



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

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

**“ESTUDIO TEÓRICO DE PROPIEDADES ELECTRÓNICAS DE ADN:
ESTIMACIÓN DE LA ENERGÍA DE PUENTES DE HIDRÓGENO Y DE
ALGUNAS PROPIEDADES REACTIVAS USANDO UN MODELO DE CAJA
TATA”**

**TESIS
PARA OPTAR POR EL GRADO DE**

MAESTRO EN CIENCIAS

PRESENTA

Q.F.B. JORGE GUTIÉRREZ FLORES

Tutor:

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

México, D.F. Mayo 2016



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.

Jurado asignado

Presidente **Dr. Rodolfo Gómez Balderas**
Facultad de Estudios Superiores Cuautitlán, UNAM

Vocal **Dr. Fernando Colmenares Landín**
Facultad de Química, UNAM

Vocal **Dr. Fernando Cortés Guzmán**
Instituto de Química, UNAM

Vocal **Dra. Carmen Nina Pastor Colón**
Universidad Autónoma del Estado de Morelos

Secretario **Dra. Karina Martínez Mayorga**
Instituto de Química, UNAM

Sitio donde se desarrolló el trabajo: Instituto de Investigaciones en Materiales, UNAM

Q.F.B. Jorge Gutiérrez Flores
Sustentante

Dra. Angélica Estrella Ramos Peña
Asesora

Este trabajo ha sido presentado en los siguientes eventos:

- **El papel de los puentes de hidrógeno en la estabilidad dúplex de ADN: implicaciones en los mecanismos de daño genómico**
XIII Reunión Mexicana de Fisicoquímica Teórica
Jorge Gutiérrez, Estrella Ramos, Enrique Hernández-Lemus
UNAM Campus Morelia, Michoacán. 6 y 7 de noviembre de 2014.
Modalidad: póster.
- **El papel de los puentes de hidrógeno en la estabilidad dúplex de ADN: implicaciones en los mecanismos de daño genómico**
3er Simposio de Estudiantes IIM
Jorge Gutiérrez, Estrella Ramos, Enrique Hernández-Lemus, Carlos I. Mendoza
Instituto de Investigaciones en Materiales, Ciudad de México. 30 de julio de 2015.
Modalidad: oral.
- **Modelando al ADN: puentes de hidrógeno y reactividad**
XIV Reunión Mexicana de Fisicoquímica Teórica
Jorge Gutiérrez, Estrella Ramos, Enrique Hernández-Lemus, Carlos I. Mendoza
CUTonalá, Jalisco. 19 de noviembre de 2015.
Modalidad: póster.

*A mi familia, especialmente
a mi papá.*

*“However difficult life may seem,
there is always something you can do and succeed at.”*
Stephen Hawking

*“Once we accept our limits,
we go beyond them.”*
Albert Einstein

*“What we need is imagination.
We have to find a new view of the world.”*
Richard Feynman

Agradecimientos

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca de maestría otorgada con el número de becario 336100 y al proyecto DGAPA-PAPIIT IB101712 por el apoyo económico. A la coordinación del posgrado de Ciencias Químicas por el apoyo y oportunidad para asistir becado al “Curso de Preparación del Examen TOEFL” en la sede de San Antonio, Texas (EUA) de la UNAM.

Agradezco infinitamente a mi familia por su incondicional cariño y apoyo en todas las etapas de mi vida. Sin ellos no podría haber llegado hasta este punto. A mi papá, esté en donde esté, por ser siempre un gran ejemplo y haber estado conmigo en todo momento; por enseñarme y seguirme enseñando muchas cosas. A mi mamá por ser una gran mujer, por compartirme de su fortaleza y ayudarme siempre a salir adelante. A mi hermana por ser una gran amiga y consejera, y estar siempre conmigo apoyándome. A mi abuelita que siempre cuidó de mí. A mis tías, tíos y primos por su inmenso apoyo y cariño.

A esa persona especial por el apoyo, cariño y orientación. A mis amigos por todo su apoyo y siempre estar conmigo en las buenas y en las malas. Sin estas personas tan especiales este camino hubiera estado vacío y no lo hubiera podido disfrutar como lo hice.

A mi querida Universidad Nacional Autónoma de México, por brindarme la oportunidad de estudiar mi licenciatura en la Facultad de Química y, posteriormente, mi maestría en el Programa de Maestría y Doctorado en Ciencias Químicas. También le agradezco todas las posibilidades brindadas para mi desarrollo académico, profesional y personal.

A la Dra. Angélica Estrella Ramos Peña, mi tutora, por la orientación y por todas las oportunidades brindadas que me han permitido conocer y desarrollarme en el área de la Química Teórica. Por todo su apoyo, tanto personal como académico, y por confiar en mí y en mi trabajo.

A todos los profesores que influyeron para que hoy me encuentre donde estoy. A la Dra. Perla Castañeda (coordinadora de la licenciatura de Q.F.B. de la Facultad de Química) por orientarme cuando lo necesité y al Dr. Carlos Mendoza por interesarse, colaborar y ayudar en este proyecto; al Dr. Roberto Salcedo y al Dr. Enrique Hernández Lemus. A Evangelina Chávez, Daizú Navarrete y Leslie Arroyo por recordarme lo que es perseverar y fijarme nuevas metas.

Al Instituto de Investigaciones en Materiales, instituto en el cual he crecido y aprendido. A Joaquín Morales y Alberto López por el apoyo y paciencia en el soporte técnico computacional.

Finalmente, agradezco a los miembros del jurado: Dr. Rodolfo Gómez Balderas, Dr. Fernando Colmenares Landín, Dr. Fernando Cortés Guzmán, Dra. Carmen Nina Pastor Colón y Dra. Karina Martínez Mayorga; por el interés en este proyecto y por las revisiones realizadas, las cuales enriquecieron a este trabajo.

Índice general

1. Resumen	1
2. Introducción	3
2.1. ADN: La molécula de la vida	3
2.2. Puentes de hidrógeno	6
2.2.1. Un punto de vista teórico	6
2.3. El ADN y la reactividad química	8
2.4. Objetivos	8
3. Marco Teórico	11
3.1. Ecuación de Schrödinger	11
3.2. Agujero de intercambio y correlación	13
3.2.1. Agujero de intercambio	15
3.2.2. Agujero de correlación	15
3.3. Teoría de los Funcionales de la Densidad (DFT)	15
3.3.1. Funcionales	15
3.3.2. Hohenberg, Kohn y el nacimiento de DFT	16
3.3.3. Ecuaciones de Kohn y Sham	18
3.3.4. Aproximaciones a los Funcionales de la Densidad: Funcionales de Intercambio y Correlación	20
Aproximación Local de la Densidad (LDA)	20
Aproximación del Gradiente Generalizado (GGA)	21
Funcionales meta-GGA	22
Saliendo de la escalera: no localidad	23
3.4. Reactividad Química: un enfoque teórico	23
3.4.1. Orbitales Moleculares Frontera (FMO)	23
La Aproximación de los Orbitales Moleculares Frontera	24
Compuestos electrofílicos y nucleofílicos	25
3.4.2. Funciones de Fukui	26
Origen de las funciones de Fukui	26
4. Metodología	29
4.1. Validación	29
4.2. Detalles computacionales	31
5. Resultados y Análisis	33
5.1. Puentes de hidrógeno	33
5.1.1. Energía promedio	33
Modelo ideal no relajado	36
5.1.2. Distancias promedio	37

5.1.3. Análisis de transferencia de carga	39
5.2. Propiedades reactivas	46
5.2.1. FMO de la caja TATA	46
5.2.2. Funciones de Fukui	48
6. Conclusiones	51
A. Bases numéricas	53
A.1. Ventajas de los conjuntos de base derivados numéricamente	53
Bibliografía	55

Capítulo 1

Resumen

La estructura espacial de la molécula de ADN determina, en gran medida, los mecanismos fisiológicos y bioquímicos en los que ésta lleva a cabo sus funciones biológicas. De especial importancia para la estructura espacial y la estabilidad de la doble hélice resultan los puentes de hidrógeno que unen a ambas hebras en la cadena. Asimismo, las propiedades químicas de la cadena de ADN determinan de forma importante la selectividad de ciertas proteínas. En este proyecto se estudia la energía asociada con los puentes de hidrógeno en cadenas de diferente tamaño (de 1 a 14 pares de base), tomando en cuenta tres diferentes modelos: modelo ideal no relajado (aquel que cumple con condiciones de simetría), modelo ideal relajado (estructuras relajadas del modelo anterior) y modelo por bloques (construcción de cadenas de ADN mediante la adición de pares de base sin restricciones de simetría). Por otro lado, se realiza una primera aproximación a la selectividad de TBP hacia la caja TATA mediante el estudio de las propiedades químicas de dicha región genómica. Todo lo anterior se realiza bajo el marco de la Teoría de los Funcionales de la Densidad (DFT) usando el código DMol³ (M06-L/DN) implementado en la suite de programas Materials Studio.

Del estudio se infiere que la energía promedio de puente de hidrógeno para el modelo ideal no relajado oscila de forma periódica dependiendo del número de pares de base que tenga la cadena. Esto se relaciona a los efectos de borde existentes asociados a la simetría del sistema, permitiendo diferenciar entre cadenas con número par o impar de pares de base. Para este modelo se propone un desarrollo analítico que permite obtener la energía promedio de puente de hidrógeno para cadenas de cualquier tamaño. En cuanto al modelo ideal relajado, se encuentra que el tamaño de la cadena influye sobre la energía promedio de puente de hidrógeno existentes entre la timina y adenina, debido a que la presencia de vecinos tiende a estabilizar dicha interacción. El modelo por bloques sobreestima la energía promedio de puente de hidrógeno para la cadena de un par de base, por lo que la energía tiende a disminuir en la medida en la que aumenta el tamaño de la cadena. Para este modelo no se encuentra un efecto de estabilización debido a que los pares de base están muy separados entre sí, menospreciando las interacciones de apilamiento $\pi - \pi$.

Finalmente, a partir de los Orbitales Moleculares Frontera y de las funciones de Fukui se logra explicar, a partir de la cadena de ocho pares de base, la forma en que la proteína de unión a la caja TATA interactúa con esta región de la cadena.

Capítulo 2

Introducción

Dada su importancia biológica es innegable el interés de estudiar a la molécula de ADN. Una de las vertientes para su estudio son las interacciones causantes de la estabilidad de esta molécula y de su participación en procesos biológicos. En esta sección se darán, de forma breve, algunos antecedentes necesarios para la comprensión del trabajo. Se hablará tanto del ADN como de su importancia biológica. Se explicará qué es la caja TATA y la razón por la cual fue elegida como modelo para el desarrollo de este proyecto. También, se mencionará el papel de las fuerzas intermoleculares en la estabilización dúplex de ADN, haciendo hincapié en los puentes de hidrógeno y en las aproximaciones teóricas existentes utilizadas para su estimación. Finalmente, se plantearán la motivación del trabajo y los objetivos a cubrir.

2.1. ADN: La molécula de la vida

El ser humano siempre ha mostrado inquietud por comprender todos los hechos a su alrededor buscando modelos, teorías, leyes y todo lo que esté a su alcance para poder entender a la naturaleza y, de cierta forma, poder manipularla. Entender cómo somos y por qué es que nuestros descendientes heredan algunos de nuestros rasgos no ha sido la excepción. En 1866, Gregor Mendel (considerado como el padre de la genética) al trabajar con plantas de chícharos y observar cómo ciertas características se transmitían entre diferentes generaciones, logró deducir lo que ahora conocemos como las “Leyes de la Genética”, las cuales explicaban, de forma empírica, cómo es que la información se transfería entre diferentes generaciones. Tiempo después, en 1869, Friedrich Miescher aisló por primera vez, sin saberlo, la molécula del ácido desoxirribonucleico (ADN). Al estudiar los componentes de células linfoides, y posteriormente otras células, aisló lo que él llamó “nucleína”, que de acuerdo con sus estudios estaba formada de oxígeno, hidrógeno, nitrógeno y fósforo [1]. Estos hechos, que sin duda son de gran importancia, marcaron el inicio de una gran cantidad de investigaciones que siguen vigentes, donde el protagonista es, por supuesto, el ADN.

El interés de cómo se transmite la información genética, sus posibles modificaciones y conocer la estructura y características del ADN fue (y sigue siendo) de gran interés científico, dando como resultado muchas explicaciones, diferentes teorías y diversos premios Nobel. Fue hasta 1953, cuando James Watson y Francis Crick¹ publicaron en la revista Nature [2] la estructura de doble hélice del ADN, marcando así una nueva era para la Biología Molecular y sus diversas ramas.

Conocer la estructura del ADN es de gran importancia debido a que a partir de ésta se pudieron entender diversos mecanismos biológicos. Por ejemplo, cómo es que la información se duplicaba o cómo es que ésta se expresaba. Hasta la fecha, el ADN sigue siendo una molécula de mucho interés en diversas

¹Con ayuda de la imagen obtenida por Rosalind Franklin por medio de la técnica de difracción de rayos X (a quien por cierto, no incluyeron en el premio).

áreas de la ciencia (no solamente la Biología) y cada año, desde diferentes enfoques, se invierten considerables esfuerzos para la investigación de esta molécula.

Uno de los puntos de vista más atractivos para el estudio de esta molécula es que ésta contiene toda la información sobre un organismo, desde cómo se ve hasta cómo funciona. Se estima que cada una de las células de nuestro cuerpo tiene alrededor de 1.5 gigabytes de información genética [3]. De forma general, el ADN es transcrito en ácido ribonucleico (ARN), algunas de estas moléculas² son, posteriormente, traducidas para la producción de proteínas, las cuales tienen diferentes funciones en el cuerpo (enzimáticas y estructurales, por ejemplo) y son la forma observable de la información contenida en el ADN.

El ADN se encuentra conformado por nucleótidos, los cuales, a su vez, están formados por un azúcar (pentosa), un grupo fosfato y bases nitrogenadas: adenina (A), guanina (G), citosina (C) y timina (T) (Figura 2.1). Éstos forman la cadena de ADN por medio de enlaces fosfodiéster entre sus grupos fosfato y azúcares para la formación de una hebra, y por medio de puentes de hidrógeno e interacciones de apilamiento $\pi - \pi$ entre sus bases nitrogenadas para la formación de la cadena doble o dúplex.

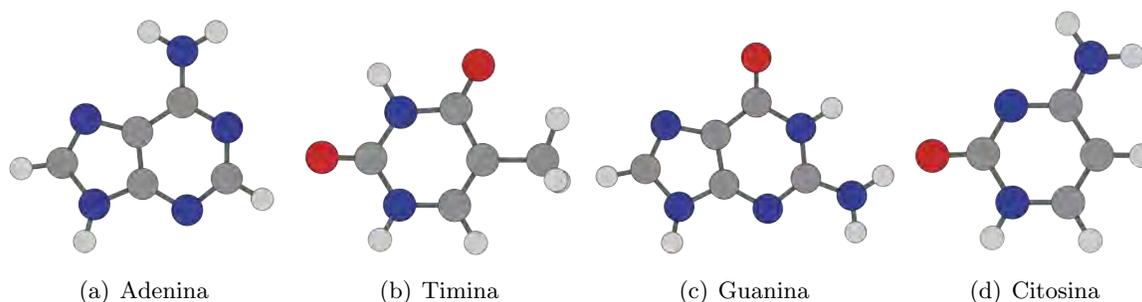


Figura 2.1: Bases nitrogenadas.

A lo largo de la cadena de ADN, existen segmentos que corresponden a zonas conocidas como genes. Un gen, de acuerdo al proyecto ENCODE [4], es la unión de secuencias genómicas que codifican un conjunto coherente de productos funcionales. Éste contiene la información de un sólo polipéptido o molécula de ARN, incluida en regiones transcritas pero no en las regiones no codificantes [5]. Esta secuencia cuenta con una región promotora, la cual regula la transcripción de la información contenida. Una región promotora conocida es la caja TATA, o caja de Goldberg-Hogness; ésta se encuentra en regiones promotoras de diversos genes eucariotes y virales [6] y está localizada a 30 pares de base río arriba (*up stream*) del sitio donde inicia la transcripción [7]. Como su nombre lo indica, la caja TATA es una cadena rica en adenina y timina (Figura 2.2) y forma parte de uno de los mecanismos más comunes para la flexión de la cadena de ADN, facilitada mediante la unión de la proteína de unión a la caja TATA (TBP, por sus siglas en inglés) alterando las torsiones que hay dentro de la cadena, permitiendo la unión de los demás factores del complejo de preinicio de la transcripción y posteriormente de la polimerasa Figura 2.3³) [5, 7].

Los principios que determinan la estructura del ADN son similares a los de las proteínas. Al igual que en las estructuras secundarias de las proteínas, muchas fuerzas intermoleculares intervienen en la estabilización de la doble hélice, tales como las interacciones dipolo-dipolo, fuerzas electrostáticas, fuerzas de van der Waals y puentes de hidrógeno, principalmente [8], siendo estos últimos los de mayor interés para este trabajo.

A pesar de que las fuerzas intermoleculares son consideradas como débiles, es sabido que muchas de las propiedades fisicoquímicas de la materia son determinadas por éstas. Estas fuerzas juegan un papel

²Existen una gran cantidad de moléculas de ARN no traducidos que cumplen con alguna función.

³Imagen tomada de la referencia [5].

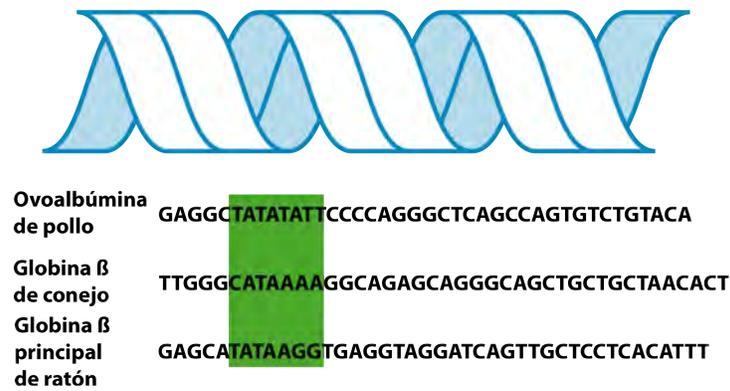


Figura 2.2: Región de ADN correspondiente a la caja TATA en diferentes organismos eucariontes [5].

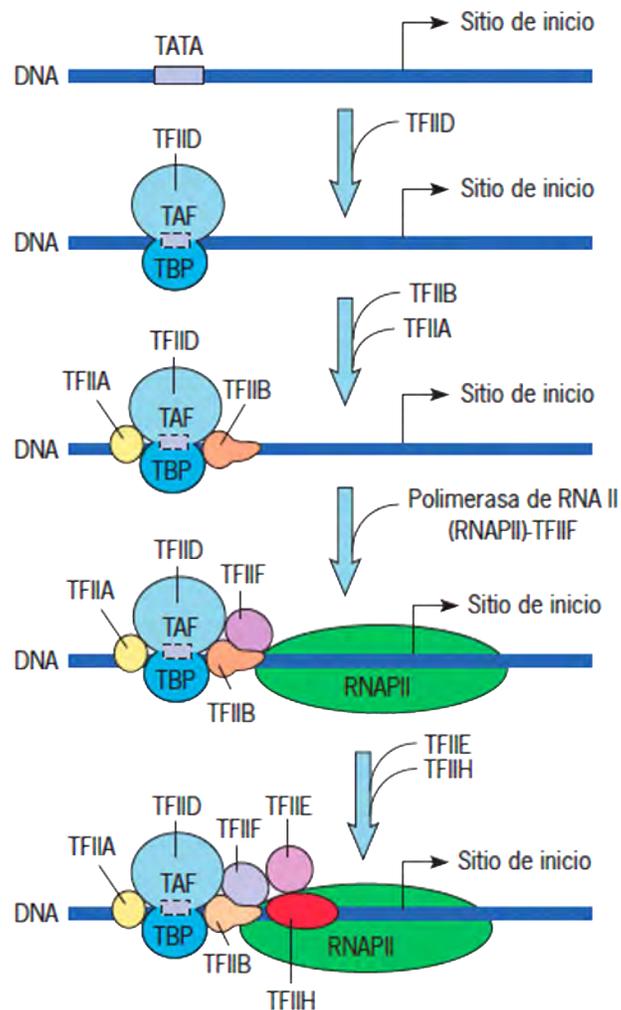


Figura 2.3: Modelo esquemático de los pasos en el ensamblaje del complejo de preinicio de la polimerasa de RNA II. El primer paso es la unión de la TBP, la cual forma parte del complejo proteínico TFIIID. Los otros componentes son los factores transcripcionales generales requeridos en el ensamble del complejo en su totalidad.

importante en la estabilidad de biomoléculas, tales como las proteínas y el ADN, siendo éste último una de las moléculas más estudiadas debido a su importancia biológica.

2.2. Puentes de hidrógeno

A pesar de que se trata de un término muy utilizado en el lenguaje científico, la definición y caracterización de los puentes de hidrógeno es un tema que hasta la fecha sigue en discusión. Para algunos autores, los puentes de hidrógeno son considerados como un caso especial de las interacciones dipolo-dipolo. Sin embargo, existe evidencia experimental que muestra que la energía asociada a estas interacciones en ocasiones es mayor que la asociada a una interacción dipolo-dipolo.

Existen diversas definiciones asociadas a la interacción de puente de hidrógeno. Una primera definición propuesta por G. N. Lewis la define de la siguiente manera: *Un átomo de hidrógeno unido al mismo tiempo por dos pares de electrones de diferentes átomos, interactuando como si hubiese un enlace no definido entre éstos* [9]. Otra forma recurrente en la que se suele describir es: *la interacción entre un átomo electronegativo (N, O y F) con un átomo de hidrógeno que se encuentra unido a otro átomo electronegativo*. Por último, una definición más “moderna” que trata de tomar aspectos teóricos y experimentales de los puentes de hidrógeno, los define como: *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* [10].

Los puentes de hidrógeno tienen un papel importante en sistemas biológicos, dado que participan en diversos procesos, tales como el reconocimiento molecular y la acción de algunos fármacos, por mencionar algunos ejemplos. Estas interacciones son de sumo interés a lo largo de la cadena de ADN, pues como se mencionó en la sección 2.1, las bases nitrogenadas que conforman a los nucleótidos interactúan entre sí por medio de puentes de hidrógeno (dos entre A y T y tres entre G y C, véase Figura 2.1); por lo que la presencia de éstos involucra estrechamente el funcionamiento y evolución del código genético. Cualquier modificación en la interacción entre las bases (ya sea por la alteración estructural o por el corrimiento de las interacciones) puede provocar errores en la lectura de la información contenida por el complejo de transcripción, obteniendo un ARN mensajero (ARNm) alterado y, por lo tanto, una proteína modificada (mutada) que pueda comprometer su función.

Este tipo de interacciones pueden ser estimadas usando como modelo a la caja TATA, esto con el objetivo de tener una descripción más realista de los puentes de hidrógeno al considerar un ambiente biológico conocido. Como veremos en la sección siguiente, muchas aproximaciones toman en cuenta únicamente la interacción entre dos bases nitrogenadas para el estudio de puentes de hidrógeno en el ADN. Usando a la caja TATA como modelo, al ser unidades repetidas de T y A, nos permitirá construir y modelar de forma relativamente sencilla cadenas de mayor tamaño.

2.2.1. Un punto de vista teórico

Debido a la importancia de los puentes de hidrógeno en la estabilización de la doble hélice del ADN, se han realizado diversos estudios teóricos con diferentes aproximaciones [11], con el objetivo de estimar la energía de esta interacción. Varios de estos estudios se realizan solamente modelando la interacción entre dos bases nitrogenadas y otros (muy pocos) utilizando, como máximo, una cadena de dos pares de bases, es decir, entre cuatro bases nitrogenadas.

Una primera aproximación es el estudio de la energía de un puente en particular en ausencia de otras interacciones [12]. Esto se logra mediante la rotación de alguna de las bases interactuantes, es decir, una de las bases rota con respecto a la otra a lo largo del eje de un puente de hidrógeno (Figura 2.4), de tal manera que las bases quedan perpendiculares entre sí. El cálculo se lleva a cabo utilizando las estructuras planas optimizadas de las bases en cuestión y fijando el ángulo entre las bases de tal forma que queden

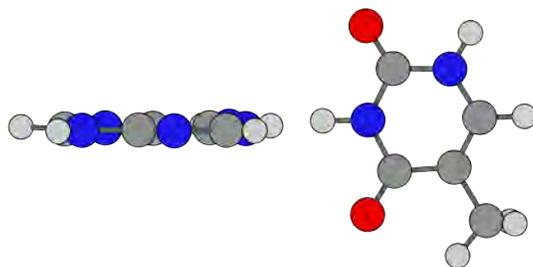


Figura 2.4: Bases perpendiculares entre sí.

perpendiculares. La suma de las energías de cada puente de hidrógeno es considerada como la energía total de interacción entre las dos bases.

Otra aproximación utilizada para el estudio energético de estas interacciones se basa en el cálculo de las constantes elásticas para todos los posibles puentes de hidrógeno [13]. Esta aproximación se basa en que a partir de estas constantes se puede llevar a cabo la descripción de la fuerza de una sola interacción. Sin embargo, al estimar las energías a partir de estas constantes, el cálculo de energía no es directo por lo que este método no proporciona explícitamente las energías asociadas a cada puente de hidrógeno.

También se ha trabajado mediante el reemplazo de átomos en las estructuras estudiadas [14]. La fuerza de la interacción $X-H \cdots Y$ es evaluada mediante el reemplazo del grupo $X-H$ por un átomo isoelectrónico (con la misma cantidad de electrones), suprimiendo al puente de hidrógeno y permitiendo estimar la fuerza de éste mediante diferencias de energías. Este reemplazo se puede entender mejor con ayuda de la Figura 2.5, donde el puente de hidrógeno entre el hidrógeno del grupo amino y el oxígeno es “eliminado” mediante la sustitución del grupo amino por un grupo hidroxilo.

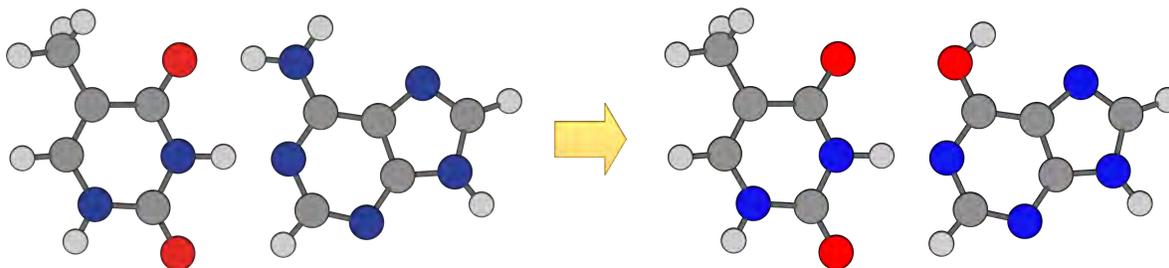


Figura 2.5: Reemplazo de un grupo amino por un grupo hidroxilo para el estudio del puente de hidrógeno.

La teoría de átomos en moléculas (AIM, por sus siglas en inglés) también ha sido utilizada para poder describir la fuerza de estas interacciones [11]. Esta teoría proporciona un método para poder obtener información sobre el comportamiento de los átomos dentro de una molécula y los enlaces e interacciones que hay entre ellos por medio del análisis de la densidad electrónica [15]; a través de un análisis topológico de la densidad y de la búsqueda de mínimos en la densidad electrónica, es posible detectar enlaces e interacciones. Mediante la comparación entre las energías de disociación calculadas y la densidad electrónica experimental (para el cálculo de puntos críticos), se logró establecer la siguiente relación $E_{HB} = 1/2V(r_{cp})$ [16] para la interacción $X-H \cdots O$, la cual indica que la energía asociada al puente de hidrógeno es proporcional a la energía potencial en el punto crítico. Esta relación (ecuación EML) ha sido utilizada para estimar la energía de puentes de hidrógeno para G-C y A-T, sin embargo, está limitada para los puentes de hidrógeno formados entre H y un átomo de O. La densidad electrónica calculada en el punto crítico, ρ_{CP} , del puente de hidrógeno también ha sido utilizada para encontrar una relación entre los

puntos críticos y la energía de disociación para poder así determinar la fuerza de los puentes de hidrógeno.

Por último, otra forma de calcular la energía de puente de hidrógeno es mediante la teoría del orbital natural de enlace (NBO, por sus siglas en inglés). Esta teoría es capaz de estudiar a los enlaces e interacciones como un fenómeno de transferencia de carga orbital [11]. Esta aproximación se basa en la hipótesis de que la interacción de puente de hidrógeno se origina debido a la transferencia de carga ocasionada por una interacción del tipo $n \rightarrow \sigma^*$ (similar a la definición establecida en la sección 2.2), es decir, involucra la débil deslocalización intermolecular de un par libre de electrones (n) de un monómero donador en la proximidad de un orbital vacío de antienlace (σ^*) del monómero aceptor; así, la deslocalización intermolecular corresponde a una transferencia parcial de carga [11].

La mayoría de los trabajos realizados al respecto solamente toman en cuenta la interacción entre dos bases nitrogenadas; algunos pocos estudios consideran la interacción de más de dos pares de base⁴. A través de un análisis topológico de la densidad electrónica de pequeñas cadenas de ADN [18] se demostró la presencia de una gran cantidad de interacciones no covalentes a lo largo de la cadena. Tomando en consideración la importancia de todas estas interacciones no covalentes, en este estudio se toma en cuenta la influencia de un mayor número de vecinos, incluyendo la presencia de grupos fosfato y azúcares, con el objetivo de estudiar de una manera más realista, el efecto asociado a la presencia de vecinos sobre la energía de puente de hidrógeno.

2.3. El ADN y la reactividad química

Como fue mencionado anteriormente, las TBP se unen selectivamente a la caja TATA. Así como este caso, existen muchas proteínas que se unen de forma selectiva a diferentes regiones del ADN para cumplir alguna función biológica. Por lo tanto conocer las propiedades químicas de la cadena permitiría, de forma indirecta, conocer las de la proteína o molécula a la cual se une.

Se tienen pocos estudios relacionados con el ADN y sus propiedades reactivas⁵, por lo que, como será detallado más adelante, otros de los objetivos que se busca en este trabajo es predecir las propiedades químicas de nuestro modelo de la caja TATA y encontrar la relación entre las propiedades reactivas y la forma en que interactúa con las TBP.

2.4. Objetivos

Como fue mencionado previamente, los modelos utilizados para los estudios reportados en la literatura cuentan con un número muy limitado de pares de base (menor a tres), por lo que la influencia de vecinos sobre los puentes de hidrógeno no se encuentra bien estudiada. De acuerdo a los estudios realizados por Fonseca et al. [17] (donde se usa como modelo una cadena de dos pares de base formada por adenina y uracilo) la distancia y energía de los puentes de hidrógeno no se ve significativamente modificada debido a la presencia de sustituyentes y otros pares de base, teniendo un panorama limitado del efecto de vecinos sobre la energía de puente de hidrógeno.

En este proyecto se busca estudiar la posible existencia de un efecto cooperativo en la energía de puente de hidrógeno en una cadena de ADN, suponiendo que dicha interacción es más fuerte que las interacciones de apilamiento $\pi - \pi$. Para este modelo se tomó en cuenta la secuencia repetitiva de adenina y timina de la caja TATA (sección 2.1), debido a que entre adenina y timina existen dos puentes de hidrógeno, siendo

⁴Véase la referencia [17]

⁵Véase la referencia [19]

un caso relativamente sencillo de estudiar.

Con el propósito de hacer una primera aproximación a la selectividad, la interacción entre las TBP y la caja TATA se estudiará usando como modelo la cadena de ocho pares de base por ser el fragmento canónico que se reporta experimentalmente para la caja TATA [20]. Consecuentemente, las propiedades químicas de la caja TATA también son de interés y éstas serán estudiadas mediante el análisis de los Orbitales Moleculares Frontera (HOMO y LUMO) y las funciones de Fukui $f^-(\vec{r})$ y $f^+(\vec{r})$ (descriptores químicos) con el propósito de tener una idea de la preferencia de la TBP hacia esta región genómica para iniciar el proceso de transcripción. Todo esto con el fundamento de que las interacciones no covalentes se deben a alteraciones en la distribución de la densidad electrónica, es decir, a la acumulación y pérdida de densidad electrónica de moléculas que interactúan entre sí.

Capítulo 3

Marco Teórico

La Química Cuántica Computacional ha avanzado de forma significativa durante las últimas décadas, gracias al desarrollo de la tecnología que ha permitido obtener dispositivos más rápidos, eficientes y económicos. Este desarrollo ha permitido que esta disciplina contribuya sustancialmente al avance de la ciencia.

La Química Cuántica Computacional está fundamentada en la versión más moderna que se tiene del átomo, sostenida en la ecuación de Schrödinger. Desgraciadamente, para dicha ecuación solamente se tiene una solución analítica para el átomo de hidrógeno por lo que se deben utilizar diferentes aproximaciones para la resolución de sistemas más complejos. A pesar de que se tienen diversos métodos que se basan en la función de onda que describen adecuadamente y con alta precisión propiedades de las moléculas, como lo son los métodos Hartree-Fock (HF) y métodos post-HF (Interacción de Configuraciones (CI, por sus siglas en inglés), Cúmulos Acoplados (CC, por sus siglas en inglés) y los métodos perturbativos, por ejemplo), éstos se vuelven cada vez más complejos y requieren mayor tiempo de cómputo (del orden de N^4 y N^5) conforme el tamaño del sistema va creciendo, haciendo impráctica su aplicación para sistemas moleculares de gran tamaño, como las biomoléculas.

Sin embargo, gracias al desarrollo de teorías “alternativas” como la Teoría de los Funcionales de la Densidad (DFT, por sus siglas en inglés) ha sido posible el cálculo de estructura electrónica y de propiedades físicas y químicas de moléculas de un número considerable de átomos. Esto debido a que, en principio, se evita el cálculo de la función de onda de N -electrones y se calcula solamente la densidad electrónica del sistema en cuestión. Dicha herramienta fue utilizada para llevar a cabo los cálculos de este proyecto, por lo que en esta sección se mencionarán las características básicas de esta teoría. Asimismo, se describirá la aproximación de los Orbitales Moleculares Frontera (FMO, por sus siglas en inglés) y las funciones de Fukui, las cuales fueron ocupadas para describir algunas propiedades químicas relevantes del sistema en estudio.

3.1. Ecuación de Schrödinger

La meta principal de diversas aproximaciones de la química cuántica es encontrar soluciones aproximadas a la ecuación de Schrödinger independiente del tiempo no relativista [21]

$$\hat{H}\Psi = E\Psi \quad (3.1)$$

donde \hat{H} es el operador Hamiltoniano de un sistema de 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, 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 que puede ser conocida de un sistema cuántico [22].

En ausencia de campos magnéticos y eléctricos y en unidades atómicas, dicho operador representa la energía total del sistema y se puede expresar de la siguiente forma [23]:

$$\hat{H} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (3.2)$$

En la ecuación anterior, A y B definen a los núcleos, mientras que 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.

Debido a que los núcleos son más pesados que los electrones¹, se puede considerar despreciable el movimiento de los núcleos con respecto al de los electrones, dado que los núcleos se mueven más lentamente. Tomando en cuenta dicha consideración, el segundo término de la ecuación (3.2) puede ser despreciado (es decir, se considera nula la energía cinética de los núcleos) y el quinto término, la repulsión entre los núcleos, es considerado una constante. Esta aproximación es conocida como la aproximación de Born-Oppenheimer [23] y permite reescribir al Hamiltoniano completo (3.2) como un Hamiltoniano electrónico, es decir, como un operador que describe el movimiento de los electrones en un campo generado por los núcleos fijos:

$$\hat{H}_{elec} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} \quad (3.3)$$

La solución a la ecuación de Schrödinger asociada al Hamiltoniano electrónico (3.3) es la función de onda electrónica Ψ_{elec} ,

$$\hat{H}_{elec} \Psi_{elec} = E_{elec} \Psi_{elec} \quad (3.4)$$

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 E_{elec} y la constante correspondiente a la repulsión nuclear.

$$E_{tot} = E_{elec} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}} \quad (3.5)$$

Este último conjunto de ecuaciones (3.3)-(3.5), constituyen el problema electrónico y es el principal interés de muchas aproximaciones. Por practicidad, en este texto se omite escribir el subíndice.

Conforme a los postulados de la mecánica cuántica, la función de onda Ψ no posee significado físico, pero el cuadrado de ésta sí [22]

$$|\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N \quad (3.6)$$

donde $\{x_i\}$ se refiere a las coordenadas espaciales y de espín de los electrones. El cuadrado de la función de onda, definido por la ecuación (3.6) representa la densidad de probabilidad de que los electrones $1, 2, \dots, N$ sean encontrados simultáneamente en los elementos de volumen $d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N$ [21]. Puesto que los electrones son indistinguibles, la densidad de probabilidad debe ser invariante al intercambio de las coordenadas de los electrones.

$$|\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \vec{x}_j, \dots, \vec{x}_N)|^2 = |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_j, \vec{x}_i, \dots, \vec{x}_N)|^2 \quad (3.7)$$

¹En el átomo de H, por ejemplo, el núcleo (protón) es aproximadamente 1800 veces más pesado el electrón.

Esto puede estar asociado a dos casos diferentes: las funciones de onda son idénticas (la función de onda es simétrica al intercambio de coordenadas) o el intercambio de coordenadas conduce a un cambio de signo (la función de onda es antisimétrica). Debido a que los electrones son fermiones con espín semientero ($1/2$), la función de onda asociada debe ser antisimétrica [21].

$$\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \vec{x}_j, \dots, \vec{x}_N) = -\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_j, \vec{x}_i, \dots, \vec{x}_N) \quad (3.8)$$

Para resolver la ecuación de Schrödinger, es necesario determinar en primera instancia la forma del Hamiltoniano de acuerdo al sistema que se está estudiando, el cual depende del número de electrones N y del potencial de atracción núcleo(s)-electrón(es), V_{Ne} (segundo término de la ecuación (3.3)). Una vez definido, se debe determinar la función propia Ψ y el valor propio E asociados al Hamiltoniano en cuestión. Una vez que se conoce la función de onda, todas las propiedades de interés pueden ser obtenidas aplicando los operadores correspondientes [21]. Sin embargo, difícilmente se logra encontrar la solución exacta para la ecuación de Schrödinger.

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

$$\langle \Psi' | \hat{H} | \Psi' \rangle = E' \geq E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \quad (3.9)$$

Por lo que es necesario minimizar el valor de E' buscando a través de todas las funciones de onda de N -electrones aceptables [21]. 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 como es el caso de la Teoría de los Funcionales de la Densidad.

3.2. Agujero de intercambio y correlación²

La densidad electrónica $\rho(\vec{r})$ es definida como la integral múltiple sobre todas las coordenadas de espín de todos los electrones y sobre todas menos una de las variables espaciales:

$$\rho(\vec{r}) = N \int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 ds_1 d\vec{x}_2 \dots d\vec{x}_N \quad (3.10)$$

$\rho(\vec{r})$ determina la probabilidad de encontrar cualquiera de los N electrones, de cualquier espín, dentro un elemento de volumen $d\vec{r}_1$ dado, mientras que el resto de los $N - 1$ electrones tiene posiciones y espines arbitrarios. Esta función se caracteriza por ser positiva para sus tres coordenadas espaciales, decrecer conforme tiende a infinito e integrar al número total de electrones.

$$\rho(\vec{r} \rightarrow \infty) = 0 \quad (3.11)$$

$$\int \rho(\vec{r}) d\vec{r}_1 = N \quad (3.12)$$

Cuando se tiene interés en conocer la probabilidad de encontrar dos electrones con el mismo espín de forma simultánea cuando el resto de los $N - 2$ electrones cuentan con posiciones y espines arbitrarios se recurre a la densidad por pares, la cual se encuentra definida por:

$$\rho_2(\vec{x}_1, \vec{x}_2) = N(N - 1) \int \dots \int |\Psi(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N)|^2 d\vec{x}_3 \dots d\vec{x}_N \quad (3.13)$$

²El desarrollo de la Teoría de los Funcionales de la Densidad puede ser presentado en términos de los agujeros de intercambio y correlación. A pesar de que se presentará un tratamiento diferente, es necesario definir dichos términos debido a que más adelante serán mencionados. Esta sección fue desarrollada con base en la información brindada por la referencia [21].

Esta cantidad es de importancia debido a que nos da información relevante sobre la correlación electrónica. Al igual que la densidad electrónica, la densidad de pares está definida positivamente. Ésta es simétrica en las coordenadas y normalizada conforme al número de pares no distinguidos, $N(N - 1)$.

Tomando en cuenta la influencia de la correlación de intercambio (debida a la antisimetría de la función de onda) y la correlación de Coulomb (debida a la repulsión entre los electrones) sobre la densidad de pares, ésta puede ser dividida en dos partes, el producto de densidades independientes y el resto:

$$\rho_2(\vec{x}_1, \vec{x}_2) = \rho(\vec{r}_1)\rho(\vec{r}_2) [1 + f(\vec{x}_1; \vec{x}_2)] \quad (3.14)$$

donde $f(\vec{x}_1; \vec{x}_2)$ es conocido como el factor de correlación. Para casos donde no exista correlación $f(\vec{x}_1; \vec{x}_2) = 0$. Sin embargo, en este caso $\rho_2(\vec{x}_1, \vec{x}_2)$ se encuentra normalizado para un número incorrecto de pares, debido a que $\iint \rho(\vec{r}_1)\rho(\vec{r}_2)d\vec{r}_1d\vec{r}_2 = N^2$, lo cual es consecuencia de la autointeracción.

Para evitar lo anterior, se recurre a la probabilidad condicional $\Omega(\vec{x}_1; \vec{x}_2)$. Ésta corresponde a la probabilidad de encontrar cualquier electrón en la posición 2 cuando un electrón es “fijado” en la posición 1.

$$\Omega(\vec{x}_1; \vec{x}_2) = \frac{\rho_2(\vec{x}_1, \vec{x}_2)}{\rho(\vec{x}_1)} \quad (3.15)$$

Tomando en cuenta las propiedades de la densidad electrónica y de la densidad de pares, tenemos que la densidad condicional integra a $N - 1$ electrones, es decir, contiene a todos los electrones menos al electrón de referencia:

$$\int \Omega(\vec{x}_1; \vec{x}_2)d\vec{x}_2 = N - 1 \quad (3.16)$$

La diferencia entre la densidad condicional y la probabilidad no correlacionada de encontrar al electrón en \vec{x}_2 describe el cambio en la probabilidad condicional debido a los efectos de la correlación de intercambio y de Coulomb, tomando en cuenta la corrección de la autointeracción.

$$h_{XC}(\vec{x}_1; \vec{x}_2) = \frac{\rho_2(\vec{x}_1, \vec{x}_2)}{\rho(\vec{x}_1)} - \rho(\vec{x}_2) = \rho(\vec{x}_2)f(\vec{x}_1; \vec{x}_2) \quad (3.17)$$

Debido a que la probabilidad condicional conduce a una disminución de la densidad electrónica en \vec{x}_2 (puesto que toma en cuenta los efectos de correlación) en comparación con el caso independiente, esta cantidad es conocida como el agujero de intercambio y correlación y, generalmente, es negativa (particularmente cerca del electrón de referencia). En otras palabras, el agujero de intercambio y correlación nos indica cómo disminuye la probabilidad de encontrar otro electrón cerca del electrón de referencia.

Tomando en cuenta las ecuaciones (3.12) y (3.16), tenemos que la integral de la ecuación (3.17) corresponde a la carga de un electrón.

$$\int h_{XC}(\vec{x}_1; \vec{x}_2) = -1 \quad (3.18)$$

Partiendo a la ecuación (3.17) en sus contribuciones de intercambio y correlación tenemos:

$$h_{XC}(\vec{x}_1; \vec{x}_2) = h_X^{\sigma_1=\sigma_2}(\vec{x}_1; \vec{x}_2) + h_C^{\sigma_1, \sigma_2}(\vec{x}_1; \vec{x}_2) \quad (3.19)$$

donde el primer término corresponde al agujero de intercambio, mientras que el segundo corresponde al agujero de correlación.

3.2.1. Agujero de intercambio

El agujero de intercambio surge como consecuencia de la antisimetría de la función de onda. Al igual que el agujero de intercambio y correlación completo, éste integra a -1 .

$$h_X(\vec{r}_1; \vec{r}_2) d\vec{r}_2 = -1 \quad (3.20)$$

Puesto que el Principio de Exclusión de Pauli establece que dos electrones del mismo espín no pueden ocupar la misma posición en el espacio, el agujero de intercambio tiene que ser igual al negativo de la densidad electrónica cuando $r_1 \rightarrow r_2$.

$$h_X(\vec{r}_2 \rightarrow \vec{r}_1; \vec{r}_1) d\vec{r}_2 = -\rho(\vec{r}_1) \quad (3.21)$$

Otra propiedad importante es que h_X es negativo en cualquier parte, dado que la probabilidad de encontrar un electrón del mismo espín que el de referencia disminuye cuando más cerca se esté de éste último.

$$h_X(\vec{r}_1; \vec{r}_2) < 0 \quad (3.22)$$

Por último, si tomamos la definición de la ecuación (3.17) y se modifica para el caso del intercambio, se tiene:

$$h_X(\vec{r}_1; \vec{r}_2) = \rho(\vec{r}_2) f_X(\vec{r}_1; \vec{r}_2) \quad (3.23)$$

Notando que la forma del agujero de intercambio no depende solamente del factor de la correlación de Fermi ($f_X(\vec{r}_1; \vec{r}_2)$) sino también de la densidad en \vec{r}_2 , teniendo como consecuencia que el agujero de intercambio no es esféricamente simétrico.

3.2.2. Agujero de correlación

El agujero de correlación proviene del término de repulsión electrostática $1/r_{12}$ y, observando las condiciones establecidas por las ecuaciones (3.18) y (3.20), éste debe ser normalizado a cero, es decir, la integral del agujero de correlación sobre todo el espacio no contiene carga:

$$\int h_C(\vec{r}_1; \vec{r}_2) d\vec{r}_2 = 0 \quad (3.24)$$

El agujero de correlación será negativo y grande conforme más cerca se esté de la posición del electrón de referencia, puesto que el término de repulsión electrostática provoca que los electrones se mantengan separados. Dado que existe una posibilidad finita de que electrones de espín diferente puedan ser encontrados en un mismo elemento de volumen, cuando $\vec{r}_2 \rightarrow \vec{r}_1$ el agujero de correlación asemejará la condición de cúspide³ para electrones de espines antiparalelos. Por último, es importante mencionar que como h_C se encuentra normalizado a cero, existen regiones donde éste es positivo. En resumen, para el agujero de correlación, la densidad disminuye en zonas cercanas al electrón de referencia y se acumula en regiones lejanas a éste.

3.3. Teoría de los Funcionales de la Densidad (DFT)

3.3.1. Funcionales

Antes de profundizar en la Teoría de los Funcionales de la Densidad, es conveniente definir qué es un funcional. Un funcional puede ser entendido como una función cuyo argumento es una función y como

³Condición en la cual el máximo de la densidad electrónica corresponde a posiciones cercanas a los núcleos, es decir, conforme más cerca se está del núcleo más será la probabilidad de encontrar a un electrón.

resultado se obtiene un número. En otras palabras, el dominio de un funcional es una función mientras que su rango es un número.

$$f(x) \xrightarrow{F[f(x)]} y \quad (3.25)$$

El valor esperado de un operador puede ser entendido como un funcional:

$$\langle \hat{O} \rangle = \int \cdots \int \Psi^* \hat{O} \Psi d\vec{x}_1 d\vec{x}_2 \dots d\vec{x}_N = \langle \Psi | \hat{O} | \Psi \rangle \quad (3.26)$$

puesto que el valor del operador \hat{O} depende de la función Ψ . Extendiendo lo dicho anteriormente al operador Hamiltoniano \hat{H} , cuyo valor esperado es la energía del sistema E (ecuación (3.1)), podemos entender a la energía como un funcional de la función de onda, es decir, $E[\Psi]$.

3.3.2. Hohenberg, Kohn y el nacimiento de DFT⁴

En 1964, Hohenberg y Kohn [24] propusieron que cualquier propiedad (como la energía o la función de onda, por ejemplo) de un sistema de N -electrones con estados basales no degenerados podría ser determinada a través de la densidad electrónica total (ρ). De tal forma que la energía del estado fundamental E_0 puede ser descrita como un funcional de la densidad electrónica: $E[\rho]$.

De acuerdo a la sección 3.1, para conocer cualquier propiedad de un sistema (como ρ) es necesario determinar la función de onda y el Hamiltoniano asociado al sistema de interés. Éste, a su vez, depende de la cantidad de electrones N y del potencial de atracción núcleo-electrón V_{Ne} (ecuación (3.3)), este último conocido, generalmente, como potencial externo V_{ext} . Lo anterior puede ser resumido en la siguiente relación:

$$V_{ext} \Rightarrow \Psi_0 \Rightarrow \rho \quad (3.27)$$

Por lo tanto, para un sistema de N -electrones interactuantes en un potencial externo V_{ext} existe solamente una función de onda asociada al estado basal Ψ_0 y una densidad electrónica ρ . De acuerdo al primer teorema de Hohenberg y Kohn, 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 lo tanto, en principio, conocemos todo acerca del sistema.

$$\rho \Rightarrow V_{ext} \Rightarrow \Psi_0 \quad (3.28)$$

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

$$\begin{aligned} E_0 &< \langle \Psi' | \hat{H} | \Psi' \rangle = \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \\ E_0 &< E'_0 + \langle \Psi' | \hat{T} + \hat{V}_{ee} + \hat{V}_{ext} - \hat{T} - \hat{V}_{ee} - \hat{V}'_{ext} | \Psi' \rangle \\ E_0 &< E'_0 + \int \rho(\vec{r}) (V_{ext} - V'_{ext}) d\vec{r} \end{aligned} \quad (3.29)$$

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

$$E'_0 < E_0 - \int \rho(\vec{r}) (V_{ext} - V'_{ext}) d\vec{r} \quad (3.30)$$

⁴Esta sección fue desarrollada con base en la información brindada por la referencia [21].

Sumando la ecuación (3.29) y (3.30), se tiene:

$$E_0 + E'_0 < E'_0 + E_0 \quad (3.31)$$

lo cual es una contradicción. Comprobando la relación uno a uno entre el V_{ext} y ρ . De esta manera, ρ puede ser utilizada como variable fundamental en el estudio de estructura electrónica y, por lo tanto, la búsqueda de la función de onda resulta innecesaria [25].

Ahora, como la energía total de un sistema es un funcional de la densidad electrónica también lo son sus componentes, por lo que la energía total puede ser escrita como:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] \quad (3.32)$$

donde el primer término representa la energía cinética, el segundo la energía potencial debida a la repulsión entre electrones y el tercero la energía potencial debida a la atracción núcleos-electrones. Simultáneamente, los primeros dos términos son independientes del potencial externo, mientras que el último sí.

Al no depender del potencial externo, los primeros dos términos de la ecuación (3.32) son considerados como universales, puesto que aplican igual tanto para una molécula de hidrógeno como para una molécula de mucho mayor tamaño (como el ADN). Estos términos pueden ser agrupados de la siguiente forma:

$$F_{HK}[\rho] = T[\rho] + V_{ee}[\rho] \quad (3.33)$$

Esta expresión es conocida como el funcional de Hohenberg y Kohn o funcional universal y a pesar de que se conoce su existencia, la forma de sus componentes es desconocida. Sin embargo, del último término se conoce al menos la parte clásica correspondiente a la interacción Coulómbica $J[\rho]$, por lo que puede ser escrito como:

$$V_{ee}[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + V_{ncl}[\rho] = J[\rho] + V_{ncl}[\rho] \quad (3.34)$$

donde $V_{ncl}[\rho]$ corresponde a la parte no clásica de la interacción entre electrones. Tomando en cuenta la ecuación (3.34), podemos reescribir al funcional universal como:

$$F_{HK}[\rho] = T[\rho] + J[\rho] + V_{ncl}[\rho] \quad (3.35)$$

donde el primer y tercer término siguen siendo desconocidos y juegan un papel importante en el desarrollo de DFT.

Sustituyendo la ecuación (3.33) en la ecuación (3.32) y considerando que:

$$V_{Ne}[\rho] = \int \rho(\vec{r})V_{ext}d\vec{r} \quad (3.36)$$

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

$$E[\rho] = F_{HK}[\rho] + \int \rho(\vec{r})V_{ext}d\vec{r} \quad (3.37)$$

Por último, para asegurar que cierta densidad electrónica corresponda a la densidad electrónica del estado basal y, por lo tanto, conduzca a la energía del correspondiente, Hohenberg y Kohn hacen uso del Principio Variacional,

$$E[\rho'] = F_{HK}[\rho'] + \int \rho'(\vec{r})V_{ext}d\vec{r} \geq F_{HK}[\rho] + \int \rho(\vec{r})V_{ext}d\vec{r} = E_0 \quad (3.38)$$

indicando que para cualquier densidad de prueba $\rho'(\vec{r})$, que cumpla con $\rho'(\vec{r}) \geq 0$ y $\int \rho'(\vec{r}) d\vec{r} = N$, asociada a un potencial externo V'_{ext} , la energía obtenida del funcional dado por la ecuación (3.37) representa una cota superior del valor real de la energía del estado basal E_0 . E_0 es obtenida sí y sólo sí la densidad electrónica corresponde a la densidad exacta del estado basal.

Lo anterior puede ser demostrado si se supone una densidad de prueba $\rho'(\vec{r})$, la cual está asociada a un Hamiltoniano \hat{H}' y a una función de onda Ψ' . Tomando a la función de onda Ψ' como función de prueba del Hamiltoniano real \hat{H} , se tiene:

$$\langle \Psi' | \hat{H} | \Psi' \rangle = T[\rho'] + V_{ee}[\rho'] + \int \rho'(\vec{r}) V_{ext} d\vec{r} = E[\rho'] \geq E_0[\rho] = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \quad (3.39)$$

llegando al resultado deseado.

3.3.3. Ecuaciones de Kohn y Sham⁵

De acuerdo a la sección anterior, Hohenberg y Kohn proponen a la energía del estado basal como funcional de la densidad electrónica, sin embargo, los teoremas propuestos no indican cómo obtener E_0 ni ρ_0 sin antes encontrar la función de onda del sistema.

Para abordar este problema, en 1965 Kohn y Sham [26] consideraron el uso de un sistema de referencia de electrones no interactuantes cuyo Hamiltoniano, el cual contiene un potencial efectivo $V_S(\vec{r}_i)$, tiene la siguiente forma:

$$\hat{H}_S = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_S(\vec{r}_i) \quad (3.40)$$

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

$$\Psi_S = \frac{1}{\sqrt{N!}} |\varphi_1(\vec{x}_1) \varphi_2(\vec{x}_2) \dots \varphi_N(\vec{x}_N)\rangle \quad (3.41)$$

el cual está compuesto por funciones monoeléctricas φ_i conocidas como los orbitales de Kohn-Sham (orbitales KS) y se obtienen al resolver:

$$\left(-\frac{1}{2} \nabla^2 + V_S(\vec{r}) \right) \varphi_i = \varepsilon_i \varphi_i \quad (3.42)$$

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.

$$\rho_S(\vec{r}) = \sum_i^N \sum_s |\varphi(\vec{r}, s)|^2 = \rho_0(\vec{r}) \quad (3.43)$$

Del sistema de referencia se conoce la energía cinética exacta,

$$T_S = -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle \quad (3.44)$$

la cual corresponde a la energía cinética de electrones no interactuantes⁶. Al conocer la forma de T_S , Kohn y Sham definieron al funcional universal como:

$$F_{HK}[\rho] = T_S[\rho] + J[\rho] + E_{XC}[\rho] \quad (3.45)$$

⁵Esta sección fue desarrollada con base en la información brindada por la referencia [21].

⁶La energía cinética de un sistema de electrones interactuantes es diferente a la energía cinética de un sistema de electrones que interactúan, es decir, $T_S \neq T$.

donde $E_{XC}[\rho]$ es conocida como la energía de intercambio y correlación. Igualando la ecuación (3.35) y la ecuación (3.45), y despejando a $E_{XC}[\rho]$, obtenemos que ésta se encuentra definida como:

$$E_{XC}[\rho] = T_C[\rho] + E_{ncl}[\rho] \quad (3.46)$$

donde $T_C[\rho]$ es la parte residual de la energía cinética real no descrita por $T_S[\rho]$. Por lo tanto, se puede entender a la energía de intercambio y correlación no sólo como la contribución no clásica de la energía potencial (como lo es la corrección de autointeracción y la correlación de intercambio y Coulombica), sino también como la parte faltante de la energía cinética no descrita por el sistema de referencia.

Ahora, sustituyendo la ecuación (3.45) en la ecuación (3.37) y tomando en cuenta los orbitales de KS, se puede reescribir a la energía total del sistema interactuante como:

$$\begin{aligned} E[\rho] &= T_S[\rho] + J[\rho] + E_{XC}[\rho] + \int \rho(\vec{r})V_{ext}d\vec{r} \\ &= T_S[\rho] + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 + E_{XC}[\rho] + \int \rho(\vec{r})V_{ext}d\vec{r} \\ &= -\frac{1}{2} \sum_i^N \langle \varphi_i | \nabla^2 | \varphi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \iint |\varphi_i(\vec{r}_1)|^2 \frac{1}{r_{ij}} |\varphi_j(\vec{r}_2)|^2 d\vec{r}_1 d\vec{r}_2 \\ &\quad + E_{XC}[\rho] - \sum_i^N \int \sum_A^M \frac{Z_A}{r_{1A}} |\varphi_i(\vec{r}_1)|^2 d\vec{r}_1 \end{aligned} \quad (3.47)$$

Aplicando el Principio Variacional, es decir, minimizando la ecuación (3.47) con respecto a la densidad electrónica, se tiene:

$$\left(-\frac{1}{2} \nabla^2 + \left[\int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}} \right] \right) \varphi_i = \varepsilon_i \varphi_i \quad (3.48)$$

Comparando la ecuación anterior con la ecuación (3.42), podemos definir a $V_S(\vec{r})$ como:

$$V_S(\vec{r}) = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A^M \frac{Z_A}{r_{1A}} \quad (3.49)$$

Así, una vez que se conocen todos los términos de la ecuación (3.49), se podrá determinar el potencial efectivo $V_S(\vec{r})$, que al ser introducido en la ecuación monoeléctronica (3.48) permitirá tener conocimiento de los orbitales de KS y, por lo tanto, de la densidad electrónica del estado basal. Una vez definida ρ_0 se podrá conocer la energía del sistema vía la ecuación (3.47).

A simple vista, el método de Kohn y Sham resuelve el problema inicial. No obstante, dentro de la ecuación (3.49) queda un término del cual no se conoce su forma, debido a su dependencia con la energía de intercambio y correlación E_{XC} :

$$V_{XC} \equiv \frac{\delta E_{XC}}{\delta \rho} \quad (3.50)$$

Por consiguiente, el problema del método propuesto por Kohn y Sham es que no se conoce la forma explícita de $E_{XC}[\rho]$, por lo que se introducen diferentes aproximaciones para su solución.

3.3.4. Aproximaciones a los Funcionales de la Densidad: Funcionales de Intercambio y Correlación⁷

Debido a que se desconoce la forma del funcional $E_{XC}[\rho]$, se han desarrollado diversos métodos para encontrar aproximaciones y, de esta manera poder resolver la ecuación (3.50) [28]. A partir del trabajo realizado por Kohn y Sham, se obtiene una primera aproximación para dicho funcional conocida como Aproximación Local de la Densidad (LDA, por sus siglas en inglés) [27], de la cual se hablará más adelante. A partir de esta aproximación, se ha logrado tener un avance sistemático en el desarrollo de dicho funcional, obteniendo cada vez mejores resultados cercanos a la precisión química. Dicho avance se puede resumir, a grandes rasgos, mediante la escalera de Jacob [29] (popularizada por Perdew) en la cual se acomodan de manera jerárquica los avances realizados en las aproximaciones de los funcionales de la densidad a la energía de intercambio y correlación. Mientras más se avanza en los peldaños de la escalera, más cerca se está de la precisión química (error menor a 0.1 eV) (Figura 3.1).

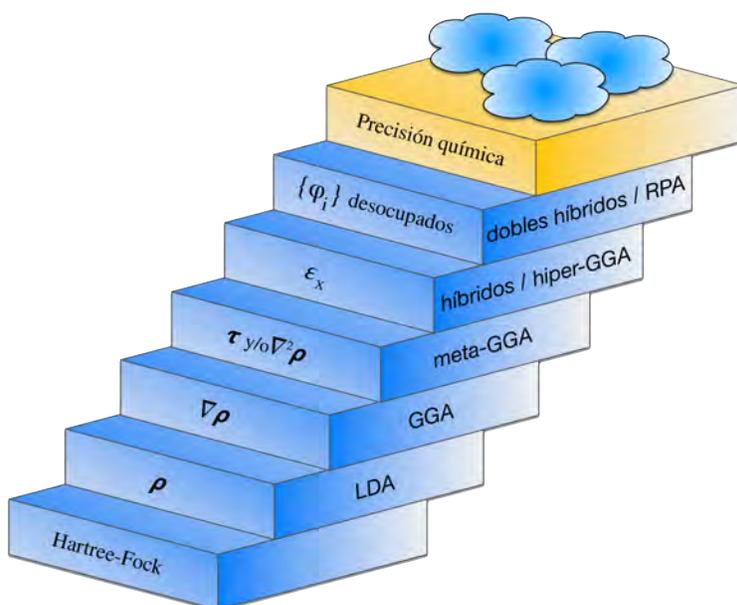


Figura 3.1: Escalera de Jacob de las aproximaciones de los funcionales de la densidad a la energía de intercambio y correlación.

Para fines de este trabajo, solamente se describirán brevemente las primeras tres aproximaciones de la escalera de Jacob. Adicionalmente, se mencionarán las correcciones a largo alcance necesarias para la descripción de interacciones débiles.

Aproximación Local de la Densidad (LDA)

Esta aproximación se basa en la idea del gas uniforme de electrones⁸, sistema en donde los electrones se mueven en un ambiente de carga positiva resultando en un ensamble eléctricamente neutro. 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 es finita ($N/V = \rho$) y es constante en cualquier punto. Con base en este modelo,

⁷Las secciones correspondientes a las aproximaciones fueron desarrolladas con base en la información proporcionada por las referencias [21], [25] y [27].

⁸También es conocido como gas homogéneo de electrones.

se propone escribir a E_{XC} de la siguiente forma:

$$E_{XC}^{LDA}[\rho] = \int \varepsilon_{XC}(\rho(\vec{r})) d\vec{r} \quad (3.51)$$

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

$$\varepsilon_{XC}(\rho(\vec{r})) = \varepsilon_X(\rho(\vec{r})) + \varepsilon_C(\rho(\vec{r})) \quad (3.52)$$

Utilizando los resultados de Thomas, Fermi y Dirac, la función de intercambio puede ser definida como:

$$\varepsilon_x = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho^{4/3} \quad (3.53)$$

Mientras que para la contribución de correlación ε_C no se cuenta con una forma explícita. Sin embargo, ε_C puede ser descrita mediante diferentes aproximaciones (por ejemplo, simulaciones Monte Carlo [30] y la parametrización de Vosko, Wilk y Nusair [31]).

A pesar de que el funcional exacto de intercambio y correlación no depende de las densidades de espín (debido a que el potencial externo no depende de las densidades de espín), generalmente, los funcionales aproximados son expresados en su forma no restringida, es decir, éstos se expresan en función de las densidades de espín, $\rho_\alpha(\vec{r})$ y $\rho_\beta(\vec{r})$ ($\rho_\alpha(\vec{r}) + \rho_\beta(\vec{r}) = \rho(\vec{r})$). Esta forma no restringida de los funcionales conduce a resultados de mayor precisión para sistemas de capa abierta (sistemas con electrones desapareados). Extendiendo LDA al caso no restringido, obtenemos la Aproximación de Densidad de Espín Local (LSD, por sus siglas en inglés):

$$E_{XC}^{LSD}[\rho_\alpha, \rho_\beta] = \int \varepsilon_{XC}(\rho_\alpha(\vec{r}), \rho_\beta(\vec{r})) d\vec{r} \quad (3.54)$$

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

$$\xi = \frac{\rho_\alpha(\vec{r}) - \rho_\beta(\vec{r})}{\rho(\vec{r})} \quad (3.55)$$

Esta aproximación funciona 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 pues la densidad varía rápidamente en estos sistemas.

Aproximación del Gradiente Generalizado (GGA)

Una forma para corregir al funcional local fue considerar no solamente la información de la densidad $\rho(\vec{r})$ en un punto particular \vec{r} , sino también el gradiente de la densidad, $\nabla\rho(\vec{r})$ con el objetivo de tomar en cuenta la inhomogeneidad de la densidad electrónica real. De tal forma, la aproximación de densidad local es interpretada como el primer elemento de la expansión de Taylor de la densidad uniforme y, con el fin de obtener una mejor aproximación del funcional de intercambio y correlación, se adicionan algunos términos más de la expansión. Así, se tiene (con σ y σ' indicando al espín α o β):

$$E_{XC}[\rho_\alpha, \rho_\beta] = E_{XC}^{LSD} + \sum_{\sigma, \sigma'} \int C_{XC}^{\sigma, \sigma'}(\rho_\alpha, \rho_\beta) \frac{\nabla\rho_\sigma}{\rho_\sigma^{2/3}} \frac{\nabla\rho_{\sigma'}}{\rho_{\sigma'}^{2/3}} + \dots \quad (3.56)$$

Esta forma del funcional 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, dicha aproximación tiene un problema: la derivada funcional (ecuación (3.50)) del

gradiente diverge para la densidad que decae exponencialmente (debido a que el agujero de intercambio no cumple con las reglas de suma (condiciones impuestas por las ecuaciones (3.20) y (3.24)) y no está definido negativamente).

Este problema es corregido haciendo cumplir las restricciones válidas para los agujeros de intercambio y correlación. Si el agujero de intercambio GEA es positivo en alguna parte, éste es ajustado a cero. Asimismo, el funcional es truncado de tal forma que se cumpla las condiciones de normalización de los agujeros de intercambio y correlación. Tomando en cuenta estas correcciones se llega a la Aproximación de Gradiente Generalizado (GGA, por sus siglas en inglés), la cual puede ser representada como:

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) d\vec{r} \quad (3.57)$$

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

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \quad (3.58)$$

de tal forma que cada término es aproximado individualmente.

A grandes rasgos, E_X^{GGA} puede expresarse como:

$$E_X^{GGA} = E_X^{LDA} - \sum_\sigma \int F(s_\sigma) \rho_\sigma^{4/3}(\vec{r}) d\vec{r} \quad (3.59)$$

donde s_σ corresponde al gradiente de densidad reducida,

$$s_\sigma(\vec{r}) = \frac{|\nabla\rho_\sigma(\vec{r})|}{\rho_\sigma^{4/3}(\vec{r})} \quad (3.60)$$

y puede ser entendido como el parámetro de inhomogeneidad local, es decir, el valor de éste varía conforme varía 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 mencionadas anteriormente.

De forma contraria, los funcionales aproximados E_C^{GGA} tienen formas analíticas más complicadas y son obtenidos a partir de datos empíricos y/o datos teóricos⁹.

Funcionales meta-GGA

A pesar de que la corrección por gradientes mejora los resultados obtenidos al hacer una mejor descripción de la variación de la densidad, sigue habiendo un problema en la forma del funcional debido a que no se corrige el problema de la autointeracción.

De acuerdo a la sección 3.3.3, la energía como funcional de la densidad puede ser escrita como:

$$E[\rho(\vec{r})] = T_S[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho] \quad (3.61)$$

donde la repulsión Coulómbica clásica corresponde a:

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \quad (3.62)$$

⁹Para más información y ejemplos específicos de los funcionales de intercambio y correlación GGA consultar referencias [21] y [25]

Este término desaparece para sistemas de un electrón (es decir, donde no hay correlación) dado que incluye un término en el cual la densidad interactúa consigo misma. Por consiguiente, para que la ecuación (3.61) sea correcta, se debe imponer que $J[\rho]$ sea exactamente igual a menos $E[\rho]$ para que la autointeracción sea cancelada.

$$\frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 = -E_{XC}[\rho] \quad (3.63)$$

Una forma de corregir la autointeracción es mediante la adición de la densidad de energía cinética (τ_σ) al funcional de correlación $E_C[\rho]$:

$$\tau_\sigma = \sum_i |\varphi_\sigma|^2 \quad (3.64)$$

Los funcionales que incluyen a τ_σ en el funcional de correlación son conocidos como meta-GGA. Estos describen mejor la correlación nula en sistemas de un electrón, tales como el átomo de hidrógeno.

Saliendo de la escalera: no localidad

Las aproximaciones a los funcionales de la densidad descritas anteriormente, y algunas otras que forman parte de la escalera, no son capaces de describir adecuadamente interacciones intermoleculares a causa de que estos funcionales son considerados como “locales” debido a que ocupan la información obtenida en cada punto de integración \vec{r}_1 . Efectos verdaderamente no locales, como las interacciones intermoleculares, deben ser descritos tomando en cuenta de forma explícita, en cada punto de integración, todos los otros puntos $\vec{r}_2 \neq \vec{r}_1$.

Una forma de atacar este problema es mediante Correcciones de Largo Alcance (LCR, por sus siglas en inglés). En dicha corrección, la interacción interelectrónica $1/r_{12}$ es dividida en una parte de corto alcance y en una de largo alcance:

$$\frac{1}{r_{12}} = \frac{1 - \text{erf}(\omega r_{12})}{r_{12}} + \frac{\text{erf}(\omega r_{12})}{r_{12}} \quad (3.65)$$

Aproximaciones de los funcionales de la densidad modificados son utilizados para tratar la parte de corto alcance, mientras que la de largo alcance es tratada de forma exacta (mediante el intercambio de Hartree-Fock). El parámetro de alcance ω no es universal, depende de la molécula de estudio y es no lineal.

Incluir dentro de $E_X[\rho]$ el intercambio exacto de Hartree-Fock es otra manera de agregar efectos no locales a las aproximaciones anteriores. Sin embargo, incluir el intercambio de Hartree-Fock puede aumentar el costo computacional, haciendo poco práctico el uso de estas correcciones en moléculas de gran tamaño, tal como en el caso de biomoléculas [32].

3.4. Reactividad Química: un enfoque teórico

Conocer las propiedades reactivas (propiedades químicas) de una molécula nos permite predecir su comportamiento durante una reacción o interacción con otra molécula. En el ámbito biológico, conocer las propiedades reactivas nos permite entender la selectividad que existe por parte de ciertas moléculas. En esta sección se describirá brevemente la aproximación de los Orbitales Moleculares Frontera y las funciones de Fukui.

3.4.1. Orbitales Moleculares Frontera (FMO)¹⁰

Un orbital molecular (OM) es una función de onda que describe a un electrón en una molécula [34]; dicha función es construida mediante la combinación lineal de orbitales atómicos. En una reacción, la

¹⁰Esta sección fue desarrollada con base en la información brindada por la referencia [33].

estructura electrónica de cada reactante puede ser descrita por una colección de orbitales moleculares (OMs), los cuales pueden estar ocupados o desocupados.

Considerando la reacción entre las moléculas A y B para dar el producto C , y que los OMs ocupados están doblemente ocupados, es decir, hay dos electrones apareados en cada orbital. Conforme A y B se van acercando se pueden manifestar tres tipos de interacciones (ver Figura 3.2):

- Interacción de cero electrones entre los orbitales desocupados.
- Interacción de dos electrones entre los orbitales desocupados de A y los ocupados de B , y viceversa.
- Interacción de cuatro electrones entre los orbitales ocupados de A y B .

Los procesos energéticos asociados a estos tres tipos de interacciones pueden ser: ninguno, estabilizadores o desestabilizadores, respectivamente. Cuando se estudia el mecanismo de reacción de alguna reacción en particular y se tienen dos o más posibilidades, es necesario comparar los procesos energéticos asociados a cada uno, al igual que las interacciones de estabilización o desestabilización, que dan lugar a la identificación del mecanismo de reacción más favorable.

En principio, llevar a cabo lo dicho anteriormente puede ser sencillo; sin embargo, la situación puede complicarse debido al número de orbitales moleculares que puede poseer una molécula y que pueden estar involucrados. Este problema es simplificado mediante el uso de la Aproximación de Orbitales Moleculares Frontera, tal como veremos a continuación.

La Aproximación de los Orbitales Moleculares Frontera

El propósito de esta aproximación es identificar, de todas las interacciones posibles, cuales son las de mayor relevancia para que ocurra una reacción. El punto clave de esta aproximación es considerar que las interacciones dominantes ocurren entre el orbital ocupado de más alta energía (HOMO, por sus siglas en inglés) y del orbital desocupado de más baja energía (LUMO, por sus siglas en inglés), generalmente conocidos como orbitales moleculares frontera (FMO, por sus siglas en inglés), de dos moléculas diferentes que forman parte de una reacción (Figura 3.2).

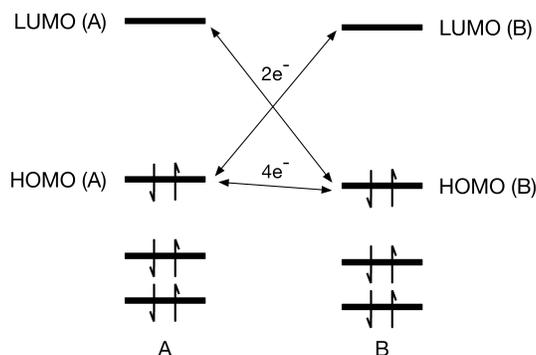


Figura 3.2: Posibles interacciones entre los FMO de dos moléculas diferentes.

La interacción de estabilización es proporcional a $S^2/\Delta\varepsilon$, donde S corresponde a la integral de traslape y $\Delta\varepsilon$ a la separación energética entre los FMO. Para evaluar qué tan favorable es una interacción es necesario analizar ambas contribuciones. En principio, la reacción debe estar permitida por simetría; con respecto a la energía, mientras más cercanos están los orbitales en energía, mayor es la estabilización debida a la interacción.

El conjunto de interacciones repulsivas, provenientes de los orbitales ocupados, corresponden a interacciones impedidas estéricamente. Estas interacciones dependen únicamente del traslape entre los orbitales, a mayor traslape mayor desestabilización. Sin embargo, para fines de esta aproximación, estas interacciones son despreciadas. Esto último no significa que la energía asociada a estas interacciones sea igual a cero, sino que para diferentes mecanismos el valor de energía asociado a éstas es similar. En caso de que se busque tomar en cuenta estas interacciones dentro de la aproximación, basta con considerar la interacción entre los dos HOMOs de las moléculas participantes.

En resumen, de acuerdo a la aproximación de orbitales moleculares frontera, para tener idea de cómo dos moléculas *A* y *B* reaccionarán, es suficiente considerar las interacciones HOMO-LUMO de dichas moléculas.

Compuestos electrofílicos y nucleofílicos

Comúnmente, en una reacción las moléculas *A* y *B* son diferentes y, por lo tanto, son caracterizadas por dos conjuntos de OMs diferentes. Una consecuencia importante es que las dos interacciones HOMO-LUMO tienen un $\Delta\varepsilon$ de diferente magnitud, permitiendo encontrar la interacción favorecida energéticamente en términos del $\Delta\varepsilon$ (Figura 3.3).

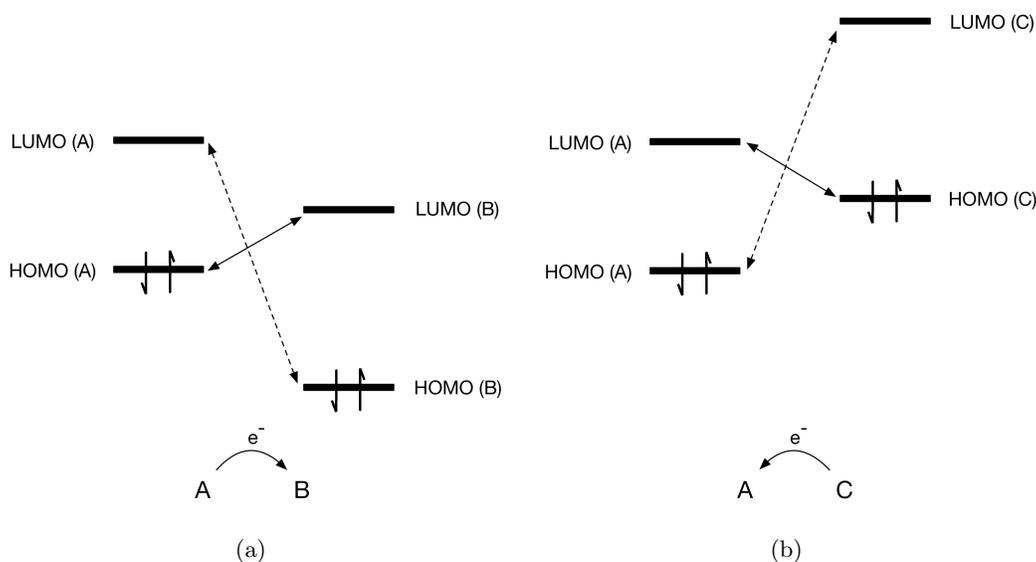


Figura 3.3: Posibles interacciones HOMO-LUMO (Líneas continuas: interacciones predominantes).

De acuerdo a la Figura 3.3(a) la interacción de dos electrones entre el HOMO de *A* y el LUMO de *B* es la predominante y resulta de la transferencia de electrones de *A* y *B*. En este caso, el reactivo *A* es un nucleófilo debido a su tendencia a donar electrones, mientras que *B* es un electrófilo por su tendencia a aceptar electrones. Habitualmente, un nucleófilo está caracterizado por un HOMO energéticamente elevado y un electrófilo por un LUMO energéticamente bajo.

El carácter nucleofílico o electrofílico de una molécula no es absoluto. Esto se puede ver ejemplificado en la Figura 3.3, donde en el par *A-B* (Figura 3.3(a)), la molécula *A* tiene el papel de nucleófilo, mientras que en el par *A-C* (Figura 3.3(b)) es un electrófilo.

La determinación de la naturaleza reactiva de una molécula es uno de los mecanismos posibles para predecir el curso de una reacción química. Si se conoce un mecanismo de reacción es relativamente más

sencillo tener control sobre el curso de una reacción. Por ejemplo, mediante modificaciones de una estructura se pueden funcionalizar moléculas alterando (modificando, mejorando, evitando) su mecanismo de actividad biológica. Asimismo, conocer estas propiedades permitiría describir como se lleva a cabo la interacción entre una proteína y su blanco.

3.4.2. Funciones de Fukui¹¹

Otros de los índices de reactividad emanados de la Teoría de los Funcionales de la Densidad, son las funciones de Fukui, propuestas por Parr y Yang en 1984 [36] como una aplicación de la aproximación de Orbitales Moleculares Frontera. Una función de Fukui, denotada por $f(\vec{r})$, es definida como el cambio diferencial en la densidad electrónica debido a un cambio infinitesimal del número de electrones:

$$f(\vec{r}) = \left(\frac{\partial \rho(\vec{r})}{\partial N} \right)_{V(\vec{r})} \quad (3.66)$$

donde $\rho(\vec{r})$, de acuerdo a la notación que se ha ocupado a lo largo de este capítulo, corresponde a la densidad electrónica y cumple con la condición establecida en la ecuación (3.12). Como veremos más adelante, estas funciones son indicadores regioselectivos, los cuales permiten conocer los sitios reactivos de una molécula en particular.

Origen de las funciones de Fukui

Como vimos en la Sección 3.3, la energía puede ser escrita como un funcional de la densidad $E[\rho]$. Sin embargo, sabemos que la densidad depende del potencial externo, en esta sección expresado como $V(\vec{r})$, y del número de electrones N . Por lo tanto, al escribir la energía como funcional del $V(\vec{r})$ y de N tenemos que la derivada total de ésta es:

$$dE = \left(\frac{\partial E_{V,N}}{\partial N} \right)_{V(\vec{r})} dN + \int \left(\frac{\delta E_{V,N}}{\delta V(\vec{r})} \right)_N \delta V(\vec{r}) d\vec{r} \quad (3.67)$$

Identificando a la derivada parcial como el potencial químico:

$$\mu_{V,N} = \left(\frac{\partial E_{V,N}}{\partial N} \right)_{V(\vec{r})} \quad (3.68)$$

y a la derivada funcional (segundo término) como la densidad electrónica:

$$\rho_{V,N}(\vec{r}) = \left(\frac{\delta E_{V,N}}{\delta V(\vec{r})} \right)_N \quad (3.69)$$

podemos reescribir a la ecuación (3.67) como:

$$dE = \mu_{V,N} dN + \int \rho_{V,N}(\vec{r}) \delta V(\vec{r}) d\vec{r} \quad (3.70)$$

Como se puede observar, el potencial químico $\mu_{V,N}$ (ecuación (3.68)) indica el cambio de la energía del sistema debido a un cambio en el número de electrones. Al igual que la energía, $\mu_{V,N}$ depende del número de electrones y del potencial externo del sistema, por lo que la derivada total del potencial químico, semejante a la de la energía, es:

$$d\mu = \left(\frac{\partial \mu_{V,N}}{\partial N} \right)_{V(\vec{r})} dN + \int \left(\frac{\delta \mu_{V,N}}{\delta V(\vec{r})} \right)_N \delta V(\vec{r}) d\vec{r} \quad (3.71)$$

¹¹Esta sección fue desarrollada con base en la información brindada por las referencias [35] y [36].

La derivada parcial del primer término corresponde a la dureza:

$$\frac{1}{2}\eta = \left(\frac{\partial \mu_{V,N}}{\partial N} \right)_{V(\vec{r})} \quad (3.72)$$

mientras que la derivada funcional del segundo término es identificada como función frontera o función de Fukui:

$$f(\vec{r}) = \left(\frac{\delta \mu_{V,N}}{\delta V(\vec{r})} \right)_N \quad (3.73)$$

Utilizando las relaciones de Maxwell en términos de la Teoría de los Funcionales de la Densidad [37], la ecuación (3.73) puede ser reescrita como:

$$f(\vec{r}) = \left(\frac{\partial \rho_{V,N}(\vec{r})}{\partial N} \right)_{V(\vec{r})} \quad (3.74)$$

La ecuación (3.71) nos ayuda a identificar la dirección por la cual una molécula B reaccionará con la molécula A para la formación de un enlace. En primera instancia, la dirección de reacción preferida va a ser aquella en la cual $d\mu$ corresponda a un máximo. Analizando los términos por separado, la dureza nos indica la facilidad de una molécula a aceptar o perder electrones y no es direccional; en cambio, la función de Fukui evalúa el cambio de la densidad electrónica debido al cambio en el número de electrones siendo más direccional que el primer término. Por lo tanto, se asume que la dirección preferida será aquella con un mayor valor de $f(\vec{r})$ en el sitio de reacción.

Al igual que $E[\rho]$, la pendiente de $\rho_{V,N}(\vec{r})$, es decir, su derivada con respecto al número de electrones, presenta discontinuidades para números enteros [38]. Por consiguiente, de la ecuación (3.74) se obtienen tres índices de reactividad diferentes:

(a) Para ataques electrofílicos:

$$f^-(\vec{r}) = \left(\frac{\partial \rho_{V,N}(\vec{r})}{\partial N} \right)_{V(\vec{r})}^- = \rho_{V,N}(\vec{r}) - \rho_{V,N-1}(\vec{r}) \quad (3.75)$$

(b) Para ataques nucleofílicos:

$$f^+(\vec{r}) = \left(\frac{\partial \rho_{V,N}(\vec{r})}{\partial N} \right)_{V(\vec{r})}^+ = \rho_{V,N+1}(\vec{r}) - \rho_{V,N}(\vec{r}) \quad (3.76)$$

(c) Ataques neutros (por radicales):

$$f^0(\vec{r}) = \left(\frac{\partial \rho_{V,N}(\vec{r})}{\partial N} \right)_{V(\vec{r})}^0 = \frac{1}{2} (f^+(\vec{r}) + f^-(\vec{r})) = \frac{1}{2} (\rho_{V,N+1}(\vec{r}) - \rho_{V,N-1}(\vec{r})) \quad (3.77)$$

Utilizando la aproximación de los Orbitales Moleculares Frontera, las derivadas anteriores pueden ser aproximadas a los cuadrados del HOMO y del LUMO:

$$f^-(\vec{r}) = |\varphi_{V,N}^{HOMO}(\vec{r})|^2 = \rho_{V,N}^{HOMO}(\vec{r}) \quad (3.78)$$

$$f^+(\vec{r}) = |\varphi_{V,N}^{LUMO}(\vec{r})|^2 = \rho_{V,N}^{LUMO}(\vec{r}) \quad (3.79)$$

$$f^0(\vec{r}) = \frac{1}{2} (\rho_{V,N}^{HOMO}(\vec{r}) + \rho_{V,N}^{LUMO}(\vec{r})) \quad (3.80)$$

Como fue mencionado anteriormente, estos índices regioselectivos son conocidos como funciones de Fukui, parámetros que permiten describir la reactividad de una molécula y reciben su nombre en honor a Kenichi Fukui, pionero de la descripción de la reactividad mediante la aproximación de los FMO (sección 3.4.1).

Capítulo 4

Metodología

4.1. Validación

Para realizar la descripción de la estructura electrónica y propiedades de las moléculas de ADN, se probaron los funcionales PW91 [39] y M06-L [32], con la finalidad de estimar la energía promedio de puente de hidrógeno entre adenina y timina (Figura 4.1), cuyo valor se encuentra reportado experimentalmente [40].

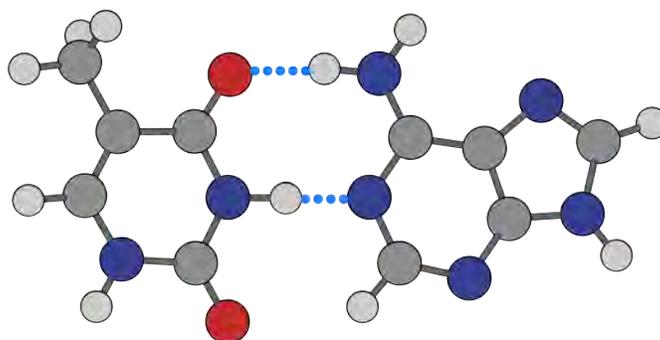


Figura 4.1: Dímero TA utilizado para la validación del modelo de cálculo.

Se eligieron estos dos funcionales debido a que:

- Fonseca *et al.* [41] reportaron que el funcional PW91 (perteneciente a la familia GGA) era adecuado para la descripción y estimación de las interacción de puente de hidrógeno entre los pares de base del ADN.
- Zhao y Truhlar [32] reportaron que el funcional meta-GGA M06-L, a pesar de ser local, era adecuado para la estimación de interacciones no covalentes en tiempos de cómputo adecuados para moléculas de gran tamaño.

Para calcular la energía promedio de puente de hidrógeno (Figura 4.1), primero se llevó a cabo la optimización de geometría del dímero; una vez optimizado se separó en sus monómeros y se calculó la energía de cada monómero (base nitrogenada). La energía promedio de puente de hidrógeno se obtuvo mediante la siguiente ecuación:

$$\Delta E = \frac{E_{TA} - (E_T + E_A)}{2} \quad (4.1)$$

donde E_{TA} corresponde a la energía total del dímero optimizado, E_T y E_A a la energía total de los monómeros con la conformación del dímero optimizado. Como es sabido, entre la interacción de las bases

A y T existen dos puentes de hidrógeno, por lo que la diferencia de energía entre el dímero y sus monómeros fue dividida entre dos para obtener de la energía promedio de puente de hidrógeno.

Para evaluar el comportamiento de los dos funcionales, se utilizó una base doble numérica (DN). Los resultados obtenidos se muestran en la Tabla 4.1. De esta tabla, se puede observar que el funcional que mejor estima la energía promedio de puente de hidrógeno es M06-L.

Modelo	Energía (kcal/mol)
PW91/DN	13.7
M06-L/DN	12.1
Experimental	12.1

Tabla 4.1: Energía de puente de hidrógeno promedio para el dímero TA obtenida por medio de dos funcionales con una base DN.

Para determinar el efecto de la base sobre la estimación de la energía promedio de puente de hidrógeno, se compararon los resultados obtenidos para dicha energía utilizando bases numéricas de diferente tamaño empleando el funcional M06-L. Los resultados se muestran en la Tabla 4.2 y, como se puede ver, con el modelo M06-L/DN se obtiene el valor experimental.

Funciones base	Energía (kcal/mol)
DN	12.1
DND	9.2
DNP	9.1
TNP	11.6
Experimental	12.1

Tabla 4.2: Energía de puente de hidrógeno promedio para el dímero TA obtenida utilizando el funcional M06-L y bases de diferente tamaño.

Finalmente, para estimar el efecto del disolvente (agua) sobre la energía promedio de puente de hidrógeno se realizó el mismo procedimiento por medio del modelo COSMO [42, 43] (*conductor-like screening model*). Los resultados (M06-L/DN) se encuentran en la Tabla 4.3. Como se puede notar, tomar en cuenta la presencia del disolvente disminuye la energía promedio de puente de hidrógeno. Este resultado puede no ser representativo de lo que sucede en una cadena de ADN, debido a que las interacciones entre las bases nitrogenadas quedan aisladas del medio al estabilizarse [44].

Fase	Energía (kcal/mol)
Gas	12.1
Acuosa	7.35
Experimental	12.1

Tabla 4.3: Energía de puente de hidrógeno promedio para el dímero TA tomando en cuenta el efecto de disolvente (ocupando el método M06-L).

Por lo tanto, tomando en cuenta los resultados anteriores, los cálculos de estructura electrónica y propiedades evaluados en este estudio se realizaron con el modelo M06-L/DN sin efectos de disolvente.

4.2. Detalles computacionales

Para llevar a cabo este trabajo, se tomaron en cuenta tres modelos diferentes para el cálculo de energías promedio de puentes de hidrógeno:

- (a) Ideal no relajado
- (b) Ideal relajado
- (c) Por bloques

Para el caso ideal, las estructuras fueron obtenidas del software de libre acceso *make-na* [45]. Dicho software toma en cuenta la distancia entre bases nitrogenadas adyacentes, distancia entre grupos fosfatos, ángulos de giro y torsión entre las bases nitrogenadas, entre otros parámetros, para obtener una cadena de ADN tipo B, es decir, que cumpla con las condiciones de simetría establecidos para este sistema (Figura 4.2). Una vez obtenidas las cadenas de ADN ideales, se realizaron cálculos de energía para cada cadena y para cada una de sus hebras con el objetivo de calcular la energía promedio de puente de hidrógeno.

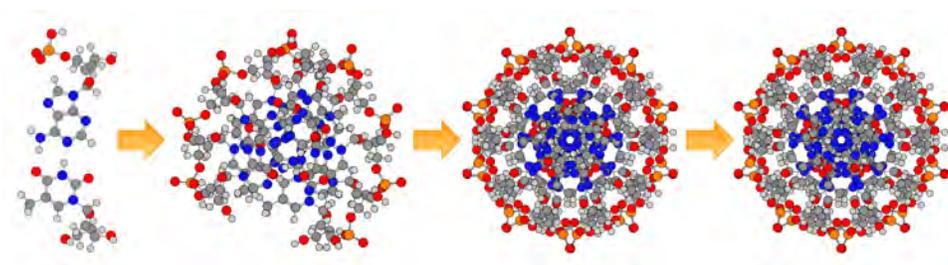


Figura 4.2: Esquema de crecimiento de estructuras ideales no relajadas generadas de 1 a 14 pares de base.

Para el modelo ideal relajado, las estructuras obtenidas del modelo anterior fueron sometidas a un proceso de optimización de geometría (Figura 4.3). Con la geometría de las cadenas optimizadas, se realizaron cálculos de energía de cada hebra para la obtención de la energía promedio de puente de hidrógeno.

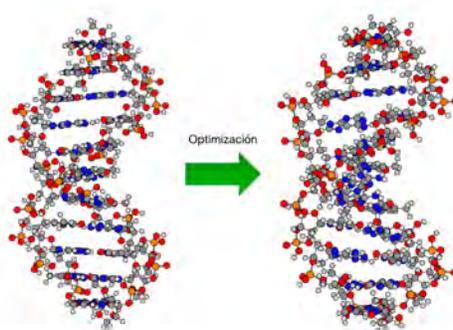


Figura 4.3: Esquemización del proceso de optimización de la cadena ideal no relajada.

Por último, para el modelo de bloques, se realizó la construcción de cadenas de ADN mediante el uso del software GaussView 5.0.8 [46], agregando un par de base (bloque) sobre la cadena anterior, cuidando que las interacciones de puente de hidrógeno se mantuvieran en todas las cadenas (Figura 4.4). Dichas estructuras, como en el modelo previo, fueron sometidas a un proceso de optimización de geometría y, una vez obtenidas las estructuras optimizadas, se realizaron cálculos de energía para cada una de las hebras para la obtención de la energía promedio de puente de hidrógeno.

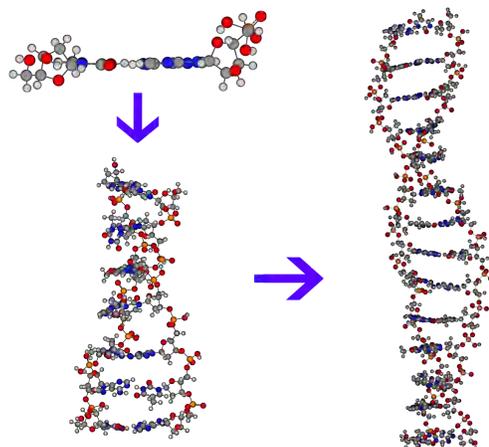


Figura 4.4: Construcción por bloques de cadenas de ADN.

En cada uno de los casos anteriores, se usó como modelo de construcción a la caja TATA, es decir, cadenas de ADN ricas en T y A. Se tomó en cuenta la dirección $5' - 3'$, así como la presencia de grupos fosfato y de azúcares. La carga de los grupos fosfato fue neutralizada mediante la adición de átomos de hidrógeno. Las cadenas generadas o construidas van desde un par de base hasta 14 pares de base (a excepción del modelo ideal relajado, donde se llegó hasta 13 pares de base). Todos los cálculos de energía fueron llevados a cabo usando el funcional M06-L con una base DN.

Para las cadenas de todos los modelos de estudio se realizó un análisis de población de Mulliken y se calcularon los mapas de potencial electrostático, los orbitales moleculares y las funciones de Fukui ($f^-(\vec{r})$ y $f^+(\vec{r})$).

Todos los cálculos, incluyendo los de la validación (sección 4.1), fueron realizados usando el código DMol³ [47] implementado en la suite de programas Materials Studio 8.0 [48]. Tanto para las optimizaciones de geometría como para las iteraciones SCF de los cálculos de energía, fue establecido el criterio de convergencia 1×10^{-4} Ha.

Capítulo 5

Resultados y Análisis

5.1. Puentes de hidrógeno

5.1.1. Energía promedio

En principio, se puede suponer que la cadena ideal conformada por unidades TA debe conservar la simetría de una cadena de ADN tipo B. Esta cadena se construye imponiendo condiciones de simetría (rotaciones y traslaciones), obteniendo una cadena como la que se muestra en la Figura 5.1. En ella es posible observar que existe un eje de simetría C_{10} y una total planaridad de las bases nitrogenadas de los pares de base.

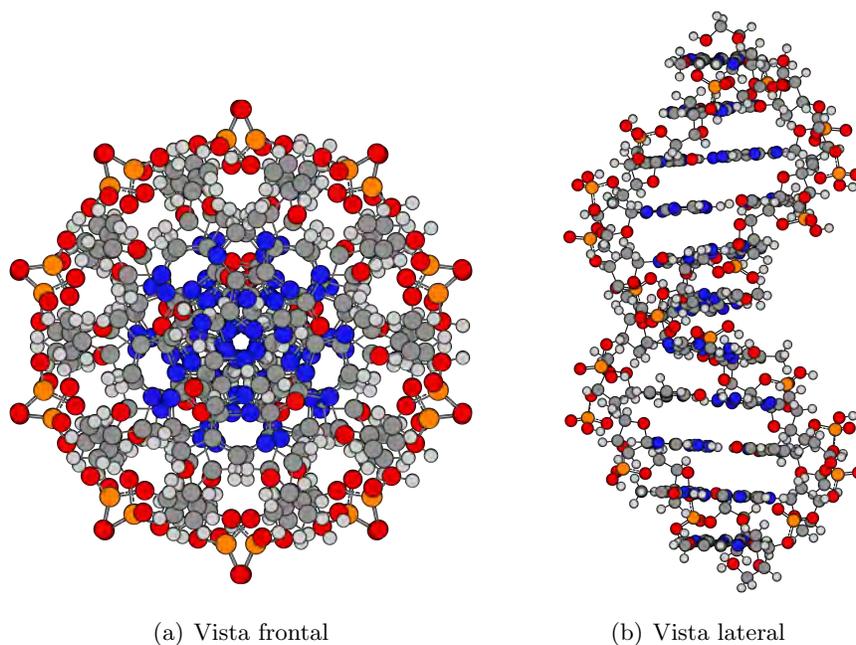


Figura 5.1: Fragmento ideal de 11 pares de base usando como modelo la caja TATA.

Al construir la cadena de esta manera no se toman en consideración las torsiones intrabase de los pares de base (propeller-twist), las cuales maximizan las interacciones de apilamiento $\pi - \pi$ entre los pares de base y evitan que las moléculas de agua interactúen con las bases nitrogenadas. Evidentemente, un cambio en la conformación debe tener efecto sobre la eficiencia de todas las interacciones involucradas en la estabilización de la doble hélice. Por este motivo, al hacer un análisis es necesario comparar el efecto de dichas interacciones sobre la energía de puente de hidrógeno de la cadena ideal con respecto a lo encontrado en la cadena ideal relajada y en la cadena por bloques.

Para estimar la energía promedio de puente de hidrógeno se evalúa la diferencia entre la energía total de la cadena y la suma de las energías totales de las hebras que conforman a la cadena. Posteriormente, esta diferencia fue dividida entre el número de puentes de hidrógeno existentes en la cadena. Lo dicho anteriormente puede expresarse en la siguiente ecuación:

$$\bar{E}_{HB} = \frac{E_n - (E_{n_{1s}} + E_{n_{2s}})}{N_{HB}} \quad (5.1)$$

donde \bar{E}_{HB} corresponde a la energía promedio de puente de hidrógeno, E_n la energía total de una cadena de ADN de n pares de base, $E_{n_{1s}}$ y $E_{n_{2s}}$ la energía total de cada hebra que forma a la cadena y N_{HB} el número de puentes de hidrógeno que hay en la cadena. Para poder calcular la energía promedio de puente de hidrógeno de esta forma, se considera que la fuerza de los puentes de hidrógeno es mucho mayor que la de otras interacciones no covalentes presentes. La variación de la energía promedio de puente de hidrógeno en función del número de unidades TA (pares de base) puede verse en la Figura 5.2 y los valores se muestran en la Tabla 5.1.

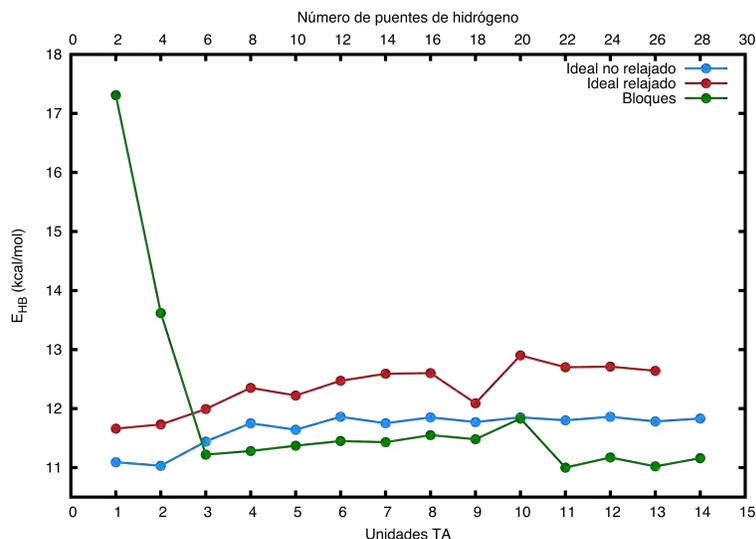


Figura 5.2: Energía promedio de puente de hidrógeno como función del tamaño de la cadena de ADN y del número de puentes de hidrógeno, para el caso ideal no relajado (azul), ideal relajado (rojo) y por bloques (verde).

Como puede verse en la Figura 5.2, la curva que representa al modelo ideal tiene energías más bajas para las cadenas de uno y dos pares de base. Considerando que este modelo tiene implícita la condición de la geometría de un par de base en la cadena infinita, es de esperarse que la energía de puente de hidrógeno para este sistema (1TA) tenga una energía menor que las cadenas de mayor tamaño debido a la ausencia de interacciones no covalentes (apilamiento $\pi - \pi$) con pares de base vecinos. Este resultado parece contradecirse con lo encontrado en 2TA que tiene una energía muy parecida a la de 1TA (la diferencia entre ellas es de 0.03 kcal/mol), esto es consecuencia de la planaridad de las bases que no permite que el apilamiento $\pi - \pi$ establezca al puente de hidrógeno [44]. Esta puede ser la razón de que en estudios previos no se encontrara diferencia significativa para los puentes de hidrógeno presentes entre uno y dos pares de base (sección 2.2.1). A partir de tres unidades TA se presentan oscilaciones periódicas que se desvanecen conforme aumenta el tamaño de la cadena. Este comportamiento periódico no se presenta para los otros modelos, por lo que en la siguiente sección se hace un análisis más detallado de este modelo.

La curva correspondiente al modelo ideal relajado sigue una tendencia similar a la encontrada en el modelo ideal no relajado, pero con energías ligeramente más grandes. Esto debido a que al permitir la reorganización de la estructura se maximizan las interacciones, favoreciendo la estabilización de

Unidades TA	No. de puentes de hidrógeno	Ideal no relajado \bar{E}_{HB} (kcal/mol)	Ideal relajado \bar{E}_{HB} (kcal/mol)	Bloques \bar{E}_{HB} (kcal/mol)
1	2	11.09	11.66	17.31
2	4	11.03	11.73	13.62
3	6	11.44	11.99	11.22
4	8	11.75	12.35	11.27
5	10	11.64	12.22	11.37
6	12	11.86	12.47	11.45
7	14	11.75	12.59	11.43
8	16	11.85	12.60	11.55
9	18	11.77	12.09	11.48
10	20	11.85	12.90	11.83
11	22	11.80	12.70	11.00
12	24	11.86	12.71	11.17
13	26	11.78	12.64	11.02
14	28	11.83	-	11.16

Tabla 5.1: Valores de la energía promedio de puente de hidrógeno en función del tamaño de la cadena de ADN y del número de puentes de hidrógeno para los diferentes modelos de estudio.

los puentes de hidrógeno. El punto correspondiente a 9TA presenta un valor que no sigue la tendencia encontrada con los demás, esta situación se discutirá al explicar lo encontrado para el modelo por bloques.

Para el modelo por bloques se encuentra que 1TA y 2TA tienen energías de puente de hidrógeno mucho mayores a las encontradas en los otros modelos. Esto puede explicarse debido a que el par de base 1TA se encuentra aislado, caso contrario a lo que sucede con 1TA en el modelo ideal. A partir de la cadena de dos pares de base, se sigue la misma tendencia que en los otros dos modelos pero con menor energía promedio de puente de hidrógeno. Esto se debe a que al estar más separados los pares de base entre sí, las interacciones de apilamiento $\pi - \pi$ son subestimadas, no contribuyendo de forma importante a la estabilización de los puentes de hidrógeno.

De la Figura 5.2, también se puede observar que tanto para modelo ideal sin relajar como para el modelo ideal relajado conforme aumenta el número de pares de base (vecinos), la energía asociada al puente de hidrógeno tiende a crecer. Lo anterior puede ser asociado a que conforme aumenta el número de pares de base aumenta el número de interacciones en la cadena, las cuales favorecen la interacción de puente de hidrógeno. Esto no ocurre para el modelo de bloques, dado a que, como se explicó anteriormente, los pares de base se encuentran más separados que en los otros modelos.

Se sabe que la doble hélice da un giro completo cada (aproximadamente) 10 u 11 pares de base [5]. En la Figura 5.2, para los modelos ideal relajado y por bloques se puede notar un cambio importante de energía alrededor de la cadena de diez pares de base. En las Figuras 5.11 y 5.12 (mostradas al final de esta sección) se muestra una vista lateral de las cadenas, donde las bases nitrogenadas del primer par de base están eclipsadas entre sí de tal forma que es posible apreciar el giro de la doble hélice. Para el modelo de bloques, se puede ver que la cadena de diez pares de base ha dado solamente medio giro, mientras que el modelo ideal relajado ha completado el giro entre las cadenas de diez y once pares de base. Por lo tanto, las variaciones de energía observadas pueden ser atribuidas al medio giro y al giro completo en los modelos de bloques e ideal relajado, respectivamente. Para la cadena de cinco pares de base del modelo ideal relajado se nota un ligero cambio en la energía promedio de puente de hidrógeno, por lo que el medio giro también puede ser identificado en este modelo. Tomando en cuenta esto, se puede decir que el medio giro y el giro completo de la cadena altera la estabilidad de los puentes de hidrógeno, provocando variaciones en la energía promedio de estas interacciones.

Enfocándose en el caso ideal no relajado, se presentan oscilaciones que se desvanecen conforme aumenta el tamaño de la cadena. Este comportamiento no se presenta en los otros dos modelos, por lo que en la siguiente sección se hace un análisis detallado de este modelo.

Modelo ideal no relajado

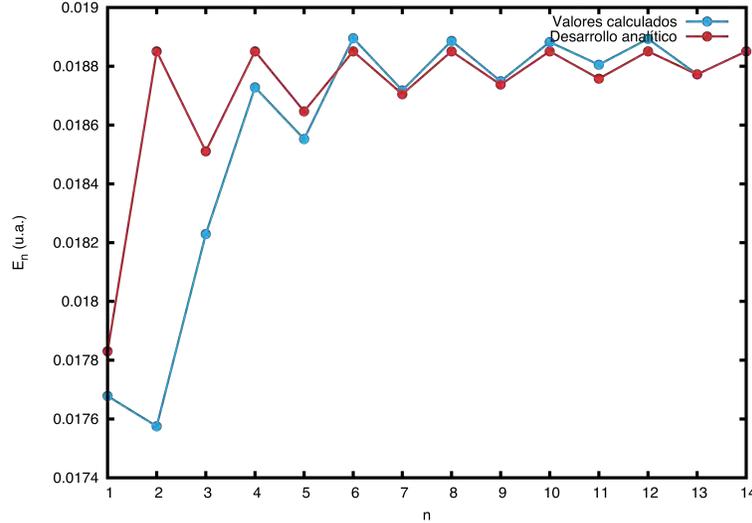


Figura 5.3: Energía promedio de puente de hidrógeno como función del tamaño de la cadena de ADN (representadas por n), para el caso ideal no relajado: valores calculados (azul) y desarrollo analítico (rojo).

La Figura 5.3 muestra la energía promedio de puentes de hidrógeno como función del número de pares de base únicamente para el caso ideal. Partiendo de la estructura ideal, por condiciones de simetría (inversión y rotación), cada fragmento AT y TA central debe ser equivalente y, por lo tanto, la energía total debe ser un múltiplo entero de una unidad fundamental. De esta forma, se puede suponer que las oscilaciones encontradas (Figura 5.3) se deben a efectos de borde. Con el objeto de estudiar estos efectos, se puede suponer que dicho efecto está relacionado con la simetría asociada con el número de pares de base dentro de la cadena, diferenciando así entre cadenas con n par o impar. De esta manera se propone el siguiente modelo.

Tomando a n como el número total de pares de base y diferenciando entre las cadenas con un número impar o par de n , la energía promedio de puente de hidrógeno (E_n) puede ser estimada de la siguiente manera:

$$E_n = \begin{cases} \frac{1}{2} \left(\epsilon_0 + \frac{\Delta E_{par}}{n} \right) & n = \text{par} \\ \frac{1}{2} \left(\epsilon_0 + \frac{\Delta E_{non}}{n} \right) & n = \text{impar} \end{cases} \quad (5.2)$$

donde el dos en el denominador corresponde a los dos puentes de hidrógeno entre cada T y A, ϵ_0 a la energía de interacción entre las bases nitrogenadas, y ΔE_{par} y ΔE_{non} al efecto de borde, si el número de pares de base es par o no, respectivamente. Por lo tanto, se tiene un sistema de dos ecuaciones con tres incógnitas (ϵ_0 , ΔE_{par} y ΔE_{non}). Del resultado numérico, se observa que E_n toma un valor aproximadamente constante a partir de $n = 4$, esto nos lleva a pensar que una buena aproximación para ΔE_{par} sería que ésta tome el valor de cero. Por lo tanto, se puede reescribir la ecuación (5.2) como:

$$E_n = \begin{cases} \frac{1}{2} (\epsilon_0) & n = \text{par} \\ \frac{1}{2} \left(\epsilon_0 + \frac{\Delta E_{non}}{n} \right) & n = \text{impar} \end{cases} \quad (5.3)$$

Considerando el valor de la energía promedio de puente de hidrógeno de las cadenas más grandes (13TA y 14TA), se tiene que $E_{14} = 0.018851$ u.a., por lo tanto de la ecuación (5.3) se obtiene:

$$\epsilon_0 = 2E_n = 2(0.018851 \text{ u.a.}) = 0.037701 \text{ u.a.} \quad (5.4)$$

Sustituyendo ϵ_0 en el caso impar de la ecuación (5.3) y sabiendo que $E_{13} = 0.018772$ u.a., se tiene que:

$$\Delta E_{non} = n(2E_n - \epsilon_0) = 13(2(0.018772 \text{ u.a.}) - 0.037701 \text{ u.a.}) = -0.001020 \text{ u.a.} \quad (5.5)$$

Así, teniendo el valor de las variables ϵ_0 y ΔE_{non} , la ecuación (5.3) se reescribe como:

$$E_n = \begin{cases} \frac{1}{2}(0.037701 \text{ u.a.}) & n = \text{par} \\ \frac{1}{2}\left(0.037701 \text{ u.a.} - \frac{0.001020 \text{ u.a.}}{n}\right) & n = \text{impar} \end{cases} \quad (5.6)$$

Los valores obtenidos del desarrollo analítico se muestran en la Figura 5.3. Como puede observarse, el desarrollo analítico se ajusta muy bien a los valores estimados para la energía promedio de puente de hidrógeno para $n > 3$. Las diferencias encontradas para $n > 3$ se deben a errores numéricos en el cálculo. Cuando $n = 1$ no existe la posibilidad de que se presenten interacciones de apilamiento $\pi - \pi$ debido a la ausencia de vecinos. De ahí que la energía promedio de puente de hidrógeno de un par aislado sea más baja de la que se presenta en cualquier otro caso. La razón por la cual el modelo analítico falla a n pequeña se debe a que nuestro modelo asume a que el puente de hidrógeno no es afectado por la presencia de los extremos de la cadena, de forma que esta aproximación es válida sólo para cadena grandes.

5.1.2. Distancias promedio

En la Tabla 5.2 se encuentran los valores de las distancias promedio de los puentes de hidrógeno entre las bases nitrogenadas para los diferentes modelos. En la Figura 5.4 se explica la notación utilizada para nombrar a las interacciones entre A y T. Es importante mencionar que para este análisis se considerará el hecho de que entre la adenina y timina se forman solamente dos puentes de hidrógeno (D_1 y D_2).

En general, de la Tabla 5.2 se puede observar que para todas las cadenas, el puente de hidrógeno central (D_2) es el más fuerte, debido a que tiene la distancia más pequeña asociada, por consiguiente el otro puente de hidrógeno es más débil (D_1). Las distancias obtenidas con los diferentes métodos concuerdan adecuadamente con las distancias reportadas experimentalmente [49] (2.93-2.95 Å para D_1 y 2.82-2.85 Å para D_2). La diferencia de 1 Å entre los datos obtenidos y lo reportado experimentalmente, corresponde a la presencia del H. Las ligeras diferencias entre estos datos se pueden deber a que a pesar de que el funcional utilizado describe adecuadamente las interacciones débiles, éste no cuenta con una corrección explícita que describa adecuadamente las distancias para estas interacciones (ver Sección 3.3.1). Otra factor importante es que los cálculos se realizaron en fase gas, por lo que el ambiente en el cual se encuentra la molécula es diferente al del cristal del cual se obtuvieron las distancias experimentales. Sin embargo, es de esperarse que, al ser una región hidrofóbica, las distancias de los puentes de hidrógeno entre las bases nitrogenadas no se vea significativamente afectada por el medio.

Para el modelo ideal no relajado, como es de esperarse, las distancias promedio de puente de hidrógeno siempre son las mismas en todas las cadenas. Al ser cadenas donde las condiciones de simetría se cumplen, es de esperarse que la distancia de puente de hidrógeno sea invariante entre los diferentes pares de base.

Comparando al modelo ideal relajado con el ideal no relajado, se puede notar un cambio en la distancia promedio de puente de hidrógeno para todos los casos. Al igual que en el análisis de energías promedio de puentes de hidrógeno, al permitir la reorganización de los átomos, las interacciones no covalentes presentes a lo largo de las cadenas son optimizadas, por lo que la distancia promedio de puente de hidrógeno se ve modificada. Uno de los cambios más importantes ocurre entre la cadena de un par de base y la de dos

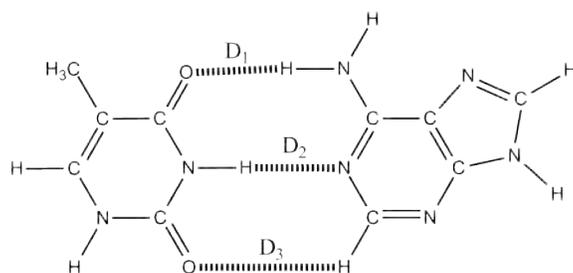


Figura 5.4: Numeración de distancias entre los átomos de A y T.

Unidades TA	Ideal no relajado		Ideal relajado		Por bloques	
	D ₁ (Å)	D ₂ (Å)	D ₁ (Å)	D ₂ (Å)	D ₁ (Å)	D ₂ (Å)
1	1.871	1.810	1.755	1.724	1.981	1.748
2	1.871	1.811	1.822	1.703	1.934	1.814
3	1.870	1.810	1.829	1.691	1.864	1.755
4	1.870	1.810	1.872	1.701	1.896	1.728
5	1.870	1.810	1.841	1.714	1.876	1.708
6	1.870	1.810	1.824	1.715	1.878	1.703
5	1.870	1.810	1.841	1.714	1.876	1.708
6	1.870	1.810	1.824	1.715	1.878	1.703
7	1.870	1.810	1.831	1.760	1.898	1.724
8	1.870	1.810	1.833	1.706	1.848	1.719
9	1.870	1.810	1.826	1.707	1.897	1.718
10	1.870	1.810	1.831	1.700	1.874	1.699
11	1.870	1.810	1.823	1.698	1.970	1.750
12	1.870	1.810	1.820	1.707	1.928	1.721
13	1.870	1.810	1.821	1.705	1.976	1.715
14	1.870	1.810	-	-	1.917	1.698

 Tabla 5.2: Distancia promedio de puente de hidrógeno para las cadenas de los tres modelos de estudio. (Ideal no relajado $\sigma = < 8.29 \times 10^{-4} \text{Å}$; ideal relajado: $\sigma = < 1.15 \times 10^{-1} \text{Å}$; por bloques: $\sigma = < 1.62 \times 10^{-1} \text{Å}$)

pares de base, donde la distancia del puente de hidrógeno central tiende a disminuir, mientras que la del puente de hidrógeno lateral tiende a aumentar, resaltando la participación del puente de hidrógeno central en la estabilización de la cadena. Conforme el tamaño de la cadena aumenta, el cambio en las distancias promedio se va haciendo menos significativo para los dos puentes de hidrógeno, lo cual está relacionado con el aumento de las interacciones no covalentes (apilamiento $\pi - \pi$), ya que conforme éstas aumentan los puentes de hidrógeno se encuentran más restringidos y, por lo tanto, el cambio en las distancias de puente de hidrógeno disminuye. Consecuentemente, considerar la presencia de vecinos es importante y permite predecir que la distancia de puente de hidrógeno tiende hacia un valor promedio, el cual puede estimarse con la cadena más grande estudiada (14TA). Por otro lado, se puede observar que en las cadenas de nueve y diez pares de base hay un cambio en las distancias promedio de puente de hidrógeno a comparación de las cadenas vecinas. Para 9TA se encuentra que D₁ disminuye, mientras que para 10TA D₁ aumenta pero D₂ disminuye. Comparando con los valores de energía promedio de puente de hidrógeno, podemos asociar este cambio a que la cadena está completando un giro.

Finalmente, para el modelo por bloques se observa que el puente de hidrógeno lateral tiene asociada una distancia más grande que la obtenida con los otros modelos. Sin embargo, la distancia del puente de hidrógeno central sigue la misma tendencia. Esto puede asociarse a que al haber mayor espacio entre las bases nitrogenadas de los pares de base, éstas se pueden mover con menos restricciones por lo que el puente de hidrógeno más fuerte se mantiene a lo largo de los diferentes pares de base, mientras que el más débil es susceptible a cambios en la distancia.

5.1.3. Análisis de transferencia de carga

En las secciones pasadas se llevó a cabo un análisis de las energías y distancia promedio asociadas a los puentes de hidrógeno en las diferentes cadenas optimizadas. Como se pudo ver, el tamaño de la cadena tuvo efecto sobre estas propiedades como consecuencia de la presencia de pares de base vecinos y, por lo tanto, de las interacciones $\pi - \pi$.

Con el propósito de estudiar el efecto del tamaño de la cadena sobre los puentes individuales, en esta sección se hará un análisis sobre la acumulación de densidad electrónica sobre las regiones de las bases nitrogenadas en las cuales ocurre la interacción entre T y A. Al no tratarse de un enlace iónico ni un enlace covalente, en una interacción intermolecular la carga de los átomos es compartida entre ellos, es decir, parte de la carga de un átomo electropositivo es transferida al átomo electronegativo; mientras mayor sea esta transferencia mayor será la fuerza de la interacción intermolecular, en este caso, del puente de hidrógeno.

Para este fin, se utiliza el Índice de Separación de Carga (CSI, por sus siglas en inglés), el cual es una medida de la polaridad local de una molécula y se obtiene mediante la suma de los valores absolutos de las cargas q de cada átomo i presente en la región de interés [50]:

$$CSI = \sum_i |q_i| \quad (5.7)$$

Otra forma de interpretar a este descriptor es como la medida del grado de separación de cargas positivas y negativas en la región de interés¹ [51]. Mientras más grande sea el valor de este descriptor se tiene que más polarizada se encuentra cierta región, es decir, mayor acumulación de densidad electrónica hay sobre ésta y, en caso de una interacción, más fuerte es.

A efecto de obtener dicho descriptor, se realizó un análisis de población de Mulliken. Para este análisis, se hace la suposición de que las cadenas crecieron a partir de un par de base central (Figura 5.5), por lo tanto, éste se hará sobre las cadenas que cuenten con un número impar de pares de base. Específicamente, se tomará en cuenta la cadena 1TA (la base del crecimiento central) y las cadenas 3TA a 9TA (aquellas donde ocurren los cambios de energía promedio de puente de hidrógeno significativos). Adicionalmente, el análisis mencionado anteriormente se realizará para la cadena de diez pares de base con el objetivo de completar la descripción de la cadena de diez pares de base, en la cual la doble hélice da un giro completo.

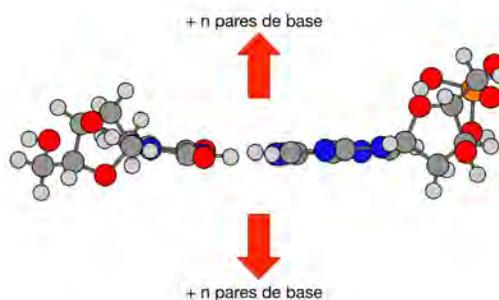


Figura 5.5: Crecimiento de la cadena a partir de un par de base central.

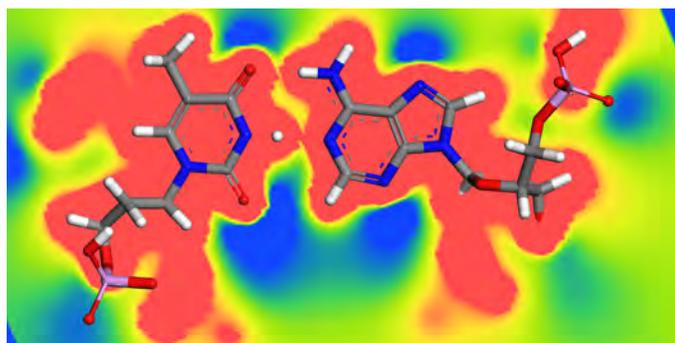
En la Tabla 5.3 (mostrada al final de este capítulo) se muestran los valores del CSI para las regiones involucradas en la interacción entre timina y adenina. Los valores se encuentran organizados de acuerdo

¹El valor de las cargas nos indica la acumulación de densidad electrónica sobre un átomo.

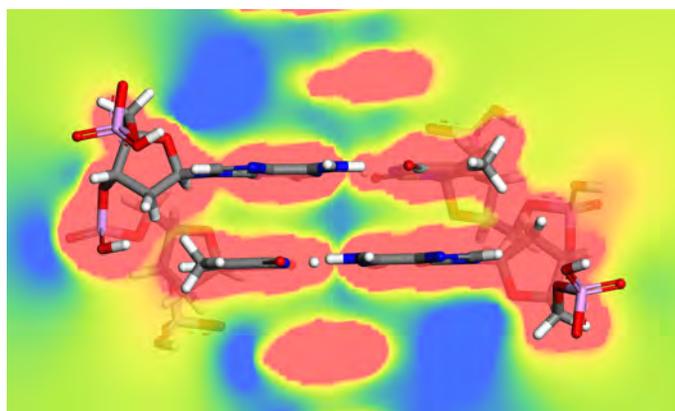
al modelo de estudio y a la nomenclatura utilizada en la Figura 5.4. De forma general, se tiene que para los tres modelos de estudio, el puente de hidrógeno central D_2 es el más fuerte debido a que se tiene un valor más grande de CSI, mientras que el puente de hidrógeno lateral D_1 es el más débil.

Para el modelo ideal no relajado las diferencias en el valor del descriptor CSI entre los diferentes pares de base no son significativamente grandes, como era de esperarse, esto se debe a la simetría del sistema. Sin embargo, se observa que los pares de base centrales cuentan con los puentes de hidrógeno más fuertes mientras que los pares de base de los extremos con los más débiles. Esto último se puede asociar a efectos de borde, ya que al ir aumentando el tamaño de la cadena, la diferencia va decreciendo, es decir, el efecto se va haciendo menos importante. Este análisis coincide con lo encontrado en las energías promedio de puente de hidrógeno, puesto que en las cadenas de menor tamaño la oscilación en la energía era mayor que en las cadenas de mayor tamaño.

Comparando al modelo ideal no relajado con el modelo ideal relajado se puede ver que las interacciones de puente de hidrógeno tienden a ser más fuertes al permitir el reacomodo de los átomos. Como fue explicado anteriormente, esta reorganización permite que las interacciones entre los pares de base (interacciones $\pi - \pi$) se optimicen. Esto último puede verse con mayor claridad en las Figuras 5.6 y 5.7, donde se muestran los potenciales electrostáticos desde dos perspectivas diferentes para los pares de base centrales de las cadenas de siete unidades TA de los modelos ideal no relajado e ideal relajado, respectivamente. Como se puede notar, tanto en la vista frontal como en la lateral, en el modelo relajado existe una mayor acumulación de carga en la interacción entre las bases nitrogenadas que en el modelo no relajado.

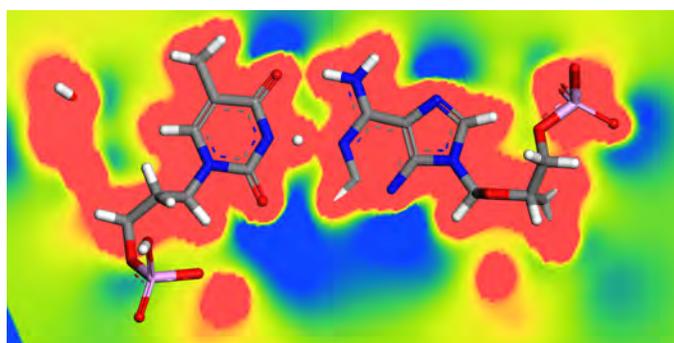


(a) Vista frontal

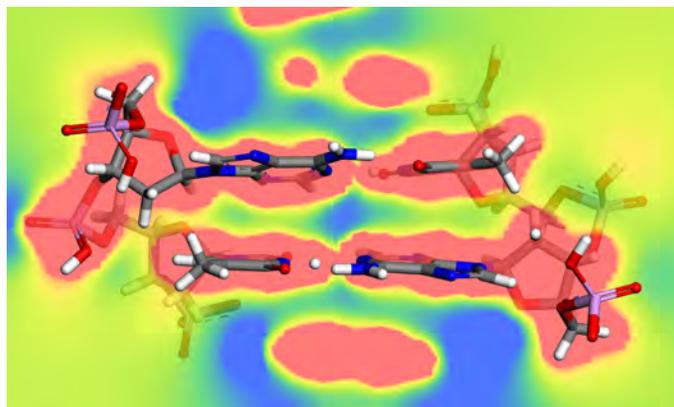


(b) Vista lateral

Figura 5.6: Potencial electrostático para los pares de base centrales de la cadena de siete unidades TA del modelo ideal no relajado (Escala: de -0.078585 u.a. (rojo) a 0.078585 u.a. (azul)).



(a) Vista frontal



(b) Vista lateral

Figura 5.7: Potencial electrostático para los pares de base centrales de la cadena de siete unidades TA del modelo ideal relajado (Escala: de -0.078585 u.a. (rojo) a 0.078585 u.a. (azul)).

Por otro lado, en el modelo ideal relajado se tiene que al menos uno de los pares de base de los extremos son uno de los más fuertes; mientras que para las cadenas de un par de base hasta la de cinco pares de base los puentes de hidrógeno más fuerte corresponden al par de base central. Esto no ocurre para las cadenas 9TA y 10TA, donde para 9TA los puentes de hidrógeno del par de base central tienden a debilitarse, al igual que algunos de sus alrededores, mientras que para 10TA tienden a fortalecerse. Al igual que en el análisis de las energías promedio de puente de hidrógeno, en donde había un alteración asociadas a dichas cadenas, este cambio en la fuerza de los puentes de hidrógeno puede estar relacionado el giro completo de la cadena.

Para el caso del modelo por bloques, se tiene la misma tendencia que en el modelo ideal relajado. Sin embargo, se encuentra que los puentes de hidrógeno están menos polarizados (son más débiles), esto se puede deber a que al estar más separados los pares de base entre sí, las interacciones $\pi - \pi$ no influyen de forma importante en la estabilización del puente de hidrógeno. Al igual que en el modelo ideal relajado, se tiene que los pares de base centrales se debilitan en comparación a las otras cadenas, lo cual, como fue mencionado anteriormente, se asocia al medio giro que da la cadena.

Finalmente, con el objetivo de complementar el análisis, en las Figuras 5.8, 5.9 y 5.10 se muestran los mapas de potencial electrostático para las cadenas 3TA, 5TA, 7TA y 9TA. Las zonas de color rojo corresponden a las zonas de mayor carga negativa, mientras que las regiones azules a las zonas de mayor carga positiva. Como se puede observar, para todos los modelos las regiones negativas se encuentran concentradas de forma más importante en las bases nitrogenadas, específicamente en los átomos involucrados en la formación de los puentes de hidrógeno. Por otra parte, para los modelos ideal no relajado

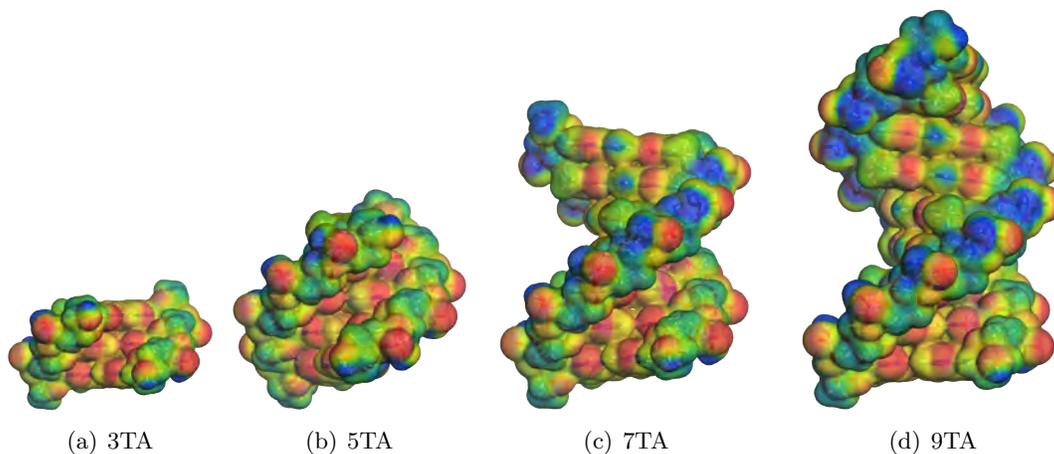


Figura 5.8: Mapas de potencial electrostático sobre la densidad electrónica de las cadenas del modelo ideal no relajado (Escala: de -0.078585 u.a. (rojo) a 0.078585 u.a. (azul)).

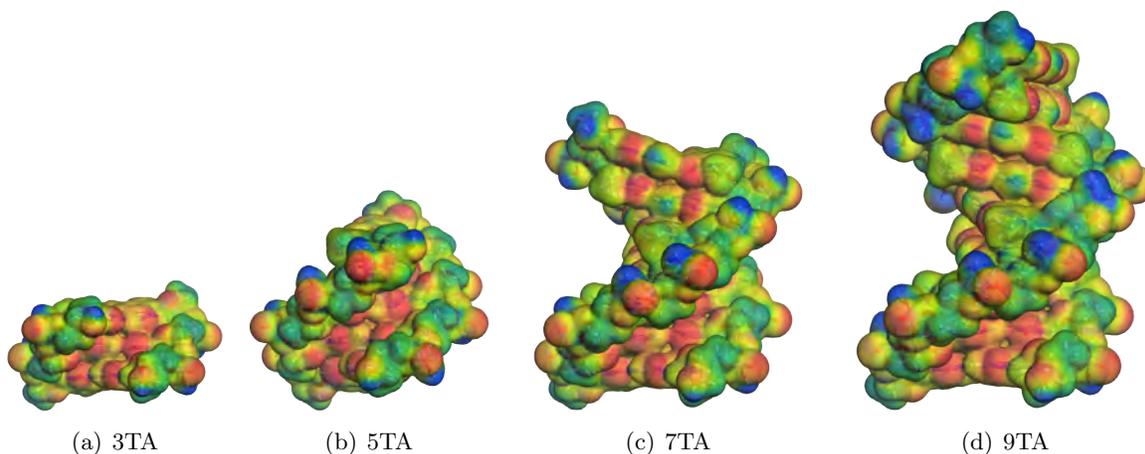


Figura 5.9: Mapas de potencial electrostático sobre la densidad electrónica de las cadenas del modelo ideal relajado (Escala: de -0.078585 u.a. (rojo) a 0.078585 u.a. (azul)).

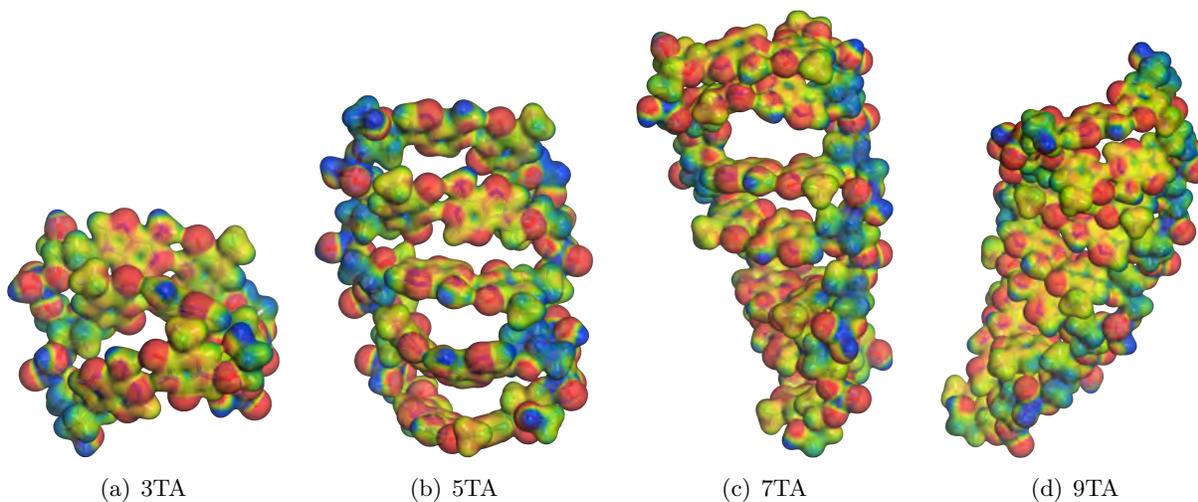


Figura 5.10: Mapas de potencial electrostático sobre la densidad electrónica de las cadenas del modelo por bloques (Escala: de -0.078585 u.a. (rojo) a 0.078585 u.a. (azul)).

y para el ideal relajado es evidente la presencia de las interacciones de apilamiento $\pi - \pi$, las cuales no están presentes en el modelo por bloques, debido al traslape de las nubes de densidad electrónica entre los pares de base. Además, se observa un cambio en la distribución de cargas entre el modelo ideal no relajado y el relajado, notándose en las regiones positivas (hidrógeno de los grupos fosfatos neutralizados) y en las regiones negativas, disminuyendo y aumentando, respectivamente, para el modelo ideal relajado, reafirmando la hipótesis de que la reorganización de los átomos de la cadena tiende a maximizar las interacciones presentes en la cadena.

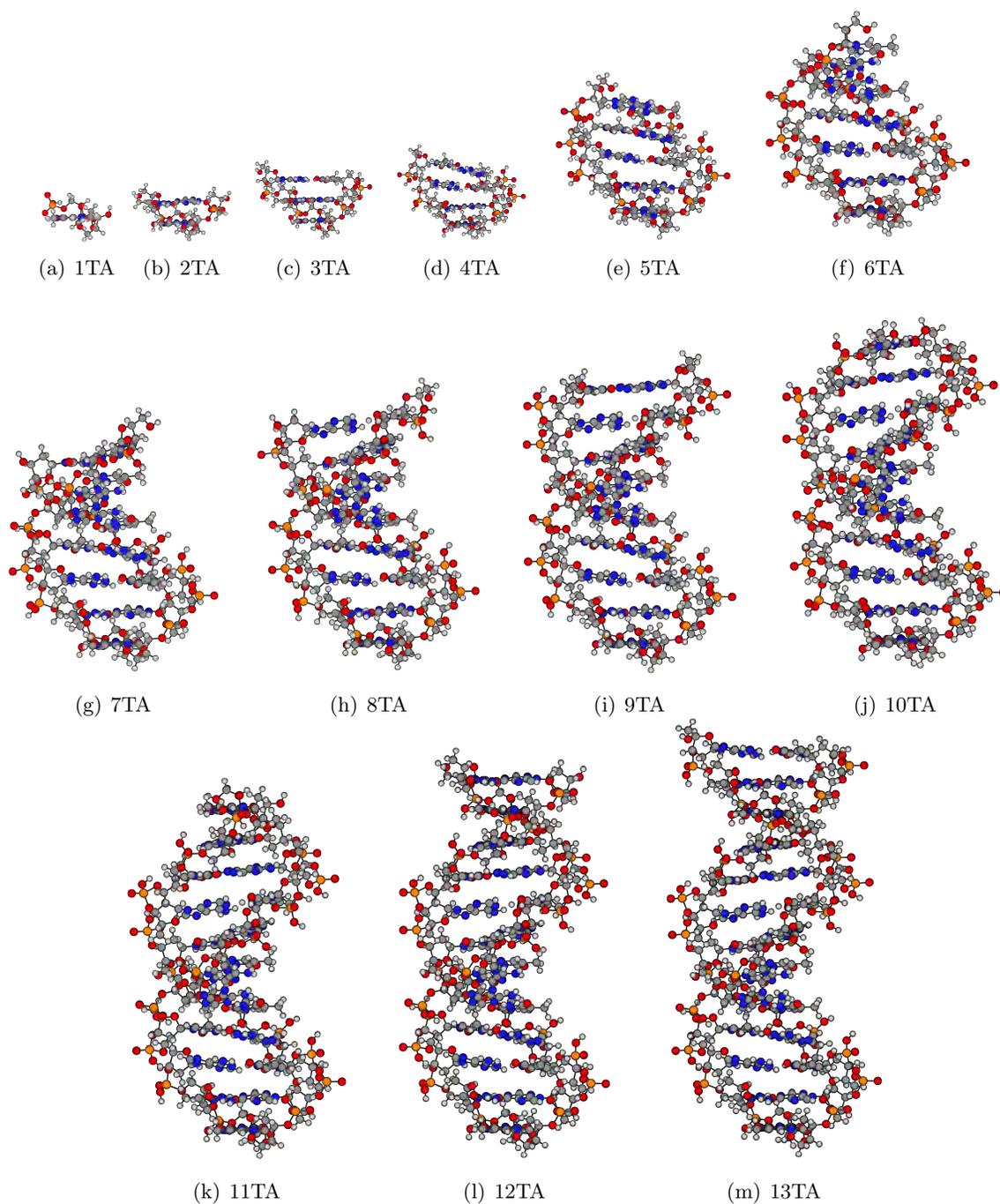


Figura 5.11: Vista lateral de las cadenas de ADN para el modelo ideal relajado (bases nitrogenadas del primer par de base eclipsadas).

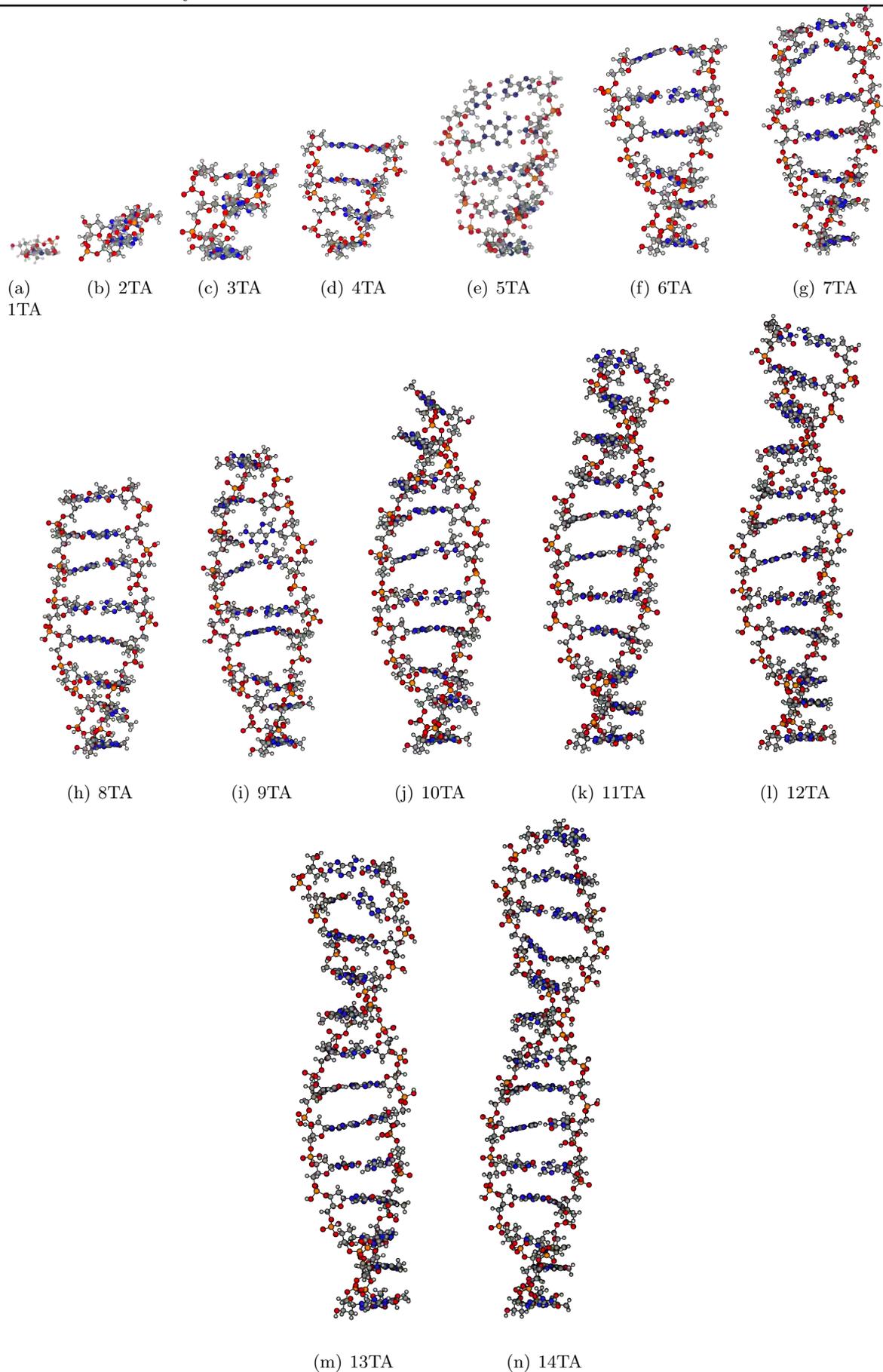


Figura 5.12: Vista lateral de las cadenas de ADN para el modelo por bloques (bases nitrogenadas del primer par de base eclipsadas).

Cadena	Par de base	Interacción	CSI			
			Ideal no relajado	Ideal relajado	Por bloques	
1TA	1	D_1	1.834	1.886	1.867	
		D_2	1.860	1.924	1.914	
		D_3	0.785	0.856	0.873	
	3TA	1	D_1	1.843	1.897	1.877
			D_2	1.890	1.946	1.847
			D_3	0.782	0.865	0.849
		2	D_1	1.845	1.864	1.867
			D_2	1.912	1.922	1.915
			D_3	0.786	0.852	0.880
5TA	3	D_1	1.843	1.861	1.863	
		D_2	1.888	1.919	1.904	
		D_3	0.796	0.855	0.872	
	1	D_1	1.841	1.896	1.866	
		D_2	1.890	1.927	1.929	
		D_3	0.781	0.852	0.883	
	2	D_1	1.847	1.859	1.843	
		D_2	1.912	1.917	1.925	
		D_3	0.787	0.843	0.878	
3	D_1	1.839	1.860	1.860		
	D_2	1.895	1.973	1.944		
	D_3	0.797	0.888	0.842		
4	D_1	1.831	1.861	1.825		
	D_2	1.902	1.972	1.860		
	D_3	0.784	0.887	0.854		
5	D_1	1.857	1.860	1.870		
	D_2	1.899	1.916	1.878		
	D_3	0.805	0.835	0.852		
7TA	1	D_1	1.841	1.887	1.896	
		D_2	1.890	1.925	1.895	
		D_3	0.782	0.850	0.872	
	2	D_1	1.846	1.864	1.844	
		D_2	1.911	1.942	1.895	
		D_3	0.786	0.860	0.852	
	3	D_1	1.836	1.857	1.831	
		D_2	1.895	1.970	1.924	
		D_3	0.796	0.885	0.896	
	4	D_1	1.832	1.856	1.808	
		D_2	1.899	1.973	1.885	
		D_3	0.788	0.886	0.864	
5	D_1	1.847	1.868	1.865		
	D_2	1.907	1.934	1.857		
	D_3	0.798	0.860	0.810		
6	D_1	1.830	1.867	1.779		
	D_2	1.902	1.933	1.548		
	D_3	0.780	0.848	0.682		
7	D_1	1.853	1.859	1.830		
	D_2	1.887	1.944	1.833		
	D_3	0.784	0.860	0.814		
9TA	1	D_1	1.841	1.880	1.875	
		D_2	1.891	1.945	1.872	
		D_3	0.782	0.867	0.849	
	2	D_1	1.845	1.866	1.831	
		D_2	1.911	1.961	1.938	
		D_3	0.787	0.877	0.870	
	3	D_1	1.836	1.863	1.875	
		D_2	1.895	1.962	1.934	
		D_3	0.797	0.883	0.876	
	4	D_1	1.831	1.976	1.847	
		D_2	1.899	1.951	1.921	
		D_3	0.789	0.860	0.868	
	5	D_1	1.845	1.869	1.855	
		D_2	1.908	1.933	1.892	
		D_3	0.797	0.843	0.864	
	6	D_1	1.831	1.855	1.849	
		D_2	1.898	1.938	1.890	
		D_3	0.783	0.868	0.838	
	7	D_1	1.842	1.852	1.846	
		D_2	1.897	1.978	1.923	
		D_3	0.779	0.892	0.882	
	8	D_1	1.840	1.857	1.868	
		D_2	1.904	1.973	1.917	
		D_3	0.807	0.892	0.907	
	9	D_1	1.838	1.857	1.846	
		D_2	1.908	1.936	1.931	
		D_3	0.798	0.852	0.890	
9	D_1	1.854	1.882	1.862		
	D_2	1.890	1.927	1.927		
	D_3	0.781	0.855	0.859		

Tabla 5.3: Índice de separación de cargas (CSI) para las regiones centrales de interacción entre A y T (Figura 5.4).

5.2. Propiedades reactivas

En esta sección se describirán las propiedades reactivas de la caja TATA a partir del uso de aproximaciones teóricas a la reactividad química (FMO y funciones de Fukui), con el objetivo de explicar la selectividad de las *proteína de unión a la caja TATA* hacia esta región. De acuerdo a diversas estructuras cristalinas [20,52,53], se sabe que el complejo TBP-caja TATA se origina cuando dos pares de fenilalanina, uno en cada dominio estructural de la molécula de TBP, doblan la hélice mediante la introducción de éstos entre los pares de base de los extremos de la caja TATA. De la literatura, es sabido que la secuencia consenso² para la caja TATA es: T-A-T-A-@-A-@-N (formada alrededor de ocho pares de base) [54]. Donde @ representa un par de base formado por A y T (sin importar su orientación: T-A o A-T) y N cualquier par de base. Por lo tanto, para esta sección se utilizará la cadena de ocho unidades TA para describir las propiedades químicas de la caja TATA, asumiendo que las bases adyacentes a esta región serán diferentes a A y T, descartando así efectos de borde. Para tener una aproximación realista de las propiedades reactivas de la caja TATA, se usarán únicamente las cadenas de los modelos ideal no relajado e ideal relajado.

5.2.1. FMO de la caja TATA

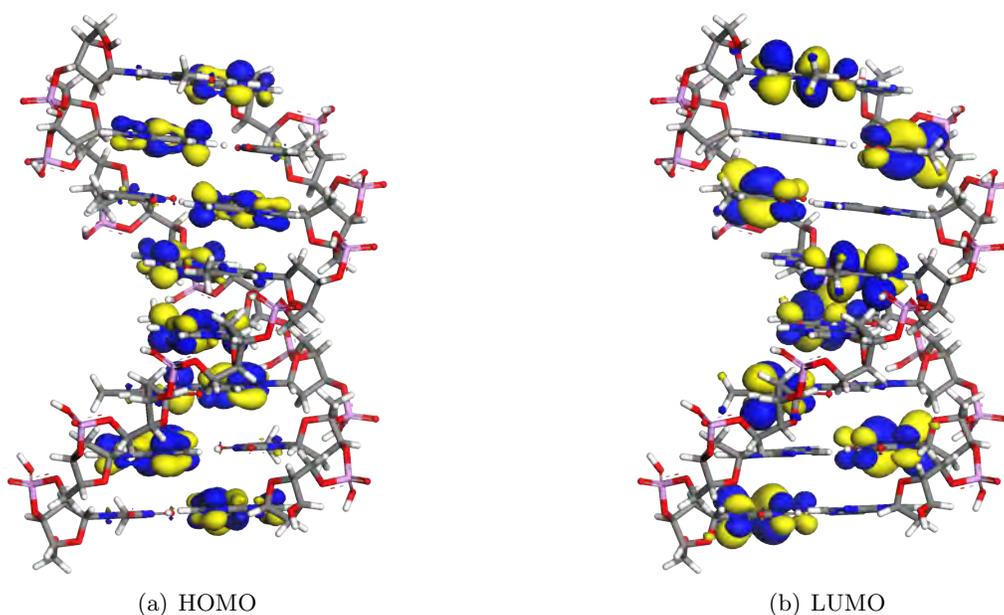


Figura 5.13: Orbitales Moleculares Frontera para la cadena 8TA del modelo ideal no relajado.

En las Figuras 5.13 y 5.14 se muestran los Orbitales Moleculares Frontera para la cadena de ocho pares de base de los modelos ideal no relajado e ideal relajado, respectivamente. Para ambos casos, de acuerdo a lo mencionado anteriormente, las cadenas cumplen con la forma consenso de la caja TATA. Al existir cierta simetría en las cadenas de ADN, no es de extrañarse que los FMO se encuentren degenerados con orbitales moleculares energéticamente cercanos a ellos.

Para el modelo ideal no relajado (Figura 5.13) se puede observar que el HOMO se encuentra localizado principalmente en las moléculas de adenina, mientras que el LUMO se ubica en las moléculas de timina. De forma similar, para el modelo ideal relajado (Figura 5.14) se puede ver que el HOMO también se

²Secuencia encontrada en diversos organismos.

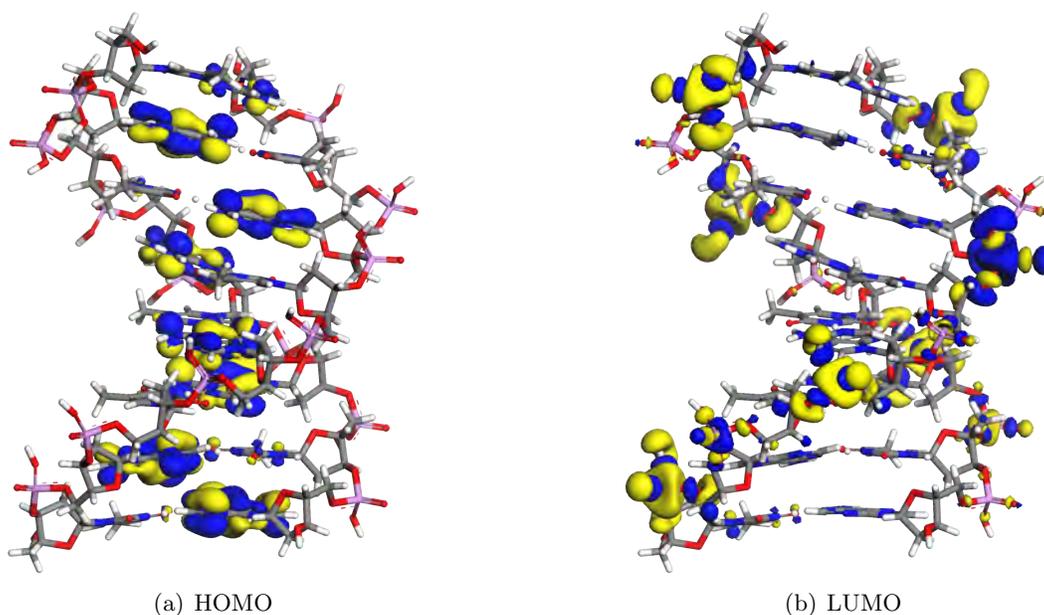


Figura 5.14: Orbitales Moleculares Frontera para la cadena 8TA del modelo ideal relajado.

localiza en las moléculas de adenina; sin embargo, el LUMO se localiza en los grupos fosfatos.

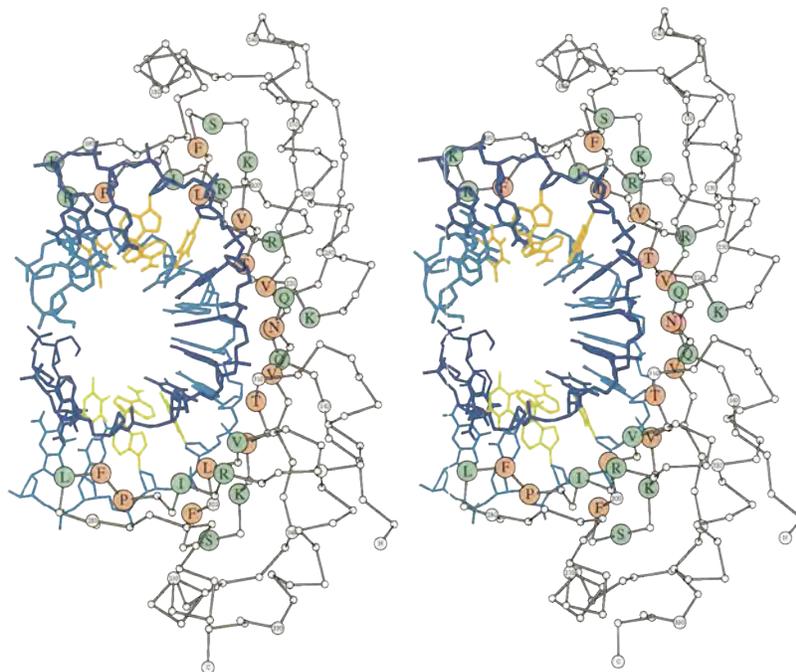


Figura 5.15: Representación gráfica del complejo TBP-TATA humano, donde la TBP se ubica del lado derecho y la secuencia CGTATATATACG del izquierdo. Los residuos de TBP que interactúan con las bases de la cadena (del lado del surco menor) se muestran de color rojo mientras que los residuos que interactúan con los grupos fosfatos y azúcares de color verde. La cadena con dirección 5' – 3' es de color azul fuerte, mientras que la complementaria de color azul claro. Los pares de base que interactúan con los pares de fenilalanina se encuentran de color amarillo (primer conjunto de pares de base) y de color naranja (último conjunto de pares de base). Figura tomada de la referencia [20].

Tomando como ejemplo a la TBP de humano [20], de acuerdo a lo que se sabe experimentalmente,

los residuos de fenilalanina Phe284 y Phe301 se intercalan entre los primeros y segundos pares de base, mientras que los residuos Phe193 y Phe210 entre el séptimo y octavo par de base, deformando a la cadena, permitiendo la interacción de otros residuos con otros pares de base y con grupos fosfatos y azúcares, dando paso a la formación del complejo de transcripción (Figura 5.15). Con base en lo dicho anteriormente y en la distribución de los FMO de los modelos ideal no relajado e ideal relajado, se puede decir que los Orbitales Moleculares Frontera obtenidos del análisis de la estructura del modelo ideal relajado describe de mejor forma la interacción entre TBP y la caja TATA; dando indicio de que el LUMO de la molécula de TBP se encuentra localizado en los pares de fenilalanina que se intercalan entre los pares de base extremos de la caja, en los cuales se encuentra localizado el HOMO; mientras que el HOMO de TBP se encuentra en los residuos que interactúan con los grupo fosfato de la cadena (puesto que ahí se localiza el LUMO). Es de esperarse que el LUMO de TBP se localice en los dos pares de fenilalanina que interactúan con la caja TATA, pues a esta molécula se le asocia una pseudo-simetría debido a la distribución casi idéntica de las estructuras secundarias en los dos dominios de la molécula [20] (ver Figura 5.15).

5.2.2. Funciones de Fukui

Para complementar el análisis de los FMO, en la Figuras 5.16 (modelo ideal no relajado) y 5.17 (modelo ideal relajado) se muestran las funciones de Fukui $f^-(\vec{r})$ y $f^+(\vec{r})$, las cuales indican las zonas probables para un ataque electrofílico y nucleofílico, respectivamente.

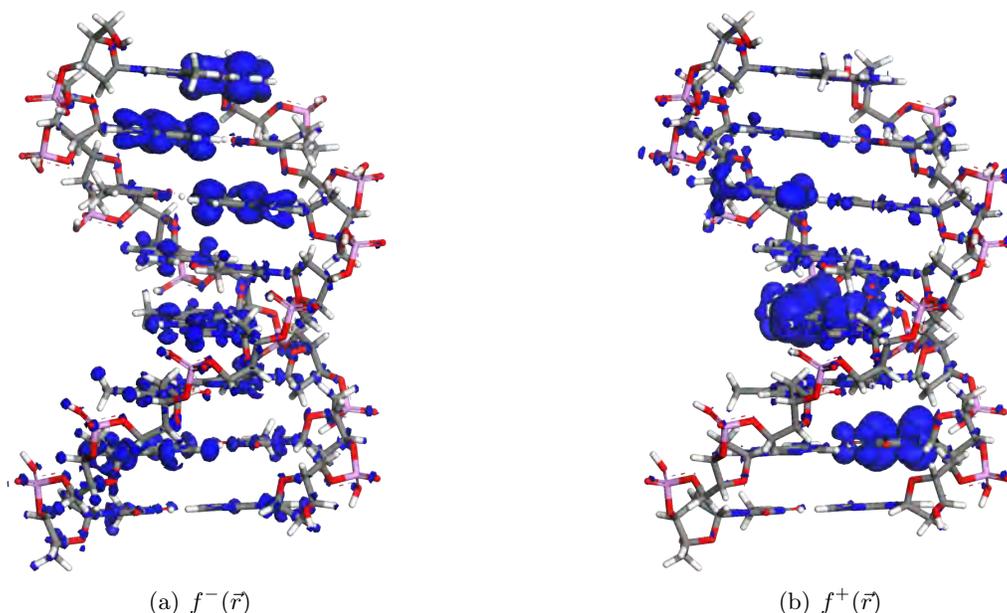


Figura 5.16: Descriptores químicos para la cadena 8TA del modelo ideal no relajado (a) representa la función de Fukui para un ataque electrofílico mientras que (b) representa la función de Fukui para un ataque nucleofílico.

Como se puede observar, para el modelo ideal no relajado (Figura 5.16) la función de Fukui $f^-(\vec{r})$ se encuentra localizada a lo largo de todas las bases nitrogenadas de los pares de base que conforman a la cadena, sin embargo, se localiza de forma más importante en las moléculas de adenina de los últimos tres pares de base. En cuanto a la función de Fukui $f^+(\vec{r})$, ésta se localiza de forma más importante sobre las moléculas de adenina del segundo y cuarto par de base. Por lo tanto, para este modelo, se tiene que los tres últimos pares de base son susceptibles a ataques de grupos electrofílicos, es decir, corresponde a las zonas que tienden a donar electrones; mientras que el segundo y cuarto par de base corresponde a zonas

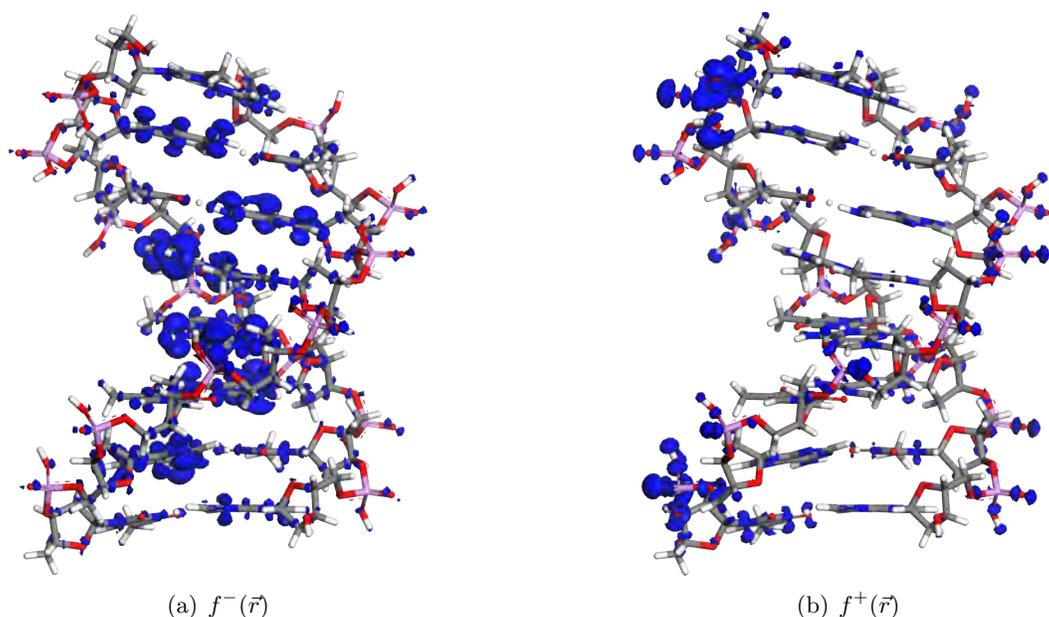


Figura 5.17: Descriptores químicos para la cadena 8TA del modelo ideal relajado (a) representa la función de Fukui para un ataque nucleofílico mientras que (b) representa la función de Fukui para un ataque electrofílico.

pobres en electrones, es decir, zonas susceptibles a sufrir ataques por grupos nucleofílicos.

Para el modelo ideal relajado (Figura 5.17) la función de Fukui $f^-(\vec{r})$ se encuentra localizada a lo largo de todas las bases nitrogenadas de la cadena, de forma más significativa, sobre las moléculas de adenina de los pares de base centrales; por consiguiente, estas zonas tienden a donar electrones. En cambio, la función de Fukui $f^+(\vec{r})$ se localiza sobre los grupos fosfato de la cadena, siendo zonas susceptibles a aceptar electrones.

Como fue mencionado en la sección anterior, ciertos residuos de la proteína de unión a la caja TATA interactúan con los pares de base de la cadena, mientras que otros residuos lo hacen con los grupos fosfato y azúcares. Por lo tanto, el modelo que mejor describe la interacción entre TBP y la caja TATA es el modelo ideal relajado, dado que las funciones de Fukui se encuentran distribuidas a lo largo de toda la molécula (pares de base y grupos fosfato), en tanto que en el modelo ideal no relajado las funciones se encuentran localizadas sobre los pares de base, no explicando por completo la interacción de los residuos con los grupos fosfato.

Por lo tanto con base en el análisis de los FMO, se tiene que los electrones más externos (los cuales pueden participar en alguna reacción o interacción con otra molécula) de la caja TATA se encuentran sobre las adeninas de la cadena; mientras que de acuerdo al análisis de carga (índice CSI) se tiene que al menos alguno de los extremos de la cadena cuenta con los puentes de hidrógeno más fuertes, es decir, hay una mayor concentración de densidad electrónica sobre los átomos que participan en la interacción. Por consiguiente, si experimentalmente se sabe que dos pares de fenilalanina se intercalan entre los pares de base extremos de la cadena (primero dos pares de base y últimos dos pares de base), con base en la función de Fukui $f^-(\vec{r})$ se puede decir que estos pares de fenilalanina prefieren interactuar con los pares de base de la cadena debido a que cuenta con los electrones más susceptibles a compartir, es decir, identifican a las regiones ricas en electrones. Por consiguiente, para que se lleve a cabo la interacción entre los residuos de fenilalanina y los pares de base correspondientes, la cadena sufre un ataque nucleofílico de los residuos, es decir, dona electrones para que se lleve a cabo la interacción.

Nuevamente, utilizando la información del análisis de los FMO se sabe que la región de más baja energía que puede aceptar electrones, es decir, la región donde se encuentra localizado el LUMO está sobre los grupos fosfato de la cadena. Por lo tanto, con base en la función de Fukui $f^+(\vec{r})$ se puede decir que los grupos fosfato interactúan con la molécula de TBP debido a que es la región que más fácilmente puede aceptar electrones, permitiendo la formación adecuada del complejo TBP-caja TATA.

Finalmente, resumiendo y conjuntando los análisis, la interacción entre la molécula TBP y la caja TATA puede ocurrir de la siguiente forma:

- a) Los residuos de fenilalanina primero interactúan con los pares de base extremos de la cadena de la caja TATA, intercalándose entre estos, alterando las interacciones $\pi - \pi$ y deformando la cadena, permitiendo la interacción de otros residuos con los demás pares de base. Aparentemente, las bases responsables de la interacción con los residuos son las adeninas.
- b) Una vez doblada la cadena, los residuos correspondientes de la TBP interactúan con los grupos fosfato, donando electrones para la formación del complejo.

Se puede pensar que al alterar la forma del dúplex de ADN de la caja TATA y de las interacciones $\pi - \pi$, los puentes de hidrógeno son debilitados permitiendo que se lleve a cabo el proceso de transcripción una vez que el complejo de transcripción se complete mediante la unión de la polimerasa.

Como trabajo a futuro, se propone analizar las propiedades reactivas de cadenas con guanina y citosina y comparar las propiedades reactivas de las cadenas, así como estudiar algunas de las propiedades reactivas de los residuos de TBP con el fin de complementar esta primera aproximación a la selectividad.

Capítulo 6

Conclusiones

En este trabajo se logró estudiar los puentes de hidrógeno involucrados en la estabilización de la cadena de ADN utilizando como base a la caja TATA. Para este fin se usaron tres modelos diferentes: modelo ideal no relajado (aquel que cumple con condiciones de simetría), modelo ideal relajado (estructuras relajadas del modelo anterior) y modelo por bloques (construcción de cadenas de ADN mediante la adición de pares de base sin restricciones de simetría). Tanto para el modelo ideal no relajado como para el modelo ideal relajado, la presencia de puentes de hidrógeno tiende a estabilizar la cadena de ADN, esto debido a que cuando se tiene más de tres pares de base, la energía de puente de hidrógeno tiende a ser mayor (más estable) que cuando se tiene un menor número de pares de base. Esto se relaciona a la estabilización conferida debido a la presencia de interacciones $\pi - \pi$ entre los pares de base, razón por la cual en el modelo ideal relajado se tiene puentes de hidrógeno de mayor fuerza, ya que al permitir la reorganización de los átomos se optimizan las interacciones no covalentes entre los pares de base, aumentando la energía de puente de hidrógeno entre las bases nitrogenadas. En cuanto al modelo por bloques, se tiene un comportamiento diferente al observado en los otros modelos, esto debido a que al ser construido y no tomar en cuenta las condiciones de simetría, el primer par de base se sobreestabiliza mientras que en las cadenas de mayor tamaño la energía de puente de hidrógeno es subestimada debido a la ausencia de interacciones $\pi - \pi$ entre los pares de base (la distancia entre ellos es grande y no permite la estabilización por dicha interacción).

Por otro lado, se tienen ligeras alteraciones en la energía de puente de hidrógeno promedio, distancias y la separación de carga en las cadenas de nueve y diez pares de base para el modelo ideal relajado y para la cadena de cinco unidades TA para el modelo por bloques. Esto puede estar relacionado con el giro y medio giro, respectivamente, que tienen las cadenas de los diferentes modelos. Se tendrían que hacer cadenas de mayor tamaño para ver que este comportamiento se repita y poder así concluir sobre el efecto del tamaño y la presencia de giros en la cadena.

En cuanto al modelo ideal no relajado, se tiene un comportamiento periódico en la energía promedio de puente de hidrógeno. Se propuso la siguiente ecuación para la estimación de la energía promedio de puente de hidrógeno de cadenas de diferente tamaño:

$$E_n = \begin{cases} \frac{1}{2} (0.037701 \text{ u.a.}) & n = \text{par} \\ \frac{1}{2} (0.037701 \text{ u.a.} - \frac{0.001020 \text{ u.a.}}{n}) & n = \text{impar} \end{cases} \quad (6.1)$$

Como se vio en la sección 5.1.1, el desarrollo analítico coincide adecuadamente con los valores calculados por lo que se puede ocupar para estimar energías promedio de puente de hidrógeno de cadenas de mayor tamaño tomando en cuenta condiciones de simetría.

Finalmente, utilizando la información proporcionada por los Orbitales Moleculares Frontera y las funciones de Fukui se logra hacer una primera aproximación de la selectividad de TBP a la caja TATA;

teniendo que el LUMO de la TBP (localizado en los residuos de fenilalanina) interactúa con el HOMO de la caja TATA (localizado en los pares de base, principalmente en las moléculas de adenina), mientras que el HOMO de la TBP (localizado en los residuos interactuantes con los grupos fosfatos, ver Figura 5.15) interactúa con el LUMO de la caja TATA (localizado en los grupos fosfato de la cadena). En cuanto a la información de las funciones de Fukui, se concluye que los pares de fenilalanina de TBP realizan un ataque nucleofílico a los pares de base extremos de la cadena, mientras que otros residuos realizan un ataque electrofílico a los grupos fosfato de la cadena. La ventaja de este tipo de análisis es que se pueden estimar de forma indirecta las propiedades reactivas de una proteína al conocer las propiedades reactivas de su molécula blanco.

Apéndice A

Bases numéricas¹

DMol³, implementado en la suite de programas Materials Studio, utiliza orbitales numéricos para las funciones base, cada función corresponde a un orbital atómico.

Las funciones base φ_μ son definidas numéricamente como valores de una malla esférica polar atómicamente centrada, en lugar de funciones analíticas, tales como los orbitales Gaussianos. La parte angular de cada función es el esférico armónico $Y_{lm}(\theta, \phi)$ apropiado. En cambio, la parte radial $F(r)$ es obtenida mediante la solución numérica de las ecuaciones de Kohn y Sham. Un nivel adecuado de precisión es, comúnmente, obtenido tomando en cuenta 300 puntos radiales desde el núcleo hasta una distancia exterior de 5.3 Å.

Las funciones radiales son almacenadas como un conjunto de coeficientes de spline cúbicos para cada una de las 300 secciones, de modo que $F(r)$ es una función analítica definida por intervalos. Esto último es una consideración importante para la generación de los gradientes analíticos de energía. Además del conjunto de funciones, el término $\nabla^2/2$ requerido para la evaluación de la energía cinética son también almacenados como coeficientes de spline.

Los conjuntos de bases atómicas son confinados con un valor de corte, r_c , apropiado para un determinado nivel del cálculo de DMol³. Ésta es una característica importante del conjunto de bases numéricas que puede conducir a cálculos más rápidos, especialmente para sistemas de estado sólido. Asimismo, el código DMol³ utiliza un potencial de confinamiento suave, el cual asegura la estricta localización del conjunto de bases dentro de un valor r_c , sin discontinuidades en las derivadas en el punto r_c . Los procesos de optimización de geometría son eficientes, incluso con valores de corte pequeños.

A.1. Ventajas de los conjuntos de base derivados numéricamente

El uso de los orbitales atómicos esféricos exactos DFT tiene diversas ventajas. Uno de ellos es que una molécula puede ser disociada exactamente en sus átomos constituyentes (dentro del marco de DFT). Debido a la calidad de estos orbitales, el efecto de superposición de bases es minimizado y es posible obtener una excelente descripción, incluso para enlaces débiles.

¹Sección basada en el manual de Materials Studio.

Bibliografía

- [1] HHMI. DNA Timeline. [En línea] Obtenido del World Wide Web: <http://www.dnai.org/timeline/>, 2003. [accesado el 8 de septiembre de 2014].
- [2] J.D. Watson y F.H.C. Crick. A Structure for Deoxyribose Nucleic Acid. *Nature*, 171:737, 1953.
- [3] David Goodsell. DNA. [En línea] Obtenido del World Wide Web: <http://www.rcsb.org/pdb/101/motm.do?momID=23>, 2001. [accesado el 20 de julio de 2013].
- [4] M. B. Gerstein, C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Welsman, y M. Snyder. What is a gene, post-ENCODE? History and updated definition. *Genome Research*, 17:669, 2007.
- [5] Gerald Karp. *Biología Celular y Molecular: Conceptos y Experimentos*. McGraw-Hill, 2009.
- [6] Laszlo Tora y Timmers H.Th.Marc. The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Cell Press*, 6:209, 2010.
- [7] L. Pardo, M. Campillo, y D. Bosch. Binding Mechanisms of TATA Box-Binding Proteins: DNA Kinking is Stabilized by Specific Hydrogen Bonds. *Biophysical Journal*, 78:1988, 2000.
- [8] Kensal E. van Holde, W. Curtis Johnson, y P. Shing Ho. *Principles of Physical Biochemistry*. Prentice Hall, 1998.
- [9] G. N. Lewis. *Valence and the Structure of Atoms and Molecules*. Chemical Catalog Co., 1923.
- [10] F. Weinhold y R.A. Klein. What is a hydrogen bond? Mutually consistent theoretical and experimental criteria for characterizing H-bonding interactions. *Molecular Physics*, 110:565, 2012.
- [11] H. Szatyłowicz y N. Sadlej-Sosnowska. Characterizing the Strength of individual Hydrogen Bonds in DNA Base Pairs. *Journal of Chemical Information and Modeling*, 50:2151, 2010.
- [12] A. Asensio, N. Kobko, y J.J. Dannenberg. Cooperative Hydrogen-Bonding in Adenine-Thymine and Guanine Cytosine Base Pairs. Density Functional Theory and Møller-Plesset Molecular Orbital Study. *Journal of Physical Chemistry A*, 107:6441, 2003.
- [13] J. Grunenberg. Direct Assessment of Interresidue Forces in Watson-Crick Base Pairs Using Theoretical Compliance Constants. *Journal of the American Chemical Society*, 126:16310, 2004.
- [14] H. Dong, W. Hua, y S. Li. Estimation on the individual Hydrogen-Bond Strength in Molecules with Multiple Hydrogen Bonds. *Journal of Physical Chemistry A*, 111:2941, 2007.
- [15] R. Gillespie y P. L. A. Popelier. *Chemical Bonding and Molecular Geometry*. Oxford University Press, 2001.
- [16] E. Espinosa, E. Molins y C. Lecomte. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chemical Physics Letters*, 285:170, 1998.

- [17] C. Fonseca, F. Matthias, J. Snijders, y E. Jan. Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment. *Journal of American Chemical Society*, 122:4117, 2000.
- [18] C. F. Matta y N. Castillo. Extended weak bonding interactions in DNA: π -stacking (base-base), base-backbone, and backbone-backbone interactions. *Journal of Physical Chemistry B*, 110:563, 2006.
- [19] I. Saito, T. Nakamura, y K. Nakatani. Mapping of Highest Occupied Molecular Orbitals of Duplex DNA by Cobalt-Mediated Guanine Oxidation. *Journal of the American Chemical Society*, 122:3001, 2000.
- [20] Z. S. Juo, T. K. Chiu, P. M. Leiberman, I. Baikalov, A. J. Berk, y E. Dickerson. How Proteins Recognize the TATA Box. *Journal of Molecular Biology*, 261:239, 1996.
- [21] W. Koch y M. C. Hothausen. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, 2001.
- [22] I. Levine. *Química Cuántica*. Prentice Hall, 2001.
- [23] A. Szabo y N. S. Ostlund. *Modern Quantum Chemistry*. Dover Publications, Inc., 1996.
- [24] P. Hohenberg y W. Kohn. Inhomogeneous Electron Gas. *Physical Review Letters*, 136:B864, 1964.
- [25] A. D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140:18A301, 2014.
- [26] W. Kohn y L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140:A1133, 1965.
- [27] A. J. Cohen, P. Mori-Sánchez, y W. Yang. Challenges for Density Functional Theory. *Chemical Reviews*, 112:289, 2012.
- [28] G. Cuevas y F. Cortés. *Introducción a la Química Computacional*. Fondo de Cultura Económica, 2003.
- [29] J. P. Perdew, A. Ruzsinsky, J. Tao, V. N. Staroverov, y G. E. Scuseria. Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits. *The Journal of Chemical Physics*, 123:062201, 2005.
- [30] D. M. Ceperley y B. J. Alder. Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters*, 45:566, 1980.
- [31] S. H. Vosko, L. Wilk, y M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58:1200, 1980.
- [32] Y. Zhao y D. G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics*, 125:194101, 2006.
- [33] Y. Jean y F. Volatron. *An Introduction to Molecular Orbitals*. Oxford University Press, year = 1993,.
- [34] R. Chang. *Fisicoquímica para las Ciencias Biológicas*. McGraw-Hill, 2008.
- [35] P. K. Chattaraj. *Chemical Reactivity Theory: a density functional view*. CRC Press, 2009.
- [36] R. G. Parr y W. Yang. Density Functional Approach to the Frontier-Electron Theory of Chemical Reactivity. *Journal of the American Chemical Society*, 106:4049, 1984.

- [37] R. F. Nalewajski y R. G. Parr. Legendre Transforms and Maxwell Relations in Density Functional Theory. *The Journal of Chemical Physics*, 77:399, 1982.
- [38] J. P. Perdew, R. G. Parr, M. Levy, y J. L. Jr. Balduz. Density-Functional Theory for Fractional Particle Number: Derivate Discontinuities of the Energy. *Physical Review Letters*, 49:1691, 1982.
- [39] J. P. Perdew y T. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45:13244, 1992.
- [40] I. K. Yanson, A. B. Teplitsky, y L. F. Sukhodub. Experimental studies of molecular interactions between nitrogen bases of nucleic acids. *Biopolymers*, 18:1149, 1979.
- [41] T. van der Wijst, C. Fonseca, M. Swart, y M. Bickelhaupt. Performance of various density functionals for the hydrogen bonds in DNA base pairs. *Chemical Physics Letters*, 426:415, 2006.
- [42] A. Klamt y G. Schüürmann. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *Journal of the Chemical Society, Perkin Transactions 2*, 5:799, 1993.
- [43] B. Delley. The conductor-like screening model for polymers and surfaces. *Molecular Simulation*, 3:117, 2006.
- [44] C. R. Calladine, H. R. Drew, B. F. Luisi, y A. A. Travers. *Understanding DNA: the molecule and how it works*. Elsevier Academic Press, 2004.
- [45] make-na. © 2006, 2011, 2012, 2013. James Stroud.
- [46] Roy Dennington, Todd Keith, y John Millam. GaussView Version 5. Semichem Inc. Shawnee Mission KS 2009.
- [47] B. Delley. From molecules to solids with the DMol3 approach. *Journal of Chemical Physics*, 113:7756, 2000.
- [48] Materials Studio, Accelrys. © 2001-2011 Accelrys Software Inc.
- [49] N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, y A. Rich. RNA double-helical fragments at atomic resolution: I. The crystal and molecular structure of sodium adenylyl-3',5'-uridine hexahydrate. *Journal of Molecular Biology*, 104:109, 1976.
- [50] E. R. Collantes y W. J. Dun III. Amino Acid Side Chain Descriptors for Quantitative Structure-Activity Relationship Studies of Peptide Analogues. *Journal of Medicinal Chemistry*, 38:2705, 1995.
- [51] C. F. Matta y 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: Structure. Function and Genetics*, 52:360, 2003.
- [52] J. L. Kim, D. B. Nikolov, y S. Burley. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, 365:520, 1993.
- [53] Y Kim, J. H. Geiger, S. Hahn, y P. B. Sigler. Crystal structure of a yeast TBP/TATA-box complex. *Nature*, 365:512, 1993.
- [54] P. Bucher. Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of Molecular Biology*, 212:563, 1990.