



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

**PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
QUÍMICAS**

**“EFECTOS DE LA COOPERATIVIDAD DE INTERACCIONES
NO-COVALENTES EN LA ESTRUCTURA Y REACTIVIDAD DEL ADN”**

TESIS
QUE PARA OPTAR POR EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

M. en C. JORGE GUTIÉRREZ FLORES

Tutor:

Dra. Angélica Estrella Ramos Peña
Instituto de Investigaciones en Materiales

Comité tutor:

Dr. Fernando Cortés Guzmán
Instituto de Química, UNAM
Dr. Enrique Hernández Lemus
Instituto Nacional de Medicina Genómica

Ciudad de México, Octubre 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS
QUÍMICAS

“EFECTOS DE LA COOPERATIVIDAD DE INTERACCIONES
NO-COVALENTES EN LA ESTRUCTURA Y REACTIVIDAD DEL ADN”

TESIS
PARA OPTAR POR EL GRADO DE

DOCTOR EN CIENCIAS

PRESENTA

M. en C. JORGE GUTIÉRREZ FLORES

Tutor:

Dra. Angélica Estrella Ramos Peña
Instituto de Investigaciones en Materiales

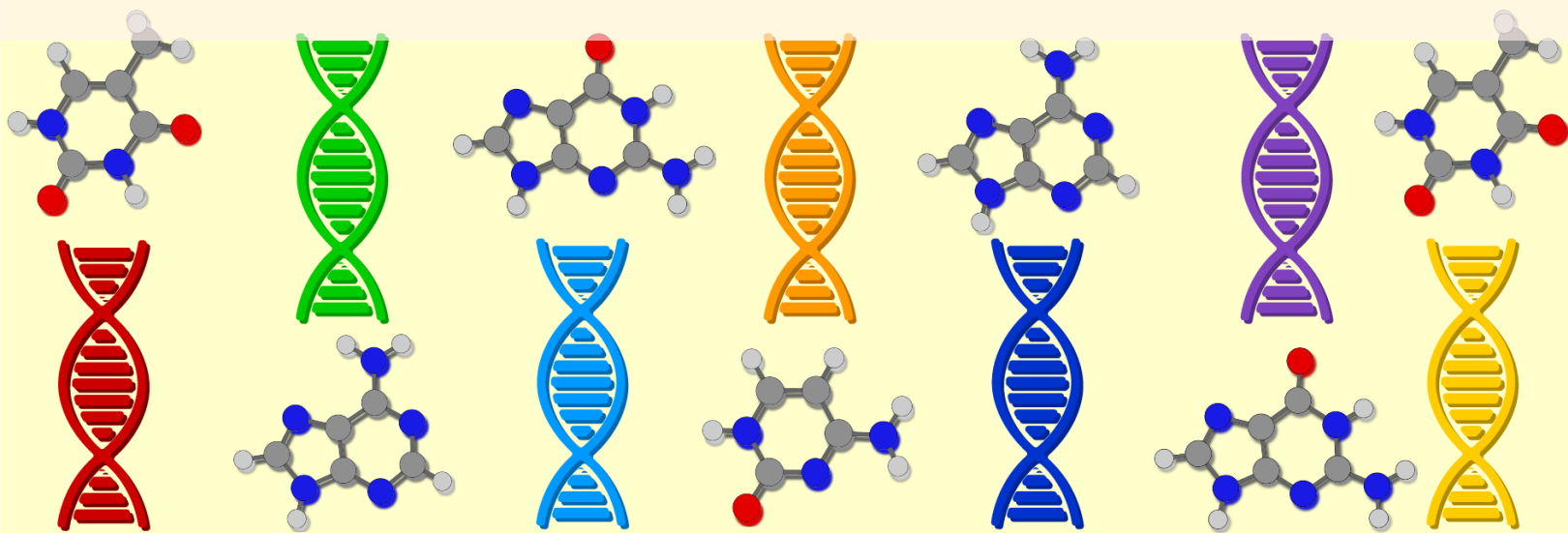
Comité tutor:

Dr. Fernando Cortés Guzmán
Instituto de Química, UNAM
Dr. Enrique Hernández Lemus
Instituto Nacional de Medicina Genómica



Ciudad de México, Octubre 2021

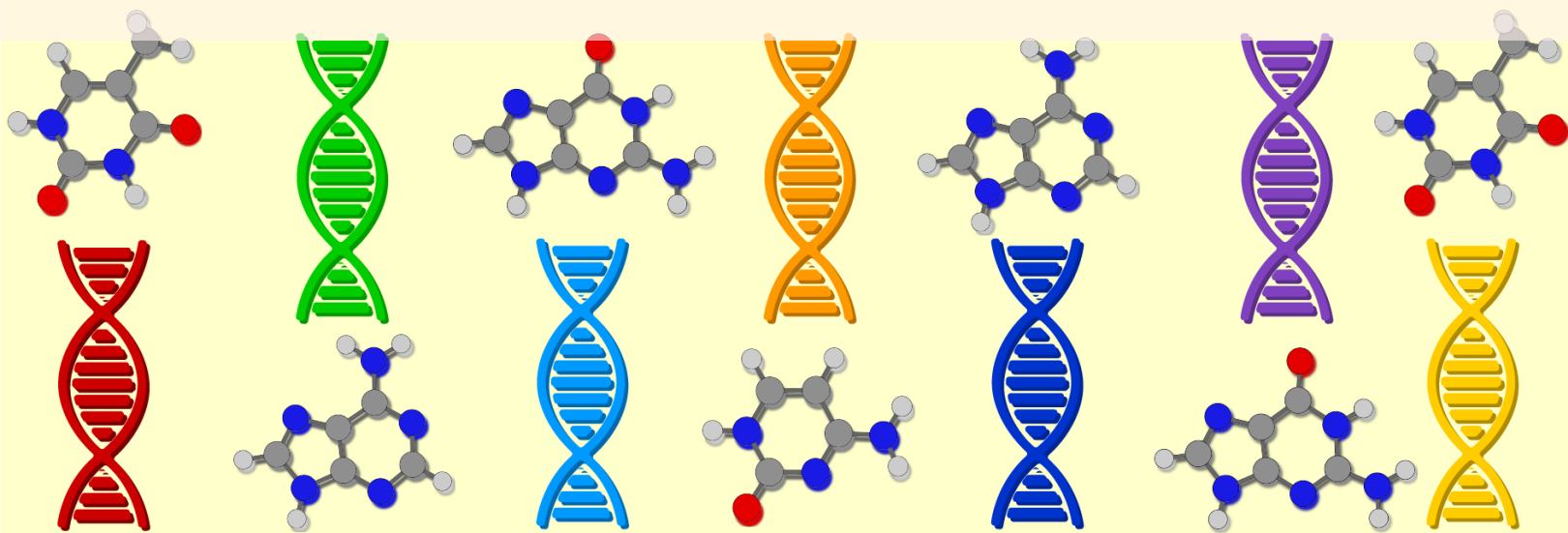
Índice general



	Agradecimientos	7
	Resumen	9
1	Introducción	11
1.1	Ácido desoxirribonucleico	12
1.2	El ADN y su estructura química y tridimensional	15
1.3	La información “oculta” en el ADN	18
1.4	Planteamiento del problema y estado del arte	19
1.5	Objetivos e hipótesis	20
2	Marco teórico	21
2.1	Interacciones débiles	22
2.2	Clasificación de las interacciones no-covalentes	23
2.2.1	Clasificación en función de la distancia	24
2.2.2	Clasificación en función del tipo de moléculas involucradas	26
2.3	Puentes de hidrógeno	30
2.4	Apilamiento π	31
2.5	Cooperatividad	32
3	Resultados y análisis	33
3.1	Interacciones en el ADN	34

3.2	Modelo biológico y químico	35
3.3	Puentes de hidrógeno	36
3.4	Apilamiento π	41
3.5	Análisis de distribución de cargas	44
3.6	Efectos cooperativos y anticooperativos	48
3.7	Contribuciones a la estabilidad de la doble hélice	49
3.8	Relevancia biológica	51
4	Conclusiones	55
4.1	Conclusiones	56
4.2	Perspectivas a futuro	56
	Artículos publicados	59
	Bibliografía	83
A	Teoría de Perturbaciones Adaptada por Simetría	89
A.1	Teoría de Perturbaciones de Rayleigh-Schrödinger	90
A.1.1	El operador resolvente	92
A.2	La Aproximación de Polarización	95
A.3	Teoría de Perturbaciones Adaptada por Simetría (SAPT)	96
B	Teoría de los Funcionales de la Densidad	105
B.1	La ecuación de Schrödinger	106
B.2	El nacimiento de DFT	108
B.3	Ecuaciones de Kohn y Sham	110
B.4	Funcionales de intercambio y correlación	112
B.4.1	Aproximación Local de la Densidad (LDA)	112
B.4.2	Aproximación de Gradiente Generalizado (GGA)	114
B.4.3	Funcionales meta-GGA	115
B.4.4	Fuera de la escalera y de la localidad	116
	Bibliografía	117

Agradecimientos



Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca para estudios de doctorado asignada (CVU: 621254, No. de Becario.: 336100), al proyecto DGAPA-PAPIIT IG100320 por el apoyo económico y a la Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC) por el tiempo de supercómputo otorgado para realizar parte de los cálculos de este trabajo (proyecto LANCAD-UNAM-DGTIC-225).

A la Dra. Angélica Estrella Ramos Peña, mi tutora, por el inmenso apoyo y gran orientación durante toda mi formación académica, por permitirme llevar a cabo este proyecto y por todas las oportunidades brindadas durante mi estancia en el grupo de investigación. Gracias por confiar en mí y en mi trabajo.

Al los integrantes del comité tutor: Dr. Enrique Hernández Lemus y Dr. Fernando Cortés, por involucrarse y formar parte de este trabajo, por enriquecerlo con sus comentarios y por ayudarme a crecer académicamente durante el doctorado.

Al Instituto de Investigaciones en Materiales de la UNAM por el tiempo de cómputo y por el apoyo para realizar este proyecto. Al Mtro. Alberto López Vivas y al L.I. Alejandro Pompa por el gran apoyo en el soporte computacional.

A la Facultad de Química de la UNAM por permitir desarrollarme como docente. Gracias a todos los alumnos a quienes he tenido la fortuna de ser su profesor, siempre aprendí algo nuevo de ustedes.

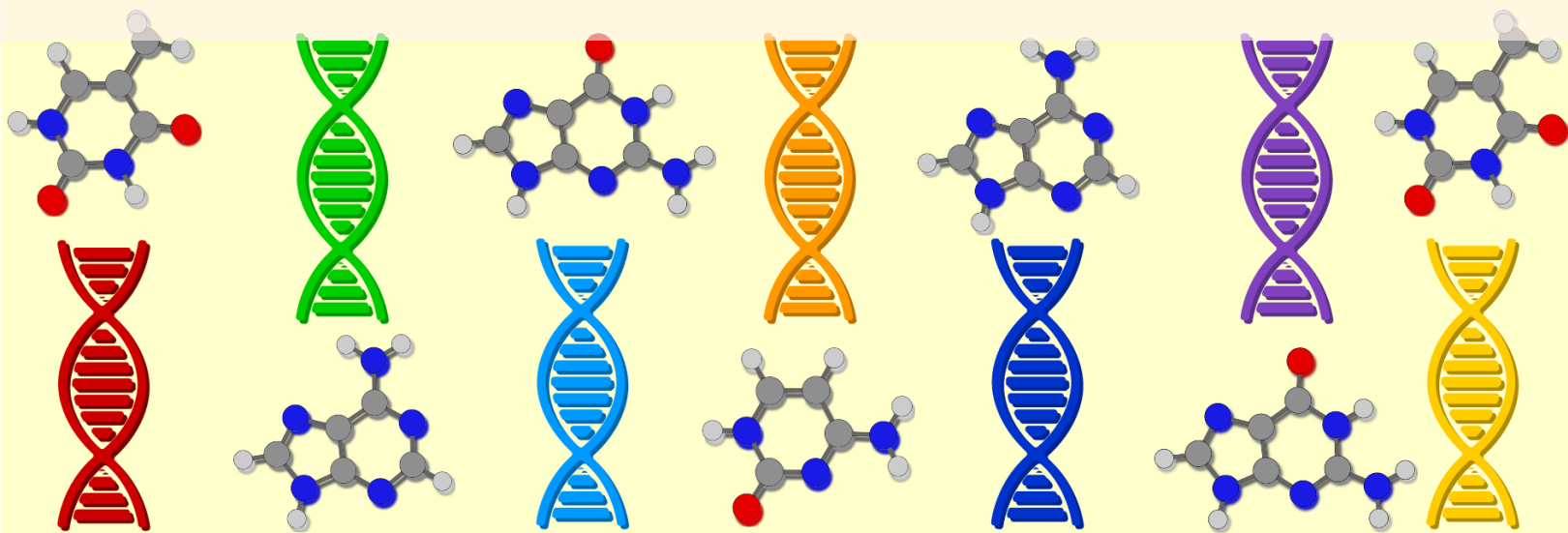
A los miembros del jurado: Dr. Vladimir A. Bassiouk Evdokimenko, Dr. Juan Raúl Álvarez Idaboy, Dra. Ana María Martínez Vázquez, Dra. Nina Pastor Colón y Dr. Fernando Cortés Guzmán por el interés en este proyecto, por su tiempo y por las revisiones y comentarios realizados. Sin duda alguna, todas las observaciones hechas enriquecieron a este trabajo.

Gracias infinitas a mi familia, todo lo logrado no hubiera sido posible sin su invaluable apoyo. Gracias mamá por tu incondicional apoyo, por tu fortaleza, por tus consejos y por

nunca dejarme tirar la toalla. Gracias hermana por siempre apoyarme, por ser una gran amiga, consejera y cómplice. Gracias papá por todas las cosas que en vida me enseñaste y por las enseñanzas que me sigues dando a pesar de no estar en este plano terrenal. Gracias abuelita por tu tiempo, cuidados y consejos.

Agradezco a Eduardo Hernández Huerta por tanto y por todo. A Ameyalli Moreno Fernández por su grandiosa amistad y por todo su inconmensurable apoyo en todo momento. A Rosa Cruz Bernardino por escucharme y apoyarme siempre que lo necesito. A Nidia Lizama Flores, Juan Miguel Arriaga Hernández, Germán Gutiérrez Campos, Allan García, Sofía Lecona Vargas y Nahiely Rosas de la Cruz por su amistad y apoyo de tantos años. A mis compañeros y amigos del grupo de trabajo: Sara Elizabeth Pérez, Enrique Gándara, Mariana Ontañón, Jennifer Salazar y Miguel Ángel Peña, por el aprendizaje en conjunto y por hacer del equipo algo único. A aquellos que, por la emoción de la conclusión de esta etapa, olvidé mencionar, pero saben que sin su amistad y apoyo esto ni hubiera sido posible.

Resumen



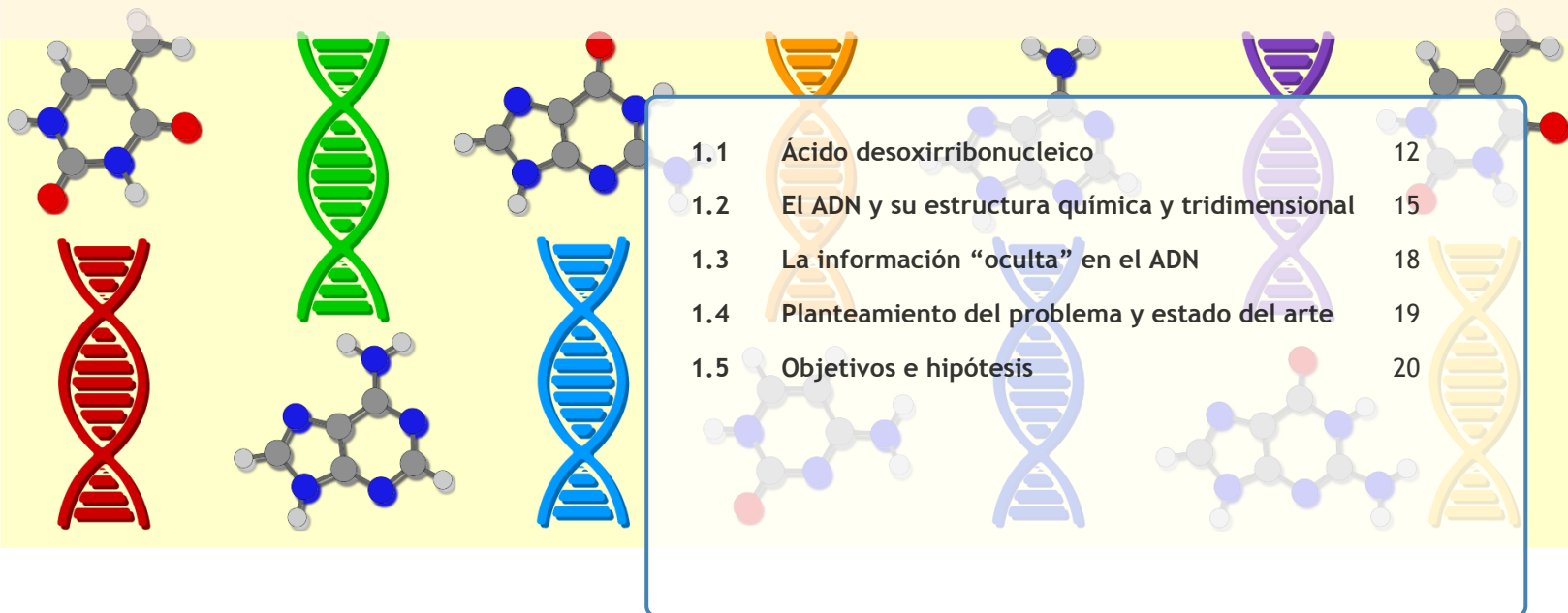
Dada la información contenida en el ácido desoxirribonucleico (ADN), y los procesos bioquímicos en los cuales está involucrado, esta biomolécula puede ser considerada como la molécula de la vida y, por lo tanto, es de vital importancia conocer sus propiedades físicas y químicas, además del impacto que éstas tendrán sobre su comportamiento biológico.

La conducta fisicoquímica del ADN está regida por su estructura, la cual, a su vez, está controlada por interacciones no-covalentes. Se cree que, de éstas, los puentes de hidrógeno y el apilamiento π son las fuerzas intermoleculares más relevantes en la formación del dúplex. Tal es la importancia de estas interacciones que algún error en la forma en la que se conectan las hebras de la doble hélice (por medio de puentes de hidrógeno y del apilamiento π) puede ocasionar alteraciones y/o daño en la estructura tridimensional y, por lo tanto, conducir a daños genéticos.

En este trabajo se describieron las interacciones de puente de hidrógeno y de apilamiento π en función del tamaño (número de pares de base) y de la conformación (secuencia de nucleótidos) de la cadena. Todo esto con la finalidad de tener un panorama más amplio del comportamiento y contribución a la estabilidad del ADN de dichas interacciones no-covalentes. Adicionalmente, como uno de los grandes retos de este proyecto, se buscó relacionar las energías estimadas con el comportamiento químico-biológico de esta biomolécula. Por ende, como parte de los objetivos, se realizó un análisis de la reactividad del ADN, el cual sirvió como indicador de la capacidad de interactuar de una determinada secuencia con una proteína.

Para cumplir los objetivos propuestos se utilizaron las secuencias de las regiones promotoras cajas TATA e islas CpG como modelos biológicos, es decir, se modelaron cadenas compuestas únicamente por pares de base con timina y adenina (cadenas tipo caja TATA) y cadenas compuestas únicamente por pares de base con guanina y citosina (cadenas tipo islas CpG). Todos los cálculos de estructura electrónica se hicieron bajo el marco de la Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés), utilizando como modelo químico Mo6-L/DNP.

1. Introducción



1.1	Ácido desoxirribonucleico	12
1.2	El ADN y su estructura química y tridimensional	15
1.3	La información “oculta” en el ADN	18
1.4	Planteamiento del problema y estado del arte	19
1.5	Objetivos e hipótesis	20

“Quisimos plantear una estructura para la sal de ácido desoxirribonucleico (ADN). Esta estructura tiene características novedosas que son de considerable interés biológico”

Rosalind Franklin

En este capítulo se hablará del ácido desoxirribonucleico (ADN), de su estructura e importancia biológica. Asimismo, se mencionarán y describirán las interacciones no-covalentes que participan en la estabilidad de la cadena, haciendo énfasis en aquellas que serán el eje central de este trabajo. Todo esto, con la finalidad de plantear el problema, los objetivos e hipótesis del proyecto.

1.1 Ácido desoxirribonucleico

“El ácido desoxirribonucleico (ADN) es el nombre químico de la molécula que contiene la información genética en todos los seres vivos. La molécula de ADN consiste en dos cadenas que se enrollan entre ellas para formar una estructura de doble hélice. Cada cadena tiene una parte central formada por azúcares (desoxirribosa) y un grupo fosfato. Cada azúcar se engancha a una base nitrogenada. Las dos cadenas se mantienen unidas por enlace entre las bases. La secuencia de estas bases a lo largo de la cadena es lo que codifica las instrucciones para formar proteínas y moléculas de ácido ribonucleico (ARN).”[1]

Christopher P. Austin, 2020

Seguramente en alguna reunión familiar has escuchado a tus tíos resaltar lo mucho que te pareces a tu mamá, papá o abuelos, mencionando de forma particular la(s) característica(s) de ellos que has *heredado*. Son justo esos rasgos, que se transmiten de generación en generación, los que despertaron el interés de los científicos y los que los encaminaron a buscar una explicación de la naturaleza y origen de esa transmisión de información genética.

Así como para toda las áreas científicas es difícil establecer sus orígenes, la genética y la biología molecular no son la excepción. Sin embargo, se cree que Charles Darwin fue uno de los precursores más importantes [2]. Derivado de sus observaciones en las islas Galápagos, en 1959, en su trabajo “El origen de las especies” [3], Darwin establece las bases de la evolución y de la selección natural, argumentando que las especies mejor adaptadas son aquellas capaces de sobrevivir, reproducirse y de transmitir esas características a su descendencia [2, 3].

Posteriormente, en 1866, Gregor Mendel establece, como resultado de sus experimentos de cruce de chícharos [4], los principios de la genética, resaltando la existencia de características hereditarias recesivas y dominantes. Mientras que, en 1869, Friedrich Miescher, al buscar las proteínas componentes de los leucocitos, descubre “accidentalmente” una nueva sustancia del núcleo celular con propiedades químicas muy distintas a la de las proteínas, incluyendo una mayor cantidad de fósforo en su estructura y una mayor resistencia a la proteólisis [5]. A dicho compuesto lo nombra nucleína, palabra antecesora de ácido nucleico y ácido desoxirribonucleico (ADN) [5].

A pesar de la relevancia de los descubrimientos de Mendel y de Miescher, sus trabajos permanecieron desapercibidos durante un buen tiempo, siendo retomados y valorados hasta los inicios del siglo XX. En 1902, Sir Archibald Edward Garrod, al estudiar el trastorno de la alcaptonuria, empleó las leyes de Mendel para explicar las bases genéticas de una enfermedad

hereditaria en humanos, publicando, por primera vez, un caso de herencia recesiva [2, 6, 7]. Más adelante, en 1919, Phoebus Levene determina los componentes químicos de los nucleótidos (grupo fosfato-azúcar-base nitrogenada), los cuales corresponden a las bases químicas estructurales del ADN; además, descubre la naturaleza del carbohidrato del ADN (desoxirribosa) [5, 8]. Todos estos antecedentes convergieron en el trabajo de Oswald Avery, quien en 1944 demuestra que el ADN es la base de los genes y, por lo tanto, la molécula hereditaria [5, 9, 10].

Teniendo una mejor idea de la composición química del ADN, y habiendo reconocido su importancia en el almacenamiento y transmisión de la información genética, parecía que el panorama de la genética y biología molecular estaba ya casi completo. Sin embargo, aún faltaban dos cuestiones, de suma importancia, por resolver. La primera de ellas era profundizar aún más en la composición química del ADN; pues, a pesar de tener conocimientos sobre la conformación química de los nucleótidos, aún no estaba claro cómo y en qué proporción estos interactuaban para dar lugar a la molécula del ADN. Aunada a esta pregunta sin contestar, la segunda cuestión residía en la estructura tridimensional del ácido desoxirribonucleico, dado que para poder explicar diversos procesos bioquímicos, era necesario saber qué distribución geométrica tenían los componentes del ADN para que éste pudiera participar en dichos mecanismos biológicos.

En 1950, Erwin Chargaff da uno de los pasos más importantes para responder a las interrogantes planteadas anteriormente. Al buscar las diferencias en la composición del ADN en función de la especie estudiada [11], Chargaff concluye que la conformación del ácido desoxirribonucleico es diferente entre las especies, es decir, la cantidad y secuencia de nucleótidos cambia según el organismo. Sin embargo, independientemente de la especie en cuestión, la cantidad de nucleótidos de adenina es la misma que la de nucleótidos de timina. De igual forma, el número de nucleótidos de guanina es el mismo que el de citosina [11, 12]. En pocas palabras, en cualquier ADN, la cantidad de purinas (adenina y guanina) es la misma que la cantidad de pirimidinas (timina y citosina) [5]. A estas proporciones se le conoce como regla de Chargaff y ésta jugó un papel fundamental en la determinación estructural del ADN.

Tuvieron que pasar dos años más para que en 1952 Rosalind Franklin, junto con Raymond Goslin [13], dieran la pieza fundamental para descubrir la estructura tridimensional del ADN. Trabajando en el área de la cristalografía de rayos X, en el King's College de Londres, Reino Unido, Rosalind logró obtener "fotografías" de alta resolución de fibras de ADN [2], siendo una de ellas la famosa "Fotografía 51". Con base en ellas, logró deducir que los fosfatos de esta molécula se encontraban hacia afuera de lo que pareciera, probablemente, una estructura helicoidal [13]. Sin consentimiento alguno, y como consecuencia de su mala relación, Maurice Wilkins, quien trabajaba en el mismo grupo de investigación y laboratorio, entregó dicha imagen a James Watson y Francis Crick [13], quienes también estaban en la búsqueda de la estructura del ADN, en el laboratorio de Cavendish en la Universidad de Cambridge.

Utilizando la Fotografía 51 de Rosalind Franklin, en 1953 Watson y Crick pusieron fin a la larga búsqueda de la estructura del ADN, concluyendo que ésta correspondía a una doble hélice, en la cual había bases nitrogenadas complementarias (adenina con timina y guanina con citosina) interactuando entre sí por medio de puentes de hidrógeno [14]. Además de que lograron corroborar, tal cual lo había propuesto Rosalind, que los grupos fosfatos se encontraban en las orillas de la hélice, por lo que entre el grupo fosfato y la base nitrogenada

se encontraba la desoxirribosa [14]. Tal fue el impacto de su descubrimiento que, en 1962 Watson, Crick y, sin mérito alguno, Wilkins fueron galardonados con el Premio Nobel de Medicina o Fisiología “por sus descubrimientos acerca de la estructura molecular de los ácidos nucleicos y su importancia para la transferencia de información en la materia viva” [15].

Sin duda alguna el hallazgo de Watson y Crick revolucionó tanto a la genética como a la biología molecular, dado que con el modelo de la doble hélice se lograron entender y explicar diversos mecanismos bioquímicos relacionados con la conservación y transferencia de la información genética, tales como la replicación, transcripción y traducción. Si bien el modelo de la doble hélice ha sufrido algunos cambios, éste fue un punto de partida crucial en la ciencia que ha permitido el desarrollo tanto de conceptos como de aplicaciones. Para poder hablar más detalladamente de esta molécula y poder explicar el interés que se tiene sobre ésta en el trabajo, será necesario hablar más tanto de su estructura geométrica (distribución tridimensional de los átomos que la conforman) como de su participación en la biología molecular.

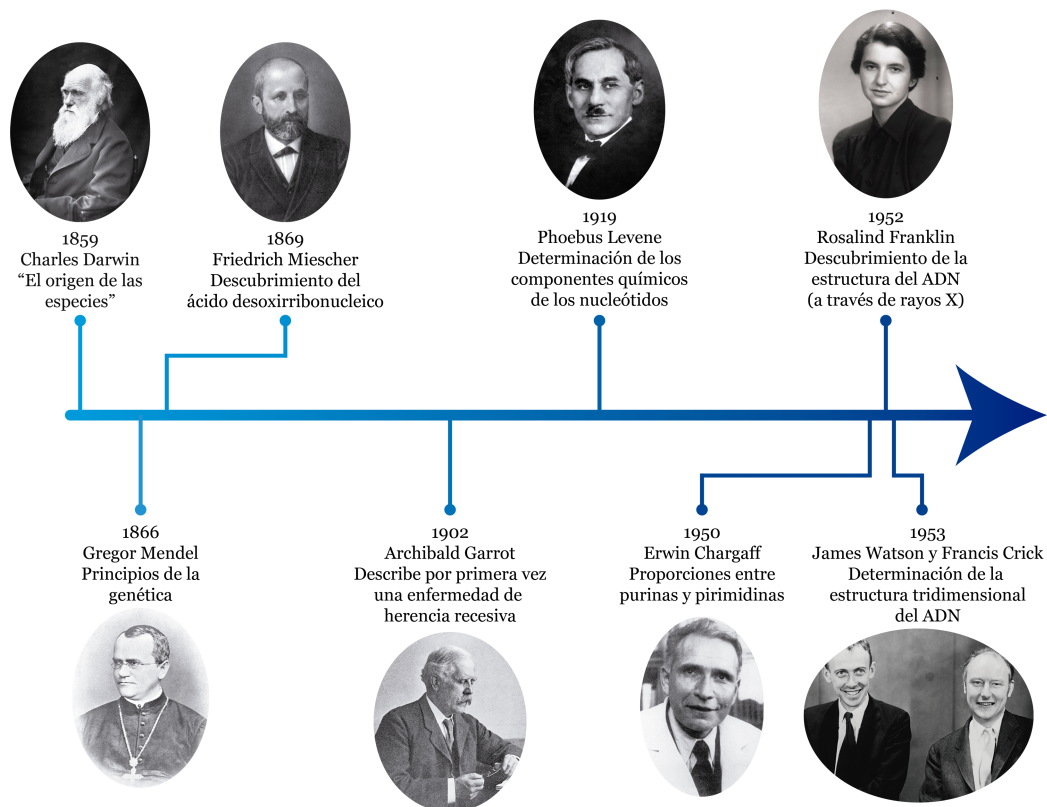


Figura 1.1. Línea del tiempo del descubrimiento de la estructura tridimensional del ADN. Todas las imágenes de los científicos involucrados fueron obtenidas de sitios web con licencia Creative Commons.

Un pequeño resumen de la historia del ADN, descrita anteriormente, está presentado en la Figura 1.1, donde se puede encontrar una línea del tiempo en la cual se resaltan las contribuciones más importantes que encaminaron el descubrimiento de la estructura tridimensional de dicha biomolécula.

1.2 El ADN y su estructura química y tridimensional

Para poder hablar de la estructura de la doble hélice es necesario, primero, recordar que ésta está formada por tres componentes químicos básicos (Figura 1.2): un grupo fosfato, un azúcar (desoxirribosa) y una base nitrogenada. Dichas unidades se encuentran enlazadas químicamente (el grupo fosfato con el azúcar y ésta, a su vez, con la base nitrogenada) para dar lugar a la formación de nucleótidos, los cuales son las bases estructurales para formar las hebras que conforman al dúplex. Es importante mencionar que existen cuatro tipos de bases nitrogenadas: adenina, timina, guanina y citosina (A, T, G y C, respectivamente), las cuales se pueden agrupar en dos conjuntos generales: purinas (adenina y guanina) y pirimidinas (timina y citosina).

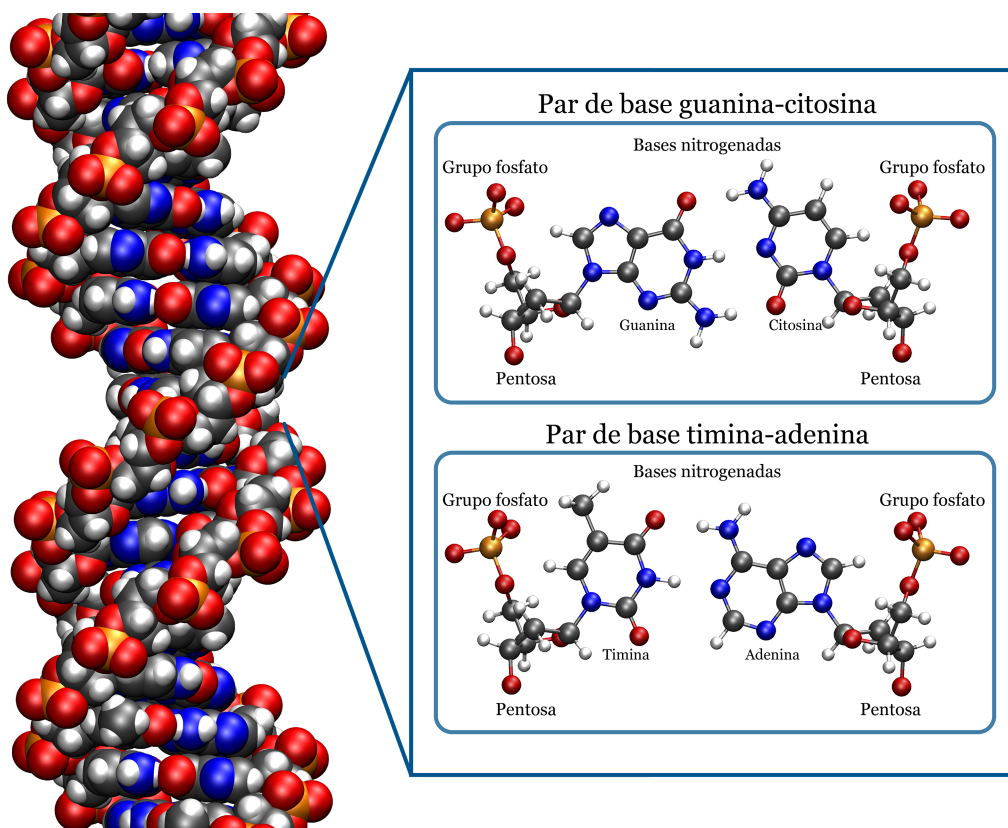


Figura 1.2. Estructura química del ADN. Una cadena de ADN está compuesta por pares de base, los cuales, a su vez, están conformados por un grupo fosfato, una pentosa (desoxirribosa) y una base nitrogenada.

Por otra parte, dos hebras de ADN interactúan entre sí principalmente a través de puentes de hidrógeno (Figura 1.2), dando lugar a la formación de la doble hélice. En función del ambiente celular se pueden obtener diferentes tipos de hélice (Figura 1.3) [16]:

- Conformación A (ADN-A): doble hélice encontrada en condiciones de deshidratación

- Conformación B (ADN-B): doble hélice encontrada en condiciones de homeóstasis
- Conformación Z (ADN-Z): doble hélice encontrada en condiciones de altas concentraciones de sales

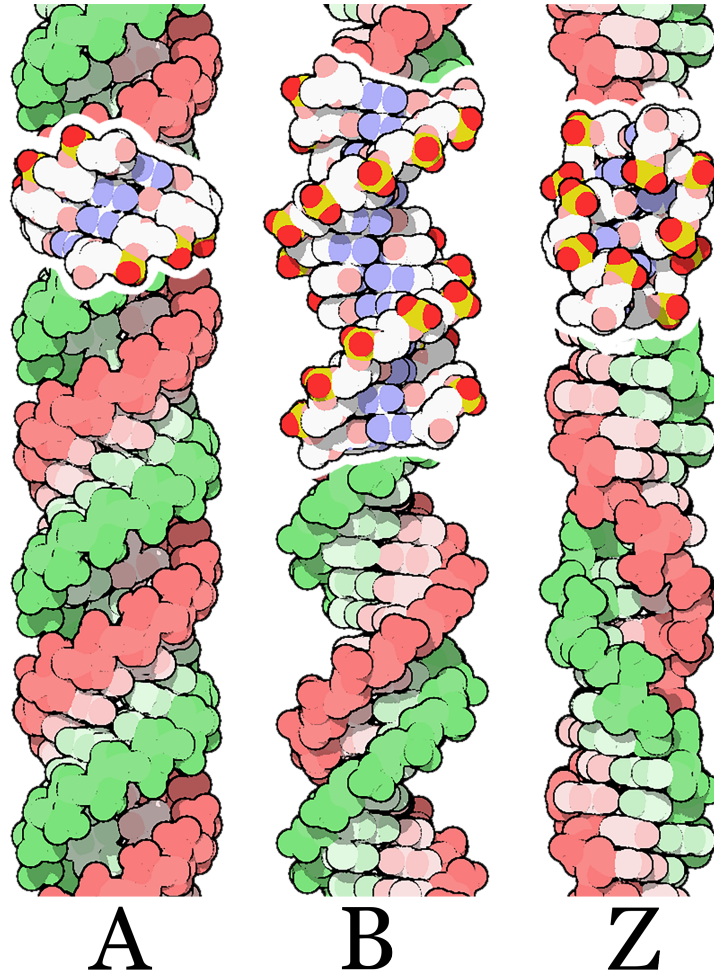


Figura 1.3. Diferentes conformaciones del ADN: ADN-A (izquierda), ADN-B (centro) y ADN-Z (derecha). Imágenes obtenidas de la página de *Protein Data Bank* [16].

Cada una de estas conformaciones presenta dos grados de libertad rotacionales y uno traslacional (Figura 1.4), los cuales son: *twist*, *roll* y *slide*, respectivamente [17]. La rotación *twist* corresponde a la rotación alrededor del eje de torsión vertical que atraviesa los centros de masa de dos pares de base vecinos [17], la rotación *roll* describe el ángulo de apertura de los pares de base a lo largo de sus ejes largos (es decir, al eje paralelo a los pares de base) [17] y, por último, la traslación *slide* compete al deslizamiento entre pares de base vecinos a lo largo de los ejes largos (eje similar a *roll*) [17].

A excepción del ADN-Z, las demás conformaciones del ADN tienen un giro a favor de las manecillas del reloj (Figuras 1.3 y 1.4). En función del ambiente celular y de la secuencia de nucleótidos, las interacciones de apilamiento π buscarán optimizarse y, en el mismo sentido, eliminar el contacto con las moléculas de agua del ambiente, pudiendo pasar de una

conformación tipo A a la tipo B, por ejemplo. Es importante aclarar que los pares de base no son totalmente planos, es decir, existe la posibilidad de que las bases nitrogenadas, que conforman a un par de base, giren sobre el eje largo, provocando que las dos bases no sean coplanares [18] y, por lo tanto, mejorando las interacciones de apilamiento π .

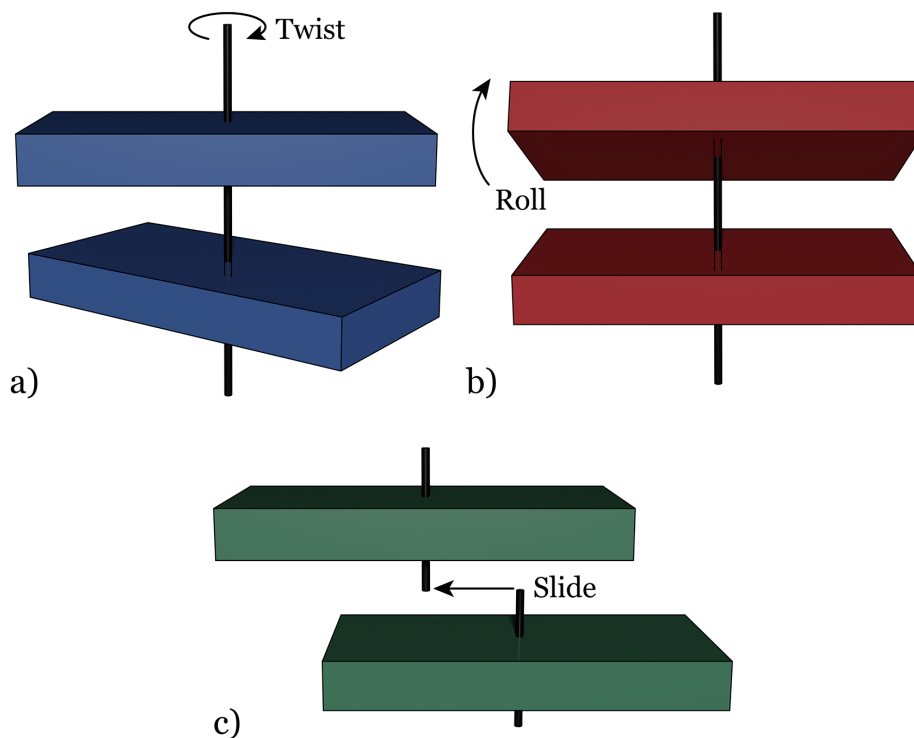


Figura 1.4. Rotaciones *twist* (a) y *roll* (b) y traslación *slide* (c) entre dos pares de base del ADN. Figura creada con base en la información de la referencia [17].

Dada la complejidad de la estructura del ADN, existe una gran cantidad de interacciones débiles a lo largo del dúplex; sin embargo, se cree que las interacciones de puente de hidrógeno y las de apilamiento π son las que participan de forma más importante en la estabilización de la cadena [19].

Generalmente, se asume que la conformación tipo B del ADN es la más común, por lo que muchos de los modelos se basan en esta estructura. En el ADN-B la cadena da un giro completo cada 10 pares de base, teniendo, en promedio, un *twist* de 36° , una rotación *roll* de 0° y un *slide* de 0 \AA [17]. Dicha conformación fue la encontrada por Watson y Crick, y es uno de los mejores modelos para describir diversos procesos biológicos, tales como la duplicación y la transcripción del ADN.

1.3 La información “oculta” en el ADN

Además de la estructura, la secuencia de bases nitrogenadas en el ADN es de suma importancia en diferentes procesos biológicos. Aproximadamente, cada célula de nuestro cuerpo tiene alrededor de 3 mil millones de nucleótidos que conforman al genoma de un organismo. Siendo el genoma toda la información genética guardada en el ADN.

A pesar de contener una gran cantidad de nucleótidos, no toda la información genética es expresada en forma de proteínas. Se estima que cerca del 1-2% de nucleótidos codifican para la producción de proteínas, llamando a esta porción de material genético ADN codificante [20]. Al 98-99% restante se le conoce como ADN no codificante, que si bien no producen proteínas, son capaces de producir otras moléculas (micro-ARN, por ejemplo), encargadas de la regulación genética [20, 21].

Las proteínas están involucradas en la mayoría de los procesos bioquímicos de un organismo, dado que son responsables del control y funcionamiento de estos mecanismos. Además, estas biomoléculas son responsables del fenotipo, es decir, de cómo se ve un individuo. Por lo que una correcta y adecuada producción de proteínas es indispensable para cualquier ser vivo.

Las secuencias de ADN que codifican para la producción de proteínas se conocen como genes. Un gen, de acuerdo al proyecto ENCODE, es la unión de secuencias genómicas que codifican un conjunto coherente de productos funcionales [22]. Dichas secuencias se encuentran en la región codificante del ADN y, a pesar de que son parte de una pequeña fracción de la cantidad total de nucleótidos, se estima que, por ejemplo, existen alrededor de 20,000 genes en un ser humano [23].

A grandes rasgos, un gen está conformado por una región promotora, una región codificadora y una región de terminación (Figura 1.5) [24]. De especial interés son las regiones promotoras, dado que éstas son identificadas por proteínas muy específicas, responsables de la formación del complejo de transcripción. Dos secuencias promotoras importantes son las cajas TATA (repeticiones de pares de base de T y A) [25] y las islas CpG (repeticiones de pares de base de G y C) [26]. Dichas regiones serán base de los modelos de este trabajo debido a la secuencia conformada por los mismos pares de base y a su relevancia biológica.



Figura 1.5. Cadena de ADN que representa la estructura básica de un gen, el cual está conformado por una región promotora (necesaria para el inicio de la transcripción), una región codificadora (la cual contiene a los nucleótidos que codifican para una proteína) y una región de terminación (donde se detiene la transcripción).

1.4 Planteamiento del problema y estado del arte

Dada la relevancia y complejidad del ADN, esta biomolécula ha sido, y sigue siendo, estudiada en diversas áreas de la ciencia, desde un enfoque biológico y fisicoquímico, hasta uno aplicado en la ciencia de los materiales. Asimismo, esta exploración aborda perspectivas tanto experimentales, así como múltiples modelos atomísticos. Todo esto con la finalidad de entender sus propiedades y, también, comprender su participación en diferentes procesos. De todas las formas en las que se puede estudiar al ADN, un común denominador de estos estudios son las interacciones no-covalentes que participan en la formación de la doble hélice. De especial interés son los puentes de hidrógeno y las interacciones de apilamiento π , puesto que, como se mencionó anteriormente, son las que participan de forma más importante en la constitución y estabilización de la estructura del ADN [19].

Un primer paso para entender el papel de las interacciones débiles en el ADN es mediante la estimación de la energía asociada a los puentes de hidrógeno y de apilamiento π , en función de las bases nitrogenadas que conforman a los pares de base. De esta manera, con base en las energías de interacción, será posible analizar la fuerza de interacción y, adicionalmente, las variaciones que hay en ésta debidas al medio en el que se encuentran las moléculas interactuantes.

Autor(es)	Aproximación/Modelo	Interacción estudiada
Asensio, et al. [27]	Rotación de bases nitrogenadas	Puentes de hidrógeno
Grunenberg [28]	Constantes elásticas	Puentes de hidrógeno
Dong, et al. [29]	Reemplazo por grupos isoelectrónicos	Puentes de hidrógeno
Espinosa, et al. [30]	Aproximación EML ^a (basada en la teoría de átomos en moléculas)	Puentes de hidrógeno
Szatyłowicz, et. al [31]	Análisis NBO ^b	Puentes de hidrógeno
Swart, et al. [32]	Diferencias de energía y análisis de descomposición de energía	Apilamiento π
Acosta-Silva, et al. [33]	Efectos del apilamiento π en puentes de hidrógeno	Puentes de hidrógeno y apilamiento π

Tabla 1.1. Principales modelos y aproximaciones utilizados por diferentes autores para la descripción y estimación de las interacciones de puente de hidrógeno y de apilamiento π . La mayoría de los trabajos toma en cuenta, únicamente, la presencia de dos bases nitrogenadas (azul) y, poco de ellos, la interacción entre dos pares de base (verde). ^a Espinosa-Molins-Lecomte (EML). ^b Aproximación de Orbital Natural de Enlace (NBO, por sus siglas en inglés).

Existen diversas aproximaciones y modelos para estimar la energía asociada a las interacciones de puente de hidrógeno y de apilamiento π . Algunas de ellas son listadas y explicadas en la Tabla 1.1. A pesar de que en varios casos se utilizan aproximaciones y métodos computacionalmente caros, cuando dichas aproximaciones no toman en cuenta la presencia de otros pares de base [34, 35], éstas ignoran los efectos cooperativos que pueden existir debido a la presencia de otras interacciones con/sin otros pares de base, teniendo, por lo tanto, un panorama limitado del comportamiento de estas interacciones y su impacto a nivel biológico.

Dado que un determinado gen o región genómica codifica para la producción de una proteína o algún tipo de ARN [36], es importante conocer cómo se ven influenciadas las interacciones de puente de hidrógeno y de apilamiento π en función del tamaño y la composición (secuencia) de la cadena. Por lo tanto, en este trabajo se buscará describir a dichas interacciones considerando tanto el efecto del tamaño de la cadena, así como el de las bases nitrogenadas que conforman a un par de base.

1.5 Objetivos e hipótesis

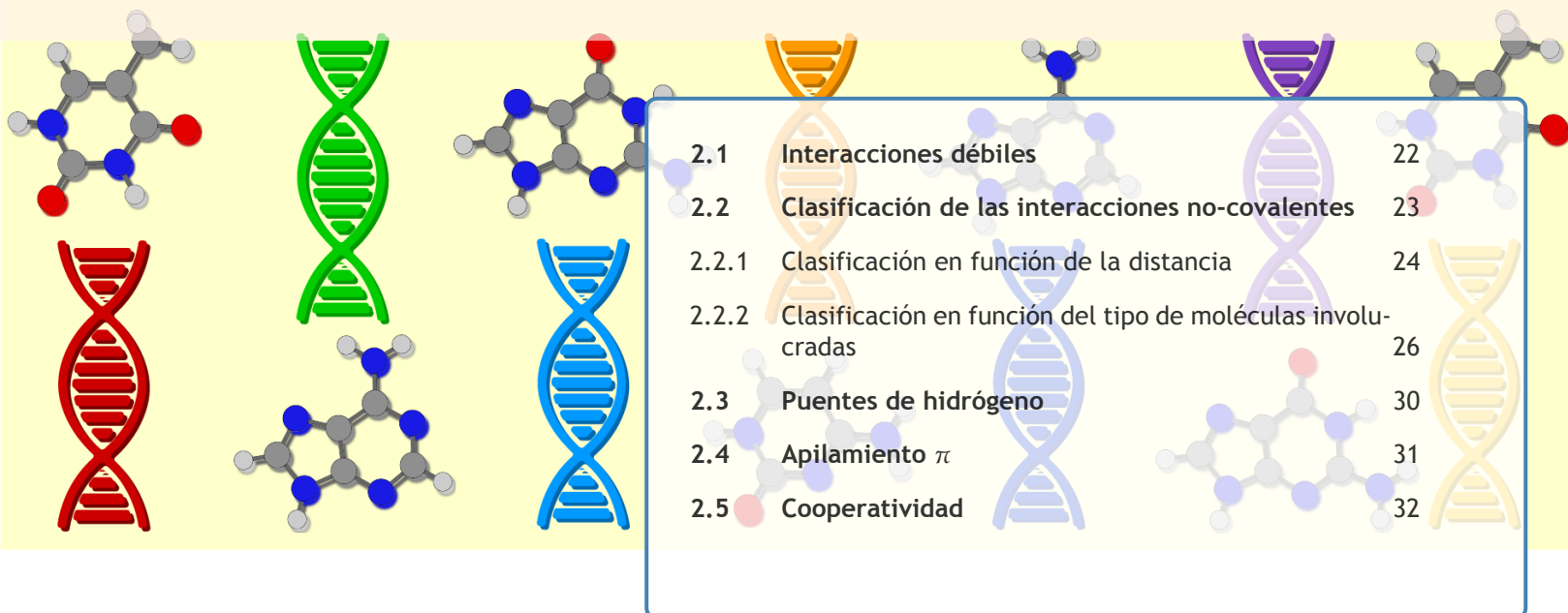
Debido a la importancia de las interacciones débiles en el ADN, es este proyecto se busca determinar el efecto del tamaño y composición de la cadena en las interacciones de puente de hidrógeno y de apilamiento π . Para tal fin, se utilizarán como modelos a las regiones promotoras caja TATA e islas CpG, y se estudiarán cadenas de diferente tamaño (distinto número de pares de base). Teniendo, por lo tanto, como hipótesis central del trabajo que: si las interacciones de puente de hidrógeno y de apilamiento π cambian en función del tamaño y composición de la cadena, entonces éstas, a su vez, influyen el comportamiento biológico de una región específica del ADN.

Un primer objetivo será describir las interacciones de puente de hidrógeno y de apilamiento π en el ADN, en función del tamaño (número de pares de base) y la composición (secuencia de nucleótidos) de la cadena. Una vez descritas estas interacciones, se buscará resolver qué interacción es más importante en el ADN, puesto que hay autores que afirman que los puentes de hidrógeno son más importantes, mientras que otros defienden que las interacciones de apilamiento, a pesar de ser más débiles, imperan sobre las de puente de hidrógeno.

Finalmente, puesto que se tendrá información tanto de los puentes de hidrógeno y del apilamiento π en cadenas tipo caja TATA y tipo islas CpG (recordando la relevancia de estas secuencias al ser regiones promotoras en el ADN) se pretende analizar si, en función de la secuencia de nucleótidos, dichas interacciones juegan un papel importante en la selectividad de las proteínas.

En el siguiente capítulo se mostrarán, brevemente, algunos conceptos necesarios en el entendimiento básico de las interacciones débiles, haciendo hincapié en los de puente de hidrógeno y el apilamiento π . Posteriormente, en otro capítulo se presentarán los resultados, análisis y conclusiones de las cadenas tipo caja TATA y tipo islas CpG estudiadas en este trabajo. Adicionalmente, en dos apéndices diferentes, se describen de manera concisa la Teoría de Perturbaciones Adaptada por Simetría (SAPT, por sus siglas en inglés) y de la Teoría de los Funcionales de la Densidad (DFT).

2. Marco teórico



2.1	Interacciones débiles	22
2.2	Clasificación de las interacciones no-covalentes	23
2.2.1	Clasificación en función de la distancia	24
2.2.2	Clasificación en función del tipo de moléculas involucradas	26
2.3	Puentes de hidrógeno	30
2.4	Apilamiento π	31
2.5	Cooperatividad	32

“La Química del siglo XX se trataba se interacciones intramoleculares [(enlace químico)]; [mientras que] la Química del siglo XXI se tratará de interacciones intermoleculares.”

Mark Ratner, 2004

Dado el interés que hay en este trabajo en las interacciones no-covalentes, en este capítulo se presentará una pequeña síntesis con información relevante de dichas fuerzas. Todo esto, con la finalidad de tener un mejor panorama del origen, características, clasificación y relevancia de las fuerzas intermoleculares. Es importante mencionar que se hará especial hincapié en las interacciones de puentes de hidrógeno y de apilamiento π .

2.1 Interacciones débiles

“La importancia de las fuerzas intermoleculares en la naturaleza es muy difícil de sobreestimar. Basta decir que la existencia de líquidos y sólidos se debe a interacciones intermoleculares. En ausencia de interacciones intermoleculares, nuestro mundo sería un gas ideal.”[37]

Ilya G. Kaplan, 2006

Conocidas como interacciones no-covalentes o fuerzas intermoleculares, estas interacciones débiles están presentes en moléculas interactuantes de distintos tipos que se encuentran a cierta distancia. Generalmente, se piensa en estas interacciones como fuerzas atractivas; sin embargo, para tener un panorama más amplio, es necesario tomar en cuenta que estas interacciones también pueden involucrar fuerzas repulsivas. Todo estará en función de la distancia a la cual se encuentren las moléculas.

Una de las mejores perspectivas, relacionadas a las interacciones intermoleculares, es aquella que nos proporciona el potencial semiempírico de Lennard-Jones, el cual tiene la siguiente forma:

$$U^{LJ}(R) = 4\varepsilon \left[\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right] \quad (2.1)$$

cuya gráfica se muestra en la Figura 2.1. Este potencial muestra la dependencia de la interacción entre partículas en función de la distancia R a la cual se encuentren separadas. Como se puede observar, a distancias grandes imperan las fuerzas atractivas (asociadas a la disminución del potencial), mientras que a distancias cortas dominan los efectos repulsivos (debidos al crecimiento rápido de la energía potencial). En dicho potencial, la distancia de equilibrio R_m corresponde a la distancia en donde ocurre la mayor atracción entre las moléculas, antes de que los efectos repulsivos comiencen a dominar (es decir, las fuerzas atractivas y repulsivas están en equilibrio). A esta distancia le corresponde un pozo de potencial atractivo, ε , el cual también está asociado a la energía de interacción (la energía necesaria para romper la atracción entre las moléculas involucradas). Finalmente, a una corta distancia, σ , las fuerzas de atracción y repulsión se cancelan entre sí y, por lo tanto, en esta zona la energía potencial de interacción es cero.

Todos los parámetros involucrados en la Ecuación (2.1) pueden ser aproximados experimental y/o computacionalmente, tanto para describir la interacción entre moléculas, como para describir la interacción entre átomos que conforman a una molécula. La diferencia principal entre estos dos tipos de interacciones será la profundidad del mínimo de potencial ε : en el caso de interacciones intramoleculares (enlaces químico), ε puede ser superior a los 200 kJ/mol; mientras que para interacciones intermoleculares, ε puede variar entre 1 – 22 kJ/mol [38]. Por tal motivo, las interacciones no-covalentes se consideran como interacciones débiles. Sin embargo, a pesar de ser débiles, las fuerzas intermoleculares son responsables de distintas propiedades termodinámicas, así como de los distintos estados de la materia.

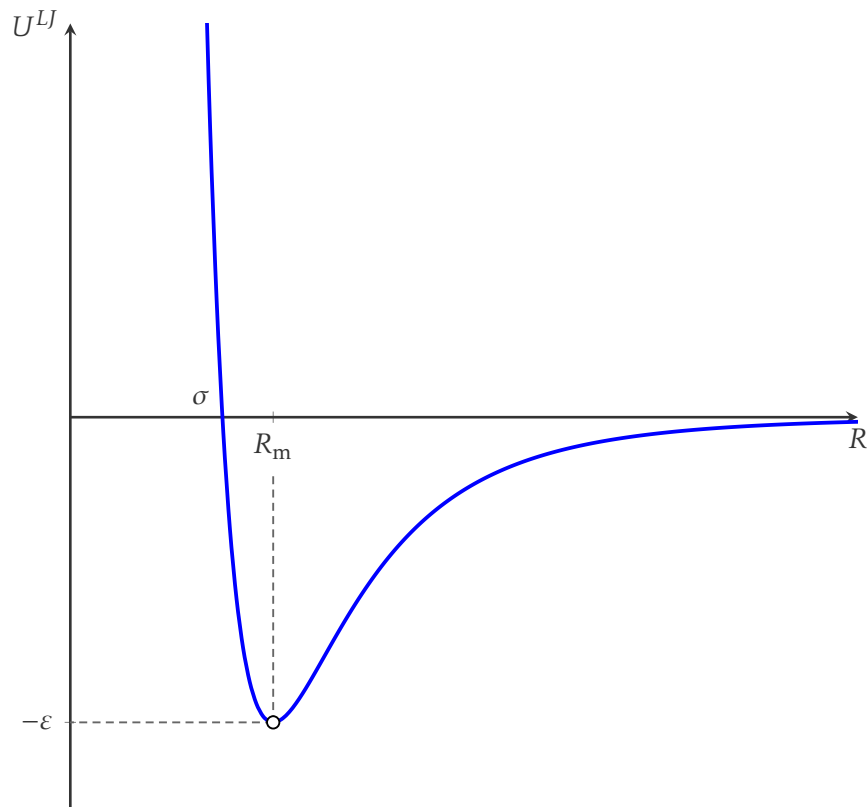


Figura 2.1. Ejemplo de la forma del potencial de Lennard-Jones para interacciones intermoleculares.

Pese a que era obvia la existencia de fuerzas (atractivas y/o repulsivas) entre moléculas interactuantes, no fue hasta 1873 que se demuestra la relevancia de las interacciones intermoleculares. Esto fue logrado por Johannes van der Waals, al proponer una ecuación para un gas real:

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \quad (2.2)$$

En dicha ecuación, van der Waals considera una desviación del comportamiento del gas ideal, al asumir que las moléculas cuentan con un volumen incompresible b (es decir, las moléculas no se comportan como partículas puntuales) y que, además, entre éstas existe una atracción continua, tomada en cuenta en la constante a . Por tal razón, a las interacciones débiles entre moléculas, generalmente, se les suele llamar también como fuerzas de van der Waals.

2.2 Clasificación de las interacciones no-covalentes

De forma general, las interacciones no-covalentes se pueden describir en función de la distancia a la que se encuentren las moléculas interactuantes. Adicionalmente, las interacciones intermoleculares pueden clasificarse en función de las moléculas y/o iones involucrados. En las siguientes secciones se hablará de forma general de dicha organización. Para consultar información más detallada, se recomienda revisar las referencias [37] y [39].

2.2.1 Clasificación en función de la distancia¹

En la Figura 2.2 se enlistan los diferentes tipos de interacciones intermoleculares, así como sus principales contribuciones, en función de la distancia existente entre las moléculas que interactúan entre sí. A grandes rasgos, éstas pueden ser clasificadas con base en los tres rangos de separación interatómica dados por cualquier potencial interatómico.

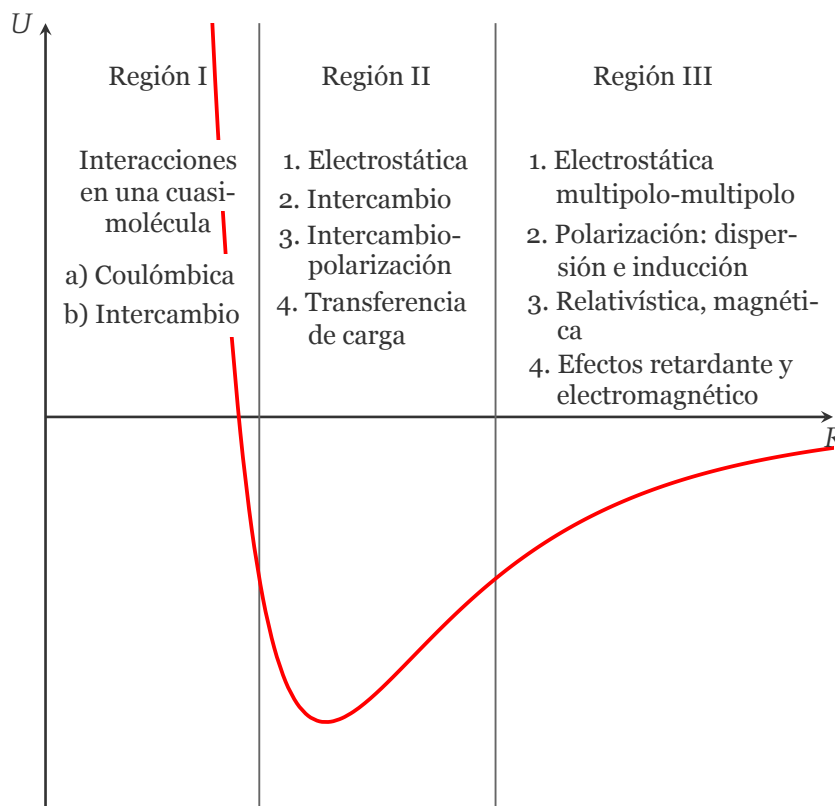


Figura 2.2. Clasificación de las interacciones intermoleculares en función de la distancia. Imagen realizada con base en la Figura 1.4 de la referencia [37].

Región I: interacciones de corto alcance

En esta región, los átomos o moléculas interactuantes pierden su individualidad, debido al gran traslape de sus capas electrónicas. La energía de interacción E_{int} en esta región está descrita por la siguiente ecuación:

$$E_{int} = E_{tot} - \sum_{a=1}^n E_a \quad (2.3)$$

donde E_a corresponde a la energía de los subsistemas aislados, mientras que E_{tot} describe la energía total del sistema interactuante completo. Debido a las distancias tan cortas, las fuerzas coulómbicas repulsivas (E_{Coul}) y el intercambio electrónico (E_{ex}), asociado al principio

¹El texto presentado en esta sección fue desarrollado con base en la información proporcionada por la referencia [37]

de antisimetría de Pauli, dominan, por lo que la energía de interacción puede ser dividida en estas dos contribuciones:

$$E_{int} = E_{Coul} + E_{ex} \quad (2.4)$$

Región II: interacciones de alcance intermedio

A una distancia intermedia (Región II de la Figura 2.2), las fuerzas repulsivas y atractivas están en equilibrio, lo cual provoca la existencia de un mínimo en la energía potencial intermolecular y, además, provee de estabilidad al sistema. Dado que la energía de interacción intermolecular es mucho menor a la energía de interacción intramolecular (enlace químico), la teoría de perturbaciones puede ser utilizada para estimar la energía de interacción (E_{int})². Con base en la Teoría de Perturbaciones Adaptada por Simetría³ (SAPT, por sus siglas en inglés), E_{int} puede ser descompuesta en una serie de perturbaciones escrita como:

$$E_{int} = \varepsilon_{el}^{(1)} + \varepsilon_{ex}^{(1)} + \sum_{n=2}^{\infty} \varepsilon_{pol.ex}^{(n)} \quad (2.5)$$

donde $\varepsilon_{el}^{(1)}$ corresponde a la energía de interacción electrostática clásica entre dos o más distribuciones de carga (densidades electrónicas); $\varepsilon_{ex}^{(1)}$ a la energía de intercambio de primer orden de la teoría de perturbaciones, cuyo origen se basa en el principio de antisimetría de Pauli: debido a la antisimetrización de las funciones de onda, los electrones tienen probabilidad de estar deslocalizados en todas las moléculas interactuantes; y, finalmente, el tercer término corresponde a términos de orden superior de SAPT, en los cuales los efectos de intercambio no pueden ser separados de la energía de polarización⁴.

Región III: interacciones de largo alcance

Finalmente, a distancias muy grandes (Región III de la Figura 2.2), la energía electrónica de intercambio ($\varepsilon_{ex}^{(1)}$) es despreciable (al no haber traslape o interacción de nubes de densidad electrónica) y las fuerzas intermoleculares son siempre atractivas. Generalmente, esto es válido a $R \gtrsim 15a_0$.

Con base en la teoría de perturbaciones, la energía para las interacciones de largo alcance puede ser expresada en función de las contribuciones de inducción y dispersión de diferentes órdenes ($\varepsilon_{ind}^{(n)}$ y $\varepsilon_{disp}^{(n)}$, respectivamente), así como de términos de órdenes superior cruzados ($\varepsilon_{ind.disp}$):

$$E_{int} = \varepsilon_{el}^{(1)} + \sum_{n=2}^{\infty} \left[\varepsilon_{ind}^{(n)} + \varepsilon_{disp}^{(n)} \right] + \sum_{n=3}^{\infty} \varepsilon_{ind.disp}^{(n)} \quad (2.6)$$

²La precisión que se alcanza con esta teoría permite distinguir entre fuerzas intra e intermoleculares, lo cual se requiere para describir a las Regiones I y II de la Figura 2.2.

³Esta teoría es explicada con mayor detalle en el Apéndice A.

⁴La polarización corresponde a la energía mínima necesaria para deformar la densidad electrónica de una molécula, debido a una perturbación externa, tal como lo puede ser un campo eléctrico provocado por la presencia de otra molécula, por ejemplo.

donde el término de energía de polarización de segundo orden perturbativo es dividido en contribuciones de energía de inducción $\varepsilon_{ind}^{(2)}$ y energía de dispersión $\varepsilon_{disp}^{(2)}$. En órdenes superiores, aparecen términos mixtos $\varepsilon_{ind.disp}$ los cuales resultan del acoplamiento de dichas energías.

El primer término de la Ecuación (2.6), corresponde, como se mencionó anteriormente, a la energía de interacción electrostática. Cuando las moléculas interactuantes se encuentran a distancias grandes, $\varepsilon_{el}^{(1)}$ puede ser descrita como una expansión de multipolos, lo cual, como consecuencia, representa la energía de una interacción directa entre multipolos.

Por otro lado, mientras que la energía de inducción $\varepsilon_{ind}^{(2)}$ se refiere a la inducción de algún momento multipolar debido a un campo eléctrico (generado por otro momento multipolar, por ejemplo) y puede ser estimada de forma clásica, la energía de dispersión $\varepsilon_{disp}^{(2)}$ es un fenómeno asociado a la correlación electrónica y, por lo tanto, netamente cuántico. Este último término se refiere a las fluctuaciones en la densidad electrónica que puedan provocar la existencia de un momento dipolo instantáneo, el cual inducirá un momento dipolo, u otros momentos de órdenes superiores, en otra molécula. Cabe destacar que los multipolos instantáneos surgen del movimiento de los electrones que minimizan la distancia entre ellos. Con base en la teoría de perturbaciones, $\varepsilon_{disp}^{(2)}$ puede ser expresada en términos de una expansión de multipolos:

$$\varepsilon_{disp}^{(2)} = - \sum_{n=6}^{\infty} \frac{C_n}{R^n} \quad (2.7)$$

donde C_n corresponde a los coeficientes de dispersión, obtenidos de métodos mecano-cuánticos sofisticados.

Por último, las contribuciones magnéticas y relativistas (incluyendo el efecto retardante) serán de importancia, principalmente, en sistemas donde las moléculas involucradas tengan electrones desapareados y, adicionalmente, en aquellos sistemas donde el tiempo de propagación de una interacción sea del orden de R/c (donde c corresponde a la velocidad de la luz).

2.2.2 Clasificación en función del tipo de moléculas involucradas⁵

A pesar de que las interacciones intermoleculares se pueden clasificar con base en la distancia que existe entre las moléculas interactuantes, esta clasificación puede ser un tanto arbitraria, pues los rangos de separación dependerán de las moléculas involucradas. Además, esta clasificación depende de la forma del potencial intermolecular analizado.

Otra forma de clasificar a las interacciones débiles es en función de las moléculas participantes en la interacción. Para esto, el potencial se describe, principalmente, como una interacción coulombica. En esta sección, se analizará las posibles interacciones entre moléculas de capa cerrada, recordando que las fuerzas de van der Waals se refieren a las fuerzas atractivas que hay entre las partículas que conforman a un sistema. Se asume que la parte repulsiva se debe principalmente a repulsiones coulombicas. Centrándonos, por lo tanto, en

⁵El texto presentado en esta sección fue desarrollado con base en la información proporcionada por la referencia [39]

la Región II de la Figura 2.2.

Ión-dipolo

La interacción entre un dipolo puntual⁶, $\mu_1 = Q_1 l$ (donde l es la distancia entre las cargas en un dipolo), y una carga puntual, Q_2 , está dada por la siguiente ecuación:

$$V = -\frac{\mu_1 Q_2}{4\pi\epsilon_0 r^2} \quad (2.8)$$

A diferencia de la interacción entre dos iones (cargas puntuales), cuya energía es proporcional a $1/r$ (donde r corresponde a la separación entre las partículas interactuantes), la interacción ión-dipolo tiende más rápidamente a cero (a grandes distancias), debido a que, a valores considerablemente grandes de r , las cargas parciales en un dipolo tienden a “fusionarse”, por lo que se cancelan entre sí.

Es importante mencionar que este tipo de interacción, así como de las que se hablará más adelante, puede ser descrita utilizando multipolos de órdenes superiores⁷. Sin embargo, la energía de interacción decae más rápidamente mientras más grande sea el orden del multipolo. La energía potencial de la interacción V , en función de la distancia, entre un momento n -polar con otro momento m -polar varía de la siguiente forma:

$$V \propto \frac{1}{r^{n+m-1}} \quad (2.9)$$

Al igual que con el momento dipolo, mientras mayor sea la distancia a la cual se observa un momento multipolar, las cargas individuales tenderán a mezclarse y, por lo tanto, a neutralizarse entre sí.

Dipolo-dipolo

La energía potencial de interacción entre dos moléculas polares (es decir, entre dos dipolos) estará en función de la orientación de los momentos dipolo. Cuando los dipolos sean paralelos entre sí, la energía de interacción estará descrita de la siguiente forma:

$$V = \frac{\mu_1 \mu_2 f(\theta)}{4\pi\epsilon_0 r^3} \quad f(\theta) = 1 - 3 \cos^2 \theta \quad (2.10)$$

donde μ_1 y μ_2 representan a los momentos dipolo de las moléculas polares participantes en la interacción, mientras que $f(\theta)$ a la orientación que hay entre éstas. Algo importante que mencionar sobre la Ecuación (2.10) es que ésta solamente es válida cuando las moléculas polares se encuentran en una posición fija (por ejemplo, en un sólido). Sin embargo, cuando las moléculas se mueven relativamente libres en un fluido, la interacción entre dipolos se

⁶Llamado así dado que la distancia (l) a la cual se encuentran separadas las cargas en un dipolo es mucho menor que la distancia (r) a la cual se observa el dipolo ($l \ll r$).

⁷Un n -polo es un arreglo de cargas puntuales con un momento n -polar, pero sin momento inferior. De tal forma, un monopolio ($n = 1$) corresponde a la carga neta de una carga puntual, un dipolo ($n = 2$) a un arreglo de cargas que no tiene monopolio (carga neta), un cuadrupolo ($n = 3$) a un arreglo que no tiene ni carga neta ni momento dipolo, un octupolo ($n = 4$) a un conjunto de cargas cuya suma es cero y, adicionalmente, no tiene ni momento dipolo ni cuadrupolo, y así sucesivamente.

describe de la siguiente forma:

$$\langle V \rangle = -\frac{C}{r^6} \quad C = \frac{2\mu_1^2\mu_2^2}{3(4\pi\epsilon_0)^2 k_b T} \quad (2.11)$$

donde μ_1 y μ_2 corresponden a los dipolos de las moléculas interactuantes, ϵ_0 a la permitividad al vacío, k_B a la constante de Boltzmann y T a la temperatura. La sexta potencia inversa en la Ecuación (2.11) surge de la tercera potencia inversa de la energía potencial de interacción (Ecuación 2.10), ponderada por la distribución de Boltzmann, la cual también es proporcional a $1/r^3$. Es importante resaltar que la relación inversa de la temperatura refleja la forma en la que los efectos térmicos superan a los efectos de orientación de los dipolos a altas temperaturas; por lo que disminuye la interacción.

Dipolo-dipolo inducido

Una molécula polar con un momento dipolo μ_1 puede inducir un momento dipolo μ_2^* a una molécula (no polar) que sea polarizable y que se encuentre cercana a ella. El dipolo inducido interactuará con el dipolo permanente, por lo que ambas se atraerán. La energía de interacción entre una molécula polar y otra no polar, que se encuentran a una distancia r , está dada por:

$$V = -\frac{C}{r^6} \quad C = \frac{\mu_1^2\alpha_2'}{4\pi\epsilon_0} \quad (2.12)$$

donde α_2' es el volumen de polarizabilidad⁸ de la molécula 2. Es relevante notar que la C de esta expresión es diferente a la que aparece en la Ecuación (2.11).

La energía de interacción dipolo-dipolo inducido es independiente de la temperatura, dado que la orientación de las moléculas no afecta dichas interacciones. Al igual que en las interacciones dipolo-dipolo, la energía potencial depende de $1/r^6$; sin embargo, en las interacciones dipolo-dipolo inducido, ésta surge de la dependencia en r del campo eléctrico \mathcal{E} generado por un dipolo (Ecuación (2.13)) y de la dependencia en $1/r^3$ de la energía potencial de la interacción entre dipolos permanentes e inducidos.

$$\mathcal{E} = \frac{\mu}{2\pi\epsilon_0 r^3} \quad (2.13)$$

Dipolo instantáneo-dipolo inducido⁹

Las moléculas no polares pueden atraerse entre sí, a pesar de no tener asociado un momento dipolo permanente. La interacción entre este tipo de moléculas ocurre como consecuencia de los dipolos instantáneos que todas las moléculas poseen como resultado de las fluctuaciones en las posiciones instantáneas de los electrones.

Cuando los electrones en una molécula cambian de posición pueden generar un dipolo instantáneo μ_1^* . Este dipolo generará un campo eléctrico que, a su vez, polarizará a una segunda molécula, induciendo un momento dipolar μ_2^* . De esta forma, ambas moléculas se

⁸La polarizabilidad es la capacidad de una molécula de redistribuir su densidad electrónica en respuesta a la aplicación de un campo eléctrico.

⁹Estas interacciones también podrían ser clasificadas como interacciones dipolo instantáneo-dipolo instantáneo dado que no se puede distinguir en qué molécula se originó primero el dipolo instantáneo.

podrán atraer entre sí. Independientemente de que en la primera molécula haya un cambio en el tamaño y/o dirección del dipolo instantáneo, la densidad electrónica de la segunda molécula también sufrirá modificaciones. Por lo tanto, se dice que los momentos dipolo están correlacionados. A este tipo de interacción también se le suele referir como interacciones de dispersión o interacciones de London.

Es importante mencionar que las fuerzas de dispersión dependen de la polarizabilidad de la primera molécula, dado que la generación del dipolo instantáneo μ_1^* dependerá de qué tan fácil es alterar la densidad electrónica de ésta. En el mismo sentido, la interacción también dependerá de la polarizabilidad de la segunda molécula, puesto que ésta determinará su respuesta frente al campo eléctrico generado por μ_1^* . Todo esto se resume en la siguiente ecuación, la cual es conocida como la ecuación de London, que permite conocer la fuerza de interacción de dispersión:

$$V = -\frac{C}{r^6} \quad C = \frac{3}{2} \alpha_1' \alpha_2' \frac{I_1 I_2}{I_1 + I_2} \quad (2.14)$$

donde I_1 , I_2 , α_1' y α_2' son las energías de ionización y volúmenes de polarización de cada molécula, respectivamente. Al igual que las interacciones descritas anteriormente, la energía de las fuerzas de dispersión es proporcional a $1/r^6$; sin embargo, estas interacciones suelen ser las más débiles puesto que dependen de los dipolos instantáneos e inducidos, los cuales suelen tener asociadas magnitudes pequeñas. Finalmente, es relevante reconocer que estas contribuciones están, generalmente, presentes en todas las interacciones entre moléculas, por lo que es un término atractivo importante.

Reuniendo a todas las interacciones

En la Tabla 2.1 se muestran a modo de resumen los tipos de interacciones descritos anteriormente, así como la dependencia de la energía potencial en la distancia entre las moléculas. Como se puede observar, para todas las interacciones¹⁰, la energía potencial es proporcional a $1/r^6$. Por lo que, en una mezcla de moléculas de diferentes tipos, todas estas interacciones estarán presentes. Cuando un sistema esté conformado únicamente por moléculas no polares, la única contribución a la energía de interacción será aquella proveniente de las fuerzas de dispersión. Por lo tanto, para un fluido con diferentes tipos de moléculas, la energía potencial intermolecular puede ser escrita como:

$$V = -\frac{C_6}{r^6} \quad (2.15)$$

donde C_6 es un coeficiente de interacción que depende del tipo de moléculas interactuantes. Cabe destacar que la Ecuación (2.15) es parecida al segundo término (atractivo) de la Ecuación (2.1), por lo que el tipo de interacciones influenciará la parte atractiva del potencial intermolecular.

Es relevante resaltar que, hasta ahora, se ha hablado de las interacciones entre momentos dipolo. Para tener una mejor descripción de éstas, es importante considerar multipolos de órdenes superiores (tal como fue discutido en la sección anterior). Adicionalmente, sólo fue considerada la interacción entre pares, es decir, entre dos moléculas, por lo que una mejor estimación de la energía de interacción será aquella que considere la interacción entre varios

¹⁰Para las interacciones dipolo-dipolo esto se cumple cuando las moléculas polares pueden rotar libremente.

cuerpos.

Tipo de interacción	Dependencia en r de la energía potencial	Energías típicas (kJ/mol)	Características
Ión-ión	$1/r$	250	Únicamente entre iones
Ión-dipolo	$1/r^2$	15	
Dipolo-dipolo	$1/r^3$	2	Entre moléculas polares con posiciones fijas
	$1/r^6$	0.6	Entre moléculas polares rotantes
Dispersión	$1/r^6$	2	Entre todos los tipos de moléculas

Tabla 2.1. Resumen del tipo de interacciones intermoleculares y de la dependencia en r de la energía potencial. Tabla generada a partir de la información de la Tabla 17.3 de la referencia [39].

2.3 Puentes de hidrógeno

Los puentes de hidrógeno son considerados como un tipo “especial” de interacciones no covalentes, dado que, por sus características, estos no pueden ser clasificados como interacciones meramente electrostáticas (interacciones entre momentos n-polares) o de dispersión (efectos cuánticos), como las interacciones descritas en la sección anterior. De hecho, algunos autores afirman que las contribuciones electrostáticas y de dispersión en los puentes de hidrógeno son muy pequeñas, mientras que las contribuciones cuánticas de transferencia de carga (traslape de nubes de densidad electrónica) son las que, principalmente, dominan dicha interacción [40, 41]. En contraste, otros investigadores proponen que existe un balance de diferentes contribuciones (tales como las electrostáticas, dispersión, transferencia de carga, repulsiones estéricas, efectos cooperativos, entre otras) para la formación de un puente de hidrógeno [42].

Una primera definición para estas interacciones fue propuesta por G. N. Lewis, quien define a los puentes de hidrógeno como “un átomo de hidrógeno que puede estar unido, al mismo tiempo, a dos pares de electrones de dos átomos diferentes, actuando como un enlace débil entre estos dos átomos” [43]. Posteriormente, Pauling describe a estas interacciones como de carácter iónico y especifica que, para que éstas se formen, es necesario que participen átomos muy electronegativos (como el N, O y F) [44].

Puesto que no siempre están involucrados átomos electronegativos en la formación de los puentes de hidrógeno, una mejor y más moderna definición es aquella propuesta por Weinhold y Klein, donde un puente de hidrógeno es “un enlace químico fraccional entre ácidos y bases de Lewis $A-H \cdots B$ (que involucra hidrógeno), asociado a una resonancia parcial intermolecular $A-H \cdots :B \longleftrightarrow A : \cdots H-B$ comúnmente originado en la interacción entre un par libre n_B de una base de Lewis y un híbrido de antienlace σ_{AH}^* de un ácido de Lewis” [40]. Adicionalmente, a este tipo de interacciones se le suele asociar una fuerza de interacción de

entre 0.6 a 45 kcal/mol [40].

Los puentes de hidrógeno son de sumo interés en diversas áreas de la ciencia (como en la ciencia e ingeniería de los materiales, así como en el área de la biología, por ejemplo) debido a su impacto en las propiedades físicas y químicas de un sistema. En la biología molecular, estas interacciones son, en gran parte, responsables del reconocimiento molecular (interacción entre proteínas y su molécula blanco), así como de la estabilidad estructural de la doble hélice de ADN. En este trabajo, las interacciones de puente de hidrógeno en el ADN serán parte fundamental del objeto de estudio.

2.4 Apilamiento π

Otro tipo de interacción “especial” es aquella que se presenta entre anillos aromáticos, la cual es conocida como apilamiento π . En los últimos años, esta interacción no-covalente ha ganado bastante atención debido a su presencia en múltiples sistemas cristalinos, así como en complejos supramoleculares.

Las conformaciones más comunes para el apilamiento π se muestran en la Figura 2.3. De éstas, los anillos aromáticos pequeños no sustituidos prefieren una conformación tipo T, mientras que los anillos sustituidos más grandes prefieren una geometría paralela desplazada¹¹. En cuanto a la configuración paralela cofacial, muy pocas veces es vista entre anillos aromáticos [46].

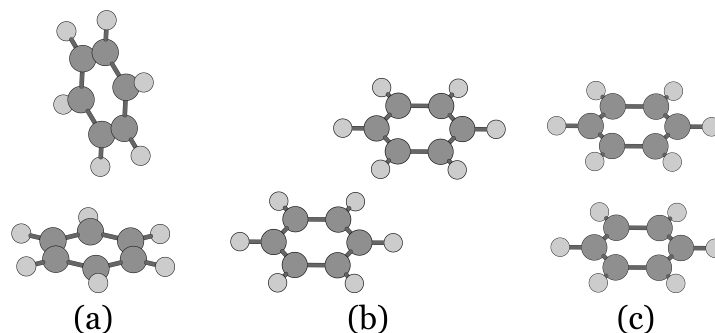


Figura 2.3. Conformaciones más probables para el apilamiento de anillos aromáticos: (a) forma de T, (b) paralela desplazada, (c) paralela cofacial.

Inicialmente, se catalogaba a estas interacciones dentro de las fuerzas de dispersión; sin embargo, éstas solían tener energías más grandes que aquellas asociadas a las interacciones de London (la energía del apilamiento π ronda de entre 1 a 15 kcal/mol, aproximadamente [47]). El origen de estas interacciones se cree que es, principalmente, electrostático (por la interacción entre momentos cuadrupolares) y de dispersión [47-49]. Sin embargo, también se ha demostrado que, además de estas contribuciones, la repulsión de Pauli y la interacción

¹¹Aunque existe evidencia que muestra que estas configuraciones están muy cercanas en energía [45]

intermolecular de orbitales, también juegan un papel importante en la formación de interacciones de apilamiento π [50].

En este trabajo, estas interacciones, al igual que los puentes de hidrógeno, serán objeto de estudio dada su relevancia en la conformación de la doble hélice del ADN, así como su posible participación en el reconocimiento molecular entre enzimas y regiones específicas dentro de la doble hélice.

2.5 Cooperatividad

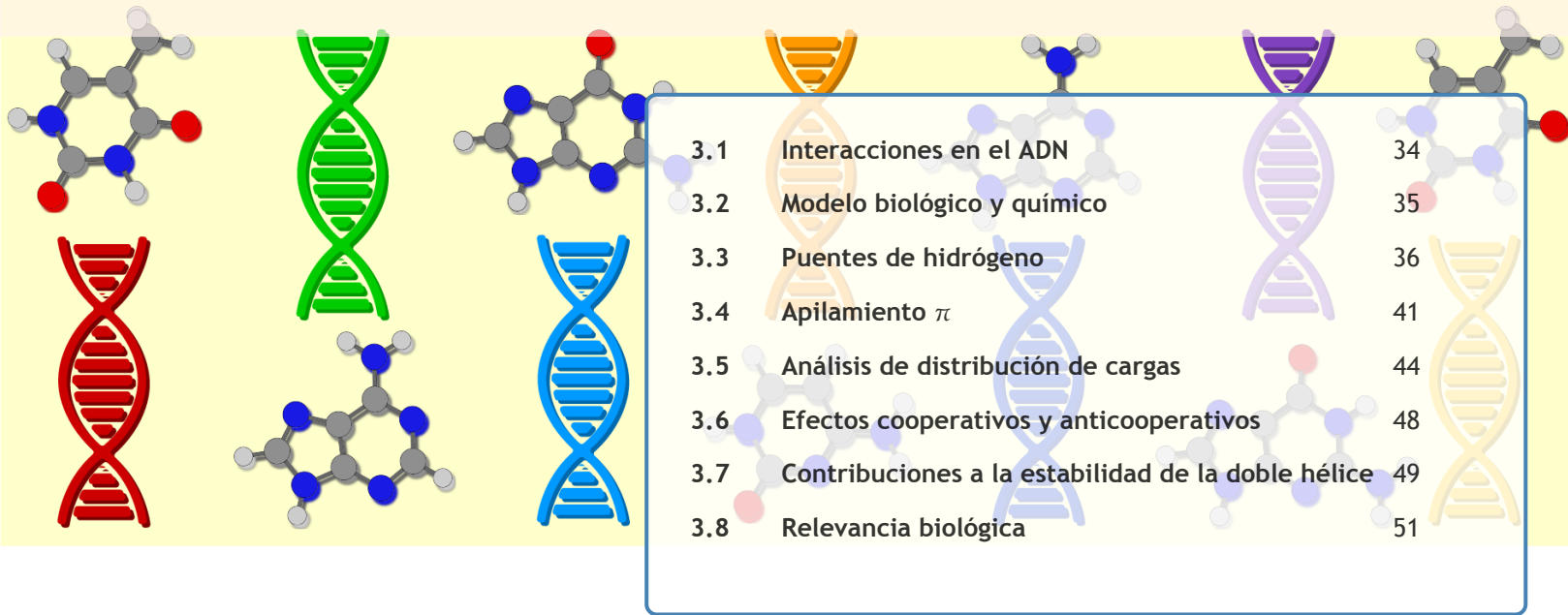
Finalmente, es relevante resaltar que las interacciones no-covalentes son no aditivas. Es decir, si tuviéramos un dímero de agua, al compararlo con un trímero, se podría esperar que la energía de interacción en el trímero correspondiera a la suma de la energía de dos puentes de hidrógeno (es decir, el doble de la energía de interacción correspondiente al dímero); sin embargo, al hacer la medición se encuentra que esto no sucede. En cambio, se observa que la energía de interacción en el trímero sería mayor que el doble de la energía de puente de hidrógeno del dímero.

Para energías asociadas al enlace químico, un esquema aditivo sería adecuado para estimar las energías de enlace. No obstante, para las interacciones débiles (dado que las interacciones no se suman) se opta por un esquema no aditivo debido a la presencia de efectos cooperativos y/o anticooperativos. Cuando un par de interacciones no-covalentes se fortalece entre sí, se dice que existe un efecto cooperativo; mientras que cuando se debilitan mutuamente, dominan los efectos anti-cooperativos [51]. La cooperatividad, por lo tanto, implica que la suma de al menos dos interacciones es más grande que la simple adición de los interacciones individuales [51].

El origen de la cooperatividad (o anticooperatividad) radica principalmente en los siguientes factores: (a) interacción de varios cuerpos, (b) interacciones secundarias, (c) efecto quelato y (d) cambios conformacionales [51]. Todos, o al menos uno de estos efectos, serán causantes del esquema no aditivo de la estimación de la energía de interacciones débiles.

En este trabajo, los efectos cooperativos y/o anticooperativos serán estudiados puesto que, por el número de pares de base en las cadenas, existen dos o más interacciones de puente de hidrógeno y, también, de apilamiento π , por lo que será de interés saber qué impacto tendrán estos efectos en el comportamiento biológico del dúplex de ADN.

3. Resultados y análisis



3.1	Interacciones en el ADN	34
3.2	Modelo biológico y químico	35
3.3	Puentes de hidrógeno	36
3.4	Apilamiento π	41
3.5	Análisis de distribución de cargas	44
3.6	Efectos cooperativos y anticooperativos	48
3.7	Contribuciones a la estabilidad de la doble hélice	49
3.8	Relevancia biológica	51

“Creo que la formación de la estructura del ADN [...] puede resultar ser uno de los mayores avances en el campo de la genética molecular en los últimos años.”

Linus Pauling

Asumiendo que los puentes de hidrógeno y de apilamiento π son las interacciones no-covalentes más importantes en la estabilización del dúplex de ADN, en este capítulo se presentará una descripción de dichas interacciones en cadenas tipo caja TATA y tipo islas CpG, considerando cadenas de diferentes longitudes (número de pares de base). Todo esto con la finalidad de determinar cuál de las interacciones participa de forma más importante en la estabilización de la doble hélice. Adicionalmente, con la información encontrada, se determinará si los puentes de hidrógeno y/o apilamiento π tienen algún impacto en el comportamiento biológico del ADN, en función de la secuencia de nucleótidos.

3.1 Interacciones en el ADN

“Aunque cualquier interacción no-covalente es bastante débil, una gran cantidad de estas interacciones entre moléculas, o diferentes partes de una molécula, puede estabilizar la estructura tridimensional de proteínas y ácidos nucleicos, además de mediar interacciones específicas de unión”[52]

Lodish y colaboradores, 2000

En el ADN, los puentes de hidrógeno se encuentran entre las bases nitrogenadas de un par de base, mientras que las interacciones de apilamiento π se encuentran entre los pares de base (Figura 3.1). Dada la enorme cantidad de pares de base en una cadena de ADN, se cree que estas interacciones son las más importantes en la estabilización de la estructura tridimensional de dicha molécula [19]. Además, en función del tipo de bases nitrogenadas que conforman a los pares de base, el ambiente que lo rodea y, principalmente, de las interacciones que hay a lo largo de la cadena, el ADN puede pasar de una conformación tipo B a una tipo A [17]. Sin embargo, en condiciones de homeostasis, la estructura más común es la tipo B [16].

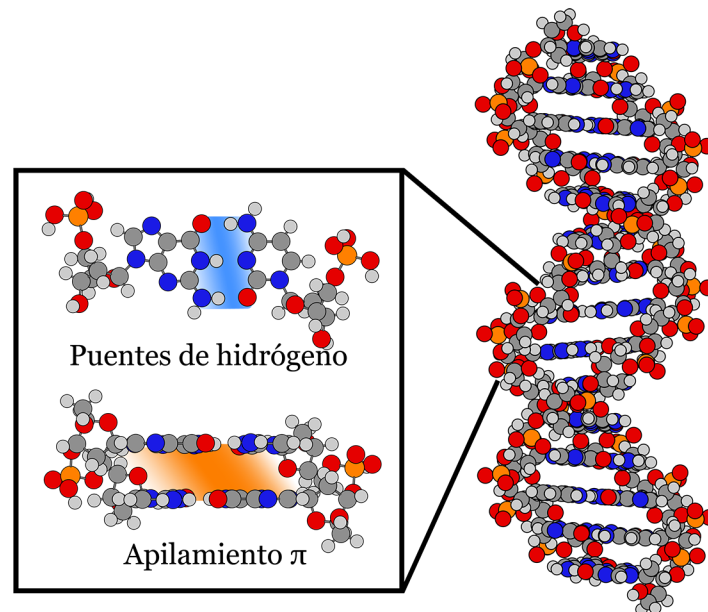


Figura 3.1. Interacciones de puente de hidrógeno (entre las bases nitrogenadas de un par de base) y de apilamiento π (entre los pares de base) a lo largo del dúplex de ADN.

Dada la complejidad de la estructura del ADN, existen discrepancias para establecer cual de las dos interacciones (puentes de hidrógeno o apilamiento π) participan de forma más significativa en la estabilización del dúplex. Por ejemplo, Smirnov y colaboradores conclu-

veron, al observar que el ADN polimerasa podía llevar a cabo el proceso de replicación, que las interacciones de apilamiento π son más importantes que los puentes de hidrógeno a pesar de que la cadena molde (la hebra utilizada como plantilla) estaba conformada por bases nitrogenadas apolares (bases sin la capacidad de formar puentes de hidrógeno) [53]. En contraste con estos hallazgos, Zhang y colaboradores, con base en sus estimaciones de la fuerza de las interacciones de puente de hidrógeno y apilamiento π obtenidas a través de la técnica de microscopia de fuerza atómica, concluyeron que los puentes de hidrógeno, al ser las interacciones más fuertes, son los que contribuyen de forma más importante en la estabilización de la doble hélice [54]. Por lo tanto, no es claro cuál de las dos juega un papel predominante en la estabilización de la estructura del ADN. Por consiguiente, uno de los objetivos particulares de este proyecto es determinar cuál interacción es más importante en la estabilización de la doble hélice.

No obstante que tanto las cadenas tipo caja TATA (cadenas conformadas por timina y adenina), como las tipo islas CpG (constituidas por guanina y citosina) son secuencias asociadas a regiones promotoras [25, 55], entre ellas existen diferencias biológicas muy importantes. Una de ellas, tal vez la más relevante, es que las islas CpG son regiones promotoras más comunes que las cajas TATA, puesto que las primeras se encuentran en un 70% de los genes asociados a organismos vertebrados [56]; además de que estas secuencias no necesitan de factores de transcripción específicos (en contraste con las cajas TATA) [57, 58] y que la probabilidad de unión de factores de transcripción es mayor en islas CpG que en las cajas TATA [56]. Con base en estas disimilitudes, una pregunta que resulta, desde un enfoque fisicoquímico, es determinar si las interacciones no-covalentes, asociadas a estas secuencias, juegan un papel importante en el comportamiento biológico de dichas regiones promotoras. Por lo tanto, con base en las descripciones realizadas de los puentes de hidrógeno y apilamiento π en las cadenas modeladas, se buscará resolver dicha cuestión.

3.2 Modelo biológico y químico

Para llevar a cabo el proyecto, se consideraron cadenas tipo caja TATA (5'-...TATA...-3') y tipo islas CpG (5'-...GCGC...-3'), modelando dúplexes de diferentes tamaño para determinar el efecto del número de pares de base (n) en las interacciones de puente de hidrógeno y de apilamiento π . Por lo tanto, se utilizaron cadenas desde $n = 1$ a $n = 14$. La carga de los grupos fosfato, a lo largo de las dobles hélices, fue neutralizada mediante la adición de un átomo de hidrógeno a un oxígeno de dicho grupo, asumiendo que no habrá un efecto importante sobre las interacciones no-covalentes estudiadas [59, 60].

Cada cadena modelada fue construida a través del software de libre acceso *make-na* [61], el cual considera diferentes parámetros, basados en datos experimentales, para la construcción de cadenas ADN-B simétricas, las cuales, de ahora en adelante, serán identificadas como cadenas ideales. Estas dobles hélices ideales, posteriormente, fueron sometidas a un proceso de optimización de geometría, con la finalidad de mejorar la descripción y estimación de las interacciones de puente de hidrógeno y apilamiento π ¹.

¹Es importante recordar que los procesos de optimización de geometría conducen al mínimo de energía más cercano asociado a la estructura inicial. Consecuentemente, podemos esperar que la descripción realizada en este trabajo será lo más parecida a lo que ocurre en una cadena ADN-B.

Todos los cálculos fueron realizados bajo el marco de la Teoría de los Funcionales de la Densidad² (DFT, por sus siglas en inglés) [62, 63]. Las optimizaciones de geometría y los cálculos de energía³ y propiedades moleculares fueron realizados con el código DMol³ [64], implementado en la paquetería de Química Computacional Materials Studio 2017 [65], utilizando el nivel de teoría Mo6-L[66]/DNP. El análisis de cargas NBO fue realizado en el código Gaussian 2016 [67], empleando el mismo funcional, pero con una base gaussiana 6-31G(d)[68-77]. Las correcciones asociadas al error de superposición de base (BSSE, por sus siglas en inglés) fueron tomadas en cuenta a través del método de counterpoise [78]. Finalmente, los criterios de convergencia considerados fueron: 2×10^{-5} Ha, 4×10^{-3} Ha/Å y 5×10^{-3} Å para la energía, el gradiente y el desplazamiento, respectivamente.

Es importante mencionar que el funcional Mo6-L fue seleccionado, como parte del modelo químico, dado que éste fue parametrizado para la adecuada descripción de interacciones no-covalentes, tales como los puentes de hidrógeno y apilamiento π [66]. De acuerdo a Goerigk, y colaboradores, dicho método es uno de los mejores meta-GGA para describir interacciones intermoleculares [79].

3.3 Puentes de hidrógeno

La energía promedio de puente de hidrógeno (\bar{D}_{HB}) fue estimada utilizando el valor de la energía promedio de formación de puente de hidrógeno (\bar{E}_{HB}), la cual fue calculada a partir de la energía total de la cadena con n pares de base (E_n), y de la energía total de cada una de las hebras ($E_{n_{1s}}$ y $E_{n_{2s}}$) que conforman al correspondiente dúplex de ADN (ver Ecuación (3.1) y Figura 3.2).

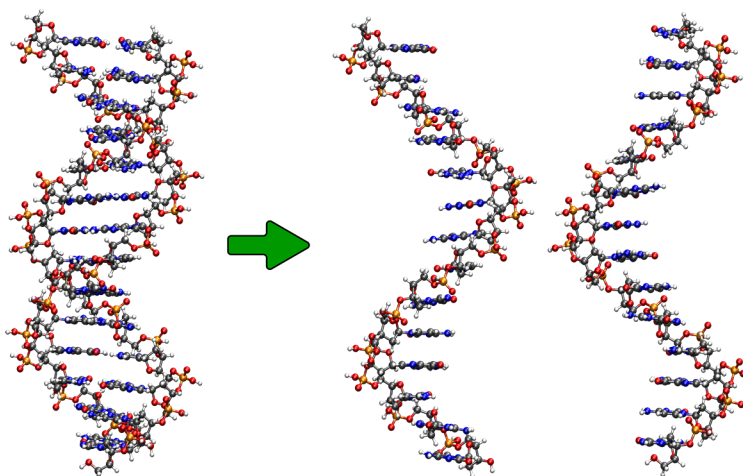


Figura 3.2. Esquematización del cálculo de la energía promedio de puente de hidrógeno (\bar{D}_{HB}). La cadena tipo isla CpG con $n = 14$ es mostrada como ejemplo.

²Esta teoría es explicada con mayor detalle en el Apéndice B.

³Cabe resaltar que no se incluyó ninguna corrección termodinámica.

$$\bar{D}_{HB} = -\bar{E}_{HB} = -\frac{E_n - (E_{n_{1s}} + E_{n_{2s}})}{n} \quad (3.1)$$

Al utilizar la Ecuación (3.1) para estimar la energía promedio de puente de hidrógeno, se asume que la mayor contribución al separar a las cadenas en sus respectivas hebras (como se muestra en la Figura 3.2) proviene de las interacciones de puente de hidrógeno. Para validar esto, se utilizó la ecuación de Espinosa-Molins-Lecomte (EML) [30]:

$$E_{HB} = 0.5 \cdot V(r) \quad (3.2)$$

La cual, basada en la Teoría Cuántica de Átomos en Moléculas (QTAIM, por sus siglas en inglés) [80], toma en cuenta la distribución electrónica en los puntos críticos de enlace (BCPs, por sus siglas en inglés) asociados a los puentes de hidrógeno, y correlaciona la densidad de energía potencial ($V(r)$) con la energía de dicha interacción (E_{HB}) [30].

Caja tipo caja TATA con $n = 7$				
Par de base	Interacción	$V(r)^{BCP}$ (a.u.)	E_{HB} (kcal/mol)	$-\sum E_{HB_i}$
1	N-H...O	-2.0346×10^{-02}	-6.38	15.01
	N-H...N	-2.6585×10^{-02}	-8.34	
	C-O...H	-8.9695×10^{-04}	-0.28	
2	N-H...O	-2.0464×10^{-02}	-6.42	14.99
	N-H...N	-2.6409×10^{-02}	-8.29	
	C-O...H	-8.9610×10^{-04}	-0.28	
3	N-H...O	-2.0428×10^{-02}	-6.41	15.03
	N-H...N	-2.6570×10^{-02}	-8.34	
	C-O...H	-8.9277×10^{-04}	-0.28	
4	N-H...O	-2.0544×10^{-02}	-6.45	15.00
	N-H...N	-2.6384×10^{-02}	-8.28	
	C-O...H	-8.9658×10^{-04}	-0.28	
5	N-H...O	-2.0352×10^{-02}	-6.39	14.99
	N-H...N	-2.6535×10^{-02}	-8.33	
	C-O...H	-8.9289×10^{-04}	-0.28	
6	N-H...O	-2.0456×10^{-02}	-6.42	14.96
	N-H...N	-2.6338×10^{-02}	-8.26	
	C-O...H	-8.9417×10^{-04}	-0.28	
7	N-H...O	-2.0388×10^{-02}	-6.40	15.06
	N-H...N	-2.6719×10^{-02}	-8.38	
	C-O...H	-8.9367×10^{-04}	-0.28	
$\bar{D}_{HB} = \left(\sum (-\sum E_{HB_i}) \right) / 7$				15.01 kcal/mol

Tabla 3.1. Energías de puente de hidrógeno para la cadena ideal tipo caja TATA con $n = 7$, calculadas con la ecuación EML (Ecuación (3.2)).

Las estimaciones de E_{HB} y el promedio (\bar{D}_{HB}), con base en la Ecuación (3.2), de las interacciones de puente de hidrógeno para las cadenas ideales y optimizadas tipo caja TATA con $n = 7$, se muestran en las Tablas 3.1 y 3.2. Mientras que la comparación de las estimaciones de \bar{D}_{HB} obtenidas mediante las Ecuaciones (3.1) y (3.2), se muestran en la Tabla 3.3.

Caja tipo caja TATA con $n = 7$				
Par de base	Interacción	$V(r)^{BCP}$ (a.u.)	E_{HB} (kcal/mol)	$-\sum E_{HB_i}$
1	N-H...O	-1.8185×10^{-02}	-5.71	14.36
	N-H...N	-2.6124×10^{-02}	-8.20	
	C-O...H	-1.4696×10^{-03}	-0.46	
2	N-H...O	-2.0331×10^{-02}	-6.38	15.96
	N-H...N	-2.8899×10^{-02}	-9.07	
	C-O...H	-1.6318×10^{-03}	-0.51	
3	N-H...O	-2.1255×10^{-02}	-6.67	15.69
	N-H...N	-2.7383×10^{-02}	-8.58	
	C-O...H	-1.3242×10^{-03}	-0.42	
4	N-H...O	-2.0107×10^{-02}	-6.31	15.19
	N-H...N	-2.6980×10^{-02}	-8.47	
	C-O...H	-1.3394×10^{-03}	-0.42	
5	N-H...O	-2.0095×10^{-02}	-6.30	15.64
	N-H...N	-2.8227×10^{-02}	-8.86	
	C-O...H	-1.5374×10^{-03}	-0.48	
6	N-H...O	-1.9071×10^{-02}	-5.98	15.70
	N-H...N	-2.9068×10^{-02}	-9.12	
	C-O...H	-1.8935×10^{-03}	-0.59	
7	N-H...O	-1.9489×10^{-02}	-6.11	15.52
	N-H...N	-2.8353×10^{-02}	-8.90	
	C-O...H	-1.6267×10^{-03}	-0.51	
$\bar{D}_{HB} = \left(\sum (-\sum E_{HB_i}) \right) / 7$				15.44 kcal/mol

Tabla 3.2. Energías de puente de hidrógeno para la cadena optimizada tipo caja TATA con $n = 7$, calculadas con la ecuación EML (Ecuación (3.2)).

Como se puede observar en la Tabla 3.3, la diferencia entre las estimaciones de la energía promedio de puente de hidrógeno es menor a 3 kcal/mol. Por lo tanto, se asume que la Ecuación (3.1) estima adecuadamente la energía de interacción de puente de hidrógeno.

En la Figura 3.3 se muestran las energías promedio de puente de hidrógeno de las cadenas tipo caja TATA y tipo islas CpG, tanto para los modelos ideales como para los optimizados. Por un lado, para las cadenas tipo caja TATA (lado izquierdo de la Figura 3.3), \bar{D}_{HB} de las cadenas ideales es mayor que los valores asociados a las cadenas optimizadas. Esto se debe a que los pares de base pierden planaridad al momento en que las estructuras son optimiza-

\bar{D}_{HB} (kcal/mol)			
Ecuación (3.1)		Ecuación (3.2)	
Ideal	Optimizada	Ideal	Optimizada
17.84	17.65	15.01	14.44

Tabla 3.3. Energías de puente de hidrógeno calculadas a través de las Ecuaciones (3.1) y (3.2) (primeras dos columnas y últimas dos columnas, respectivamente), para las cadenas ideales y optimizadas, tipo caja TATA, con $n = 7$.

das. Dado que la interacción C-H...O es débil (Tablas 3.1 y 3.2), la geometría de los pares de base no se encuentra muy restringida, razón por la cual los pares de base dejan de ser planos fácilmente, afectando a las interacciones de puente de hidrógeno. Por el contrario, en las cadenas tipo islas CpG (lado derecho de la Figura 3.3) las energías promedio de puente de hidrógeno del modelo optimizado son mayores que las del modelo ideal, puesto que al existir un tercer puente de hidrógeno (más fuerte) la planaridad de los pares de base está más restringida. Esto se ve reflejado en las diferencias energéticas de \bar{D}_{HB} que hay entre los modelos ideales y optimizados de las cadenas tipo caja TATA y tipo islas CpG: ésta es menor en las cadenas tipo islas CpG, dado que los pares de base no pierden planaridad al optimizar las cadenas. Cabe destacar que las energías promedio de puente de hidrógeno calculadas coinciden con las tendencias encontradas experimentalmente para los pares de base TA y GC [81].

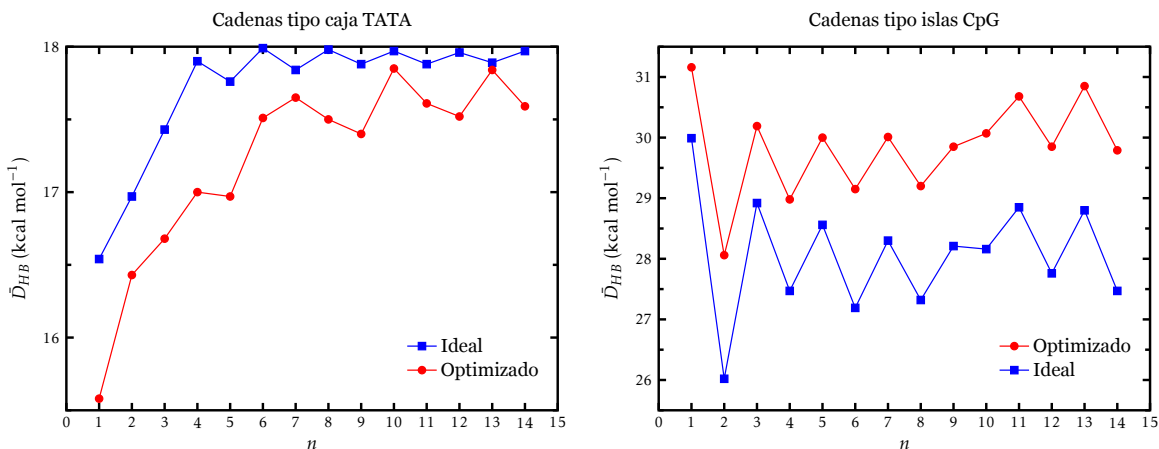


Figura 3.3. Energías promedio de puente de hidrógeno (\bar{D}_{HB}) en función del número de pares de base. La gráfica de la izquierda muestra las estimaciones para las cadenas tipo caja TATA, mientras que la de la derecha las correspondientes para las cadenas tipo islas CpG. Los cuadrados azules representan los resultados para las cadenas ideales y los círculos rojos los de las cadenas optimizadas. Las líneas conectan a los puntos solamente por claridad.

En cuanto a la correlación de la energía promedio de puente de hidrógeno y el número de pares de base (tamaño de la cadena), podemos observar que \bar{D}_{HB} tiende a aumentar conforme aumenta n para las cadenas tipo caja TATA. No obstante, para las cadenas tipo islas CpG, \bar{D}_{HB} parece no ser afectado por la longitud de la cadena, puesto que no hay cambios signifi-

cativos en la energía de puente de hidrógeno. Esto puede asociarse a qué tan restringida esté la planaridad de los pares de base: mientras menos cambios pueda haber en la conformación de los pares de base, menor efecto tendrá sobre \bar{D}_{HB} el número de pares de base. Como veremos más adelante, esto estará asociado a los efectos cooperativos entre las interacciones de puente de hidrógeno en función de la secuencia de nucleótidos. Finalmente, es importante mencionar que las oscilaciones en \bar{D}_{HB} (presentes tanto para las cadenas tipo caja TATA y tipo islas CpG) cuando n es par, o impar, se deben a efectos de borde.

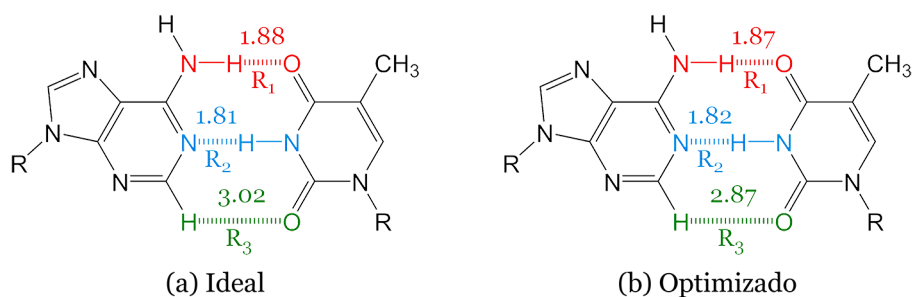


Figura 3.4. Distancias promedio de puente de hidrógeno de las cadenas ideales y optimizadas tipo caja TATA.

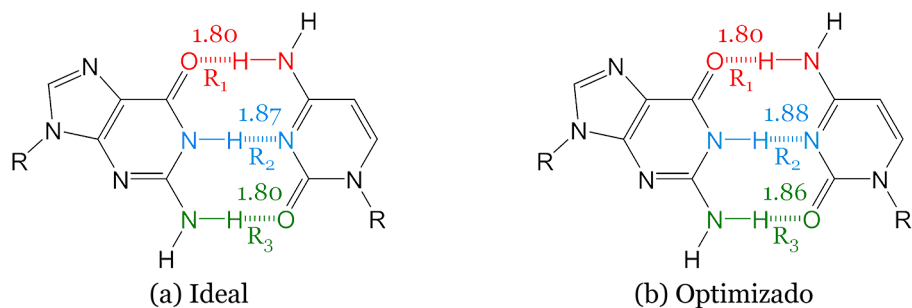


Figura 3.5. Distancias promedio de puente de hidrógeno de las cadenas ideales y optimizadas tipo islas CpG.

Para profundizar en la descripción de las interacciones de puente de hidrógeno, en las Figuras 3.4 y 3.5 se muestran las distancia promedio de puente de hidrógeno de las cadenas ideales y optimizadas tipo caja TATA y tipo islas CpG, respectivamente. En este análisis se asume que mientras menor sea la distancia entre los átomos involucrados en la interacción, mayor será la fuerza de los puentes de hidrógeno [82]. Para las cadenas tipo caja TATA, el puente de hidrógeno central (R₂) es el más fuerte, mientras que en las cadenas tipo islas CpG los puentes de hidrógeno laterales (R₁ y R₃) son los de menor distancia (mayor fuerza); todo esto en concordancia con las tendencias reportadas experimentalmente [83, 84]. Cabe resaltar que las interacciones de puente de hidrógeno entre los pares de base GC son más energéticas que en los pares TA (Figura 3.3) debido a la existencia de una interacción más fuerte entre los átomos involucrados en R₃ (Figuras 3.4 y 3.5).

3.4 Apilamiento π

De forma similar a los puentes de hidrógeno, la energía promedio de apilamiento π (\bar{D}_π) fue calculada a partir de la energía de formación de dicha interacción (\bar{E}_π). Para esta última, se tomó en cuenta la energía total de las cadenas sin grupos fosfato ni azúcares⁴ (E_{n_π}), la energía total de cada par de base ($E_{i_{bp}}$) y el número de espacios entre los pares de base (N_π). Todo esto se resume y representa en la Ecuación (3.3) y Figura 3.6.

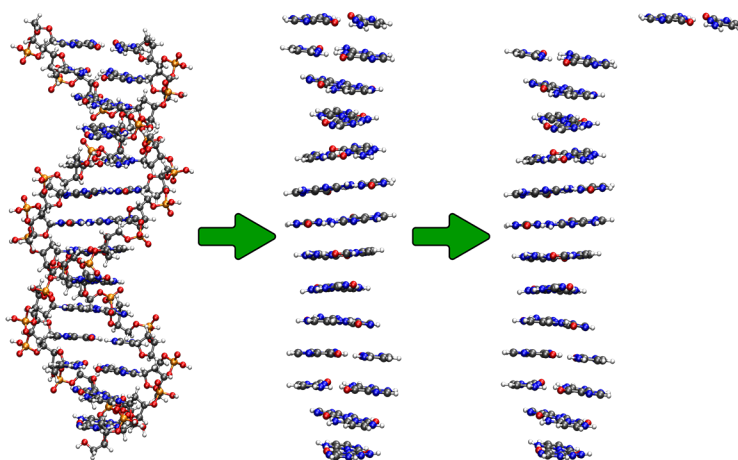


Figura 3.6. Esquematzación del cálculo de la energía promedio de apilamiento \bar{D}_π . La cadena tipo isla CpG con $n = 14$ es mostrada como ejemplo. Los grupos fosfato y los azúcares fueron reemplazados por un átomo de hidrógeno para la estimación de esta interacción.

$$\bar{D}_\pi = -\bar{E}_\pi = -\frac{E_{n_\pi} - \sum_{i=1}^n E_{i_{bp}}}{N_\pi} \quad (3.3)$$

Las estimaciones de \bar{D}_π con base en la Ecuación (3.3), para las cadenas ideales y optimizadas tipo caja TATA y tipo islas CpG, se muestran en la Figura 3.7. Para ambos tipos de cadena, en general, la energía promedio de apilamiento π es mayor en las dobles hélices optimizadas que en las ideales. Esta diferencia es mayor en los dúplexes tipo islas CpG debido a que los pares de base suelen ser más planos, lo que permite que estas interacciones se vean favorecidas entre pares de base con guanina y citosina. A diferencia de los puentes de hidrógeno, la diferencia de energía entre las cadenas ideales y optimizadas asociadas al apilamiento π es mayor, por lo que se puede sospechar que las interacciones de apilamiento π son las que se ven mayormente favorecidas cuando el dúplex de ADN busca su configuración más estable, lo cual coincide con lo concluido en el trabajo de Smirnov y colaboradores [53].

Cabe destacar que no existe una relación muy clara entre el número de pares de base y la energía promedio de apilamiento π (Figura 3.7), puesto que, en general, no existen diferencias de energía significativas en \bar{D}_π conforme n aumenta. Sin embargo, tanto para las cadenas tipo caja TATA como islas CpG, alrededor de $n = 10$ se puede observar un aumento en \bar{D}_π . Con base en lo descrito por Swart y colaboradores [32], la estructura de más baja energía de

⁴Dichos grupos fueron reemplazados por un átomo de hidrógeno.

dos pares de base apilados tiene un mínimo cuando el ángulo *twist* es de 36° . Por lo tanto, este incremento encontrado en $n = 10$ puede estar relacionado con el giro completo de 360° de la cadena de ADN, debido al eje de simetría C_{10} asociado a los dúplexes ADN-B.

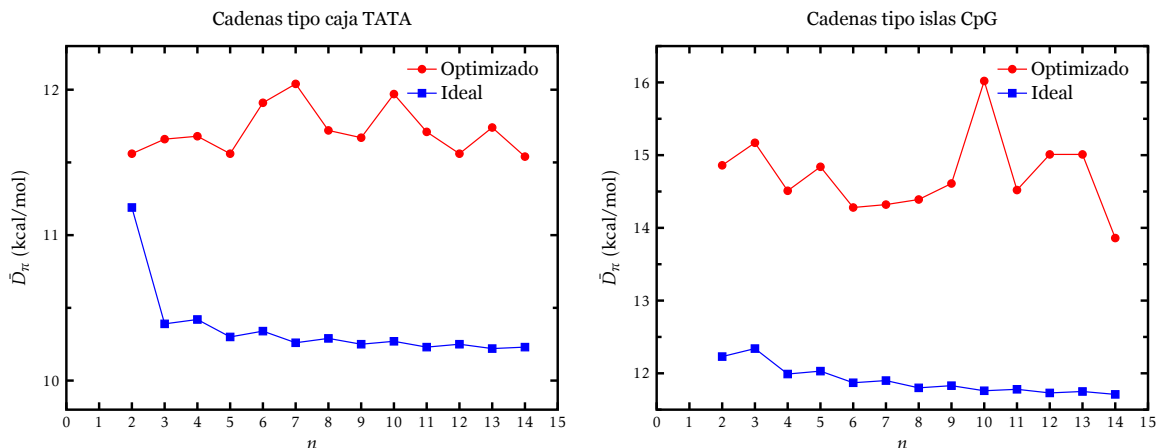


Figura 3.7. Energías promedio de apilamiento π (\bar{D}_π) en función del número de pares de base. La gráfica de la izquierda muestra las estimaciones para las cadenas tipo caja TATA, mientras que la de la derecha las correspondientes para las cadenas tipo islas CpG. Los cuadrados azules representan los resultados para las cadenas ideales y los círculos rojos los de las cadenas optimizadas. Las líneas conectan a los puntos solamente por claridad.

Las oscilaciones en \bar{D}_π , observadas en la Figura 3.7, pueden ser explicadas con base en las rotaciones *roll* y *twist* y en la traslación *slide*. Los valores de dichos grados de libertad se presentan en las Tablas 3.4 y 3.5, tanto para las dobles hélices ideales y optimizadas de las cadenas tipo caja TATA y tipo islas CpG. Al trabajar con modelos basados en dúplexes tipo ADN-B, los valores esperados de *roll*, *twist* y *slide* son de 0° , 36° y 0 \AA , respectivamente. Por lo tanto, de acuerdo a las Tablas 3.4 y 3.5, las dobles hélices estudiadas en este trabajo corresponden, como era de esperarse, a cadenas tipo B. Las desviaciones a los valores esperados se deben a que las cadenas fueron modeladas en fase gas. De estos grados de libertad analizados, la rotación *roll* y la traslación *slide* están relacionadas con la energía promedio de apilamiento π , puesto que están directamente involucradas con el acomodo correlacionado de los pares de base a lo largo de la cadena de ADN. En el caso de las cadenas tipo caja TATA, mientras más negativo sea el *slide* mayor será \bar{D}_π . En este caso, al no haber cambios significativos, el *slide* no provoca cambios importantes en la energía promedio de apilamiento π . Por otro lado, en el caso de las cadenas tipo islas CpG, mientras más negativo sea el *slide* y más cercano a 0° sea el *roll*, mayor será la magnitud de \bar{D}_π . Las diferencias encontradas en la dependencia de \bar{D}_π en función de la rotación *roll*, entre las cadenas tipo caja TATA y tipo islas CpG, se puede deber a la planaridad de las bases: al estar más restringida la planaridad de un par de base GC, debido a la existencia de un tercer puente de hidrógeno (más fuerte), un cambio en dicha rotación provocará un alejamiento entre dos pares de base, lo cual disminuirá el contacto entre las nubes de densidad π y, por lo tanto, un descenso en la energía promedio de apilamiento π .

Cadenas tipo caja TATA						
<i>n</i>	Ideal			Optimizada		
	<i>Roll</i>	<i>Slide</i>	<i>Twist</i>	<i>Roll</i>	<i>Slide</i>	<i>Twist</i>
2	-0.86	0.01	36.87	3.13	-0.26	38.25
3	-2.81	-0.21	35.90	-2.60	-0.35	36.21
4	-2.81	-0.20	35.84	-1.90	-0.35	37.29
5	-3.63	-0.28	36.38	-3.63	-0.28	36.38
6	-2.81	-0.20	35.86	-2.96	-0.40	36.97
7	-2.81	-0.20	35.89	-3.49	-0.47	35.86
8	-2.81	-0.20	35.87	-3.68	-0.38	36.41
9	-2.81	-0.20	35.89	-3.60	-0.39	35.95
10	-2.81	-0.20	35.87	-3.66	-0.49	35.96
11	-2.81	-0.20	35.89	-3.64	-0.42	36.04
12	-2.81	-0.20	35.88	-3.46	-0.38	36.35
13	-2.81	-0.20	35.89	-4.10	-0.49	35.89
14	-2.81	-0.20	35.88	-3.52	-0.39	36.28

Tabla 3.4. Magnitudes de la traslación *slide* y de los ángulos de rotación *slide* y *twist* de las cadenas tipo caja TATA con $n > 2$. Análisis realizado mediante el software *Web 3DNA 2.0* [85].

Cadenas tipo islas CpG						
<i>n</i>	Ideal			Optimizada		
	<i>Roll</i>	<i>Slide</i>	<i>Twist</i>	<i>Roll</i>	<i>Slide</i>	<i>Twist</i>
2	-2.78	-0.12	37.95	-1.80	-0.30	39.31
3	-2.80	-0.18	35.90	-1.44	-0.27	37.43
4	-2.80	-0.16	36.58	-2.00	-0.30	37.34
5	-2.80	-0.18	35.90	-0.37	-0.17	37.85
6	-2.80	-0.16	36.31	-2.23	-0.22	37.16
7	-2.80	-0.18	35.90	-1.71	-0.21	37.26
8	-2.80	-0.17	36.19	-0.77	-0.24	37.61
9	-2.80	-0.18	35.89	-0.83	-0.13	37.73
10	-2.80	-0.17	36.12	-2.02	-0.33	36.87
11	-2.80	-0.18	35.89	-1.62	-0.21	37.07
12	-2.80	-0.17	36.08	-2.30	-0.29	36.56
13	-2.80	-0.18	35.89	-1.63	-0.30	37.12
14	-2.80	-0.17	36.05	-1.51	-0.24	36.92

Tabla 3.5. Magnitudes de la traslación *slide* y de los ángulos *slide* y *twist* de las cadenas tipo islas CpG con $n > 2$. Análisis realizado mediante el software *Web 3DNA 2.0* [85].

3.5 Análisis de distribución de cargas

Para completar la descripción de las interacciones de puente de hidrógeno y de apilamiento π en las dobles hélices estudiadas, en esta sección se hará un análisis de la distribución de cargas por medio del índice de separación de cargas (CSI, por sus siglas en inglés) y del potencial electrostático.

El CSI puede ser interpretado como una medida de la polaridad local en una molécula, y se obtiene de la suma de los valores absolutos de las cargas q de cada átomo i en la región de interés [86]:

$$CSI = \sum_i |q_i| \quad (3.4)$$

Por lo tanto, este descriptor puede ser utilizado para cuantificar el grado de separación entre cargas negativas y positivas [87] y relacionarlo con la fuerza de interacción de, por ejemplo, puentes de hidrógeno: mientras mayor sea la magnitud del CSI, mayor será la polarización en la región de interés y, por ende, más fuerte será el puente de hidrógeno descrito.

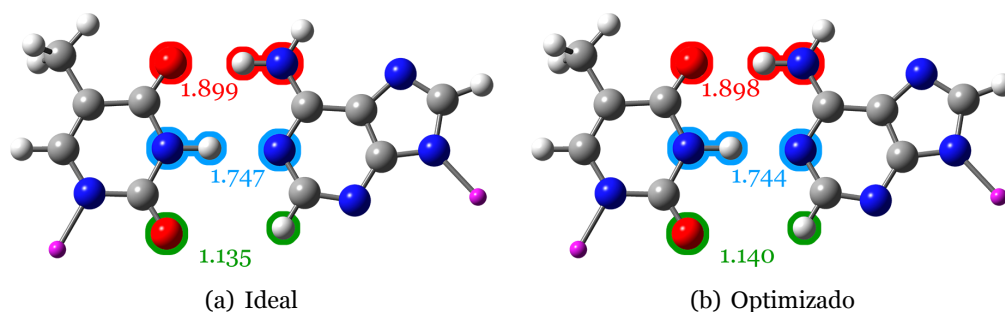


Figura 3.8. Promedio global del índice de separación de cargas (CSI, por sus siglas en inglés) asociado a los puentes de hidrógeno entre T y A, para las cadenas ideales y optimizadas tipo caja TATA con $n = 1, 3, 5$ y 7 . Solamente se consideraron los valores de los pares de base centrales con la finalidad de evitar efectos de borde.

La magnitud del CSI fue obtenida por medio de la Ecuación (3.4) para los pares de base centrales de todas las cadenas con $n = 1, 3, 5$ y 7 . Se eligieron estas cadenas dado que cuentan con un número impar de pares de base y, por lo tanto, con un par de base central en el cual, se asume, los efectos de borde son mínimos. Las cargas utilizadas fueron obtenidas a partir del análisis de poblaciones NBO. En las Figuras 3.8 y 3.9 se muestra el promedio global de los valores de CSI de los puentes de hidrógeno del par de base de los dúplexes mencionados.

Por un lado, para las cadenas tipo caja TATA (Figura 3.8) la diferencia entre los valores de CSI de las cadenas ideales y optimizadas no es significativa. Sin embargo, las magnitudes de dicho índice son, en general, menores en las cadenas optimizadas que en las ideales, lo cual está en concordancia con el análisis de las energías promedio de puente de hidrógeno (Figura 3.3). Al igual que en las distancias (donde R_3 tiende a disminuir en las cadenas optimizadas Figura 3.4), el valor de R_3 tiende a aumentar, lo que significa que la interacción tiende a crecer. Sin embargo, con base en los valores de \bar{D}_{HB} y de CSI, R_1 y R_2 son las interacciones de

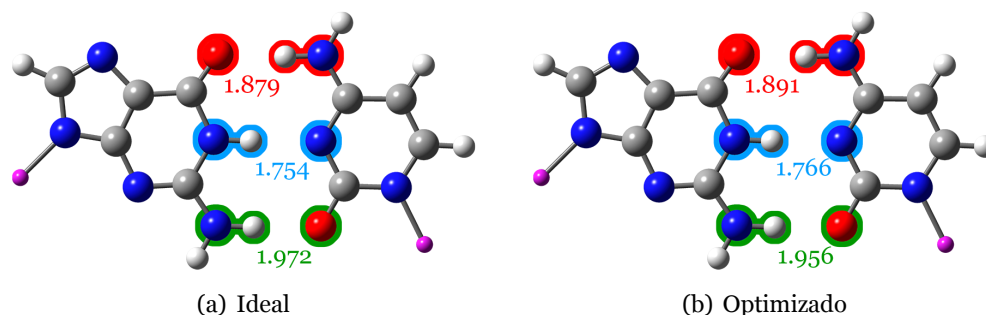


Figura 3.9. Promedio global del índice de separación de cargas (CSI, por sus siglas en inglés) asociado a los puentes de hidrógeno entre G y C, para las cadenas ideales y optimizadas tipo islas CpG con $n = 1, 3, 5$ y 7 . Solamente se consideraron los valores de los pares de base centrales con la finalidad de evitar efectos de borde.

puente de hidrógeno más importantes entre A y T.

Por otro lado, de forma similar que en las cadenas tipo caja TATA, el valor de CSI no cambia significativamente para las cadenas tipo islas CpG. No obstante, en contraste con los pares de base formados con T y A, las magnitudes de este descriptor tienden a aumentar en las cadenas optimizadas, lo cual coincide con el aumento de \bar{D}_{HB} al momento de optimizar la geometría de las cadenas ideales (ver Figura 3.6). El índice CSI es mayor en los puentes de hidrógeno R_1 y R_3 , coincidiendo con las distancias reportadas en la Figura 3.5 (donde las distancias más pequeñas están asociadas a R_1 y R_3), por lo que, en concordancia con la definición más básica de un puente de hidrógeno, mientras más electronegativo sea un átomo, mayor será la fuerza de interacción de puente de hidrógeno.

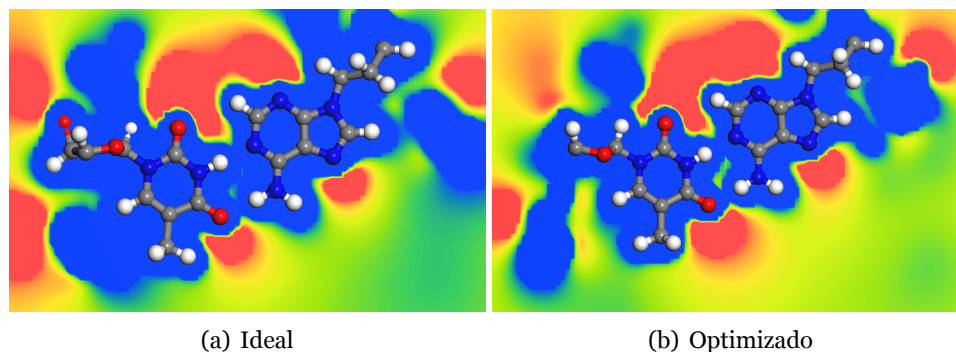


Figura 3.10. Planos de potencial electrostático del par de base central de las cadenas ideal y optimizada tipo caja TATA con $n = 7$, asumiendo que los efectos de borde no son significativos. La perspectiva mostrada en la imagen corresponde a las interacciones de puente de hidrógeno entre las bases nitrogenadas del par de base central. Escala del potencial electrostático: -0.04 u.a. (rojo) a 0.04 u.a. (azul).

En las Figuras 3.10 y 3.11 se muestran los planos de potencial electrostático asociados a las interacciones de puente de hidrógeno para las cadenas ideales y optimizadas tipo caja TATA y tipo islas CpG, respectivamente. Estos planos de potencial se graficaron sobre las bases ni-

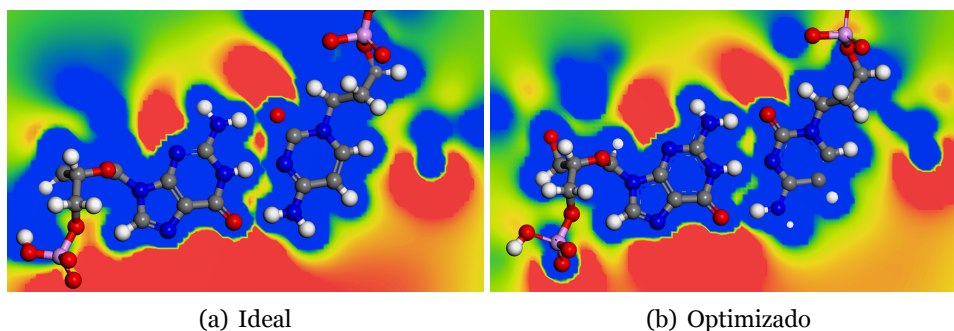


Figura 3.11. Planos de potencial electrostático del par de base central de las cadenas ideal y optimizada tipo islas CpG con $n = 7$, asumiendo que los efectos de borde no son significativos. La perspectiva mostrada en la imagen corresponde a las interacciones de puente de hidrógeno entre las bases nitrogenadas del par de base central. Escala del potencial electrostático: -0.04 u.a. (rojo) a 0.04 u.a. (azul).

trogenadas del par de base central de los dúplexes con $n = 7$, con el propósito de evitar los efectos de borde. En acuerdo con lo analizado para la energía promedio de puente hidrógeno (Figura 3.3), en los planos de potencial se puede observar cómo los puentes hidrógeno entre T y A tienden a debilitarse al momento en que la cadena ideal es optimizada (Figura 3.10), puesto que la región negativa asociada al par libre del oxígeno (en el puente de hidrógeno $O \cdots H - N$) tiende a aumentar en la cadena optimizada, siendo un indicio de que el par libre se encuentra menos compartido. En contraste, en los planos de potencial de los puentes de hidrógeno entre G y C (Figura 3.11) las zonas negativas entre las bases nitrogenadas tienden a desaparecer al optimizar la cadena ideal, lo cual puede ser relacionado con el hecho de que dicha carga se concentra en las regiones de los puentes de hidrógeno.

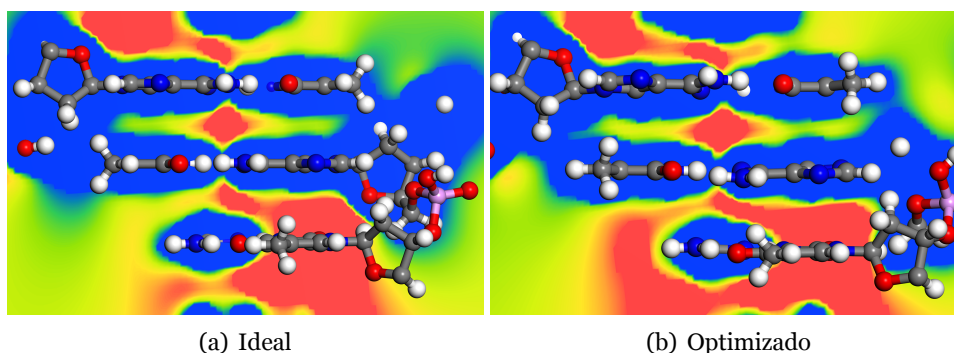


Figura 3.12. Planos de potencial electrostático del par de base central de las cadenas ideal y optimizada tipo caja TATA con $n = 7$, asumiendo que los efectos de borde no son significativos. La perspectiva mostrada en la imagen corresponde a las interacciones de apilamiento π entre los pares de base del dúplex. Escala del potencial electrostático: -0.04 u.a. (rojo) a 0.04 u.a. (azul).

Por otra parte, en las Figuras 3.12 y 3.13 se muestran los planos de potencial electrostáticos asociados a las interacciones de apilamiento π para las cadenas ideales y optimizadas tipo caja TATA y tipo islas CpG, respectivamente. Los planos se encuentran graficados de forma

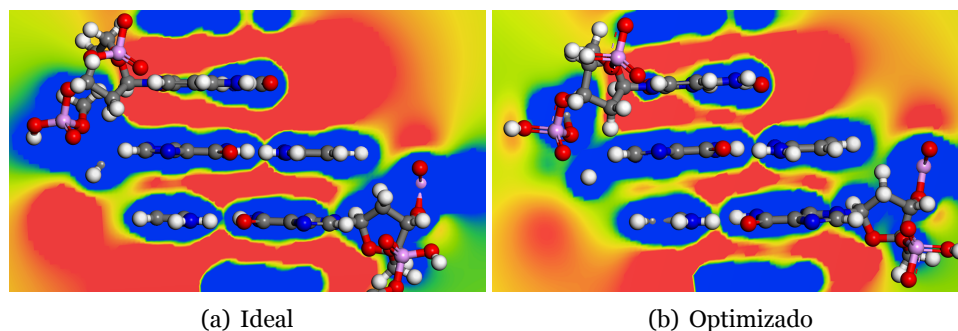


Figura 3.13. Planos de potencial electrostático del par de base central de las cadenas ideal y optimizada tipo islas CpG con $n = 7$, asumiendo que los efectos de borde no son significativos. La perspectiva mostrada en la imagen corresponde a las interacciones de apilamiento π entre los pares de base del duplex. Escala del potencial electrostático: -0.04 u.a. (rojo) a 0.04 u.a. (azul).

perpendicular al puente de hidrógeno R_2 de los pares de base centrales de las dobles hélices $n = 7$, con la finalidad de observar el potencial electrostático perteneciente a las regiones donde las interacciones de interés están presentes. Tanto para las cadenas tipo caja TATA (Figura 3.12) y las tipo islas CpG (Figura 3.13), las interacciones de apilamiento π tienden a favorecerse al momento en que los dúplexes ideales son optimizados. Esto se concluye debido a que en las cadenas optimizadas la región negativa entre los pares de base se concentra entre las bases nitrogenadas de cadenas opuestas, siendo indicio del apilamiento π en una configuración paralela desplazada (ver Figura 2.3) [46]. Este análisis está en concordancia con las conclusiones asociadas a la energía promedio de apilamiento π (Figura 3.6).

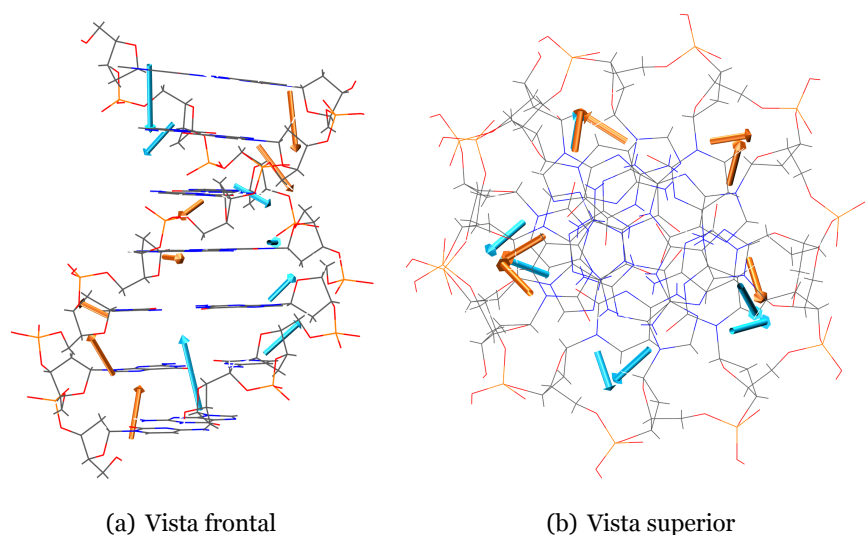


Figura 3.14. Momentos dipolo locales de cada base nitrogenada y su correspondiente pentosa. Los vectores mostrados en azul pertenecen a la hebra con sentido $5' - 3'$; los vectores mostrados en naranja corresponden a la hebra $3' - 5'$. La cadena tipo isla CpG con $n = 7$ fue utilizada para este análisis.

Finalmente, a partir del análisis de carga NBO, se estimaron los momentos dipolo locales $\vec{\mu}$ (Ecuación (3.5)) de cada base nitrogenada (con su respectiva pentosa) para cada par de base de la cadena tipo isla CpG con $n = 7$.

$$\vec{\mu} = \sum_i q_i \vec{r}_i \quad (3.5)$$

Los resultados se muestran en la Figura 3.14. Como se puede observar, los momentos dipolo giran entre cada par de base, por lo que la región negativa de cada base nitrogenada (punta de cada vector) apunta a una dirección diferente, disminuyendo, por lo tanto, las repulsiones entre las bases nitrogenadas, favoreciendo la conformación tipo ADN-B para cadenas con pares de base alternantes de un solo tipo (T-A o G-C), tales como las dobles hélices modeladas en este trabajo.

3.6 Efectos cooperativos y anticooperativos

Una vez analizadas las interacciones de puente de hidrógeno y de apilamiento π de forma individual, en esta sección se determinará si la presencia de otros pares de base (y, por lo tanto, de otras interacciones) tiene un efecto cooperativo o anticooperativo. Para tal fin, se tomará en cuenta la energía total de puente de hidrógeno y de apilamiento π (D_{HB} y D_{π} , respectivamente), calculadas a partir de \bar{D}_{HB} y \bar{D}_{π} (Ecuaciones (3.1) y (3.3), respectivamente) y del número correspondiente n de pares de base y espacios N_{π} entre los pares de base:

$$D_{HB}(n) = \bar{D}_{HB}(n) \cdot n \quad (3.6)$$

$$D_{\pi}(n) = \bar{D}_{\pi}(n) \cdot N_{\pi} \quad (3.7)$$

Adicionalmente, con base en la multiplicación de las energías D_{HB} y D_{π} estimadas para $n = 1$ y $N_{\pi} = 1$ (Ecuaciones (3.6) y (3.7), respectivamente), consideradas como la unidad, por el número de pares de base n y de espacios N_{π} entre los pares de base correspondientes, se calcularon las energías aditivas de puente de hidrógeno y de apilamiento π (D_{HB}^{add} y D_{π}^{add} , respectivamente):

$$D_{HB}^{add} = D_{HB}(n = 1) \cdot n \quad (3.8)$$

$$D_{\pi}^{add} = D_{\pi}(N_{\pi} = 1) \cdot N_{\pi} \quad (3.9)$$

Las energías aditivas corresponden al caso ideal donde los efectos cooperativos, o anticooperativos, son despreciados, es decir, en las energías aditivas se considera que las interacciones no-covalentes serán múltiplos de una unidad; mientras que en las energías totales dichos efectos cooperativos sí son tomados en cuenta, por lo que la energía de interacción será diferente al múltiplo de una unidad. La comparación gráfica entre las energías aditivas y totales de las interacciones de puente de hidrógeno y de apilamiento π , para las cadenas optimizadas tipo caja TATA y tipo islas CpG, se muestra en las Figuras 3.15 y 3.16, respectivamente.

Por un lado, en el caso de los puentes de hidrógeno (Figura 3.15), para las cadenas tipo caja TATA existen efectos cooperativos entre las interacciones de puente de hidrógeno, dado que $D_{HB} > D_{HB}^{add}$, en tanto que los efectos anticooperativos imperan en las cadenas tipo islas CpG debido a que $D_{HB} < D_{HB}^{add}$. Esta diferencia puede estar asociada a la rigidez de los pares de base, puesto que en los pares de base GC hay un tercer puente de hidrógeno (más fuerte), por lo que el movimiento de las bases nitrogenadas se encuentra más restringido, provocando

que haya un comportamiento anticooperativo.

Por otro lado, para las interacciones de apilamiento π , de forma contraria a las interacciones de puente de hidrógeno, en las cadenas tipo caja TATA existe un comportamiento aditivo, puesto que $D_\pi \approx D_\pi^{add}$; mientras que en las cadenas tipo islas CpG predominan, de forma similar a los puentes de hidrógeno, los efectos anticooperativos, dado que $D_\pi < D_\pi^{add}$. Nuevamente, estos diferentes comportamientos se pueden deber a la rigidez de los pares de base: mientras más restringida esté la planaridad de las bases nitrogenadas, mayor será la contribución de los efectos anticooperativos.

Lo discutido anteriormente se encuentra en concordancia con el análisis de \bar{D}_{HB} (Figura 3.3) y \bar{D}_π (Figura 3.7), en donde \bar{D}_{HB} aumenta en función de n para las cadenas ricas en T y A (efecto cooperativo), mientras que para las cadenas ricas en G y C no hay un cambio significativo (efecto anticooperativo). Para las interacciones de apilamiento π , \bar{D}_π no cambia significativamente conforme n aumenta en las cadenas tipo caja TATA (efecto aditivo), mientras que en las cadenas tipo CpG \bar{D}_π tiende, ligeramente, a disminuir (efecto anticooperativo).

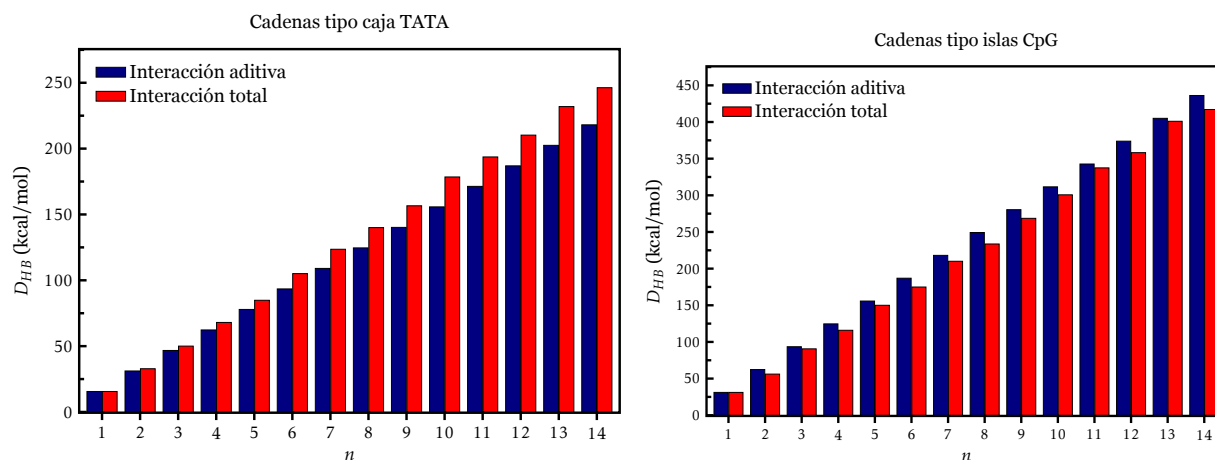


Figura 3.15. Comparación entre las energías aditivas (barras azules) y las energías totales (barras rojas) de puente de hidrógeno de las cadenas optimizadas tipo caja TATA (izquierda) y tipo islas CpG (derecha).

3.7 Contribuciones a la estabilidad de la doble hélice

Para determinar cuál de las dos interacciones participa de forma más importante en la estabilización de la doble hélice, se determinó el porcentaje de contribución a la estabilización asumiendo que los puentes de hidrógeno y el apilamiento π son las interacciones más relevantes a lo largo del dúplex de ADN [19]. Para tal fin, se utilizaron las energías totales de puente de hidrógeno y de apilamiento π (D_{HB} y D_π , respectivamente), estimadas mediante las Ecuaciones (3.6) y (3.7), respectivamente. Posteriormente, se calculó el porcentaje de contribución a la estabilidad de la siguiente forma:

$$\%Contibución = \frac{D_i}{D_{HB} + D_\pi} \cdot 100 \quad (3.10)$$

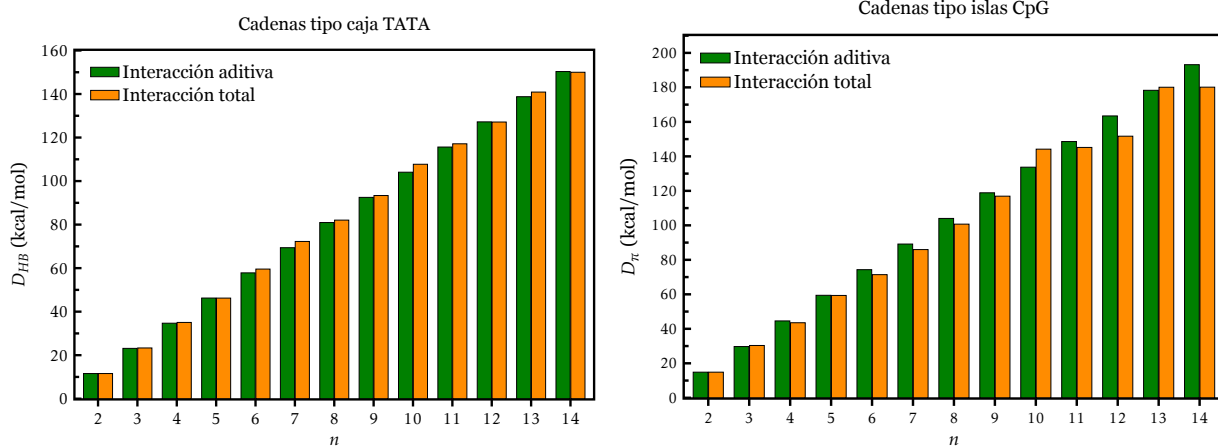


Figura 3.16. Comparación entre las energías aditivas (barras verdes) y las energías totales (barras naranjas) de apilamiento π de las cadenas optimizadas tipo caja TATA (izquierda) y tipo islas CpG (derecha).

donde D_i representa a D_{HB} o D_{π} , dependiendo de cual de las dos interacciones está siendo descrita.

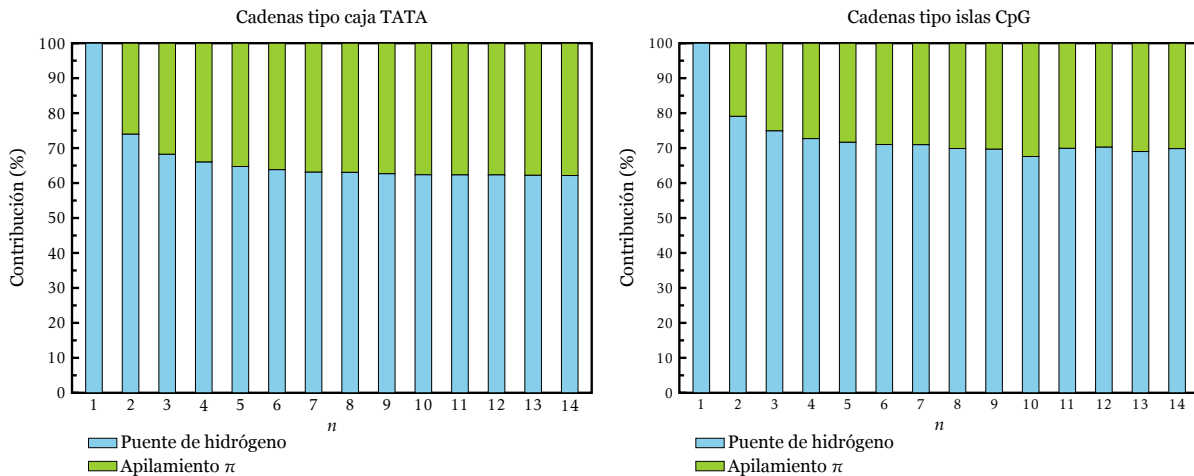


Figura 3.17. Porcentaje de contribución a la estabilidad de la doble hélice proveniente de las interacciones de puente de hidrógeno y apilamiento π . Análisis realizado en las cadenas optimizadas tipo caja TATA y tipo islas CpG.

En la Figura 3.17 se presentan las contribuciones a la estabilización provenientes de las interacciones de puente de hidrógeno y de apilamiento π para las cadenas optimizadas tipo caja TATA y tipo islas CpG. Como se puede notar, tanto en las cadenas tipo caja TATA como tipo islas CpG, la contribución más importante en la estabilidad de los dúplexes tiene origen en los puentes de hidrógeno. Sin embargo, conforme aumenta el número de pares de base, la contribución de apilamiento π es cada vez más significativa. Esto se encuentra en concordancia con lo reportado por Zhang y colaboradores [54], donde, desde un punto de vista físico,

las interacciones de puente de hidrógeno contribuyen de forma más importante en la estabilización de la cadena de ADN.

Por otro lado, desde un punto de vista biológico, las interacciones más importantes serán las de apilamiento π , dado que, probablemente, se optimizan antes que las de puente de hidrógeno (razón por la cual suele haber un aumento en \bar{D}_π al momento de optimizar las cadenas ideales, ver Figuras 3.2 y 3.6), por lo que al interactuar con alguna proteína (el ADN-polimerasa, por ejemplo), los residuos de esta última formarán primero interacciones de apilamiento π con los pares de base de la cadena de ADN. Esto último concuerda con los hallazgos de Smirnov y colaboradores [53].

Por último, cabe destacar que la contribución a la estabilidad de los puentes de hidrógeno es mayor en las cadenas tipo islas CpG que en las cadenas tipo caja TATA. Esto se debe a la existencia de un tercer puente de hidrógeno (más fuerte) entre los pares de base conformados por G y C. Por lo tanto, es de esperarse que en las cadenas ricas en G y C sea más difícil romper las interacciones de puente de hidrógeno, pero más fácil desestabilizar a las interacciones de apilamiento π . Ésta puede ser una de las razones por las cuales estas secuencias suelen ser más comunes que las tipo caja TATA.

3.8 Relevancia biológica

En las secciones anteriores se hizo un análisis y descripción del comportamiento de las interacciones de puente de hidrógeno y de apilamiento π en cadenas de ADN-B de diferente longitud (número de pares de base) y secuencia de nucleótidos. Sin embargo, de las preguntas planteadas en ese proyecto, queda una sin resolver: *¿Las interacciones débiles afectan el comportamiento biológico del ADN?* Con este propósito, en esta sección se hará un análisis de la dureza química (η) y de las densidades de estado (DOS, por sus siglas en inglés) de las cadenas tipo caja TATA y tipo islas CpG con $n = 8$, dado que, comúnmente, la longitud de la región promotora caja TATA es de 8 pares de base [88, 89].

Se eligieron como modelos cadenas tipo ADN-B considerando que, antes de adquirir la conformación tipo ADN-TA al interactuar con la proteína de unión a TATA (TBP, por sus siglas en inglés), se ha propuesto que la región promotora caja TATA inicia con una conformación tipo ADN-B, posteriormente cambia a una ADN-A y, finalmente, adquiere una conformación tipo ADN-TA en el complejo caja TATA-TBP [90]. Por otro lado, las islas CpG, antes de sufrir cambios epigenéticos, se pueden encontrar en dicha conformación [91]. Por lo tanto, como una primera aproximación, se tomaron en cuenta cadenas tipo caja TATA y tipo islas CpG en la conformación ADN-B para la exploración de las propiedades reactivas (capacidad de interacción) de estas regiones.

Para relacionar el análisis de las interacciones débiles con la reactividad (posibilidad de interactuar con alguna proteína) asociada a una secuencia específica de ADN, se recurrió al concepto de polarizabilidad, el cual puede ser entendido como la tendencia de una molécula a modificar su densidad electrónica [92] como consecuencia de la inducción de momentos dipolo instantáneos, debidos a la presencia de un campo eléctrico externo [93]. Por lo tanto, mientras mayor sea la polarizabilidad de un sistema, mayor será su capacidad de formar in-

teracciones no-covalentes [94-96]. Con esto en mente, y dada la buena correlación que hay entre esta propiedad y la dureza química (η)⁵ [97], el descriptor η fue seleccionado para determinar cuál de las dos secuencias tiene mayor probabilidad de formar interacciones débiles. Este índice de reactividad fue calculado a partir de las energías de ionización (EI) y afinidades electrónicas (AE) verticales [98]:

$$\eta = EI - AE \quad (3.11)$$

Cadena	η (eV)
Tipo caja TATA ($n = 8$)	4.79
Tipo islas CpG ($n = 8$)	4.21

Tabla 3.6. Dureza química (η) para las cadenas tipo caja TATA y tipo islas CpG con $n = 8$.

En la Tabla 3.6 se muestran los valores de la dureza química, estimados mediante la Ecuación (3.11), para las cadenas tipo caja TATA y tipo islas CpG. De ésta, se puede concluir que la cadena tipo islas CpG es más blanda que la tipo caja TATA. Considerando que si una distribución de carga es menos polarizable, se espera que la respuesta a una perturbación externa (alguna interacción con una proteína, por ejemplo) sea menor que para un sistema más polarizable. Por lo tanto, se puede sugerir que, en la transcripción, el complejo transcripcional podrá interactuar más fácilmente con una región rica en GC que con una región rica en TA, resultado que coincide con lo conocido de la cinética transcripcional [56]. Cabe resaltar que el análisis de la dureza química concuerda con el análisis de las energías promedio de puente de hidrógeno y de apilamiento π (Figuras 3.3 y 3.6, respectivamente), ya que las cadenas tipo islas CpG cuentan con las interacciones más fuertes, pero no se ven afectadas por el número de pares de base dada la ausencia de efectos cooperativos (Figuras 3.15 y 3.16). Probablemente, esta sea la razón por la cual las islas CpG no tienen un tamaño definido [99], contrario a las cajas TATA.

Por último, como se mencionó anteriormente, tanto en los trabajos experimentales como en este análisis teórico, se ha concluido que, desde un punto de vista biológico, las interacciones de apilamiento π son más importantes que las interacciones de puente de hidrógeno. Por lo tanto, se puede asumir que cuando una proteína se acerca a la cadena de ADN, la densidad electrónica asociada a las nubes π podría ser la primera en ser afectada. Para corroborar lo dicho, en esta parte del análisis se recurrirá a las densidades de estado totales (DOS, por sus siglas en inglés) y parciales (pDOS, por sus siglas en inglés) de las cadenas tipo caja TATA y tipo islas CpG con $n = 8$, con la finalidad de determinar en cuál de los dos sistemas se encuentran energéticamente más disponibles los electrones π y, por ende, cuál de las dos secuencias puede establecer más fácilmente interacciones con alguna proteína, por ejemplo.

En la Figura 3.18 se muestran las DOS y pDOS de las cadenas tipo caja TATA y tipo islas CpG con $n = 8$. Como se puede observar, para ambas secuencias, los últimos niveles ocupados corresponden, principalmente, a orbitales tipo p , asociados a las interacciones de apilamiento π . Como era de esperarse, y en concordancia con el análisis de la dureza química, los orbitales p de la cadena tipo islas CpG son energéticamente más externos que aquellos pertenecientes a la cadena tipo caja TATA. Consecuentemente, una proteína con residuos aromáticos podrá

⁵A mayor dureza, menor polarizabilidad, y viceversa [97].

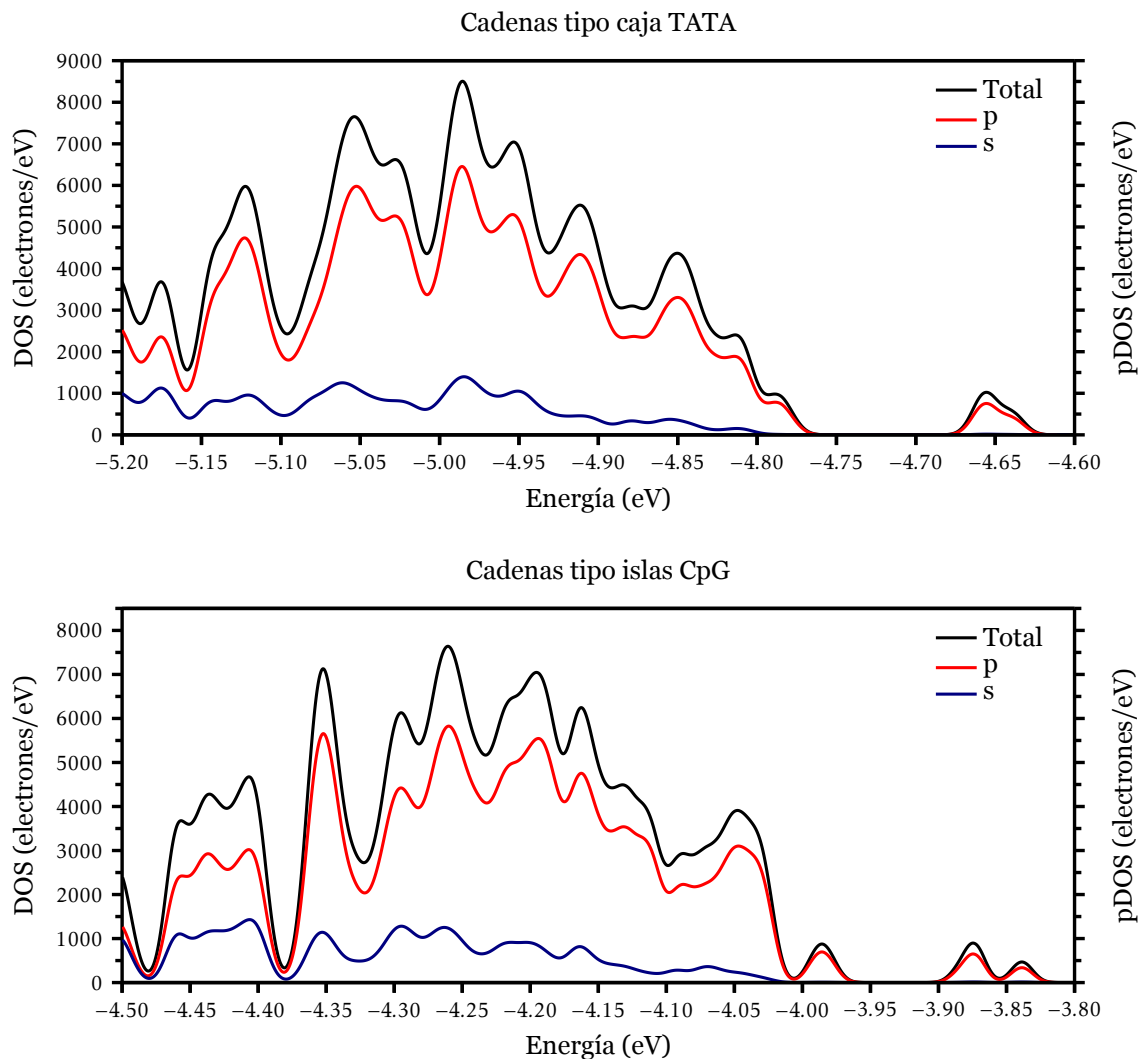


Figura 3.18. Densidades de estado totales (DOS, por sus siglas en inglés) y parciales (pDOS, por sus siglas en inglés) de las cadenas tipo caja TATA (arriba) y tipo islas CpG (abajo). Los niveles de Fermi de las cadenas tipo caja TATA y tipo islas CpG son: $E_F = -4.78$ eV y $E_F = -3.98$ eV, respectivamente.

interactuar con mayor facilidad con una región del ADN rica en G y C que con una rica en T y A, puesto que en las primeras las densidades electrónicas asociadas al apilamiento π se encuentran más disponibles. Estas secuencias suelen tener asociadas densidades electrónicas más deformables (menos duras). Por consiguiente, se puede concluir que las regiones con guanina y citosina serán las que formen interacciones no-covalentes más favorablemente.

4. Conclusiones

4.1 Conclusiones

4.2 Perspectivas a futuro

56

56

“Me he convertido en una especie de máquina para observar los hechos y extraer conclusiones.”

Charles Darwin

Una vez realizado el análisis de las interacciones de puente de hidrógeno y de apilamiento π en función del tamaño y composición de la cadena, en este capítulo se presentan las conclusiones a las cuales se llegó.

4.1 Conclusiones

En este trabajo se describieron las interacciones de puente de hidrógeno y de apilamiento π en el ADN en cadenas tipo caja TATA y tipo islas CpG, encontrando que ambas interacciones cambian en función del número de pares de base y del tipo de bases nitrogenadas que conforman a la doble hélice. Además, se demostró la existencia de efectos cooperativos, anti-cooperativos y aditivos en función del tamaño y composición de la cadena. Por otro lado, se determinó que las interacciones de puente de hidrógeno participan de forma más importante en la estabilización del dúplex; sin embargo, la contribución del apilamiento π no es despreciable. Adicionalmente, se encontró que los estados energéticos involucrados en la formación de las interacciones de apilamiento π son más externos, por lo que una molécula interactuará con ADN, en primer lugar, afectado dichas fuerzas. Finalmente, cabe destacar la existencia de una correlación entre el comportamiento biológico, la reactividad (capacidad de interacción) y las interacciones débiles en el ADN, dado que se observó que las cadenas tipo islas CpG cuentan con estados asociados a orbitales tipo p energéticamente más externos que aquellos provenientes de las cadenas tipo caja TATA.

4.2 Perspectivas a futuro

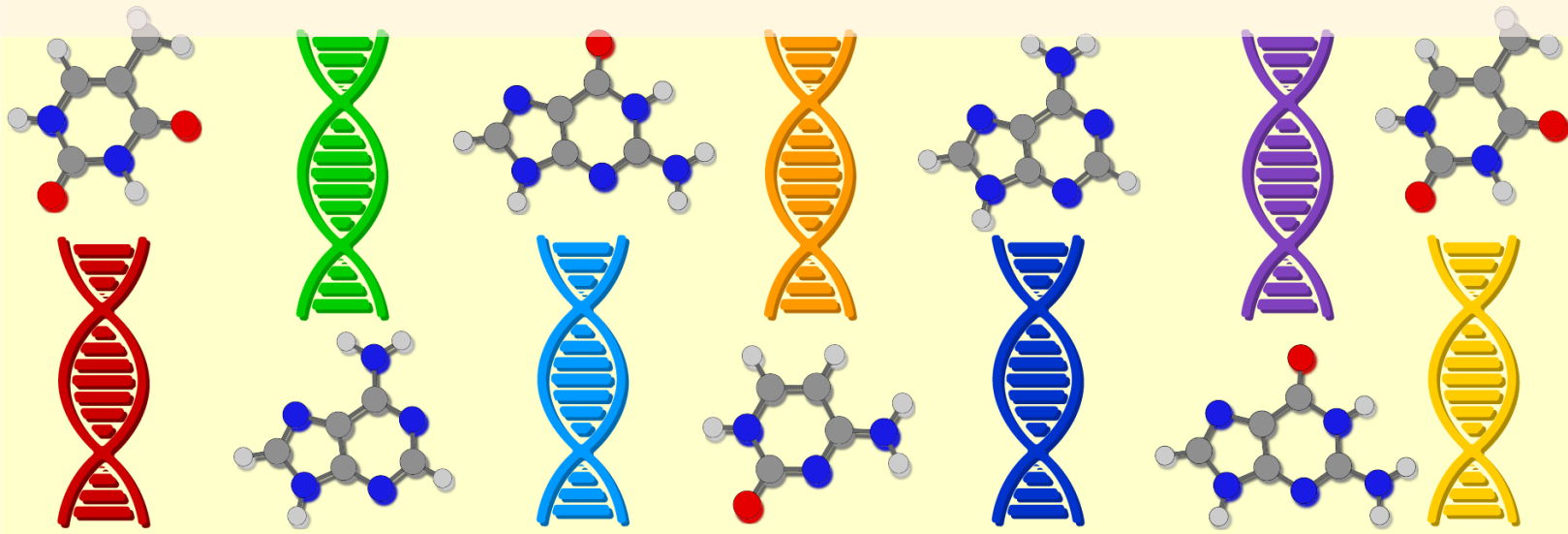
Pese a que al inicio de un proyecto se plantean los objetivos y los alcances del mismo, ciertos imprevistos y/o resultados inesperados provocan que algunos objetivos queden parcialmente resueltos y, adicionalmente, surge la generación de nuevas perspectivas y objetivos a alcanzar.

En el caso de este trabajo, las perspectivas a futuro pueden ser planteadas desde los siguientes puntos de vista:

- **Mejoramiento de la aproximación:** las aproximaciones utilizadas para estimar las energías de puente de hidrógeno y de apilamiento π implicaron tanto la “contaminación” energética debida a la presencia de otras interacciones, así como la remoción estructural de algunos componentes que pueden contribuir a la energía de dichas interacciones. Por lo tanto, a futuro, queda mejorar el método por el cual se estiman los valores energéticos asociados a los puentes de hidrógeno y al apilamiento π . Esto es de vital importancia puesto que, biológicamente, diferencias energéticas pequeñas pueden ser causa de la preferencia de alguna conformación o proceso bioquímico específico.
- **Interacciones débiles, secuencia y aplicaciones:** en este trabajo, las interacciones de puente de hidrógeno y de apilamiento π fueron estimadas para cadenas con pares de base alternantes de un solo tipo (T-A y G-C), por lo que aún falta describir a dichas interacciones en cadenas no alterantes ($5' - \dots TTTT \dots - 3'$ y $5' \dots - GGGG \dots - 3'$, por ejemplo) y en cadenas que contengan los cuatro tipos de bases nitrogenadas en su secuencia. Esto con el propósito de ampliar el panorama de los puentes de hidrógeno y del apilamiento π en las cadenas de ADN. Adicionalmente, aún queda por estimar el efecto de los cambios epigenéticos sobre las interacciones débiles de interés. Toda esta información es de relevancia puesto que brindará conocimiento, desde un punto de vista mecano-cuántico, sobre el rol de la secuencia de los pares de base y de los cambios epigenéticos en la lectura de la información contenida en el ADN, así como del papel

de las interacciones de puente de hidrógeno y de apilamiento π en los cambios (mutaciones) no deseados sobre una región genómica dada. Por otro lado, puesto que se puede analizar el efecto de estas fuerzas intermoleculares en las propiedades físicas y químicas del ADN, queda por investigar el uso de la molécula de ADN en dispositivos nano-electrónicos.

Artículos publicados



“Sólo cuando ha publicado sus ideas y hallazgos, el científico ha hecho su contribución; y sólo cuando lo ha hecho parte del dominio público de la erudición, [éste] puede realmente reclamarlo como propio. Porque su argumentación reside únicamente en el reconocimiento otorgado por sus pares en el sistema social de la ciencia, a través de la referencia de su trabajo”

Robert King Merton, 1993

En este capítulo se muestran las publicaciones científicas realizadas a partir de los resultados de este trabajo. La primera de ellas [100] trata sobre las interacciones de puente de hidrógeno y de apilamiento π en cadenas tipo caja TATA y sus contribuciones a la estabilización del dúplex. El segundo artículo [101] describe a las mismas interacciones, pero ahora en cadenas tipos islas CpG, haciendo una comparativa entre cadenas con timina y adenina, y cadenas con guanina y citosina, todo eso con el propósito de comparar sus propiedades y realizar conclusiones acerca de su comportamiento biológico.



ELSEVIER

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

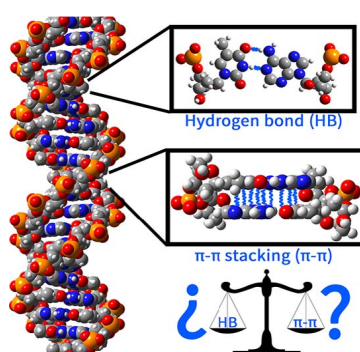
Electronic properties of DNA: Description of weak interactions in TATA-box-like chains

Jorge Gutiérrez-Flores^a, Estrella Ramos^{a,*}, Carlos I. Mendoza^a, Enrique Hernández-Lemus^b^a Instituto de Investigaciones en Materiales, Universidad Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, 04510 Mexico City, Mexico^b Computational Genomics, National Institute of Genomic Medicine, Mexico City, Mexico

HIGHLIGHTS

- The influence of the chemical environment on hydrogen bonds and π - π stacking in TATA-box-like B-DNA chains is studied.
- Novel strategies are proposed to estimate the non-covalent interactions energies present in B-DNA.
- An analysis of the contribution of hydrogen bond and pi-pi stacking to B-DNA stabilization is made.

GRAPHICAL ABSTRACT



ABSTRACT

DNA is one of the most important biomolecules since it contains all the genetic information about an organism. The tridimensional structure of DNA is a determinant factor that influences the physiological and biochemical mechanisms by which this molecule carries out its biological functions. It is believed that hydrogen bonds and π - π stacking are the most relevant non-covalent interactions regarding DNA stability. Due to its importance, several theoretical works have been made to describe these interactions, however, most of them often consider only the presence of two nitrogenous bases, having a limited overview of the participation of these in B-DNA stabilization. Furthermore, due to the complexity of the system, there are discrepancies between which involved interaction is more important in duplex stability. Therefore, in this project we describe these interactions considering the effect of chain length on the energy related to both hydrogen bonds and π - π stacking, using as model TATA-box-like chains with n base pairs ($n = 1$ to 14) and taking into consideration two different models: ideal and optimized B-DNA. We have found that there is a cooperative effect on hydrogen bond and π - π stacking mean energies when the presence of other base pairs is considered. In addition, it was found that hydrogen bonds contribute more importantly than π - π stacking to B-DNA stability; nevertheless, the participation of π - π stacking is not negligible: when B-DNA looks for a conformation of lower energy, π - π stacking interaction are the first to be optimized. All work was realized under the framework of DFT using the DMol³ code (M06-L/DNP).

* Corresponding author.

E-mail address: eramos@iim.unam.mx (E. Ramos).

1. Introduction

The hereditary basis of almost all life form is its genome, which consists of a long DNA chain that provides a complete set of all genetic information [1]. It is estimated that each of our cells has, approximately, 1.5 gigabytes of genetic information stored in this biomolecule [2]. Hence, DNA is considered as one of the molecules of greater biological relevance, since from this is possible to get several functional products (as proteins and different types of RNA) necessary for many biological processes.

DNA is conformed by nucleotides which are formed by a sugar, a phosphate group and a nitrogenous base: adenine (A), guanine (G), thymine (T) and cytosine (C). Through phosphodiester bonds between phosphate groups and sugars, these units lead to the formation of strands that interact non-covalently with each other to originate different DNA structures [3]. Under normal physiological conditions, DNA adopts a so-called B-DNA conformation [2,4] whose structure and stability is mainly determined by the number and type of weak interactions. There are different kinds of weak interactions, nevertheless, it is believed that hydrogen bonds (between nitrogenous bases) and π - π stacking (between base pairs) are the most relevant for DNA stabilization [5]. Therefore, understanding these interactions would explain the physical and chemical properties of DNA, and its role in biological

processes involving this molecule.

Due to the importance of π - π stacking and hydrogen bonds in stabilizing DNA duplex, several theoretical studies have been made with different approximations in order to estimate their energy [6–13]. Most of them consider only the presence of two nitrogenous bases, having a limited overview of the participation of these interactions in B-DNA stabilization.

On the other hand, due to the complexity of the system, there are discrepancies between which of the two interactions (hydrogen bonds or π - π stacking) participates more significantly in the stabilization of DNA duplex. For example, according to Smirnov et al. [14] π - π stacking interactions are more important than hydrogen bonds (a similar conclusion was obtained by Sen et al. for PNA-DNA duplexes [15]), while Zhang et al. [16] establish that hydrogen bonds contributes more in the B-DNA stabilization. Consequently, it is not clear which one plays the most important role in the duplex stabilization or if both are equally relevant.

In view of the above, the goal of this study is to estimate the associated energy to both hydrogen bonds and π - π stacking interactions in order to get a better description of these, considering the effect of chain length (presence of other base pairs). Furthermore, with the obtained results, the contribution of each interaction to B-DNA stabilization is estimated.

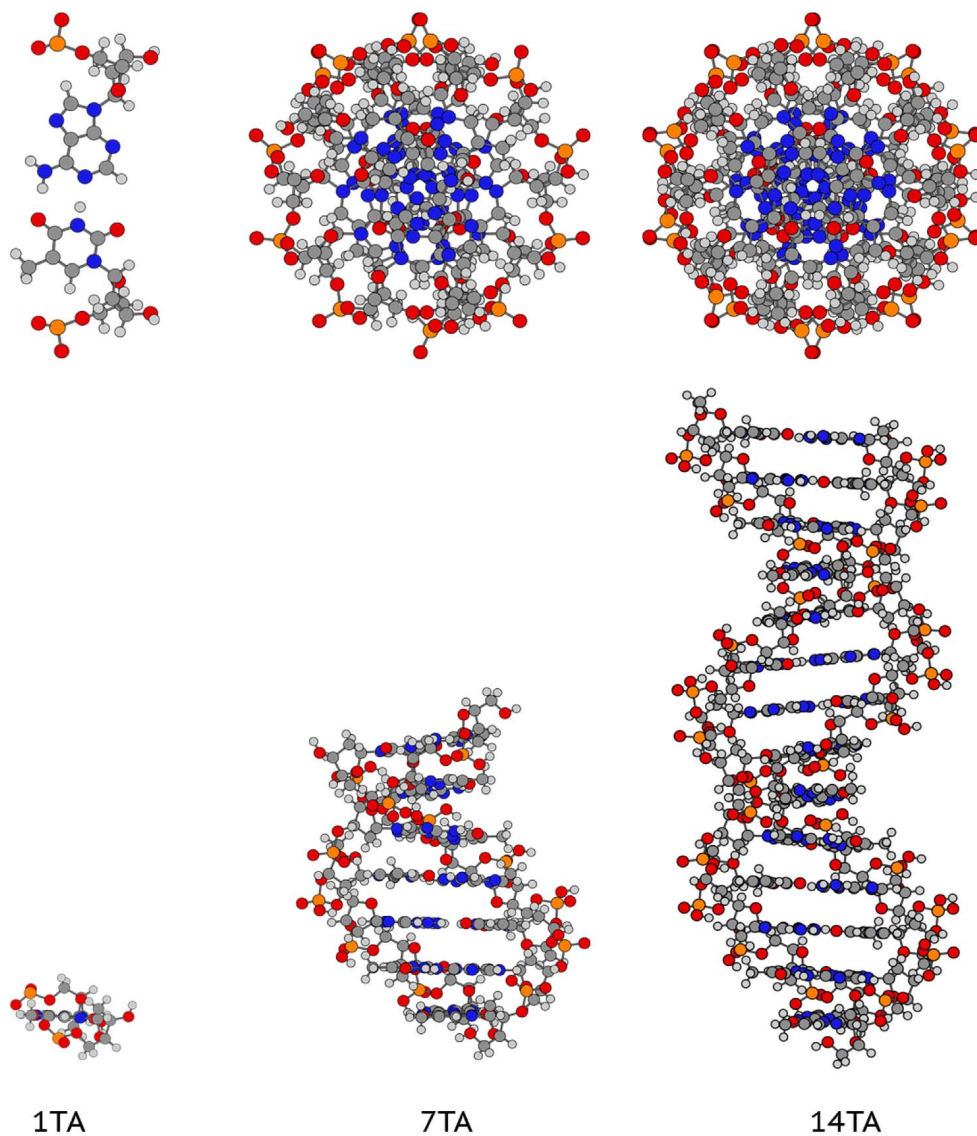


Fig. 1. Examples of B-DNA chains with different base pairs (n) as indicated. Frontal (up) and lateral (down) view.

2. Methods

To carry out this work, based on TATA box's sequence (a well known promoter region conserved in a wide range of eukaryotic and viral genes [17] which consists in a repetition of T and A bases), we modeled TATA-box-like chains of different length, i.e. different number of base pairs (n), from $n = 1$ to $n = 14$. The charge of the phosphate groups was neutralized by the addition of a hydrogen atom, assuming that the presence of these atoms would not affect significantly the hydrogen bond and π - π stacking interactions [11,18].

In order to estimate the energy associated to hydrogen bonds and π - π stacking, we took into account two different cases:

- (a) Ideal
- (b) Optimized

For the ideal case, structures were obtained with the open-source software *make-na* [19]. This software considers the distance between adjacent nucleobases and between phosphate groups, rotation and torsion angles, among other parameters (based on experimental data), in order to get ideal B-DNA chains, i.e., chains that obey the symmetry restrictions required for these systems (Fig. 1). Let us stress that even if these chains satisfy the symmetry restrictions, they are not relaxed into their minimum energy configuration; however, they provide a good starting point to estimate weak interaction energies. Subsequently, ideal structures were submitted to a geometry optimization process (Fig. 2) to get a better description of the stabilizing interactions present in the chain.

For hydrogen bond mean energy (\bar{D}_{HB}) estimation (see Fig. 3), single point energy calculations for each chain and for each of their strands were performed, and the following equation was proposed:

$$\bar{D}_{HB} = -\bar{E}_{HB} = -\frac{E_n - (E_{n_1s} + E_{n_2s})}{n} \quad (1)$$

where \bar{E}_{HB} is the mean hydrogen bond formation energy; E_n represents the total energy of a n -base pair chain; E_{n_1s} and E_{n_2s} the total energy of chain's strands and n the number of base pairs.

The Espinosa-Molins-Lecomte (EML) equation [20] was used to estimate hydrogen bond energies:

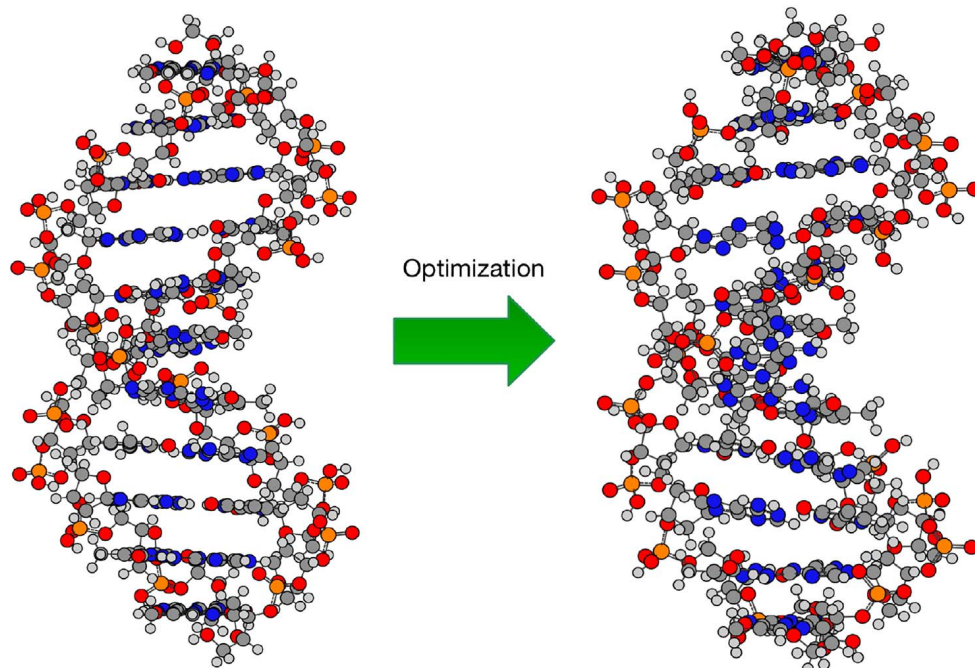


Fig. 2. Lateral view of a B-DNA chain before (left) and after (right) optimization.

$$E_{HB} = 0.5V(r) \quad (2)$$

Based on Quantum Theory of Atoms in Molecules (QTAIM) [21], this equation takes into account the electron distribution at $(3, -1)$ bond critical points (BCPs) of hydrogen bonds and correlates local potential energy density ($V(r)$) with hydrogen bond energy (E_{HB}).

Whereas for π - π staking mean energy (\bar{D}_π) estimation, assuming that the major contribution for this interaction corresponds to nitrogenous bases, phosphate groups and sugar moieties were removed from DNA chains, while hydrogen atoms were added instead, leaving nitrogenous bases in the chain configuration (see Fig. 4). Single point energy calculations were performed for each base pair and the following equation was employed:

$$\bar{D}_\pi = -\bar{E}_\pi = -\frac{E_{n_\pi} - \sum_{i=1}^n E_{ibp}}{N_\pi} \quad (3)$$

where \bar{E}_π is the mean π - π stacking formation energy; E_{n_π} corresponds to the chain total energy without sugars and phosphate groups; E_{ibp} the total energy of base pairs and N_π the number of spaces between base pairs.

All calculations were carried out under Density Functional Theory (DFT) framework [22,23], using the DMol³ code [24] implemented in the Materials Studio 8.0 software suite [25], employing the M06-L functional [26] (a meta-GGA functional that, among other things, was parameterized for the good description of weak interactions like π - π staking and hydrogen bonds) and a double numeric basis set with polarized functions (DNP). Basis set superposition error (BSSE) corrections were performed by means of counterpoise method [27] implemented in DMol³ code. The occupied convergence tolerance for all calculations were 1×10^{-5} Ha, 2×10^{-5} Ha, 4×10^{-3} Ha/ and 5×10^{-3} for SFC cycles, energy, gradient and displacement, respectively.

For topology analysis, single point energy calculations were carried out using the Gaussian 09 suit of programs of computational chemistry [28] (M06-L/6-31G(d,p)). Finally, electron density was analyzed utilizing Multwfn software [29].

3. Results and discussion

It is important to mention that ideal B-DNA chains are built by imposing symmetry restrictions [19,30], obtaining chains such as those

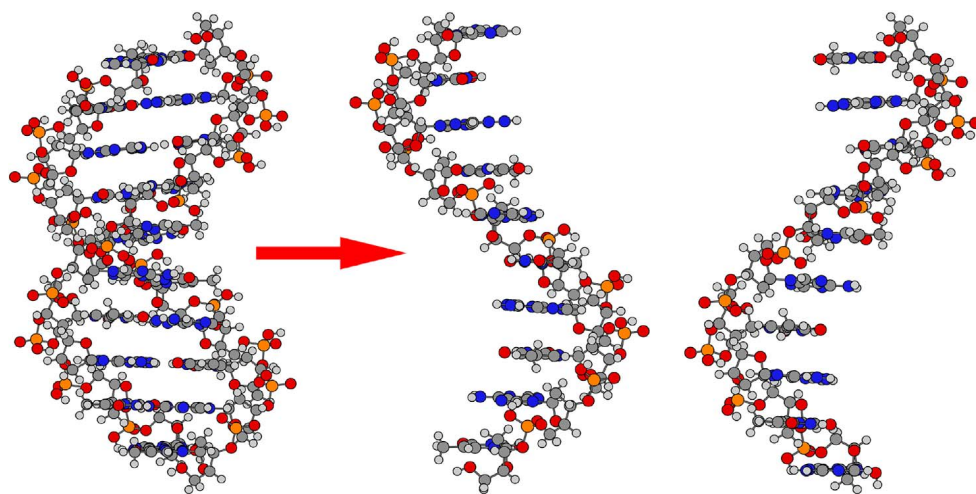


Fig. 3. Illustration of how hydrogen bond mean energy (\bar{D}_{HB}) is estimated.

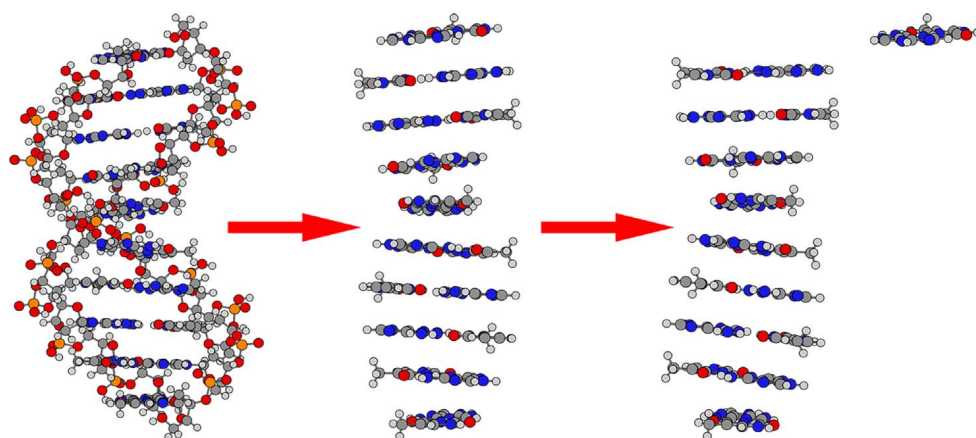


Fig. 4. Illustration of how π - π staking mean energy (\bar{D}_{π}) is estimated.

shown in Fig. 1. In this figure, it is possible to observe a C_{10} axis of symmetry and total planarity of nitrogenous bases. Due to the relation between tridimensional arrangement of nitrogenous bases and the interactions involved in double helix stabilization [30], a change in conformation will affect the efficiency and strength of the latter. For that reason, it is interesting to study the effect of the rearrangement of the chain (optimization process) on hydrogen bonds and π - π stacking.

As mentioned before, it is considered that hydrogen bonds and π - π stacking interactions are the most important in B-DNA stabilization [5]. For that reason, we start with the estimation of interaction energies and the analysis of chain length effect over these. In a later step, a charge distribution analysis is made and, finally, the relative contribution of these interactions to B-DNA stabilization is estimated.

3.1. Hydrogen bonds

Eq. (1) was used in order to estimate the hydrogen bond mean energy (\bar{D}_{HB}), i.e. the required energy to break up these interactions. When calculating the mean hydrogen bond energy by means of Eq. (1), we are neglecting the presence of other noncovalent interactions when strands are separated like shown in Fig. 3. In order to validate this approach, we made a topology analysis of TATA-box-like chain with $n = 7$ (as an example) and estimated the hydrogen bond mean energy using Eq. (2) (see Tables S1 and S2 in SI). As the energy is calculated with $V(r)$ at hydrogen bonds (3, -1) BCPs, we assume that this corresponds only to hydrogen bonds. The difference between the energies calculated with Eqs. (1) and (2) (Table 1) is less than 3 kcal mol⁻¹ (around 16%); therefore, we consider that Eq. (1) adequately estimates the hydrogen bond mean energy. Hydrogen bond mean energy as a function of the

Table 1
Hydrogen bond mean energies calculated by means of Eqs. (1) (first two columns) and (2) (last two columns) for chains with $n = 7$.

		\bar{D}_{HB} (kcal mol ⁻¹)			
	Eq. (1)	Optimized	Eq. (2)	Optimized	
Ideal			Ideal		
17.84		17.65	15.01	14.44	

number of base pairs (n) is shown in Fig. 5 (values are available in Table S3 in the SI).

As seen in Fig. 5, \bar{D}_{HB} increases as n does. Consequently, the presence of other base pairs tends to raise the hydrogen bond energy, i.e. there is a cooperative effect. From the same figure, it is clear that the curve representing the ideal case has the greatest hydrogen bond mean energies. When the structure is relaxed, the planarity of base pairs is lost. As a consequence, the strength of hydrogen bonds is altered, resulting in weaker hydrogen bonds for optimized chains. Nevertheless, hydrogen bond mean energy differences between ideal and optimized cases are less than 1 kcal mol⁻¹ and, when number of base pairs increases difference, tends to vanish. These differences could be associated with the fact that when ideal structure is submitted to a geometry optimization process, base pairs are distributed according to neighbors pairs. Therefore, as n grows the movements of base pairs are more restricted and the energy estimated is more similar to ideal case.

Oscillations present in curves (Fig. 5) could be attributed to finite size effects. Indeed, by symmetry restrictions in an ideal chain with a very large number of base pairs, any TA or AT base pair would be

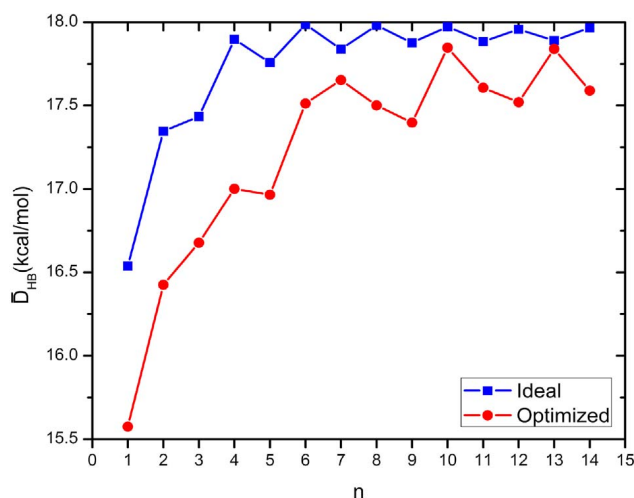


Fig. 5. Hydrogen bond mean energy as a function of DNA chain length. Blue symbols correspond to ideal chains while red symbols to optimized structures. Lines connect the points for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

equivalent. In the case of a finite length ideal B-DNA chain, the units at the extremes will contribute with a different quantity to hydrogen bond energy.

Using n as the total number of base pairs and distinguishing chains by its parity, the hydrogen bond mean energy can be written as

$$\bar{D}_{HB} = \begin{cases} \epsilon_0 + \frac{\Delta E_{even}}{n} & n \text{ even} \\ \epsilon_0 + \frac{\Delta E_{odd}}{n} & n \text{ odd} \end{cases} \quad (4)$$

where ϵ_0 is the interaction energy between the nitrogenous bases and ΔE_{even} and ΔE_{odd} are the contribution from the base pairs at the edges of the chain (border effects), when the number of base pairs is even or odd, respectively. The numerical values of the edge effects are unknown, however, from the calculated values (Fig. 5 and Table S1), it can be observed that \bar{D}_{HB} takes an approximately constant value for even n , with $n \geq 4$, suggesting that $\Delta E_{even} \approx 0$.

Since we expect that this analysis would be more accurate for large chains, to evaluate ϵ_0 and ΔE_{odd} we take into account the values of the largest chains ($n=13$ and $n=14$). We have that $E_{n=13} \approx 17.89 \text{ kcal mol}^{-1}$, and $E_{n=14} \approx 17.97 \text{ kcal mol}^{-1}$, consequently, from Eq. (4) we get

$$\epsilon_0 = \bar{D}_{HB} \approx 17.97 \text{ kcal mol}^{-1}$$

and

$$\Delta E_{odd} = n(\bar{D}_{HB} - \epsilon_0) \approx -1.01 \text{ kcal mol}^{-1}$$

Hence, with values of ϵ_0 and ΔE_{odd} , Eq. (4) can be rewritten as

$$\bar{D}_{HB} = \begin{cases} 17.97 \text{ kcal mol}^{-1} & n \text{ even} \\ 17.97 \text{ kcal mol}^{-1} - \frac{1.01 \text{ kcal mol}^{-1}}{n} & n \text{ odd} \end{cases} \quad (5)$$

The values obtained from Eq. (5) (Table S4 in SI) are shown in Fig. 6. As can be observed, this expression adjusts satisfactorily to the estimated values for the hydrogen bond mean energy for $n > 3$. The reason why the analytical development fails for small n is that it assumes that chain edges do not affect the hydrogen bond mean energy and this is only valid in larger chains.

From the results shown before (Figs. 5 and 6), it is clearly important to consider more than 4 base pairs for the adequate estimation of the hydrogen bond energy; otherwise, hydrogen bond energy could be underestimated.

In Table 2 we show the values of the hydrogen bond mean distances

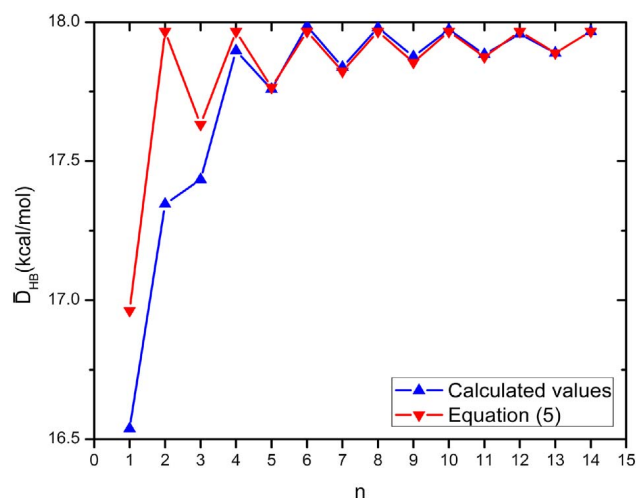


Fig. 6. Hydrogen bond mean energy as a function of the DNA chain size (represented by n) for the ideal case: theoretical values (blue) and analytical development (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Hydrogen bond mean distances for the studied cases. (ideal chains: $\sigma \leq 9.00 \times 10^{-4} \text{ \AA}$; optimized chains: $\sigma \leq 1.40 \times 10^{-2} \text{ \AA}$).

TA Units	Ideal case		Optimized case	
	R_1 O	R_2 O	R_1 O	R_2 O
1	1.871	1.810	1.884	1.835
2	1.870	1.810	1.856	1.833
3	1.870	1.810	1.868	1.819
4	1.870	1.810	1.869	1.825
5	1.870	1.810	1.867	1.816
6	1.870	1.810	1.868	1.810
7	1.870	1.810	1.882	1.799
8	1.870	1.810	1.868	1.810
9	1.870	1.810	1.855	1.823
10	1.870	1.810	1.889	1.801
11	1.870	1.810	1.862	1.816
12	1.870	1.810	1.851	1.826
13	1.870	1.810	1.878	1.802
14	1.870	1.810	1.855	1.825

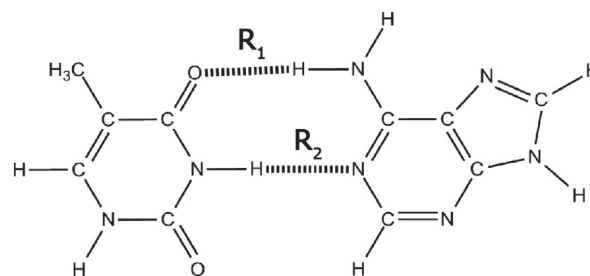


Fig. 7. Sketch indicating the distances between the atoms for both, A and T.

between nitrogenous bases for the ideal and optimized chains. Notation used to name interactions between A and T is given in Fig. 7. It is important to remark that we considered only the presence of two hydrogen bonds (with separation distances R_1 and R_2 , respectively).

In general, from Table 2, it can be noticed that central hydrogen bonds (R_2) are stronger than the outer ones (R_1), i.e., they have smaller associated distances for a given chain (in compliance with discussion in Ref. [31]). Distances are in concordance with calculated values using Eq. (2) (see Tables S1 and S2 in SI). Calculated distances obtained for both cases have the same trend of experimentally reported values [32]

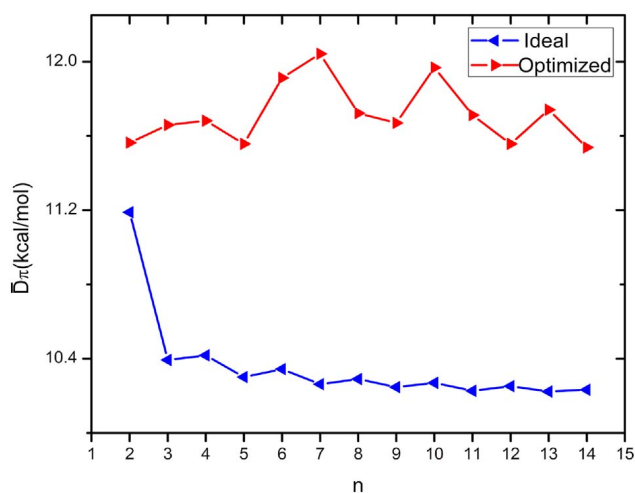


Fig. 8. π - π stacking mean energy as a function of DNA chain length. Blue symbols correspond to ideal chains while red symbols to optimized structures. Lines connect the points for clarity. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2.93–2.95 for R_1 and 2.82–2.85 for R_2). Difference of ≈ 1 Å between the calculated data and the reported experimentally is expected due to hydrogen atoms are not observable by X-ray diffraction. Additionally, these differences could be associated to errors related to the calculation method used (there is not an explicit correction for the noncovalent interactions) and to the fact that all calculations were performed in gas phase. For ideal chains, as it can be expected, mean distances are essentially the same for all the chains. For all of these chains symmetry restrictions are satisfied, thus distance will not change with the addition of more base pairs.

3.2. π - π stacking

π - π stacking usually refers to attractive interactions between π systems [33] and, therefore, are often present in unsaturated organic groups [34]. They depend on both geometry and force factors that favor the juxtaposition of two aromatic molecules, in order to allow the direct contact between their corresponding π systems [5]. Eq. (3) was used to estimate the energy related to these interactions along TATA-box-like DNA chains (between base pairs). Results are shown in Fig. 8 (values are available in Table S5 in SI).

In contrast to hydrogen bonds (Fig. 5), π - π stacking interactions are favored when geometry optimization is carried out, i.e. π - π stacking mean energy is greater in optimized chains than in ideal chains (on average, difference between ideal and optimized cases is around $1.36 \text{ kcal mol}^{-1}$). As mentioned before, ideal chains have smaller hydrogen bond associated distances. Hence, the base pairs are arranged in order to maximize the contact between them, and, therefore, the π - π stacking mean energy is increased.

Due to differences between hydrogen bond and π - π stacking mean energies in optimized cases, it can be suspected that when B-DNA reaches its minimum energy configuration, π - π stacking interactions are optimized first, before hydrogen bonds. This may be the reason why Smirnov et al. [14] concluded that π - π stacking contribution was more important than that proceeding from hydrogen bonds when DNA polymerase was able to add non-polar nitrogenous bases (bases without the capability to interact with other nitrogenous base by means of hydrogen bonds) during DNA replication process.

3.3. Charge transfer

Aiming to a deeper examination of hydrogen bonds, an analysis of the electronic density accumulation as a function of chain length was

made over regions where hydrogen bonds are present between A and T bases ($\text{N-H}\cdots\text{X}$, where $\text{X} = \text{N}, \text{O}$; see Fig. 7). In an intermolecular interaction the atomic charge is shared between them, i.e., part of the electropositive atoms charge is transferred to more electronegative atoms. Hence, the larger the charge transfer, the larger the hydrogen bond strength.

For this purpose, Charge Separation Index (CSI), a measure of the local polarity of a molecule, is obtained from the sum of absolute values of charges q of each atom i in the region of interest [35]:

$$\text{CSI} = \sum_i |q_i| \quad (6)$$

A different way to interpret this descriptor is as the measure of the degree of separation between negative and positive charges of region of interest [36]. The larger the CSI value, the larger the polarization in that region and larger the hydrogen bond strength.

To obtain this descriptor, we performed a Mulliken population analysis. We considered chains of $n = 1, 3, 5$ and 7 , since only chains with an odd number of base pairs have a central one which is equidistant to both chain edges which is convenient to reduce their influence as much as possible.

In Table 3 we show CSI descriptor values for $\text{N-H}\cdots\text{X}$ (where $\text{X} = \text{N}, \text{O}$) regions. Values are organized according to the nomenclature introduced in Fig. 7. It can be observed that for the two cases studied, the central hydrogen bond (R_2) is stronger than the external one (R_1). This fact is in accord with distances analyzed before (Table 2) and with the topology analysis performed for ideal and optimized chains with $n = 7$ (Tables S1 and S2 in SI). Other works have reported the same behavior [7].

Due to system symmetry, as it could be expected, differences in CSI values in ideal chains are not significantly large. However, it can be

Table 3

Charge Separation Index (CSI) for hydrogen bond regions ($\text{N-H}\cdots\text{X}$, where $\text{X} = \text{N}, \text{O}$; see Fig. 7) between T and A bases.

Chain	Base pair	Interaction	CSI		
			Ideal	Optimized	
1TA	1	R_1	1.426	1.422	
		R_2	1.589	1.531	
3TA	1	R_1	1.426	1.416	
		R_2	1.580	1.521	
	2	R_1	1.413	1.417	
		R_2	1.576	1.512	
	3	R_1	1.427	1.435	
		R_2	1.581	1.530	
5TA	1	R_1	1.426	1.413	
		R_2	1.579	1.522	
	2	R_1	1.412	1.419	
		R_2	1.577	1.520	
	3	R_1	1.419	1.418	
		R_2	1.572	1.521	
	4	R_1	1.415	1.412	
		R_2	1.575	1.509	
	5	R_1	1.425	1.436	
		R_2	1.581	1.528	
	7TA	1	R_1	1.425	1.405
			R_2	1.578	1.520
2		R_1	1.411	1.410	
		R_2	1.575	1.529	
3		R_1	1.414	1.413	
		R_2	1.573	1.522	
4		R_1	1.414	1.410	
		R_2	1.575	1.520	
5		R_1	1.417	1.411	
		R_2	1.572	1.527	
6		R_1	1.414	1.408	
		R_2	1.575	1.519	
7		R_1	1.425	1.416	
		R_2	1.581	1.536	

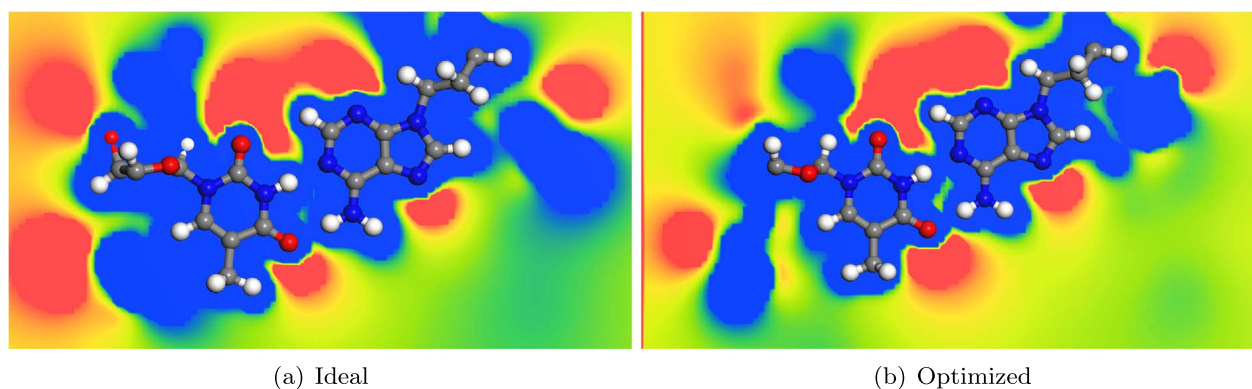


Fig. 9. Electrostatic potential planes for central base pairs of ideal (a) and optimized (b) chains with $n = 7$ (scale: from -0.04 a.u. (red) to 0.04 a.u. (blue)). Top view. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

noted that when these structures are relaxed (optimized cases) there is a loss of charge (CSI values decrease), and hydrogen bonds tend to be weaker. This is in agreement with hydrogen bond mean energies analysis (Fig. 5), where hydrogen bonds in optimized cases were weaker than those in ideal cases due to a loss of planarity.

3.4. Electrostatic potentials

With the aim to complete the analysis made before for hydrogen bond and π - π stacking interactions, electrostatic potentials and molecular electrostatic potentials are analyzed in order to shed light on charge distribution along studied B-DNA chains. For potential electrostatic planes, central base pairs are studied in order to avoid border effects.

As previously highlighted, when chains are relaxed, the hydrogen bond mean energy tends to decrease (Fig. 5 and Table 3) due to base pairs planarity loss. Likewise, this can be observed in the electrostatic potential planes showed in Fig. 9. In this figure, a top view of the electrostatic potential for central base pair of ideal and optimized chains with $n = 7$ is shown. As can be seen, in the optimized case there are more negative (rich in electrons) zones around hydrogen bonds than in ideal case, this is more clear in oxygen atom of N-H...O hydrogen bond. As stronger hydrogen bonds are associated with ideal chains, electrons of electronegative atoms are more shared in ideal chains hydrogen bonds, therefore the associated area will be smaller. Hence, planarity of base pairs plays an important role in how electrons are shared and, consequently, in hydrogen bond strength.

Contrary to hydrogen bonds, as showed before (Fig. 8), π - π stacking interactions are favored when chains are optimized. This is reflected in Fig. 10, where a lateral view of central base pairs of ideal and optimized chains with $n = 7$ is shown. Negative zones between base pairs are bigger in optimized case, indicating that contact of π electron density is

greater in this case. To get a better overview, electrostatic potentials for some ideal and optimized chains with an odd number of base are shown in Figs. 11 and 12, respectively. Yellow regions correspond to electrostatic potential minima (negative zones) while the blue ones correspond to maxima. From these figures, it is clear that electron density overlap between base pairs is greater in optimized chains. Therefore, the presence of other base pairs tends to optimize, in first place, π - π stacking.

Based on the results shown previously, it could be possible to confirm that π - π stacking interactions are firstly improved along DNA chain and lately hydrogen bonds. That is the reason why DNA polymerase has the capability to add nitrogenous bases during DNA replication despite bases are polar or non-polar, as reported by Smirnov et al. [14].

3.5. Contribution to stabilization

Finally, we determined the contributions of hydrogen bonds and π - π stacking in B-DNA stabilization, assuming that these are the most important non-covalent interactions along DNA duplex. To this end, we used total hydrogen bond energy (required energy to separate strands in DNA duplex):

$$D_{HB} = -E_{HB} = -(E_n - (E_{n_{1s}} + E_{n_{2s}})) \quad (7)$$

and total π - π stacking energy (required energy to separate all base pairs that form the chain):

$$D_{\pi} = -E_{\pi} = -E_{n_{\pi}} - \sum_{i=1}^n E_{i_{bp}} \quad (8)$$

Then, the contribution rate was estimated using the following equation:

$$\text{Contribution} = \frac{D_i}{D_{HB} + D_{\pi}} \cdot 100 \quad (9)$$

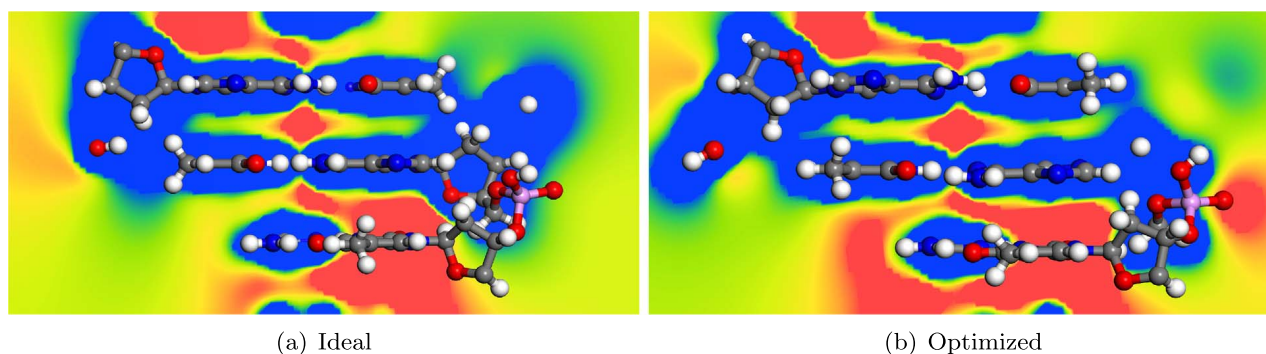


Fig. 10. Electrostatic potential planes for central base pairs of ideal (a) and optimized (b) chains with $n = 7$ (scale: from -0.04 a.u. (red) to 0.04 a.u. (blue)). Lateral view. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

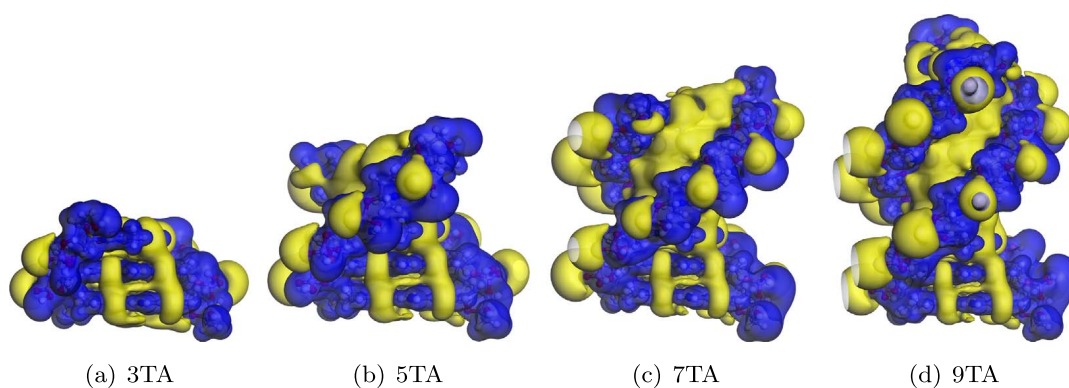


Fig. 11. Molecular electrostatic potential maps for ideal chains. Yellow regions represent minima, while blue ones represent maxima. Isovalue = 0.03 a.u. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

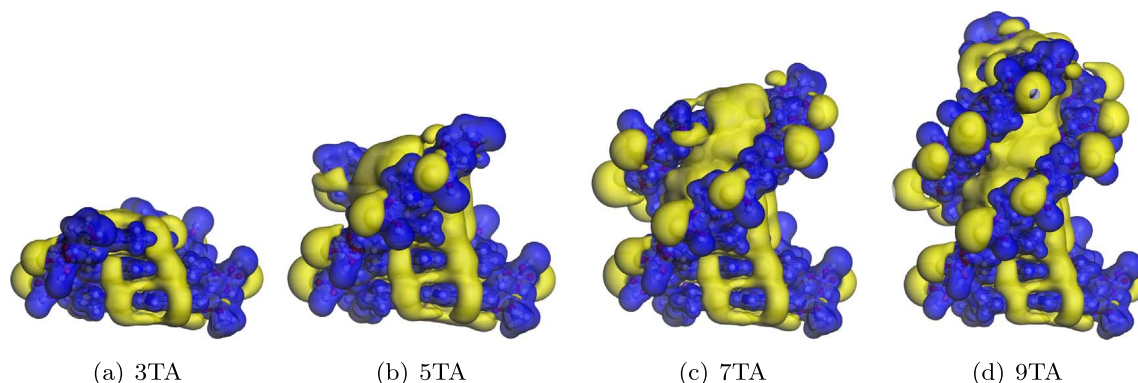


Fig. 12. Molecular electrostatic potential maps for optimized chains. Yellow regions represent minima, while blue ones represent maxima. Isovalue = 0.03 a.u. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

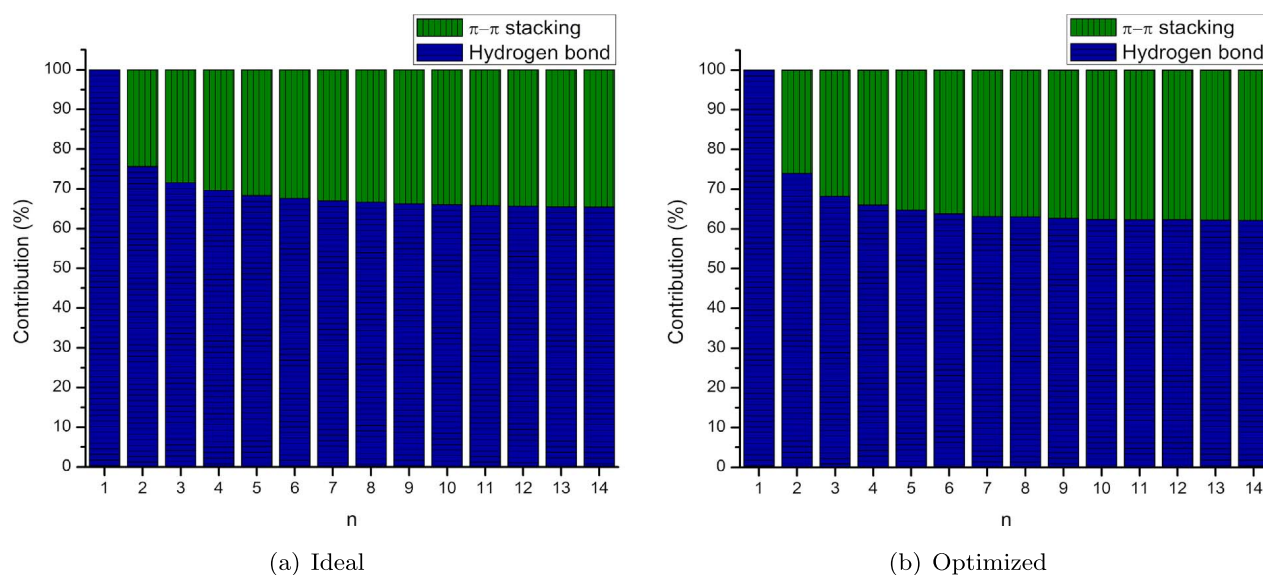


Fig. 13. Contribution rate of hydrogen bond and π - π stacking interactions for ideal (a) and optimized (b) cases.

where D_i represents D_{HB} or D_{π} , depending on which energy is being described. Results are shown in Fig. 13.

For both ideal and optimized cases, hydrogen bonds has the greater contribution to B-DNA stabilization, in concordance with what was reported by Zhang et al. [16]. However, the contribution of π - π stacking is not negligible.

Discrepancies between which interaction is more important in B-DNA stabilization could be related in how they are examined. For

example, when B-DNA is studied out of biological processes (like Atomic Force Microscopy (AFM) [16], for example), hydrogen bonds predominate over π - π stacking due to B-DNA strands are separated artificially. On the other hand, when interactions are studied by means of biological processes (like DNA replication [14], for example) π - π stacking is considered more important than hydrogen bond because of, in this approximation, interactions contribution is determined in how an enzyme adds a nucleotide and how this is stabilized in the growing

chain. All the mentioned before is in agreement with our results: π - π stacking interactions are firstly optimized, explaining why DNA polymerase is able to add nucleotides through π - π stacking no matter if nucleotides can or not form hydrogen bonds (as found by Smirnov et al. [14]). Nevertheless, when interactions are studied as proposed in Figs. 3 and 4 it is found that hydrogen bonds are more important than π - π stacking for B-DNA stabilization as reported by Zhang et al. [16]. All depends on the context and approximation used to determine interaction contributions.

It would be very interesting to contrast our results with experimental data. However, at the best of our knowledge, no experimental data of TATA-like-box chains is available to make comparisons with such results.

We have some proposals of experimental designs that may provide quantitative comparisons with our work. To this end we may look at a number of techniques which have resulted useful to determine experimentally the energetics of DNA duplex formation. A first common step to all these methods would consist in the laboratory synthesis of short DNA sequences with the required amount of TA base pairs that may provide the genomic sample for all the different methods. One possible approach would be using thermal DNA denaturation, tracked by either isothermal titration calorimetry coupled with spectrophotometric assessment [37] or by the use of a compounded NanoDrop device combining spectrophotometric with fluorometric measurements on microvolumes [38]. Another approach is based in elastic properties of DNA sequences and the use of atomic force microscopy [39] to provide approximations based on DNA stretching [40,41].

In brief, once that synthetic sequence specific constructs are available there are a number of experimental techniques that may be used as experimental comparisons to our theoretical calculations. Most of these are based on well-established principles in DNA analytics.

4. Conclusions

In this work, we performed an analysis of the influence of chain length in both, hydrogen bonds and π - π stacking, assuming that these are the most important non-covalent interactions in B-DNA stabilization. For this purpose, TATA-box-like chains (ideal and optimized cases) were used.

For hydrogen bonds, we found that they become weaker when a geometry optimization process is carried out. Whereas, π - π stacking interactions tends to be stronger. Therefore, it is proposed that when B-DNA looks for a conformation of lower energy, π - π stacking interactions are the first to be optimized, lately hydrogen bonds. This fact could explain DNA polymerase ability to add non-polar nitrogenous bases during DNA replication, as reported by Smirnov et al. [14].

With our results, it is clear that the presence of other base pairs have a cooperative effect on both hydrogen bond and π - π stacking mean energies, i.e., as chain length increases, these interactions become stronger. Important periodic oscillations were observed only for hydrogen bond mean energies due to border effects. Eq. (5) is proposed to estimate the hydrogen bond mean energies of TATA-box-like chains of any size.

Finally, the contribution of each interaction to B-DNA stabilization was calculated. According to these results, hydrogen bonds contribute more importantly in duplex stabilization than π - π stacking, in agreement with that reported by Zhang et al. [16]. Nevertheless, the contribution of π - π stacking is not negligible. The disagreements between which interaction is more important are due to the method and approximation used to study these interactions.

Acknowledgment

We want to thank to Carmen Nina Pastor Colon, Ateet Dutt and Julio César Aguilar Cordero for their useful comments. We also thank the computational support of Joaquín Morales, Alberto López and

Alejandro Pompa from IIM-UNAM and to DGTIC-UNAM for their supercomputing services. We acknowledge partial financial support from DGAPA, UNAM PAPIIT grant numbers IB-101712 (ER) and IN-110516 (CIM). The authors were also funded by grant number 179431/2012 from CONACyT, as well as Mexican Federal funding (grant number 173-2017) from the National Institute of Genomic Medicine (INMEGEN). JGF acknowledges support from CONACyT through the scholarship grant 336100. EHL also acknowledges support from the 2016 Marcos Moshinsky Research Chair in Physical Sciences.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bpc.2017.11.008>.

References

- [1] J.E. Kreb, E.S. Goldstein, S.T. Kilpatrick, Lewin's Genes, Johns and Bartlett Publishers, 2010.
- [2] D. Goodsell, DNA, (2001) <http://www.rcsb.org/pdb/101/motm.do?momID=23> Retrieved January, 2017.
- [3] T. Ohyama, DNA Conformation and Transcription, Springer US, 2005.
- [4] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids, Nature 171 (1953) 737–738.
- [5] E.T. Kool, Hydrogen bonding, base stacking, and steric effects in DNA replication, Annu. Rev. Biophys. Biomol. Struct. 30 (2001) 1–22.
- [6] C. Fonseca, F. Matthias, J. Snijders, E. Jan, Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment, J. Am. Chem. Soc. 122 (2000) 4117–4128.
- [7] H. Szatyłowicz, N. Sadlej-Sosnowska, Characterizing the strength of individual hydrogen bonds in DNA base pairs, J. Chem. Inf. Model. 50 (2010) 2151–2161.
- [8] A. Asensio, N. Kobko, J. Dannenberg, Cooperative hydrogen-bonding in adenine-thymine and guanine cytosine base pairs. Density functional theory and Möller-Plesset molecular orbital study, J. Phys. Chem. A 107 (2003) 6441–6443.
- [9] J. Grunenberg, Direct assessment of interresidue forces in Watson-Crick base pairs using theoretical compliance constants, J. Am. Chem. Soc. 126 (2004) 16310–16311.
- [10] H. Dong, W. Hua, S. Li, Estimation on the individual hydrogen-bond strength in molecules with multiple hydrogen bonds, J. Phys. Chem. A 111 (2007) 2941–2945.
- [11] C. Matta, N. Castillo, R. Boyd, Extended weak bonding interactions in DNA: π -stacking (base-base), base-backbone, and backbone-backbone interactions, J. Phys. Chem. B 110 (2006) 563578.
- [12] P. Mignon, S. Loverix, J. Steyaert, P. Geerlings, Influence of the π - π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases, Nucleic Acids Res. 2005 (2005) 17791789.
- [13] M. Swart, T. van der Wijst, C. Fonseca Guerra, F.M. Bickelhaupt, π - π stacking tackled with density functional theory, J. Mol. Model. 13 (2007) 1245–1257.
- [14] S. Smirnov, T.J. Matray, E.T. Kool, C. de los Santos, Integrity of duplex structures without hydrogen bonding: DNA with pyrene paired at abasic sites, Nucleic Acids Res. 30 (2002) 5561–5569.
- [15] A. Sen, P.E. Nielsen, Hydrogen bonding versus stacking stabilization by modified nucleobases incorporated in PNA-DNA duplexes, Biophys. Chem. 141 (2009) 29–33.
- [16] T. Zhang, C. Zhang, Z. Dong, Y. Guan, Determination of base binding strength and base stacking interaction of DNA duplex using atomic force microscope, Sci. Rep. 5 (2014) 9143.
- [17] L. Tora, T. H.Th.Marc, The TATA box regulates TATA-binding protein (TBP) dynamics in vivo, Trends. Biochem. Sci. 6 (2010) 309–314.
- [18] J. Duguid, V. Bloomfield, J. Benevides, G. Thomas Jr, Raman spectroscopy of DNA-metal complexes. I. Interactions and conformational effects of the divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd, Biophys. J. 65 (1993) 1916–1928.
- [19] make-na. ©2006, 2011, 2012, 2013. James Stroud.
- [20] E. Espinosa, E. Molins, C. Lecomte, Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities, Chem. Phys. Lett. 285 (1998) 170–173.
- [21] R. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, 1990.
- [22] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 46 (1964) B864.
- [23] W. Kohn, L. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133.
- [24] B. Delley, From molecules to solids with the DMol3 approach, J. Chem. Phys. 113 (2000) 7756.
- [25] Materials Studio, Accelrys, Biovia ©2014 Systèmes, Biovia Software Inc.
- [26] Y. Zhao, D.G. Truhlar, A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions, J. Chem. Phys. 125 (2006) 194101.
- [27] F.B. van Duijneveldt, J.G.C.M. van Duijneveldt-van de Rijdt, J.H. van Lenthe, State of the art in counterpoise theory, Chem. Rev. 94 (1994) 1873–1885.
- [28] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini,

- F. Egidí, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 09, Revision E.01, Gaussian, Inc., Wallingford CT, 2016.
- [29] T. Lu, F. Chen, Multiwfn: a multifunctional wavefunction analyzer, *J. Comp. Chem.* 33 (2012) 580–592.
- [30] C.R. Calladine, H.R. Drew, B.F. Luisi, A.A. Travers, *Understanding DNA: The Molecule and How It Works*, Elsevier Academic Press, 2004.
- [31] P. Yurenko, J. Novotný, R. Sklenář, R. Marek, Exploring non-covalent interactions in guanine and xanthine-based model DNA quadruplex structures: a comprehensive quantum chemical approach, *Phys. Chem. Phys.* 16 (2014) 2072–2084.
- [32] N.C. Seeman, J.M. Rosenberg, F.L. Suddath, J.J.P. Kim, A. Rich, RNA double-helical fragments at atomic resolution: I. The crystal and molecular structure of sodium adenylyl-3',5'-uridine hexahydrate, *J. Mol. Bio.* 104 (1976) 109–144.
- [33] C.A. Hunter, J.K.M. Sanders, The nature of π - π interactions, *J. Am. Chem. Soc.* 112 (1990) 5525–5534.
- [34] S. Grimme, Do special noncovalent π - π stacking interactions really exist? *Angew. Chem. Int. Ed.* 47 (2008) 3430–3434.
- [35] E.R. Collantes, W.J. Dun III, Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues, *J. Med. Chem.* 38 (1995) 2705–2713.
- [36] C.F. Matta, R.F.W. Bader, Atoms-in-molecules study of the genetically encoded amino acids. III. Bond and atomic properties and their correlations with experiment including mutation-induced changes in protein stability and genetic coding, *Proteins: Struct., Funct., Genet.* 52 (2003) 360–399.
- [37] O. Khakshoor, S.E. Wheeler, K.N. Houk, E.T. Kool, Measurement and theory of hydrogen bonding contribution to isosteric DNA base pairs, *J. Am. Chem. Soc.* 134 (2012) 3154–3163.
- [38] P. Desjardins, D. Conklin, NanoDrop microvolume quantitation of nucleic acids, *J. Vis. Exp.* 45 (2010) 2565.
- [39] T. Boland, B.D. Ratner, Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy, *Proc. Natl. Acad. Sci. U S A* 92 (1995) 5297–5301.
- [40] S. Smith, Y. Cui, C. Bustamante, Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules, *Science* 271 (1996) 795–799.
- [41] S.B. Bustamante, C. Smith, J. Liphardt, D. Smith, Single-molecule studies of DNA mechanics, *Curr. Opin. Struct. Biol.* 10 (2000) 279–285.



Do weak interactions affect the biological behavior of DNA? A DFT study of CpG island-like chains

Jorge Gutiérrez-Flores¹ · Enrique Hernández-Lemus² · Fernando Cortés-Guzmán³ · Estrella Ramos¹

Received: 6 April 2020 / Accepted: 3 August 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The origin, stability, and contribution to the formation of noncovalent interactions, such as hydrogen bonds and $\pi - \pi$ stacking, have been already widely discussed. However, there are few discussions about the relevance of these weak interactions in DNA performance. In this work, we seek to shed light on the effect of hydrogen bonds and $\pi - \pi$ stacking interactions on the biological behavior of DNA through the description of these intermolecular forces in CpG island-like (GC-rich) chains. Furthermore, we made some comparisons with TATA box-like (TA-rich) chains in order to describe hydrogen bond and $\pi - \pi$ stacking interactions as a function of the DNA sequence. For hydrogen bonds, we found that there is not a significant effect related to the number of base pairs. Whereas for $\pi - \pi$ stacking interactions, the energy tended to decrease as the number of base pairs increased. We observed anticooperative effects for both hydrogen bonds and $\pi - \pi$ stacking interactions. These results are in contrast with those of TATA box-like chains since cooperative and additive effects were found for both hydrogen bonds and $\pi - \pi$ stacking, respectively. Based on the chemical hardness and density of states, we can conclude that proteins may interact easier with GC-rich chains. We conclude that regardless of the chain length, a protein could interact more easily with these genomics regions because the $\pi - \pi$ stacking energies did not increase as a function of the number of base pairs, making, for the first time, a first approximation of the influence of noncovalent interaction on DNA behavior. We did all this work by means of DFT framework included in the DM \circ 1³ code (M06-L/DNP).

Keywords DNA · DFT · Hydrogen bonds · $\pi - \pi$ stacking · Promoter regions · Cooperativity

Introduction

The deoxyribonucleic acid (DNA) is the hereditary material in all organisms [1]. We can think that DNA is the most relevant biomolecule since it contains all the information about a life-form. Like a memory storage device in

computers, DNA in cells is a system in which the data is stored and can be read by different cells' mechanisms as proteins, among other many examples.

Similar to how we classify and organize our information in directories, in DNA, the info is organized into genes, making it easier to read and interpret the content. A gene, according to the ENCODE project, is a union of genomic sequences encoding a coherent set of potentially overlapping functional products [2]. At one edge of the genes, there is a sequence known as the promoter region, which is identified by various transcription factors that allow the formation of the protein complex necessary to carry out the transcription process. TATA box (thymine (T)-adenine (A) rich) and CpG island (cytosine (C)-guanine (G) rich) sequences are common in diverse promoter regions [3]. However, the frequency of the CpG islands as promoters is higher than that of the TATA boxes since it is believed that approximately 70% of gene promoters, for vertebrate genomes, are CpG islands [4].

As in the computers' architecture, the structure of the DNA is of vital importance to determine if cells' proteins

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00894-020-04501-6>) contains supplementary material, which is available to authorized users.

✉ Estrella Ramos
estrella@materiales.unam.mx

¹ Instituto de Investigaciones en Materiales, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, 04510, CDMX, México

² Computational Genomics, National Institute of Genomic Medicine, Mexico City, Mexico

³ Instituto de Química, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, 04510, CDMX, México

can read the information written on it. Under normal conditions, DNA adopts a B-helix structure [5], which is stabilized by a different kind and number of noncovalent interactions (NCIs).

NCIs are ubiquitous in chemical, biological, and material systems [6]. They have reached significant attention in recent years since they are associated with several physical and chemical phenomena. Of particular interest are those interactions present in biological systems, with a broad set of roles ranging from structural parameters [7–11] to the determination of biochemical and biological processes [12–17]. In DNA, it is commonly assumed that hydrogen bonds and $\pi - \pi$ stacking are the essential weak interactions alongside the base pairs [18].

Different approximations have been made in order to describe hydrogen bond and $\pi - \pi$ stacking interactions in DNA. Nevertheless, many of them only consider a small number of base pairs [19–29]. Moreover, the relationship between hydrogen bond and $\pi - \pi$ stacking of DNA and RNA bases have been studied; nonetheless, such studies usually only take into account nitrogenous bases [25] or base pair dimers (without the effect of the backbone) [26, 30]. Therefore, the effect of the presence of a greater number of base pairs on these interactions is not yet well understood and the overview of the biological impact of these interactions is limited.

In previous work, our group determined the impact of the chain length (the presence of additional sequence base pairs) over hydrogen bonds and $\pi - \pi$ stacking within TATA box-like chains. There we found that these interactions were enhanced when the number of base pairs in the DNA chain increased. Nevertheless, nothing was discussed about chains mainly composed of G and C, and the relevance of these interactions in the biological response of DNA.

In the present work, we aim to inquire if weak interactions affect the biological behavior of DNA. For this, we make a description of the behavior of hydrogen bond and $\pi - \pi$ stacking interactions in CpG island-like chains. Furthermore, as CpG islands are found in gene-promoting areas in a more significant proportion than TATA boxes [3], we will make some comparisons between CpG island-like chains' results with those of TATA box-like chains in order to make a first approximation on the biological differences between CpG islands and TATA boxes. Additionally, to complete our analysis, we seek to explore whether these features can be described with some reactivity indices, such as chemical hardness (η). Previous reports have found a good correlation between polarizability (deformability of the electron density cloud), hardness, and the tendency to form weak interactions [31–34]; therefore, the hardness index will be useful to describe the differences between these genomic regions.

Computational details

To carry out this work, we considered CpG island-like chains (5'-...GCGC...-3'), i.e., the modeled DNA duplexes were based on the sequence of the CpG islands (a promoter region found in several vertebrate genes, which consists of repeated units of guanine and cytosine [35]). DNA duplexes of different sizes were studied in order to determine the effect of the number of base pairs (n) on the hydrogen bond and $\pi - \pi$ stacking interactions. Therefore, chains with $n = 1$ to $n = 14$ were examined. The charge associated with the phosphate groups were neutralized by the addition of a hydrogen atom to each phosphate group, assuming that these would not have a substantial effect over the noncovalent interactions of interest of this work [24, 36, 37].

Ideal and optimized structures were analyzed. The ideal duplexes were constructed employing the open-source software *make-na* [38], which considers different parameters based on experimental data for the construction of symmetrical B-DNA chains. These chains were optimized to improve the description of hydrogen and $\pi - \pi$ stacking interactions since the weak interactions are not optimal in this configuration. It is important to remember that geometry optimization leads to a minimum that depends strongly on the starting structure. So, if the modeled chains were constructed based on experimental reports, we can expect that the description realized in this work will be analogous to what happens in a real DNA chain.

The mean energy of hydrogen bonds (\bar{D}_{HB}) was estimated from the value of the average hydrogen bond formation energy (\bar{E}_{HB}), which was calculated from the total energy of the chain with n -base pairs (E_n) and the total energy of each of the strands ($E_{n_{1s}}$ and $E_{n_{2s}}$) that make up the corresponding DNA duplex (see Eq. 1 and Fig. 1).

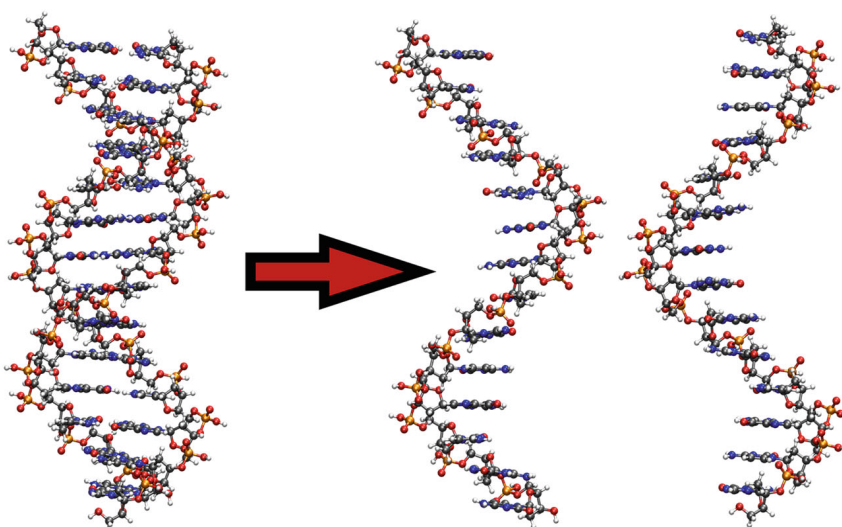
$$\bar{D}_{HB} = -\bar{E}_{HB} = -\frac{E_n - (E_{n_{1s}} + E_{n_{2s}})}{n} \quad (1)$$

Similar to the hydrogen bond mean energy, the $\pi - \pi$ stacking mean energy (\bar{D}_π) was calculated from the formation energy of these interactions (\bar{E}_π). For the latter, the total energy of the chain (E_{n_π}) without phosphate groups neither sugars (adding hydrogen atoms in its place), the total energy of each base pair ($E_{i_{bp}}$) and the number of spaces between base pairs (N_π) were considered (see Eq. 2 and Fig. 2).

$$\bar{D}_\pi = -\bar{E}_\pi = -\frac{E_{n_\pi} - \sum_{i=1}^n E_{i_{bp}}}{N_\pi} \quad (2)$$

The total hydrogen bond ($D_{HB}(n)$, Eq. 3) and $\pi - \pi$ stacking ($D_\pi(n)$, Eq. 4) interaction energies were estimated

Fig. 1 Schematization of the energy calculation of the hydrogen bond mean energy (\bar{D}_{HB}). GC chain with $n = 14$ is used as an example



by multiplying Eqs. 1 and 2 by n and N_π , respectively.

$$D_{HB} = \bar{D}_{HB}(n) \cdot n \quad (3)$$

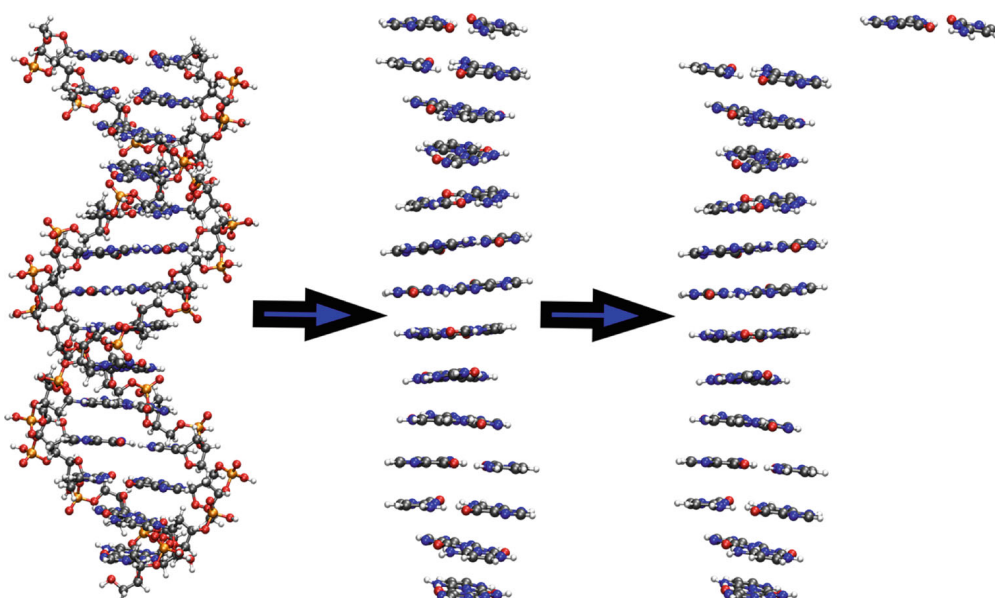
$$D_\pi = \bar{D}_\pi(n) \cdot N_\pi \quad (4)$$

The charge separation index (CSI) was employed in order to measure the local polarity of the hydrogen bond regions [39], since the greater the CSI value, the stronger the hydrogen bond [36]. This parameter was obtained from the sum of the absolute values of charges q of each i atom of the region of concern. Charges were taken from NBO population analysis.

$$CSI = \sum_i |q_i| \quad (5)$$

All calculations were done by means of the density functional theory (DFT) [40, 41] framework. DMol³ code [42], implemented in the modeling and simulation software Materials Studio 2017 [43], was used. Geometry optimizations, single-point energies, and molecular properties were obtained at the M06-L [44]/DNP level of theory, except for the NBO population analysis which was done in the Gaussian 2016 code [45] with the same functional but with a 6-31G(d) basis set [46–55]. We selected M06-L since it was parameterized to describe adequately noncovalent interactions, such as hydrogen bonds and $\pi - \pi$ stacking [44]. Even though there are more precise methodologies for the estimation and description of hydrogen bond [27] and $\pi - \pi$ stacking [28] interactions, we used the meta-GGA-M06-L functional to have acceptable predictions at an adequate computational cost. Furthermore, this

Fig. 2 Schematization of the energy calculation of the $\pi - \pi$ stacking mean energy (\bar{D}_π). GC chain with $n = 14$ is used as an example



chemical model was used and validated in a previously published work to describe these interactions in TATA box-like chains. According to Goerigk et al. [56], M06-L functional is one of the best meta-GGA to describe intermolecular forces. In our experience, DMol³ (in Materials Studio 2017) is the only algorithm capable of optimizing large molecules such as the studied DNA chains, although the code does not have dispersion corrections for the M06-L functional. Basis set superposition error (BSSE) corrections were taken into account through the counterpoise method [57]. The admitted convergence criteria were 1×10^{-5} Ha, 2×10^{-5} Ha, 4×10^{-3} Ha/Å, and 5×10^{-3} Å for SCF cycles, energy, gradient, and displacement, respectively.

Results and analysis

We start this section with the analysis of hydrogen bonds and $\pi - \pi$ stacking interactions in the studied CpG island-like chains. Then, we examine the charge distributions and the cooperative effects in order to complete the study of the noncovalent interactions of interest in this work. Later, the biological behavior of DNA will be discussed. For these purposes, comparisons will commonly be made with the results belonging to TATA box-like chains reported in reference [36].

Hydrogen bonds

A hydrogen bond, commonly, is referred to as a hydrogen atom shared between two different atoms. Generally, these correspond to electronegative atoms (N, O, and F) [58]. In a more sophisticated way, this interaction can be understood as a fractional chemical bond of coordinative $A - H \cdots B$ Lewis acid–base type, associated with the partial intermolecular $A - H \cdots : B \leftrightarrow A : \cdots H - B$ resonance commonly originating in the $n_B \rightarrow \sigma_{AH}^*$ donor–acceptor interaction between the lone pair n_B of the Lewis base and the hydride antibond σ_{AH}^* of the Lewis acid [59]. In the DNA, these interactions are present between the nitrogenous bases that belong to a base pair.

The hydrogen bond mean energies (\bar{D}_{HB}) were estimated by means of Eq. 1. One consequence of using Eq. 1 for this purpose is that the presence of other interactions when the strands are separated is neglected, as displayed in Fig. 1. However, as was shown by using the Espinosa-Molins-Lecomte (EML) approximation [60], Eq. 1 gives us reasonable magnitudes for the hydrogen bond strength [36].

The \bar{D}_{HB} values for the studied chains of different length are shown in Fig. 3 (see Table S1 at Supporting Information (SI) to see \bar{D}_{HB} values). From this figure, we can observe that the hydrogen bond mean energies of the optimized chains are slightly higher than those of ideal

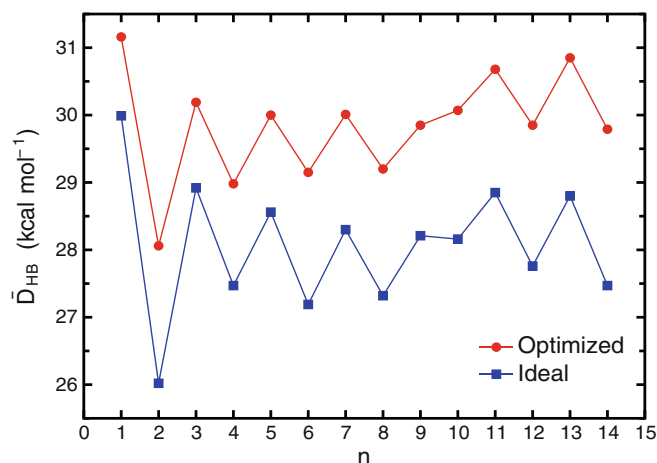
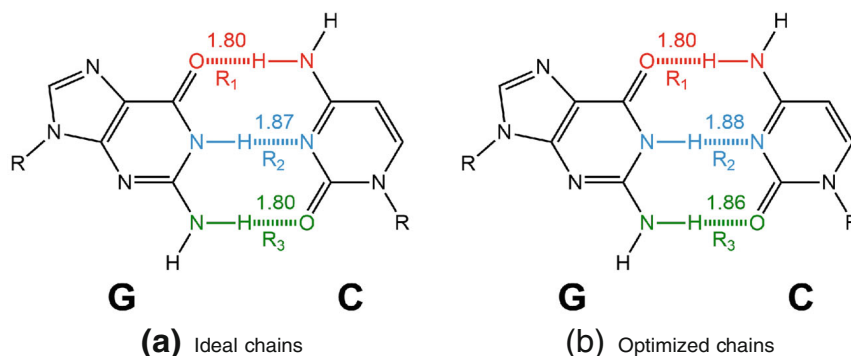


Fig. 3 Hydrogen bond mean energy (\bar{D}_{HB}) as a function of the number of base pairs. Blue squares represent the results for ideal chains, while red circles those for optimized chains. Lines connect the points for clarity. \bar{D}_{HB} of the optimized chains are slightly higher than those of ideal chains. Furthermore, \bar{D}_{HB} seems to not increase as the number of base pairs augment. The oscillations are associated with border effects. See Table S1 at Supporting Information to see \bar{D}_{HB} values

chains. As expected, at the moment of carrying out the geometry optimization process, the interactions throughout the chain are favored. There is not a substantial increase in \bar{D}_{HB} since, if we consider the existence of three hydrogen bonds between the nitrogenous bases, the structure is rigid; consequently, the structural changes in the optimized chains are not meaningful. The increase of ~ 1 kcal mol⁻¹ can be associated with small contamination of $\pi - \pi$ stacking, due to the way we estimated it (see Fig. 1 and Eq. 1). Additionally, from Fig. 3, it can be perceived that there is not a significant increment in \bar{D}_{HB} as a function of the number of base pairs (n). The observed oscillations, at odd and even number of n , are correlated with border effects: as we worked with duplexes of a finite number of base pairs, in the CpG island-like chains with an odd n , the 3' and 5' ends will be different (for example, 3GC's sequence is 3'-GCG-5', while for 4GC is 3'-GCGC-5'), impacting D_{HB} .

To get a better insight into the behavior of hydrogen bonds within the CpG island-like duplexes, the global average distances belonging to these interactions were measured for both ideal and optimized chains. Values are shown in Fig. 4 (see Table S4 at SI for individual average distances). For this analysis, it is assumed that the smaller the distance between the atoms involved in the hydrogen bond, the higher the strength of this interaction [61]. For both ideal and optimized cases, the central hydrogen bond R_2 is the weakest, whereas the strongest are those formed between the guanine's oxygens and cytosine's nitrogens. The difference in distances and strength associated with hydrogen bonds may be related to the electronegativities

Fig. 4 Global average of hydrogen bond distances for both ideal and optimized chains. Assuming that the smaller the distance between atoms forming hydrogen bond, the stronger the interaction; R_2 is the weakest hydrogen bond, while R_1 is the strongest hydrogen bond. See Table S2 at SI for individual average distances. Distance units are in Å



of the involved atoms: the greater the electronegativity difference, the stronger the hydrogen bond and the smaller the correlated length. Distances are in concordance with the reported trend from experimental values [62], where the shorter interaction space is related to $O-H\cdots N$ interactions. Furthermore, our predictions are in agreement with other studies [19]. It should be noted that even for the energy analysis, there are no significant differences between the distances of the ideal and optimized chains since given the structural rigidity of the nitrogenous bases (of each base pair) associated with the presence of three hydrogen bonds, the structural parameters to these interactions do not undergo any relevant change. It is important to mention that, on average, the hydrogen bonds' angles did not undergo significant changes when optimizing the chains (see Table S3 at SI).

When comparing GpG island-like chains' \bar{D}_{HB} and distances with those obtained for the TATA box-like chains ($\bar{D}_{HB} < 18 \text{ kcal mol}^{-1}$, $\bar{R}_1 = 1.87 \text{ Å}$, $\bar{R}_2 = 1.82 \text{ Å}$; no reported value for $C-H\cdots O$ (R_3) interaction) [36], it can be inferred that the hydrogen bonds in GC-rich duplexes are stronger than those of the TA-rich chains, which is related to the existence of a third hydrogen bond between the G-C bases.

$\pi - \pi$ stacking interactions

The $\pi - \pi$ stacking interactions are, regularly, referred to as the interaction between aromatic π clouds [63, 64]. In DNA, these kinds of intermolecular forces are present between the base pairs, and, due to the nitrogenous bases are substituted multiring aromatic compounds, they prefer to be in a parallel displaced geometry [64]. The $\pi - \pi$ stacking mean energy (\bar{D}_π) was calculated through Eq. 2, and results are shown in Fig. 5 for both ideal and optimized cases (see Table S4 at SI to see \bar{D}_π values). When the duplexes are optimized, \bar{D}_π tends to increase, which is a consequence of the planarity of the base pairs due to the existence of three hydrogen bonds that restrains the propeller twist of the nitrogenous bases that conform the base pair. Although

there is not a significant effect associated with the number of base pairs, the $\pi - \pi$ stacking energy seems to decrease as n rises, which means that there are no cooperative effects due to the increment of the stacking interactions throughout the chains. Besides, it can be observed that at $n = 10$, the $\pi - \pi$ stacking energy tends to increase. According to Fonseca et al. [26], two stacked Watson-Crick base pairs have a minimum when there is a twist angle of 36° between them. Therefore, this increment can be related to the 360° full turn of the DNA chain, due to the C_{10} symmetry axis associated with B-DNA.

Importantly, TATA box-like chains have weaker interactions ($\bar{D}_\pi < 12 \text{ kcal mol}^{-1}$ [36]) than CpG island-like chains. Between the TA base pairs, it is assumed that there are only two hydrogen bonds; therefore, the propeller twist between the A and T nitrogenous bases can change more loosely. Consequently, $\pi - \pi$ stacking interactions are not as favored as in GC pairs.

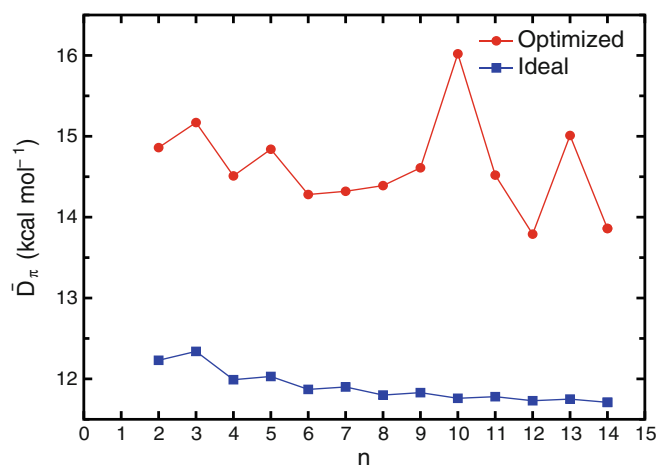


Fig. 5 $\pi - \pi$ stacking mean energy (\bar{D}_π) as a function of the number of base pairs. Blue squares represent the results for ideal chains, while red circles those for optimized chains. Lines connect the points for clarity. $\pi - \pi$ stacking interactions are stronger for optimized chains than for ideal chains. Additionally, \bar{D}_π seems to decrease as n rises. See Table S4 at SI to see \bar{D}_π values

Charge distribution analysis

To make a more in-depth analysis of the hydrogen bond and $\pi - \pi$ stacking interactions, in this section, an analysis of the charge distribution along the studied DNA duplexes is made. For this, at first, the charge separation index (CSI) is used, which gives us an idea of the local polarity [39]. This index was calculated via Eq. 5, and it was only estimated for chains with $n = 1, 3, 5,$ and 7 since chains with an odd number of base pairs have a central base pair and, therefore, the border effects associated with finite DNA chains can be examined.

Figure 6 shows the global average values of CSI for the hydrogen bond regions of the central base pairs, for chains with $n = 1, 3, 5,$ and 7 , of both ideal and optimized structures (see Table S5 at SI to see all CSI values). In concordance to the previous analysis, there are no significant differences between the ideal and optimized chains since CSI values are almost the same. Furthermore, as in the analysis of the distances, N – H...O hydrogen bonds (R_1 and R_3) are stronger than N – H...N hydrogen bonds (R_2). Besides, R_3 's CSI tends to decrease when the chains are optimized. This may be related to the electron-withdrawing inductive effects in the R_3 neighborhood.

It is important to mention that the most external base pairs (those found at the edges) have the highest CSI values (see Table S5 at Supporting Information). This may be due to the border effects caused when modeling finite DNA chains. These effects are almost negligible for the central base pairs, given that they have practically the same CSI values.

Figures 7 and 8 show planes of electrostatic potential in the areas where the hydrogen bonding and $\pi - \pi$ stacking interactions are located, respectively. These analyses correspond to the central base pairs of the 7GC duplexes ($n = 7$). For hydrogen bonds, it can be seen that there are no significant differences between the ideal and the optimized chains. While for the $\pi - \pi$ stacking, the charge (red zone) is more concentrated between the base pairs of the optimized chain. This charge concentration occurs

between the nitrogenous bases of the opposite strands, which can be related to a displaced parallel interaction between the nitrogenous bases of different base pairs, which is believed to be one of the most favorable configurations for the stabilization of this interaction [64].

Cooperativity and non-cooperativity effects

Although it is commonly supposed that there is an additive effect as the number of weak interactions increases in a system, it has been demonstrated that the effect due to the increase of weak interactions is actually cooperative (or anticooperative) [8, 65], since “1+1 is more than 2” [65] (or less, if it is anticooperative). It has been shown indeed that for both TATA box-like chains [36] and CpG island-like chains, an effect of the number of base pairs on hydrogen bonds and $\pi - \pi$ stacking exists. Now, based on the estimation of these interactions, in this section, we propose a way to determine if the presence of other base pairs has a cooperative or anticooperative effect.

The hydrogen bond and $\pi - \pi$ stacking additive energies were calculated from the smallest chains to compare them with the total interaction energies, in order to analyze the existence of cooperative or anticooperative effects. To estimate the additive energy of hydrogen bonds (D_{HB}^{add}), the hydrogen bond energy (D_{HB} , see Eq. 3) of the chains with $n = 1$ was considered as the unit, and then it was multiplied by n (Eq. 6).

$$D_{HB}^{add} = D_{HB}(n = 1) \cdot n \quad (6)$$

For the additive energy associated with the $\pi - \pi$ stacking interactions (D_{π}^{add}), the $\pi - \pi$ stacking energy (D_{π} , see Eq. 4) of the chains with $n = 2$ was thought as the unit, then it was multiplied by the number of spaces between base pairs (Eq. 7).

$$D_{\pi}^{add} = D_{\pi}(n = 2) \cdot N_{\pi} \quad (7)$$

Comparisons between the additive energies and total interactions energies, for both hydrogen bond and $\pi - \pi$ stacking, are shown in Figs. 9 and 10, respectively. In the

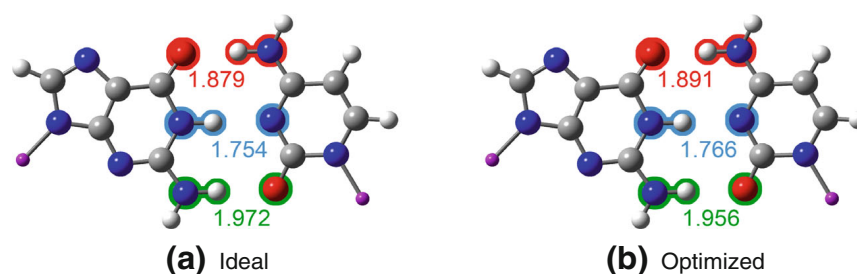


Fig. 6 Global average of the charge separation index (CSI) for hydrogen bond regions between G and C ($X - H \cdots Y$, where $X, Y = N, O$). Only values of the central base pairs (of the chains with $n = 1, 3, 5,$

and 7) were considered for the average (see Table S5 at SI to see all CSI values). Not significant between the ideal and optimized chains are present, which is in concordance with the energy analysis

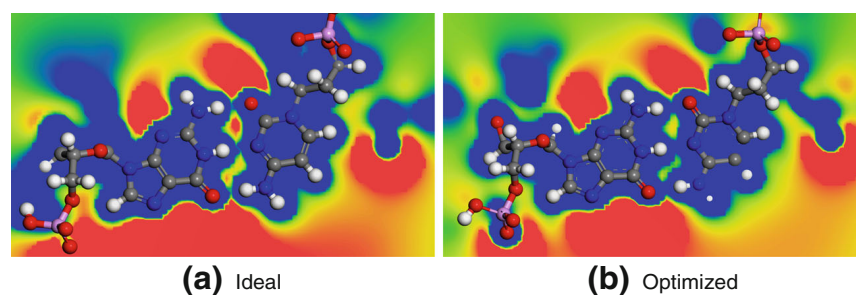


Fig. 7 Electrostatic potential planes for the central base pair of the ideal (a) and optimized (b) chains with $n = 7$ (it is assumed that no significant border effects are present in these chain). View corresponding to the hydrogen bond interaction between the nitrogenous bases of

the central base pair. Scale: from -0.04 a.u. (red) to 0.04 a.u. (blue). There are no relevant changes between the ideal and optimized chain, which is consistent with the previous analysis

case of hydrogen bonds (Fig. 9), for TATA box-like chains (\bar{D}_{HB} values were obtained from reference [36]), there is a cooperative effect between hydrogen bond interactions ($D_{HB} > D_{HB}^{add}$), whereas for CpG island-like chains, an anticooperative effect is observed ($D_{HB} < D_{HB}^{add}$). The difference between TATA box-like chains and CpG island-like chains could be to the “rigidity” of the base pairs, since, due to the third hydrogen bond, in the CpG island-like chains, the movement of the bases, to favor interactions, is restricted, having an anticooperative behavior for these interactions.

As distinct from hydrogen bonds, $\pi - \pi$ stacking interactions (Fig. 10) have different behavior for TATA box-like (\bar{D}_{π} values were obtained from reference [36]) since neither cooperativity nor anticooperativity exists between these interactions ($D_{\pi} \simeq D_{\pi}^{add}$). Instead, for CpG island-like chains, the same pattern as for hydrogen bonds was observed (in general, $D_{\pi} < D_{\pi}^{add}$). Again, the more restrained the movement of the base, the less the cooperativity between interactions.

All this analysis is in agreement with the energy analysis of hydrogen bonds and $\pi - \pi$ stacking for both TATA box-like chains and CpG island-like chains. In the case of hydrogen bonds, TA-rich chains tended to increase

\bar{D}_{HB} as the number of base pairs (interactions) increased (cooperative effect) [36], contrary to what is seen for GC-rich chains, in which the increasing number of base pairs did not provoke a large increase on \bar{D}_{HB} (anticooperative effect). As for the $\pi - \pi$ stacking interactions, in TATA box-like chains, n did not have a significant change on \bar{D}_{π} (additive). While for CpG island-like chains, \bar{D}_{π} tends to decrease as the number of base pairs increases (anticooperative).

TATA box-like vs CpG island-like chains: biological relevance

In the previous sections, we have presented a systematic analysis of the weak interactions in CpG island-like chains and made some comparisons with the results obtained previously for TATA box-like chains. However, the central question is not yet answered at all: Do weak interactions affect the biological behavior of DNA? For this purpose, it is necessary to remember that both the TATA box and CpG islands act as promoter regions [66, 67]. Nevertheless, the frequency of the CpG islands as promoters is higher than that of the TATA boxes [4]. Furthermore, it has been established that these promoter regions often lack common

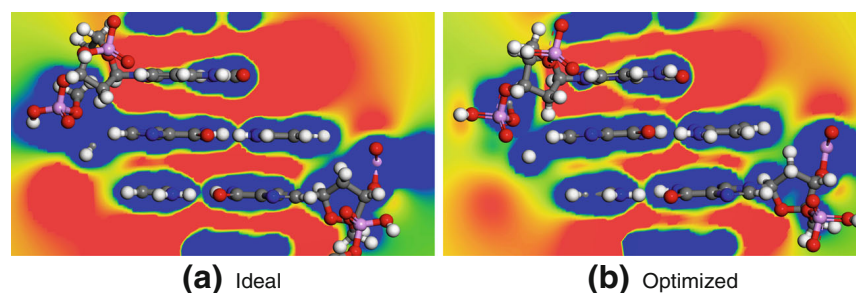


Fig. 8 Electrostatic potential planes for the central base pairs of the ideal (a) and optimized (b) chains with $n = 7$ (it is assumed that no significant border effects are present in these chains). View corresponding to the $\pi - \pi$ stacking interaction between the base pairs of the

duplex. Scale: from -0.04 a.u. (red) to 0.04 a.u. (blue). For the optimized chain, the charge is more concentrated between the base pairs. Also, the base pairs conformation seem to be displaced, which is one of the most propitious conformation for $\pi - \pi$ stacking interactions

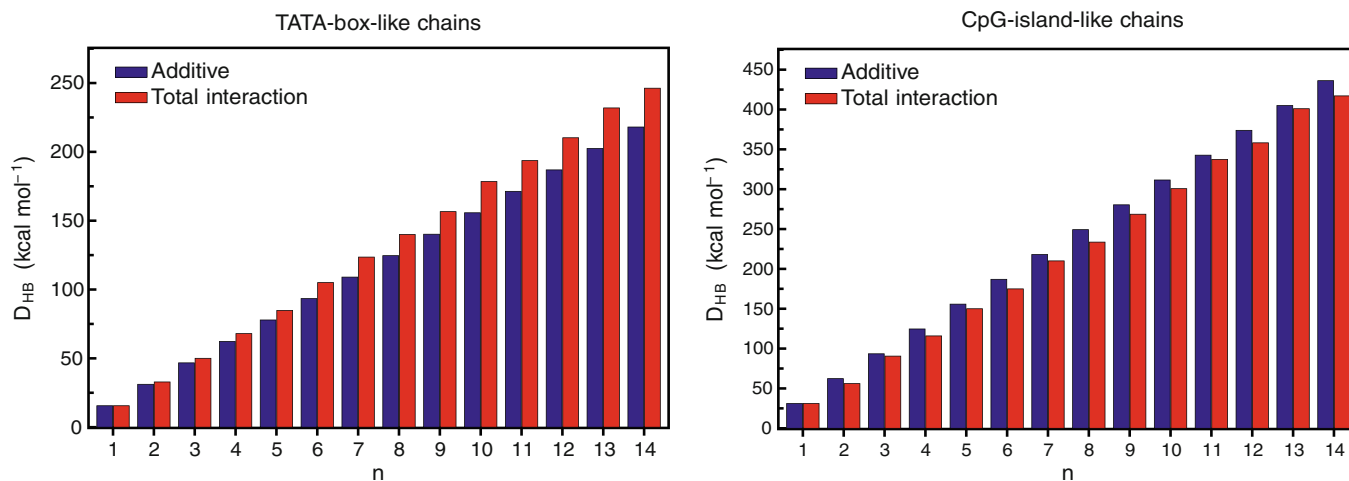


Fig. 9 Comparison between hydrogen bond additive energies D_{HB}^{add} (blue bars) and total hydrogen bond energies $D_{HB}(n)$ (red bars) for both TATA box-like [66] (left) and CpG island-like chains (right). For

TATA box-like chains, a cooperative effect between hydrogen bonds ($D_{HB} > D_{HB}^{add}$) exists, while for CpG island-like chains, this effect is anticooperative ($D_{HB} < D_{HB}^{add}$)

core promoter elements [68, 69] and the probability of transcription factor binding is more considerable due to its GC richness [4].

Aside from the purely theoretical relevance, the present analysis on the interplay between $\pi - \pi$ stacking interactions and hydrogen bonding in CpG complexes gains importance given the fact that such interactions are able to modify the chemical, mechanical, and electrostatic properties of DNA [25, 30, 70]. Such influence also extends to chemical modifications of DNA, such as cytosine methylation [30, 71], which are of foremost importance for gene regulatory process [72, 73], in particular for the epigenetic control of gene expression via chemical and structural modifications [74–76]. The balance

between stacking interactions, hydrogen bonds, and charge distribution is also related to mechanisms of DNA damage as, for example, via intercalation [70, 77].

To make a connection between the analysis of weak interactions and the reactivity (possibility to interact with a protein) of a specific DNA sequence, we turn to the polarizability, which could be understood as the tendency of a molecule to modify its electron density [78], as a consequence of the induced instantaneous dipole moments due to an external electric field [79]. Therefore, the higher the polarizability of a system, the bigger the tendency to form weak interactions [31–33]. Given that a good correlation between polarizability and chemical hardness has been demonstrated [34], we believe that this index can

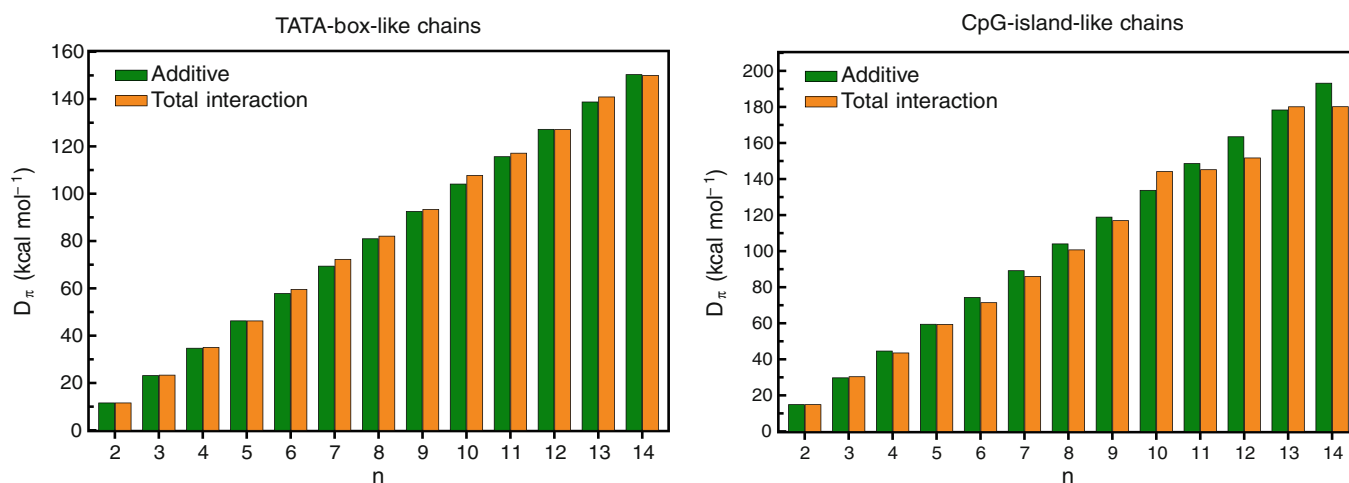
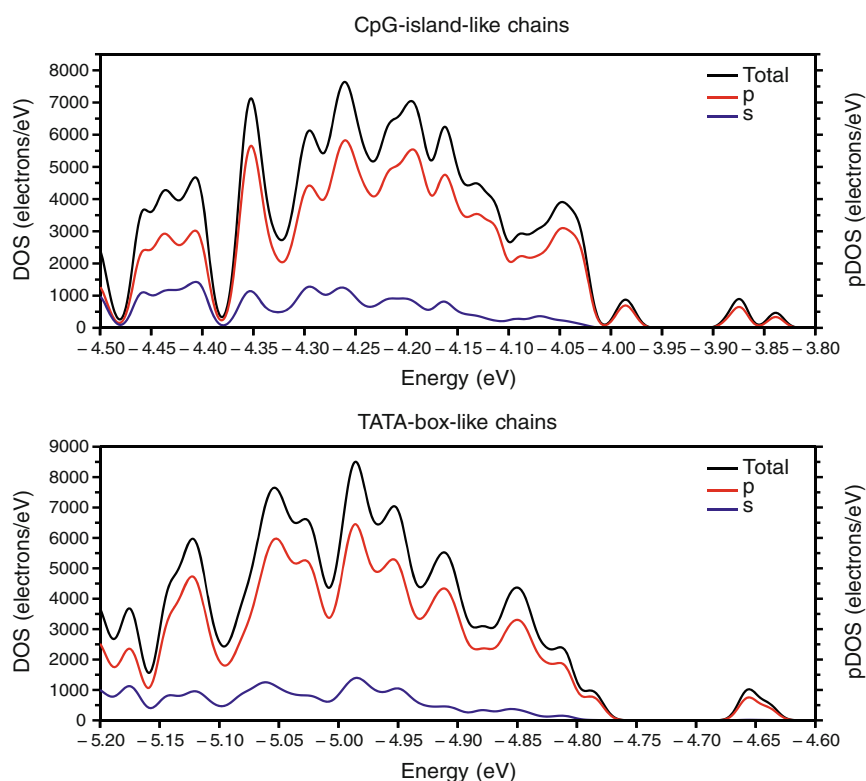


Fig. 10 Comparison between $\pi - \pi$ stacking additive energies D_{π}^{add} (green bars) and total $\pi - \pi$ energies $D_{\pi}(n)$ (orange bars) for both TATA box-like [66] (left) and CpG island-like chains (right). For TATA box-like chains there is neither cooperative nor anticooperative

effect between $\pi - \pi$ stacking interactions ($D_{\pi} \simeq D_{\pi}^{add}$), while for CpG island-like chains, in general, this effect is anticooperative (in most cases $D_{\pi} < D_{\pi}^{add}$)

Fig. 11 Total (DOS) and partial (pDOS) density of states for 8TA (down) and 8GC (up) chains. Fermi levels (the last occupied level): $E_F = -4.780$ eV and $E_F = -3.978$ eV for 8TA and 8GC, respectively. Last occupied levels are mainly conformed by p-orbitals. Furthermore, 8TA's last occupied states are energetically outer than those from 8TA. Hence, it is expected that other molecules, like proteins, will interact first with GC-rich chains



be useful to describe the differences between the TATA box and the CpG islands. Therefore, chemical hardness (η) [80] is calculated by means of the vertical ionization energies (IE) and electron affinities (EA) (see Eq. 8), for 8TA [36], (since eight base pairs is the standard length of the TATA box canonical sequence), and for the sake of comparison for 8GC chains [81, 82].

$$\eta = IE - EA \quad (8)$$

Table 1 shows the values of η for both 8TA and 8GC chains. From this, it can be concluded that 8GC is softer than 8TA. If a charge distribution is less polarizable, its response to an external disturbance (an interaction with a protein, for example) is expected to be lower than for a more polarizable system. Hence, as 8TA is harder than 8GC, it can be suggested that in the transcription process, it would be more accessible to the transcriptional complex to interact with a GC-rich region than with a TA-rich region, a result in compliance with known transcriptional kinetics.

It is necessary to notice that the chemical hardness is in agreement with the energy analysis since GpC island-like chains have stronger hydrogen bonds and $\pi - \pi$ stacking interactions, but they do not increase as a function of the chain length since there are no cooperative effects for the interactions in the CpG island-like chains. Probably, this is the reason why CpG islands do not have a defined size [83].

In both experimental [84, 85] and theoretical [36] works, it has been concluded that, from a biological viewpoint,

$\pi - \pi$ stacking interactions are more relevant than hydrogen bonds. Therefore, we can assume that, for instance, when a protein gets close to a DNA chain, the first change suffered by the electron density is going to be at the level of π cloud electron density, which is associated with $\pi - \pi$ stacking interactions. Therefore, an analysis of the total (DOS) and partial (pDOS) density of states was performed to characterize the electrons in both 8TA and 8GC chains and to determine in which system the π electrons are energetically more available to interact with another system (a protein, for example).

In Fig. 11, DOS and pDOS for both 8TA chain and 8GC chain are presented. As can be observed, the last occupied levels correspond, mainly, to p-orbitals, which, as mentioned before, are assumed to belong to π electron density clouds involved in $\pi - \pi$ stacking interactions. As foreseen, the 8GC's p-orbitals are energetically outer than those belonging to 8TA. Consequently, if a protein is

Table 1 Chemical hardness (η) for 8TA and 8GC chains

Chain	η (eV)
8TA	4.79
8GC	4.21

8TA is harder than 8TA; therefore, 8GC's electronic density is more polarizable, having better properties that will allow a more favorable interaction with a protein, for example

placed against two DNA sequences, one rich in thymine and adenine and another rich in guanine and cytosine, it can be expected that it preferentially binds the GC chain, since its electronic density can be more easily deformable (it is less hard than TA chains) and its respective densities associated with π clouds become energetically more available, leading to the formation of more efficient weak interactions that allow the interplay between the protein and the DNA chain.

Finally, it is important to mention that there is still much debate as to what are the actual contributions of each interaction type on the structure, reactivity, and conformational properties of CpG (and general DNA) complexes. What is not questioned, however, is the relevance such phenomena may have on the biological processes—mostly gene expression regulation—involving CpG-rich promoter regions. To name but a few, experimental studies have found that the binding activity of important transcription factors VEGF and SP1 are highly dependent on the energetic profiles at the CpG enriched promoter regions of their targets [72], and also that there are huge differences in the binding energies of general transcription factors in methylated CpG and non-CpG (i.e., CC, CA, or CT base pairs) promoter sites and that such differences are key to fine-tune gene regulatory processes [73]. It is not adventurous to hypothesize that at least some of these differences are due to the different balances of hydrogen bonding and $\pi - \pi$ stacking interactions.

Conclusions

DNA is one of the most important biomolecules since it contains all the genetic information about an organism, and, undoubtedly, noncovalent interactions (as hydrogen bonds and $\pi - \pi$ stacking) contribute significantly to the stabilization of the duplex. Despite the vast number of theoretical work done to describe these interactions better, few discussions focus on examining the effect of hydrogen bonds and $\pi - \pi$ stacking on the biological behavior of DNA. Therefore, in this work, we attempted to solve this inquiry through the description of these intermolecular interactions in CpG island-like (GC-rich) chains.

To accomplish our purpose, first, we studied the relation between the hydrogen bond and $\pi - \pi$ stacking mean energies. For hydrogen bonds, we did not find significant effects due to the number of base pairs, whereas for $\pi - \pi$ stacking interactions, we detected that the energy tended to decrease as the number of base pairs increased. These behaviors were also observed at the charge distribution analysis. To get a better overview of hydrogen bonds and $\pi - \pi$ stacking interactions as a function of the DNA sequence, we made some comparisons between CpG island-like chains' results and those of TATA box-like

(TA-rich) chains reported at reference [36]. In general, the studied interactions are stronger in GC-rich duplexes.

Furthermore, in order to analyze the existence of cooperative or anticooperative effects on hydrogen bond and $\pi - \pi$ stacking interactions, we estimated hydrogen bond and $\pi - \pi$ stacking additive energies and compared it with the total interaction energies. For both hydrogen bonds and $\pi - \pi$ stacking for CpG island-like chains, an anticooperative effect was observed (due to the rigidity of the bases), whereas a cooperative influence for hydrogen bonds and an additive effect for $\pi - \pi$ stacking interactions were observed for TATA box-like chains.

Finally, in order to get a better insight into the biological differences between the TATA box and CpG island promoters, we estimated the chemical hardness for both 8TA and 8GC chains. We found that 8GC is softer (more polarizable) than 8TA. To relate these reactivity differences with the electronic structure of the chains, an analysis of the total (DOS) and partial (pDOS) density of states was made. We discovered that the last unoccupied states of 8GC were energetically outer than those from 8TA. Based on hardness and DOS, we can expect that a protein, for example, could interact easier with GC-rich chains since it could be more favorable to form weak interactions with this kind of chains. Besides, this asseveration is in coherence with the energetic analysis, since cooperative effects are absent in GC-rich chains. Therefore, regardless of the chain length, a protein could interact more easily with these genomics regions because the $\pi - \pi$ stacking energies would not increase.

In summary, we determined the effect of the number of base pairs on both hydrogen bond and $\pi - \pi$ stacking interactions and analyzed these interactions as a function of the DNA sequence. Furthermore, for the first time, we made a first approximation of the influence of noncovalent interactions in the behavior of DNA. Therefore, answering the central question of this work: Do weak interactions affect the biological behavior of DNA?, we conclude that, indeed, hydrogen bond and $\pi - \pi$ stacking interactions have an impact on the DNA behavior, and we demonstrated it by the analysis of two different promoter regions.

There is, of course, much to be fathomed in terms of molecular interactions such as NCIs and its relationship to DNA structure and reactivity and how this influences gene expression regulation. We believe, however, that rooting future studies on the solid tenets of theoretical frameworks (such as DFT) may result quite enlightening to understand these fundamental biological phenomena better.

Acknowledgments The authors thank the computational support of Alberto López and Alejandro Pompa from IIM-UNAM and to DGTIC-UNAM (LANCAD-UNAM-DGTIC-194 and LANCAD-UNAM-DGTIC-225 projects) for their supercomputing services. This work was supported by the Consejo Nacional de Ciencia y

Tecnología SEP-CONACYT-2016-285544 and FRONTERAS-2017-2115, the National Institute of Genomic Medicine, México and by PAPIIT-UNAM (BG100320). Additional support has been granted by the Laboratorio Nacional de Ciencias de la Complejidad, from the Universidad Nacional Autónoma de México. EHL is the recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences. JGC acknowledges the CONACyT's support through the 336100 scholarship grant.

References

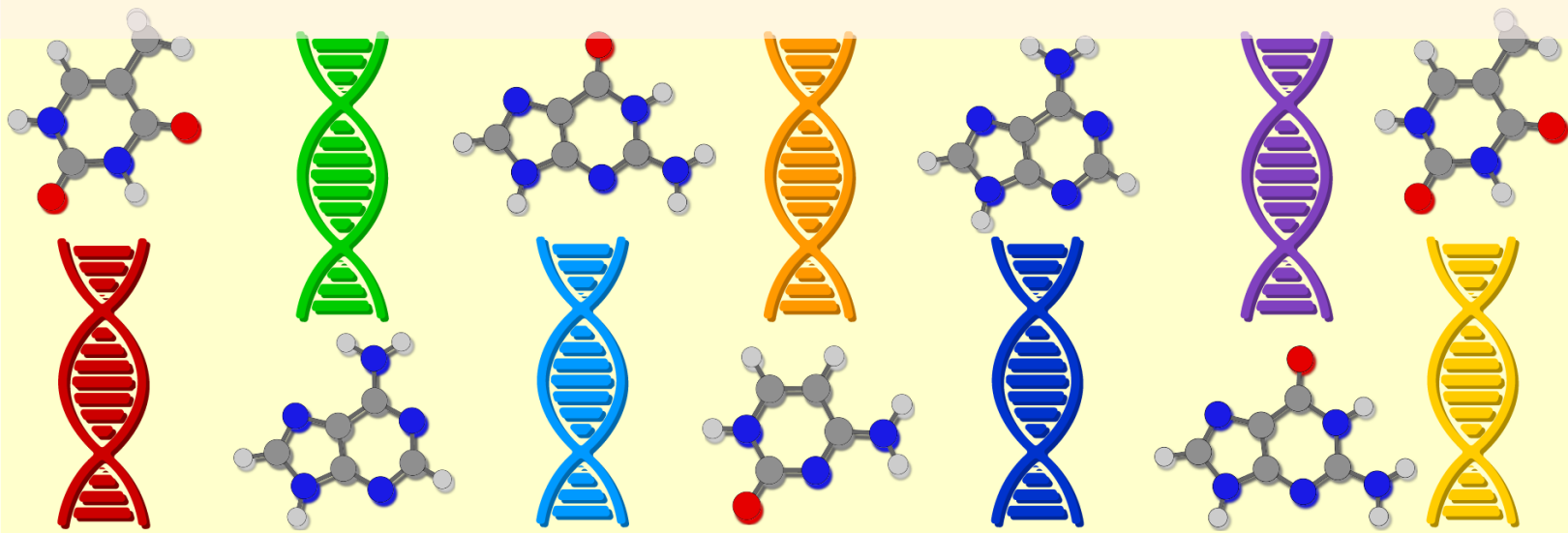
1. U.S. National Library of Medicine (2020) What is DNA? Retrieved March, 2020, <https://ghr.nlm.nih.gov/primer/basics/dna>
2. Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbelt JO, Emanuelsson O, Zhang ZD, Weissman S, Snyder M (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res* 17(6):669–681
3. Haberle V, Stark A (2018) Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* 19(10):621–637
4. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev* 25:1010–1022
5. Goodsell D (2001) DNA. Retrieved March, 2020, <http://www.rcsb.org/pdb/101/motm.do?momID=23>
6. Hermann J, DiStasio RA Jr, Tkatchenko A (2017) First-principles models for van der waals interactions in molecules and materials: concepts, theory, and applications. *Chem Rev* 117:4714–4758
7. Zhou P, Huang J, Tian F (2012) Specific noncovalent interactions at protein-ligand interface: implications for rational drug design. *Curr Med Chem* 19(2):226–238
8. Mahadevi AS, Sastry GN (2016) Cooperativity in noncovalent interactions. *Chem Rev* 116:2775–2825
9. Mundlapati VR, Sahoo DK, Bhaumik S, Jena S, Chandrakar A, Biswal HS (2018) Noncovalent carbon-bonding interactions in proteins. *Angew Chem Int Ed* 57(50):16496–16500
10. Hirota S, Lin Y-W (2018) Design of artificial metalloproteins/metalloenzymes by tuning noncovalent interactions. *J Biol Inorg Chem* 23(1):7–25
11. Fingerhut BP, Elsaesser T (2019) Coherent multidimensional spectroscopy, volume 226 of Springer Series in Optical Sciences. Springer, Singapore, pp 171–195
12. Christensen DE, Brzovic PS, Kleivit RE (2007) E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol* 14(10):941–948
13. Aran M, Ferrero DS, Pagano E, Wolosiuk RA (2009) Typical 2-Cys peroxiredoxins - Modulation by covalent transformations and noncovalent interactions. *FEBS J* 276(9):2478–2493
14. Riley KE, Hobza P (2011) Noncovalent interactions in biochemistry. *WIREs Comput Mol Sci* 1:3–17
15. Wright JD, Mace PD, Day CL (2016) Noncovalent ubiquitin interactions regulate the catalytic activity of ubiquitin writers. *Trends Biochem Sci* 41(11):924–937
16. Chi C-B, Tang Y, Zhang J, Dai Y-N, Abdalla M, Chen Y, Zhou C-Z (2018) Structural and biochemical insights into the multiple functions of yeast Grx3. *J Mol Biol* 430(8):1235–1248
17. Andrade H, Lin W, Zhang X (2019) Specificity from nonspecific interaction: regulation of tumor necrosis factor- α activity by DNA. *J Biol Chem* 294(16):6397–6404
18. Kool ET (2001) Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu Rev Biophys Biomol Struct* 30:1–22
19. Fonseca C, Matthias F, Snijders J, Jan E (2000) Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. *J Am Chem Soc* 122:4117–4128
20. Szatyłowicz H, Sadlej-Sosnowska N (2010) Characterizing the strength of individual hydrogen bonds in DNA Base pairs. *J Chem Inf Model* 50:2151–2161
21. Asensio A, Kobko N, Dannenberg JJ (2003) Cooperative hydrogen-bonding in adenine-thymine and guanine cytosine base pairs. Density functional theory and møller-pleiset molecular orbital study. *J Phys Chem A* 107:6441–6443
22. Grunenberg J (2004) Direct assessment of interresidue forces in watson-crick base pairs using theoretical compliance constants. *J Am Chem Soc* 126:16310–16311
23. Dong H, Hua W, Li S (2007) Estimation on the individual Hydrogen-Bond Strength in Molecules with Multiple Hydrogen Bonds. *J Phys Chem A* 111:2941–2945
24. Matta C, Castillo N, Boyd RJ (2006) Extended weak bonding interactions in DNA: π -stacking (base-base), base-backbone, and backbone-backbone interactions. *J Phys Chem B* 110:563–578
25. Mignon P, Loverix S, Steyaert J, Geerlings P (2005) Influence of the π - π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res* 2005:17791789
26. Swart M, van der Wijst T, Fonseca Guerra C, Bickelhaupt FM (2007) π - π stacking tackled with density functional theory. *J Mol Model* 13:1245–1257
27. Jurečka P, Šponer J, Černý J, Hobza P (2006) Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys Chem Chem Phys* 8(17):1985–1993
28. Kruse H, Banáš P, Šponer J (2019) Investigations of stacked DNA base-pair steps: highly accurate stacking interaction energies, energy decomposition, and many-body stacking effects. *J Chem Theory Comput* 15(1):95–115
29. Halder A, Data D, Seelam PP, Bhattacharyya D, Mitra A (2019) Estimating strengths of individual hydrogen bonds in RNA base pairs: toward a consensus between different computational approaches. *ACS Omega* 4(4):7354–7368
30. Acosta-Silva C, Branchadell V, Bertran J, Oliva A (2010) Mutual relationship between stacking and hydrogen bonding in dna. Theoretical study of guanine-cytosine, guanine-5-methylcytosine, and their dimers. *J Phys Chem B* 114:10217–10227
31. Hwang J, Dial BE, Li P, Kozik ME, Smith MD, Shimizu KD (2015) How important are dispersion interactions to the strength of aromatic stacking interactions in solution? *Chem Sci* 6:4358–4364
32. Sinnokrot MO, Valeev EF, Sherrill CD (2002) Estimates of the Ab Initio limit for π - π interactions: the benzene dimer. *J Am Chem Soc* 124:10887–10893
33. Zeinalipour-Yazdi CD, Pullman DP (2006) Correlation of polarizabilities with Van Der Waals interactions in π -systems. *J Chem Phys* 110:24260–24265
34. Gupta K, Ghanty TK, Ghosh SK (2012) Polarizability, ionization potential, and softness of water and methanol clusters: an interrelation. *J Phys Chem A* 116:6831–6836
35. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
36. Gutiérrez-Flores J, Ramos E, Mendoza CI, Hernández-Lemus E (2018) Electronic properties of DNA: description of weak interactions in TATA-box-like chains. *Biophys Chem* 233:26–35
37. Duguid J, Bloomfield VA, Benevides J, Thomas Jr. GJ (1993) Raman spectroscopy of DNA-metal complexes. I. Interactions and conformational effects of the divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. *Biophys J* 65:1916–1928

38. make-na. © 2006, 2011, 2012, 2013. James Stroud
39. Collantes ER, Dun WJ III (1995) Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues. *J Med Chem* 38:2705–2713
40. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 46:B864
41. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:A1133
42. Delley B (2000) From molecules to solids with the DMol3 approach. *J Chem Phys* 113:7756
43. Materials Studio, Accelrys, Biovia © 2014 Systèmes, Biovia Software Inc
44. Zhao Y, Truhlar DG (2006) A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J Chem Phys* 125:194101
45. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) Gaussian¹⁶ Revision C.01. Gaussian Inc. Wallingford CT
46. Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods. IX. An extended gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys* 54(2):720–723
47. Hehre WJ, Ditchfield K, Pople JA (1972) Self-consistent molecular orbital methods. XII. Further extensions of gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J Chem Phys* 56(5):2257–2261
48. Hariharan PC, Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theor Chim Acta* 28(3):213–222
49. Hariharan PC, Pople JA (1974) Accuracy of AHn equilibrium geometries by single determinant molecular orbital theory. *Mol Phys* 27(1):209–214
50. Gordon MS (1980) The isomers of silacyclopropane. *Chem Phys Lett* 76(1):163–168
51. Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ, Pople JA (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J Chem Phys* 77(7):3654–3665
52. Binning RC, Curtiss LA (1990) Compact contracted basis sets for third-row atoms: Ga–Kr. *J Comput Chem* 11(10):1206–1216
53. Blaudeau J-P, McGrath MP, Curtiss LA, Radom L (1997) Extension of Gaussian-2 (G2) theory to molecules containing third-row atoms K and Ca. *J Chem Phys* 107(13):5016–5021
54. Rassolov VA, Pople JA, Ratner MA, Windus TL (1998) 6-31G* basis set for atoms K through Zn. *J Chem Phys* 109(4):1223–1229
55. Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA (2001) 6-31G* basis set for third-row atoms. *J Comput Chem* 22(9):976–984
56. Goerigk L, Hansen A, Bauer C, Ehrlich S, Najibi A, Grimme S (2017) A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys Chem Chem Phys* 19(48):32184–32215
57. van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van Lenthe JH (1994) State of the Art in Counterpoise Theory. *Chem Rev* 94:1873–1885
58. Lewis GN (1923) Valence and the structure of atoms and molecules. The Chemical Catalog Co., New York
59. Weinhold F, Klein RA (2012) What is a hydrogen bond? Mutually consistent theoretical and experimental criteria for characterizing H-bonding interactions. *Mol Phys* 110(9-10):565–579
60. Espinosa E, Molins E, Lecomte C (1998) Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chem Phys Lett* 285:170–173
61. Herschlag D, Pinney MM (2018) Hydrogen bonds: simple after all? *Biochemistry* 57:3338–3352
62. Rosenberg JM, Seeman NC, Day RO, Rich A (1976) RNA double-helical fragments at atomic resolution: II. The crystal structure of sodium guanylyl-3',5'-cytidine nonahydrate. *J Mol Biol* 104:145–167
63. Martinez CR, Iverson BL (2012) Rethinking the term “pi-stacking”. *Chem Sci* 3:2191–2201
64. Thakuria R, Nath NK, Saha BK (2019) The nature and applications of π - π interactions: a perspective. *Cryst Growth Des* 19(2):523–528
65. Prins LJ, Reinhoudt DN, Timmerman P (2001) Noncovalent synthesis using hydrogen bonding. *Angew Chem Int Ed* 40:2382–2426
66. Tora L, H.Th.Marc T (2010) The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends Biochem Sci* 6:309–314
67. Antequera F (2003) Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* 60:1647–1658
68. Juven-Gershon T, Hsu J-Y, Theisen JWM, Kadonga JM (2008) The RNA polymerase II core promoter — the gateway to transcription. *Curr Opin Cell Biol* 20:253–259
69. Zhang L, Yu H, Wang P, Ding Q, Wang Z (2013) Screening of transcription factors with transcriptional initiation activity. *Gene* 531:64–70
70. Hihath J, Guo S, Zhang P, Tao N (2012) Effects of cytosine methylation on DNA charge transport. *J. Phys. Condens. Matter* 24(16):164204
71. Song Q, Qiu Z, Wang H, Xia Y, Shen J, Zhang Y (2013) The effect of methylation on the hydrogen-bonding and stacking interaction of nucleic acid bases. *Struct Chem* 24(1):55–65
72. Tsukakoshi K, Saito S, Yoshida W, Goto S, Ikebukuro K (2018) CpG methylation changes g-quadruplex structures derived from gene promoters and interaction with VEGF and SP1. *Molecules* 23(4):944
73. Jin J, Lian T, Gu C, Yu K, Gao YQ, Su XD (2016) The effects of cytosine methylation on general transcription factors. *Sci Rep* 6:29119
74. Teng X, Hwang W (2018) Effect of methylation on local mechanics and hydration structure of DNA. *Biophys J* 114(8):1791–1803
75. Norberg J, Vihinen M (2001) Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides. *J Mol Struct* 546(1-3):51–62
76. Guza R, Kotandeniya D, Murphy K, Dissanayake T, Lin C, Giambasu GM, Lad RR, Wojciechowski F, Amin S, Sturla SJ, et al. (2011) Influence of c-5 substituted cytosine and related nucleoside analogs on the formation of benzo [a] pyrene diol epoxide-dg adducts at cg base pairs of DNA. *Nucleic Acids Res* 39(9):3988–4006
77. Gil A, Branchadell V, Calhorda MJ (2016) A theoretical study of methylation and CH/ π interactions in DNA intercalation:

- methylated 1, 10-phenanthroline in adenine–thymine base pairs. *RSC Adv* 6(89):85891–85902
78. Clark T (2017) Polarization, donor–acceptor interactions, and covalent contributions in weak interactions: a clarification. *J Mol Model* 23:297
 79. Le Fèvre RJW (1965) *Advances in physical organic chemistry*, vol 3. Academic Press, London, pp 1–90
 80. Parr RG, Pearson RG (1983) Absolute hardness: companion parameter to absolute electronegativity. *J Am Chem Soc* 105(26):7512–7516
 81. Juo ZS, Chiu TK, Leiberman PM, Baikalov I, Berk AJ, Dickerson RE (1996) How proteins recognize the TATA box. *J Mol Biol* 261(2):239–254
 82. Bucher P (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J Mol Biol* 212(4):563–578
 83. Elango N, Yi SV (2011) Functional relevance of CpG island length for regulation of gene expression. *Genetics* 187(4):1077–1083
 84. Smirnov S (2002) Integrity of duplex structures without hydrogen bonding: DNA with pyrene paired at abasic sites. *Nucleic Acids Res* 30(24):5561–5569
 85. Feng B, Sosa RP, Mårtensson AKF, Jiang K, Tong A, Dorfman KD, Takahashi M, Lincoln P, Bustamante CJ, Westerlund F, Nordén B (2019) Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. *Proc Natl Acad Sci USA* 116(35):17169–17174

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Bibliografía



- (1) ADN (Ácido Desoxirribonucleico) | NHGRI <https://www.genome.gov/es/genetics-glossary/ADN-acido-Desoxirribonucleico> (visitado 13-08-2020).
- (2) The History of DNA Timeline | DNA Worldwide <https://www.dna-worldwide.com/resource/160/history-dna-timeline> (visitado 18-08-2020).
- (3) Darwin, C. y Darwin, C., *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life* /; John Murray, London : 1859, págs. 1-564.
- (4) Abbott, S. y Fairbanks, D. J. *Genetics* **2016**, 204, 407 LP -422.
- (5) Pray, L. A. *Nature Education* **2008**, 1, 100.
- (6) Sir Archibald Edward Garrod :: DNA from the Beginning <http://www.dnaftb.org/13/bio.html> (visitado 18-08-2020).
- (7) Garrod, A. E. *Molecular Medicine* **1996**, 2, 274-82.
- (8) Levene, P. A. *Journal of Biological Chemistry* **1919**, 40, 415-424.
- (9) Avery, O. T.; Macleod, C. M. y McCarty, M. *Journal of Experimental Medicine* **1944**, 79, 137-158.
- (10) Vigue, C. L. *The American Biology Teacher* **1984**, 46, 207-211.
- (11) Chargaff, E. *Experientia* **1950**, 6, 201-209.
- (12) Chargaff, E. *Science* **1971**, 172, 637-642.
- (13) Scott, M. Rocky Road: Rosalind Franklin <http://www.strangescience.net/rfranklin.htm> (visitado 03-09-2020).
- (14) Watson, J. D. y Crick, F. H. C. *Nature* **1953**, 171, 737-738.
- (15) All Nobel Prizes <https://www.nobelprize.org/prizes/lists/all-nobel-prizes/> (visitado 03-09-2020).
- (16) Goodsell, D. *RCSB Protein Data Bank* **2001**, DOI: 10.2210/rcsb_pdb/mom_2001_11.

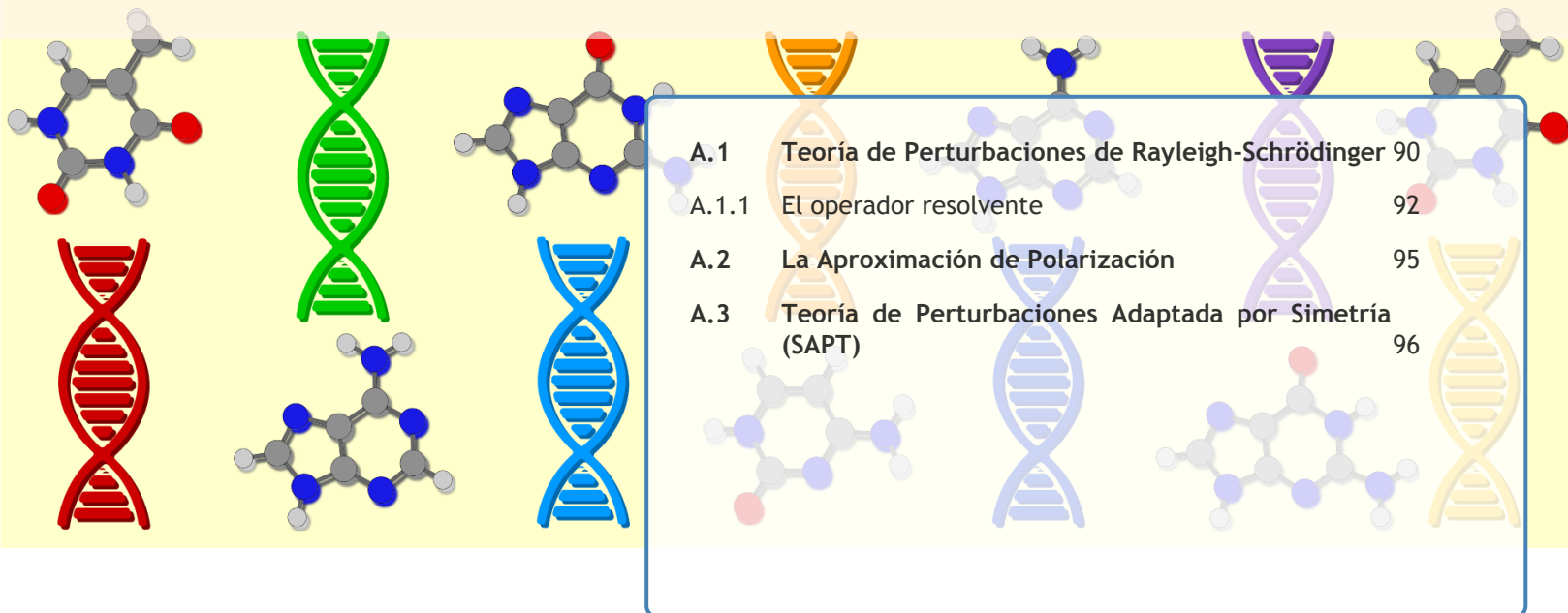
- (17) Calladine, C. R.; Drew, H. R.; Luisi, B. F. y Travers, A. A., *Understanding DNA: The Molecule and How it Works: Third Edition*; Elsevier Ltd: 2004, págs. 1-334.
- (18) Hassan, M. A. E. y Calladine, C. R. *J. Mol. Biol* **1996**, *259*, 95-103.
- (19) Kool, E. T. **2001**, *30*, 1-22.
- (20) The 99 Percent... of the Human Genome - Science in the News <http://sitn.hms.harvard.edu/flash/2012/issue127a/> (visitado 29-10-2020).
- (21) Non-Coding DNA <https://www.genome.gov/genetics-glossary/Non-Coding-DNA> (visitado 29-10-2020).
- (22) Gerstein, M. B.; Bruce, C.; Rozowsky, J. S.; Zheng, D.; Du, J.; Korbel, J. O.; Emanuelsson, O.; Zhang, Z. D.; Weissman, S. y Snyder, M. *Genome Research* **2007**, *17*, 669-681.
- (23) Gene <https://www.genome.gov/genetics-glossary/Gene> (visitado 29-10-2020).
- (24) University, C. S. P. Gene Structure <https://elearning.cpp.edu/learning-objects/gene-structure/> (visitado 04-11-2020).
- (25) Tora, L. y Timmers, H. *Trends in Biochemical Sciences* **2010**, *35*, 309-314.
- (26) Gardiner-Garden, M. y Frommer, M. *Journal of Molecular Biology* **1987**, *196*, 261-282.
- (27) Asensio, A.; Kobko, N. y Dannenberg, J. J. *The Journal of Physical Chemistry A* **2003**, *107*, 6441-6443.
- (28) Grunenberg, J. *Journal of the American Chemical Society* **2004**, *126*, 16310-16311.
- (29) Dong, H.; Hua, W. y Li, S. *The Journal of Physical Chemistry A* **2007**, *111*, 2941-2945.
- (30) Espinosa, E.; Molins, E. y Lecomte, C. *Chemical Physics Letters* **1998**, *285*, 170-173.
- (31) Szatyłowicz, H. y Sadlej-Sosnowska, N. *Journal of Chemical Information and Modeling* **2010**, *50*, 2151-2161.
- (32) Swart, M.; van der Wijst, T.; Fonseca Guerra, C. y Bickelhaupt, F. M. *Journal of Molecular Modeling* **2007**, *13*, 1245-1257.
- (33) Acosta-Silva, C.; Branchadell, V.; Bertran, J. y Oliva, A. *The Journal of Physical Chemistry B* **2010**, *114*, 10217-10227.
- (34) Jurečka, P.; Šponer, J.; Černý, J. y Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985-1993.
- (35) Kruse, H.; Banáš, P. y Šponer, J. *Journal of Chemical Theory and Computation* **2019**, *15*, 95-115.
- (36) DNA Sequencing <https://www.genome.gov/genetics-glossary/DNA-Sequencing> (visitado 09-11-2020).
- (37) Kaplan, I. G., *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*; John Wiley & Sons Ltd: Engalnd, 2006, pág. 389.
- (38) Stone, A., *The Theory of Intermolecular Forces*, Second Edi; Oxford University Press: Oxford, 2013, pág. 338.
- (39) Atkins, P. y De Paula, J., *Physical Chemistry*, Ninth Edit; W.H. Freeman y Company: New York, 2010, pág. 1053.
- (40) Weinhold, F. y Klein, R. A. *Molecular Physics* **2012**, *110*, 565-579.

- (41) Weinhold, F. y Klein, R. A. *Angewandte Chemie International Edition* **2014**, *53*, 11214-11217.
- (42) Van der Lubbe, S. C. C. y Fonseca Guerra, C. *Chemistry – An Asian Journal* **2019**, *asia.201900717*.
- (43) Lewis, G., *Valence and the structure of atoms and molecules*; The Chemical Catalog Co.: New York, 1923, pág. 172.
- (44) Pauling, L., *Nature of Chemical Bond*, Tercera Ed; Cornell University Press: Nueva York, 1960, pág. 449.
- (45) Sinnokrot, M. O. y Sherrill, C. D. *The Journal of Physical Chemistry A* **2006**, *110*, 10656-10668.
- (46) Thakuria, R.; Nath, N. K. y Saha, B. K. *Crystal Growth & Design* **2019**, *19*, 523-528.
- (47) Molčanov, K. y Kojić-Prodić, B. *IUCrJ* **2019**, *6*, 156-166.
- (48) Hunter, C. A. y Sanders, J. K. M. *Journal of the American Chemical Society* **1990**, *112*, 5525-5534.
- (49) Cao, M.; Fu, A.; Wang, Z.; Liu, J.; Kong, N.; Zong, X.; Liu, H. y Gooding, J. J. *The Journal of Physical Chemistry C* **2014**, *118*, 2650-2659.
- (50) Zhao, R. y Zhang, R.-Q. *Physical Chemistry Chemical Physics* **2016**, *18*, 25452-25457.
- (51) Mahadevi, A. S. y Sastry, G. N. *Chemical Reviews* **2016**, *116*, 2775-2825.
- (52) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D. y Darnell, J., *Molecular Cell Biology*, Cuarta Edi; W.H. Freeman y Company: Nueva York, 2000, pág. 1184.
- (53) Smirnov, S. *Nucleic Acids Research* **2002**, *30*, 5561-5569.
- (54) Zhang, T.-b.; Zhang, C.-l.; Dong, Z.-l. y Guan, Y.-f. *Scientific Reports* **2015**, *5*, 9143.
- (55) Antequera, F. *Cellular and Molecular Life Sciences (CMLS)* **2003**, *60*, 1647-1658.
- (56) Deaton, A. M. y Bird, A. *Genes & Development* **2011**, *25*, 1010-1022.
- (57) Juven-Gershon, T.; Hsu, J.-Y.; Theisen, J. W. y Kadonaga, J. T. *Current Opinion in Cell Biology* **2008**, *20*, 253-259.
- (58) Zhang, L.; Yu, H.; Wang, P.; Ding, Q. y Wang, Z. *Gene* **2013**, *531*, 64-70.
- (59) Duguid, J.; Bloomfield, V.; Benevides, J. y Thomas, G. *Biophysical Journal* **1993**, *65*, 1916-1928.
- (60) Matta, C. F.; Castillo, N. y Boyd, R. J. *The Journal of Physical Chemistry B* **2006**, *110*, 563-578.
- (61) Stroud, J. make-na, 2013.
- (62) Hohenberg, P. y Kohn, W. *Physical Review* **1964**, *136*, B864-B871.
- (63) Kohn, W. y Sham, L. J. *Physical Review* **1965**, *140*, A1133-A1138.
- (64) Delley, B. *J. Chem. Phys.* **2000**, *113*, 7756.
- (65) Materials Studio, Accelrys, Biovia © 2014 Systèmes, Biovia Software Inc.
- (66) Zhao, Y. y Truhlar, D. G. *The Journal of Chemical Physics* **2006**, *125*, 194101.
- (67) Frisch, M. J. y col. Gaussian~16 Revision C.01, Gaussian Inc. Wallingford CT, 2016.
- (68) Ditchfield, R.; Hehre, W. y Pople, J. J. *J. Chem. Phys.* **1971**, *54*, 720-723.

- (69) Hehre, W.; Ditchfield, K. y Pople, J. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
- (70) Hariharan, P. y Pople, J. *Theor. Chim. Acta* **1973**, *28*, 213-222.
- (71) Hariharan, P. y Pople, J. *Mol. Phys.* **1974**, *27*, 209-214.
- (72) Gordon, M. *Chem. Phys. Lett.* **1980**, *76*, 163-168.
- (73) Francl, M.; Pietro, W.; Hehre, W.; Binkley, J.; Gordon, M.; DeFrees, D. y Pople, J. *J. Chem. Phys.* **1982**, *77*, 3654-3665.
- (74) Binning, R. y Curtiss, L. *J. Comput. Chem.* **1990**, *11*, 1206-1216.
- (75) Blaudeau, J.-P.; McGrath, M.; Curtiss, L. y Radom, L. *J. Chem. Phys.* **1997**, *107*, 5016-5021.
- (76) Rassolov, V.; Pople, J.; Ratner, M. y Windus, T. *J. Chem. Phys.* **1998**, *109*, 1223-1229.
- (77) Rassolov, V.; Ratner, M.; Pople, J.; Redfern, P. y Curtiss, L. *J. Comput. Chem.* **2001**, *22*, 976-984.
- (78) Van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M. y van Lenthe, J. H. *Chemical Reviews* **1994**, *94*, 1873-1885.
- (79) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A. y Grimme, S. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184-32215.
- (80) Bader, R. F. W. *Accounts of Chemical Research* **1985**, *18*, 9-15.
- (81) Yanson, I. K.; Teplitsky, A. B. y Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149-1170.
- (82) Herschlag, D. y Pinney, M. M. *Biochemistry* **2018**, *57*, 3338-3352.
- (83) Seeman, N. C.; Rosenberg, J. M.; Suddath, F.; Kim, J. J. P. y Rich, A. *Journal of Molecular Biology* **1976**, *104*, 109-144.
- (84) Rosenberg, J. M.; Seeman, N. C.; Day, R. O. y Rich, A. *Journal of Molecular Biology* **1976**, *104*, 145-167.
- (85) Li, S.; Olson, W. K. y Lu, X.-J. *Nucleic Acids Research* **2019**, *47*, W26-W34.
- (86) Collantes, E. R. y Dunn, W. J. *Journal of Medicinal Chemistry* **1995**, *38*, 2705-2713.
- (87) Matta, C. F. y Bader, R. F. W. *Proteins: Structure, Function, and Bioinformatics* **2003**, *52*, 360-399.
- (88) Bucher, P. *Journal of Molecular Biology* **1990**, *212*, 563-578.
- (89) Juo, Z. S.; Chiu, T. K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J. y Dickerson, R. E. *Journal of Molecular Biology* **1996**, *261*, 239-254.
- (90) Hunter, C. A. *Journal of Molecular Biology* **1993**, *230*, 1025-1054.
- (91) Temiz, N. A.; Donohue, D. E.; Bacolla, A.; Luke, B. T. y Collins, J. R. *PLoS ONE* **2012**, *7*, ed. por Mayer, C., e35558.
- (92) Clark, T. *Journal of Molecular Modeling* **2017**, *23*, 297.
- (93) Le Fèvre, R. en *Advances in Physical Organic Chemistry*; Academic Press: London, 1965; vol. 3, págs. 1-90.
- (94) Sinnokrot, M. O.; Valeev, E. F. y Sherrill, C. D. *Journal of the American Chemical Society* **2002**, *124*, 10887-10893.
- (95) Zeinalipour-Yazdi, C. D. y Pullman, D. P. *The Journal of Physical Chemistry B* **2006**, *110*, 24260-24265.

- (96) Hwang, J.; Dial, B. E.; Li, P.; Kozik, M. E.; Smith, M. D. y Shimizu, K. D. *Chemical Science* **2015**, 6, 4358-4364.
- (97) Gupta, K.; Ghanty, T. K. y Ghosh, S. K. *The Journal of Physical Chemistry A* **2012**, 116, 6831-6836.
- (98) Parr, R. G. y Pearson, R. G. *Journal of the American Chemical Society* **1983**, 105, 7512-7516.
- (99) Elango, N. y Yi, S. V. *Genetics* **2011**, 187, 1077-1083.
- (100) Gutiérrez-Flores, J.; Ramos, E.; Mendoza, C. I. y Hernández-Lemus, E. *Biophysical Chemistry* **2018**, 233, 26-35.
- (101) Gutiérrez-Flores, J.; Hernández-Lemus, E.; Cortés-Guzmán, F. y Ramos, E. *Journal of Molecular Modeling* **2020**, 26, 266.

A. Teoría de Perturbaciones Adaptada por Simetría



A.1	Teoría de Perturbaciones de Rayleigh-Schrödinger	90
A.1.1	El operador resolvente	92
A.2	La Aproximación de Polarización	95
A.3	Teoría de Perturbaciones Adaptada por Simetría (SAPT)	96

“Cambios pequeños [...] [la] energía pueden conducir a alteraciones masivas en tu resultado final.”

Kevin Michel, 2013

En este apéndice se presentará, de forma básica, las ideas centrales de la Teoría de Perturbaciones Adaptada por Simetría (SAPT, por sus siglas en inglés), la cual permite una adecuada descripción de las interacciones no-covalentes, dado que su formulación da pie a encontrar diferentes términos energéticos que permiten obtener un panorama más amplio de las contribuciones energéticas involucradas en la formación de dichas interacciones débiles. La información de este apéndice fue desarrollada con base en la referencia [1]. Se recomienda consultar las referencias [2-4] para una discusión más detallada y profunda de la teoría SAPT.

A.1 Teoría de Perturbaciones de Rayleigh-Schrödinger ¹

La teoría de perturbaciones es el segundo método aproximado más utilizado en la química cuántica. Ésta permite estimar las contribuciones y cambios en los niveles energéticos, y modificaciones en las funciones de onda, que ocurren cuando un campo externo o un campo que surge cuando algún término previamente ignorado en el Hamiltoniano se aplica a especies cuyos estados no perturbados son conocidos [6].

Para incluir alguna perturbación causada por un campo, el primer paso es modificar el Hamiltoniano original (\mathcal{H}_0) de algún sistema, mediante la introducción de un efecto nuevo:

$$\mathcal{H} = \mathcal{H}_0 + \lambda\mathcal{H}' \quad (\text{A.1})$$

donde \mathcal{H}' corresponde al operador que representa a la perturbación externa más allá del potencial normal que experimenta el sistema, y λ a un parámetro de ordenamiento sin significado físico. Puesto que \mathcal{H}_0 es el Hamiltoniano original y, por lo tanto, no cuenta con ninguna perturbación, éste también es conocido como el Hamiltoniano de orden cero.

Ahora, se asume que existe un conjunto ortogonal completo de soluciones de \mathcal{H}_0 :

$$\mathcal{H}_0|n\rangle = E_n^0|n\rangle \quad (\text{A.2})$$

al cual se le asocia un estado de más baja energía (E_0^0) cuya función propia es $|0^0\rangle$. Este conjunto de funciones comprende a las soluciones no perturbadas, también conocidas como soluciones de orden cero. Cuando se agrega la perturbación \mathcal{H}' al Hamiltoniano original, el sistema se relajará produciendo, como resultado, una función de respuesta ($|0\rangle$) asociada al estado basal, la cual puede ser expandida como una serie de potencias del parámetro de ordenamiento:

$$|0\rangle = |0^0\rangle + \lambda|0^1\rangle + \lambda^2|0^2\rangle + \dots \quad (\text{A.3})$$

Adicionalmente, cabe destacar que la función de respuesta del estado basal está específicamente construida para ser intermedicamente normalizada a la función del estado basal de orden cero:

$$\langle 0^0|0\rangle = \langle 0|0^0\rangle = 1 \quad (\text{A.4})$$

Por lo que al multiplicar la Ecuación (A.3) por el ket $\langle 0^0|$, integrar y usar el criterio de normalización intermedia (Ecuación (A.4)), se concluye que:

$$\langle 0^0|0^1\rangle = \langle 0^0|0^2\rangle = \dots = 0 \quad (\text{A.5})$$

En cuanto a los términos energéticos, de forma similar a la función de respuesta, la energía perturbada del estado basal puede ser expandida de la siguiente forma:

$$E_0 = E_0^0 + \lambda E_0^1 + \lambda^2 E_0^2 + \dots \quad (\text{A.6})$$

Sustituyendo las Ecuaciones (A.1), (A.3) y (A.6) en la Ecuación (A.2), se tiene que:

$$\begin{aligned} (\mathcal{H}_0 + \lambda\mathcal{H}') [|0^0\rangle + \lambda|0^1\rangle + \lambda^2|0^2\rangle + \dots] = \\ (E_0^0 + \lambda E_0^1 + \lambda^2 E_0^2 + \dots) [|0^0\rangle + \lambda|0^1\rangle + \lambda^2|0^2\rangle + \dots] \end{aligned} \quad (\text{A.7})$$

¹El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por la referencia [5]

Al multiplicar por la izquierda a la Ecuación (A.7) con la función ket $\langle 0|$ (Ecuación (A.3)), integrar y ocupar el criterio de la normalización intermedia (Ecuación (A.4)), los términos sobrevivientes pueden ser agrupados en función del orden del parámetro λ . Por lo tanto, la ecuación de orden cero (independiente de λ) es:

$$E_0^0 = \langle 0^0 | \mathcal{H}_0 | 0^0 \rangle \quad (\text{A.8})$$

la cual es la energía del estado basal no perturbada. Además de los términos de la Ecuación (A.8), los términos de orden cero restantes se encuentran asociados a λ^1 , y estos, al agruparlos, corresponden a la energía del estado basal de primer orden de la teoría de perturbaciones:

$$E_0^1 = \langle 0^0 | \mathcal{H}' | 0^0 \rangle \quad (\text{A.9})$$

La Ecuación (A.9) muestra que la energía de primer orden corresponde al valor esperado de la función de onda del estado basal no perturbada, evaluada sobre el Hamiltoniano perturbante (\mathcal{H}').

De forma similar que en las Ecuaciones (A.8) y (A.9), al multiplicar por la Ecuación (A.7) con la función ket $\langle 0|$, integrar y ocupar el criterio de la normalización intermedia (Ecuación (A.4)), se pueden obtener los términos de energía de cualquier orden en función del factor λ . Por ejemplo, las expresiones de las energías de segundo y tercer orden (obtenidas del agrupamiento de los términos asociados a λ^2 y λ^3 , respectivamente) son:

$$E_0^2 = \langle 0^0 | \mathcal{H}' | 0^1 \rangle \quad (\text{A.10})$$

$$E_0^3 = \langle 0^0 | \mathcal{H}' | 0^2 \rangle \quad (\text{A.11})$$

Como se puede observar en las Ecuaciones (A.10) y (A.11), se requiere de una representación de las funciones de onda de respuesta de primer orden y de órdenes superiores para determinar la energía de órdenes superiores al orden 1. Estas funciones son expandidas a partir de los elementos complementarios del vector propio de orden cero. Por ejemplo, la función de respuesta de primer orden puede ser expresada como:

$$|0^1\rangle = \sum_{n \neq 0} c_n^1 |n\rangle \quad (\text{A.12})$$

Teniendo ecuaciones semejantes para las funciones de respuesta de órdenes superiores. La base $|n\rangle$ constituye un conjunto ortonormal; por lo tanto, los coeficientes de expansión c_n^1 están relacionados con la función de respuesta $|0^1\rangle$ de la siguiente manera:

$$\langle n | 0^1 \rangle = \sum_{k \neq 0} c_k^1 \langle n | k \rangle = \sum_{k \neq 0} \delta_{kn} c_k = c_n^1 \quad (\text{A.13})$$

Lo cual permite escribir:

$$|0^1\rangle = \sum_{n \neq 0} |n\rangle \langle n | 0^1 \rangle \quad (\text{A.14})$$

Mostrando que los coeficientes de expansión de la función respuesta son obtenidos del operador de proyección $|n\rangle \langle n|$.

Los coeficientes de expansión de un orden arbitrario k pueden ser determinados mediante la agrupación de todos los elementos con el mismo orden en el parámetro de ordenamiento λ^k (Ecuación (A.7)). De esta forma, la función de respuesta de primer orden es:

$$\mathcal{H}_0|0^1\rangle + \mathcal{H}'|0^0\rangle = E_0^0|0^1\rangle + E_0^1|0^0\rangle \quad (\text{A.15})$$

Multiplicando por la izquierda a la Ecuación (A.15) por un elemento arbitrario del complemento ortogonal $\langle n|$ e integrando, se tiene que:

$$E_n^0\langle n|0^1\rangle + \langle n|\mathcal{H}'|0^0\rangle = E_0^0\langle n|0^1\rangle \quad (\text{A.16})$$

Rearreglando la Ecuación (A.16) y sustituyendo en la Ecuación (A.14), se obtiene:

$$|0^1\rangle = \sum_{n \neq 0} \frac{|n\rangle\langle n|\mathcal{H}'|0^0\rangle}{E_0^0 - E_n^0} \quad (\text{A.17})$$

Al comparar las Ecuaciones (A.14) y (A.17), y utilizando lo establecido en la Ecuación (A.13), se concluye que los coeficientes de expansión de primer orden tienen la siguiente forma:

$$c_n^1 = \frac{\langle n|\mathcal{H}'|0^0\rangle}{E_0^0 - E_n^0} \quad (\text{A.18})$$

Tomando en cuenta todo lo mencionado anteriormente, al insertar la Ecuación (A.17) en la Ecuación (A.10), se tiene que la expresión para la energía perturbativa del estado basal de segundo orden es:

$$E_0^2 = \sum_{n \neq 0} \frac{\langle 0^0|\mathcal{H}'|n\rangle\langle n|\mathcal{H}'|0^0\rangle}{E_0^0 - E_n^0} = \frac{|\langle 0^0|\mathcal{H}'|n\rangle|^2}{E_0^0 - E_n^0} \quad (\text{A.19})$$

Esta ecuación es conocida también como una expresión de suma sobre estados. Siguiendo de forma similar el procedimiento utilizado para encontrar la expresión de la Ecuación (A.19), se puede concluir que la energía de tercer orden tendrá la siguiente forma:

$$E_0^3 = \sum_{n \neq 0} \frac{\langle 0^0|\mathcal{H}'|n\rangle\langle n|\mathcal{H}'|m\rangle\langle m|\mathcal{H}'|0^0\rangle}{E_0^0 - E_n^0} - E_0^1 \sum_{n \neq 0} \frac{|\langle 0^0|\mathcal{H}'|n\rangle|^2}{(E_0^0 - E_n^0)^2} \quad (\text{A.20})$$

A.1.1 El operador resolvente

Una forma más elegante de desarrollar la Teoría de Perturbaciones es mediante la construcción de un operador de proyección conocido como *resolvente*. Para este fin, se seguirá asumiendo que existe una solución no perturbada asociada al estado basal ($|0^0\rangle$) y, además, funciones de respuesta (de diferente orden) del estado basal ($|0^k\rangle$), las cuales pueden ser expandidas a partir de un conjunto complementario de vectores propios de orden cero.

En primer lugar, se define el operador de proyección del estado basal:

$$\mathcal{P}_0 = |0^0\rangle\langle 0^0| \quad (\text{A.21})$$

Al multiplicar por la derecha por $|0\rangle$ y utilizar el criterio de normalización intermedia (Ecuación (A.4)), se obtiene la siguiente expresión equivalente:

$$\mathcal{P}_0|0\rangle = |0^0\rangle \quad (\text{A.22})$$

Cualquier corrección de orden superior puede ser proyectada de $|0\rangle$ al usar el operador de proyección del complemento ortogonal de orden cero:

$$\mathcal{P}_n = \sum_{n \neq 0} |n\rangle\langle n| \quad (\text{A.23})$$

Ahora, se añade un parámetro de energía (ε) arbitrario a la ecuación de Schrödinger:

$$(\varepsilon + \mathcal{H})|0\rangle = (\varepsilon + E_0)|0\rangle \quad (\text{A.24})$$

donde \mathcal{H} , $|0\rangle$ y E_0 se encuentran definidos por las Ecuaciones (A.1), (A.3) y (A.6), respectivamente. Con base en el desarrollo de la Teoría de Perturbaciones de Rayleigh-Schrödinger² (RS PT, por sus siglas en inglés), $\varepsilon = E_0^0$, lo cual permite reescribir a la Ecuación (A.24) como:

$$(E_0^0 - \mathcal{H}_0)|0\rangle = (E_0^0 - E_0 + \mathcal{H}')|0\rangle \quad (\text{A.25})$$

Proyectando a la Ecuación (A.25) con el complemento ortogonal del estado basal (Ecuación (A.23)):

$$\mathcal{P}_n(E_0^0 - \mathcal{H}_0)|0\rangle = \mathcal{P}_n(E_0^0 - E_0 + \mathcal{H}')|0\rangle \quad (\text{A.26})$$

El operador \mathcal{P}_n conmuta con la energía de orden cero y con el Hamiltoniano de orden cero (puesto que está formado de sus funciones propias), la Ecuación (A.26) puede reescribirse como:

$$\mathcal{P}_n|0\rangle = (E_0^0 - \mathcal{H}_0)^{-1} \mathcal{P}_n(E_0^0 - E_0 + \mathcal{H}')|0\rangle \quad (\text{A.27})$$

la cual introduce el concepto de un operador inverso, el cual (cuando existe un operador \mathcal{O} para una función propia $|\psi_k\rangle$, con un valor propio σ_k) posee la siguiente propiedad:

$$(c - \mathcal{O})^{-1}|\psi_k\rangle = \frac{1}{c - \sigma_k}|\psi_k\rangle \quad (\text{A.28})$$

Dado que el operador de proyección exhibe la propiedad de idempotencia: $\mathcal{P}_n^2 = \mathcal{P}_n$, esto permite reescribir a la Ecuación (A.27) (al multiplicarla por la izquierda por \mathcal{P}_m) como:

$$\mathcal{P}_m|0\rangle = \mathcal{R}_{mn}(E_0^0 - E_0 + \mathcal{H}')|0\rangle \quad (\text{A.29})$$

La cual introduce al operador resolvente:

$$\mathcal{R}_{mn} = \mathcal{P}_m(E_0^0 - \mathcal{H}_0)^{-1} \mathcal{P}_n = \sum_{m, n \neq 0} |m\rangle\langle m|(E_0^0 - \mathcal{H}_0)^{-1} |n\rangle\langle n| \quad (\text{A.30})$$

A esta expresión también suele referirse como *operador de reacción de onda*. El complemento ortogonal es construido a partir de un conjunto ortonormal de funciones propias de \mathcal{H}_0 , por lo que la operación del bra-ket inverso muestra el siguiente comportamiento:

$$\langle m|(E_0^0 - \mathcal{H}_0)^{-1}|n\rangle = \frac{1}{E_0^0 - E_n^0} \langle m|n\rangle = \frac{1}{E_0^0 - E_n^0} \delta_{mn} \quad (\text{A.31})$$

la cual colapsa la suma en la Ecuación (A.30), y la simplifica de la siguiente forma:

$$\mathcal{R}_n = \sum_{n \neq 0} |n\rangle\langle n|(E_0^0 - \mathcal{H}_0)^{-1} |n\rangle\langle n| \quad (\text{A.32})$$

² $\varepsilon = E_0$ conduce al desarrollo de la Teoría de Perturbaciones de Brillouin-Wigner.

Una consecuencia importante, de cualquiera de las Ecuaciones (A.30) o (A.32), es que el estado $|0^0\rangle$ no debe estar degenerado en energía con ningún miembro de su complemento; de lo contrario, el resolvente puede introducir una singularidad en las expresiones.

De acuerdo a las Ecuaciones (A.3) y (A.14), la función de onda de respuesta del estado basal puede ser expresada en términos del operado de proyección del complemento ortogonal (Ecuación (A.23)) como:

$$|0\rangle = |0^0\rangle + |0^1\rangle + |0^2\rangle + \dots = |0^0\rangle + \mathcal{P}_n|0\rangle \quad (\text{A.33})$$

Insertando la Ecuación (A.29) del lado derecho de la Ecuación (A.33) se tiene que:

$$|0\rangle = |0^0\rangle + \mathcal{R}_n(E_0^0 - E_0 + \mathcal{H}')|0\rangle \quad (\text{A.34})$$

Esta expresión, al iterarla de la siguiente forma, proporciona una expansión de orden infinito de la función de respuesta del estado basal:

$$|0\rangle = \sum_{k=0}^{\infty} [\mathcal{R}_{mn}(E_0^0 - E_0 + \mathcal{H}')]^k |0^0\rangle \quad (\text{A.35})$$

De la cual, por ejemplo, los tres primeros términos son:

$$\begin{aligned} |0\rangle &= |0^0\rangle + \mathcal{R}_n \mathcal{H}' |0^0\rangle + \mathcal{R}_n (E_0^0 - E_0^1 + \mathcal{H}') \mathcal{R}_n |0^0\rangle + \dots \\ &= |0^0\rangle + \sum_{n \neq 0} \frac{|n\rangle \langle n | \mathcal{H}' | 0^0\rangle}{E_0^0 - E_n^0} \\ &\quad + \left[\sum_{m, n \neq 0} \frac{|m\rangle \langle m | \mathcal{H}' | n\rangle \langle n | \mathcal{H}' | 0^0\rangle}{(E_0^0 - E_n^0)(E_0^0 - E_m^0)} \right. \\ &\quad \left. - (E_0^1 - E_0^0) \sum_{n \neq 0} \frac{|n\rangle \langle n | \mathcal{H}' | 0^0\rangle}{(E_0^0 - E_n^0)^2} \right] + \dots \end{aligned} \quad (\text{A.36})$$

Cuando el Hamiltoniano perturbado actúa sobre esta función respuesta:

$$\mathcal{H}|0\rangle = E_0|0\rangle \quad (\text{A.37})$$

y se multiplica la Ecuación (A.37) por $\langle 0|$ y se sustituye el Hamiltoniano con la Ecuación (A.1), se encuentra que:

$$E_0 = E_0^0 + \langle 0^0 | \mathcal{H}' | 0^0 \rangle \quad (\text{A.38})$$

Utilizando la secuenciación de un parámetro de ordenamiento, se puede definir de la siguiente manera la energía de cualquier orden $n > 0$:

$$E_0^n - E_0^0 = \langle 0^0 | \mathcal{H}' | 0^{n-1} \rangle \quad (\text{A.39})$$

De la Ecuación (A.35), se obtiene una expansión arreglada en términos del orden perturbativo que adquiere la misma forma de, por ejemplo, las Ecuaciones (A.19) y (A.20). Por lo tanto, las energías de primer, segundo y tercer orden tendrán la siguiente forma, respectivamente:

$$E_0^1 = \langle 0^0 | \mathcal{H}' | 0^0 \rangle \quad (\text{A.40})$$

$$E_0^2 = \langle 0^0 | \mathcal{H}' \mathcal{R}_n \mathcal{H}' | 0^0 \rangle \quad (\text{A.41})$$

$$E_0^3 = \langle 0^0 | \mathcal{H}' \mathcal{R}_n \mathcal{H}' \mathcal{R}_m \mathcal{H}' | 0^0 \rangle - E_0^1 \langle 0^0 | \mathcal{H}' \mathcal{R}_n^2 \mathcal{H}' | 0^0 \rangle \quad (\text{A.42})$$

A.2 La Aproximación de Polarización³

Considerando dos moléculas (o fragmentos de una molécula) A y B, a los cuales se les asocia los hamiltonianos \mathcal{H}^A y \mathcal{H}^B , y las ecuaciones de Schrödinger $\mathcal{H}^A|i^A\rangle = E_i^A|i^A\rangle$ y $\mathcal{H}^B|j^B\rangle = E_j^B|j^B\rangle$ (con $i, j = 0, 1, 2, \dots$), respectivamente, la suma $\mathcal{H} = \mathcal{H}^A + \mathcal{H}^B$ y los productos $|i^A j^B\rangle = |i^A\rangle|j^B\rangle$ son el Hamiltoniano y los eigenestados la supermolécula A + B, por lo que la ecuación de Schrödinger asociada a este sistema sería:

$$\mathcal{H}|i^A j^B\rangle = (E_i^A + E_j^B)|i^A j^B\rangle = E_{ij}|i^A j^B\rangle \quad (\text{A.43})$$

Si A y B interactúan entre ellos, pero la energía de interacción (E_{int}) es mucho menor a la energía total de A y B (E_i^A y E_j^B , respectivamente), se puede utilizar la Teoría de Perturbaciones para determinar la magnitud de E_{int} al definir un Hamiltoniano de perturbación o de interacción (\mathcal{H}') como una expresión (estándar) electrostática (en unidades atómicas):

$$\mathcal{H}' = - \sum_{i \in A} \sum_{\beta \in B} Z_{\beta} r_{i\beta}^{-1} - \sum_{j \in B} \sum_{\alpha \in A} Z_{\alpha} r_{j\alpha}^{-1} + \sum_{i \in A} \sum_{j \in B} r_{ij}^{-1} \quad (\text{A.44})$$

donde α y β corresponden a las etiquetas asociadas a los núcleos, mientras que i y j las correspondientes a los electrones. Por lo tanto, las correcciones de a la energía de primer y segundo orden (dadas por las Ecuaciones (A.40) y (A.41)) adquieren la siguiente forma:

$$\mathcal{U}^{(1)} = \langle 0^A 0^B | \mathcal{H}' | 0^A 0^B \rangle \quad (\text{A.45})$$

$$\mathcal{U}^{(2)} = \langle 0^A 0^B | \mathcal{H}' \mathcal{R}_n \mathcal{H}' | 0^A 0^B \rangle \quad (\text{A.46})$$

donde $\mathcal{U}^{(1)}$ representa la interacción electrostática entre los fragmentos no perturbados A y B (E_{elstat}):

$$\mathcal{U}^{(1)} = E_{elstat} = \iint d\mathbf{r}_1 d\mathbf{r}_2 r_{12}^{-1} \rho_t^A(\mathbf{r}_1) \rho_t^B(\mathbf{r}_2) \quad (\text{A.47})$$

donde $\rho_t^A(\mathbf{r}_1)$ es la densidad de carga total de A (incluyendo las contribuciones nucleares), teniendo una definición similar para $\rho_t^B(\mathbf{r}_2)$. Por otro lado, \mathcal{R}_n puede ser escrito de la siguiente forma:

$$\mathcal{R}_n = \mathcal{R}_{n,ind}^{A \rightarrow B} + \mathcal{R}_{n,ind}^{B \rightarrow A} + \mathcal{R}_{n,disp} = \mathcal{R}_{n,ind} + \mathcal{R}_{n,disp} \quad (\text{A.48})$$

donde los subíndices *ind* y *disp* se refieren a contribuciones de inducción⁴ y dispersión⁵, respectivamente:

$$\mathcal{R}_{n,ind}^{A \rightarrow B} = \sum_{i \neq 0} \frac{|i^A 0^B\rangle \langle i^A 0^B|}{E_{00}^0 - E_{i0}^0} \quad (\text{A.49})$$

$$\mathcal{R}_{n,ind}^{B \rightarrow A} = \sum_{j \neq 0} \frac{|0^A j^B\rangle \langle 0^A j^B|}{E_{00}^0 - E_{0j}^0} \quad (\text{A.50})$$

$$\mathcal{R}_{n,disp} = \sum_{i,j \neq 0} = \frac{|i^A j^B\rangle \langle i^A j^B|}{E_{00}^0 - E_{ij}^0} \quad (\text{A.51})$$

³El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por la referencia [1].

⁴La inducción se refiere a la deformación de la densidad electrónica debido al campo eléctrico generado por todos los vecinos alrededor de una molécula. Esta contribución siempre es atractiva [7]

⁵La dispersión es una contribución no clásica, la cual tiene origen en las fluctuaciones constantes de la distribución de carga de las moléculas, asociadas al movimiento de los electrones [7].

$\mathcal{R}_{n,ind}^{A \rightarrow B}$ incluye únicamente términos en la Ecuación (A.32) asociados al estado basal de B (0^B) y a estados excitados de A ($i^A \neq 0^A$), $\mathcal{R}_{n,ind}^{B \rightarrow A}$ los términos de A en el estado basal (0^A) y los estados excitados de B ($j^B \neq 0^B$) y, por último, $\mathcal{R}_{n,disp}$ a los estados excitados tanto de A como de B ($i^A \neq 0^A$ y $j^B \neq 0^B$, respectivamente).

Sustituyendo la Ecuación (A.48) en (A.46), se tiene que $\mathcal{U}^{(2)}$ es:

$$\mathcal{U}^{(2)} = \mathcal{U}_{ind}^A + \mathcal{U}_{ind}^B + \mathcal{U}_{disp} = \mathcal{U}_{ind} + \mathcal{U}_{disp} \quad (\text{A.52})$$

donde \mathcal{U}_{ind}^A y \mathcal{U}_{ind}^B representan la energía de inducción de la molécula A en el campo generado por B y la energía de inducción de B en el campo de A, respectivamente. \mathcal{U}_{disp} corresponde a la energía de dispersión de segundo orden. Todos estos términos, en conjunto con E_{elstat} , son la base para la descripción de las interacciones no-covalentes.

Esta parte de la Teoría de Perturbaciones para la descripción de las interacciones intermoleculares es solamente válida para las interacciones de largo alcance (Región III de la Figura 2.2), donde el traslape de la densidad electrónica es nulo. Es decir, deben existir dos regiones no-traslapantes Ω_A y Ω_B , de tal forma que $|i^A\rangle \neq 0$ sólo cuando todos los electrones de A se encuentren en Ω_A y, viceversa, $|j^B\rangle \neq 0$ sólo cuando todos los electrones de B residan en Ω_B . Sin embargo, incluso en el límite de largo alcance, la función de onda de orden cero correcta para la supermolécula A + B debe tener la forma $\mathcal{A}|i^A\rangle|j^B\rangle$, donde \mathcal{A} es un operador antisimetrizador que intercambia los electrones de A y B en todas las formas posibles⁶. Bajo este régimen, se puede asignar un número fijo de electrones N_A al fragmento A y otro número N_B al fragmento B. Además, se puede despreciar los fenómenos de transferencia de carga dado que los fragmentos no se superponen. Fuera de la zona de largo alcance, pero donde la superposición es todavía pequeña, la identidad $E_{int} = \mathcal{U}^{(1)} + \mathcal{U}^{(2)}$ sigue siendo, aproximadamente, válida. Al menos hasta los términos de segundo orden de la Teoría de Perturbaciones.

A.3 Teoría de Perturbaciones Adaptada por Simetría (SAPT)⁷

Conforme las moléculas (o fragmentos) A y B se acercan entre sí, y el traslape de sus densidades electrónicas se vuelve más significativo, la Aproximación de Polarización comienza a fallar puesto que cuando las densidades electrónicas de los fragmentos A y B empiezan a traslaparse, el papel del operador \mathcal{A} sobre $|i^A\rangle|j^B\rangle$ ya no puede ser ignorado; es decir, el intercambio de electrones entre ambas moléculas debe ser tomado en cuenta (de otra forma, el principio de exclusión de Pauli no se cumpliría).

El intercambio electrónico es la principal causa de la repulsión entre moléculas a distancias cortas, evitando así que se interpenetren. Sin embargo, la antisimetrización de cada

⁶De otra forma, aún cuando $|i^A\rangle|j^B\rangle$ permanezcan igual o cambien de signo al realizar una permutación arbitraria de los electrones de A o de B ($|i^A\rangle$ y $|j^B\rangle$ son individualmente antisimétricas), el producto simple no es ni simétrico ni antisimétrico con respecto al intercambio de electrones entre los fragmentos, por lo que no se satisface el principio de antisimetría de Pauli. Sin embargo, se puede demostrar que esta falta de simetría de $|i^A\rangle|j^B\rangle$ no tiene ninguna consecuencia energética; por lo tanto, las Ecuaciones (A.45) y (A.46) son exactamente válidas cuando A y B no se traslapan

⁷El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por la referencia [1].

función $|i^A\rangle|j^B\rangle$ trae consigo un problema. El conjunto $\{|i^A\rangle|j^B\rangle\}$ es ortogonal, siendo cada uno de sus elementos un estado propio del Hamiltoniano no perturbado \mathcal{H}_0 . Desde este punto de vista, este conjunto ortonormal es el que debería utilizarse dentro de la Teoría de Perturbaciones. No obstante, el conjunto antisimetrizado $\{\mathcal{A}|i^A\rangle|j^B\rangle\}$ no es ortonormal, por lo que sus elementos no son funciones propias de ningún operador Hamiltoniano y no pueden ser usadas en la Teoría de Perturbaciones de Rayleigh-Schrödinger. Un método que trata de resolver este percance es la Teoría de Perturbaciones Adaptada por Simetría (SAPT, por sus siglas en inglés), la cual evita los problemas asociados a la ortonormalidad del conjunto $\{\mathcal{A}|i^A\rangle|j^B\rangle\}$ al trabajar con productos simples como $\{|i^A\rangle|j^B\rangle\}$. Además, evita el incumplimiento del principio de Pauli al trabajar también con $\{\mathcal{A}|i^A\rangle|j^B\rangle\}$. Por lo tanto, ambos conjuntos son necesarios en SAPT.

En una primera formulación de SAPT, propuesta por Jeziorski and Kołos, la ecuación de Schrödinger es resuelta iterativamente, forzando en cada paso la simetría correcta de las soluciones intermedias. Para esto, el punto de partida son las Ecuaciones (A.34) y (A.38), escritas ahora de la siguiente forma:

$$\psi_0(n) = \phi^0 + \mathcal{R}_n[-\varepsilon_0(n) + \mathcal{H}']\psi_0(n-1) \quad (\text{A.53})$$

$$\varepsilon_0(n) = \langle \phi^0 | \mathcal{H}' | \psi_0(n-1) \rangle \quad (\text{A.54})$$

donde $n = 1, 2, 3, \dots$ es el número de iteración, ψ_0 es la solución exacta del estado basal de la Ecuación de Schrödinger para el sistema perturbado-interactuante ($\mathcal{H}\psi_0 = E_0\psi_0$), $\phi^0 \equiv |0^A 0^B\rangle$ es la función propia del estado basal del Hamiltoniano no perturbado \mathcal{H}^0 , y $\varepsilon_0 = E_0 - E_0^0$ ($E_0 \equiv E_{00}^0$) es el valor propio de \mathcal{H}^0 . El proceso iterativo inicia cuando $n = 1$ y $\psi_0(0) = \phi^0$. Utilizando esta función de onda en la Ecuación (A.54) se obtiene:

$$\varepsilon_0(1) = \langle \phi^0 | \mathcal{H}' | \phi^0 \rangle = E_{elstat} \equiv E_{pol}^{(1)} \quad (\text{A.55})$$

la cual corresponde a la energía de interacción electrostática entre los fragmentos no perturbados (E_{elstat}) o a la energía de primer orden en la Aproximación de Polarización ($E_{pol}^{(1)}$).

Sustituyendo el resultado anterior y $\psi_0 = \phi^0$ en la Ecuación (A.53), $\psi_0(1)$ resulta ser $\psi_0(1) = \phi^0 + \mathcal{R}_1 \mathcal{H}' \phi^0$, expresión en la cual sea ha tomado en cuenta que $\mathcal{R}_0 \phi^0 = 0$. La función $\psi_0(1)$ coincide con la función de onda de primer orden de la Teoría de Perturbaciones de Rayleigh-Schrödinger. Usando ahora $\psi_0(1)$ en la Ecuación (A.54), se tiene que:

$$\varepsilon_0(2) = \langle \phi^0 | \mathcal{H}' (1 + \mathcal{R}_0 \mathcal{H}') | \phi^0 \rangle = E_{pol}^{(1)} + \langle \phi^0 | \mathcal{H}' \mathcal{R}_0 \mathcal{H}' | \phi^0 \rangle = E_{pol}^{(1)} + E_{pol}^{(2)} \quad (\text{A.56})$$

donde $E_{pol}^{(2)} \equiv \mathcal{U}^{(2)}$ es la energía de segundo orden en la Aproximación de Polarización. Al repetir las iteraciones para n 's de órdenes superiores, y agrupar los términos de la misma potencia de perturbación de \mathcal{H}' , se obtendrán las mismas expresiones que en la Teoría de Perturbaciones de Rayleigh-Schrödinger, más otros términos de órdenes superiores. La velocidad a la que converge este proceso iterativo dependerá principalmente en la función $\psi^0(0)$ de inicio. Sin embargo, $\psi_0(0)$ no es, particularmente, la mejor elección, incluso a distancias intermoleculares grandes. En efecto, en el límite de largo alcance, $\varepsilon_0(n)$ se convierte en un aproximación arbitrariamente buena de la energía de interacción para valores grandes de n . No obstante, dada la falta de simetría en $\phi^0 = |0^A 0^B\rangle$, la energía de interacción debida al intercambio de electrones entre dos moléculas (energía de interacción de intercambio) no se reproduce en

ningún orden finito.

Tomando en cuenta lo mencionado anteriormente, una mejor opción de $\psi_0(0)$ para iniciar las iteraciones en la Ecuación (A.54) es $\psi_0(0) = \mathcal{N}\mathcal{A}\phi^0 = \mathcal{B}\phi^0$, donde $\mathcal{N} = \langle \phi^0 | \mathcal{A} | \phi^0 \rangle^{-1}$ y \mathcal{A} es un antisimetrizador parcial que opera sobre las $N!/(N_A!N_B!)$ permutaciones de electrones intersistema, intercambiando las coordenadas de los electrones asociados a A con las coordenadas de los electrones asociados a B. Adicionalmente, \mathcal{N} asegura el cumplimiento de la condición de normalización intermedia $\langle \phi^0 | \psi_0 \rangle = 1$. Por lo tanto, la primera iteración en la Ecuación (A.54) da como corrección de primer orden a la energía:

$$\varepsilon_0(1) = \frac{\langle \phi^0 | \mathcal{H}' \mathcal{A} | \phi^0 \rangle}{\langle \phi^0 | \mathcal{A} | \phi^0 \rangle} \quad (\text{A.57})$$

la cual puede ser dividida en una contribución de polarización ($E_{pol}^{(1)}$) (Ecuación (A.55)) y el resto:

$$\varepsilon_0(1) = E_{pol}^{(1)} + E_{exh}^{(1)} \quad (\text{A.58})$$

Siendo las Ecuaciones (A.57) y (A.58) definen la energía de interacción de intercambio.

La función de onda de la primera iteración ($\psi_0(1)$) puede ser derivada al utilizar la propiedad de conmutación $[\mathcal{A}, \mathcal{H}] = 0$, el hecho de que ϕ^0 es el estado propio de \mathcal{H}^0 con el valor propio E_0^0 , y las propiedades del resolvente \mathcal{R}_n . Todo esto nos conduce a [8]:

$$\psi_0(1) = \psi_0(0) - \mathcal{N}\mathcal{R}_n\mathcal{A}(\varepsilon_0(1) - \mathcal{H}')\phi^0 \quad (\text{A.59})$$

Al sustituir la Ecuación (A.59) en la Ecuación (A.54) se obtiene:

$$\varepsilon_0(2) = \varepsilon_0(1) - \mathcal{N}\langle \phi^0 | \mathcal{H}'\mathcal{R}_n\mathcal{A}(\varepsilon_0(1) - \mathcal{H}') | \phi^0 \rangle \quad (\text{A.60})$$

Estas ecuaciones corresponden a la Teoría de Perturbaciones derivadas por Murrell, Sahw, Musher y Amos (MS-MA). Esta propuesta es un caso más general del esquema iterativo propuesto por Jeziorski y Kołos, cuyas ecuaciones básicas son modificaciones de las Ecuaciones (A.53) y (A.54). La diferencia recae en el hecho de que $\psi_0(n-1)$ en las Ecuaciones (A.53) y (A.54) es reemplazado por $\mathcal{F}\psi_0(n-1)$ y $\mathcal{G}\psi_0(n-1)$, respectivamente, donde \mathcal{F} y \mathcal{G} son operadores que deben cumplir las siguientes condiciones después de la convergencia: $\mathcal{F}\psi_0 = \psi_0$ y $\mathcal{G}\psi_0 = \psi_0$. El planteamiento MS-MA corresponde a $\mathcal{F} = \mathcal{G} = 1$ y $\psi_0(0) = \mathcal{B}\phi^0 = \langle \phi^0 | \mathcal{A} | \phi^0 \rangle^{-1}$. La aproximación de polarización también encaja en el esquema anterior, eligiendo que $\mathcal{F} = \mathcal{G} = 1$ y $\psi_0(0) = \phi^0$.

Una comparación detallada entre los diferentes términos energéticos obtenidos de los diferentes esquemas de SAPT con los provenientes de la Aproximación de Polarización requiere un estudio detallado del antisimetrizador \mathcal{A} . Este operador puede ser escrito de la siguiente forma:

$$\mathcal{A} = C(1 - p^{(1)} + p^{(2)} - \dots) = C(1 + \mathcal{P}) \quad (\text{A.61})$$

donde C es una constante y $p^{(1)}$ denota al operador que intercambia uno de los electrones de A con uno de los electrones de B, $p^{(2)}$ el operador que intercambia dos electrones de A con dos electrones de B, y así sucesivamente. Particularmente, para $p^{(1)}$ se tiene:

$$p^{(1)} = \sum_{i \in A} \sum_{j \in B} p_{ij}^{(1)} \quad (\text{A.62})$$

donde $p_{ij}^{(1)}$ intercambia las coordenadas del electrón i en el fragmento A y del electrón j en el fragmento B. Se puede demostrar que $\langle 0^A 0^B | p_{ij}^{(1)} | 0^A 0^B \rangle \sim s^2$, donde s representa una medida del traslape entre los orbitales de A y B. Teniendo esto en cuenta, y despreciando los términos $S^{(4)}$ de orden superior y superiores, la Ecuación (A.57) queda de la siguiente forma:

$$\varepsilon_0(1) \simeq \frac{\langle 0 | \mathcal{H}'(1 - p^{(1)} + \dots) | 0 \rangle}{\langle (1 - p^{(1)} + \dots) | 0 \rangle} \quad (\text{A.63})$$

$$\simeq \left[\langle 0 | \mathcal{H}' | 0 \rangle - \langle 0 | \mathcal{H}' p^{(1)} | 0 \rangle \right] \left[1 + \langle 0 | p^{(1)} | 0 \rangle \right] + \mathcal{O}(s^4) \quad (\text{A.64})$$

$$\simeq E_{pol}^{(1)} - \langle 0 | \mathcal{H}' p^{(1)} | 0 \rangle + E_{pol}^{(1)} \langle 0 | p^{(1)} | 0 \rangle + \mathcal{O}(s^4) \quad (\text{A.65})$$

donde la propiedad $1/(1-x) \simeq 1+x$ fue utilizada para expandir el denominador de la Ecuación (A.65), $|0^A 0^B\rangle \equiv \phi^0$ fue abreviado como $|0\rangle$, y $E_{elstat} = E_{pol}^{(1)} = \langle 0 | \mathcal{H}' | 0 \rangle$. Al comparar la Ecuación (A.65) con la Ecuación (A.65), $E_{exch}^{(1)}$ resulta ser:

$$E_{exch}^{(1)} = -\langle 0 | \mathcal{H}' p^{(1)} | 0 \rangle + E_{pol}^{(1)} \langle 0 | p^{(1)} | 0 \rangle + \mathcal{O}(s^4) \quad (\text{A.66})$$

De las Ecuaciones (A.57), (A.58) y (A.61), $E_{exch}^{(1)}$ queda expresada como (antes de expandir \mathcal{A} como en la Ecuación (A.61)):

$$E_{exch}^{(1)} = \frac{\langle \phi^0 | (\mathcal{H}' - E_{pol}^{(1)}) \mathcal{P} | \phi^0 \rangle}{1 + \langle \phi^0 | \mathcal{P} | \phi^0 \rangle} \quad (\text{A.67})$$

Un cálculo preciso de la expresión anterior es difícil debido a los múltiples intercambios de electrones implicados en \mathcal{P} . No obstante, la Ecuación (A.67) representa una buena aproximación a distancias intermoleculares correspondientes a mínimos de van der Waals típicos, puesto que su error es de cuarto orden en las integrales de traslape intermolecular.

Se puede demostrar que la corrección a la energía de primer orden de SAPT ($\varepsilon_0(1)$, Ecuación (A.57)) es idéntica a la energía de Heitler-London (HL), definida por [8]:

$$E_{HL} = \frac{\langle \mathcal{A} \phi^0 | \mathcal{H} | \mathcal{A} \phi^0 \rangle}{\langle \mathcal{A} \phi^0 | \mathcal{A} \phi^0 \rangle} - E_0^A - E_0^B \quad (\text{A.68})$$

Adicionalmente, $\varepsilon_0(1)$ coincide con la energía de primer orden de la Teoría de Perturbaciones Intermolecular (IMPT, por sus siglas en inglés) de Stone y Hayes. Para esto, se debe mencionar que la energía de orden cero más la de primer orden en IMPT puede ser obtenida como la energía del determinante de Slater no ortogonal, formado por todos los orbitales moleculares ocupados (ϕ) de los fragmentos en sus estados basales y, dado que los orbitales del fragmento A (o B) son ortonormales entre sí, pero los orbitales del fragmento A no son ortogonales a los del fragmento B, esta energía está dada por:

$$(E_0 + E_1)^{IMPT} = \sum_{i,j} h_{i,j} S_{ji}^{-1} + \frac{1}{2} \sum_{i,k,j,m} \langle ik | jm \rangle (S_{ji}^{-1} S_{mk}^{-1} - S_{mi}^{-1} S_{jk}^{-1}) \quad (\text{A.69})$$

donde $h_{ij} = \langle \phi_i | \hat{h} | \phi_j \rangle$ y \hat{h} es el Hamiltoniano monoeléctrico que incluye la contribución de energía cinética (\hat{t}), la atracción hacia los núcleos de la molécula A ($-\sum_{\alpha \in A} Z_{\alpha}/r$) y la

atracción hacia el núcleo B ($-\sum_{\beta \in B} Z_{\beta}/r$); $\langle ik|jm \rangle$ es la integral bielectrónica de repulsión $\langle \phi_i(1)\phi_k(2)|r_{12}^{-1}|\phi_j(1)\phi_m(2) \rangle$; y $S_{ij} = \langle \phi_i|\phi_j \rangle$, con $S_{ij} \neq \delta_{ij}$ si $\phi_i \in A$ y $\phi_j \in B$. Sin embargo, la expresión anterior coincide con el primer término de la Ecuación (A.68) cuando las $N!/(N_A!N_B!)$ permutaciones intersistema son incluidas en \mathcal{A} . Al separar la energía de los fragmentos aislados de la Ecuación (A.69), se tiene que $E_0^{IMPT} = E_0^A + E_0^B$, mientras que E_1^{IMPT} puede ser escrito como:

$$E_1^{IMPT} = E_{elstat} + E_X + E_{rep} \quad (\text{A.70})$$

donde $E_{elstat} = E_{pol}^{(1)}$ es la energía de interacción electrostática descrita anteriormente, y

$$E_X = - \sum_{i \in A, j \in B} \langle \phi_i \phi_j | \phi_j \phi_i \rangle \quad (\text{A.71})$$

es la energía de interacción de intercambio entre los orbitales del fragmento A y los orbitales del fragmento B. La energía de repulsión (E_{rep}) es una combinación de correcciones (de uno y dos electrones) a la energía ortogonal dada por la Ecuación (A.69) si $S_{ij} = \delta_{ij}$:

$$E_{rep} = \sum_{i,j} h_{ij}(S_{ji}^{-1} - \delta_{ji}) + \frac{1}{2} \sum_{i,k,j,m} \langle ik|jm \rangle (S_{ji}^{-1} S_{mk}^{-1} - S_{mi}^{-1} S_{jk}^{-1} - \delta_{ji} \delta_{mk} + \delta_{mi} \delta_{jk}) \quad (\text{A.72})$$

Comúnmente, las energías de intercambio y de repulsión suelen sumarse ($E_X + E_{rep} = E_{xr}$) y referírseles como energía de intercambio-repulsión. Para sistemas simples y pequeños (como el estado triplete de la molécula de H_2 , por ejemplo) E_{xr} coincide exactamente con la energía de intercambio de primer orden ($E_{xr} \equiv E_{exch}^{(1)}$); por lo tanto, las Ecuaciones (A.58) y (A.70) son equivalentes. Al observar la Ecuación (A.72), resulta evidente que E_{rep} incluye correcciones de antisimetrización en todos los términos energéticos del Hamiltoniano (energía cinética, atracciones nucleares y contribuciones Coulómbicas y de intercambio), tanto a nivel intra como interfragmento.

Para calcular $\varepsilon_0(2)$ en SAPT, se puede utilizar la Ecuación (A.60), en la cual $\varepsilon_0(1)$ ya está separada de las correcciones de segundo orden:

$$\varepsilon_0(2) - \varepsilon_0(1) = - \frac{\langle 0 | \mathcal{H}' \mathcal{R}_n \mathcal{A} (\varepsilon_0(1) - \mathcal{H}') | 0 \rangle}{\langle 0 | \mathcal{A} | 0 \rangle} \quad (\text{A.73})$$

Utilizando las propiedades de \mathcal{R}_n , se puede demostrar que $\langle 0 | \mathcal{H}' \mathcal{R}_n \mathcal{A} \varepsilon_0(1) | 0 \rangle = 0$, por lo tanto:

$$\varepsilon_0(2) - \varepsilon_0(1) = \frac{\langle 0 | \mathcal{H}' \mathcal{R}_n \mathcal{A} \mathcal{H}' | 0 \rangle}{\langle 0 | \mathcal{A} | 0 \rangle} \quad (\text{A.74})$$

Si el antisimetrizador \mathcal{A} (Ecuación (A.61)) es insertado en la Ecuación (A.74), y utilizando la misma estrategia que en la Ecuación (A.64), se obtiene que:

$$\varepsilon_0(2) - \varepsilon_0(1) = \mathcal{U}_{ind} + \mathcal{U}_{disp} + E_{exch}^{(2)} \quad (\text{A.75})$$

donde \mathcal{U}_{ind} y \mathcal{U}_{disp} son las energías de interacción de inducción y de dispersión de la Aproximación de Polarización (Ecuación (A.52)), y $E_{exch}^{(2)}$ representa la corrección de intercambio que surge de tener adecuadamente considerada la antisimetría de la función de onda. Si el segundo término de la Ecuación (A.48) es usado, $E_{exch}^{(2)}$ puede ser dividido en sus contribuciones de dispersión y de inducción:

$$E_{exch}^{(2)} = E_{exch-ind}^{(2)} + E_{exch-disp}^{(2)} \quad (\text{A.76})$$

con:

$$E_{exch-ind}^{(2)} = -\langle 0|\mathcal{H}\mathcal{R}_{n,ind}p^{(1)}\mathcal{H}'|0\rangle + \langle 0|\mathcal{H}'\mathcal{R}_{n,ind}\mathcal{H}'|0\rangle\langle 0|p^{(1)}|0\rangle + \mathcal{O}(s^4) \quad (\text{A.77})$$

$$E_{exch-disp}^{(2)} = -\langle 0|\mathcal{H}'\mathcal{R}_{n,disp}p^{(1)}\mathcal{H}'|0\rangle + \langle 0|\mathcal{H}'\mathcal{R}_{0,disp}\mathcal{H}'|0\rangle\langle 0|p^{(1)}|0\rangle + \mathcal{O}(s^4) \quad (\text{A.78})$$

En resumen, dentro de SAPT se obtiene las mismas correcciones a la energía encontradas en la Aproximación de Polarización (las cuales desprecian el traslape entre orbitales moleculares de los dos fragmentos interactuantes y la antisimetría de la función de onda), más las contrapartes de intercambio, las cuales toman en cuenta los efectos despreciados en la Aproximación de Polarización:

$$E_{int} = E_{pol}^{(1)} + E_{exch}^{(1)} + E_{pol}^{(2)} + E_{exch}^{(2)} + \dots = \sum_n \left[E_{pol}^{(n)} + E_{exch}^{(n)} \right] \quad (\text{A.79})$$

Las correcciones de polarización en SAPT ($E_{pol}^{(n)}$) son los coeficientes en la expansión de Taylor del valor propio del estado basal ($E_0(\lambda)$) del Hamiltoniano $\mathcal{H} = \mathcal{H}^0 + \lambda\mathcal{H}'$, y estos son idénticos a sus análogos en la Teoría de Perturbaciones de Rayleigh-Schrödinger. El sufijo *pol* se refiere al hecho que estos término ignoran por completo el intercambio de electrones entre los fragmentos A y B. De forma contraria a $E_{pol}^{(n)}$ y $E_{exch}^{(1)}$, los términos de intercambio con $i > 1$, y, consecuentemente, la rapidez de convergencia de la Ecuación (A.79), difieren en las diferentes formulaciones de SAPT. Sin embargo, en el caso de las interacciones de van der Waals entre fragmentos poco polares, estas diferencias implican grandes problemas, dado que los primeros cuatro términos de la Ecuación (A.79) dominan en esta interacción y $E_{exch}^{(2)}$ no difiere en las diferentes aproximaciones de SAPT.

Hasta este punto, se ha asumido que la ecuación de Schrödinger para los fragmentos aislados (no interactuantes) se ha resuelto de forma exacta y, consecuentemente, los valores propios y las funciones propias del Hamiltoniano \mathcal{H}^0 son conocidas. Sin embargo, como es de esperarse, esto en general no es cierto. Para moléculas con más de un electrón, no es posible tener la solución exacta, por lo que solamente se pueden conocer energías y funciones de ondas aproximadas de cada fragmento. Esto complica, considerablemente, es esquema de solución de SAPT. LA forma habitual de proceder es comenzando con funciones de ondas aproximadas que describan los fragmentos, como, por ejemplo, los determinantes de Slater obtenidos de resolver las ecuaciones de campo autoconsistente (SCF, por sus siglas en inglés) de Hartree-Fock (HF). Posteriormente, se hace que ambos fragmentos interactúen y se aplica la Teoría de Perturbaciones de tres parámetros, la cual es capaz de determinar la energía de interacción entre los fragmentos, así como aproximar la energía de correlación dentro de cada fragmento. El Hamiltoniano no perturbado \mathcal{H}^0 se toma como la suma de los operadores de Fock de ambos fragmentos ($\mathcal{H} = F^A + F^B = F$), la perturbación (\mathcal{H}') incluye la interacción electrostática entre las partículas (electrones + núcleos) de A y de las partículas de B (Ecuación (A.44))⁸, y la diferencia entre el Hamiltoniano verdadero y el operador de Fock de cada molécula ($W^X = \mathcal{H}^X - F^{X(X=A,B)}$). Denotando como ζ , λ_A y λ_B como los parámetros formales de expansión de la teoría, \mathcal{H} , por lo tanto, puede ser escrito como:

$$\mathcal{H} = F + [\zeta V + \lambda_A W^A + \lambda_B W^B] = F + \mathcal{H}' \quad (\text{A.80})$$

⁸En este contexto, \mathcal{H}' suele llamarse V .

Aplicando la Teoría de Perturbaciones de tres términos (Ecuación (A.80)), se obtiene la siguiente expansión:

$$E = \sum_{n=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \left[E_{pol}^{(nij)} + E_{exch}^{(nij)} \right] = \sum_{n=0}^{\infty} \left[E_{pol}^{(n)} + E_{exch}^{(n)} \right] \quad (\text{A.81})$$

donde $E_{pol}^{(nij)}$ y $E_{exch}^{(nij)}$ son de orden n en V , y de orden i y j en λ_A y λ_B , respectivamente. Los términos con $n = 0$ carecen de las contribuciones de intercambio y son el equivalente a apagar la perturbación interfragmento V o a hacer que $\zeta = 0$. Si se establece que $n = i = j = 0$, el término E_{pol}^{000} corresponde a la suma de las energías de los orbitales HF SCF de ambos fragmento. Esto es semejante a la suma de las energías totales de orden cero de los fragmentos A y B dentro de la Teoría de Perturbaciones de Møller-Plesset (MP) de muchos cuerpos. Si el término E_{pol}^{011} es añadido a E_{pol}^{000} , el resultado es la suma de las energías totales de HF de ambas moléculas. Finalmente, si se suman los términos E_{pol}^{0ij} hasta ciertos valores máximos de i y j , E corresponde a la suma de la energía total del fragmento A (descrita en el nivel MP*i*) y la energía total del fragmento B (descrita en el nivel MP*j*).

Las correcciones en la Ecuación (A.81) con $n > 0$ describen los diferentes tipos de interacción entre los dos fragmentos. Entre estos, los términos con $i = j = 0$ son los más importantes: $E_{pol}^{(n00)}$ y $E_{exch}^{(n00)}$ corresponden a la interacciones de polarización e intercambio entre los fragmentos A y B, descritos en el nivel de teoría HF. Como se mencionó anteriormente, $E_{pol}^{(100)} \equiv E_{elstat}$ es la interacción electrostática clásica entre las densidades electrónicas HF de los fragmentos, mientras que $E_{exch}^{(100)}$ es el principal componente de la repulsión de intercambio o de Pauli. En el mismo sentido, $E_{pol}^{(200)}$ y $E_{exch}^{(200)}$ recopilan las interacciones de inducción más dispersión e intercambio, respectivamente. $E_{pol}^{(nij)}$ y $E_{exch}^{(nij)}$ con i y/o j mayor que cero recopilan todos los efectos de la correlación interfragmento sobre las energías no-correlacionadas HF. Por ejemplo, $E_{pol}^{(1ij)}$ y $E_{exch}^{(1ij)}$ toman en cuenta los efectos de correlación interfragmento en las energías de interacción electrostáticas y de repulsión de intercambio, respectivamente; y $E_{pol}^{(2ij)}$ y $E_{exch}^{(2ij)}$ los efectos de la correlación interfragmento en las interacciones de inducción más dispersión y de intercambio.

Es habitual sumar los índices interfragmento y utilizar $E^{(nk)}$ para identificar el n -ésimo orden en V y el k -ésimo orden en W^A y W^B . Los términos debidos a la correlación interfragmento ($k > 0$) son mucho más difíciles de evaluar, en comparación con el término principal ($k = 0$), correspondiente a una descripción HF de los monómeros. La dificultad aumenta con k , y es mayor para el intercambio que para los términos directos, ya que el primero requiere un cálculo de múltiples integrales de traslape entre orbitales no-ortogonales. En muchos de los casos, entre todas las contribuciones de intercambio y correlación interfragmento, únicamente las correcciones de primer orden en V ($n = 1$) dan una contribución no despreciable a E_{int} .

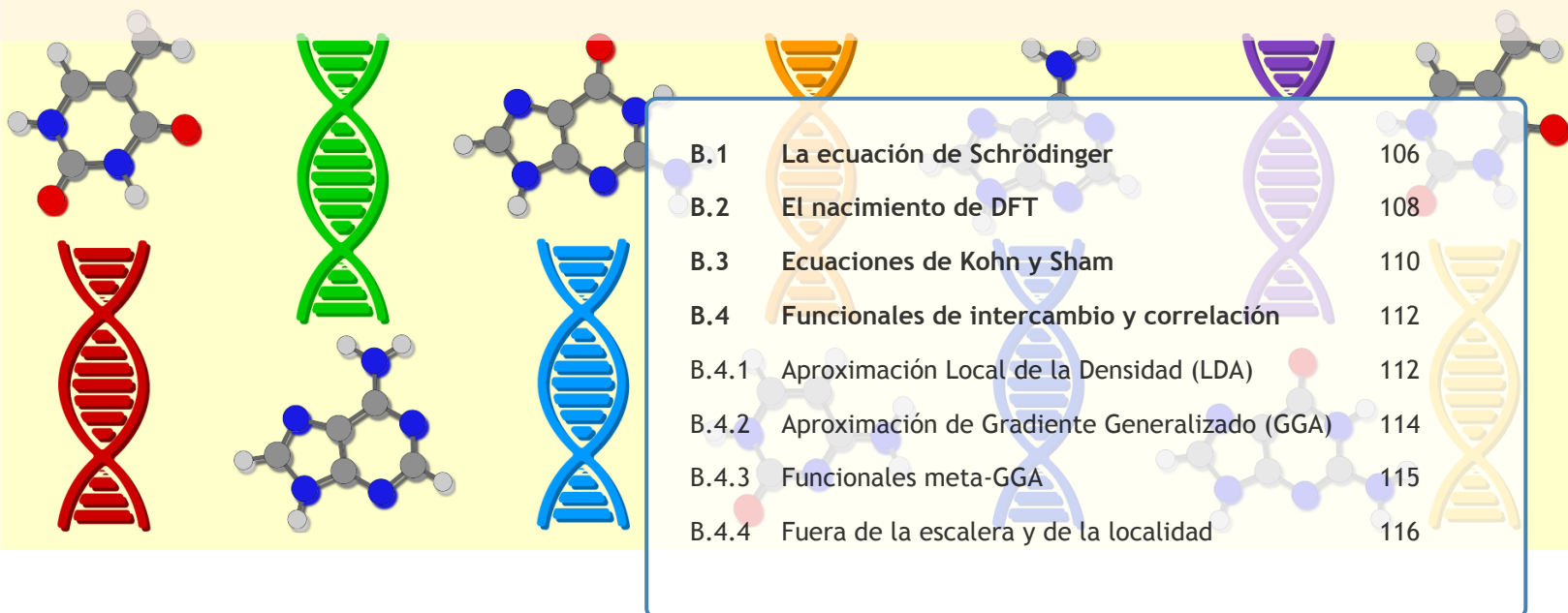
Desde su inicio, hasta finales de la década de 1990, se han tomado en cuenta las correcciones de la energía de correlación en SAPT mediante métodos de estructura electrónica de función de onda (Cúmulos Acoplados (CC, por sus siglas en inglés), por ejemplo). Sin embargo, a principios de los 2000, se formuló la metodología SAPT que incluía sólo las expresiones

de órdenes más bajos (computacionalmente menos exigentes), reemplazando los orbitales y las energías HF SCF por sus contrapartes de la Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés), y utilizando el Hamiltoniano $\mathcal{H} = \mathcal{K} + \mathcal{W}^{KS} + V$, donde $\mathcal{K} = \mathcal{K}^A + \mathcal{K}^B$ es la suma de los operadores de Kohn-Sham (KS) de los dos fragmentos y $\mathcal{W}^{KS} = \mathcal{H}^A + \mathcal{H}^B - \mathcal{K}$. Las correcciones a la energía incluidas fueron:

$$E_{int} = E_{pol}^{(10)} + E_{exch}^{(10)} + E_{ind}^{(20)} + E_{disp}^{(20)} + E_{exch-ind}^{(20)} + E_{exch-disp}^{(20)} \quad (\text{A.82})$$

La aproximación SAPT-DFT es computacionalmente menos demandante que el evaluar las expresiones regulares de SAPT (no DFT), las cuales involucran el tratamiento de orden superior de la correlación electrónica. Aunque los primeros resultados de SAPT-DFT fueron poco satisfactorios (incluso para el componente $E_{pol}^{(10)}$), más tarde se encontró que las deficiencias se debían a la incorrecta conducta asintótica de los potenciales de intercambio-correlación. Después de su corrección, la aproximación revisada recuperó con precisión los primeros tres términos de la Ecuación (A.82), a pesar de que los términos de dispersión siguieron incorrectos. Sin embargo, después de que las energías de dispersión fueron estimadas utilizando un método basado en el uso de polarizabilidades de KS acopladas, el esquema SAPT-DFT corregido se volvió extremadamente preciso.

B. Teoría de los Funcionales de la Densidad



B.1	La ecuación de Schrödinger	106
B.2	El nacimiento de DFT	108
B.3	Ecuaciones de Kohn y Sham	110
B.4	Funcionales de intercambio y correlación	112
B.4.1	Aproximación Local de la Densidad (LDA)	112
B.4.2	Aproximación de Gradiente Generalizado (GGA)	114
B.4.3	Funcionales meta-GGA	115
B.4.4	Fuera de la escalera y de la localidad	116

“La Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés) es un negocio sutil, seductor y provocativo. Su premisa básica, que todos los movimientos intrincados y correlaciones de pares en un sistema de muchos electrones están contenidos, únicamente, de alguna manera en la densidad total electrónica, es tan convincente que puede volver loco a uno.”

Axel D. Becke, 2014 [9]

Es este apéndice se mostrará un pequeño resumen de las ideas centrales de la Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés), dado que ésta es la filosofía elegida para el abordaje de los cálculos de estructura electrónica presentados en este trabajo. La información presentada en este apéndice fue desarrollada con base en la referencia [10]. Para una revisión más detallada, se recomienda consultar la referencia [11].

B.1 La ecuación de Schrödinger¹

La meta principal de diversas aproximaciones de la Química Cuántica es encontrar soluciones aproximadas a la ecuación de Schrödinger independiente del tiempo no relativista:

$$\mathcal{H}\Psi = E\Psi \quad (\text{B.1})$$

donde \mathcal{H} es el operador Hamiltoniano de un sistema conformado por M núcleos y N electrones, descritos por los vectores de posición R_A y r_i , respectivamente; Ψ corresponde a la función de onda del sistema, la cual depende de las $3N$ coordenadas espaciales y de las N coordenadas de espín de los electrones, y de las $3M$ coordenadas espaciales de los núcleos. De acuerdo a los postulados de la Mecánica Cuántica, dicha función contiene toda la información de un sistema cuántico.

En ausencia de campos eléctricos y magnéticos externos, dicho operador representa la energía total del sistema, y se puede expresada se la siguiente forma (en unidades atómicas):

$$\mathcal{H} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (\text{B.2})$$

donde A y B definen a los núcleos, i y j a los electrones. Los primeros dos términos corresponden a los operadores de energía cinética para los electrones y núcleos, respectivamente; el tercer término representa la atracción Coulómbica entre los electrones y núcleos; el cuarto la repulsión entre los electrones; y, por último, el quinto término la repulsión existente entre los núcleos.

Dado que los núcleos son más pesados que los electrones (aproximadamente, 1836 veces más pesados), se puede despreciar el movimiento nuclear. Por lo tanto, el término de energía cinética de los núcleos en la Ecuación (B.2) puede ser despreciado, mientras que la repulsión entre los núcleos es considerado como una constante. Lo anterior es conocido como aproximación de Born-Oppenheimer y permite reescribir al Hamiltoniano completo (Ecuación (B.2)) como un Hamiltoniano electrónico, el cual es un operador de energía que describe el movimiento de los electrones en un potencial generado por los núcleos fijos:

$$\mathcal{H}_{elec} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (\text{B.3})$$

Por ende, la solución a la ecuación de Schrödinger asociada al Hamiltoniano electrónico (Ecuación (B.3)) es la función de onda electrónica Ψ_{elec} :

$$\mathcal{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec} \quad (\text{B.4})$$

la cual depende explícitamente de las coordenadas electrónicas y paramétricamente de las coordenadas nucleares. La energía total para un sistema cuyos núcleos se encuentran fijos es, entonces, la suma de la energía electrónica total (E_{elec}) y la constante correspondiente a la repulsión nuclear:

$$E = E_{elec} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (\text{B.5})$$

¹El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por las referencias [10, 12, 13].

Las Ecuaciones (B.3), (B.4) y (B.5) forman parte del problema electrónico central de muchas aproximaciones. En este texto, por practicidad, se omitirá el subíndice *elec*.

Con base en los postulados de la Mecánica Cuántica, la función de onda Ψ no posee significado físico; sin embargo, el cuadrado de ésta sí:

$$|\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N \quad (\text{B.6})$$

donde $\{x_i\}$ se refiere a las coordenadas espaciales y de espín de los electrones. La Ecuación (B.6) representa la densidad de probabilidad de que los electrones $1, 2, \dots, N$ se encuentren simultáneamente en los elementos de volumen $d\vec{x}_1, d\vec{x}_2, \dots, d\vec{x}_N$.

Dado a que los electrones son indistinguibles, la densidad de probabilidad debe ser invariante al intercambio de las coordenadas de los electrones:

$$\left| \Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \vec{x}_j, \dots, \vec{x}_N) \right|^2 = \left| \Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_j, \vec{x}_i, \dots, \vec{x}_N) \right|^2 \quad (\text{B.7})$$

Lo cual puede estar asociado a dos posibles casos: las funciones de onda no cambian de signo al intercambiar coordenadas electrónicas (funciones de onda simétricas) o las funciones de onda presentan un cambio de signo cuando esto ocurre (funciones de onda antisimétricas). Puesto que los electrones son fermiones (espín semientero $1/2$), la función de onda debe ser antisimétrica.

Para resolver la ecuación de Schrödinger es necesario conocer, en primera instancia, la forma del Hamiltoniano de acuerdo al sistema que se está estudiando. Este operador dependerá del número de electrones N y del potencial de atracción núcleo(s)-electrón(es) V_{Ne} (segundo término de la Ecuación (B.3)). Una vez definido el Hamiltoniano, se debe determinar la función propia Ψ y el valor propio E asociados. Una vez que se conoce la función de onda, todas las propiedades de interés pueden ser obtenidas, aplicando los operadores correspondientes. A pesar de parecer un procedimiento sencillo, esta ecuación sólo tiene solución analítica para el caso de un núcleo interactuando con un electrón, debido a la interacción (repulsión) entre electrones (tercer término de la Ecuación (B.3)). Para cualquier otro sistema, es necesario hacer aproximaciones para encontrar una solución.

Una forma de encontrar la función de onda real asociada al estado basal (Ψ_0), es decir, al estado de mínima energía E_0 , es mediante el Principio Variacional. Con base en éste, el valor esperado obtenido de aplicar el operador Hamiltoniano \mathcal{H} (y, en general, cualquier operador asociado a un observable) a una función de prueba Ψ , será una cota superior del valor esperado de la función del estado base:

$$\langle \Psi' | \mathcal{H} | \Psi' \rangle = E' \geq E_0 = \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle \quad (\text{B.8})$$

Por lo que es necesario minimizar el valor de E' buscando a través de todas las funciones de onda aceptable para N -electrones. La búsqueda de dichas funciones de onda torna difícil la resolución de la Ecuación de Schrödinger, por lo que una alternativa es el uso de métodos que eviten la determinación de la función de onda para poder predecir las propiedades del sistema de interés, tal es el caso de la Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés).

B.2 El nacimiento de DFT²

Un funcional puede ser entendido como una función cuyo argumento es otra función, de la cual se obtiene como resultado un número. Es decir, el dominio de un funcional es una función, mientras que su rango es un número:

$$f(x) \xrightarrow{F[f(x)]} y \quad (\text{B.9})$$

Por otro lado, el valor esperado de un operador puede ser entendido como un funcional:

$$\langle \mathcal{O} \rangle = \int \dots \int \Psi^* \hat{\mathcal{O}} \Psi d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N = \langle \Psi | \hat{\mathcal{O}} | \Psi \rangle \quad (\text{B.10})$$

puesto que el valor propio del operador \mathcal{O} depende de la función Ψ . Extendiendo esto al operador Hamiltoniano \mathcal{H} , cuyo valor esperado es la energía total E del sistema, podemos entender a la energía como un funcional de la función de onda, es decir, $E[\Psi]$.

En 1964, Hohenberg y Kohn (HK) propusieron que cualquier propiedad de un sistema de N -electrones, cuyo estado basal no está degenerado, puede ser determinada a través de la densidad electrónica total (ρ). Así, la energía del estado fundamental E_0 será descrita como un funcional de la densidad electrónica: $E[\rho]$.

Para conocer cualquier propiedad de un sistema (como la densidad electrónica ρ , por ejemplo) es necesario determinar la función de onda y el Hamiltoniano asociado. Éste, a su vez, depende de la cantidad de electrones (N) y del potencial de atracción núcleo-electrón (V_{Ne}), el cual, generalmente, es conocido como potencial externo (V_{ext}):

$$V_{ext} \Rightarrow \Psi_0 \Rightarrow \rho \quad (\text{B.11})$$

Por lo tanto, para un sistema de N -electrones interactuantes, en un potencial externo V_{ext} dado, solamente existe una función de onda asociada al estado basal (Ψ_0) y una densidad electrónica ρ . De acuerdo al primer teorema de HK, si el mapeo de V_{ext} a ρ es uno a uno y reversible, entonces ρ determina inequívocamente V_{ext} , el cual tiene una única Ψ_0 y, por ende, en principio conocemos todo acerca del sistema:

$$\rho \Rightarrow V_{ext} \Rightarrow \Psi_0 \quad (\text{B.12})$$

Para demostrar lo anterior, se supone la existencia de dos potenciales externos V_{ext} y V'_{ext} que conducen a la misma densidad electrónica $\rho(\vec{r})$. Cada potencial externo es parte de dos Hamiltonianos diferentes: $\mathcal{H} = \mathcal{T} + V_{ee} + V_{ext}$ y $\mathcal{H}' = \mathcal{T} + V_{ee} + V'_{ext}$ ³. A su vez, cada Hamiltoniano pertenece a dos funciones de onda de estado basal diferente: Ψ_0 y Ψ'_0 , éstas corresponden a las energías del estado basal E_0 y E'_0 , respectivamente. Haciendo uso del Principio Variacional, y considerando a Ψ'_0 como la función de prueba para \mathcal{H} , se tiene lo siguiente:

$$\begin{aligned} E_0 &< \langle \Psi'_0 | \mathcal{H} | \Psi'_0 \rangle = \langle \Psi'_0 | \mathcal{H}' | \Psi'_0 \rangle + \langle \Psi'_0 | \mathcal{H} - \mathcal{H}' | \Psi'_0 \rangle \\ E_0 &< E'_0 + \langle \Psi'_0 | \mathcal{T} + V_{ee} + V_{ext} - \mathcal{T} - V_{ee} - V'_{ext} | \Psi'_0 \rangle \\ E_0 &< E'_0 + \int \rho(\vec{r}) (V_{ext} - V'_{ext}) d\vec{r} \end{aligned} \quad (\text{B.13})$$

²El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por la referencia [10]

³ \mathcal{T} representa al operador de energía cinética y V_{ee} la repulsión entre electrones.

De forma semejante, considerando ahora a Ψ_0 como función de prueba para \mathcal{H}' , se obtiene:

$$E'_0 < E_0 - \int \rho(\vec{r})(V_{ext} - V'_{ext}) d\vec{r} \quad (\text{B.14})$$

Sumando las Ecuaciones (B.13) y (B.14), se concluye que:

$$E_0 + E'_0 < E'_0 + E_0 \quad (\text{B.15})$$

lo cual es una contradicción, comprobando la relación uno a uno entre V_{ext} y ρ . De esta manera, ρ puede ser utilizada como una variable fundamental en el estudio de estructura electrónica y, por lo tanto, la búsqueda de la función de onda es innecesaria.

Por otro lado, como la energía total de un sistema es un funcional de la densidad electrónica, también lo serán cada uno de sus componentes. Por lo tanto, la energía total puede ser escrita como:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] \quad (\text{B.16})$$

donde el primer término representa la energía cinética, el segundo la energía potencial asociada a la repulsión entre electrones y el tercero la energía potencial debida a la atracción entre núcleos y electrones. Tanto $T[\rho]$ como $V_{ee}[\rho]$ son independientes al potencial externo, mientras que $V_{Ne}[\rho]$ depende directamente de V_{ext} .

Al no depender del potencial externo, los dos primeros términos de la Ecuación (B.16) son considerados como universales dado que aplican igual para cualquier sistema (desde átomos aislados hasta moléculas de gran tamaño). Estos términos pueden ser agrupados en la siguiente expresión:

$$F_{HK}[\rho] = T[\rho] + V_{ee}[\rho] \quad (\text{B.17})$$

la cuales es llamada como el funcional universal o el funcional de HK. A pesar de que en DFT se propone su existencia, la forma de sus componentes es desconocida. Sin embargo, de $V_{ee}[\rho]$ se conoce, al menos, la contribución clásica correspondiente a la interacción Coulómbica $J[\rho]$, por lo que el segundo término de la Ecuación (B.17) puede ser escrito como:

$$V_{ee}[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + V_{ncl}[\rho] = J[\rho] + V_{ncl}[\rho] \quad (\text{B.18})$$

donde $V_{ncl}[\rho]$ corresponde a la contribución no clásica de la interacción entre electrones. Sustituyendo la Ecuación (B.18) en la Ecuación (B.17), se tiene que el funcional universal puede ser reescrito como:

$$F_{HK}[\rho] = T[\rho] + J[\rho] + V_{ncl}[\rho] \quad (\text{B.19})$$

En la Ecuación (B.19), el primer y tercer término siguen siendo desconocidos; sin embargo, estos juegan un papel fundamental para el desarrollo de DFT. Sustituyendo la Ecuación (B.17) en la Ecuación (B.16), y considerando que:

$$V_{Ne}[\rho] = \int \rho(\vec{r})V_{ext} d\vec{r} \quad (\text{B.20})$$

se puede reescribir a la energía como funcional de la densidad de la siguiente forma:

$$E[\rho] = F_{HK}[\rho] + \int \rho(\vec{r})V_{ext} d\vec{r} \quad (\text{B.21})$$

Finalmente, para asegurar que la densidad electrónica corresponda a la densidad electrónica del estado basal y, por lo tanto, conduzca a la energía de dicho estado, HK hacen uso del Principio Variacional:

$$E[\rho'] = F_{HK}[\rho'] + \int \rho'(\vec{r}) V_{ext} d\vec{r} \geq F_{HK}[\rho] + \int \rho(\vec{r}) V_{ext} d\vec{r} = E_0 \quad (\text{B.22})$$

asegurando que la energía obtenida a partir de la Ecuación (B.21), utilizando cualquier densidad electrónica de prueba $\rho'(\vec{r})$, representará una cota superior del valor real de la energía del estado basal E_0 , siempre y cuando se cumpla que $\rho'(\vec{r}) \geq 0$ y $\int \rho'(\vec{r}) d\vec{r} = N$. Cabe destacar que E_0 solamente será obtenida cuando la densidad electrónica empleada en la Ecuación (B.21) corresponda a la densidad electrónica exacta del estado basal. Esto puede ser demostrado si se supone la existencia de una densidad de prueba $\rho'(\vec{r})$ asociada a un Hamiltoniano \mathcal{H}' y a una función de onda Ψ' . Tomando a la función de onda Ψ' como función de prueba del Hamiltoniano real \mathcal{H} , se demuestra que:

$$\langle \Psi' | \mathcal{H} | \Psi' \rangle = T[\rho'] + V_{ee}[\rho'] + \int \rho'(\vec{r}) V_{ext} d\vec{r} = E[\rho'] \geq E_0[\rho] = \langle \Psi_0 | \mathcal{H} | \Psi_0 \rangle \quad (\text{B.23})$$

B.3 Ecuaciones de Kohn y Sham⁴

A pesar de que HK proponen a la energía del estado basal, de cualquier sistema, como funcional de la densidad electrónica, los autores no proporcionan el camino para obtener E_0 ni ρ_0 sin antes encontrar la función de onda. Para atacar este problema, en 1965 Kohn y Sham (KS) consideraron el uso de un sistema de referencia de electrones no interactuantes cuyo Hamiltoniano \mathcal{H}_S (el cual en su formulación incluye un potencial efectivo $V_S(\vec{r}_i)$) tiene la siguiente forma:

$$\mathcal{H}_S = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_S(\vec{r}_i) \quad (\text{B.24})$$

Al ser un sistema no interactuante, la función de onda puede ser representada como un solo determinante de Slater:

$$\Psi_S = \frac{1}{\sqrt{N!}} |\varphi_1(\vec{x}_1) \varphi_2(\vec{x}_2) \dots \varphi_N(\vec{x}_N)\rangle \quad (\text{B.25})$$

el cual está compuesto por funciones monoeléctricas φ_i conocidas como orbitales de KS, y se obtienen al resolver la siguiente ecuación:

$$\left(-\frac{1}{2} \nabla^2 + V_S(\vec{r}) \right) \varphi_i = \varepsilon_i \varphi_i \quad (\text{B.26})$$

Para establecer una conexión entre el sistema de referencia y el sistema real, $V_S(\vec{r})$ debe ser tal que la densidad electrónica del sistema de referencia sea igual a la del sistema real con electrones interactuantes:

$$\rho_S(\vec{r}) = \sum_i^N \sum_s |\varphi(\vec{r}, s)|^2 = \rho_0(\vec{r}) \quad (\text{B.27})$$

⁴El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por la referencia [10].

Del sistema de referencia se conoce la energía cinética T_S exacta:

$$T_S = -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle \quad (\text{B.28})$$

la cual corresponde a la energía cinética de los electrones no interactuantes⁵. Al conocer la forma de T_S , KS definieron al funcional universal del sistema de referencia como:

$$F_{HK}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \quad (\text{B.29})$$

donde $E_{XC}[\rho]$ es conocida como la energía de intercambio y correlación. Igualando las Ecuaciones (B.19) y (B.29), y despejando a $E_{XC}[\rho]$, se tiene que ésta se encuentra definida como:

$$E_{XC}[\rho] = T_C[\rho] + E_{ncl}[\rho] \quad (\text{B.30})$$

donde $T_C[\rho]$ es la parte residual de la energía cinética del sistema interactuante no descrita por $T_S[\rho]$. Por lo tanto, la energía de intercambio y correlación puede ser entendida no sólo como la contribución no clásica de la energía potencial (tal como la corrección de autointeracción y la correlación de intercambio y Coulómbica), sino, también, como la parte faltante de la energía cinética no descrita por el sistema de referencia.

Al sustituir la Ecuación (B.29) en la Ecuación (B.21), y tomando en cuenta los orbitales de KS (Ecuación (B.25)), se puede reescribir la energía total del sistema interactuante como:

$$\begin{aligned} E[\rho] &= T_S[\rho] + J[\rho] + E_{XC}[\rho] + \int \rho(\vec{r}) V_{ext} d\vec{r} \\ &= T_S[\rho] + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{XC}[\rho] + \int \rho(\vec{r}) V_{ext} d\vec{r} \\ &= -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \iint |\varphi_i(\vec{r}_1)|^2 \frac{1}{r_{ij}} |\varphi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \\ &\quad + E_{XC}[\rho] - \sum_i^N \int \sum_A^M \frac{Z_A}{r_{1A}} |\varphi_i(\vec{r}_1)|^2 d\vec{r}_1 \end{aligned} \quad (\text{B.31})$$

Al aplicar el Principio Variacional en la Ecuación (B.31), es decir, al minimizar esta expresión con respecto a la densidad electrónica, se tiene que:

$$\left(-\frac{1}{2} \nabla^2 + \left[\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}} \right] \right) \varphi_i = \varepsilon_i \varphi_i \quad (\text{B.32})$$

Comparando la Ecuación (B.32) con la Ecuación (B.26), $V_S(\vec{r})$ queda definido como:

$$V_S(\vec{r}) = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}} \quad (\text{B.33})$$

Así, una vez conocidos todos los términos de la Ecuación (B.33), se podrá determinar el potencial efectivo $V_S(\vec{r})$, que al ser introducido en la Ecuación (B.32), permitirá tener conocimiento

⁵La energía cinética del sistema de referencia (electrones no interactuantes) es diferente de la energía cinética del sistema real (electrones interactuantes), es decir, $T_S \neq T$.

de los orbitales de KS y, por ende, de la densidad electrónica correspondiente al estado basal. Una vez definida ρ_0 , se podrá conocer la energía del sistema por medio de la Ecuación (B.31).

A simple vista, el método de KS resuelve el problema inicial: evitar la búsqueda de la función de onda. Sin embargo, dentro de la Ecuación (B.33) resta un término del cual no se conoce su forma debido a la dependencia de éste con la energía de intercambio y correlación:

$$V_{XC} \equiv \frac{\delta E_{XC}}{\delta \rho} \quad (\text{B.34})$$

Consecuentemente, el problema central del método propuesto por KS es la búsqueda de la forma explícita de $E_{XC}[\rho]$. Por lo tanto, para conocer la forma de esta contribución, se introducen diferentes aproximaciones para su descripción.

B.4 Funcionales de intercambio y correlación⁶

Dado que se desconoce la forma del funcional $E_{XC}[\rho]$, diversos grupos de trabajo han desarrollado una gran cantidad de métodos para encontrar aproximaciones a dicho funcional y, de tal forma, poder resolver la Ecuación (B.34). A partir del trabajo de KS se obtiene una primera aproximación para este funcional conocida como Aproximación Local de la Densidad (LDA, por sus siglas en inglés). A partir de ésta, posteriormente, se ha logrado tener un avance sistemático en el desarrollo de la Ecuación (B.34), teniendo cada vez mejores resultados, cercanos a la precisión química requerida (0.1 eV). El progreso alrededor del funcional de intercambio y correlación se puede resumir, a grandes rasgos, mediante la escalera de Jacob (Figura B.1), popularizada por J. Perdew [15]. En ésta se acomodan de manera jerárquica las aproximaciones a $E_{XC}[\rho]$: mientras más se avanza en los peldaños de la escalera, más cerca se estará de la precisión química deseada.

En este Apéndice, para fines prácticos, solamente se describirán las primeras tres aproximaciones de la escalera de Jacob. Además, se mencionarán brevemente las correcciones de largo alcance necesarias para la descripción de las interacciones no-covalentes.

B.4.1 Aproximación Local de la Densidad (LDA)

Esta primera aproximación se basa en la idea del gas uniforme de electrones (o gas homogéneo de electrones), el cual consiste de electrones que se mueven en un ambiente de carga positiva, siendo el sistema en estudio un ensamble eléctricamente neutro. Para esto se considera que tanto el número de electrones N , así como el volumen V del gas tienden a infinito ($N \rightarrow \infty$ y $V \rightarrow \infty$), mientras que la densidad electrónica ρ es finita y constante en cualquier punto ($\rho = N/V$). Con base en esto, se propone escribir a E_{XC} de la siguiente forma:

$$E_{XC}^{LDA}[\rho] = \int \varepsilon_{XC}(\rho(\vec{r})) d\vec{r} \quad (\text{B.35})$$

⁶El texto presentado en esta sección fue escrito y desarrollado con base en la información proporcionada por las referencias [9, 10, 14].

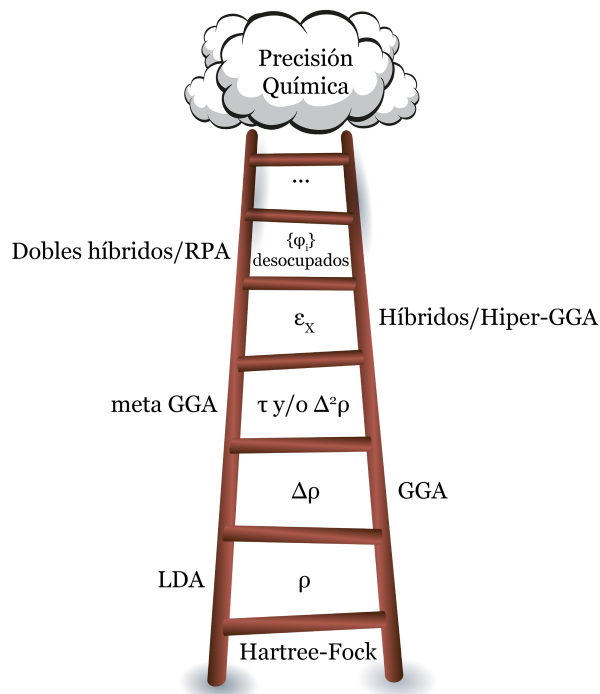


Figura B.1. Escalera de Jacob de las diferentes aproximaciones del funcional de intercambio y correlación $E_{XC}[\rho]$.

donde $\varepsilon_{XC}(\rho(\vec{r}))$ es la energía de intercambio y correlación por partícula para el gas uniforme de electrones con densidad $\rho(\vec{r})$. Esta cantidad puede ser dividida en sus contribuciones de intercambio ($\varepsilon_X(\rho(\vec{r}))$) y de correlación ($\varepsilon_C(\rho(\vec{r}))$):

$$\varepsilon_{XC}(\rho(\vec{r})) = \varepsilon_X(\rho(\vec{r})) + \varepsilon_C(\rho(\vec{r})) \quad (\text{B.36})$$

Basándose en la propuesta de Thomas [16], Fermi [17] y Dirac [18], la función de intercambio puede ser definida como:

$$\varepsilon_x = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho^{4/3} \quad (\text{B.37})$$

Mientras que, en contraste, para la energía de correlación ε_C no se cuenta con una forma explícita. Sin embargo, ε_C puede ser descrita mediante diferentes aproximaciones, tales como las simulaciones Monte Carlo [19] y la parametrización de Vosko, Wilk y Nusair [20].

Dado que el potencial externo no depende de las densidades de espín, el funcional exacto de intercambio y correlación no depende de éstas. Sin embargo, generalmente, los funcionales aproximados son expresados en su forma no restringida, es decir, estos se expresan en función de las densidades de espín $\rho_\alpha(\vec{r})$ y $\rho_\beta(\vec{r})$ ($\rho_\alpha(\vec{r}) + \rho_\beta(\vec{r}) = \rho(\vec{r})$) con la finalidad de obtener resultados de mayor precisión para sistemas con electrones desapareados. Por lo tanto, al extender la aproximación de LDA al caso no restringido, se llega a la Aproximación de Densidad de Espín Local (LSD, por sus siglas en inglés):

$$E_{XC}^{LSD}[\rho_\alpha, \rho_\beta] = \int \varepsilon_{XC}(\rho_\alpha(\vec{r}), \rho_\beta(\vec{r})) d\vec{r} \quad (\text{B.38})$$

Para el caso de sistemas de capa cerrada, se tiene que $\rho_\alpha(\vec{r}) = \rho_\beta(\vec{r}) = 1/2\rho(\vec{r})$. En contraste, para sistemas polarizados (es decir, cuando $\rho_\alpha(\vec{r}) \neq \rho_\beta(\vec{r})$), el grado de polarización es determinado mediante el siguiente factor:

$$\xi = \frac{\rho_\alpha(\vec{r}) - \rho_\beta(\vec{r})}{\rho(\vec{r})} \quad (\text{B.39})$$

Cabe destacar que LDA y LSD funcionan adecuadamente para sistemas donde la densidad varía lentamente con la posición. Sin embargo, como es de esperarse, en muchos átomos y moléculas esto no ocurre, puesto que la densidad varía rápidamente.

B.4.2 Aproximación de Gradiente Generalizado (GGA)

Una manera de improvisar al funcional local de la densidad es considerando no solamente la información de la densidad $\rho(\vec{r})$ en un punto \vec{r} particular, sino también el gradiente de la densidad $\nabla\rho(\vec{r})$, con el propósito de tomar en cuenta la inhomogeneidad de la densidad electrónica real. De tal forma, la aproximación de la densidad local puede ser interpretada como el primer elemento de la expansión de Taylor de la densidad uniforme, adicionando algunos términos más de la expansión con el fin de obtener una mejor aproximación de E_{XC} . Por lo tanto, se tiene que:

$$E_{XC}[\rho_\alpha, \rho_\beta] = E_{XC}^{LDS} + \sum_{\sigma, \sigma'} \int C_{XC}^{\sigma, \sigma'}(\rho_\alpha, \rho_\beta) \frac{\nabla\rho_\sigma}{\rho_\sigma^{2/3}} \frac{\nabla\rho_{\sigma'}}{\rho_{\sigma'}^{2/3}} + \dots \quad (\text{B.40})$$

La forma del funcional mostrada en la Ecuación (B.40) es conocida como la Aproximación de Expansión en Gradientes (GEA, por sus siglas en inglés) y resulta adecuada para sistemas donde la densidad no es uniforme pero varía lentamente. Sin embargo, la Ecuación (B.40) tiene un inconveniente: la derivada funcional del gradiente diverge para densidades que decaen exponencialmente, debido que el agujero de intercambio no cumple con las reglas de suma y no está definido negativamente⁷.

Este detalle de la aproximación GEA es corregido al hacer cumplir las restricciones válidas para los agujeros de intercambio y correlación al hacer que el agujero de intercambio en GEA sea cero en las regiones donde sea positivo. Asimismo, el funcional se trunca de tal forma que se cumpla las condiciones de normalización de los agujeros de intercambio y correlación. Teniendo en cuenta estas correcciones se llega a la Aproximación de Gradiente Generalizado (GGA, por sus siglas en inglés), la cual es representada de la siguiente manera:

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \quad (\text{B.41})$$

donde la función del integrando suele ser obtenida a partir de los parámetros teóricos y/o experimentales. Generalmente, E_{XC}^{GGA} se divide en sus contribuciones de intercambio y correlación:

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \quad (\text{B.42})$$

⁷El agujero de intercambio y correlación h_{XC} puede ser dividido en sus contribuciones de intercambio y correlación: $h_{XC}(\vec{r}_1; \vec{r}_2) = h_X^{\sigma_1=\sigma_2}(\vec{r}_1; \vec{r}_2) + h_C^{\sigma_1, \sigma_2}(\vec{r}_1; \vec{r}_2)$ y éste integra a -1. Adicionalmente, se sabe que el agujero de intercambio integra a -1 ($\int h_X(\vec{r}_1; \vec{r}_2) d\vec{r}_2 = -1$) y, por lo tanto el agujero de correlación integra a cero ($\int h_C(\vec{r}_1; \vec{r}_2) d\vec{r}_2 = 0$).

con el objetivo de aproximar cada término de forma individual.

A grandes rasgos, E_X^{GGA} puede ser expresada de la siguiente forma:

$$E_X^{GGA} = E_X^{LDA} - \sum_{\sigma} \int F(s_{\sigma}) \rho_{\sigma}^{4/3}(\vec{r}) d\vec{r} \quad (\text{B.43})$$

Para definir $F(s_{\sigma})$, es necesario explicar que s_{σ} corresponde al gradiente de la densidad reducida:

$$s_{\sigma}(\vec{r}) = \frac{|\nabla \rho_{\sigma}(\vec{r})|}{\rho_{\sigma}^{4/3}(\vec{r})} \quad (\text{B.44})$$

Esta expresión puede ser entendida como el parámetro de inhomogeneidad local, es decir, el valor de ésta varía conforme cambia la densidad electrónica en una región dada. La función $F(s_{\sigma})$ se escoge de tal forma que la expansión de gradientes cumpla con las condiciones del agujero de intercambio y correlación.

En contraste a E_X^{GGA} , E_C^{GGA} tiene formas analíticas más complicadas y ésta es obtenida a partir de datos experimentales y/o teóricos⁸.

B.4.3 Funcionales meta-GGA

Aunque la corrección por gradientes mejora los resultados obtenidos al improvisar la descripción de la inhomogeneidad de la densidad electrónica, existe un problema en dicha aproximaciones asociado a la autointeracción. Como se describió anteriormente, la energía como funcional de la densidad puede ser escrita como:

$$E[\rho(\vec{r})] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \quad (\text{B.45})$$

donde la repulsión Coulómbica clásica se expresa de la siguiente forma:

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (\text{B.46})$$

Esta contribución debe desaparecer para sistemas monoeléctricos (es decir, donde no hay correlación electrónica); sin embargo, ésta incluye un término en el cual la densidad interactúa consigo misma. Por ende, para que la Ecuación (B.45) se estime adecuadamente, se debe imponer que $J[\rho]$ sea exactamente igual a $-E_{XC}[\rho]$ para que la autointeracción sea cancelada:

$$\frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 = -E_{XC}[\rho] \quad (\text{B.47})$$

Una manera de corregir la autointeracción es mediante la adición de un término asociado a la densidad de energía cinética τ_{σ} al funcional de correlación $E_C[\rho]$:

$$\tau_{\sigma} = \sum_i |\varphi_{\sigma}|^2 \quad (\text{B.48})$$

Los funcionales que incluyen a τ_{σ} , en la estructura del funcional de correlación, son conocidos como funcionales meta-GGA. Estos describen mejor la nula correlación en sistemas de un sólo electrón, tales como el átomo de hidrógeno y los átomos hidrogenoides, por ejemplo.

⁸Para un desarrollo completo de E_C^{GGA} y ejemplos específicos de los funcionales de intercambio y correlación GGA, se recomienda consultar las referencias [9, 10].

B.4.4 Fuera de la escalera y de la localidad

La mayoría de las aproximaciones al funcional de intercambio y de correlación no son capaces de describir adecuadamente las interacciones no-covalentes a causa de que los funcionales son considerados como “locales”, puesto que gran parte de estos ocupan la información obtenida en cada punto de integración \vec{r}_1 . Los efectos no locales, como los son las interacciones intermoleculares, deben ser descritos tomando en cuenta explícitamente, en cada punto de integración, todos los otros puntos $\vec{r}_2 \neq \vec{r}_1$.

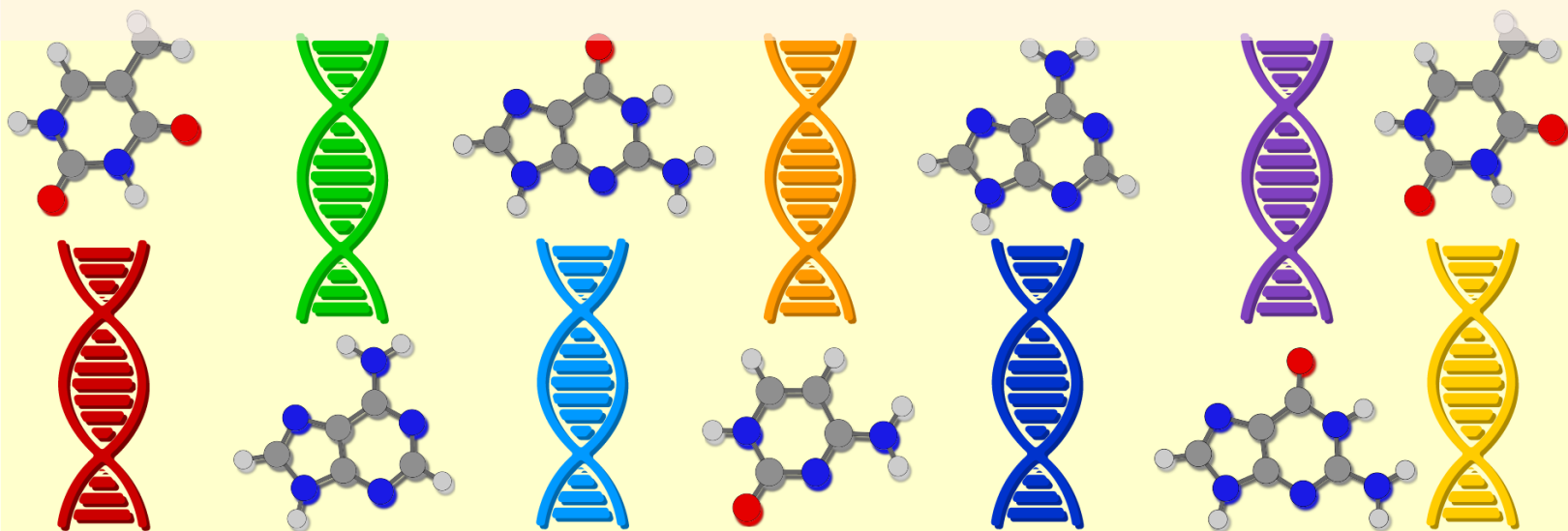
Una forma de abordar el problema de no localidad es mediante Correcciones de Largo Alcance (LCR, por sus siglas en inglés). En éstas, la interacciones interelectrónica $1/r_{12}$ es dividida en una parte de corto alcance y en una de largo alcance:

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\omega r_{12})}{r_{12}} + \frac{\text{erf}(\omega r_{12})}{r_{12}} \quad (\text{B.49})$$

Para considerar las LCR, las diferentes aproximaciones de los funcionales de la densidad pueden ser utilizadas para tratar la parte de corto alcance, mientras que el largo alcance es tratado de forma exacta mediante el intercambio de Hartree-Fock. El parámetro ω no es universal dado que depende de la molécula en estudio y no es lineal.

Incluir dentro de $E_X[\rho]$ el intercambio exacto de Hartree-Fock suele ser una buena solución para integrar los efectos no locales. Sin embargo, esto puede incrementar el costo computacional, haciendo poco práctico el uso de estas correcciones en moléculas de gran tamaño, por lo que se recurre a modificaciones o ajustes en los parámetros de los diferentes funcionales o en la adición de correcciones empíricas.

Bibliografía



- (1) Francisco, E. y Martín Pendás, A. en *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier: 2017, págs. 27-64.
- (2) Jeziorski, B.; Moszynski, R. y Szalewicz, K. *Chemical Reviews* **1994**, *94*, 1887-1930.
- (3) Kaplan, I. G., *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*; John Wiley & Sons Ltd: Engalnd, 2006, pág. 389.
- (4) Stone, A., *The Theory of Intermolecular Forces*, Second Edi; Oxford University Press: Oxford, 2013, pág. 338.
- (5) Parkinson, W. en *What's the Matter with Waves?*; 2053-2571; Morgan & Claypool Publishers: 2017, págs. 12-1 -12-17.
- (6) Simons, J. Perturbation Theory <https://chem.libretexts.org/@go/page/60529> (visitado 28-04-2021).
- (7) Stone, A. J. en *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier: 2017, págs. 3-26.
- (8) Piel, L., *Ideas of Quantum Chemistry*, 2da Edición; Elsevier: United Kingdom, 2013, pág. 1978.
- (9) Becke, A. D. *The Journal of Chemical Physics* **2014**, *140*, 18A301.
- (10) Koch, W. y Holthausen, M. C., *A Chemist's Guide to Density Functional Theory*; Wiley: 2001.
- (11) Parr, R. G. en *Horizons of Quantum Chemistry*; Springer Netherlands: Dordrecht, 1980, págs. 5-15.
- (12) Szabo, A. y Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Nueva York, 1996, pág. 466.
- (13) Levine, I. N., *Quantum Chemistry*, 7ma edición; Pearson: Nueva York, 2014, pág. 700.
- (14) Cohen, A. J.; Mori-Sánchez, P. y Yang, W. *Chemical Reviews* **2012**, *112*, 289-320.
- (15) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E. y Csonka, G. I. *The Journal of Chemical Physics* **2005**, *123*, 062201.

- (16) Thomas, L. H. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, 23, 542-548.
- (17) Fermi, E. *Rendiconti Accademia Dei Lincei* **1927**, 6, 602-607.
- (18) Dirac, P. A. M. *Mathematical Proceedings of the Cambridge Philosophical Society* **1930**, 26, 376-385.
- (19) Ceperley, D. M. y Alder, B. J. *Physical Review Letters* **1980**, 45, 566-569.
- (20) Vosko, S. H.; Wilk, L. y Nusair, M. *Canadian Journal of Physics* **1980**, 58, 1200-1211.