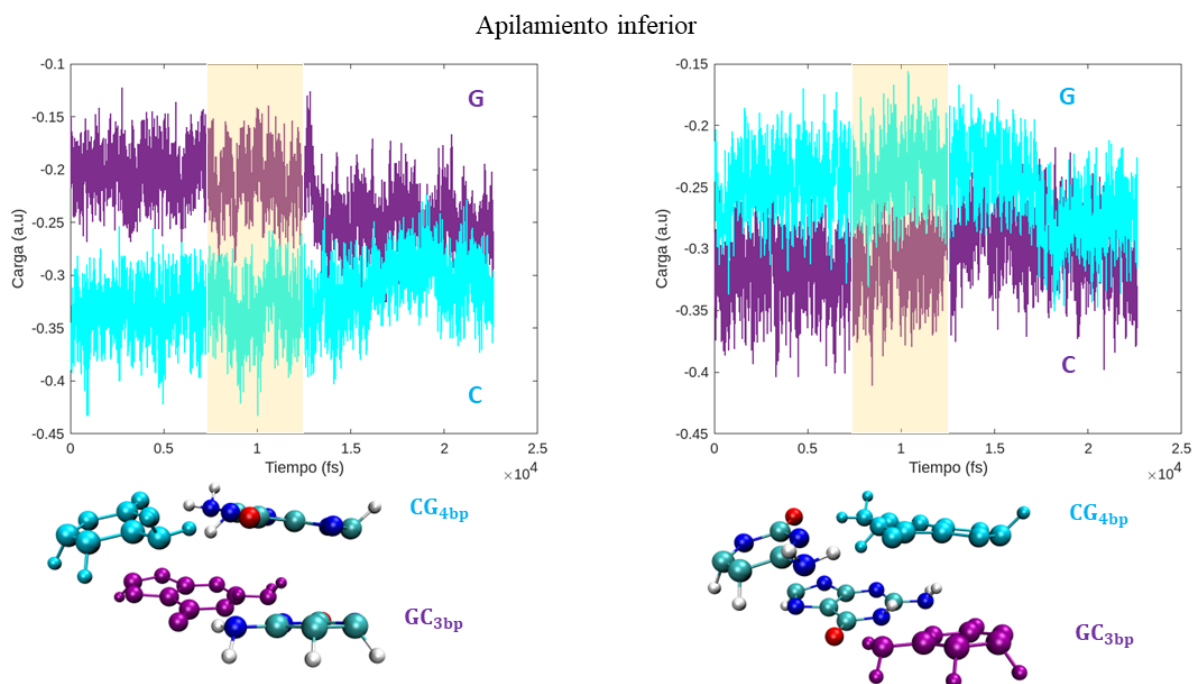


Figura 3.27: Variaciones de carga para bases apiladas (tercer y cuarto par de base) para el segundo ADN (secuencia: T-C-G-C-G). La parte sombreada se refiere al momento en el que la base central (tercer par de bases [GC] de la secuencia) comienza a separarse. Se muestran apilamientos para bases adyacentes por izquierda y por derecha, respectivamente.

Finalmente, se presentan los comportamientos de cargas eléctricas entre las bases involucradas en el par de bases central (GC) de la segunda secuencia T-C-G-C-G. Vemos que similarmente a lo encontrado para la primera estructura, existe una polarización inicial, aunque para este caso, al término de la disociación no ocurrió una inversión entre los valores, simplemente hay una estabilización alrededor de un valor promedio similar para ambas bases.

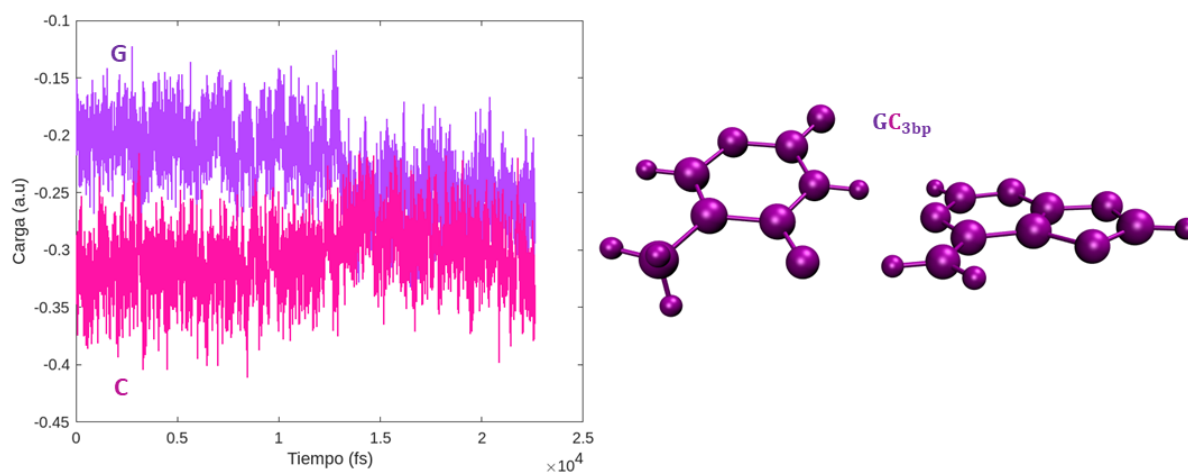


Figura 3.28: Distribuciones de carga para bases nucleicas del par de base central (GC).

Es importante señalar que las gráficas de distribuciones de cargas para bases apiladas son obtenidas por un método de análisis poblacional de Mulliken y existe una cierta ambigüedad para la determinación exacta de los electrones.

### 3.3.3. Dependencia de la secuencia de bases nucleicas

Es importante mejorar el entendimiento de la relación que existe entre el comportamiento del ADN en procesos celulares y la secuencia específica de la cadena. En particular, nos interesa estudiar las posibles dependencias con la secuencia durante la acción de las helicasas en la disociación de la doble hélice.

Las estructuras de ADN presentadas en las secciones anteriores, nos permiten analizar diferencias en los valores obtenidos para las energías, cargas eléctricas y fuerzas, al cambiar la secuencia de bases nucleicas.

Podemos notar que la primera estructura (secuencia: **T-T-G-C-G**) y la segunda (secuencia: **T-C-G-C-G**) difieren en un solo par de base. El segundo par de base para el primer ADN es de tipo T-A, y para el segundo ADN del tipo C-G.

A manera de resumen se presenta una tabla con valores particulares para cada par de bases de cada estructura de ADN. Además, se presenta una tabla comparando valores directos entre estructuras.

Cuadro 3.6: Fuerzas de disociación<sup>1</sup>

Pares de base	$F_{max}$ (pN)	$F_{rup}$ (pN)	Tiempo de separación (fs)	$F_{exp}$ , <sup>2</sup> $F_{teo}$ <sup>3</sup> (pN)
<b>Primer ADN</b>				
1bp ( <b>T-A</b> )	377.509	307.973	1860	(630,550)
2bp ( <b>T-A</b> )	416.067	379.980	2090	(630,550)
3bp ( <b>G-C</b> )	958.274	853.969	4800	(1080,860)
4bp ( <b>C-G</b> )	194.027	64.016	990	(1080,860)
5bp ( <b>G-C</b> )	569.971	494.008	2860	(1080,860)
<b>Segundo ADN</b>				
1bp ( <b>T-A</b> )	591.969	492.031	2970	(630,550)
2bp ( <b>C-G</b> )	1137.966	1033.99	5700	(1080,860)
3bp ( <b>G-C</b> )	933.969	836.000	4680	(1080,860)
4bp ( <b>C-G</b> )	831.970	691.990	4170	(1080,860)
5bp ( <b>G-C</b> )	1041.982	897.960	5220	(1080,860)

<sup>1</sup> Valores comparativos entre las 2 estructuras de ADN. Fuerzas disociativas en la columnas tercer y cuarta. Tiempo de separación por par de base en la quinta columna.

<sup>2</sup> Valores experimentales comparativos obtenidos de [93]. Los valores se obtuvieron sobre bases nucleicas aisladas mediante microscopía de fuerza atómica (AFM, por sus siglas en inglés)

<sup>3</sup> Valores teóricos comparativos obtenidos de [94]. Los valores se obtuvieron sobre bases nucleicas aisladas con el software AMBER 4.1

De las primeras tablas podemos notar que las fuerzas de disociación para cada par de bases están por debajo de valores teóricos para el caso en que las bases se encuentran aisladas. Es decir, el considerar apilamiento y columna vertebral entera, disminuye los valores de las fuerzas

Cuadro 3.7: Valores comparativos entre estructuras<sup>1</sup>

Estructura	t-separación (fs)	$(T_{max}, E_{max})$ (K),(H)	$F_{max-central}$ (pN)	$ \Delta E $ (Kcal/mol)
T-T-G-C-G	12600	(356.2,-0.7658)	958.274 (1430 <sup>2</sup> , 860 <sup>3</sup> )	31.37
T-C-G-C-G	22740	(372.4,-0.0123)	933.969 (1430 <sup>2</sup> , 860 <sup>3</sup> )	53.33

<sup>1</sup> Valores de fuerzas disociativas para cada par de bases.

<sup>2</sup> Valores teóricos comparativos obtenidos de [22] con solvatación explícita.

<sup>3</sup> Valores teóricos comparativos obtenidos de [93].

de ruptura de los pares de base. Por otro lado, el tiempo de separación total de la segunda estructura es mucho mayor al de la primera. Este tiempo extra, de  $\approx 10000$  fs para la segunda estructura, puede deberse a que es una estructura rica en pares de base tipo GC. Como sabemos, este tipo de par de bases cuenta con tres puentes de hidrógeno y lograr la disociación es más difícil. Adicionalmente, la fuerza de apilamiento es mayor entre bases G-C, lo cual contribuye a la estabilidad de la estructura.

El par de bases central para las cadenas cortas utilizadas es del tipo G-C, y cuenta con dos bases apiladas por arriba y por abajo. Este par central sirve de referencia para analizar la dependencia de secuencia. Entre las dos estructuras, uno de los primeros vecinos del par central cambio de T-A a G-C. Vemos que para la primera estructura de ADN, este par de bases tiene una fuerza de disociación de **958.274 pN** y para la segunda estructura de **933.969 pN**, una diferencia de 24.30 pN. A pesar de que la diferencia entre estos valores es relativamente pequeña, los valores para las fuerzas de los pares de base vecinos varían mucho más y su dependencia de posición dentro de la cadena es fuerte.

Además, se realizó una simulación de separación para la segunda secuencia (TCGCG), abriendo únicamente el tercer par de base. Las dos bases apiladas por debajo y las dos bases apiladas por arriba permanecieron cerradas. El valor para la fuerza de disociación encontrada fue de **790.034 pN**, un valor más pequeño en comparación con los casos en el que la cadena de ADN esta sujeta a una separación completa.

El cambio de energía entre la etapa de estructuras equilibradas y la etapa final de estructuras de ADN con hebras separadas fue de, **31.37 Kcal/mol** y **53.33 Kcal/mol**, para la primera y segunda estructura, respectivamente. Vemos que el cambio de energía para la estructura con más pares de bases GC es mayor, justo como uno esperaría. El factor de escala observado, entre el cambio de secuencia (un solo par de base alterado) es de  $|\Delta E_2| = 1.7|\Delta E_1|$ .

Es importante tener en cuenta, que estos valores son susceptibles a la manera que se realiza la separación; ángulo de aplicación de las fuerzas, posición sobre la cadena, y velocidad de disociación.

## Capítulo 4

# Conclusiones

En este trabajo se investigó el comportamiento mecánico de estructuras de ADN durante procesos de disociación de la doble hélice, derivados de la presencia de fuerzas armónicas externas. Se realizaron dinámicas moleculares de primeros principios bajo teoría de funcionales de la densidad, que permitieron obtener valores de fuerzas de disociación, cargas eléctricas, fluctuaciones energéticas y térmicas, durante el proceso de separación. Las estructuras se encuentran inmersas en solvatación implícita acuosa de tipo Langevin, lo que nos permite aproximar condiciones celulares realistas.

Las fuerzas externas utilizadas tienen la peculiaridad de actuar secuencialmente sobre cada par de bases de la estructura, similar al comportamiento de las helicasas en procesos celulares. Además, el posicionamiento específico de las fuerzas sobre las bases nucleicas se puede adaptar dependiendo del tipo de estudio que se quiere realizar.

Las fuerzas de disociación encontradas para cada par de bases toman en cuenta efectos de apilamiento, de la columna vertebral de fosfatos, y de la posición específica sobre la cadena. En particular, los trabajos teóricos sobre estructuras de ADN encontrados en la literatura hacen uso de campos de fuerza, que fallan en incluir efectos cuánticos ó analizan los pares de bases nucleicas de manera aislada. Por otro lado, los estudios experimentales, tanto de pinzas ópticas/magnéticas y AFM, muestran deficiencias en la colocación precisa del amarre sobre las estructuras y ninguno logra implementar fuerzas secuenciales.

Las simulaciones se realizaron sobre dos estructuras cortas de ADN, ampliamente utilizadas en trabajos previos. Los valores de energía de disociación y tiempos totales de separación de la doble cadena fueron, ( $31.37 \text{ Kcal/mol}$ ,  $12600 \text{ fs}$ ) y ( $53.33 \text{ Kcal/mol}$ ,  $22740 \text{ fs}$ ), para la secuencia **T-T-G-C-G** y **T-C-G-C-G**, respectivamente. Esta diferencia de tiempo nos habla de una dependencia de secuencia, donde la estructura rica en bases G-C, presenta un costo de separación más alto. El análisis de fuerzas de disociación de cada par de bases en particular, nos entregó comportamientos similares en cada caso. Fuerzas de separación relativamente pequeñas para pares de base posicionados al inicio de la cadena y fuerzas mayores para pares que se encuentran en posiciones donde están sujetos a interacciones de apilamiento superiores e inferiores. Esto nos muestra que la fuerza que se requiere para separar a cada par de base específico de una secuencia depende de la posición en la cadena nucleica y la naturaleza de los pares de base apilados que lo rodean.

Los resultados de las gráficas de las fuerzas de disociación en función de la distancia entre

base nucleicas (figuras 3.16 y 3.18), asemejan a las encontradas utilizando agua explícita, lo que nos habla de la efectividad de la solvatación implícita de Langevin para simular las propiedades reales a un bajo costo computacional. Adicionalmente, el análisis del comportamiento de cargas eléctricas entre las hélices del ADN y cargas de bases apiladas, nos muestran las fluctuaciones de carga durante la disociación de cada pares de base dentro de la secuencia. Los valores de cargas de bases apiladas muestran fluctuaciones y posibles transferencias de carga durante el proceso de apertura de cada par de bases. Esta caracterización se pudo alcanzar gracias a que se realizó un tratamiento cuántico en las dinámicas y nos ayuda a comprender de mejor manera la distribución de las cargas durante este tipo de procesos y la importancia de considerar apilamiento.

Cabe señalar, que tratar con simulaciones moleculares cuánticas nos restringe las magnitudes de velocidad de estiramiento y de escala temporal. Por lo que, los valores encontrados estarán sujetos a estas discrepancias respecto a valores reales biológicos.

Esta investigación nos muestra la importancia de considerar la columna de fosfatos, bases apiladas y la posición específica dentro de la estructura, para caracterizar las fuerzas de disociación de los pares de base.

El esquema teórico implementado destaca en su exhaustividad, donde distintos campos científicos coexisten, como la física, biología, química y ciencias computacionales. En particular, dentro de la física, el modelo implementa varias áreas: la mecánica clásica durante la dinámica de los núcleos de las moléculas, la mecánica cuántica bajo DFT para analizar el movimiento electrónico, y la mecánica estocástica y estadística dentro de la construcción del reservorio térmico. Como podemos ver, gracias al gran rango de aplicación de las metodologías utilizadas, este tipo de trabajos son importantes para el entendimiento de los procesos biológicos esenciales durante el flujo de información genética, erradicación de enfermedades hereditarias y técnicas nuevas de nanotecnología.

Esta pagina se dejo en blanco intencionalmente



# Apéndice A

## Hartree-Fock

### A.1. Método de Hartree-Fock

Debido a la importancia conceptual y para entender de donde proviene la inspiración que culminó en el desarrollo de DFT y otras teorías post-Hartree-Fock, es de suma importancia entender esta metodología primero, para sentir de una manera más natural la construcción de DFT.

Podemos ver de la ecuación de Schrödinger electrónica (2.11), que debido al término de las repulsiones interelectrónicas la ecuación no es separable, nuevamente se hace uso de una aproximación, al despreciar estas repulsiones. Es decir, una aproximación de electrones independientes, esto nos permite separar la ecuación electrónica al escribir a la función de onda como un producto de orbitales (funciones de onda de un solo electrón).

Los electrones son fermiones y los consideramos bajo la teoría de la mecánica cuántica, por lo que los orbitales deben de cumplir el *principio de exclusión de Pauli* y tomar en consideración el espín del electrón ( $s = 1/2$ ) de manera explícita. Para esto se hace uso del concepto de espin-orbital, el cual es un producto de una función de onda orbital y una función de espín. Se introduce un determinante de Slater<sup>1</sup>, el cual nos garantiza tener un producto antisimétrico (equivalente a cumplir el principio de antisimetría de Pauli), esta observación fue hecha por Fock y Slater en 1930; por lo que podemos entonces escribir a la función de onda molecular de HF, como un producto antisimétrico de espin-orbitales:

$$\psi^{DS}(\vec{x}_1, \dots, \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\vec{x}_1) & \dots & \psi_N(\vec{x}_1) \\ \vdots & & \vdots \\ \psi_1(\vec{x}_N) & \dots & \psi_N(\vec{x}_N) \end{vmatrix} \quad (\text{A.1})$$

El término de las repulsiones es muy importante y debe ser considerado, en el caso del método de HF, esto se toma en cuenta de forma promedio, centramos nuestra atención a cada electrón y lo consideramos moviéndose bajo un campo electrostático generado por los electrones restantes.

---

<sup>1</sup>Podemos notar que los elementos de la diagonal es el producto de Hartree. Los elementos fuera de la diagonal aparecen debido a que consideramos la indistinguibilidad de los electrones y se permite que cada electrón este en cualquiera de los orbitales.

Posteriormente, lo que este método busca, es encontrar a los espin-orbitales que minimicen al funcional de la energía, dado por (A.2)<sup>2</sup>:

$$E = \left\langle \psi^{DS}(\vec{x}_1, \dots, \vec{x}_N) / \hat{H}_{elec} / \psi^{DS}(\vec{x}_1, \dots, \vec{x}_N) \right\rangle \quad (\text{A.2})$$

El mínimo valor para  $E$  se relaciona con la energía electrónica para el estado base, para cierta configuración nuclear. El proceso para encontrar estos espin-orbitales fue propuesto por Hartree en 1928 y se utiliza un método variacional.

Este proceso de minimización nos lleva a las ecuaciones de Hartree-Fock para espin-orbitales individuales, (la minimización se realiza usando multiplicadores de Lagrange, ya que los orbitales deben cumplir la condición de ser ortonormales).

$$\hat{f}\psi_1 = \epsilon_1\psi_1 \quad (\text{A.3})$$

$$\vdots \quad (\text{A.4})$$

$$\hat{f}\psi_N = \epsilon_N\psi_N, \quad (\text{A.5})$$

donde  $\hat{f}_j = \hat{h}_j + \hat{J} - \hat{K}$  se le conoce como el operador de Fock, con  $\hat{J}$  y  $\hat{K}$  los operadores totales de Coulomb e intercambio respectivamente, que tienen la siguiente forma,

$$\hat{J}_j f(x) = f(x) \int \frac{|\psi_j(\vec{x}')|^2}{|\vec{r} - \vec{r}'|} dx' \quad (\text{A.6})$$

$$\hat{K}_j f(x) = \psi_j(x) \int \frac{f(\vec{x}')\psi_j^*(\vec{x}')}{|\vec{r} - \vec{r}'|} dx dx', \quad (\text{A.7})$$

y además,

$$\hat{J} = \sum_{j=1}^N \hat{J}_j \quad \hat{K} = \sum_{j=1}^N \hat{K}_j. \quad (\text{A.8})$$

El operador de Coulomb toma en cuenta la repulsión coulombiana entre los electrones y el de intercambio representa la modificación a esta energía y se le adjudica a los efectos de correlación de espín.

Las ecuaciones de HF, son un conjunto de ecuaciones integro-diferenciales y dependen de las mismas soluciones (por lo que se consideran pseudo ecuaciones de eigenvalores), como podemos ver de las definiciones de los operadores. Se resuelven mediante un proceso iterativo hasta alcanzar la autoconsistencia, de ahí el nombre del método autoconsistente de Hartree-Fock [95, 96].

<sup>2</sup>Conocemos  $\hat{H}_{elec}$  y ocupando las reglas de Slater-Condon podemos calcular los valores esperados [62]

## A.2. Método de Roothaan

En 1951, *C.C J Roothaan* y *G.G Hall* [97] independientemente, propusieron representar a los spin-orbitales de HF (en particular la parte espacial) en términos de un conjunto de funciones base [58, 59]. Este procedimiento nos permite encontrar la función de onda de HF usando álgebra matricial y, por lo tanto, hacer más fácil las implementaciones computacionales del método, algo que sabemos es de suma importancia para aumentar el alcance de la teoría a sistemas más complejos.

Se expanden los orbitales espaciales  $\phi_i$  como combinaciones lineales de un conjunto de funciones base de un electron  $\xi_s$  entonces:

$$\phi_i = \sum_{s=1}^b C_{si} \xi_s \quad (\text{A.9})$$

Por lo que el problema de calcular la función de onda (ecuación A.1) se transforma al de encontrar los coeficientes  $C_{si}$ . Al sustituir esta última ecuación en las ecuaciones de HF, multiplicamos por  $\xi_r^*$  e integramos sobre  $\vec{r}_1$ , y acomodando obtenemos:

$$\sum_{s=1}^b C_{si} (F_{rs} - \epsilon_i S_{rs}) = 0 \quad (\text{A.10})$$

donde se hicieron las definiciones siguientes:

$$F_{rs} = \langle \xi_r / \hat{f} / \xi_s \rangle \quad S_{rs} = \langle \xi_r / \xi_s \rangle \quad (\text{A.11})$$

Como los elementos de matriz  $F_{rs}$ , involucran integrales que dependen de los orbitales  $\xi_i$  se usa nuevamente un proceso iterativo para la solución de las ecuaciones. La matriz de solapamiento (*overlap*)  $S_{rs}$  nos da una medida de la cantidad de orbitales de átomos cercanos, que están en la misma región del espacio.

Al introducir una notación más compacta obtenemos la forma conocida de las ecuaciones de Roothaan.

$$\mathbf{F}\mathbf{c} = \mathbf{S}\mathbf{c}\epsilon \quad (\text{A.12})$$

Donde  $c$  es una matriz  $b \times b$  compuesta de los elementos  $C_{si}$  y  $\epsilon$  una matriz diagonal de energías orbitales  $\epsilon_i$ .

Las ecuaciones de Roothaan (A.12) tienen solución no trivial solo si se satisface la ecuación secular:

$$\det|F_{rs} - \epsilon_i S_{rs}| = 0 \quad (\text{A.13})$$

El proceso de solución, el cual es más flexible computacionalmente es el siguiente:

1. Se empieza con un conjunto de prueba para las expresiones de los orbitales moleculares en términos de las funciones base.
2. El conjunto inicial de orbitales se usa para construir el operador de Fock.
3. Los elementos de matriz  $F_{rs}$  y  $S_{rs}$  son construidos.

4. Se resuelve la ecuación secular para tener un conjunto inicial de  $\epsilon_i$ .
5. Estas energías orbitales se usan para resolver las ecuaciones de Roothaan nuevamente para un conjunto de coeficientes mejorados.
6. Se repite el proceso hasta alcanzar un criterio de convergencia especificado.

## Apéndice B

# Derivadas de la energía y matriz Hessiana

Como podemos ver, en la geometría de equilibrio las fuerzas ejercidas por los electrones sobre los núcleos son cero (el gradiente es cero). El hecho que  $\nabla\vec{U} = 0$  nos caracteriza un punto estacionario, pero no nos distingue entre mínimos, máximos y puntos silla (representan estados de transición <sup>1</sup>). Para distinguir entre los tipos de puntos estacionarios, necesitamos considerar segundas derivadas de la energía respecto a las coordenadas nucleares (se conoce como cálculo de frecuencias). En general se hace uso de la matriz *Hessiana*:

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 U}{\partial q_1^2} & \cdots & \frac{\partial^2 U}{\partial q_1 \partial q_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 U}{\partial q_p \partial q_1} & \cdots & \frac{\partial^2 U}{\partial q_p^2} \end{pmatrix} \quad (\text{B.1})$$

Por lo que, los puntos estacionarios de la PES, están caracterizados por los eigenvalores de la matriz Hessiana:

- mínimo= todos los eigenvalores positivos.
- máximo= todos los eigenvalores negativos.
- punto silla (edo. transición)= un eigenvalor es negativo y todos los demás positivos.

Las segundas derivadas de la energía respecto a las coordenadas nucleares son las *constantes de fuerzas* para las frecuencias de los modos normales (dentro la aproximación armónica). Particularmente, en trabajos de dinámica molecular, esto es importante, ya que las constantes de fuerzas son valores conocidos de manera experimental (por medio de espectroscopia infrarroja o Raman) y pueden ser usado para comparar con nuestros cálculos computacionales. Esto sería equivalente a decir que los valores obtenidos de las segundas derivadas son buenos, y por ende

---

<sup>1</sup>El tipo de puntos silla que interesan son cuando tenemos mínimos en todas las variables excepto una, en esa tenemos un máximo. Esto corresponde a un estado de transición [66]

los de las primeras derivadas también los son. Como las primeras derivadas de los gradientes en realidad son las fuerzas, esto nos diría que nuestra dinámica molecular esta dentro de valores correctos. Es decir, es una manera para comprobar que nuestra dinámica se esta realizando correctamente y poder notar si hay valores que se desvían de los datos experimentales [59, 66].




# Apéndice C

## Publicaciones

Research article Draft (May 22, 2023)

---

### Simulating the helicase enzymatic action on ds-DNA: a first principle molecular dynamic simulation

 Ivan León <sup>\*α</sup>,  Cristian Ordoñez <sup>†β</sup>, and  Rubén Santamaría <sup>‡α</sup>

<sup>α</sup> Department of Theoretical Physics, Institute of Physics, Universidad Nacional Autónoma de México.

<sup>β</sup> A second affiliation; note how it ends in a full stop.

#### ABSTRACT

In this study, we employ first-principle molecular dynamics to investigate the mechanical properties of short *ds-DNA* molecules during the separation of two single strands. The double strands feature different nucleic sequences and are submerged in an implicit Langevin aqueous solvent that emulates biological cellular environments. By utilizing external harmonic forces in a sequential and helical manner, the two strands are opened. Our study reports on the changes in energy, thermal fluctuations, charge, and individual forces observed during the opening of each base pair in the analyzed sequences. These findings highlight the importance of combining quantum mechanical approaches with aqueous solvent models to obtain a better understanding of *DNA* behavior at the atomic level, especially in the context of cellular replication processes.

**KEYWORDS:** Dna dissociation; Helicase; Nucleic acid sequences; First principles molecular dynamics; Density functional theory; Langevin implicit solvent.

Figura C.1: Página principal del artículo próximo a enviar para su publicación.

# Bibliografía

- [1] C. Q. Choi y S. Dutfield. *7 theories on the origin of life*. 2022. URL: <https://www.livescience.com/13363-7-theories-origin-life.html>.
- [2] C. Sagan. «The origin of life in a cosmic context». En: *Origins of Life* 5 (1974). DOI: [10.1007/BF01207649](https://doi.org/10.1007/BF01207649).
- [3] R. Shapiro. *A simpler origin for life*. *Scientific American*. 2007. URL: <https://www.scientificamerican.com/article/a-simpler-origin-for-life/>.
- [4] W. Isaacson. *The Code Breaker*. Simon Schuster, 2021.
- [5] R. M. Brosh y Steven W. Matson. «History of DNA Helicases». En: *American Chemical Society* 11, 255 (2020). DOI: [10.3390/genes11030255](https://doi.org/10.3390/genes11030255).
- [6] C. A. Laughton y S. A. Harris. «The atomistic simulation of DNA». En: *John Wiley Sons, Ltd* (2011). DOI: [10.1002/wcms.46](https://doi.org/10.1002/wcms.46).
- [7] A. Singh, T. Modi y N. Singh. «Opening of DNA chain due to force applied on different locations». En: *Physical Review E* 94, 032410 (2016). DOI: [10.1103/PhysRevE.94.032410](https://doi.org/10.1103/PhysRevE.94.032410).
- [8] F. Kilchherr et al. «Single-molecule dissection of stacking forces in DNA». En: *Science* 353, aaf5508 (2016). DOI: [10.1126/science.aaf5508](https://doi.org/10.1126/science.aaf5508).
- [9] K. DE, M. Narayan y Patel SS. «T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA». En: *J Mol Biol* 321(5):807-19 (2002). DOI: [10.1016/s0022-2836\(02\)00733-7](https://doi.org/10.1016/s0022-2836(02)00733-7). PMID:12206763.
- [10] J. Sponer, K. E. Rileya y P. Hobza. «Nature and magnitude of aromatic stacking of nucleic acid bases». En: *Phys. Chem. Chem. Phys.* 10, 2595–2610 (2008). DOI: [10.1039/b719370j](https://doi.org/10.1039/b719370j).
- [11] P. Hobza y J. Sponer. «Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab Initio Calculations». En: *Chem. Rev.* 99, 32473276 (1999). DOI: [10.1021/cr9800255](https://doi.org/10.1021/cr9800255).
- [12] M. Levitt. *Nobel Prize Outreach AB*. Downloaded Aug 2022. URL: <https://www.nobelprize.org/prizes/chemistry/2013/levitt/facts/>.
- [13] M. Levitt. «Computer simulation of DNA double-helix dynamics». En: *Cold Spring Harb Symp Quant Biol* (1983).
- [14] N. Singh, A. Singh y T. Modi. «Opening of DNA chain due to force applied on different locations». En: *Physical Review E* 94 (3 2016). DOI: [10.1103/PhysRevE.94.032410](https://doi.org/10.1103/PhysRevE.94.032410).



- [15] S. Nath y T. Modi. «Statistical mechanics of DNA rupture: Theory and simulations». En: *Journal of Chemical Physics* 139.16 (2013). DOI: [10.1063/1.4824796](https://doi.org/10.1063/1.4824796).
- [16] S. Harris, A. Sands y et al. «Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of DNA». En: *Biophysical Journal* 88(3), 1684–1691 (2005). DOI: [10.1529/biophysj.104.046912](https://doi.org/10.1529/biophysj.104.046912).
- [17] J. Rezac, S. Harris y et al. «Stretched DNA investigated using molecular-dynamics and quantum-mechanical calculations». En: *Biophysical Journal* 98(1), 101–110 (2010). DOI: [10.1016/j.bpj.2009.08.062](https://doi.org/10.1016/j.bpj.2009.08.062).
- [18] C. Fonseca Guerra, F. Matthias Bickelhaupt y Evert Jan Baerends. *Tackling DNA with Density Functional Theory: Development and Application of Parallel and Order-N DFT Methods*. 2010. URL: [Computational % 20Chemistry : % 20Reviews % 20of % 20Current % 20Trends, % 20Downloaded % 20from % 20www.worldscientific.com](https://www.worldscientific.com/Computational%20Chemistry:%20Reviews%20of%20Current%20Trends,%20Downloaded%20from%20www.worldscientific.com).
- [19] J. Spöner y A. Mladek. «The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies». En: *Phys. Chem. Chem. Phys* 14, 15257–15277 (2012). DOI: [10.1039/c2cp41987d](https://doi.org/10.1039/c2cp41987d).
- [20] R. Santamaría y A. Vazquez. «Structural and electronic property changes of the nucleic acid bases upon base pair formation». En: *Journal of Computational Chemistry* 15(9), 981–996 (1994). DOI: [10.1002/jcc.540150907](https://doi.org/10.1002/jcc.540150907).
- [21] M. Zapata y R. Santamaría. «The interaction of sodium sulfite with the DNA nucleic acid bases: A first principles molecular dynamics study». En: *Computational and Theoretical Chemistry* 1099, 116–122 (2017). DOI: [10.1016/j.comptc.2016.11.021](https://doi.org/10.1016/j.comptc.2016.11.021).
- [22] C. Ordoñez, D. Zapata y R. Santamaría. «Dissociation of the Watson-Crick base pairs in vacuum and in aqueous solution: a first-principles molecular dynamics study». En: *Journal of Biomolecular Structure and Dynamics* (2021). DOI: [10.1080/07391102.2021.1987988](https://doi.org/10.1080/07391102.2021.1987988).
- [23] S. Mogurampelly y et al. «Overstretching of B-DNA with various pulling protocols: Appearance of structural polymorphism and S-DNA». En: *The Journal of Chemical Physics* 147, 225102 (2017). DOI: [10.1063/1.4991862](https://doi.org/10.1063/1.4991862).
- [24] A. M. Naserian-Nik, M. Tahani y M. Karttunen. «Molecular dynamics study of DNA oligomers under angled pulling». En: *RSC Adv* 4, 10751 (2014). DOI: [10.1039/c3ra45604h](https://doi.org/10.1039/c3ra45604h).
- [25] S. B. Dixit, N. Arora y B. Jayaram. «How Do Hydrogen Bonds Contribute to Protein-DNA Recognition?» En: *Journal of Biomolecular Structure and Dynamics* 17:sup1, 109–112 (2000). DOI: [10.1080/07391102.2000.10506610](https://doi.org/10.1080/07391102.2000.10506610).
- [26] A. D. Godbeer, J. S. Al-Khalili y P. D. Stevenson. «Modelling proton tunnelling in the adenine–thymine base pair». En: *Phys. Chem. Chem. Phys* 17, 13034 (2015).
- [27] M. Monajjemi y Chahkandi. «Theoretical investigation of hydrogen bonding in Watson–Crick, Hoogsteen and their reversed and other models: Comparison and analysis for configurations of adenine–thymine base pairs in 9 models». En: *Journal of Molecular Structure: THEOCHEM* 714(1), 43–60 (2005). DOI: [10.1016/j.theochem.2004.09.048](https://doi.org/10.1016/j.theochem.2004.09.048).

- [28] S. Madariaga, J. Contreras y et al. «Theoretical study of the non Watson-Crick base pair Guanine-Guanine». En: *Journal of the Chilean Chemical Society* 55(1), 50–52 (2010). DOI: [10.4067/S0717-97072010000100012](https://doi.org/10.4067/S0717-97072010000100012).
- [29] M. Palafox, V. Rastogi y et al. «Density functional computations on 6-aminouracil: Effect of amino group in the 6th position on the Watson-Crick base pair Uridine-Adenosine.» En: *Australian Journal of Chemistry* 69(8), 881–889 (2016). DOI: [10.1071/CH15793](https://doi.org/10.1071/CH15793).
- [30] A. M. Naserian-Nik, M. Tahani y M. Karttunen. «Pulling of double-stranded DNA by atomic force microscopy: a simulation in atomistic details». En: *RSC Advances* 3,10516 (2013). DOI: [10.1039/c3ra23213a](https://doi.org/10.1039/c3ra23213a).
- [31] I.S. Ufimtsev y T.J. Martinez. «Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients and First Principles Molecular Dynamics». En: *J. Chem. Theory Comput.*, 5, p2619 (2009).
- [32] NVIDIA. *Quantum Chemistry (QC) on GPUs*. 2016. URL: <https://www.nvidia.com/en-gb/data-center/tesla-v100/>.
- [33] G. Stovall. *Cells Under A Microscope*. Diciembre 2013.
- [34] Cortesia: "National Human Genome Research Institute. Acceso 2022. URL: [genome.gov](http://genome.gov).
- [35] V. Rodwell y et al. *Harper's Illustrated Biochemistry*. 30.<sup>a</sup> ed. y McGraw-Hill Global Education Holdings, 2016.
- [36] W. Ma y K. D Whitley. «Free-energy simulations reveal molecular mechanism for functional switch of a DNA helicase». En: *eLife* 7:e34186 (2018).
- [37] J. Mcfadden y J. Al-Khalili. *Life on the Edge*. 7.<sup>a</sup> ed. Crown Publishers, 2014.
- [38] S. C. West. «DNA helicases get physical». En: *Nature* 384 (1996).
- [39] C. Calladine y H. Drew. *The Physics of DNA and Chromosomes*. IOP Publishing Ltd, 2018. DOI: [10.1088/978-0-7503-1602-6](https://doi.org/10.1088/978-0-7503-1602-6).
- [40] R. Franklin y R. Gosling. «Molecular Configuration in Sodium Thymonucleate». En: *Nature* 171,4356 (1953).
- [41] J. Watson y F. Crick. «Genetical Implications of the Structure of Deoxyribonucleic Acid». En: *Nature* 171,964-967 (1953).
- [42] J. D. Watson. *The Double Helix*. Scribner Classics, 1968.
- [43] Compound Interest. *The Chemical Structure of DNA*. 2015. URL: <https://www.compoundchem.com/2015/03/24/dna/>.
- [44] N. Tuteja y R. Tuteja. «Helicases as molecular motors: An insight». En: *Physica A* 372 70–83 (2006).
- [45] T. Chauhan. Cortesia: *Genetic Education Inc*. URL: <https://geneticeducation.co.in/what-is-helicase-and-how-it-unwinds-dna/>.
- [46] M. W. van der Kamp y A. J. Mulholland. «Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology». En: *American Chemical Society* (2013). DOI: [10.1021/bi400215w](https://doi.org/10.1021/bi400215w).

- [47] C. Calladine y H. Drew. *Understanding DNA*. 3.<sup>a</sup> ed. Elsevier Academic Press, 2004.
- [48] C. Luskin. *Back to Basics: Reader Asks How We Know DNA Exists*. 2015. URL: [https://evolutionnews.org/2015/04/back\\_to\\_basics\\_/](https://evolutionnews.org/2015/04/back_to_basics_/).
- [49] M. Santosh y P. K. Maiti. «Force induced DNA melting». En: *Journal of Physics: Condensed Matter* (2009). DOI: [10.1088/0953-8984/21/3/034113](https://doi.org/10.1088/0953-8984/21/3/034113).
- [50] Bustamante y C.J. «Optical tweezers in single-molecule biophysics». En: *Nat Rev Methods Primers* 1,25 (2021). DOI: [10.1038/s43586-021-00021-6](https://doi.org/10.1038/s43586-021-00021-6).
- [51] D. Song y et al. «Tethered particle motion with single DNA molecules». En: *American Journal of Physics* 83, 418 (2015). DOI: [10.1119/1.4902187](https://doi.org/10.1119/1.4902187).
- [52] J. G. Yodh, M. Schlierf y T. Ha. «Insight into helicase mechanism and function revealed through single-molecule approaches». En: *Quarterly Reviews of Biophysics* 43, 2 (2010). DOI: [10.1017/S0033583510000107](https://doi.org/10.1017/S0033583510000107).
- [53] B. Jagannathan y S. Marqusee. «Protein Folding and Unfolding Under Force». En: *Biopolymers* 99,11 (2013). DOI: [10.1002/bip.22321](https://doi.org/10.1002/bip.22321).
- [54] E. Maletz. *NISE Network-Nanoscale informal Science Education Network*. 2016.
- [55] S. Gupta y P. K. Varadwaj. «Computational enzymology». En: *Biochimie* 79,687-689 (1997).
- [56] S. Gupta y P. K. Varadwaj. «A brief overview on molecular dynamics simulation of biomolecular system: Procedure, Algorithms and Applications». En: *IJPSR* 9(4): 1333-1350 (2018).
- [57] C. Cohen-Tannoudji, B. Diu y F. Laloë. *Quantum Mechanics*. Volume 1, Second Edition. Wiley-VCH, 20.
- [58] I. N. Levine. *Quantum Chemistry*. 7.<sup>a</sup> ed. Pearson, 2014.
- [59] P. Atkins y R. Friedman. *Molecular Quantum Mechanics*. 4.<sup>a</sup> ed. Oxford University Press, 2005.
- [60] J. P. Lowe y K. A. Peterson. *Quantum Chemistry*. 3.<sup>a</sup> ed. Elsevier Academic Press, 2006.
- [61] W. Koch y M. C. Holthausen. *A Chemist's Guide to Density Functional Theory*. 2.<sup>a</sup> ed. Wiley, 2014.
- [62] L. Piel. *Ideas of Quantum Chemistry*. ELSEVIER, 2007.
- [63] P. Hohenberg y W. Kohn. «Inhomogeneous Electron Gas». En: *Physical Review* 136, 3B (1964).
- [64] W. Kohn y L. J. Sham. «Self-Consistent Equations Including Exchange and Correlation Effects». En: *Physical Review* 140, number 4A (1965).
- [65] D. C. Young. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. John Wiley Sons, Inc., 2001.
- [66] A. R. Leach. *Molecular Modeling Principles and Applications*. 2.<sup>a</sup> ed. Pearson Education Limited, 2001.

- [67] R. S. Mulliken. «Electronic Population Analysis on LCAO-MO Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies». En: *The Journal of Chemical Physics* Vol. 23 num. 10 (1955).
- [68] A. Hinchliffe. *Molecular Modelling for Beginners*. 2.<sup>a</sup> ed. John Wiley Sons Ltd, 2008.
- [69] R. L. Liboff. «The correspondence principle revisited». En: *Physics Today* 37, 2, 50 (1984). DOI: [10.1063/1.2916084](https://doi.org/10.1063/1.2916084).
- [70] J. C. Tully. «Mixed quantum–classical dynamics». En: *Faraday Discuss* 110, 407–419 (1998).
- [71] K. Jung. «Is the de Broglie-Bohm interpretation of quantum mechanics really plausible?». En: *Journal of Physics: Conference Series* 442 (2013). DOI: [10.1088/1742-6596/442/1/012060](https://doi.org/10.1088/1742-6596/442/1/012060).
- [72] H. J. C. Berendsen. *Simulating the Physical World*. Chapter 3, pag. 39. Cambridge University Press, 2007.
- [73] L. Verlet. «Computer Experiments on Classical Fluids.» En: *Phys. Rev.* 159, 98 (1967).
- [74] P. H. Cowell y A. C. D. Crommelin. «The Perturbations of Halley’s Comet». En: *Monthly Notices of the Royal Astronomical Society* Volume 67, Issue 8 (1907). DOI: [10.1093/mnras/67.8.511](https://doi.org/10.1093/mnras/67.8.511).
- [75] W.C. Swope y et al. «A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters.» En: *The Journal of Chemical Physics* 76, 637-649 (1982). DOI: [10.1063/1.442716](https://doi.org/10.1063/1.442716).
- [76] D. Frenkel y B. Smit. *Understanding Molecular Simulation*. Academic Press, 2002.
- [77] R. Santamaria, L. Adamowicz y H. Rosas-Acevedo. «Microscopic pressure-cooker model for studying molecules in confinement». En: *Molecular Physics* (2014). DOI: [10.1080/00268976.2014.968649](https://doi.org/10.1080/00268976.2014.968649).
- [78] W. Coffey, Yu. Kalmykov y J. Waldron. *The Langevin Equation*. 2.<sup>a</sup> ed. World Scientific Publishing Co. Pte. Ltd, 2004.
- [79] T. Schlick. *Molecular Modeling and Simulation*. 2.<sup>a</sup> ed. Springer, 2004.
- [80] W.F. van Gunsteren y H.J.C. Berendsen. «Algorithms for brownian dynamics». En: *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics* 45:3, 637-647 (1982). DOI: [10.1080/00268978200100491](https://doi.org/10.1080/00268978200100491).
- [81] P. Ullersma. «An exactly solvable model for brownian motion: I. derivation of the langevin equation». En: *Physica* 32(1):27 – 55 (1966).
- [82] A.O. Caldeira. *An Introduction to Macroscopic Quantum Phenomena and Quantum Dissipation*. Cambridge University Press, 2014.
- [83] R. Zwanzig. «Nonlinear generalized Langevin equations.» En: *J Stat Phys* 9, 215–220 (1973). DOI: [10.1007/BF01008729](https://doi.org/10.1007/BF01008729).
- [84] H. Goldstein y J. Safko C. Poole. *Classical Mechanics*. 3.<sup>a</sup> ed. Addison Wesley, 2000.

- [85] S. Chandrasekhar. «Stochastic problems in physics and astronomy.» En: *Reviews of Modern Physics* 15(1), 1–89 (1943). DOI: [10.1103/RevModPhys.15.1](https://doi.org/10.1103/RevModPhys.15.1).
- [86] E. Hernández-Huerta, R. Santamaria y Tomás Rocha-Rinza. «Thermodynamics from Lagrangian theory and its applications to nanosize systems». En: *Molec. Phys.* (2021). DOI: [10.1080/00268976.2021.1940333](https://doi.org/10.1080/00268976.2021.1940333).
- [87] M.G. Paterlini y D.M. Ferguson. «Constant temperature simulations using the Langevin equation with velocity Verlet integration». En: *Chemical Physics* 243–252 (1998).
- [88] W. F. Van Gunsteren y H. J. C. Berendsen. «Algorithms for brownian dynamics». En: *Molecular Physics* Vol. 45 No. 3,637-647 (1982).
- [89] I.S. Ufimtsev y T. J. Martínez. «Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients and First Principles Molecular Dynamics». En: *J. Chem. Theo. Comp.* 5, p2619 (2009).
- [90] A. Titov y et al. «Generating Efficient Quantum Chemistry Codes for Novel Architectures». En: *J. Chem. Theo. Comp.* 9, p213 (2013).
- [91] W. Shi y W. Zhou. «Frequency distribution of TATA Box and extension sequences on human promoters». En: *BMC Bioinformatics* 7(Suppl 4):S2 (2006). DOI: [10.1186/1471-2105-7-S4-S2](https://doi.org/10.1186/1471-2105-7-S4-S2).
- [92] O. V. Shishkina, J. Sponer y P. Hobza. «Intramolecular flexibility of DNA bases in adenine–thymine and guanine–cytosine Watson–Crick base pairs». En: *Journal of Molecular Structure.* 477, 15–21 (1999). DOI: [S0022-2860\(98\)00603-6](https://doi.org/S0022-2860(98)00603-6).
- [93] T. Boland y B. D. Ratner. «Direct Measurement of Hydrogen Bonding in DNA Nucleotide Bases by Atomic Force Microscopy». En: *Proc. Natl. Acad. Sci.* Vol. 92, pp. 5297-5301 (1995).
- [94] E. Stofer, C. Chipot y R. Lavery. «Free Energy Calculations of Watson-Crick Base Pairing in Aqueous Solution». En: *J. Am. Chem. Soc.* 121, 9503-9508 (1999). DOI: [10.1021/ja991092zCCC](https://doi.org/10.1021/ja991092zCCC).
- [95] D. Hartree. «The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods.» En: 24(1), 89-110 (1928). DOI: [10.1017/S0305004100011919](https://doi.org/10.1017/S0305004100011919).
- [96] V. Fock. «Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems.» En: *Z. Physik* 61, 126–148 (1930). DOI: [doi.org/10.1007/BF01340294](https://doi.org/10.1007/BF01340294).
- [97] C. C. Roothaan. «A Study of TwoCenter Integrals Useful in Calculations on Molecular Structure». En: *Journal of Chemical Physics* (1951). DOI: [10.1063/1.1748100](https://doi.org/10.1063/1.1748100).

