

00362



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS FÍSICAS

INSTITUTO DE FÍSICA

“ISÓMEROS ESTABLES DE LAS BASES DEL ADN”

T E S I S

Que para obtener el grado de MAESTRO EN CIENCIAS (FÍSICA)

PRESENTA:

FÍS. JOSÉ VEGA MUNGUÍA

Director de Tesis:

Dr. KARO MICHAELIAN PAUW

Ciudad Universitaria, México D.F.

2002



posgrado en ciencias físicas unam

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: José Vega Munguía

FECHA: 26 de Agosto de 2002

FIRMA: José Vega Munguía

TESIS CON FALLA DE ORIGEN



Universidad Nacional
Autónoma de México

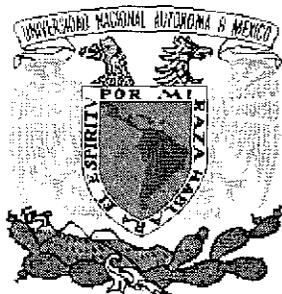


UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

*División de Estudios de Posgrado
Posgrado en Ciencias Físicas
Instituto de Física*

ISÓMEROS ESTABLES DE LAS BASES
DEL ADN

TESIS QUE PARA OBTENER EL GRADO
DE:

Maestro en Ciencias

P r e s e n t a

José Vega Munguía

D i r e c t o r:

Dr. Karo Michaelian Pauw

Ciudad Universitaria, México D. F. 2002



posgrado en ciencias físicas
u n a m

TESIS CON
FALLA DE ORIGEN

I MET A TRAVELLER FROM AN ANTIQUE LAND
WHO SAID... TWO VAST AND TRUNKLESS LEGS OF STONE
STAND IN THE DESERT ... NEAR THEM, ON THE SAND,
HALF SUNK. A SHATTERED VISAGE LIES, WHOSE FROWN,
AND WRINKLED LIP, AND SNEER OF COLD COMMAND
TELL THAT ITS SCULPTOR WELL THOSE PASSIONS READ
WHICH YET SURVIVE, STAMPED ON THESE LIFELESS THINGS
THE HAND THAT MOCKED THEM, AND THE HEART THAT FED-
AND ON THE PEDESTAL THESE WORDS APPEAR
MY NAME IS OZYMANDIAS, KING OF KINGS
LOOK ON MY WORKS YE MIGHTY AND DESPAIR.
NOTHING BESIDE REMAINS, ROUND THE DECAY
OF THAT COLOSSAL WRECK. BOUNDLESS AND BARE
THE LONE AND LEVEL SANDS STRETCH FAR AWAY

SHELLEY (en *Gaussian98*)

THE ONE-EYED VIEW OF OUR UNIVERSE SAYS YOU
MUST NOT LOOK FAR AFIELD FOR PROBLEMS.
SUCH PROBLEMS MAY NEVER ARRIVE.
INSTEAD TEND TO THE WOLF WITHIN YOUR FENCES.
THE PACKS RANGING OUTSIDE MAY NOT EVEN EXIST.

THE AZHAR BOOK SHAMRA I:4
CHILDREN OF DUNE, BY FRANK HERBERT (en *Gaussian98*)

Dedicatoria:

Especialmente a mis padres. Ing. José Adolfo Juvenal Vega Martell y Sofía Munguía Bravo, gracias por todo su apoyo, afecto y cariño.

A mis hermanos M.I. Elio Vega, Biól. Lila Vega y Nadia Jasel Vega.

A mis tíos Hugo, Lila, Virginia, Octavio, Dolores y Reyes, a mis primos y abuelitos.

A mis compañeros y amigos Juan Andrés Reyes, Manuel V., a todos los miembros del grupo de Din. Mol.: Dr. Ignacio Garzón y Dra. Marcela Beltrán.

A mi asesor Dr. Karo Michaelian Pauw, gracias por todo su apoyo para terminar bien este trabajo.

Al Instituto de Física de la UNAM.

A la UNAM, máxima casa de estudios.

INDICE GENERAL
MOTIVACION Y OBJETIVOS

1 LAS BASES DEL ADN 4

1.1 Introducción 4

1.2 Código Genético 11

1.3 Mutaciones en el código genético 13

1.4 Nucleótidos en otras moléculas biológicas 16

2 EL POTENCIAL UTILIZADO 19

2.1 Introducción 19

2.2 El campo de fuerza 23

2.3 Componentes del campo de fuerza 23

2.4 Variación de longitud de enlace 24

2.5 Doblamiento de ángulo 24

2.6 Torsion 26

2.7 Doblamientos fuera del plano 28

2.8 Interacciones de Van der Waals 28

2.9 Parametrización 32

2.10 Términos específicos 36

2.11 Enlaces de hidrógeno y efectos del medio 37

2.12 Energía de conformación 37

2.13 El modelo AMBER 38

3 LA BÚSQUEDA CON EL ALGORITMO GENÉTICO 42

3.1 Introducción 42

3.2 El algoritmo genético 44

3.3 La búsqueda global con el Algoritmo Genético 46

3.4 Resultados 53

3.5 Mínimos del potencial que convergieron a estructuras estables
con una posterior optimización local 67

TESIS CON
FALLA DE ORIGEN

4	OPTIMIZACIÓN CON CÁLCULOS CUÁNTICOS	86
4.1	Justificación	86
4.2	Campos auto consistentes	87
4.3	Relajación con cálculos de primeros principios	90
4.4	MP2	92
4.5	Justificación de cálculos MP2	93
4.6	Resultados	95
4.7	Resultados para adenina	96
4.8	Resultados para guanina	106
4.9	Resultados para citosina	128
4.10	Resultados para timina	137
5	CONSECUENCIAS Y CONCLUSIONES	138
5.1	Posibles consecuencias de la existencia de nuevos isómeros . . .	138
5.2	Efectos en mutaciones	143
5.3	Mutación+selección=evolución	149
5.4	Conclusiones	151
6	APÉNDICES	154
6.1	Apéndice A: Glosario breve de términos	154
6.2	Apéndice B: Descripción del ajuste por mínimos cuadrados para la obtención de los parámetros del potencial utilizado . . .	157
6.3	Apéndice C: Métodos de optimización cuánticos	163

BIBLIOGRAFIA

RESUMEN

Existen cuatro bases nitrogenadas del ADN: adenina, guanina, citosina y timina. El primer capítulo trata acerca de estas moléculas, se muestran las estructuras conocidas en la naturaleza, las cuales son planas, así como su función importantísima que estas tienen en los seres vivos, ya que en base al orden en que están dispuestas estas bases en los nucleótidos del ADN es como se codifica cada tipo de proteína que se producirá en la célula, además de ser también componentes de otras moléculas biológicas importantes.

Realizando una búsqueda en diversas fuentes de información, no encontramos reportes de que estas moléculas tuvieran isómeros diferentes de las conformaciones planas; de existir otros isómeros (lo cual es probable), estos podrían tener consecuencias importantes para los sistemas biológicos ya que tendrían un comportamiento químico distinto a las configuraciones conocidas; aunque actualmente no se conocen otros isómeros de las bases del ADN, esto no significa que no existen en la naturaleza, tal vez a lo largo de millones de años de historia biológica, estos isómeros desconocidos de las bases han intervenido en el origen de cambios de la información celular dando origen a mutaciones que tendrían consecuencias en la evolución de las especies así como en posibles enfermedades. De lo anterior surgió nuestro interés en investigar la posible existencia en la naturaleza de otras conformaciones distintas a las ya conocidas, ya que de encontrar nuevas conformaciones de las bases del ADN, estas serían las primeras reportadas.

El desarrollo de la búsqueda de los isómeros estables consistió en hacer primero una búsqueda global con un algoritmo genético de los mínimos locales de un potencial; este potencial describe las interacciones entre los átomos de las bases del ADN. Posteriormente a las configuraciones interesantes encontradas les aplicamos métodos de relajación cuánticos. El trabajo de investigación requirió un gran componente computacional. Deseando ir mas allá de meramente presentar los resultados, se tratarán de exponer en los siguientes capítulos de la tesis los fundamentos de los 3 principales recursos utilizados: la formulación y desarrollo del potencial, la forma en que se

implementó el algoritmo genético, el cual es uno de los métodos de búsqueda global que más éxito ha tenido últimamente por su eficacia y finalmente una breve introducción a los cálculos cuánticos *ab-initio* desarrollados por Gaussian98. En el último capítulo se presentarán los resultados y conclusiones. A continuación describiremos brevemente estos capítulos

En el segundo capítulo se expone la manera en que se construyó el potencial utilizado para representar la interacción entre átomos en las bases: Partiendo de aproximaciones elementales (tales como representar a la energía de la molécula igual a la suma de las contribuciones armónicas por alargamiento de enlace, doblamiento de ángulo y torsiones; puentes de hidrógeno: términos coulombianos y de Lennard-Jones), trataremos de mostrar el método empírico llamado "Campo de fuerza consistente", el cual fué desarrollado por Lifson y Warshel (en la tesis se encuentran todas las referencias). En esta aproximación la función potencial se escoge no solamente para ajustar las frecuencias vibracionales, si no también las estructuras de conformación y energías de muchas moléculas orgánicas. La palabra "consistente" justamente significa que la función potencial es consistente con estos conjuntos de datos. En este capítulo se muestra como se utilizan la ecuación secular de modos normales para análisis vibracional y las frecuencias de vibración en los cálculos energéticos para su posterior utilización en el método de mínimos cuadrados para ajuste de los parámetros del potencial.

En el tercer capítulo mostramos una descripción del algoritmo genético que se implementó para encontrar los mínimos locales del potencial. esta es una técnica de optimización global desarrollada en los 60's por John H. Holland y diseñada en analogía a la teoría de evolución por selección natural de Darwin. Por medio de tres operaciones muy simples, el entrecruzamiento, mutación y selección, operando sobre poblaciones de individuos (llamados cromosomas, los cuales resultan de la codificación de las coordenadas cartesianas de los átomos de las bases del ADN generadas al azar), es posible llegar a encontrar los mejores (en nuestro caso, las configuraciones de más baja energía). Hemos visto a lo largo de la evolución natural que la sobre-

vivencia de las especies a ambientes adversos y cambiantes ha llevado a la continua especialización de las mismas, siendo finalmente aquellos individuos mejor dotados los que logran triunfar y reproducirse. Con esta técnica se encontraron muchas configuraciones mínimas, al final del capítulo se muestran aquellas que con una posterior optimización cuántica local convergieron a estructuras estables.

El objetivo de la química cuántica computacional *ab initio* es el tratar de determinar las energías electrónicas y funciones de onda de átomos, moléculas y demás especies químicas. La frase *ab initio* implica que uno intenta resolver la ecuación de Schroedinger de primeros principios, tratando la molécula como una colección de núcleos positivos y electrones negativos moviéndose bajo la influencia de potenciales coulombianos; Un uso práctico de tal punto de vista requiere que ciertas aproximaciones sean introducidas, ya que a la ecuación de Schroedinger aún no se le ha encontrado una solución analítica. Los cálculos cuánticos para realizar la optimización local de las moléculas fueron realizados en las aproximaciones de Hartree-Fock más MP2 para incluir la correlación electrónica. Para optimizar la posición de los núcleos se utiliza el algoritmo de Berny.

En el cuarto capítulo se muestran algunos isómeros estables dentro de estas aproximaciones. Gaussian98 es un programa muy sofisticado, también en este capítulo de la tesis mostramos como a través de la técnica de campo auto consistente el programa interactúa hasta que los orbitales moleculares y sus eigenvalores sean consistentes. Aquí utilizamos el modelo de Hartree-Fock para resolver la ecuación de Schroedinger para un simple electrón moviéndose en un potencial donde los movimientos complicados de los restantes electrones son “promediados”. En el caso de moléculas se usa el modelo LCAO (Linear Combination of Atomic Orbitals), donde cada orbital molecular puede ser escrito como una combinación lineal de n orbitales atómicos, es importante recordar que en este caso estamos trabajando dentro de la aproximación de Born-Oppenheimer, con los núcleos fijos en sus posiciones en el espacio.

En el método de campo autoconsistente los electrones se mueven en un

potencial promedio originado por los otros electrones y así la posición instantánea de un electrón no es influida por la presencia de un electrón vecino. De hecho, los movimientos de los electrones están correlacionados y ellos tienden a evitarse uno a otro mucho más de lo que la teoría de Hartree-Fock sugeriría, dando lugar a una energía menor, un posterior refinamiento utilizando MP2 se aplicó a las estructuras que convergieron con Hartree-Fock a formas estables, en MP2 uno desarrolla una ecuación de Schroedinger modelo, la cual en algún sentido es diseñada para representar el sistema cuya ecuación de Schroedinger completa no puede resolverse. La diferencia entre el hamiltoniano completo y el modelo Hartree-Fock, H y H^o respectivamente, es usado para definir una perturbación $V = H - H^o$. La teoría de perturbaciones es entonces empleada para aproximar los niveles de energía y funciones de onda del completo hamiltoniano H en términos de los niveles de energía y funciones de onda del sistema modelo H^o de Hartree-Fock (los cuales se asumen que pueden ser encontradas). La imperfección en el problema modelo es por lo tanto usada como la perturbación. El éxito de tal aproximación depende fuertemente de que tan bien el modelo Hartree-Fock representa el problema verdadero, esto es, que tan pequeño es V . Aquí también utilizamos el programa Gaussian98.

Finalmente, en el quinto capítulo se discuten las posibles consecuencias y las conclusiones, se presenta la unión estables de un isómero de la adenina unido con la citosina y como esto afectaría al proceso de duplicación y transcripción.

ISÓMEROS ESTABLES DE LAS BASES DEL ADN

José Vega Munguía
Instituto de Física, UNAM
A.P. 20-364, 01000 México D.F., México.

Asesor:

Dr. Karo Michaelian Pauw.
Instituto de Física, UNAM.

Participantes:

Dr. Ignacio L. Garzón Sosa.
Instituto de Física, UNAM.
Dra. Marcela R. Beltrán Sánchez.
Instituto de Inv. en Materiales, UNAM.

Junio, 2002.

MOTIVACION

Las moléculas biológicas adenina, citosina, guanina y timina son las bases del ácido nucleico ADN, la molécula esencial de la vida y la evolución, existe también otra base, Uracil, pero esta solo aparece en el RNA, que es una molécula para transcripción, y no vamos a discutir mas en este trabajo. Utilizando estas cuatro bases el ADN codifica toda la información necesaria para dirigir las actividades celulares. Es el ordenamiento de estas moléculas en el ADN lo que produce la especificidad de los organismos vivos, en consecuencia la transferencia de información depende mucho de la actividad biológica y química de estas moléculas. Aparentemente en la naturaleza hay una única configuración planar para las bases; si existieran otras configuraciones distintas de las conocidas estas tendrían un comportamiento químico muy distinto y traería consecuencias muy importantes para la vida. Es aquí donde surgió la motivación de este trabajo: ¿las bases biológicas tienen otros isómeros?, y si es así. ¿porqué aparentemente en la naturaleza hay una sola configuración?. El responder estas preguntas podría ser útil en estudios biológicos o de enfermedades genéticas.

OBJETIVOS

1. Búsqueda global en la superficie de un potencial propuesto para las bases del ADN
2. Relajación cuántica de configuraciones interesantes.
3. Investigación preliminar de las configuraciones que sobrevivan relajación cuántica y sus posibles implicaciones biológicas.
4. Tratar de investigar la posible unión entre dos nuevas conformaciones.

Esta investigación es la continuación de un trabajo empezado en licenciatura con adenina, lo cual está publicado en: "Isomers of Adenine", J Vega, K. Michaelian, L. Hernández, M.R. Beltrán, I.L. Garzón, *Journal of Molecular Structure*, THEOCHEM. 493 (1999) 275-285.

Capítulo 1

LAS BASES DEL ADN

1.1 Introducción

El análisis químico del ADN muestra que pertenece a la clase de moléculas denominadas ácidos nucleicos. Cada molécula de ADN está compuesta de unidades llamadas nucleótidos (más específicamente, desoxirribonucleótidos) cada uno de los cuales tiene tres componentes principales: un fosfato, un azúcar de 5 carbonos, la desoxirribosa, y una base nitrogenada (figura 1.1). Las porciones de fosfato y azúcar del nucleótido son una característica constante en cualquier ADN, pero hay cuatro variedades de bases nitrogenadas: adenina, guanina, citosina y timina y (figura 1.2). La adenina y la guanina son estructuralmente similares y constan de dos anillos fusionados: se denominan purinas. La timina y la citosina, tienen un solo anillo, son estructuralmente semejantes, pero diferentes de las purinas y se llaman pirimidinas. Todas las bases son planas en su configuración natural. La configuración espacial o tridimensional de los nucleótidos se relacionan con sus funciones. Así, para comprender las propiedades funcionales del ADN, es necesario comprender el ordenamiento espacial de los nucleótidos en la molécula. Como veremos, la naturaleza tridimensional del ADN es la clave para saber realmente como funciona. Como ocurre con las proteínas que están compuestas de largas cadenas de aminoácidos, el ADN consta de nucleótidos enlazados en una larga cadena. El problema de la estructura del ADN fue resuelto en 1953 por Maurice Wilkins, James Watson y Francis Crick.

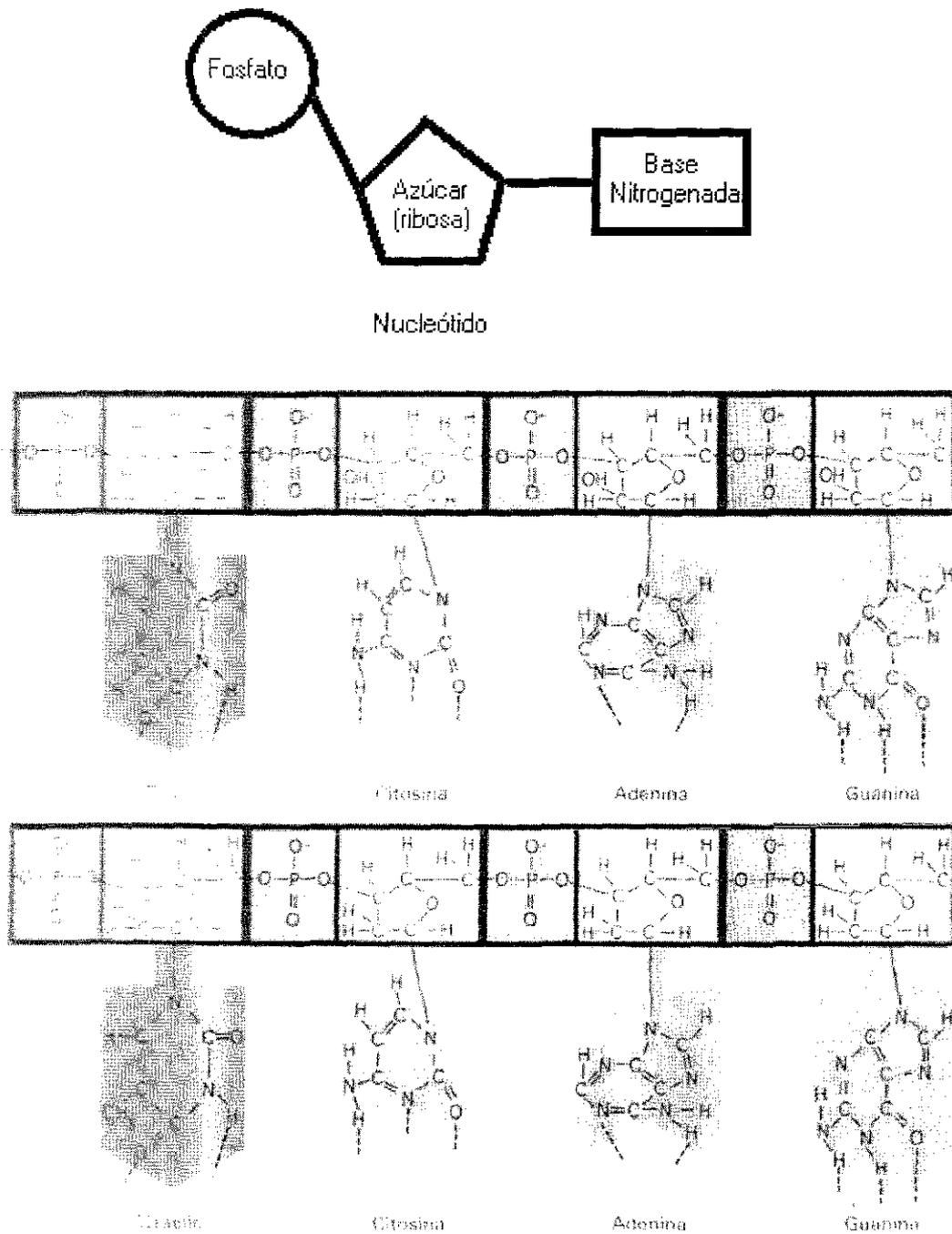


Figura 1.1: Estructura de los nucleótidos (Desoxirribonucleótido). Figura tomada del libro: Biología. una Perspectiva Humana, de Irwin W. Sherman, Vilia G. Sherman, tercera edición. Mc Graw Hill.

TESIS CON
FALLA DE ORIGEN

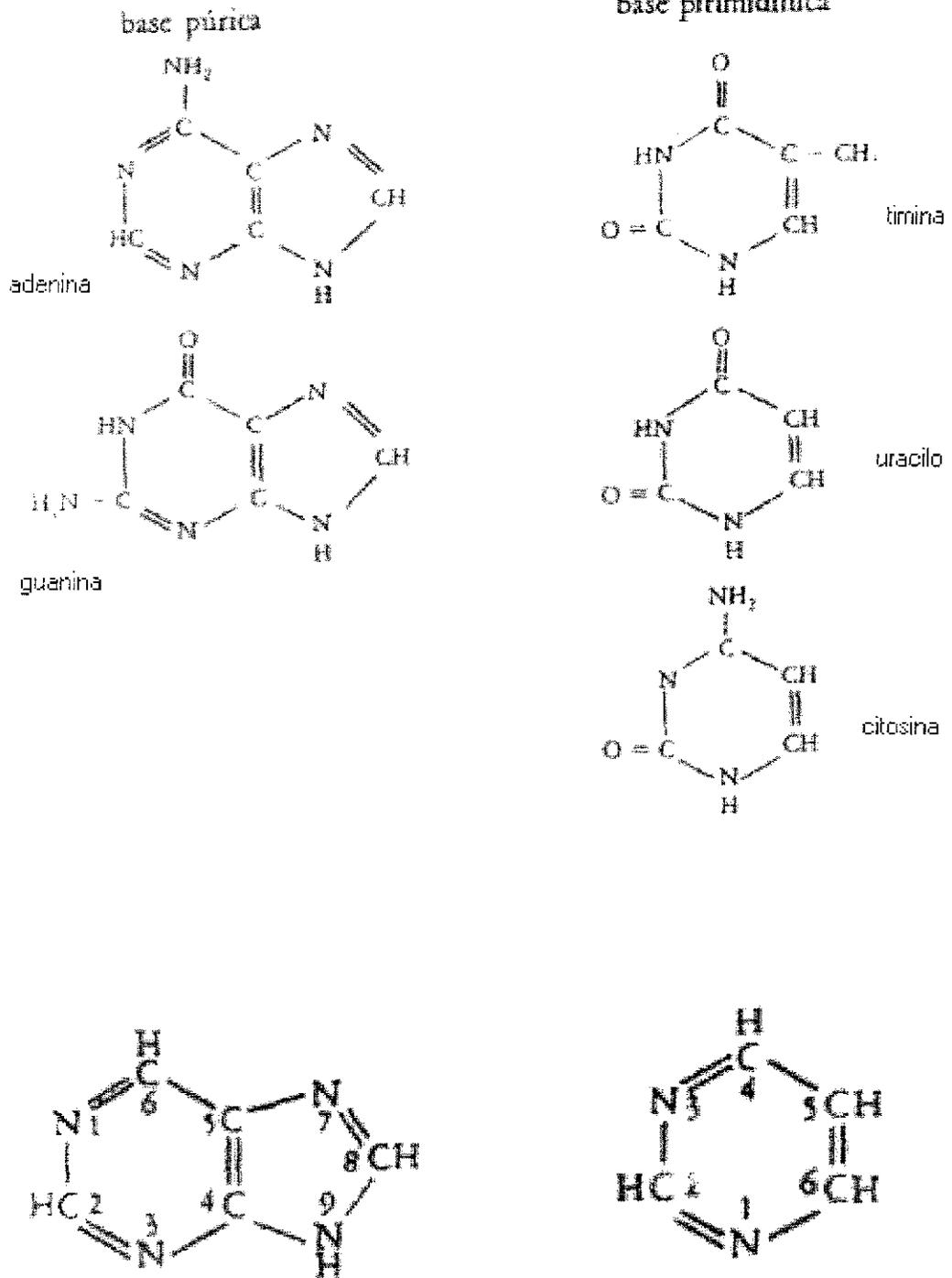


Figura 1.2 Bases nitrogenadas del ADN y las estructuras de la purina y pirimidina. [Irwin et al. [1]]

**TESIS CON
FALLA DE ORIGEN**

Para resolver este espaciamiento repetitivo Watson y Crick construyeron modelos a escala del ADN, utilizando el tamaño de las purinas y pirimidinas y las dimensiones del cristal del ADN. Ordenaron los nucleótidos en una secuencia lineal en la que el eje de la cadena estaba formada por secuencias repetitivas de azúcar-fosfato-azúcar-fosfato; mas aún, propusieron que la molécula de ADN constaba de dos largas cadenas, cada una de ellas con un eje de azúcar-fosfato y ambas estaban enlazadas por puentes de hidrógeno. La adenina unida con la timina (mediante 2 puentes de hidrógeno) y la guanina unida con la citosina (mediante 3 puentes de hidrógeno), (figura 1.3 y 1.4).

Los pares de bases forman los escalones de una escalera y los ejes constan de una secuencia ordenada en forma alterna de azúcares y fosfatos; la distancia entre bases sucesivas (peldaños) era de 3.4\AA . La periodicidad de 34\AA que Wilkins había encontrado en los cristales de ADN es 10 veces 3.4\AA . Watson y Crick propusieron que la escalera estaba enrollada en forma de hélice (semejante a una escalera en espiral), con un giro completo cada décima base figura 1.5. Una de las propiedades que el ADN debe tener para funcionar en la herencia es la capacidad de duplicarse. El modelo de Watson-Crick incluye esto. Una cadena del ADN es complemento de la otra, puesto que cuando la adenina se presenta en una cadena, la timina lo hace en la opuesta, y la guanina en una cadena está frente a la citosina en la otra. Las cadenas helicoidales pueden ser separadas por las bases apareadas; el "abrirse" es el resultado de la reducción de los hidrógenos que formaban puentes entre las bases, esto permite que se forme una nueva cadena complementaria de nucleótidos y otros materiales presentes en la célula, bajo la dirección de una enzima específica llamada ADN polimerasa. Toda la información necesaria para ordenar estas bases en una secuencia lineal con el fin de complementar las cadenas originales del ADN está proporcionada por este mecanismo de apareamiento complementario de bases. Por ejemplo, si la cadena del ADN contiene adenina, un nucleótido que contenga timina se apareará con éste; y si la siguiente base en la cadena de ADN es guanina, el nucleótido que hará

TESIS CON
FALLA DE ORIGEN

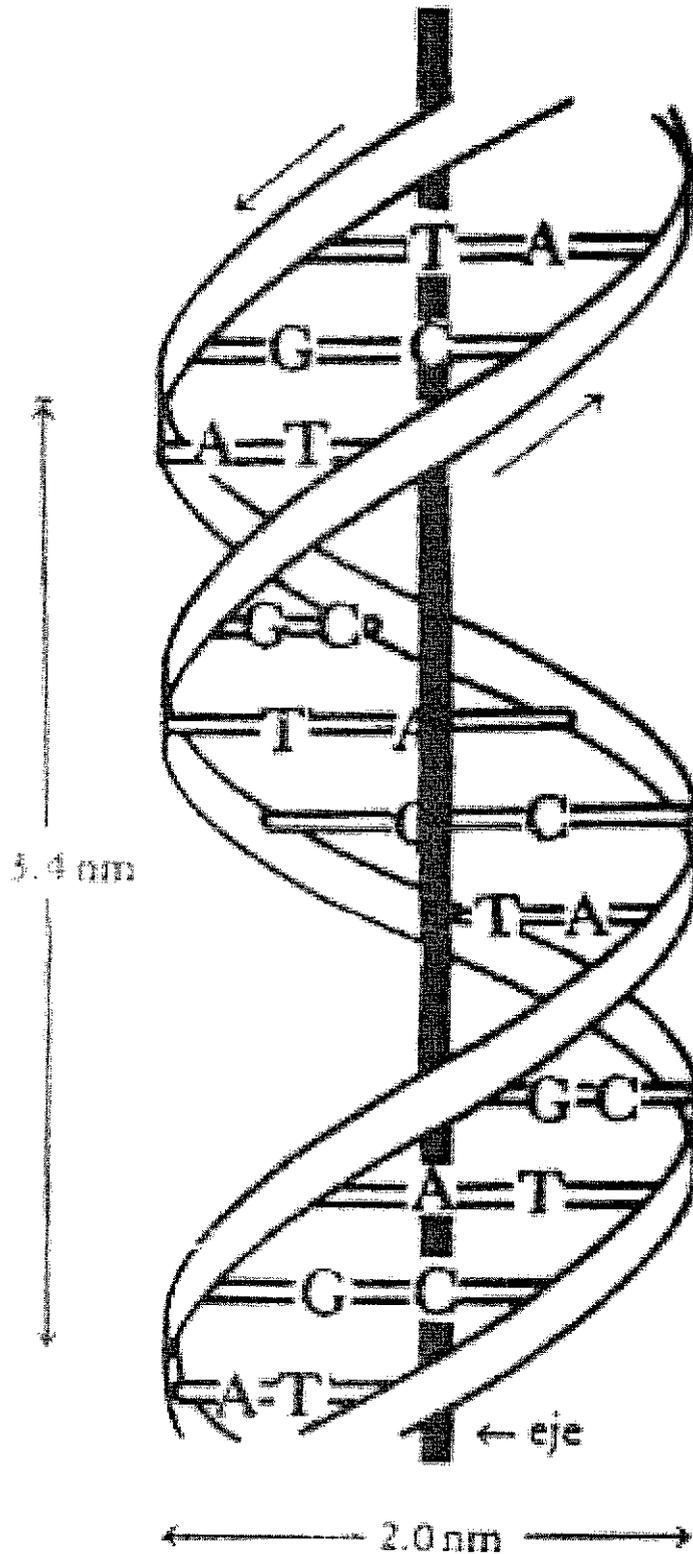


Figura 1.3: Organización de la doble hélice del ADN. Los soportes de hélice constan de grupos alternados de carbohidratos fosfatados, y los escalones están formados por pares de bases unidas por puentes de hidrógeno. [Irwin et al. [1]]

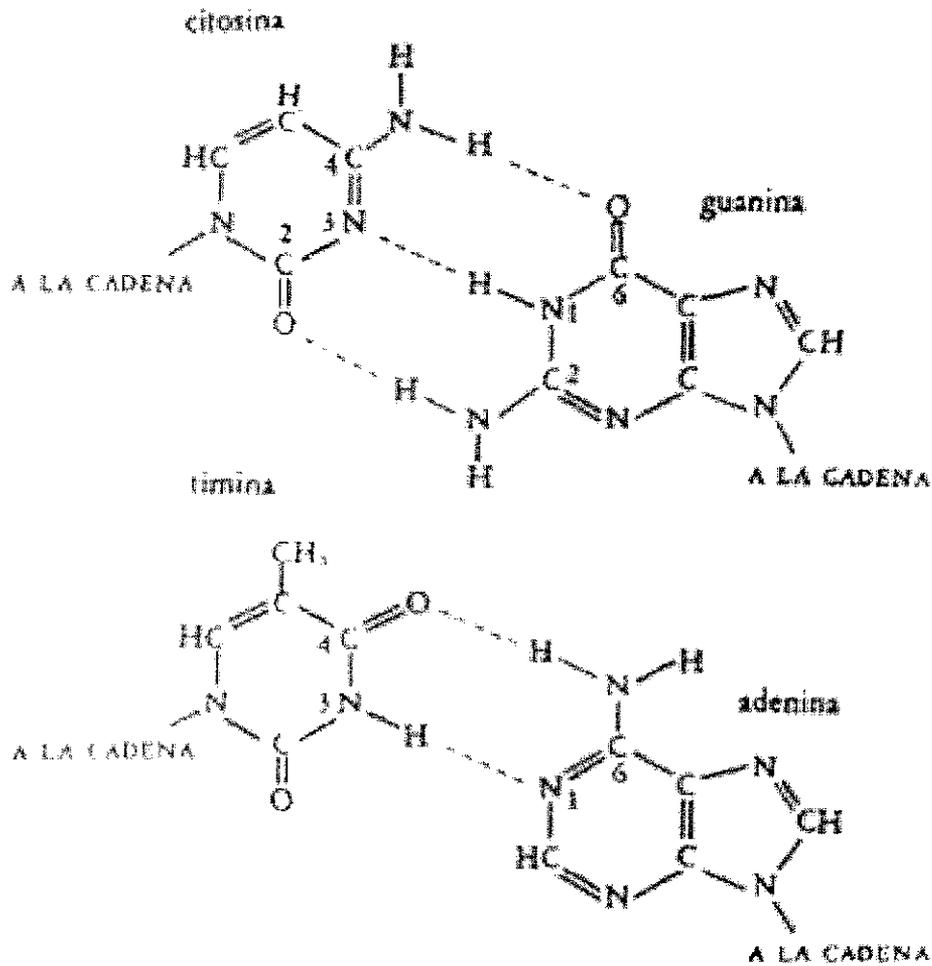


Figura 1.4: Formación de puentes de hidrógeno en la molécula de ADN entre la citosina y la guanina y entre la adenina y la timina.

TESIS CON
FALLA DE ORIGEN

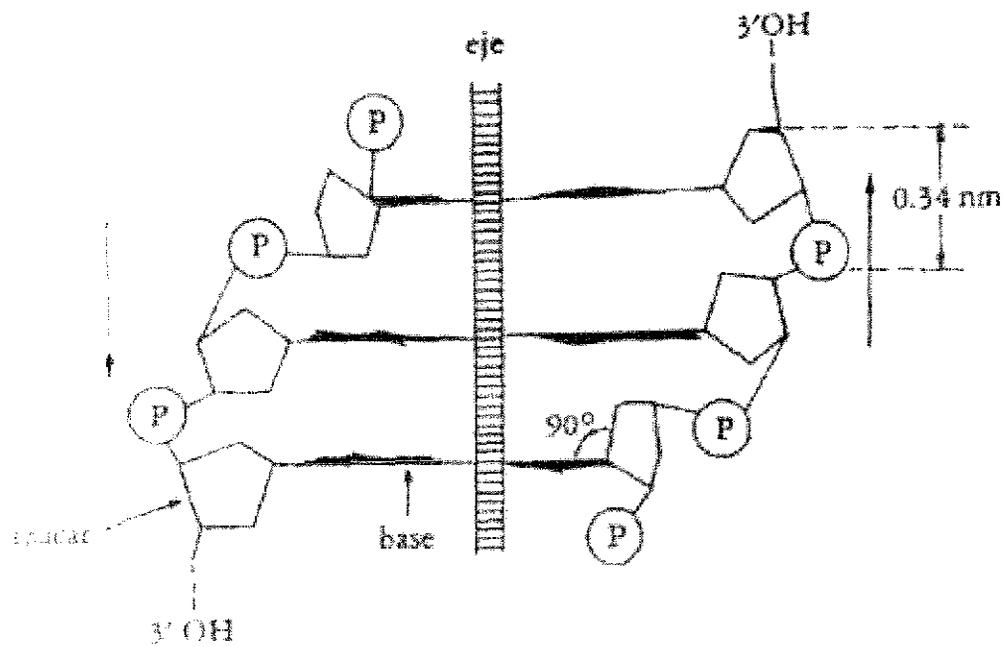


Figura 1.5: Esquema del arreglo espacial entre la base y el azúcar en la molécula del ADN. [154]

TESIS CON
FALLA DE ORIGEN

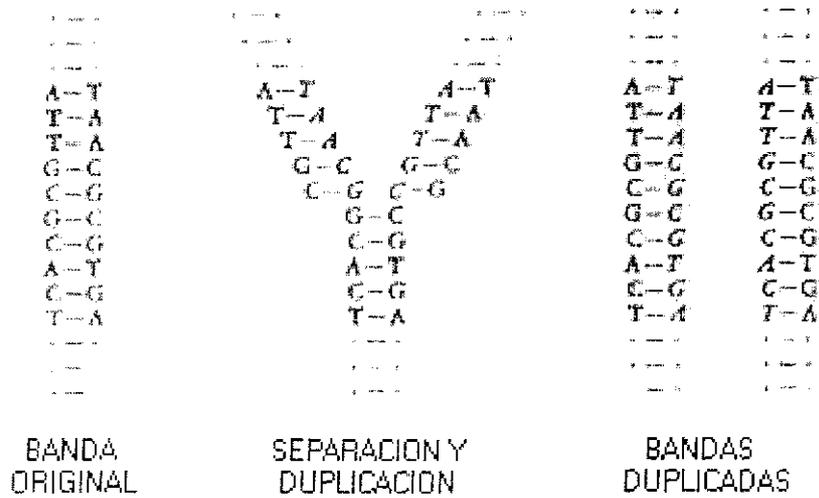
par con él debe contener citosina; y así sucesivamente. En esta forma las cadenas originales del ADN dirigen la secuencia u orden de los nucleótidos que forman las nuevas cadenas complementarias; la nueva cadena es una copia complementaria de la original. La formación de dos nuevas cadenas idénticas a las dos originales se efectúa así, pero además y esto es lo más importante, la secuencia exacta de los pares de bases nitrogenadas en las cadenas originales se habrá reproducido exactamente, excluyendo errores raros para ser discutidos posteriormente (figura 1.6). Este proceso de duplicación del ADN se llama replicación.

1.2 Código Genético: Alfabeto de Cuatro Letras, Palabras de Tres Letras

Los genes dirigen la producción de proteínas. Regiones específicas de la hélice del ADN presentes en la célula son la materia de la cual están hechos los genes: en una forma simple, una porción de la molécula de ADN es igual a un gen. (La cantidad de ADN que constituye un gen, es una cuestión compleja que no trataremos aquí).

El ADN de casi todos los organismos existe como una cadena doble helicoidal larga y sin ramificar, localizada en el núcleo. En cada cadena de doble hélice hay cuatro clases de moléculas; las bases nitrogenadas. Es la secuencia de estas bases la que determina la naturaleza de la información disponible para la célula. El segmento de ADN que compone un gen o genes produce una proteína, la secuencia de bases codifica las sub-unidades de una proteína: los aminoácidos. Existen 20 diferentes tipos de aminoácidos en la mayor parte de las proteínas; por tanto, la secuencia de bases debe poder codificar por lo menos 20 diferentes aminoácidos.

Con palabras de tres letras (o “palabras” compuestas de tres bases) el código genético puede codificar 64 palabras, más de lo que se necesita. Se ha establecido que el código genético es un código por tripletes; esto es, consta de secuencias de tres bases nitrogenadas. La naturaleza de las palabras del código por tripletes ha sido descifrada completamente. Esto se muestra



**TESIS CON
FALLA DE ORIGEN**

Figura 1.6: Representación esquemática de la forma en que se replica el ADN. al extremo izquierdo se muestra la doble hélice intacta de ADN, enseguida las cadenas complementarias se separan y los nucleótidos libres se alinean en un orden específico (en forma complementaria a los que se encuentran en la cadena de ADN); mediante la enzima ADN polimerasa, se forman enlaces fosfato entre nucleótidos adyacentes. Finalmente resultan dos nuevas cadenas complementarias de ADN. [Irwin et al. [1]]

en la figura 1.7. Como puede observarse, hay alguna redundancia en el código, puesto que algunas combinaciones significan el mismo aminoácido; por ejemplo, las palabras del código del ADN que especifican la alanina son: CGA, CGG, CGT y CGC, el código genético es un código degenerado.

Se ha descubierto que el código genético es universal, esto es, se da el mismo triplete para el mismo aminoácido en todos los organismos desde los virus hasta los seres humanos. Esto sugiere que el código genético probablemente evolucionó sólo una vez. Las diferencias entre virus, bacterias, seres humanos y los demás organismos vivos no se deben a diferencias en las clases de bases nitrogenadas en su ADN, sino a variaciones en la secuencia de estas bases. Es muy parecido a usar las letras del alfabeto para formar palabras diferentes. Las letras utilizadas son siempre las mismas, sin embargo, las palabras formadas y su significado (mensaje) pueden diferir según la manera en que se hayan ordenado las letras. Similarmente, el orden diferente de los tripletes (letras) a lo largo de la cadena de ADN, originará una secuencia variable de aminoácidos (palabras) en una proteína (enunciado); es posible producir un número casi infinito de ordenamientos de tripletes en el ADN, y así es posible tener una variedad ilimitada de secuencias de aminoácidos en las proteínas. La especificidad de las especies tiene como base las diferentes secuencias de aminoácidos en una proteína (esto es, la hemoglobina del gato difiere de la del humano en el número de aminoácidos), pero debe hacerse hincapié en que éste es, en último término, un reflejo del orden específico de las palabras de tres letras de la clave genética, el ADN.

1.3 Mutaciones en el código genético

Ocasionalmente ocurre en la naturaleza una alteración en el orden de las bases en la molécula de ADN. Tales mutaciones pueden ser causadas por radiaciones de alta energía, o por influencias de naturaleza química. Las mutaciones son indudablemente importantes en la evolución. La información para un determinado aminoácido corresponde a un triplete de nucleótidos. Si se altera una de las bases del triplete, se cambia la codificación para un

Segunda base en el triplete de DNA

A		G		T		C	
AAA	metionina	AGA	serina	ATA	metionina	ACA	tirosina
AAU		AGC		ATG		ACG	histidina
AAT	asparagina	AGT	asparagina	ATT	codon de terminación	ACT	tirosina
AAC		AGC		ATC	terminación	ACC	tirosina
GAA	glutamina	GGA	prolina	GTA	histidina	GCA	arginina
GAG		GGG		GTG		GCG	
GAT	glutamina	GGT	prolina	GTT	glutamina	GCT	arginina
GAC		GCC		GTC		GCC	
TAA	isoleucina	TGA	isoleucina	TTA	asparagina	TCA	serina
TAG		TGG		TTG		TCC	
TAT	isoleucina	TGT	isoleucina	TTT	lisina	TCT	serina
TAC	isoleucina	TGC		TTC		TCC	asparagina
CAA	valina	CGA	valina	CTA	ácido aspártico	CCA	glicina
CAG		CGC		CTG		CCG	
CAT	valina	CGT	valina	CTT	ácido glutámico	CCT	glicina
CAC		CGC		CTC		CCC	

Primera base en el triplete de DNA

(a)

TESIS CON FALLA DE ORIGEN

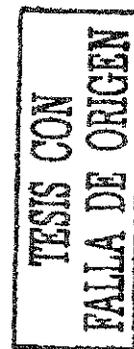
Figura 1.7: Código genético descifrado. Palabras del código del ADN. [Irwin et al. [1]]

nuevo aminoácido, alterando ligeramente la estructura de la proteína de la cual contiene el gen toda la información.

Un aminoácido cambiado en una parte poco importante de la molécula de proteína no tiene efecto sobre el fenotipo del organismo. Pero una mutación en una zona crucial de la molécula puede tener consecuencias inesperadas. El grado de modificación de determinado carácter del fenotipo dependerá del número de genes que sean responsables de ese carácter. Si depende de un solo gen, la mutación en este gen puede tener efectos drásticos, es decir, puede tener una variación discontinua en el individuo; pero si son una serie de genes los que conforman el carácter, la mutación en uno de ellos tendrá un efecto pequeño sobre el carácter.

Es evidente que una forma de producir un cambio genético en una célula u organismo es mediante la sustitución de un nucleótido por otro. Esta sustitución podría deberse a un tipo de error "tipográfico". Es concebible que durante la replicación de la cadena de ADN, cuando esta se separa, se reordenen los átomos de hidrógeno del nucleótido, y se originen puentes de hidrógeno defectuosos que permitan el enlace del nucleótido equivocado en la nueva cadena que se va a sintetizar. Por ejemplo, en lugar de que A se enlace con T, debido quizá a un puente de hidrógeno defectuoso, se enlaza a C. Cuando algo así ocurre en la replicación siguiente, la cadena con la C se aparea con G; la cadena central original se aparea con T. Así, surgirá una molécula de ADN en la siguiente generación en la que GC sustituiría a AT. Hay estudios experimentales donde se muestran como las bases del ADN pueden unirse por puentes de hidrógeno a otras bases diferentes a las que se unen en el ADN [156].

También podría ocurrir que al cambiar la molécula a un isómero, entonces debido a su nueva configuración esta solo se pueda unir a una base no complementaria a la que se une en el ADN, de esto no hay estudios experimentales ya que no se sabe si las bases tienen otros isómeros distintos de las conformaciones planas. Nosotros presentamos en los resultados un posible caso donde se presenta lo anterior: un isómero de la adenina unido a la molécula citosina.



La configuración se comprobó que es estable; de confirmarse esta situación entonces también ocurriría que en la siguiente generación GC sustituiría a AT.

Otra posible fuente de error son las transiciones en las que una base cambia a un tautómero (esto es, un hidrógeno se enlaza a otro átomo de la molécula). Por ejemplo, podría ocurrir una transición en la que guanina o adenina cambien de estructura y esto podría resultar en la formación de pares de bases guanina-timina o adenina-citosina de forma similar a los pares A:T y G:C. Tales pares de bases no han sido predichas únicamente por cálculos teóricos [123][124], sino que también han sido encontradas en cristales y soluciones [125].

1.4 Nucleótidos en otras moléculas biológicas

Los nucleótidos también existen en otras moléculas biológicas además del ADN y ARN, pero con papeles muy diferentes en la función celular. Por ejemplo, una molécula relativamente sencilla, denominada ATP (adenosin trifosfato), consta de una molécula adenina, un azúcar y de tres grupos “fosfato” formados por fósforo y oxígeno: esta molécula es fuente inmediata de la mayor parte de la energía utilizada por los seres vivos. Los dos grupos de fosfatos terminales se unen al nucleótido por enlaces ricos en energía. La energía biológicamente útil de estos enlaces puede ser transferida a otras moléculas; la mayor parte de la energía química de la célula es almacenada en estos enlaces, para ser liberada cuando el grupo de fosfatos es transferido a otra molécula (figura 1.8). Una conformación diferente de la adenina podría tener consecuencias en la energía de enlace de los grupos fosfatos que se unen con estas moléculas. El trifosfato de guanosina, GTP, se requiere específicamente en ciertas etapas de la síntesis de las proteínas; el trifosfato de uridina, UTP, se requiere en ciertas etapas del metabolismo de los carbohidratos, como en la síntesis del glucógeno, y el trifosfato de citidina, CTP, se necesita específicamente para ciertas etapas de la síntesis de grasas y fosfolípidos. Estos cuatro trifosfatos nucleótidos son necesarios para

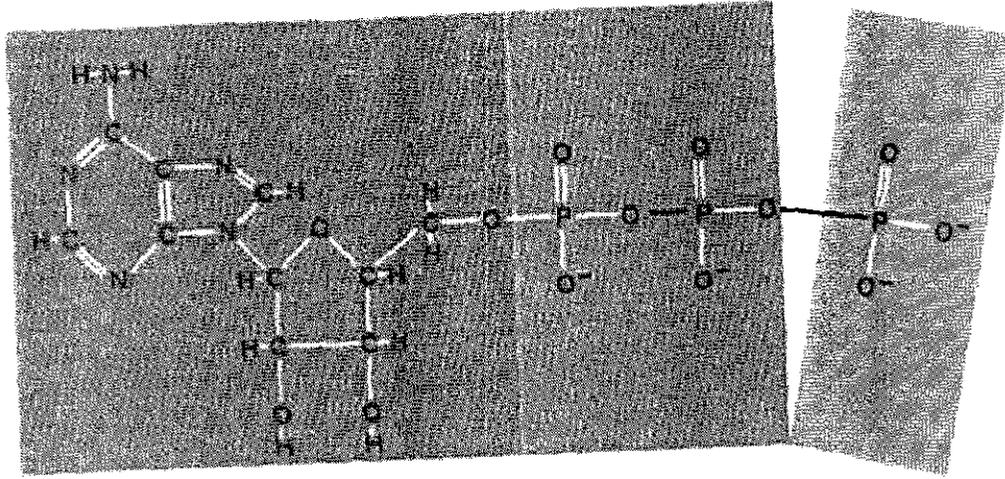


Figura 1.8: La molécula ATP es fuente de energía en la mayoría de los procesos biológicos.

la síntesis de RNA y los cuatro trifosfatos nucleósidos desoxirribosa -dATP, dGTP, dCTP y dTTP- se necesitan para la síntesis del ADN. Diferentes conformaciones de las 4 bases podría repercutir en estas importantes funciones que desempeñan en los seres vivos, las cuales como se puede ver son muy diversas.

Completando esta lista de nucleótidos importantes en los procesos metabólicos se encuentran los dinucleótidos NAD, NADP y FAD. El dinucleótido de niacina-adenina, abreviado NAD (llamado también nucleótido difosfopiridina. DPN), consta de nicotinamida, ribosa y fosfato unidos a un nucleótido de adenina, fosfato, ribosa y adenina. El NAD es de gran importancia en las reacciones de oxidación celulares. Enzimas llamadas deshidrogenasas eliminan electrones e hidrógeno de moléculas como las de ácido láctico y los transforman en NAD, el cual a su vez los transfiere a otros aceptores de electrones.

El dinucleótido de niacina-adenina-fosfato, abreviado NADP (llamado también nucleótido trisfosfopiridina, TPN) es un aceptor de electrones e hidrógeno para otras enzimas. El NADP es exactamente igual que el NAD, excepto que

TESIS CON
 FALLA DE ORIGEN

posee un tercer grupo de fosfatos unidos a la ribosa del nucleótido de adenina.

El dinucleótido de adenina flavina, FAD, consta de riboflavina-ribitol-fosfato-fosfato-ribosa-adenina y sirve como aceptor de hidrógeno y electrones para otras deshidrogenasas. Es importante el notar que estos dinucleótidos tiene vitaminas -nicotinamida o riboflavina- como partes componentes. Estas moléculas NAD, NADP y FAD se denominan coenzimas; son cofactores necesarios para el funcionamiento de ciertos sistemas enzimáticos, pero sólo están flojamente unidos a la molécula de enzima y son fácilmente eliminados. Cuando aceptan electrones e hidrógenos, cambia su forma oxidada (por ejemplo NAD) a su forma reducida (NADH). Se convierte de nuevo en la forma oxidada cuando transfieren sus electrones al siguiente aceptor de la cadena de enzimas respiratorias.

Capítulo 2

EL POTENCIAL UTILIZADO

Para realizar la búsqueda de los isómeros de las bases del ADN utilizamos un potencial que a continuación presentaremos y el cual fué desarrollado específicamente para tratar estas moléculas así como todos los aminoácidos.

2.1 Introducción

La geometría molecular es el punto de inicio para estudiar la mayoría de las propiedades de los mas diversos compuestos, tanto orgánicos como inorgánicos. Aunque técnicas modernas (NMR, etc.) nos dan algunas indicaciones acerca de las distancias interatómicas para algunas muestras en estado líquido, la información geométrica, para moléculas de tamaño grande o mediano, ha sido hasta ahora principalmente derivada de datos cristalográficos, obtenidos de mediciones directas.

Sin embargo, tal información sólomete corresponde a los cristales, donde la geometría puede de algún modo afectarse por efectos de empaquetamiento, la cual no correspondería a moléculas aisladas o en una fase gaseosa, o por ejemplo para sustancias reactivas en medios biológicos. Además, tales datos no son alcanzables para muestras que son difíciles de obtener en forma de cristales, para estructuras hipotéticas (conformaciones diferentes del estado base) o para moléculas aún no sintetizadas. Finalmente, si los datos con rayos X dan la geometría para la forma mas estable, estos estudios no nos dirán nada acerca de la posible existencia de otras estructuras estables de mayor

energía. Es aquí donde técnicas computacionales adecuadas se requieren para tratar algunos de los problemas anteriores.

Los métodos de química cuántica son una herramienta adecuada para predecir las características geométricas, electrónicas o energéticas de moléculas conocidas o aún de conformaciones desconocidas. Sin embargo, estos métodos son hasta ahora y probablemente también lo serán en un futuro cercano, demasiados costosos en términos de tiempo computacional e intratables, aún en el nivel más simple, el semi-empírico, para estudiar muchas moléculas orgánicas o estructuras biológicas macromoleculares. En este contexto, un interés cada vez más creciente se ha enfocado en otros modelos. En este campo un rol cada vez más importante ha sido jugado por técnicas conocidas como mecánica molecular o métodos de potencial empíricos (empirical force field). Básicamente la mecánica molecular trata a las moléculas como si estuviera compuesta de masas y resortes. Esta técnica usa las leyes de la mecánica clásica para tratar las diversas interacciones en la molécula, de acuerdo a un modelo que ha sido parametrizado empíricamente.

Estos potenciales empíricos en términos de átomos y enlaces extienden las capacidades para tratar situaciones fuera de lo común que no pueden ser representadas fácilmente (distancias o ángulos de enlace diferentes de los valores usuales, distorsiones geométricas, etc.), también nos dan información importante acerca de la energía de conformación así como posibles geometrías más estables o de más baja energía. Aunque este método no nos dice nada acerca de las características electrónicas, su implementación y velocidad computacionales y su habilidad para tratar con sistemas grandes lo hacen muy atractivo siempre que las propiedades electrónicas no sean necesarias o en casos que se requieran geometrías optimizadas para enviarlas subsecuentemente a métodos de química cuántica.

Estos métodos son a menudo considerados como la otra cara de la separación de Born-Oppenheimer de los movimientos electrónicos y nucleares. En los cálculos de mecánica cuántica, uno comienza con una posición dada para los núcleos y busca la mejor repartición de los electrones en el poten-

cial generado por los núcleos (este cálculo puede continuarse modificando las posiciones de los átomos para obtener el confórmero de mejor estabilidad en un proceso de minimización de energía). En la superficie potencial del potencial empírico, por el contrario, uno estudia la posición de los núcleos en un campo generado por los electrones. Los electrones no son considerados explícitamente (con algunas excepciones, sin embargo), y el campo que ellos generan no es realmente calculado, sino que es representado por un potencial “efectivo” tratado de acuerdo a la mecánica clásica.

Diversas publicaciones bien documentadas acerca de los métodos de campos de fuerza empíricos han sido publicadas [4]-[12], por lo que aquí únicamente presentaremos una breve revisión de los principios en que se basan. La aproximación de Allinger et al es una de las más famosas utilizadas [4],[12]-[18].

La energía interna de un sistema molecular, se puede representar, cerca de un punto de mínima energía, por medio de una expansión de serie de Taylor que involucre las coordenadas atómicas. Este es más fácilmente llevado a cabo por medio del uso de coordenadas internas: distancias de enlace, ángulos de torsión y enlace, junto con términos de interacción entre átomos no enlazados, expresados por interacciones de Van der Waals.

Esto puede ilustrarse con el ejemplo simple de una molécula diatómica tratada por el modelo de mecánica clásica de dos esferas unidas por un resorte (figura 2.1) para una pequeña variación de la distancia internuclear (r) alrededor de la posición de equilibrio (r_o) la energía potencial es expresada por:

$$V = V_o + \frac{dV}{dr}(r - r_o) + \frac{1}{2} \frac{d^2V}{dr^2}(r - r_o)^2 + \dots \quad (2.1)$$

Si le asignamos a la energía un valor nulo en el mínimo (V_o) y ya que en la posición de equilibrio la primera derivada es cero:

$$V = \frac{1}{2} \frac{d^2V}{dr^2}(r - r_o)^2 + \dots = \frac{1}{2} k(r - r_o)^2 + \dots \quad (2.2)$$

Esto corresponde a la bien conocida ley de Hooke, familiar a los espectroscopistas, k siendo la constante de fuerza de alargamiento.

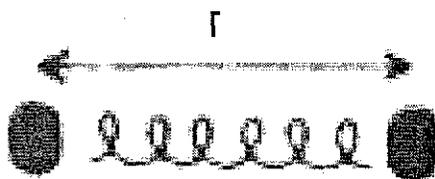


Figura 2.1: Modelo de mecánica clásica.

Una expresión similar puede ser escrita para sistemas más complejos:

$$V = \frac{1}{2} \sum k_{ij} \Delta r_i * \Delta r_j + \dots \quad (2.3)$$

donde las k_{ij} son las constantes de fuerza cuadráticas y Δr_i las coordenadas de desplazamiento del núcleo i (cambios de las coordenadas que definen la localización del núcleo de los valores de referencia).

El problema es ahora definir los coeficientes k_{ij} que son las constantes de fuerza. Este conjunto constituye el campo de fuerza, el cual nos permite definir los cambios en la energía dependiendo de la geometría molecular. Los principios fundamentales de los campos de fuerza son:

- La energía potencial de una molécula puede ser representada como una suma de términos asociados respectivamente con los varios tipos de deformaciones moleculares (cambios en longitudes de enlace, ángulos de valencia o torsión) o interacciones átomo-átomo. La “energía de conformación” calculada a partir de la suma de estos términos representa la energía adicional asociada a la estructura con respecto a una situación ideal donde todos los elementos geométricos estarían en un estado de referencia (el infinito).

- Los parámetros necesarios para calcular esta “energía de conformación” pueden ser derivados de información ganada en base a pequeñas moléculas (longitudes de enlace, ángulos), su transferibilidad a sistemas grandes se asume.

TESIS CON
 FALTA DE ORIGEN

2.2 El campo de fuerza

La clave para exitosas predicciones de el campo de fuerza y la mecánica molecular está basada en una buena representación de estos diferentes términos energéticos, en este punto nos debe quedar claro que los cálculos del campo de fuerza estaán basados en modelos. Estos modelos tienen características que reproducen hechos experimentales, pero esto no significa que en todos los aspectos sea una reproducción fiel de la molécula bajo estudio. Esto únicamente significa que una información particular que ha sido utilizada para crear el modelo es reproducida por el campo de fuerza [4].

Aunque todos están de acuerdo en que el campo de fuerza sea “consistente”, esto es, de acuerdo con Lifson y Warshel [19, 20], un campo de fuerza que debería de ser optimizado al ajustar toda clase de datos experimentales disponibles: en realidad en la práctica la situación es demasiado compleja, ya que hasta ahora solo se han podido tratar una cantidad limitada de datos. Así, varios campos de fuerza han sido propuestos, los cuales varían unos de otros en las propiedades experimentales a ser reproducidas: geometrías, calores de formación, frecuencias de vibración, así como el alcance de aplicación.

Sin embargo, conforme los campos de fuerza se vuelvan más y más completos, estos tenderán a ser mas similares conforme sus áreas de aplicación se traslapen.

2.3 Componentes del Campo de Fuerza

En cálculos de mecánica molecular, los campos de fuerza toman generalmente la siguiente forma:

$$V = V(r) + V(\theta) + V(\Phi) + V(nb) + (\text{terminos específicos}) \quad (2.4)$$

en la cual los términos sucesivos están asociados con alargamiento o disminución de enlace, doblamiento de ángulo de enlace, torsión de enlace, interacciones entre átomos no ligados más términos específicos tales como

doblamiento fuera de plano, interacciones electrostáticas y posible enlaces de hidrógeno. Enfocaremos nuestra atención al nivel de aproximación mas simple para enfatizar los principios básicos del modelo, posteriormente nos enfocaremos en los refinamientos del campo de fuerza.

2.4 Variación de longitud de enlace

La contribución por variación de longitud de enlace está representada por la ley de Hooke:

$$\frac{1}{2} \sum k_r (r - r_o)^2 \quad (2.5)$$

el índice de la sumatoria cubre todos los enlaces, k_r y r_o son constante de fuerza armónica y la longitud de enlace de referencia respectivamente para el enlace correspondiente.

Alternativamente, un potencial de Morse se puede utilizar:

$$V(r) = \sum D[1 - \exp(-a(r - r_o))]^2 \quad (2.6)$$

D y a son parámetros que caracterizan el enlace (figura 2.2). El uso de tal potencial parece útil para enlaces de hidrógeno alargados que de otro modo tenderían a disociarse.

Para grandes deformaciones o en el estudio de gran concentración de moléculas, un término cúbico adicional (figura 2.3) $\frac{1}{2}k'_r(r - r_o)$ puede ser introducido (términos cuárticos son rara vez considerados). Sin embargo hay un inconveniente en el uso de términos cúbicos: debido a la forma de la función cúbica, la energía va a menos infinito conforme los átomos se separan. Si el comienzo de la geometría es demasiado malo, el enlace tendería a disociarse [21]

2.5 Doblamiento de ángulo

Una función de doblamiento cuadrática es usualmente utilizada:

$$V(\theta) = \frac{1}{2} \sum_{\text{angulos}} (\theta - \theta_o)^2 \quad (2.7)$$

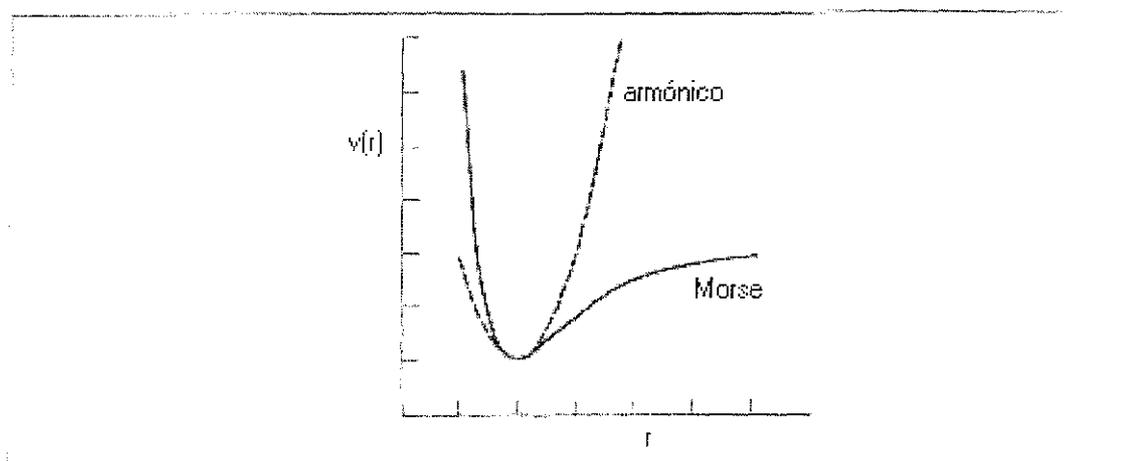


Figura 2.2: Comparación del potencial armónico simple (ley de Hooke) con la curva de Morse.

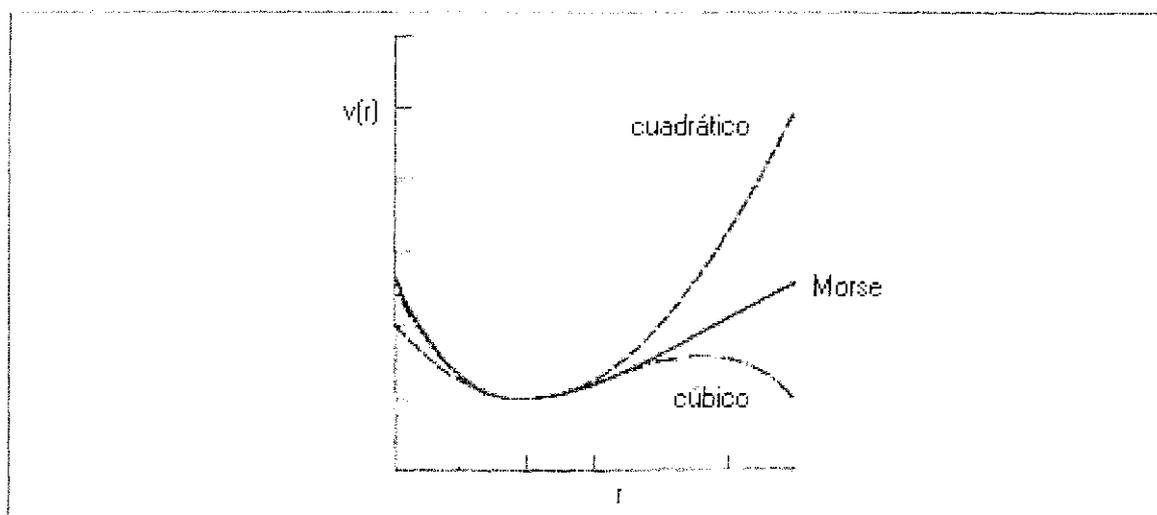


Figura 2.3: Aunque un potencial de alargamiento de enlace cúbico posee un máximo este da una mejor aproximación que la forma cuadrática a la curva de Morse cerca de la estructura de equilibrio

TESIS CON
FALLA DE ORIGEN

(el subíndice o también corresponde a los valores de referencia. Tal expresión parece trabajar bien para deformaciones en un rango alrededor de $10 - 15^\circ$, el cual cubre la mayoría de los casos usuales.

Para distorsiones mayores que ocurren en pequeñas moléculas cíclicas, términos adicionales podrían ser introducidos, pero lo que usualmente se prefiere es mantener un término cuadrático con diferentes constantes de fuerza, específicas para cada molécula cíclica investigada.

2.6 Torsión

Los efectos de torsión de enlace pueden ser bien reproducidos con una serie de Fourier:

$$V(\Phi) = \sum \left[\frac{V_1}{2}(1 + \cos\Phi) + \frac{V_2}{2}(1 - \cos 2\Phi) + \frac{V_3}{2}(1 + \cos 3\Phi) + \dots \right] \quad (2.8)$$

Φ representa el ángulo torsional, esto es, para un fragmento A-P C-D formado por 4 átomos, es igual al ángulo diedro entre los planos ABC y BCD. En una proyección de Newman a lo largo del enlace central BC, Φ también es ángulo entre las proyecciones de los enlaces A-B y C-D. De acuerdo a la convención propuesta por Klyne y Prelog [22], un ángulo positivo diedro corresponde a una rotación en el sentido de las manecillas del reloj del enlace A-B de tal forma que este llegue al segundo enlace CD. Este adquiere el mismo valor cuando uno mira a lo largo del enlace de un lado o del otro (figura 2.4). En mecánica molecular, gracias a la inclusión de las repulsiones de Van der Waals entre dos átomos no ligados, y dependiendo del grado de precisión requerida, parece ser que tres términos son suficientes para cubrir todas las situaciones (figura 2.5). La torsión alrededor de un enlace $sp^3 - sp^3$ altamente simétrico, como en el etano, esta bien reproducida cuando utilizamos únicamente el tercer término correspondiente a una barrera con 3-fold. Se ha demostrado que, en los casos mas generales, inclusión de términos de 1-fold y 2-fold (aunque mas pequeños) mejoran los resultados [13]. Por otra parte, para un enlace $sp^2 - sp^2$ (como en el etileno), la situación preferida es el término 1-fold. Notemos aquí que no hacemos una mención explícita a

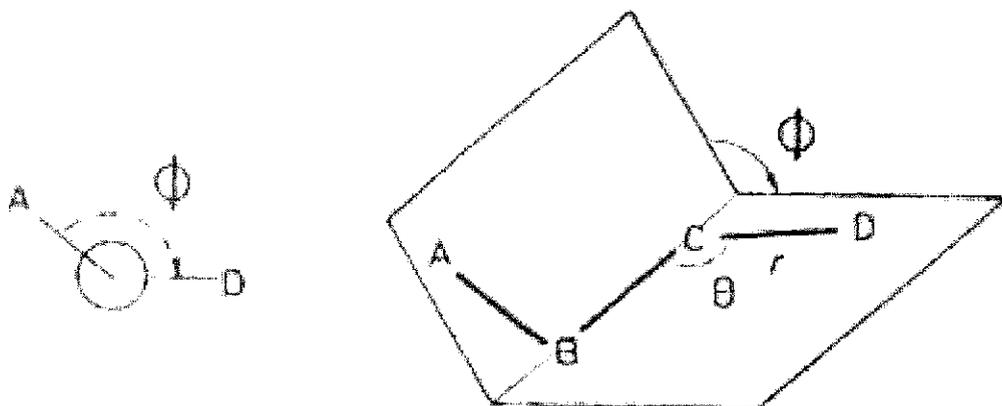


Figura 2.4: Torsión.

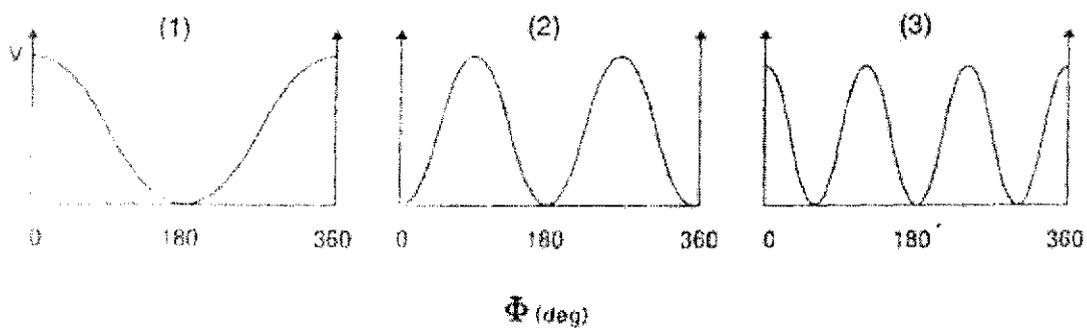


Figura 2.5: Forma de los potenciales de torsión V_1 , V_2 y V_3 .

TESIS CON
FALLA DE ORIGEN

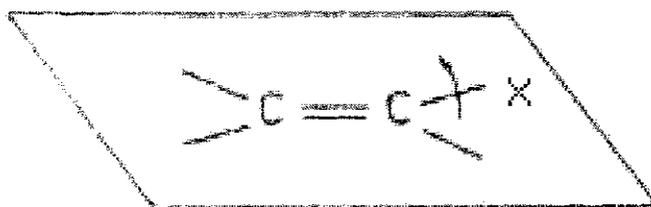


Figura 2.6: Doblamiento fuera de plano.

TESIS CON
FALLA DE ORIGEN

los electrones π para tratar la rigidez del doble enlace.

En moléculas más complejas, los tres términos de arriba tienen que ser considerados. V_2 se toma en cuenta para barreras 2-fold a lo largo de enlaces dobles, pero este término también interviene para enlaces $sp^3 - sp^3$ (en el butano).

2.7 Doblamientos fuera del plano

Para sistemas con carbonos sp^2 , es necesario distinguir entre deformaciones en el plano y afuera del mismo. En alcanos, por ejemplo, doblamientos fuera del plano son descritos por un término cuadrático (figura 2.6 y 2.7):

$$\frac{1}{2} \sum k' x^2 \quad (2.9)$$

Tales términos son importantes para pequeñas moléculas cíclicas (ciclobutano) donde los ángulos de valencia son completamente diferentes de los usuales con un valor de 120° . Sin ellos, el átomo exocíclico tendería a moverse exageradamente afuera del plano para restaurar un ángulo de valencia de 120° (figura 2.8 y 2.9).

2.8 Interacciones de Van der Waals

Para asegurar la transferibilidad del campo de fuerza de una molécula a otra la mecánica molecular toma en cuenta explícitamente las interacciones entre

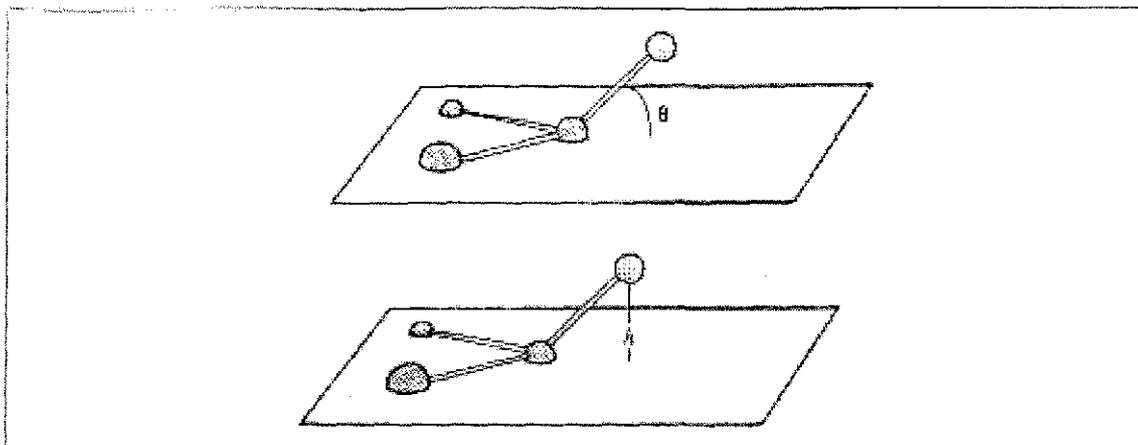


Figura 2.7: Dos modos de modelar las contribuciones de doblamiento fuera de plano



Figura 2.8: Sin términos cuadráticos, el átomo exocíclico trataría de salirse del plano para adquirir un valor de 120° .

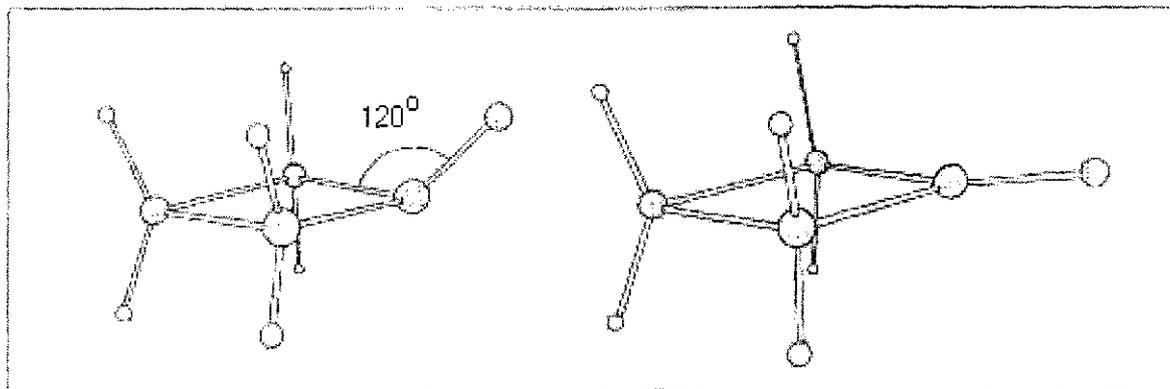


Figura 2.9: La situación mostrada en la figura 2.8 se presenta en el átomo de oxígeno del ciclobutano, donde se predice que este átomo debe de estar fuera del plano del anillo en lugar de estar en el plano.

TESIS CON
FALLA DE ORIGEN

átomos no ligados. Estas interacciones se les refiere usualmente como interacciones de Van der Waals. Estos términos son desgraciadamente omitidos en cálculos espectroscópicos, con la consecuente variación de constantes para cada molécula individual considerada.

Estas interacciones de Van der Waals son representadas asumiendo la aditividad de términos pares: para átomos i,j :

$$V(nb) = \sum [A(r_{ij})^{-n} - B(r_{ij})^{-6}] \quad (2.10)$$

o, alternatively un potencial Buckingham $\exp(-6)$:

$$V(nb) = \sum [A \exp(-br_{ij}) - B(r_{ij})^{-6}] \quad (2.11)$$

donde la sumatoria es tomada sobre todos los pares de átomos no ligados (excluyendo, por supuesto, átomos ligados a un átomo común, ya que estas interacciones ya son tomadas en cuenta en los términos de alargamiento y doblamiento de ángulo de enlace.

Estas expresiones involucran dos términos:

1) Una parte atractiva, correspondiente a interacciones dipolo inducido-dipolo inducido, proporcional a r^{-6} , donde r es la distancia entre los dos átomos

2) Una parte repulsiva, correspondiente a términos de dispersión de London y aumentando rápidamente a distancias cortas (recordemos que en un modelo de esferas duras cuando las dos esferas se acercan demasiado la repulsión sería infinita cuando la distancia de los centros se vuelve menor a la suma de los radios). Varias expresiones han sido escogidas para esta parte repulsiva [23], usando ya sea una función exponencial o una potencia inversa de r . Para el valor del exponente n un valor igual a 12 es correcto para expresar interacciones en los gases raros o entre moléculas de capa cerrada consideradas como un todo (potencial 6-12 de Lennard-Jones). Un término repulsivo más suave con $n = 9$ es algunas veces preferido en mecánica molecular.

La ecuación de Hill

$$V(nb) = \epsilon [-c_1 \left(\frac{r^*}{r}\right)^6 + c_2 \exp(-c_3 \frac{r}{r^*})] \quad (2.12)$$

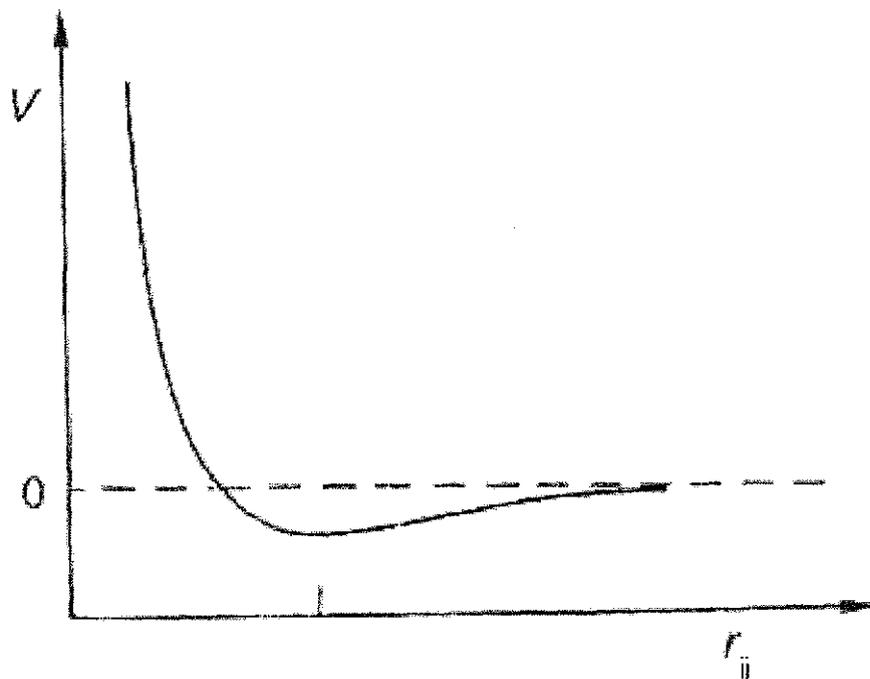


Figura 2.10: Variación de un potencial interatómico en función de la distancia internuclear.

la cual trabaja bien para los gases raros y pequeñas moléculas, ha sido también usada frecuentemente en sistemas moleculares más grandes. En la ecuación anterior, los coeficientes c_i son tomados como constantes universales, ϵ es un parámetro de energía relacionado con la profundidad del potencial. Para un par de átomos i, j , $\epsilon = \sqrt{\epsilon_i \epsilon_j}$, donde ϵ_k caracteriza la “dureza” del átomo k , r^* se define como la suma de los radios de Van der Waals del par de átomos interactuantes [24](figura 2.10). Debido a que los coeficientes c_i son tratados como constantes universales, únicamente dos parámetros son necesarios para ajustar la repulsión de Van der Waals para cada par de átomos (ϵ y la suma de los radios de Van der Waals (r^*)) [4]. Algunas otras aproximaciones implícitas son:

- Se considera que estas interacciones de pares no son dependientes de su ambiente, esto es, de los otros átomos presentes.

TESIS CON
FALLA DE ORIGEN

- Además, los átomos son tratados como esferas (con una distribución isotrópica electrónica alrededor de ellos). Esto no es completamente correcto para átomos en moléculas, y particularmente para átomos de hidrógeno donde el único electrón está concentrado en el enlace. En tal caso resultados buenos han sido obtenidos al cambiar el centro de la esfera alrededor del 0.1Å hacia el átomo más pesado [25]. Con átomos con pares solos (tales como nitrógeno en las aminas u oxígeno en alcoholes o éteres), donde las interacciones entre átomos no ligados se esperan anisotrópicas, los pares solos pueden ser tomados explícitamente en cuenta como átomos virtuales [26, 27, 28], a los cuales se les asigna un radio de Van der Waals apropiado.

También para bio-moléculas más grandes, algunos programas proveen la posibilidad de usar “átomos unidos”: átomos de hidrógeno (excepto aquellos que intervienen en enlaces de hidrógeno) son omitidos. Uno considera únicamente los átomos pesados con sus hidrógenos unidos como una entidad esférica, a la cual se le adaptan radios de Van der Waals.

2.9 Parametrización

La selección de los parámetros adecuados para determinar el campo de fuerza es de primordial importancia para la precisión que se espera en las predicciones de energía o geometría así como en el alcance estructural cubierto. Los campos de fuerza espectroscópicos merecen una especial mención, debido a fuertes similitudes entre algunos términos energéticos (alargamiento de enlace, doblamiento de enlace) y las expresiones utilizadas es espectroscopía vibracional (ley de Hooke). En análisis vibracional, desarrollado con la aproximación del campo de fuerza de valencia, la expresión de la energía potencial en términos de coordenadas internas únicamente involucra términos diagonales. Esto resulta en constantes de fuerza diferentes de un compuesto a otro. Sin embargo, tomando en cuenta algunos términos de interacción (términos no diagonales) en un campo de fuerza de valencia generalizado, alguna transferibilidad puede asegurarse dentro de una clase limitada de compuestos cercanamente relacionados (familias químicas como alcanos, etc.). Un resultado

TESIS CON
FALLA DE ORIGEN

similar puede también ser alcanzado por medio de la aproximación Urey-Bradley, donde interacciones 1-3 (interacciones entre dos átomos ligados a un tercero común) son introducidas [29].

En mecánica molecular, obtener parámetros transferibles es de primera importancia, ya que un modelo simple sería aplicado a un gran conjunto de moléculas. Esto conduce a diferentes definiciones del campo de fuerza con, por ejemplo, la inclusión de los términos de Van der Waals.

Las longitudes de enlace y los ángulos de enlace son usualmente disponibles de información estructural ya existente. La cristalografía de rayos-X es una fuente principal para obtención de distancias interatómicas (difracción de electrones o neutrones, y para moléculas más pequeñas por espectroscopía de microondas). sin embargo, tales datos deben ser estudiados con mucha precaución. Las fuerza en las celdas cristalinas (lattice force) pueden modificar distancias derivadas de difracción de rayos-X, cuando las comparamos con medidas de difracción por electrones llevadas a cabo con moléculas aisladas en fase gaseosa. Sin embargo, estas diferencias son pequeñas y usualmente se desprecian. Los enlaces H-C constituyen una excepción notable, ya que la posición nuclear y el centro de densidad electrónica están separados por aproximadamente 1Å .

Los parámetros de variación de enlace pueden ser directamente derivados de constantes de fuerza vibracionales, mientras que para las constantes de doblamiento, los valores correspondientes tienen que ser escalados alrededor de 0.8 a 0.5. Los coeficientes de barreras torsionales pueden confiablemente ser estimados de las alturas de barreras (obtenidas por medio de espectroscopía de microondas o estudios termodinámicos, y más recientemente por medio de estudios con espectroscopía infrarroja y Raman).

Una parte más difícil es la evaluación de las repulsiones de Van der Waals, un punto realmente crucial ya que estas interacciones son de primordial importancia en determinar la estabilidad de moléculas altamente ramificadas. Además los parámetros de Van der Waals están altamente correlacionados con otros términos. Las moléculas en las cuales las interacciones de Van der

TESIS CON
FALLA DE ORIGEN

waals son grandes y generalmente complicadas, muestran muchos grados de libertad en los cuales la molécula se puede relajar [4].

Notemos también que, hablando únicamente de los términos de Van der Waals, hay alguna interdependencia entre el parámetro ϵ y el radio interviniendo en la ecuación de Hill, así, diferentes modelos pueden trabajar bien. Tal problema es ejemplificado en la definición del campo de fuerza AMBER (Assisted Model Building and Energy Refinement) [32], para carbonos $r^* = 1.81\text{\AA}$ y $\epsilon = 0.184\text{kcal/mol}$ son convenientes en un campo de fuerza 6-9. Sin embargo, en un potencial 6-12, estos valores tienen que ser modificados a $r^* = 2.175\text{\AA}$ y $\epsilon = 0.039\text{kcal/mol}$.

La elección de los radios de Van der Waals es un problema difícil, ampliamente discutido por Allinger [4]. Una primera fuente de información viene del estudio de datos de contacto de los gases raros, o en compuestos estudiados en la fase cristalina (por ejemplo, en el grafito, la distancia entre los planos de carbono resulta de las interacciones de Van der Waals). Verdaderamente, para los gases raros, estos valores realmente representan la suma de los radios atómicos de Van der waals. Sin embargo, para moléculas, la influencia de otros átomos inducen algunas interacciones atractivas y esto conduce a alguna interpenetración de los radios de Van der Waals para los átomos más cercanos. Tales distancias debidas a las aproximaciones en los cristales son fuentes de grandes tabulaciones [31].

Para la mecánica molecular, cálculos basados en el espaciado atómico en cristales alcanos selectos permitieron valorar los radios de Van der Waals para carbono e hidrógeno. Estimaciones razonables de los radios para otros átomos más comunes han sido propuestas [4]. Notemos que estos radios (usados en el campo de fuerza de Allinger, por ejemplo) son, en una aproximación burda, 0.3\AA mayor que los valores de longitud de enlace.

La optimización del campo de fuerza es un trabajo difícil, ya que este involucra muchos parámetros que no son completamente independientes unos de otros. Algunos procedimientos automáticos pero "ciegos" han sido propuestos donde la optimización es desarrollada por una computadora por medio

TESIS CON
FALLA DE ORIGEN

de cálculos de regresiones. Una aproximación iterativa donde el usuario monitorea los refinamientos sucesivos parece ser mas eficiente. Diversos protocolos de parametrización han sido propuestos [33, 34], el programa PEFF [35, 36] permite ajustar los parámetros del campo de fuerza a un conjunto de datos experimentales, usando un método de mínimos cuadrados. Ultimamente la conversión de campos de fuerza *ab-initio* a parámetros de energía de mecánica molecular han sido propuestos [37, 38].

Debido al aspecto empírico de estos métodos, podría ocurrir que algunos parámetros no podrían ajustar moléculas no previamente consideradas en el alcance de ciertos campos de fuerza. Los parámetros tomados de algún otro campo de fuerza no pueden ser introducidos directamente en el campo, ya que únicamente pueden ser utilizados como una guía, ya que cualquier campo de fuerza está definido como un todo para reproducir ciertos hechos experimentales y cada uno de sus términos por separados no son independientes.

Los valores de referencia en las ecuaciones (2.5)-(2.7) no son los valores observados experimentalmente en sistemas simples. Ellos son los valores que condujeron a predicciones correctas utilizando un campo de fuerza tomando en cuenta las otras contribuciones de energía. Por ejemplo, en el programa MM2, las longitudes de un enlace simple (usualmente 1.53 Å, 1.54 Å) son bien reproducidas utilizando un valor de referencia $r_o = 1.523\text{Å}$ (para el etano, el valor experimental es 1.534Å, mientras que el calculado es 1.532Å).

Más allá de las contribuciones fundamentales a el campo de fuerza (alargamiento, doblamiento de ángulo, torsión, interacciones entre átomos no ligados), términos adicionales pueden ser considerados para refinar la exactitud de las precisiones. Por ejemplo, la formulación Urey-Bradley incorpora interacciones 1-3 [29]. Alternativamente, algunos términos consideran variaciones simultáneas de dos parámetros, tales como cambios en la longitud de enlace y variaciones en el ángulo que ellos definen (este término de longitud-ángulo es importante en sistemas donde la longitud de enlace se varía cuando se comprime el ángulo). Tales términos tienen la forma general (para enlaces adyacentes i y j haciendo un ángulo θ_{ij}):

$$\frac{1}{2}k_{r(i,j,0)}[r_i - r_{j0} + r_j - r_{j0}][\theta_{ij} - \theta_{ij0}] \quad 2.13$$

2.10 Términos específicos

Algunos mecanismos adicionales pueden ser también tomados en cuenta. El enlace de polarización es débil en los alcanos y puede ser despreciado, pero este tiene que ser introducido en sistemas heteroatómicos. Estos términos electrostáticos son de primordial importancia para la evaluación de interacciones intermoleculares en moléculas polares. sin embargo, para una representación mas exacta, parece ser mejor tratar las interacciones electrostáticas utilizando un conjunto de cargas atómicas parciales. Así, un especial interés ha sido dedicado para tratar de determinar los valores de las cargas atómicas parciales.

Para reproducir los efectos electrostáticos, la Ley de Coulomb es utilizada con el uso de cargas puntuales localizadas en los núcleos:

$$V(e) = \sum \frac{q_i q_j}{r_{ij}} \frac{1}{D} \quad (2.14)$$

donde q_i y q_j son las cargas atómicas y r_{ij} la distancia interatómica para el par de átomos i, j . Una constante dieléctrica local D (usualmente variando de 1 a 4) puede introducirse.

Las cargas atómicas pueden ser derivadas de cálculos de química cuántica. Sin embargo, tales métodos (a nivel *ab-initio* o aún a un nivel semi-empírico o HMO) no son fácilmente disponibles para grandes moléculas como las proteínas. Además, el concepto de carga atómica está todavía abierto a discusión. El análisis de población de Mulliken, a menudo utilizado en métodos MO, ha sido ampliamente cuestionado ya que es gran dependiente del conjunto base utilizado. El método de cargas consistentes con el potencial (esto es, cargas puntuales centradas en átomos y definidas de tal forma que sean capaces de reproducir el potencial electrostático en una serie de puntos que rodean a la molécula [39] también ha sido cuestionado).

TESIS CON
FALLA DE ORIGEN

Así, una posible solución para tratar grandes moléculas (adoptada en AMBER) es llevar a cabo cálculo de mecánica cuántica en fragmentos y entonces “pegarlos” [32], la transferibilidad de sub-unidades se asume.

2.11 Enlaces de hidrógeno y efectos del medio

Los enlaces de hidrógeno merecen un especial interés debido a su importancia en la estabilidad de varias biomoléculas. Ellos pueden ser tratados como parte de las interacciones electrostáticas y de Van der Waals [49]. Para esta interacción diversas funciones han sido introducidas [50]-[54]. Para obtener un mejor resultado, ha sido propuesto reducir el radio de Van der Waals de los hidrógenos involucrados en puentes-H e incrementar el término atractivo hacia el átomo electronegativo alrededor de 1-3 kcal/mol, dependiendo del átomo particular involucrado. En AMBER, un potencial diferente (10-12 en vez de 6-12) se utiliza para hidrógenos involucrados en puentes-H [32].

Tomar en cuenta la influencia del solvente también es de primordial importancia para una predicción confiable de las energías de conformación. El modo más simple es tratar estas interacciones soluto-solvente con la ayuda de un modelo continuo el cual representa globalmente los efectos del medio sin considerar explícitamente las moléculas de solvente. Por ejemplo, constantes dieléctricas con valor 1-1.5 son generalmente utilizadas para estudios en la fase gaseosa, mientras que $D = 4$, posiblemente con una dependencia de la distancia (en el potencial AMBER que mostramos mas adelante) es usualmente mas apropiado para cristales o soluciones acuosas.

2.12 Energía de conformación

La suma de de todas estas contribuciones define la energía de conformación. Para un sistema molecular, considerado en una geometría dada, la energía de conformación representa la diferencia debida a deformaciones internas e interacciones entre átomos no ligados con respecto a un sistema hipotético donde todos los átomos están situados en el infinito y las interacciones son

nulas

La minimización de esta energía de conformación dará las conformaciones estables (optimización geométrica); la comparación entre estas energías de conformación asociadas con diversas conformaciones indican su estabilidad relativa. Las variaciones de esta energía de conformación dependiendo únicamente de uno o dos parámetros escogidos (generalmente ángulos de torsión) permiten dibujar perfiles energéticos o mapas para estudiar los procesos de inter-conversión. Este es uno de los usos más comunes en mecánica molecular

Sin embargo, para una comparación de datos energéticos con información experimental, así como con el estudio de diferentes moléculas (esto es, sistemas con diferente número de átomos o aún diferente topología), otros parámetros energéticos tienen que ser introducidos.

Los calores de formación, directamente comparables a datos experimentales, nos permiten juzgar la exactitud del método que se utilizó para obtener el campo de fuerza en términos de predicciones energéticas.

Para evaluar los calores de conformación, la energía de conformación tiene que ser completada por la contribución de entalpías de traslación, rotación y vibración obtenidas a partir de funciones de partición. Estas funciones de partición son teóricamente alcanzables vía mecánica estadística. Sin embargo, esta evaluación todavía está sujeta a fuertes controversias.

2.13 El modelo AMBER

La aproximación en mecánica molecular de Allinger et al. [4], Ermer y Lifson [93] comenzó con el estudio de hidrocarburos saturados vistos como moléculas aisladas (en solventes inertes o en fase gaseosa). De aquí surgió el interés por el mejor campo de fuerza para moléculas iónicas o polares en fases condensadas, particularmente para sistemas bio-orgánicos tales como proteínas y péptidos por un lado [102, 103], y para ácidos nucleicos por otro [104, 105]. El campo de fuerza AMBER, de Weiner, Kollman y demás colegas [32, 84], se puede describir como un campo de fuerza general desarrollado en un modo

TESIS CON
FALLA DE ORIGEN

consistente tanto para proteínas como ácidos nucleicos. Básicamente, aunque originalmente uso “átomos unidos” (se omitieron los hidrógenos unidos a carbonos), este campo de fuerza no es muy diferente en su expresión fundamental de la fórmula (2.4) indicada anteriormente. Similarmente, sus parámetros fueron primero extraídos de medidas físicas sobre compuestos cercanamente relacionados a las estructuras bajo investigación (microondas, difracción de neutrones, información de estructuras cristalinas, espectros de vibración, ..) y refinados a través de subsecuentes optimizaciones. El potencial semiempírico tiene la forma:

$$V = \sum_{bonds} K_R(R - R_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedral\ angles} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{332.17q_iq_j}{DR_{ij}} \right] + \sum_{II-bonds} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right]$$

Para las proteínas y ácidos nucleicos, moléculas relativamente planas, las funciones cuadráticas son suficientes para la variación en longitud de enlace o doblamiento de ángulo, junto a una serie de Fourier para la energía de torsión. Los términos de Van der Waals son expresados con un potencial 6-12. Se incluyen interacciones electrostáticas, así como la introducción de un potencial 10-12 complementario para describir los enlaces de hidrógeno. El radio y la profundidad del potencial de Van der Waals (interviniendo en A_{ij} y B_{ij}) son inter-dependientes y dependen de la expresión utilizada (potencial 6-9 o 6-12).

La evaluación del término electrostático involucra la característica interesante de que utiliza una constante dieléctrica dependiente de la distancia ($D = R_{ij}$) la cual se escogió para simular el hecho de que en medios acuosos, las interacciones electrostáticas se atenúan más rápidamente con la distancia que en fase gaseosa. En vez de usar cargas de Mulliken semi-empíricas (del tipo CNDO), los valores de las cargas atómicas se escogieron para ajustar un potencial electrostático *ab-initio* que rodea a la molécula [39, 106]. Estos cálculos fueron llevados a cabo en sub-unidades y entonces “pegadas”.

La elección del radio de Van der Waals es un problema difícil, una primera fuente de información viene del estudio de datos sobre los gases raros, o sobre

TESIS CON
FALLA DE ORIGEN

compuestos selectos en la fase cristalina (por ejemplo, en el grafito, la distancia entre planos de carbono resulta de las interacciones de Van der Waals). Verdaderamente, para los gases raros, estos valores representan realmente la suma de los radios de Van der Waals de los átomos.

Hagler et al. [109] compararon las funciones no ligadas 6-12 y 6-9 en cálculos de empaquetamiento cristalino y encontraron que ninguna es claramente superior. Mientras que la distancia interatómica no esté debajo de la suma de los radios de Van der Waals, la forma 6-12 será adecuada. Las cantidades R_{ij} , q_i y q_j son las distancias de no-enlace y las cargas respectivamente para el par de átomos i y j . El factor ϵ es un parámetro dieléctrico efectivo.

Weiner et al. usaron $D = R_{ij}$ con la razón de que usar una constante dieléctrica dependiente de la distancia nos ayuda a crear los efectos de polarización en interacciones atractivas, con interacciones más fuertes multiplicadas por un factor de peso mayor. Además, esto ayuda a compensar la pérdida de solvatación explícita al amortiguar más el rango mayor de las interacciones de carga y a su vez le da un mayor peso a las interacciones de rango más corto. Hay un soporte computacional y empírico para tal modelo, dado que el solvente (agua) en los experimentos no está explícitamente incluido en el cálculo. Sin embargo, cuando el agua está explícitamente incluida, una D constante es probablemente más apropiada.

Por ejemplo, constantes dieléctricas de 1-1.5 son generalmente usadas para fases gaseosas mientras que un valor de 4 es más apropiado para cristales o soluciones acuosas.

El último término es la energía de los enlaces de hidrógeno, la cual aparece en la unión de los pares de bases (adenina-timina y citosina-guanina).

Este campo de fuerza ha sido exitosamente evaluado en diversas situaciones tales como el plegamiento del azúcar furanosa, enlaces de hidrógeno, el apareamiento de bases y en el refinamiento de la insulina (una proteína de 500 átomos). Estos éxitos han comprobado que este modelo trabaja bien tanto para proteínas como para ácidos nucleicos [32]. Las unidades de V son

kcal/mol con R en angstroms, θ y ϕ en radianes, q_i en unidades electrostáticas (dada la carga Mulliken), y 332.17 es un factor de conversión.

Capítulo 3

LA BÚSQUEDA GLOBAL CON EL ALGORITMO GENÉTICO

Para realizar la búsqueda de configuraciones mínimas en la superficie del potencial utilizamos un algoritmo genético, el cual es una técnica de búsqueda inteligente. La forma en que lo implementamos se describe a continuación.

3.1 Introducción

los métodos analíticos de optimización buscan los óptimos locales al situarse en la superficie de la función y luego moverse en la dirección dada por el gradiente. Esta categoría de método incluiría a cualquier método donde la generación y evaluación de una conformación de una molécula dada está determinada por la conformación precedente en un modo determinista. Por ejemplo, los métodos de dinámica molecular son deterministas porque cada configuración es determinada por la precedente en un modo determinado por la ley de Newton.

Casi todos los problemas de optimización se pueden resolver, al menos en principio, por métodos analíticos; sin embargo, dificultades en el cómputo casi siempre aparecen cuando la dimensionalidad del sistema se vuelve grande, estos métodos analíticos tienen la restricción de que la función y sus derivadas deben ser continuas, también requieren tener una solución para los puntos

donde la primera derivada desaparece.

También existe la desventaja de que una función no siempre puede ser expresada en una forma matemática, como por ejemplo cuando los valores de la función son obtenidos por resultados experimentales, además de que aún contando con una función, muchas veces esta no puede ser derivable fácilmente.

Debido a estas condiciones restrictivas, diferentes métodos alternativos han sido desarrollados, entre los más comunes son los llamadas comúnmente métodos de búsqueda numéricos y las llamadas técnicas de búsqueda inteligentes.

Existen diversas técnicas de búsqueda numéricas, pero casi todas están basadas en la misma táctica: se escogen varios o un simple punto de comienzo de variables independientes (representados por cualquier solución posible al problema, no necesariamente la mejor), se calculan y comparan los valores de la función a optimizar, basado en esta comparación y dependiendo de cada método de búsqueda entonces se elige un nuevo conjunto de variables independientes y se prueba de nuevo la función para ver si este da un mejor valor, la elección se hace de tal forma que aquellos puntos que dieron un valor de la función con baja calidad son desechados, este proceso se vuelve iterativo, se re-evalúa la función y se re-examinan los valores de las variables independientes. Los varios métodos de búsqueda esencialmente difieren únicamente en la técnica de elección, pero básicamente en los métodos numéricos de optimización la elección de una nueva dirección de búsqueda es dirigida usualmente al azar o por intuición, así estos tienen la desventaja de que probablemente la búsqueda no será particularmente eficiente en el número de cálculos requeridos para alcanzar una precisión dada.

Las técnicas de búsqueda inteligentes están basadas en el método numérico, pero usan información obtenida del espacio de búsqueda para guiar la búsqueda. Las dos principales subclases son “simulated annealing” (recocido simulado) y algoritmos evolutivos o genéticos.

3.2 El Algoritmo genético

A finales de los años 50's y principios de la década de los 60's se desarrollaron las bases de los que después serían conocidos como algoritmos genéticos. Los algoritmos genéticos combinan la explotación de soluciones conocidas con la exploración de nuevas áreas del espacio de posibles soluciones. La eficacia de los algoritmos genéticos depende de una apropiada combinación entre la exploración y explotación de los datos (figura 3.1).

El algoritmo genético, originalmente propuesto por Holland [131], está basado en la técnica desarrollada por la naturaleza para resolver problemas al hacer evolucionar un código genético microscópico a través de mutación y entrecruzamiento, con la selección basada en la fineza de los correspondientes individuos. Esta es una técnica eficiente para optimizaciones globales multi-variables y ha sido aplicada exitosamente a un gran número de problemas complejos de la física [117] [118]

Los algoritmos genéticos son métodos de optimización basados en diversas estrategias tomadas de la evolución biológica. La primera de tales estrategias es el uso de una población en la cual los individuos que son mas "capaces" en algún sentido tienen mas posibilidades de sobrevivir y producir descendencia y así pasar su información genética a la siguiente generación. La segunda es el uso de entrecruzamiento en la cual el material genético del "hijo" es una mezcla de sus padres. La tercera es mutación, significando que el material genético es ocasionalmente alterado para mantener un cierto nivel de diversidad genética en la población. Los detalles del algoritmo genético que implementamos está descrito en la siguiente sección. Durante el proceso de búsqueda el algoritmo genético implícitamente aprende del pasado al escoger los individuos dependiendo de su fineza. Esta es la propiedad fundamental de los algoritmos genéticos.

La principal ventaja de los algoritmos genéticos radica en su sencillez. Se requiere poca información sobre el espacio de búsqueda ya que se trabaja sobre un conjunto de soluciones o parámetros codificados (hipótesis o individuos). Se busca una solución por aproximación de la población, en lugar

TESIS CON
FALLA DE ORIGEN

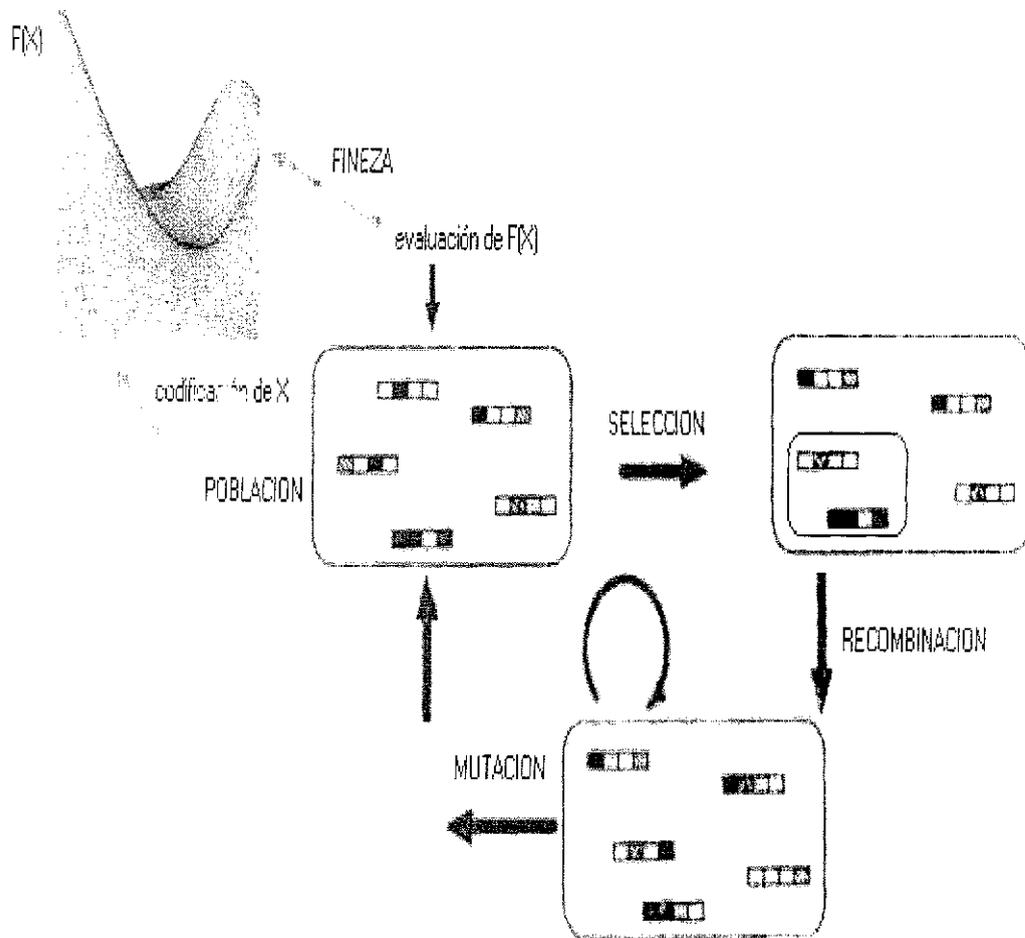


Figura 3.1: Flujo básico de un algoritmo evolutivo.

TESIS CON
FALLA DE ORIGEN

de una aproximación punto a punto.

3.3 La búsqueda global con el algoritmo genético

Los algoritmos genéticos generan un conjunto de “individuos” mediante la mutación y recombinación de parte del conjunto de “individuos” conocidos. En cada paso el conjunto de individuos conocido como “población actual” se renueva reemplazando una proporción de esta población por los sucesores de los individuos más “adecuados” (mediante el uso de una función de evaluación)

McGarrah y Judson [132] fueron los primeros en aplicar el algoritmo genético para tratar de determinar la conformación molecular de la molécula cíclica hexaglicina. Judson *et al*, [133] han observado la efectividad de los algoritmos genéticos en encontrar el estado base de un polímero plano comparada con las técnicas de búsqueda global de simulated annealing y al azar. Ellos llegaron a la conclusión general de que mientras más grande sea la molécula, mayor es la efectividad de los algoritmos genéticos sobre las otras aproximaciones. Diversos estudios han sido publicados mostrando que los algoritmos genéticos pueden ser usados para encontrar las conformaciones de baja energía para unas moléculas elegidas [2],[121],[126]-[130]. Un método para evaluar el éxito del algoritmo genético como método de búsqueda conformacional es comprobar si puede encontrar las conformaciones de baja energía naturales ya conocidas

La idea más importante es que nosotros trabajamos con una población de “individuos” los cuales van a interactuar a través de operadores genéticos para llevar a cabo un proceso de optimización.

El comienzo de la búsqueda global genética se llevó a cabo de la siguiente manera: primero generamos una población de “individuos” al azar de distintas configuraciones de los átomos que conforman las bases del ADN (nitrógenos, carbonos, oxígenos e hidrógenos). Estas son generadas en posiciones distribuidas al azar dentro de una esfera de radio $r = 5\text{Å}$ centrada en el origen de un sistema de coordenadas cartesianas tridimensional. Este radio

TESIS CON
FALLA DE ORIGEN

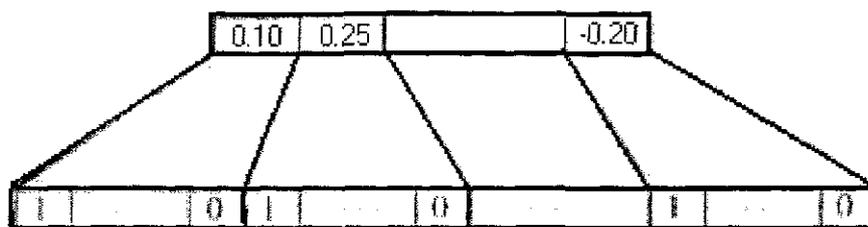


Figura 3 2: Digitalización de los valores reales (que representan las coordenadas de los átomos de una molécula) de un cromosoma.

es un poco mayor al de una esfera que encerraría por completo la conocida estructura “natural” de la base más grande, la guanina.

Los individuos en los algoritmos genéticos se suelen representar mediante cadenas de bits, de forma que puedan ser fácilmente manipulables por los operadores genéticos de mutación y cruce. Un individuo es especificado por un cromosoma el cual es una cadena de bits que corresponden a las coordenadas de una cierta configuración. A cada individuo de la población le corresponde una cadena genética, esta última se formó al juntar los 8 bits de las 3 coordenadas espaciales de cada uno de los átomos que conforman la molécula, siguiendo un orden preestablecido para los átomos (figura 3.2). Un código Gray [134] binario de 8 bits para las coordenadas fué escogido, implicando una discretización del espacio de $10\text{Å}/2^8$. Un entero que es representado como un número binario en código Gray tiene la propiedad de que para cambiar el valor en ± 1 solo es necesario cambiar un solo bit a la vez. En otras palabras, la “distancia hamming” = 1.

Como ejemplo, como la adenina tiene 14 átomos, entonces en una cadena hay $3 * 14 * 8 = 336$ bits que corresponden a las coordenadas espaciales de una configuración dada. Estas cadenas se hacen evolucionar de acuerdo a las técnicas estándar de los algoritmos genéticos empleando mutación y entrecruzamiento, con la selección basada en los resultados de una función

TESIS CON
FALLA DE ORIGEN

de fineza para la molécula [119], la cual, por supuesto, se tomó como la energía de interacción entre los átomos dada por el potencial de interacción. Las cadenas escogidas para formar parte de la siguiente generación son las que tengan más baja energía potencial.

Cada generación consta de una población formada por 10 individuos, subsecuentes generaciones son formadas como sigue Utilizando una estrategia “elitista”, los 5 individuos con más baja energía de la población son escogidos para ser “padres” de los individuos de la siguiente generación por medio de entrecruzamiento y mutación, solo el mejor individuo (el que tenga mas baja energía de los 10) no sufre mutación y entra elitístamente a entrecruzamiento, los 5 restantes individuos son desechados. Enseguida, pares de individuos (de los 5 individuos con más baja energía) son elegidos al azar y sus cromosomas son entrecruzados para formar los cromosomas de los individuos de la siguiente generación. El entrecruzamiento consiste de tomar algún subconjunto de los bits del padre 1 y el conjunto complementario de bits del padre 2 y combinar ellos para formar los cromosomas del hijo 1 (figura 3.3). Los restantes bits de los padres son combinados para formar los cromosomas del hijo 2. Es importante enfatizar que el entrecruzamiento combinado con la selección es la clave que distingue el algoritmo genético de otros métodos de búsqueda global estocásticos.

Adicionalmente, durante la reproducción hay una pequeña probabilidad (en nuestro caso fué 1/1000) de cambio de bit (mutación) en un cromosoma. Esta sirve principalmente para mantener la diversidad y prevenir convergencia prematura, la cual ocurre cuando un individuo predomina sobre la población entera en las primera etapas del proceso de evolución. Así, si cada individuo es especificado por una cadena de 336 bits y hay 10 individuos en la población para dar un total de 3360 bits, habrá entonces $3360 \cdot 0.001$ o alrededor de 3 mutaciones de un bit en cada generación. Por lo tanto, menos de 1 individuo en 10 sufre una mutación de un bit en cada generación.

La evolución de las cadenas prosigue hasta que la energía más baja de de la molécula no hubiese cambiado durante 7 generaciones consecutivas. Es

TESIS CON
FALLA DE ORIGEN

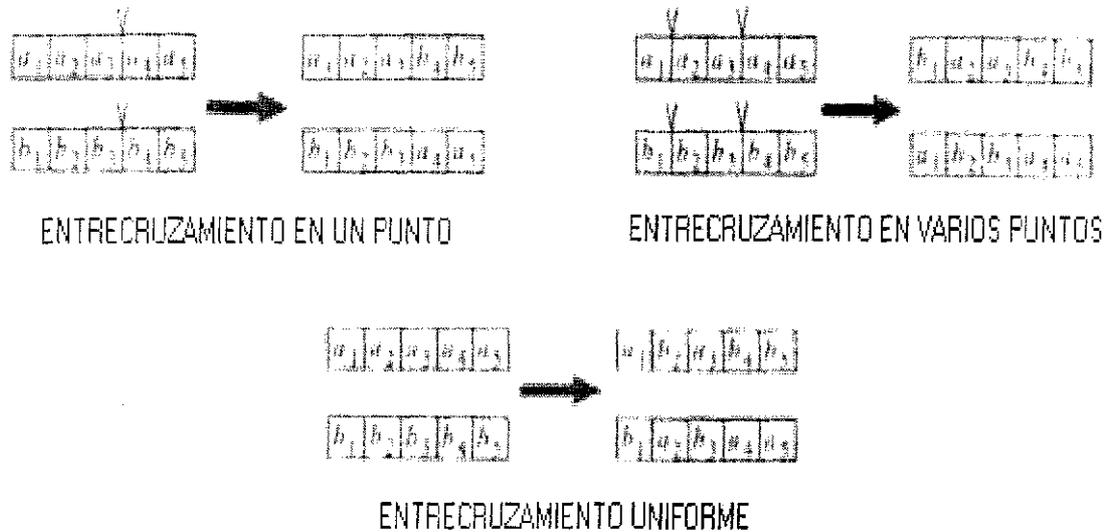


Figura 3.3: El operador de cruce genera dos nuevos hijos a partir de dos cadenas padre recombinando sus bits.

entonces cuando suponemos que la configuración ha caído en una cuenca de atracción (o mínimo local de energía) de la superficie del potencial y aplicamos gradiente conjugado.

El gradiente indica la mejor dirección para moverse a través de la superficie de energía potencial. El proceso iterativo del gradiente conjugado puede ser presentado como sigue: comenzando en un punto dado de la hipersuperficie de energía, uno camina en la dirección hacia la cual la energía disminuye mas; esta corresponde al opuesto del vector gradiente en ese punto.

En resumen, el algoritmo genético consiste transformar sucesivamente una generación de individuos a la siguiente utilizando las operaciones de selección, entrecruzamiento y mutación. Debido a que el proceso de selección apunta hacia individuos con más alta fitness, nuevos individuos son producidos con un parecido cada vez mas cercano a las soluciones óptimas de la función de interés.

Este trabajo utilizó el procedimiento descrito arriba implementado en el programa "EVOL" de Michaelian con Fortran90 [121, 122]. Un programa

TESIS CON
FALLA DE ORIGEN

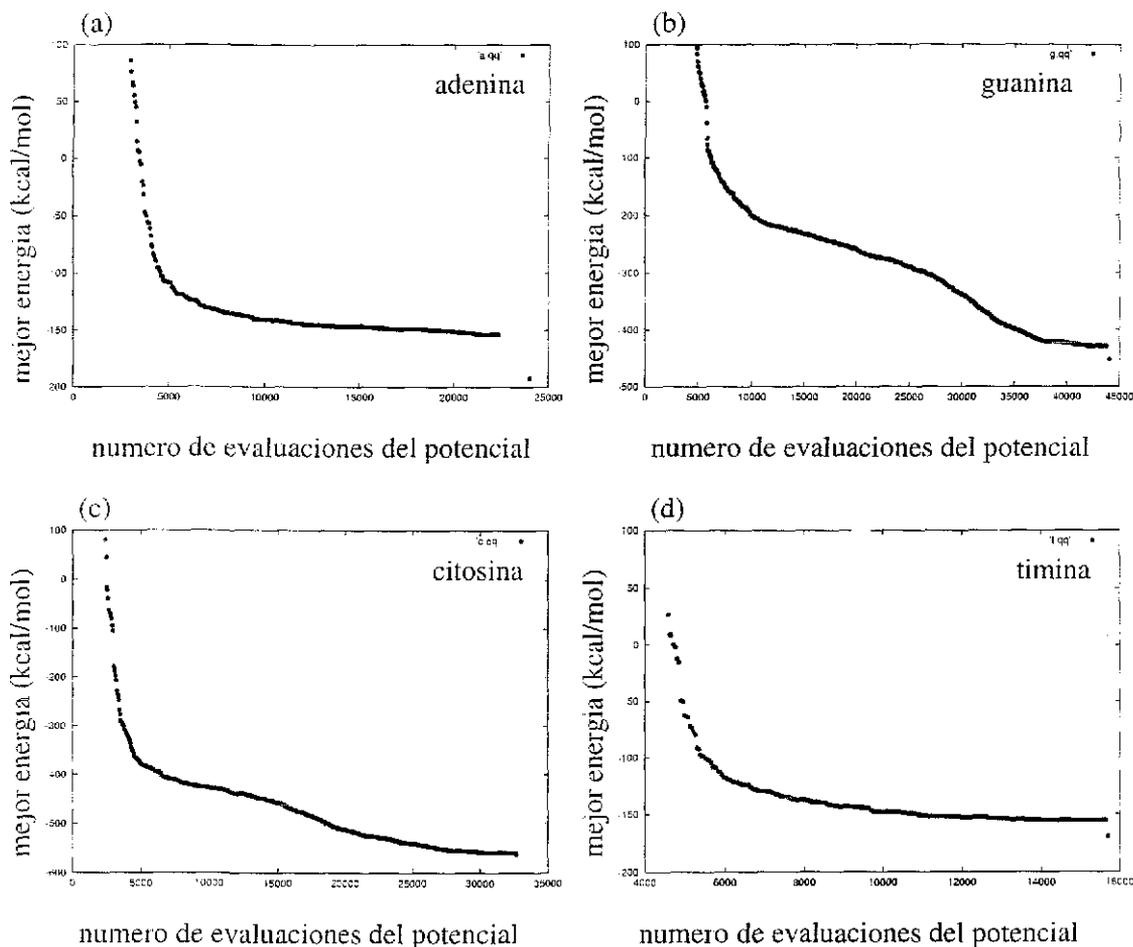


Figura 3.4: Optimización de la energía en una corrida del algoritmo genético para cada una de las bases. Se muestra como cambia la mejor energía (en este caso la menor) en kcal/mol en función del número de evaluaciones del potencial hechas por el algoritmo genético, en una generación se hacen 10 evaluaciones. El último punto corresponde al valor de energía después de aplicar gradiente conjugado, solo se introdujo como referencia.

TESIS CON
FALLA DE ORIGEN

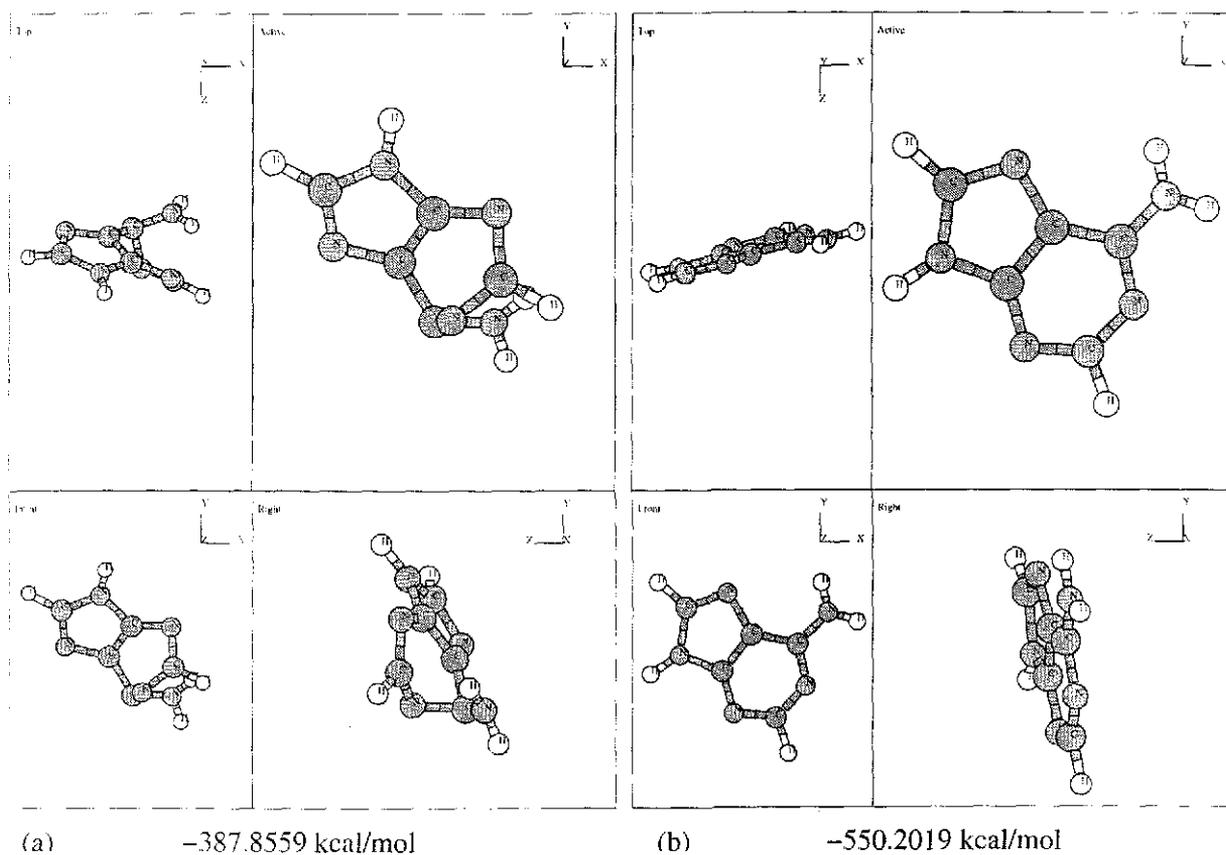


Figura 3.5: Comparación de la geometría y energía de la adenina antes y después de aplicar gradiente conjugado. (a) muestra un mínimo encontrado con el algoritmo genético y (b) muestra la geometría después de aplicar gradiente conjugado, en este caso convergió al estado base.

TESIS CON
FALLA DE ORIGEN

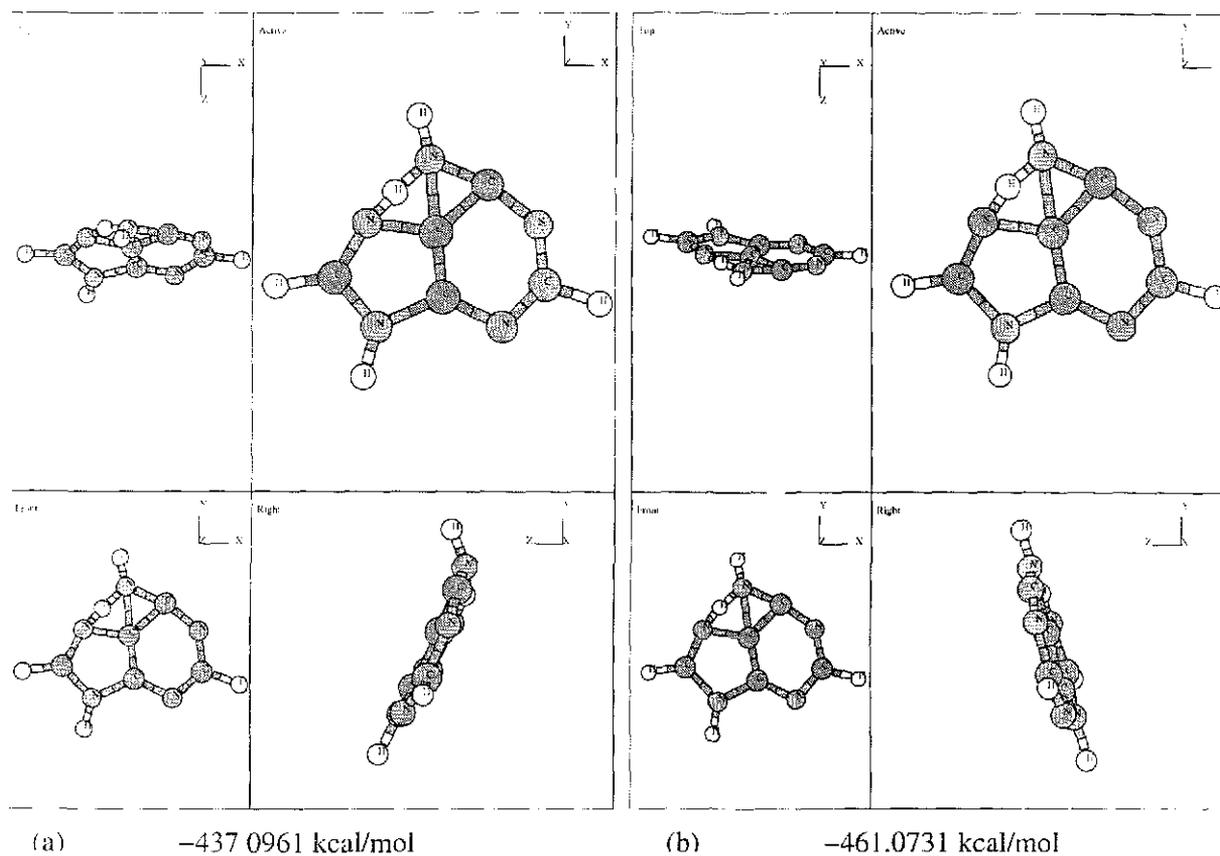


Figura 3.6: Comparación de la geometría y energía de la adenina antes y después de aplicar gradiente conjugado. (a) muestra un mínimo encontrado con el algoritmo genético y (b) muestra la geometría después de aplicar gradiente conjugado. Se observa como al variar un poco la posición de un átomo H la estructura se vuelve plana y es diferente del estado base.

TESIS CON
FALLA DE ORIGEN

para graficar en tres dimensiones llamado Xmol [138], al cual se suministran los datos (coordenadas) obtenidas mediante la optimización genética, permite visualizar las geometrías correspondientes de las conformaciones de mínima energía.

3.4 Resultados

En la Figura 3.4 mostramos como la menor energía de la población cambia durante las sucesivas generaciones en el transcurso de una corrida con los algoritmos genéticos para adenina, guanina, citosina y timina, respectivamente. Se muestra como la energía es alta y positiva durante la primera etapa de la corrida, pero disminuye al principio exponencialmente. Los detalles de las curvas mostrados en la figura 3.4 difieren por cada corrida distinta, pero el patrón es esencialmente el mismo. La población comienza con una amplia distribución de energías que convergen en un pequeño número de generaciones conforme el algoritmo genético busca en las regiones de baja energía del espacio de conformación y las configuraciones con alta energía son removidas de la población.

Para optimizar los detalles de la conformación en la última generación, una optimización local del tipo gradiente conjugado Polak-Ribiere [140] se llevó a cabo, ver figura 3.5 y 3.6. La última evaluación en la energía en la figura 3.4 corresponde al valor que se obtuvo después de aplicar la minimización con gradiente conjugado, se puede observar un mejoramiento significativo del valor final con respecto al último valor obtenido con el algoritmo genético en algunos casos, esta solo se introdujo como referencia pero no tiene nada que ver con el número de evaluaciones.

El procedimiento así descrito nos permitió hacer converger el proceso iterativo en cualquier precisión de la energía de configuración (aquí se escogió $\Delta V/V = 10^{-8}$). Finalmente, para checar que la configuración no estuviera en un punto de silla y que fuera realmente una estructura estable, se obtuvo y diagonalizó el Hessiano a una alta precisión numérica y se descartaron aquellas estructuras que presentaron eigenvalores negativos (puntos de silla).

En 13.000 corridas del algoritmo genético para cada base con la función del potencial, comenzando de configuraciones generadas completamente al azar, se encontraron más de 100 estructuras estacionarias (mínimos o puntos de silla) para cada base.

En las tablas 3.1 - 3.4 se muestran las energías y el número de veces que se encontraron para las 20 configuraciones de más baja energía para guanina, adenina, guanina, citosina y timina, respectivamente. En las figuras 3.7- 3.14 se muestran el número de veces que se encontró una configuración con una energía dada en las 13000 corridas del algoritmo genético para la adenina, citosina, guanina y timina. El número de veces que se encontró una estructura mínima local o punto de silla es proporcional, en primera aproximación, al tamaño de su cuenca de atracción sobre la superficie de energía potencial

La configuración con energía mas baja encontrada correspondió a la bien conocida configuración "natural" para todas las bases, y esta configuración fué la que se encontró un mayor número de veces, lo cual significa que la conformación natural tiene la cuenca de atracción más grande y más profunda en la superficie de potencial

La mayoría de las configuraciones encontradas tienen una estructura gemela formada por el intercambio de los átomos de hidrógeno en el enlace H-N-H. Esto en realidad es un artificio debido a las distintas cargas asignadas a los dos átomos de hidrógeno y al hecho de que las cargas no fueron optimizadas en la aproximación del potencial modelo. Por ejemplo, en el caso de la guanina, hubo 2 conformaciones (una con energía -650.35346 kcal/mol y otra con energía -650.23180 kcal/mol) y ambas corresponden al estado natural, lo cual significa que el estado base se encontró $1528+1551= 3079$ veces en las 13000 corridas del algoritmo genético, situaciones similares ocurrieron para las otras bases.

Tabla 3.1. Las energías y número de veces encontrados de los 20 más bajos puntos estacionarios encontrados en 13000 optimizaciones globales de configuraciones iniciales al azar de los átomos de la **adenina**, utilizando el potencial de campo de fuerza con el algoritmo genético de búsqueda.

adenina Energía kcal/mol	número de veces encontrados
-550.20188	2087
-550.15406	1950
-497.41885	1
-497.41721	2
-497.41621	2
-497.41225	2
-497.40997	1
-497.40868	1
-497.40651	5
-497.40011	2
-497.39840	1
-497.39717	1
-497.39578	1
-497.38930	1
-497.37284	1
-497.37004	1
-497.36620	1
-497.36389	3
-497.36004	1
-497.35891	2

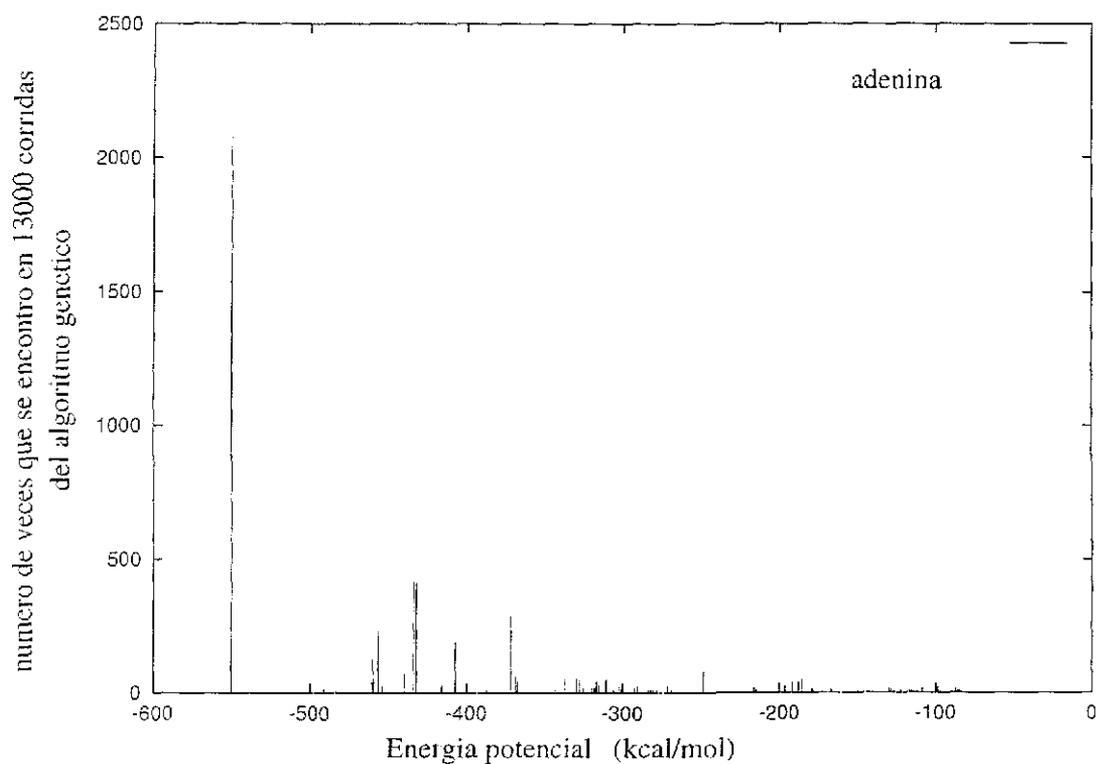


Figura 3.7: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base adenina.

TESIS CON
FALLA DE ORIGEN

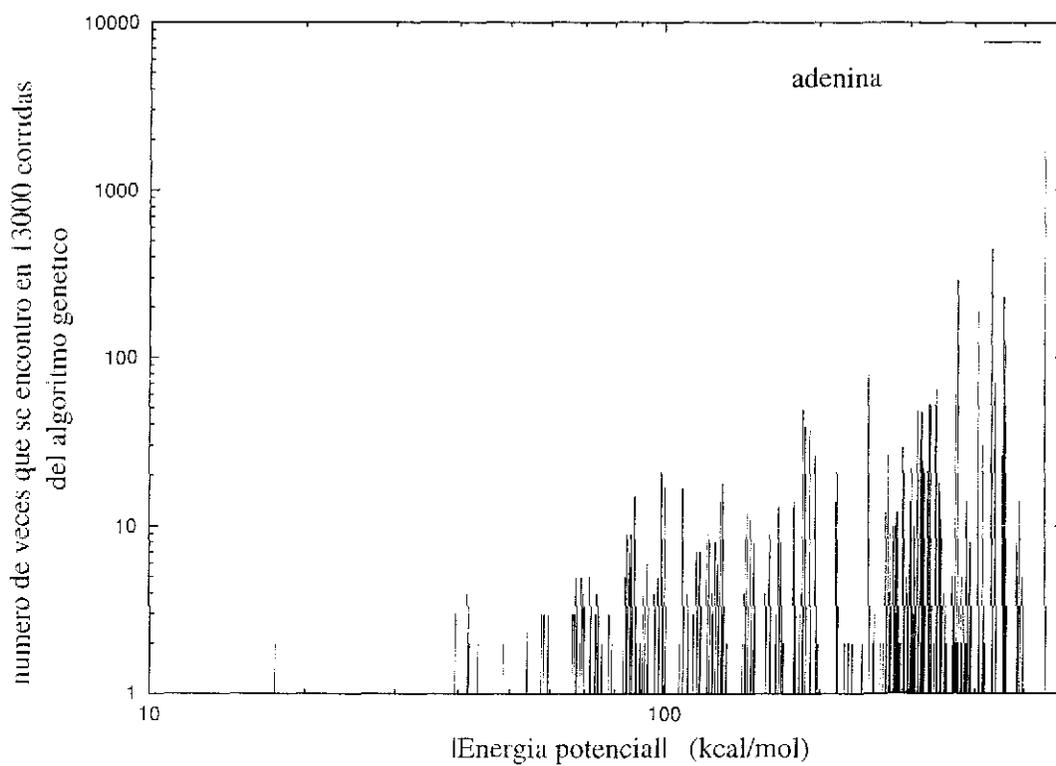


Figura 3 8: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base adenina (escala logarítmica).

TESIS CON
FALLA DE ORIGEN

Tabla 3.2. Las energías y número de veces encontrados de los 20 más bajos puntos estacionarios encontrados en 13000 optimizaciones globales de configuraciones iniciales al azar de los átomos de la **guanina**, utilizando el potencial de campo de fuerza con el algoritmo genético de búsqueda.

guanina Energía kcal/mol	número de veces encontrados
-650.35346	1528
-650.23180	1551
-650.23077	174
-599.52620	25
-599.52509	34
-599.52407	16
-599.52300	4
-599.41714	34
-599.41600	30
-599.41374	1
-597.97491	1
-597.83155	1
-597.79810	1
-597.62898	1
-597.44707	1
-597.31491	1
-597.27242	1
-597.25470	1
-596.92225	1
-596.90662	1

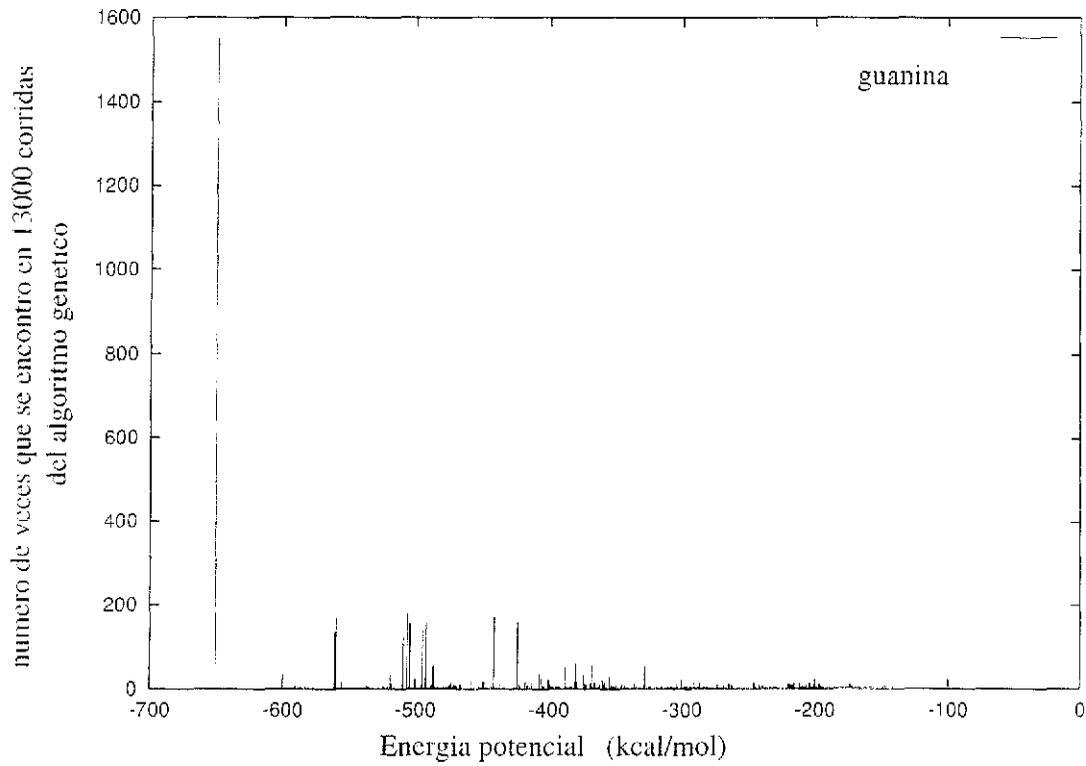


Figura 3.9: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base guanina.

TESIS CON
FALLA DE ORIGEN

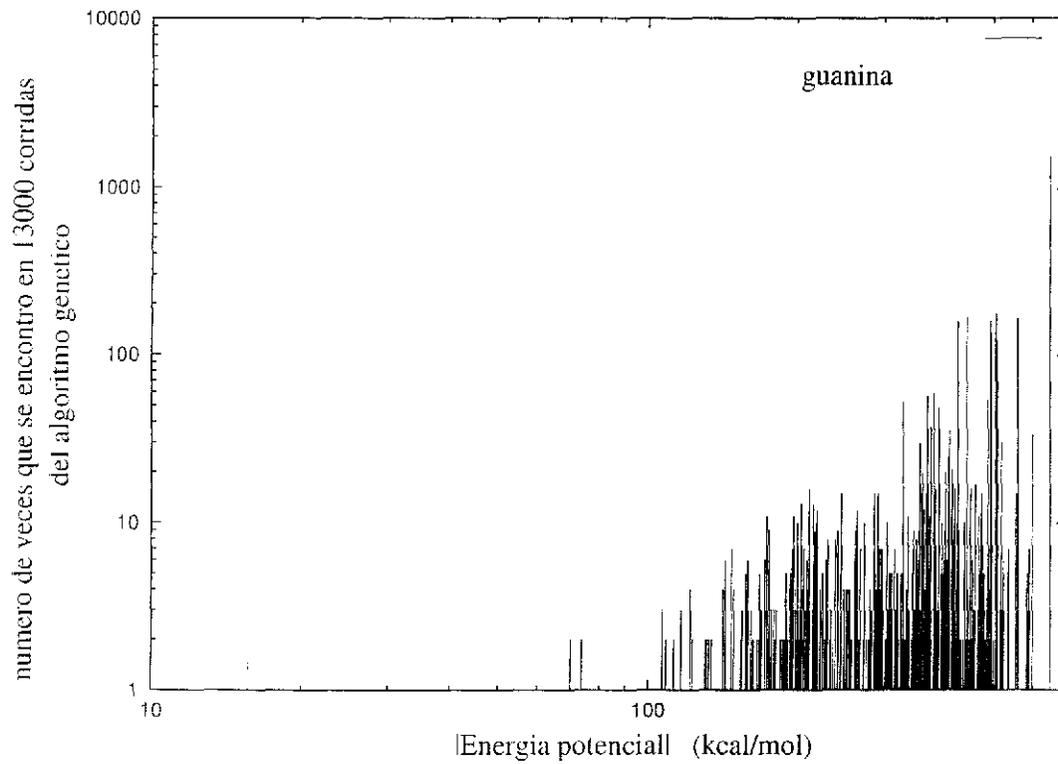


Figura 3.10: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base guanina (escala logarítmica).

TESIS CON
FALLA DE ORIGEN

Tabla 3.3. Las energías y número de veces encontrados de los 20 más bajos puntos estacionarios encontrados en 13000 optimizaciones globales de configuraciones iniciales al azar de los átomos de la **citosa**, utilizando el potencial de campo de fuerza con el algoritmo genético de búsqueda.

citosa Energía kcal/mol	número de veces encontrados
-685.29793	980
-685.11157	1046
-648.55192	42
-627.65731	4
-626.12521	4
-620.66303	30
-620.45335	23
-615.63634	128
-615.50962	112
-615.43993	5
-615.43154	7
-615.43044	9
-615.42937	12
-615.42770	17
-615.42669	6
-615.42551	3
-615.42316	6
-615.42157	5
-615.41972	4
-615.41784	1

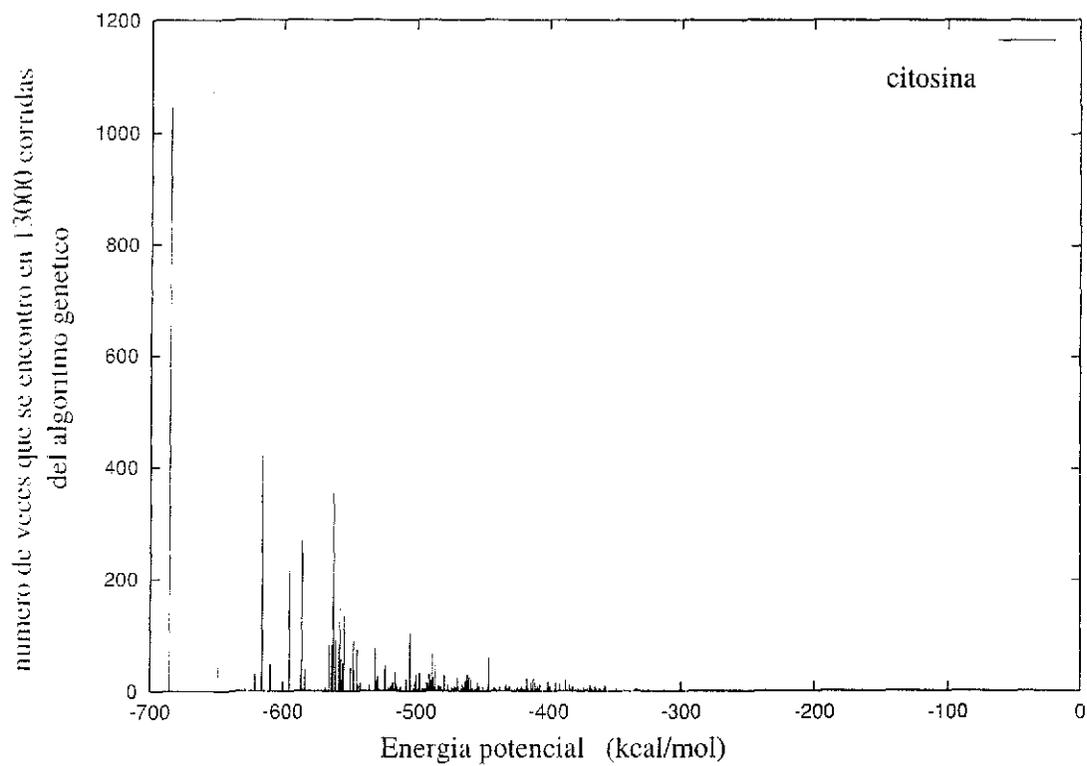


Figura 3.11: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base citosina.

TESIS CON
FALLA DE ORIGEN

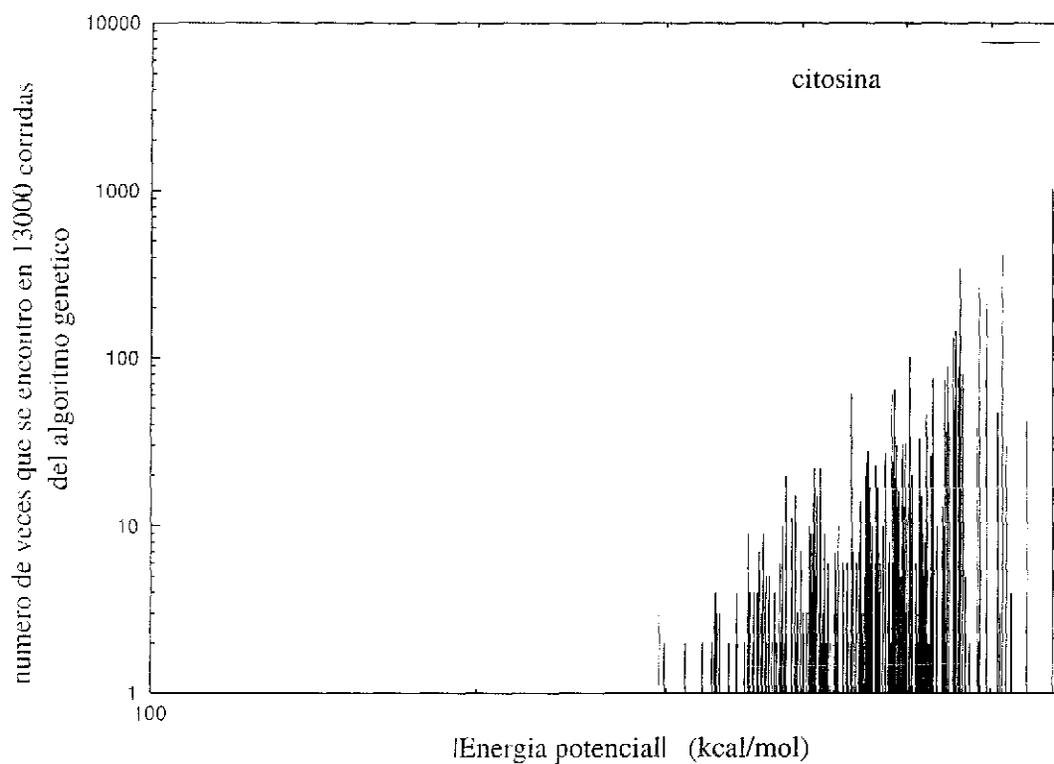


Figura 3.12: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base citosina (escala logarítmica).

TESIS CON
FALLA DE ORIGEN

Tabla 3.4. Las energías y número de veces encontrados de los 20 más bajos puntos estacionarios encontrados en 13000 optimizaciones globales de configuraciones iniciales al azar de los átomos de la **timina**, utilizando el potencial de campo de fuerza con el algoritmo genético de búsqueda.

timina Energía kcal/mol	número de veces encontrados
-476.13629	215
-476.13528	332
-476.13420	211
-476.13310	140
-476.13209	87
-476.13073	52
-476.12943	41
-476.12837	35
-476.12736	34
-476.12611	36
-476.12508	27
-476.12385	18
-476.12235	3
-476.11967	102
-476.11837	98
-476.11708	73
-476.11570	36
-476.11385	26
-476.11185	18
-476.11012	21

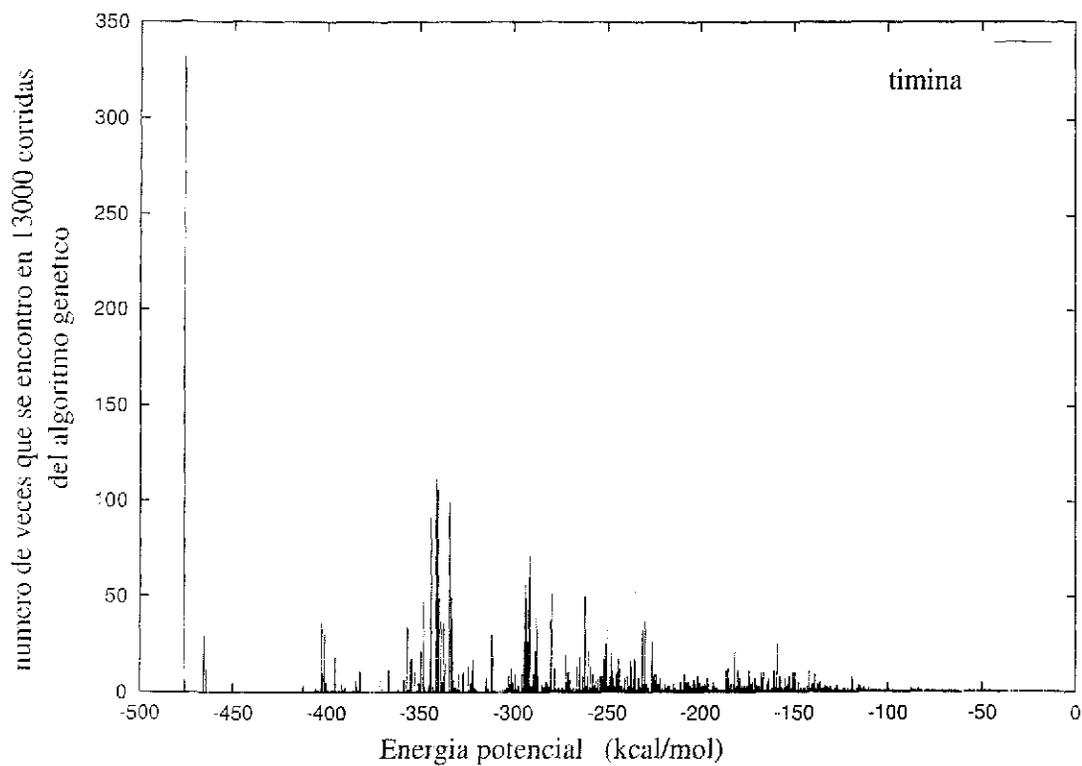


Figura 3 13: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base timina

TESIS CON
FALLA DE ORIGEN

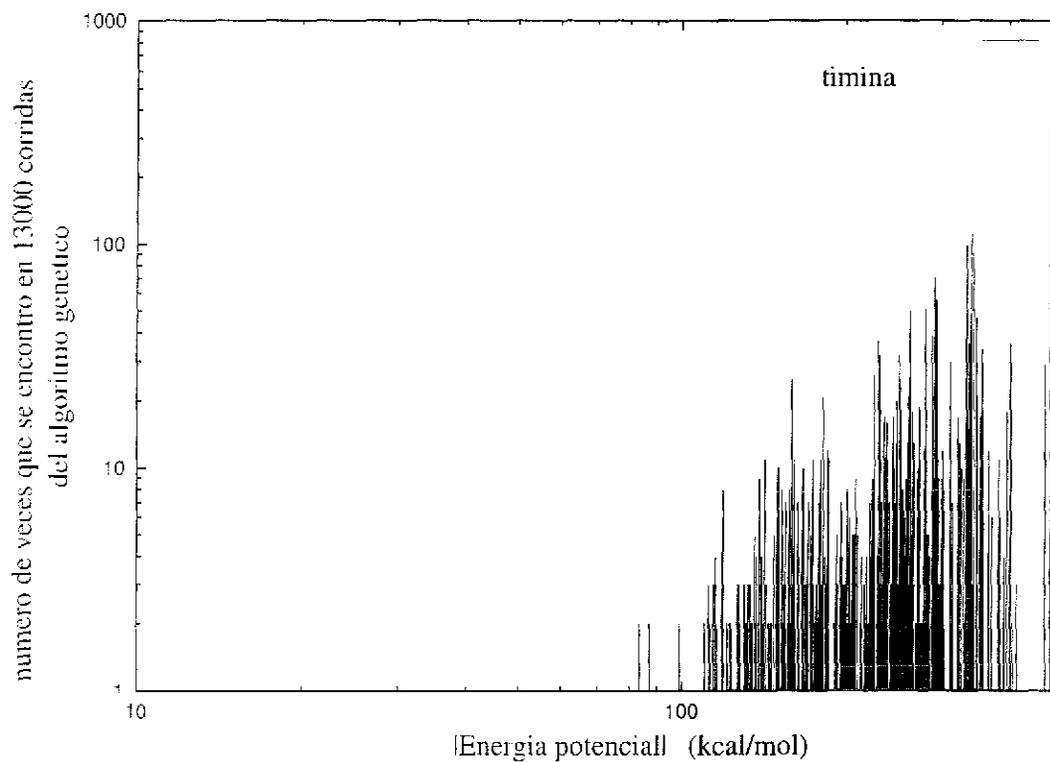


Figura 3.14: Distribución de mínimos de energía que se encontraron en 13000 corridas del algoritmo genético para la base timina (escala logarítmica).

TESIS CON
FALLA DE ORIGEN

3.5 Mínimos del potencial que convergieron a estructuras estables con una posterior optimización local

A continuación mostramos las configuraciones mínimas que se escogieron de las corridas de los algoritmos genéticos y que posteriormente utilizando Gaussian98 [136] para una optimización cuántica local convergieron a estructuras estables, es decir, las coordenadas de las conformaciones siguientes sirvieron como conformaciones de entrada para llevar a cabo la optimización con Gaussian98, siempre la estructura con menor energía correspondió a la conformación natural y son las primeras que se muestran: A1,G1,C1 y T1 (figuras 3.15, 3.19, 3.27 y 3.31 respectivamente).

Aunque se escogieron más de 10 mínimos locales del potencial por base, y a todos estos los optimizamos cuánticamente, solo los mostrados enseguida convergieron con optimización cuántica a mínimos diferentes del estado base y además son estables.

En el proceso de selección de configuraciones tomamos aquellas que no se alejaban mucho de las naturales, que presentaran enlaces entre átomos factibles (por ejemplo, hubo muchas configuraciones en donde los átomos de hidrógeno presentaban doble enlace con otros átomos, donde aparecían átomos en el centro de los anillos, todas estas configuraciones se descartaron), que no presentaran anillos rotos, etc.

Para la adenina mostramos las tres configuraciones de entrada denominadas en adelante A2-A4 distintas de la natural (figuras 3.16- 3.18); para la guanina mostramos seis configuraciones (G2-G7) distintas de la natural (figuras 3.20- 3.26); para la citosina fueron tres configuraciones (C2-C4) (figuras 3.28- 3.30) y para la timina sólo mostramos la configuración natural (T1) (figura 3.31) porque con las optimizaciones posteriores todas las que se escogieron convergieron a esta.

Para verificar que estas conformaciones fueran mínimos del potencial se determinó y diagonalizó numéricamente el Hessiano a una alta precisión

TESIS CON
FALLA DE ORIGEN

numérica. Todas estas configuraciones tuvieron todos sus eigenvalores positivos, con seis eigenvalores igual a cero por los seis grados de libertad traslacional y rotacional no removidos durante el proceso de optimización.

Aquí enfatizamos que el potencial semiempírico fué ajustado a partir de datos experimentales de las configuraciones “naturales”. Aunque el número y tipos de moléculas fué grande, es probable que éste potencial nos dé una buena descripción únicamente para las configuraciones similares a las naturales.

TESIS CON
FALLA DE ORIGEN

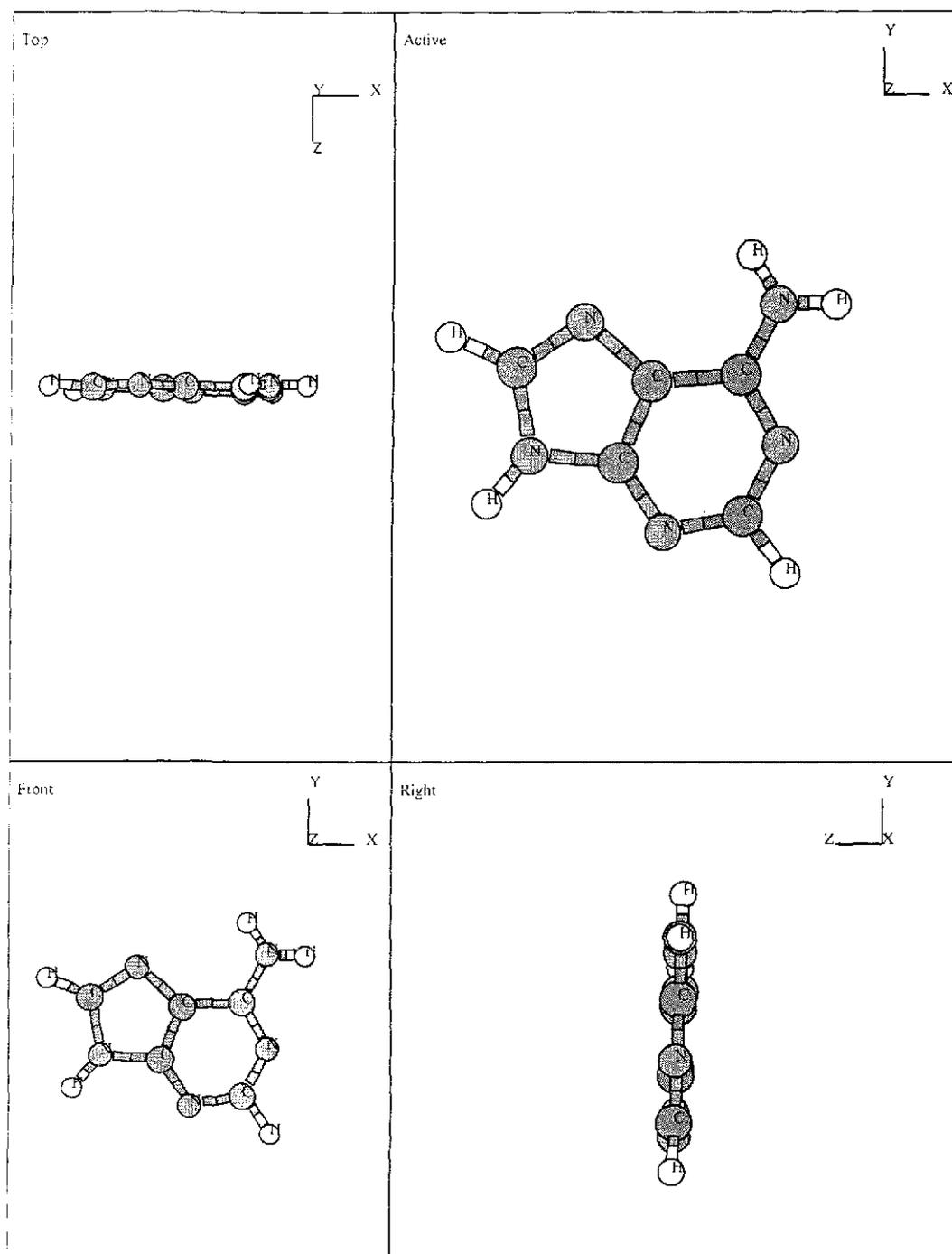


Figura 3.15: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base adenina. A1, energía potencial= -550.201 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la adenina: 4037.

TESIS CON
FALLA DE ORIGEN

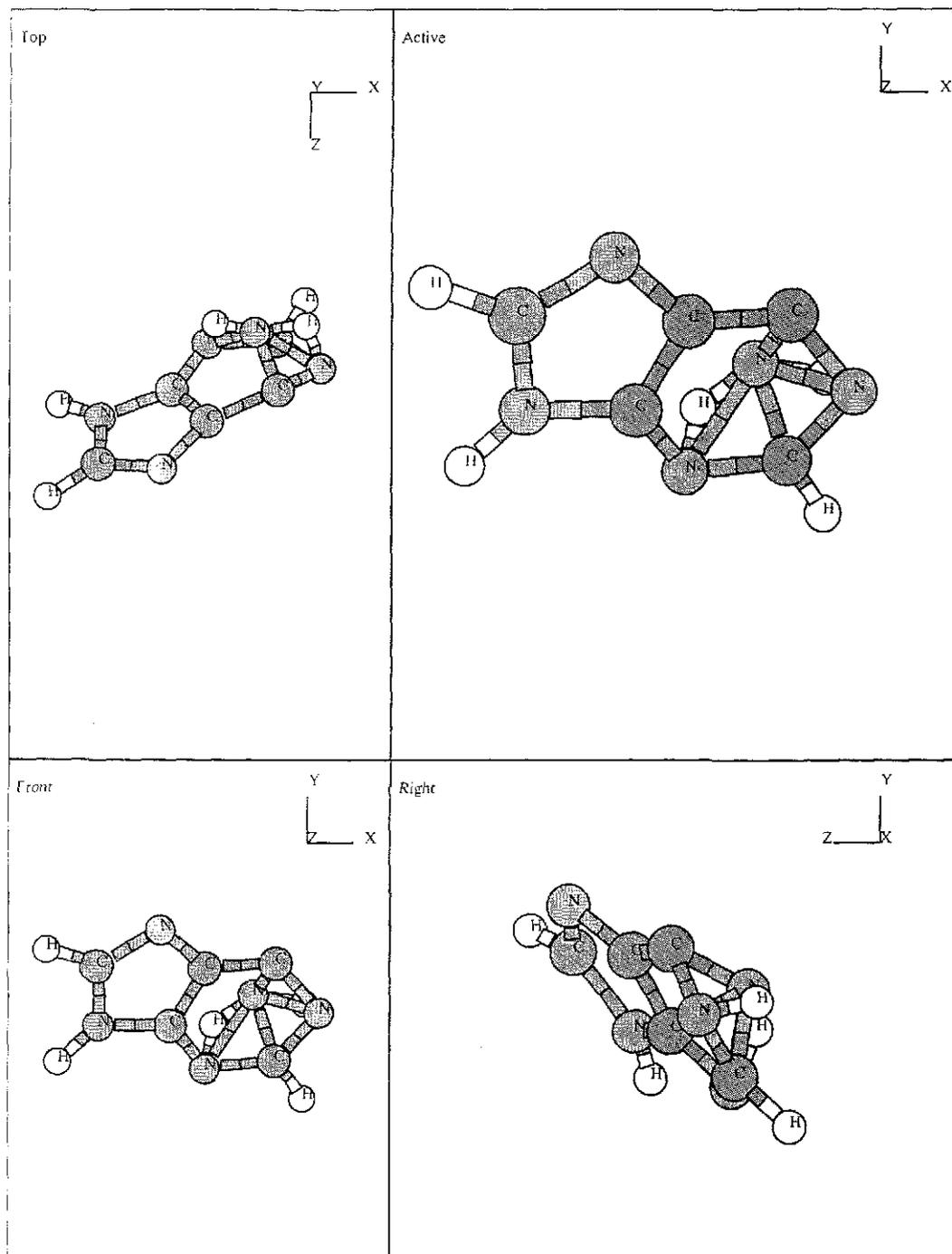


Figura 3 16: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base adenina. A2, energía potencial= -439.897 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la adenina: 127.

TESIS CON
FALLA DE ORIGEN

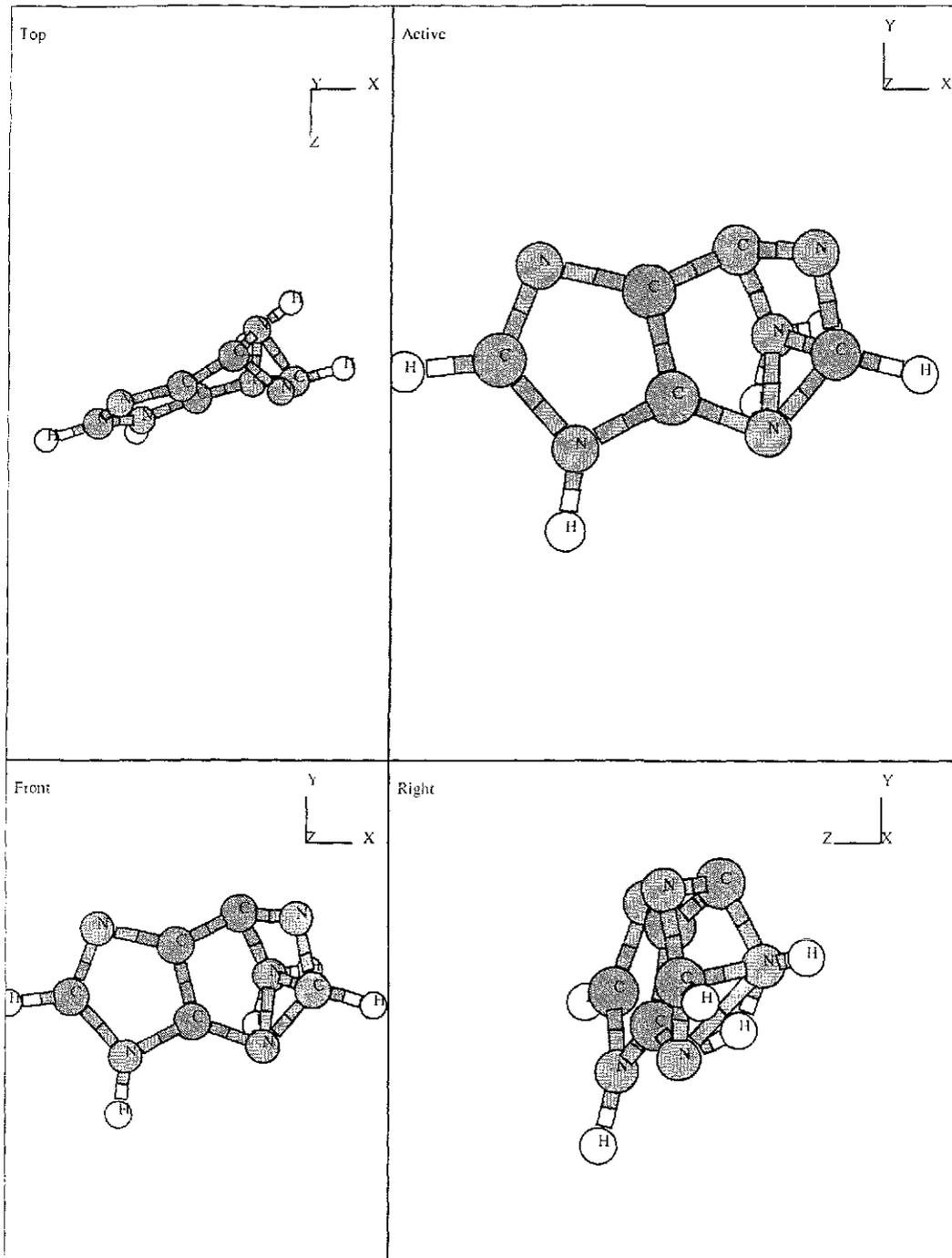


Figura 3.17: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base adenina A3, energía potencial= -433.733 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la adenina: 448.

TESIS CON
FALLA DE ORIGEN

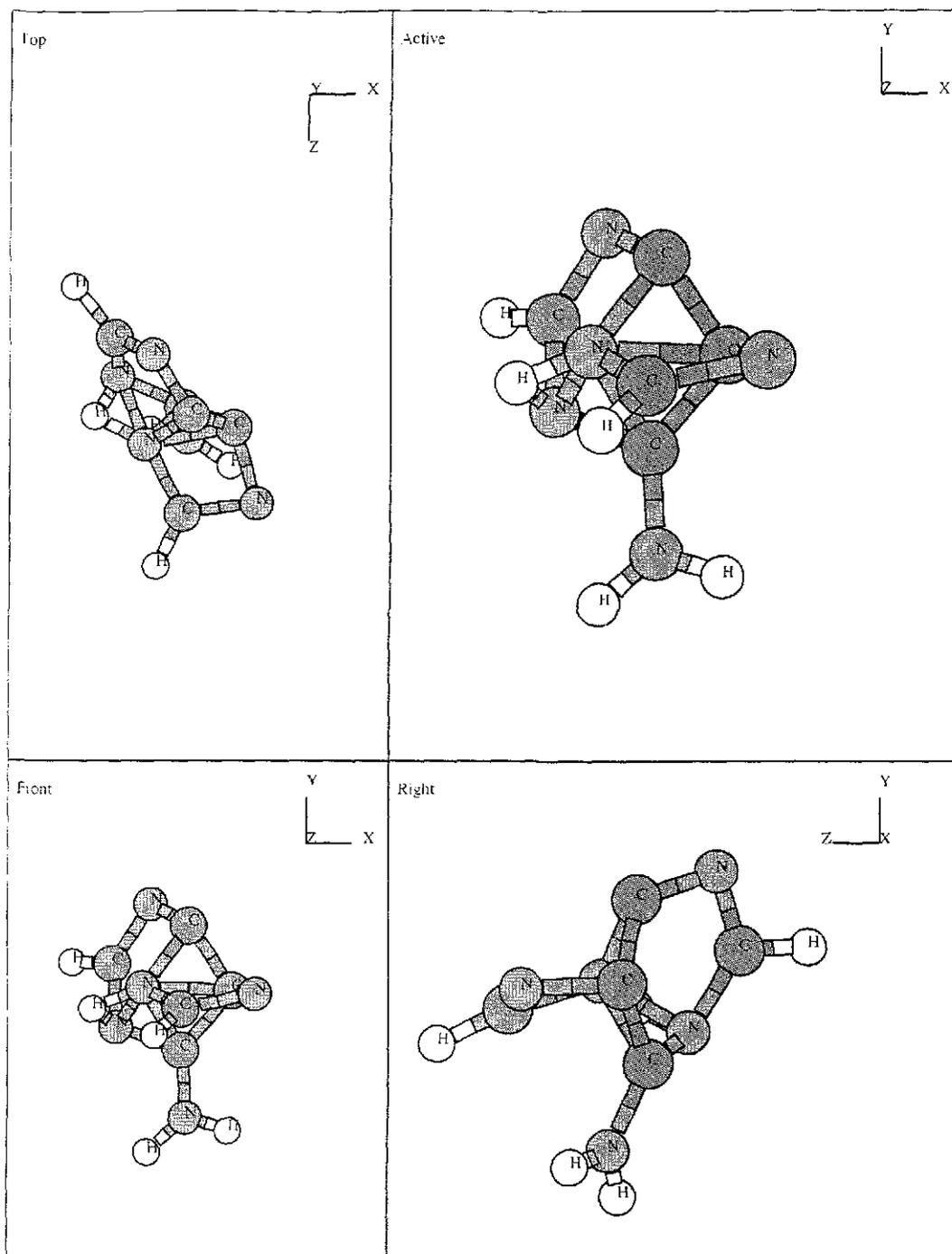


Figura 3.18: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base adenina A4, energía potencial= -372.105 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la adenina: 570.

TESIS CON
FALLA DE ORIGEN

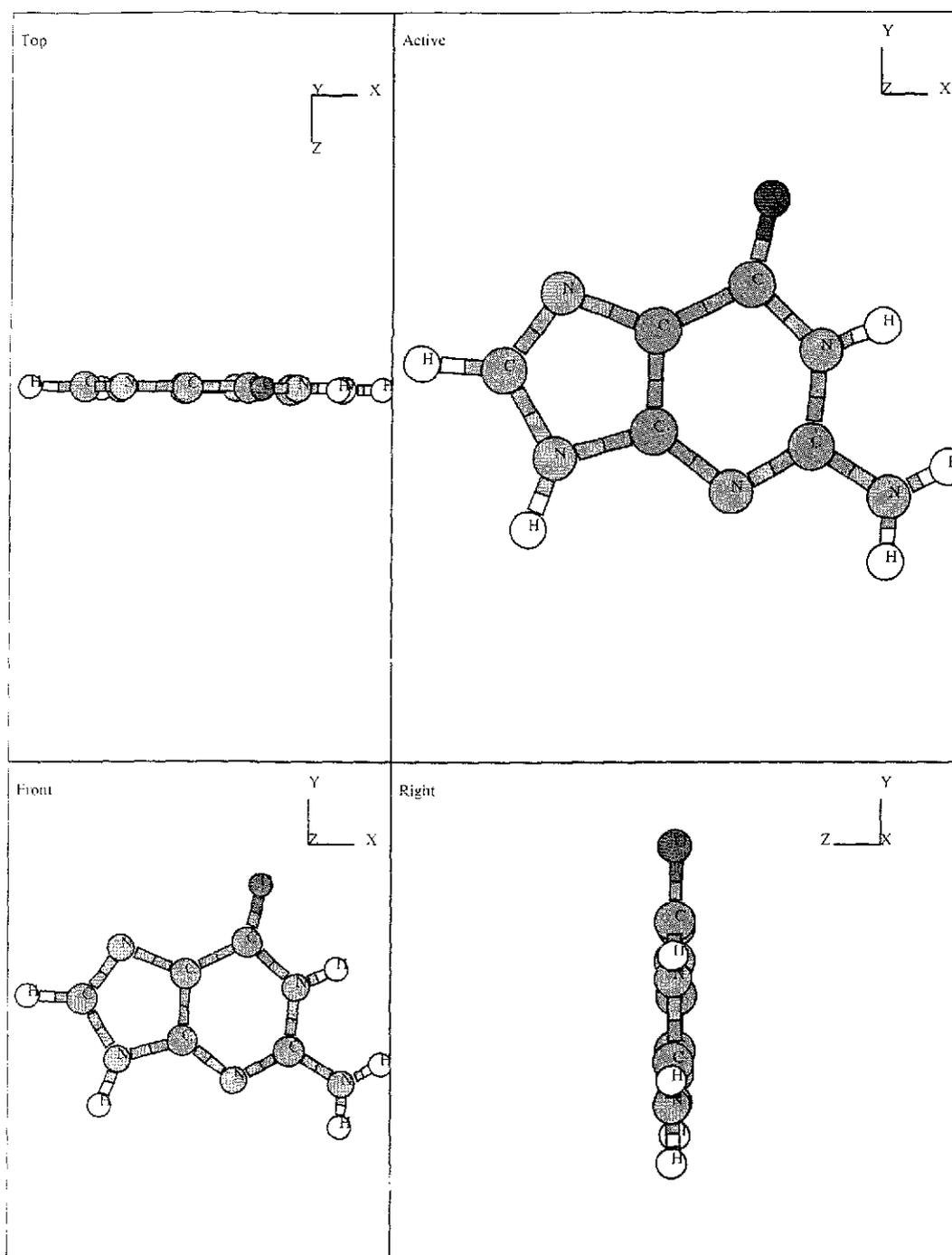


Figura 3.19: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G1, energía potencial= -650.353 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 3079.

TESIS CON
FALLA DE ORIGEN

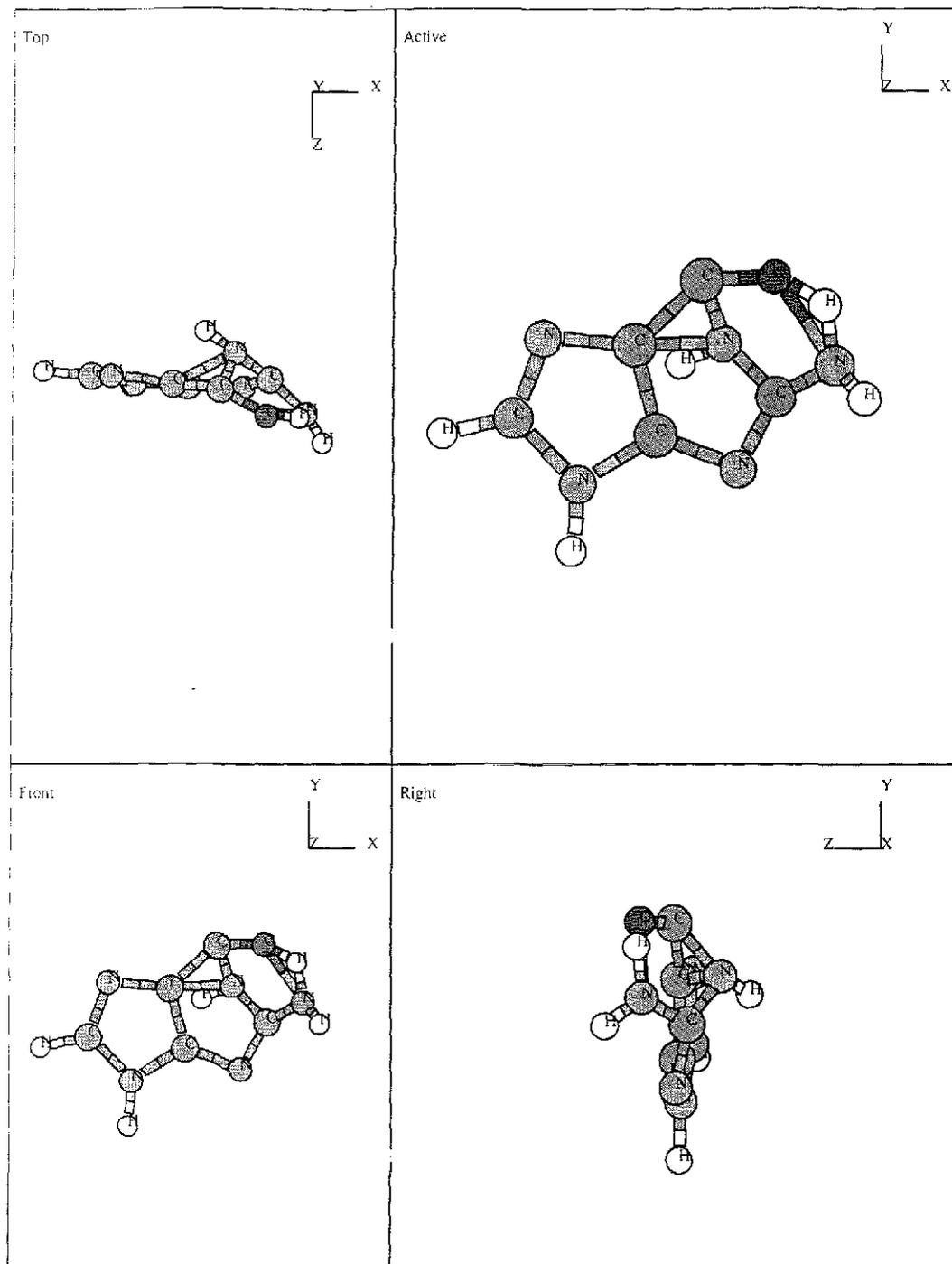


Figura 3.20: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G2, energía potencial = -505.953 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 178.

TESIS CON
FALLA DE ORIGEN

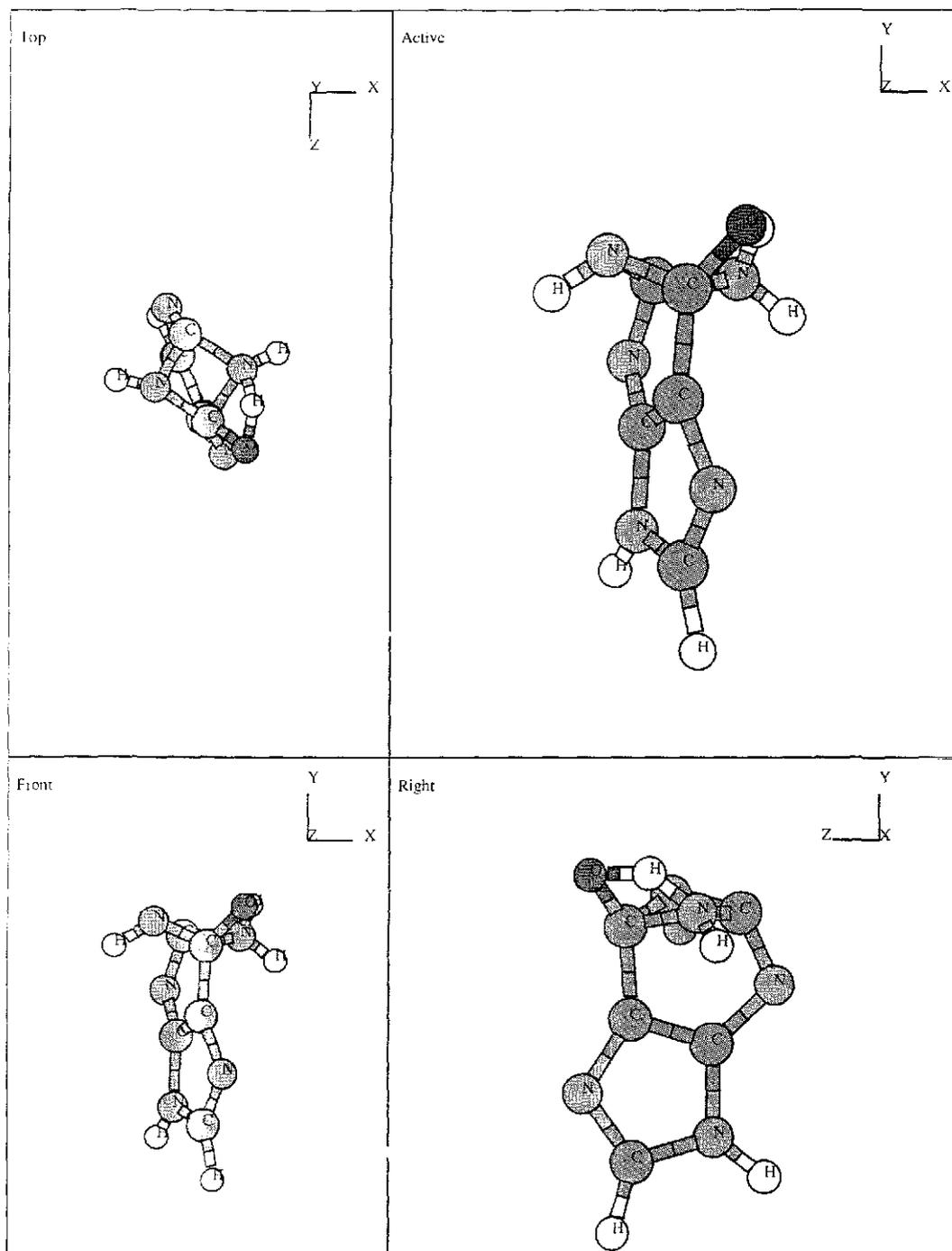


Figura 3.21: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G3, energía potencial= -486.428 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 62.

TESIS CON
FALLA DE ORIGEN

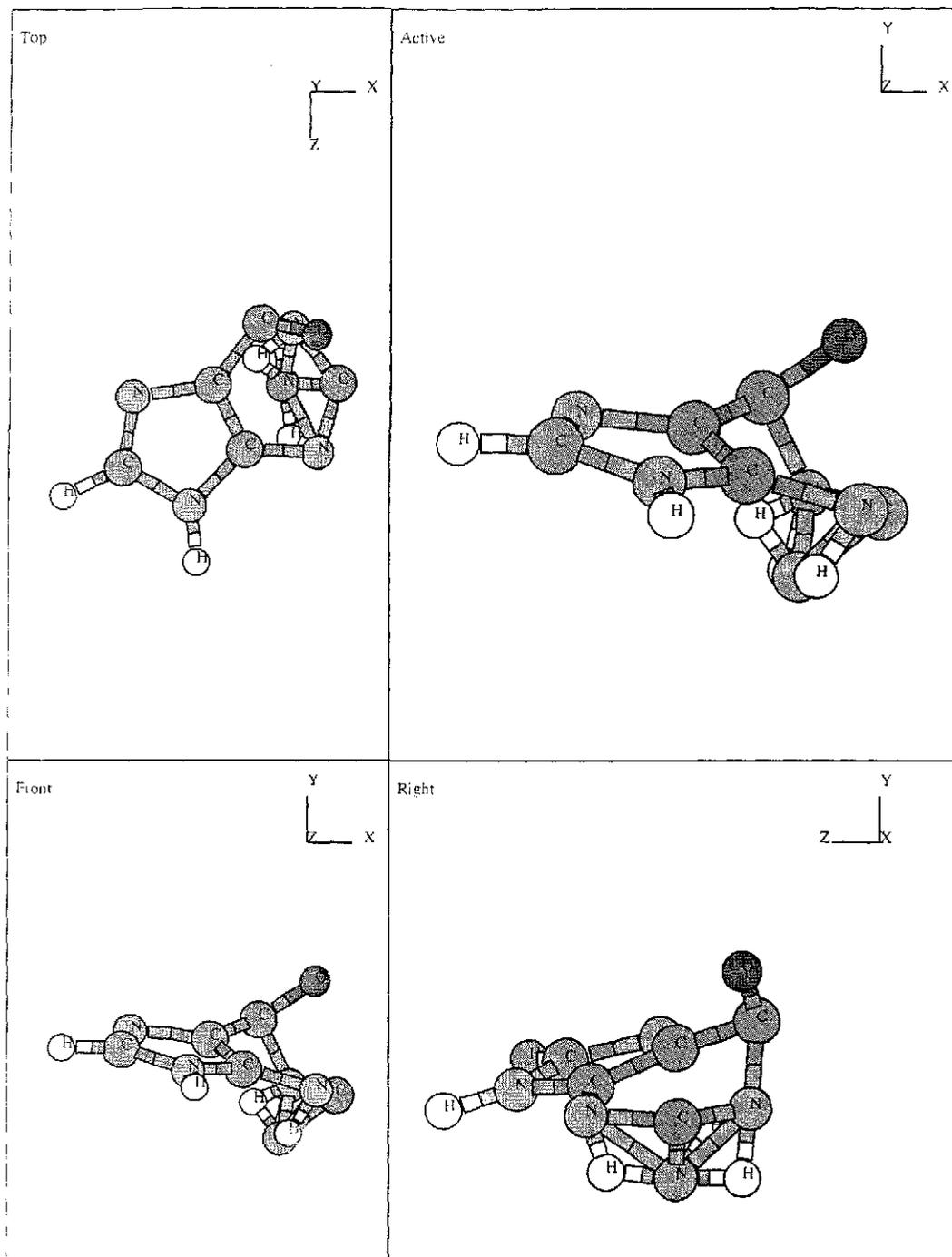


Figura 3 22: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G4, energía potencial= -380.099 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 1.

TESIS CON
FALLA DE ORIGEN

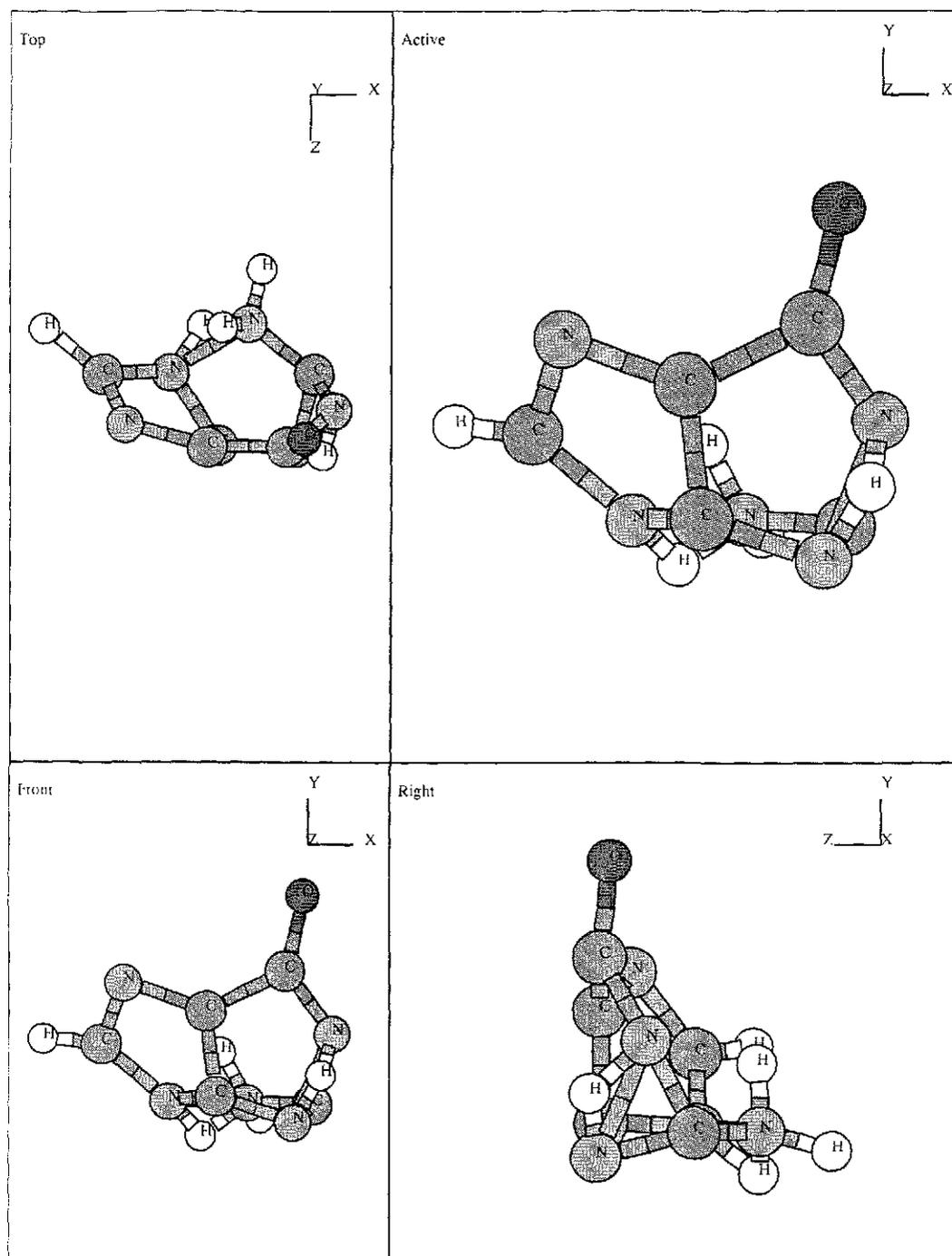


Figura 3.23: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G^* , energía potencial = -368.564 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 11.

TESIS CON
FALLA DE ORIGEN

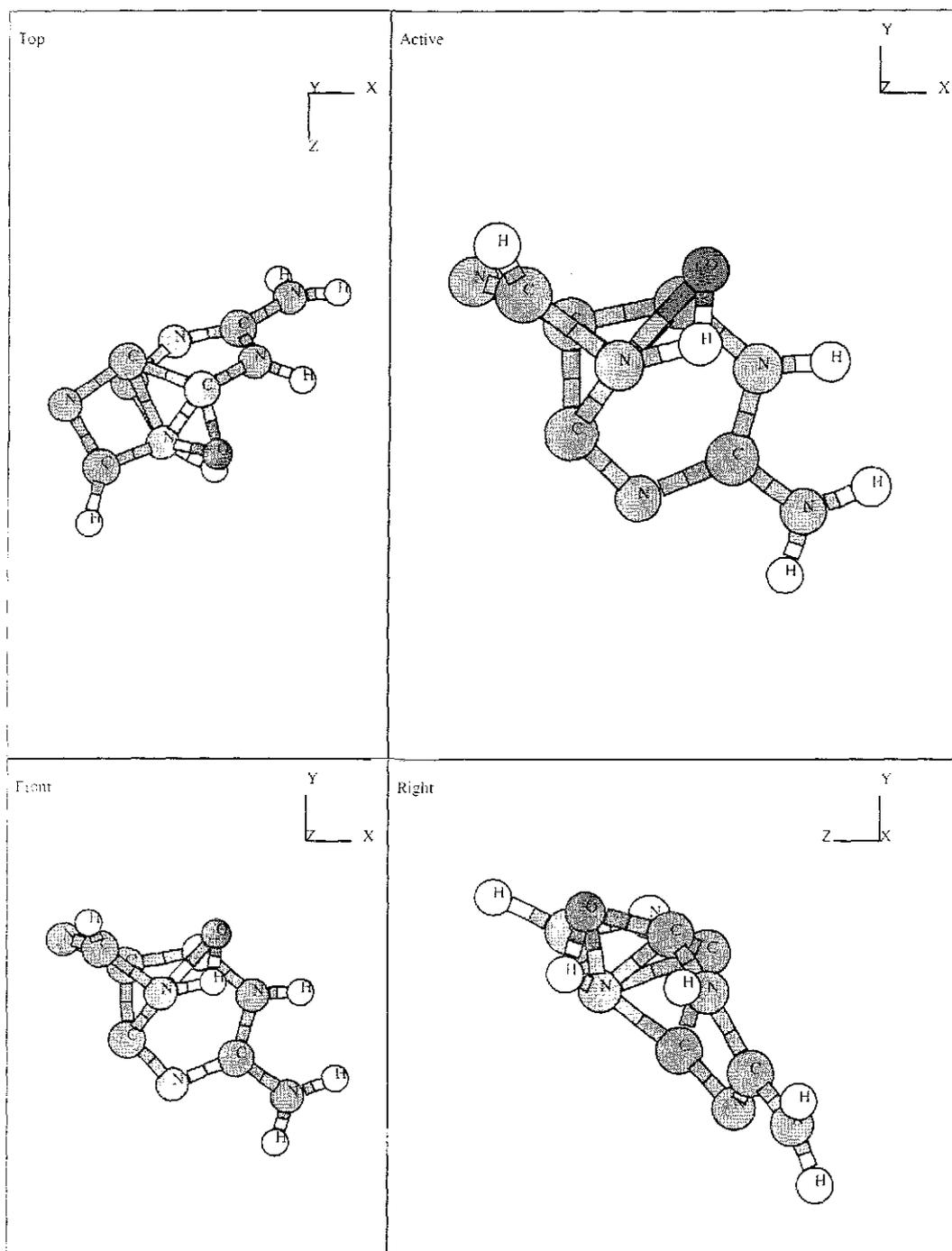


Figura 3.24: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina G5, energía potencial= -367.476 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina: 57.

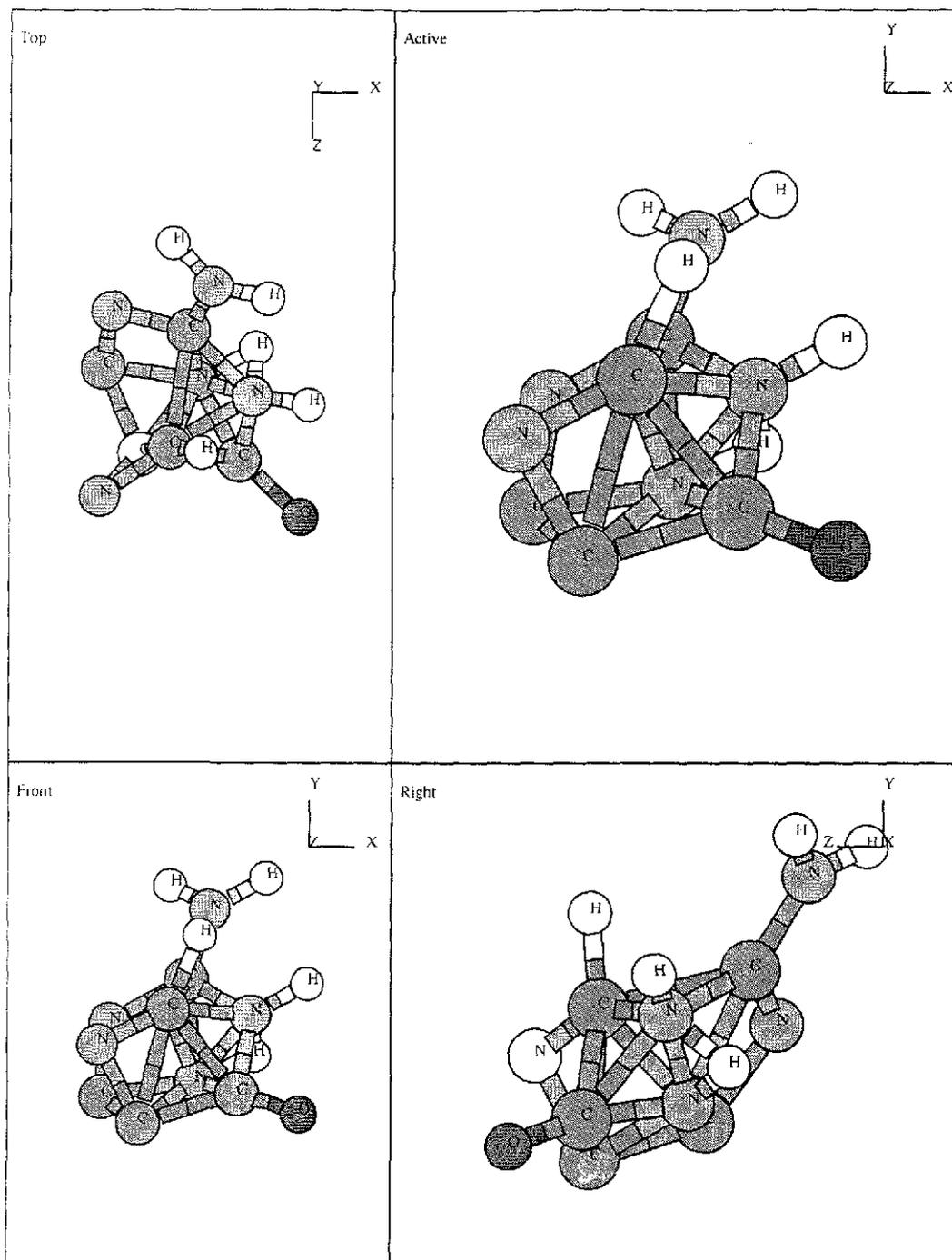


Figura 3.25: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base guanina. G6, energía potencial = -327.967 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la guanina:

53

TESIS CON
 FALLA DE ORIGEN

ESTA TESIS NO SALDRA
 DE LA BIBLIOTECA

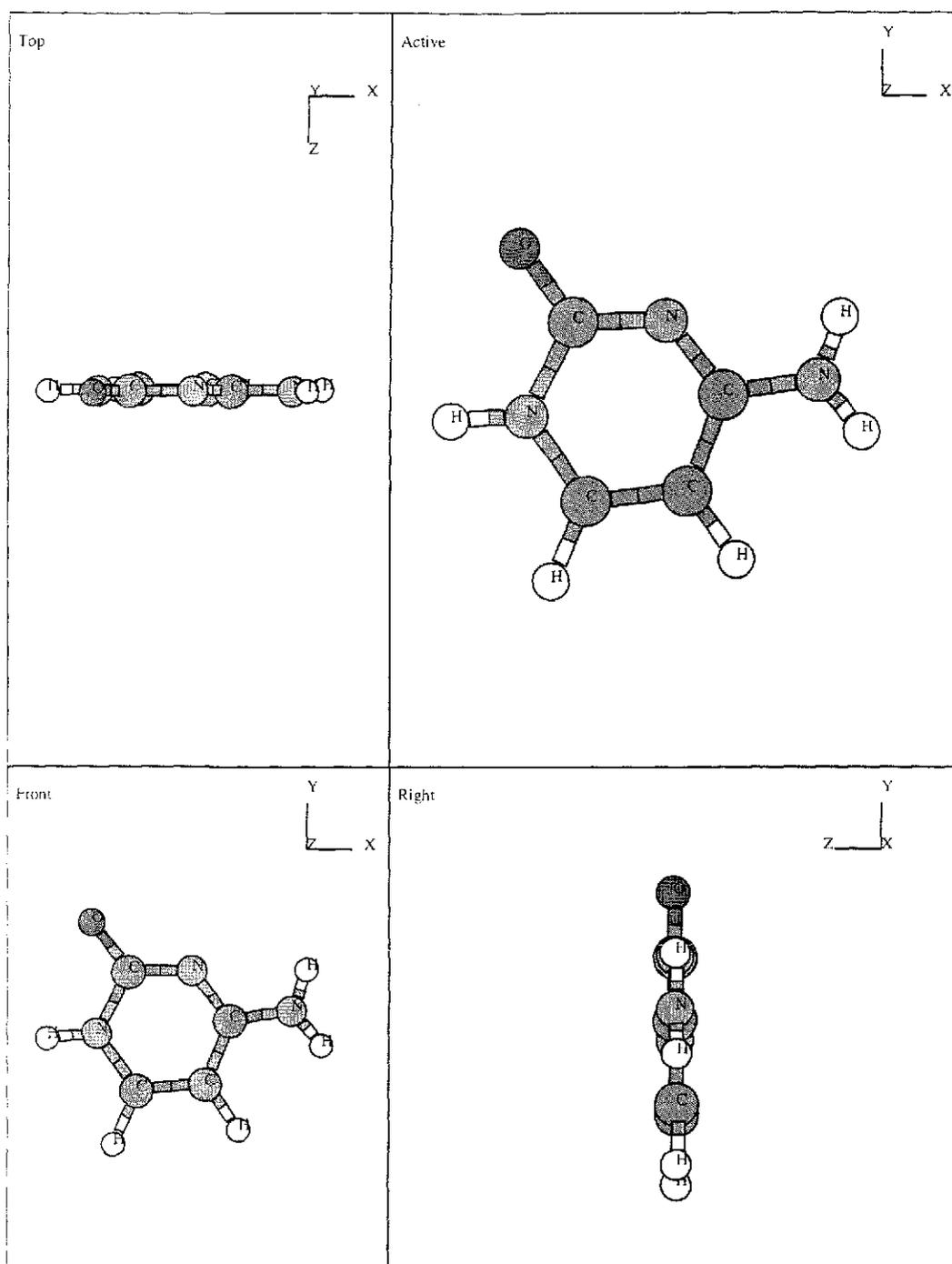


Figura 3.27: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base citosina. C1, energía potencial= -685.298 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la citosina: 2026.

TESIS CON
FALLA DE ORIGEN

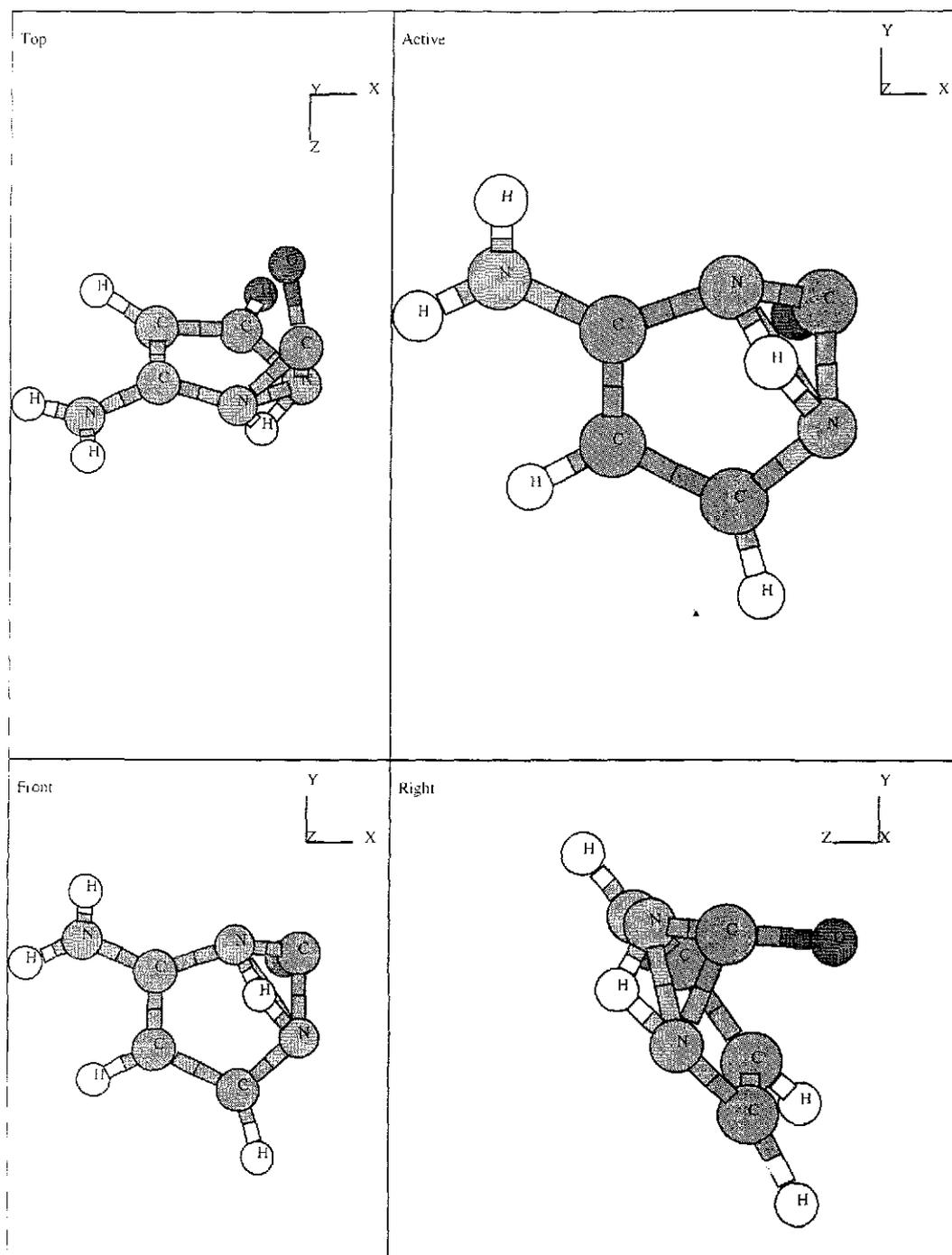


Figura 3.28: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base citosina C2, energía potencial = -585.517 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la citosina: 517

TESIS CON
FALLA DE ORIGEN

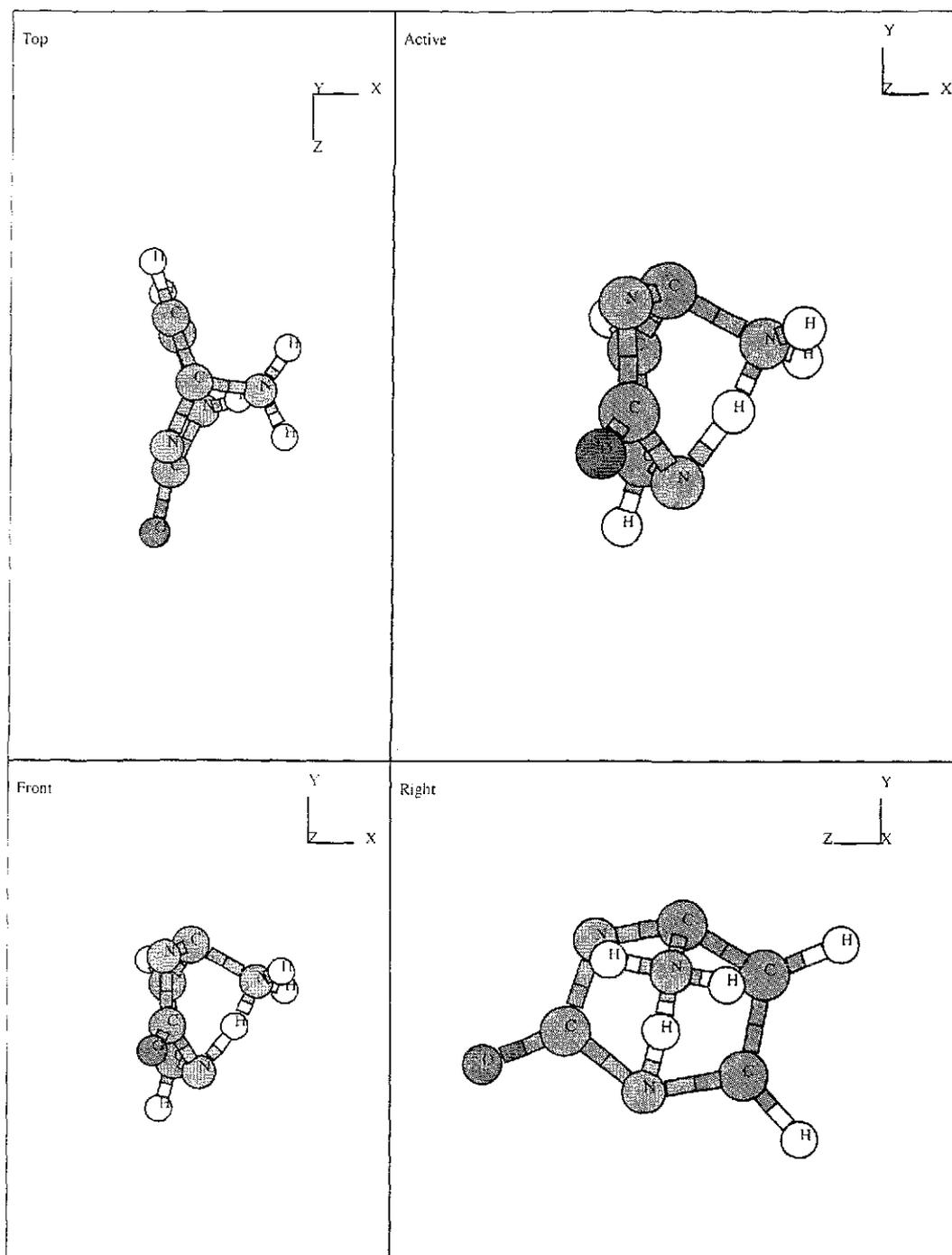


Figura 3.29: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base citosina. C3, energía potencial= -561.758 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la citosina: 354.

TESIS CON
FALLA DE ORIGEN

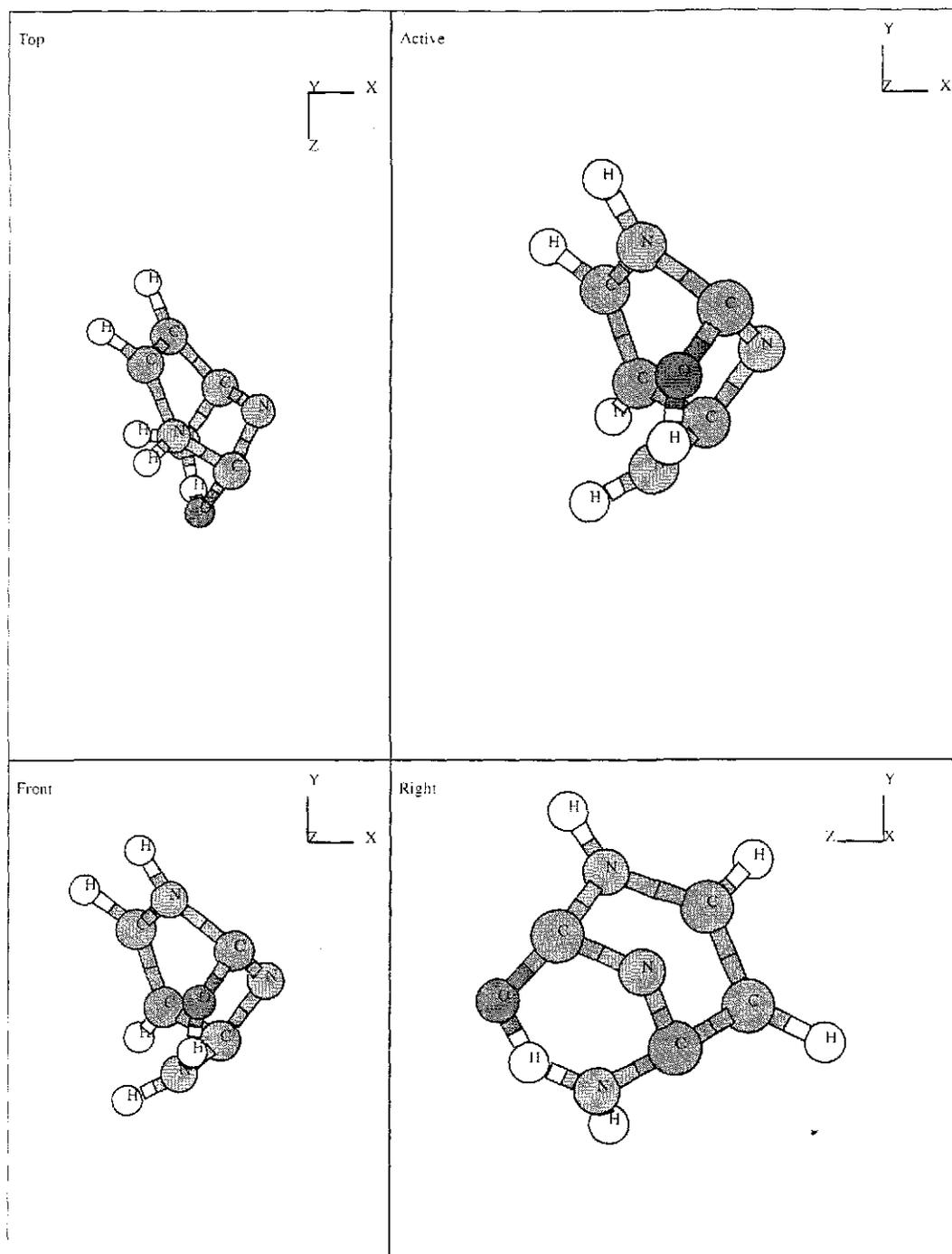


Figura 3 30: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base citosina. C4, energía potencial= -556 664 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la citosina: 148.

TESIS CON
FALLA DE ORIGEN

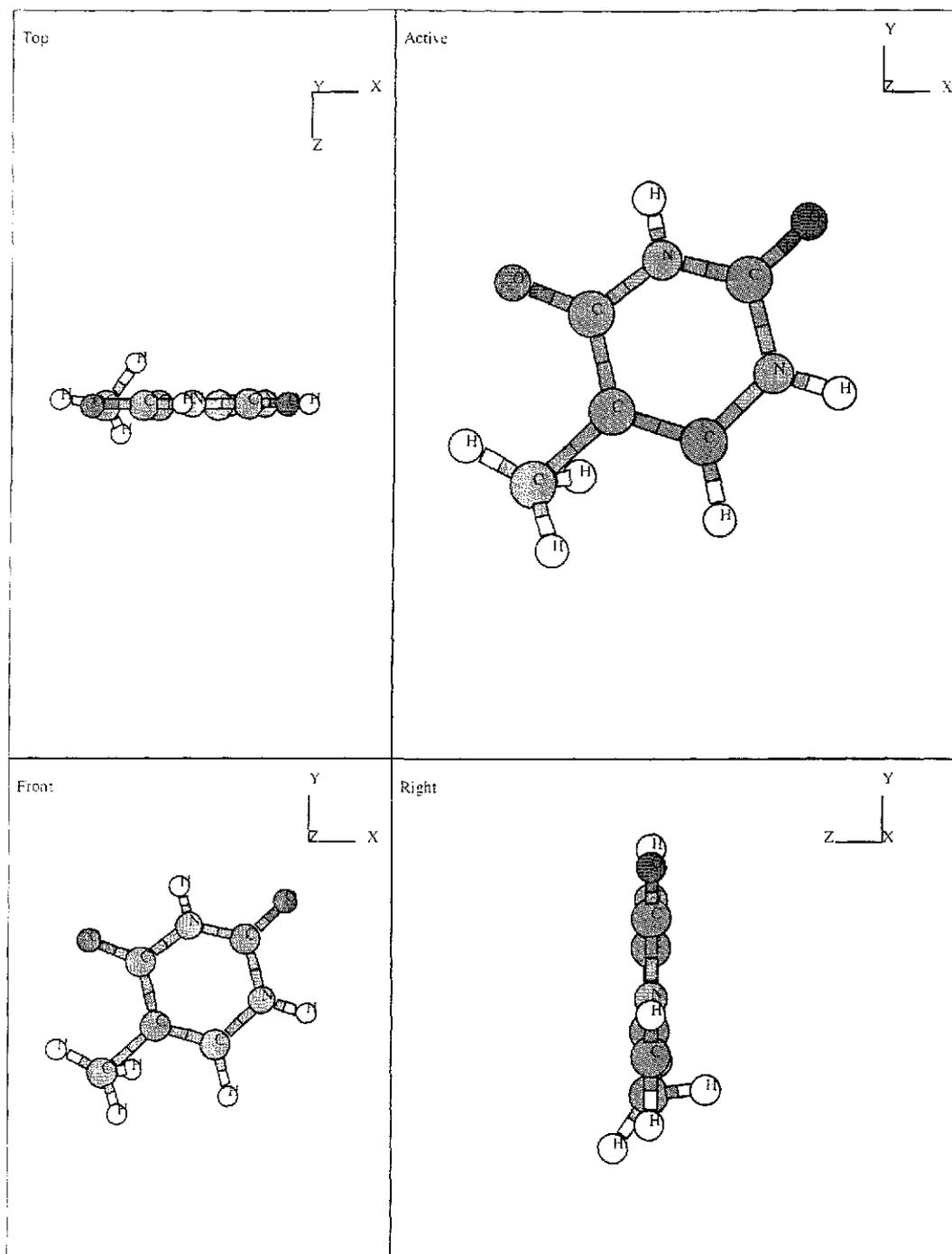


Figura 3.31: Mínimo del potencial encontrado con la optimización del algoritmo genético para la base timina. T1, energía potencial= -476.136 kcal/mol, número de veces que se encontró en 13000 corridas del algoritmo genético para la timina: 898.

TESIS CON
FALLA DE ORIGEN

Capítulo 4

OPTIMIZACIÓN CON CÁLCULOS CUÁNTICOS

En la actualidad, a pesar del avance en las técnicas de optimización y de búsqueda, el tratar de encontrar las estructuras de un conjunto de átomos de baja energía estables es todavía un problema difícil. Aún con los sorprendentes márgenes de eficiencia que ha presentado la maduración de técnicas de dinámica molecular con cálculos *ab-initio* (de primeros principios), las optimizaciones globales llevadas a cabo con este método son muy costosas computacionalmente debido a la gran cantidad de grados de libertad. Las aproximaciones que emplean potenciales clásicos con parámetros ajustados a datos experimentales o resultados *ab-initio* seleccionados son una alternativa efectiva. Tales potenciales a menudo son la puerta de entrada a búsquedas más refinadas [135], ya que los resultados obtenidos con la optimización clásica pueden ser usados como una entrada configuracional hacia un completo cálculo *ab-initio* o funcional de la densidad.

4.1 Justificación

Las estructuras encontradas con el algoritmo genético son mínimos locales de una función matemática, si esta función fuera la que en realidad describe la superficie de energía potencial debido a las interacciones atómicas entonces podríamos asegurar que estas conformaciones en realidad existen en la naturaleza, desafortunadamente esto no es verdad ya que el potencial se ajustó

para las configuraciones naturales. Es justificable por lo tanto recurrir a una optimización local utilizando cálculos de primeros principios para comprobar si estas nuevas formas son factibles, hay que destacar aquí que estos métodos dan una solución aproximada a la ecuación de Schroedinger, cuya solución correspondería al valor experimental.

4.2 Campos Autoconsistentes

Mientras que es verdad que una solución exacta de la ecuación de Schrodinger es todavía un problema matemático imposible, en la actualidad soluciones aproximadas aceptables pueden ser rutinariamente obtenidas para moléculas que contienen muchos átomos.

La esencia del modelo de Hartree-Fock (HF) o de Campo Autoconsistente (Self-Consistent Field, SCF) es resolver la ecuación de Hartree-Fock que resulta de minimizar la energía electrónica [152]. El operador de esta ecuación depende de las coordenadas de un simple electrón moviéndose en un potencial donde los movimientos complicados de los restantes electrones son "promediados". Estas interacciones no son ciertamente tomadas como cero, pero el modelo obviamente pierde los detalles mas finos causados por las interacciones instantáneas electrón-electrón.

La idea es que la expresión de la energía, la cual involucra las coordenadas de todos los electrones, puede ser reducida a una suma mucho mas simple de términos involucrando las coordenadas de uno o dos electrones (reglas de Slater-Condon-Shortley [153]). Lo que normalmente se hace para moléculas es usar el modelo LCAO (Linear Combination of Atomic Orbitals) donde cada orbital molecular ϕ_i puede ser descrito como una combinación lineal de n orbitales atómicos $\chi_1, \chi_2, \dots, \chi_n$. Esta formulación de las ecuaciones de HF, donde se usa la aproximación LCAO, se acredita a Roothaan [151].

Los orbitales LCAO son usualmente ortogonales. Usando el principio variacional se buscan los mejores valores posibles de los coeficientes y energías orbitales. Recordemos que para una molécula estamos trabajando dentro de la aproximación de Born-Oppenheimer donde los movimientos de

TESIS CON
FALLA DE ORIGEN

los electrones y los núcleos se consideran independientes unos de otros.

Los mejores orbitales moleculares son obtenidos al imponer que la energía electrónica sea un mínimo, sujetos a la restricción de que sean ortonormales. Esto conduce a las bien conocidas ecuaciones de Hartree-Fock de un electrón

$$F(1)\phi_i(1) = \epsilon_i\phi_i(1) \quad 4.1$$

donde ϵ_i es la energía del orbital ϕ_i y F , el operador de Fock dado por:

$$F(1) = h(1) + \sum_{j=1}^n [2J_j(1) - K_j(1)] \quad 4.2$$

1 representa las coordenadas del electrón uno, los operadores de un electrón están definidos por:

$$h(1) = -\frac{\hbar^2}{8\pi^2m}\nabla_1^2 - \sum_{\nu=1}^N \frac{Z_\nu e^2}{r_{1\nu}} \quad 4.3$$

$$J_j(1)\phi_i(1) = \left[\int \phi_j(2) \frac{e^2}{r_{12}} \phi_j(2) dV_2 \right] \phi_i(1) \quad 4.4$$

$$K_j(1)\phi_i(1) = \left[\int \phi_j(2) \frac{e^2}{r_{12}} \phi_i(2) dV_2 \right] \phi_j(1) \quad 4.5$$

$h(1)$ esta compuesto de los operadores de energía nuclear y cinética para el electrón 1 mientras que $J_j(1)$ y $K_j(1)$ son los operadores de Coulomb y de intercambio, respectivamente. Estos dos operadores últimos son responsables de la repulsion inter-electrónica, pero se observa de las ecuaciones anteriores que el electrón 1 interactúa con el electrón 2 localizado en una posición promedio en su orbital molecular ϕ , no en una instantánea como debería ser. Gaussian98 resuelve estas ecuaciones utilizando el procedimiento iterativo descrito en el esquema de la figura 4.1 y conocido como técnica de campo auto-consistente. En el método SCF, uno comienza a partir de una solución aproximada inicial propuesta para los orbitales $\phi_i^{(0)}$, $i = 1, 2, \dots, n$; los cuales son usados para calcular las integrales involucradas en los operadores de Coulomb y de intercambio de $F^{(1)}$, el operador de Fock de la iteración 1. Después, las ecuaciones de Hartree-Fock son resueltas, conduciendo a un nuevo conjunto de orbitales $\phi_i^{(1)}$, los cuales son usados subsecuentemente

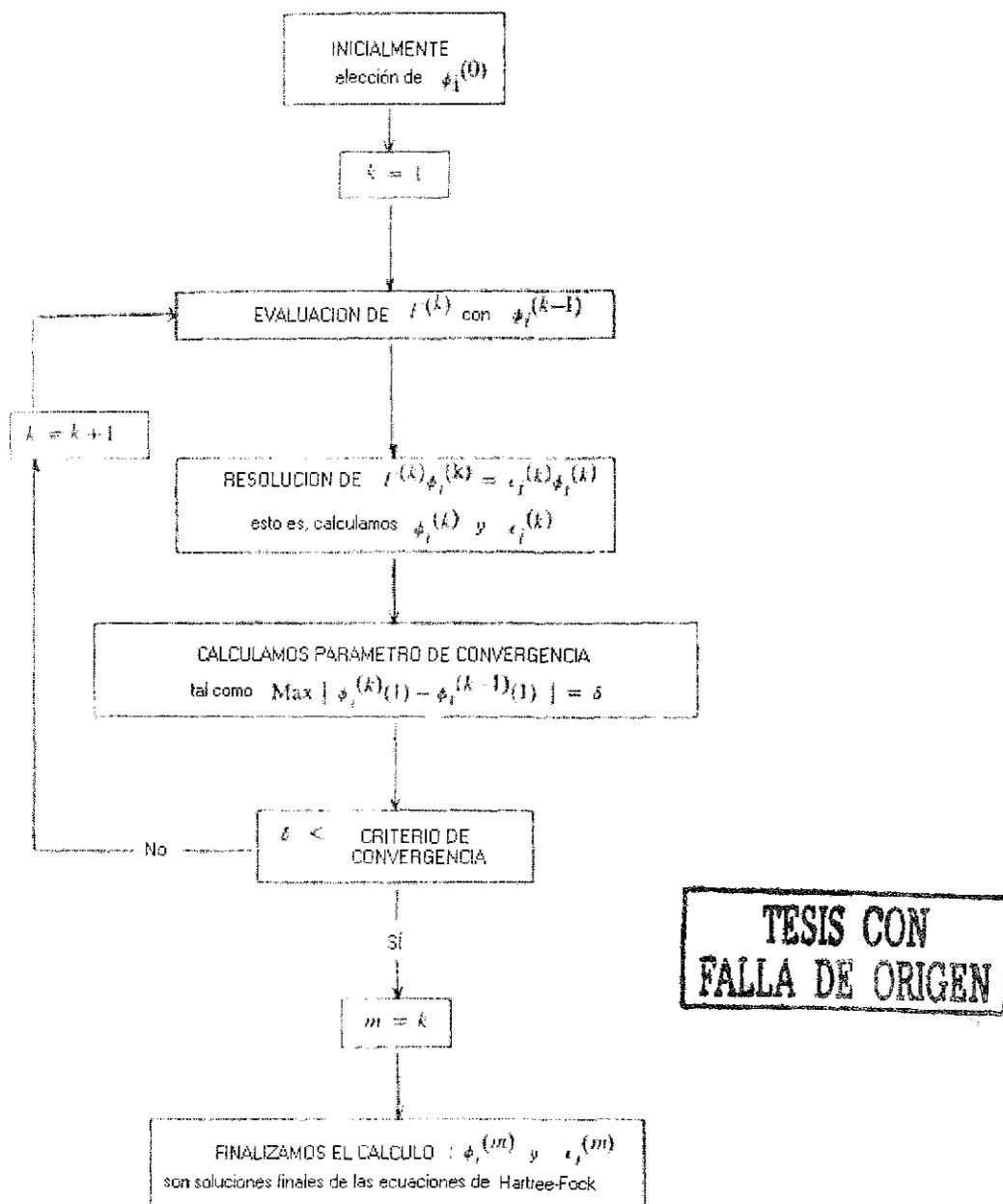


Figura 4.1: Técnica de campo auto-consistente.

para construir el nuevo operador de Fock $F^{(2)}$. Este proceso se repite hasta que el valor de un parámetro de prueba de convergencia δ , el cual puede ser definido como la diferencia máxima entre los valores de la función de onda obtenidos en iteraciones consecutivas, sea mas pequeño que un criterio dado.

4.3 Relajación con cálculos de primeros principios

Dos distintos cálculos cuánticos: Hartree-Fock más MP2, fueron empleados en una optimización local de las configuraciones interesantes encontradas con el potencial. El método Hartree-Fock utiliza orbitales moleculares de campo auto-consistentes (SCF-MO)

En el conjunto de cálculos cuánticos, optimizaciones de orbital molecular *ab-initio* fueron llevadas a cabo con el campo auto-consistente Hartree-Fock (esquema computacional de combinación lineal de orbitales atómicos HF-LCAO) implementado con el programa Gaussian 98 [136]. La diagonalización del Hamiltoniano fué desarrollada usando como conjunto base la denominada 6-31G*, En una molécula, la densidad de carga de un átomo se distorsiona de la simetría esférica. Para tomar en cuenta los puntos más finos de esta distorsión, incluimos funciones de polarización al conjunto base (figura 4.2).

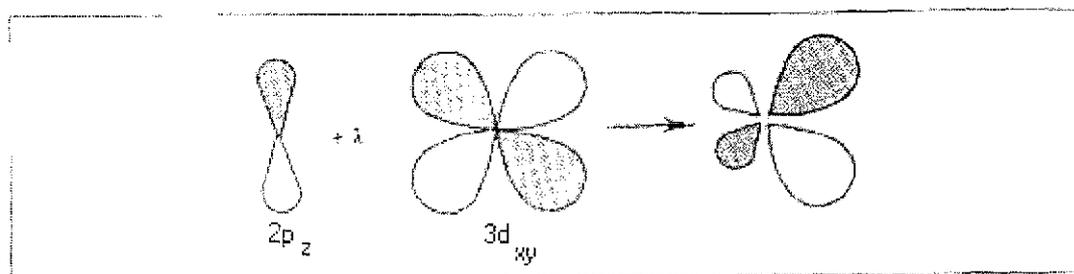


Figura 4.2: La adición de un orbital $3d_{xy}$ a uno $2p_z$ da un orbital distorsionado

Tal ingreso permite a un átomo libre tomar en cuenta la polarización inducida por un campo eléctrico externo o por la formación de la molécula. El asterisco denota la adición al conjunto base 6-31G de 6 primitivas Gaussianas del tipo-d cartesianas para actuar como polarización para las funciones de valencia s y p .

En la base 6-31G se representan los orbitales de Slater de cada capa inter-

na por una función base tomada como la suma de seis Gaussianas (figura 4.3). Cada orbital de valencia de Slater es dividido en partes internas y externas descritas por tres y una funciones Gaussianas, respectivamente.

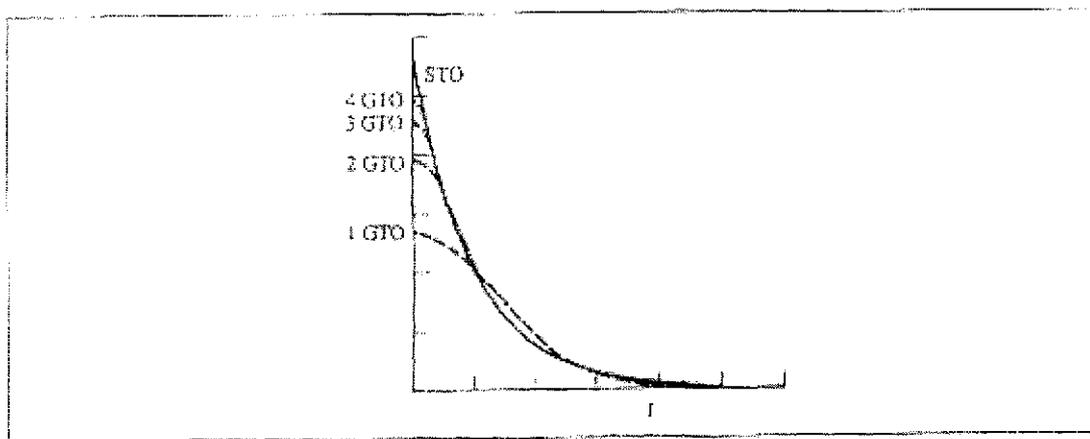


Figura 4.3: Para átomos con muchos electrones no existe una solución exacta a la ecuación de Schroedinger, la solución propuesta por Slater varía los exponentes orbitales de las funciones de onda radiales, en esta figura se compara el orbital 1s de Slater y las expansiones Gaussianas con hasta cuatro términos.

El conjunto base 6-31G* de C, N y O consiste de 15 funciones:

$$1s, 2s, 2s', 2p_x, 2p'_x, 2p_y, 2p'_y, 2p_z, 2p'_z, 3d_{xx}, 3d_{yy}, 3d_{zz}, 3d_{xy}, 3d_{xz}, 3d_{yz}$$

para los átomos H son dos orbitales atómicos: $1s, 1s'$. Por ejemplo, para la adenina, con fórmula $C_5N_5H_5$, en el conjunto base 6-31G* el número de funciones bases es 160, 15 para cada átomo de carbono y nitrógeno y 2 para cada átomo de hidrógeno.

Para la optimización de las coordenadas nucleares un algoritmo de Berny [137] fué incorporado sin imponer unas restricciones de simetría sobre la geometría, uno comienza con una posición dada para los núcleos y busca la mejor repartición de los electrones en un potencial generado por los núcleos y los mismos electrones, esto se hace de forma iterativa, buscando en la superficie de Born-Oppenheimer los mínimos. Un criterio de convergencia en

el cual a las fuerzas se les impuso que fueran menor que 10^{-4} au fué usado para detener la optimización de la geometría.

Para comprobar si efectivamente son estructuras estables, se calculó numéricamente y diagonalizó la matriz Hessiana de segundas derivadas. Solamente aquellas configuraciones que tuvieran todos los eigenvalores positivos (excepto los 6 eigenvalores iguales a cero correspondientes a los grados de libertad de traslación y rotación) fueron considerados interesantes. El programa Gaussian desarrolla el análisis vibracional dentro del formalismo matricial estándar de Wilson, usando la ecuación

$$GF = I\lambda \quad 4.6$$

Donde G es la matriz de energía cinética inversa, calculada a partir de los parámetros estructurales de equilibrio y las masas atómicas, F es la matriz de energía potencial (constantes de fuerza), I es la matriz identidad y λ es la matriz diagonal cuyos elementos son $\lambda_i = 4\pi^2c^2\nu_i^2$, donde c es la velocidad de la luz y ν_i es el i -ésimo número de onda del modo normal [155].

4.4 MP2

Asumiendo que todas las funciones de onda ϕ_k^0 y energías ϵ_k^0 pertenecientes al hamiltoniano Hartree-Fock son conocidas

$$F(1)\phi_k^0(1) = \epsilon_k^0\phi_k^0(1) \quad 4.7$$

y dado que uno desea encontrar los eigenestados del Hamiltoniano electrónico

$$H_{el}\Psi_{el}^{(k)} = \epsilon^{(k)}\Psi_{el}^{(k)} \quad 4.8$$

entonces

$$H_{el} = F(1) + V \quad 4.9$$

de la teoría de perturbaciones se expresa $\Psi_{el}^{(k)}$ y $\epsilon^{(k)}$ como

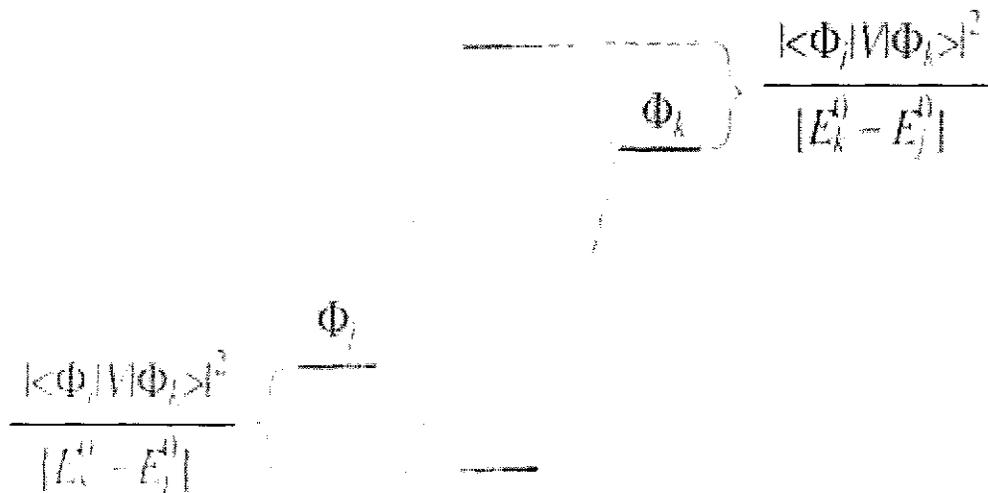
$$\Psi_{el}^{(k)} = \phi_k^0 + \sum_{n=1} \Psi_k^{(n)} \quad 4.10$$

TESIS CON
FALLA DE ORIGEN

$$\epsilon^{(k)} = \epsilon_k^0 + \sum_{n=1} \epsilon_k^{(n)} \quad 4.11$$

Las correcciones de la energía en primer y segundo orden son (figura 4.4):

$$\epsilon_k^{(1)} = \langle \phi_k^0 | V | \phi_k^0 \rangle \quad \epsilon_k^{(2)} = \sum_j \frac{|\langle \phi_j^0 | V | \phi_k^0 \rangle|^2}{\epsilon_k^0 - \epsilon_j^0} \quad 4.12$$



TESIS CON FALLA DE ORIGEN

Figura 4.4: Perturbación de segundo orden debido a la interacción entre los orbitales moleculares ϕ_k y ϕ_j .

4.5 Justificación de cálculos MP2

Anteriormente ya se había mencionado que los operadores de Coulomb (ec. 4.4) y de intercambio (ec. 4.5) de las ecuaciones de Hartree-Fock (ec. 4.1) involucran las interacciones promedio, no las instantáneas como debería de ser, entre pares de electrones. Esta deficiencia es debido a la aproximación del electrón independiente, y como consecuencia de esto por la función de

onda de un determinante simple de Slater, postulada como la solución para la configuración del estado base del sistema de n electrones. La energía de correlación E_c se define como la diferencia entre la verdadera energía no-relativista E_o del sistema y la energía E_{HF} de la función de onda de Hartree-Fock

$$E_c = E_o - E_{HF} \quad 4.13$$

Como el principio variacional establece que $E_o < E_{HF}$, de aquí se sigue que E_c es siempre negativa. Aunque en la práctica E_c vale típicamente solo 1% de E_{HF} , su valor absoluto es grande debido al orden de magnitud de E_{HF} , como ejemplo, la diferencia de energía de Hartree-Fock con respecto a MP2 del estado base de la adenina es de -895.4425 kcal/mol!

Para refinar la estructura utilizamos el tratamiento perturbativo post-Hartree-Fock de MP2, este método es conveniente ya que requiere menor componente computacional que otros. Por ejemplo, el tiempo de CPU necesario en un cálculo de MP2 es generalmente mayor que un SCF únicamente por un factor de 2 a 4. Además MP2 ha probado ser muy eficiente en la predicción precisa de geometrías de equilibrio y de frecuencias vibracionales [159].

Los cálculos MP trabajan bien cerca de la geometría de equilibrio, en [157] se muestra como en diversas conformaciones de equilibrio de varias moléculas un cálculo MP2 obtuvo un 94% de la energía de correlación del conjunto base, cálculos MP con un conjunto base pequeño son de poco valor práctico, un cálculo MP con el programa Gaussian98 debería de utilizar un conjunto base 6-31G* o mayor para resultados útiles.

No siempre es necesario ir más allá del tratamiento de Hartree-Fock (las distancias de enlace están dentro de un error de 0.02-0.03 Å y los ángulos de enlace cambian dentro de 5°. Sin embargo, diversos estudios han mostrado que estas variaciones afectan en el cálculo de las frecuencias normales [158]-[160].

4.6 Resultados

De las configuraciones encontradas por medio de los algoritmos genéticos nosotros escogimos aquellas que se veían interesantes para su posterior optimización con cálculos de primeros principios. Seleccionamos éstas en base a baja energía, que fueran encontradas varias veces, y que no fueran tan distintas de las configuraciones conocidas; además nos basamos en configuraciones que podrían alterar la unión conocida A-T y G-C, por ejemplo que un isómero de guanina pudiera hacer una unión con una molécula de timina al adquirir una forma tal que en las zonas de puentes de hidrógeno se pareciera a esta zona de la adenina.

Se encontró que para las moléculas de anillo simple citosina y timina, aunque en la búsqueda global en la superficie del potencial se encontraron varias configuraciones distintas de la natural, estas mismas configuraciones no sobrevivían a las optimizaciones cuánticas y convergían finalmente al estado natural.

Para las moléculas de doble anillo adenina y guanina se encontraron varias configuraciones distintas de la natural. En la tabla 4.1 en la primera columna se muestran los valores de las energías que se encontraron utilizando el potencial semiempírico. En la segunda columna se muestran los valores de energía después de optimizar con Hartree-Fock que se encontraron de la configuración encontrada con los algoritmos genéticos; esta es la energía “total”, incluyendo la parte electrónica y la repulsión nuclear.

Finalmente en la tercera columna se muestran los valores de energía MP2 que se encontraron tomando en cuenta la correlación electrónica, para esto se utilizó como geometría inicial la configuración ya optimizada con los cálculos de HF. En el cálculo MP2 la idea es utilizar teoría de perturbación y tomar el hamiltoniano no perturbado como el modelo de Hartree-Fock y la perturbación como la diferencia entre el hamiltoniano verdadero y el hamiltoniano de Hartree-Fock. Recordemos que el hamiltoniano de Hartree-Fock promedia sobre el movimiento electrónico; MP2 se refiere a un tratamiento a segundo orden en la teoría de perturbaciones.

4.7 Resultados para adenina

Para la base adenina los resultados que se obtuvieron fueron:

Tabla 4.1. Energías totales obtenidas con métodos Hartree-Fock y MP2 para las diferentes conformaciones de la **adenina** encontradas con el potencial y el algoritmo genético.

Núm. de isómero	Potencial kcal/mol	Hartree-Fock Hartrees	MP2 Hartrees
A1	-550.201	-464.520	-465.947
A2	-439.897	-464.297	-465.724
A3	-433.733	-464.290	-465.752
A4	-372.105	-464.313	-465.731

Las diferencias en energías con respecto al valor mas bajo es lo que realmente tiene sentido para el cálculo de la energía de enlace, debido a que en la energía de Hartree se toma en cuenta tanto la energía nuclear así como la energía electrónica, y como el término electrónico es casi constante para todas las configuraciones, entonces al tomar la resta se eliminará, obteniendo así la diferencia de energía en la superficie de Born-Oppenheimer (de "bonding" o de enlace). Hay que notar que 1 Hartree=627 5095 kcal/mol

Tabla 4.2. Comparación de las diferencias en la energía total entre la configuración natural (correspondiente a la energía más baja) de la **adenina** (isómero A1) y los isómeros de más alta energía listados en las Figs. A1-A4 para el potencial modelo, Hartree-Fock y cálculos MP2.

Núm de isómero	Ener. pot. kcal/mol	Ener. HF kcal/mol	Ener. MP2 kcal/mol
A1	0.0	0.0	0.0
A2	110.304	139.935	139.935
A3	116.468	144.327	122.364
A4	178.096	129.894	135.542

TESIS CON
FALLA DE ORIGEN

Estas diferencias de energías de excitación son relativamente grandes si las comparamos con la energía térmica media (del orden de kT), la cual a temperatura ambiente (298 K) es aproximadamente de $2.4 * 10^{-2}$ eV = 0.55324 kcal/mol: en consecuencia y aún en caso de existir estos isómeros, las moléculas de adenina estarían normalmente en su estado fundamental, el cual corresponde al natural; lo mismo ocurre para la guanina. Pero hay que notar aquí que estas consideraciones se aplican a sistemas en equilibrio, cosa que no ocurre en los sistemas biológicos, por ejemplo, la molécula del ADN no es una conformación de mínima energía, ver las razones de esto iría más allá de nuestra investigación que realizamos.

Los enlaces de estas moléculas son covalentes, los cuales debido a su estructura electrónica rígida, son extremadamente duros y difíciles de deformar. Debido a la rigidez de estos enlaces se requiere una cantidad de energía relativamente grande para producir vibraciones en toda la molécula. En nuestro trabajo anterior sobre la búsqueda de los isómeros de la adenina utilizando el mismo potencial también encontramos isómeros estables diferentes a los que aquí se presentan [2]. Esta búsqueda la hicimos de nuevo para la adenina desde las optimizaciones con el algoritmo genético ya que en [2] no incluimos el átomo de hidrógeno en el enlace, que se une normalmente a la molécula de azúcar. Un dato interesante es que la energía del estado base aumento de -550.871 kcal/mol a -550.201 kcal/mol al agregarle un átomo de hidrógeno: nosotros esperábamos que disminuyera ya que normalmente al formar un enlace químico se invierte energía. Las energías de la conformación natural en [2] y en el presente trabajo son las únicas que es factible comparar ya que aunque en esta nueva búsqueda obtuvimos algunas configuraciones similares a las encontradas en [2], en realidad estas nuevas conformaciones presentaron detalles distintos.

Las estructuras que se presentan enseguida para la adenina (A2-A4) son diferentes del estado base A1.

Isómeros de adenina utilizando aproximación Hartree-Fock.

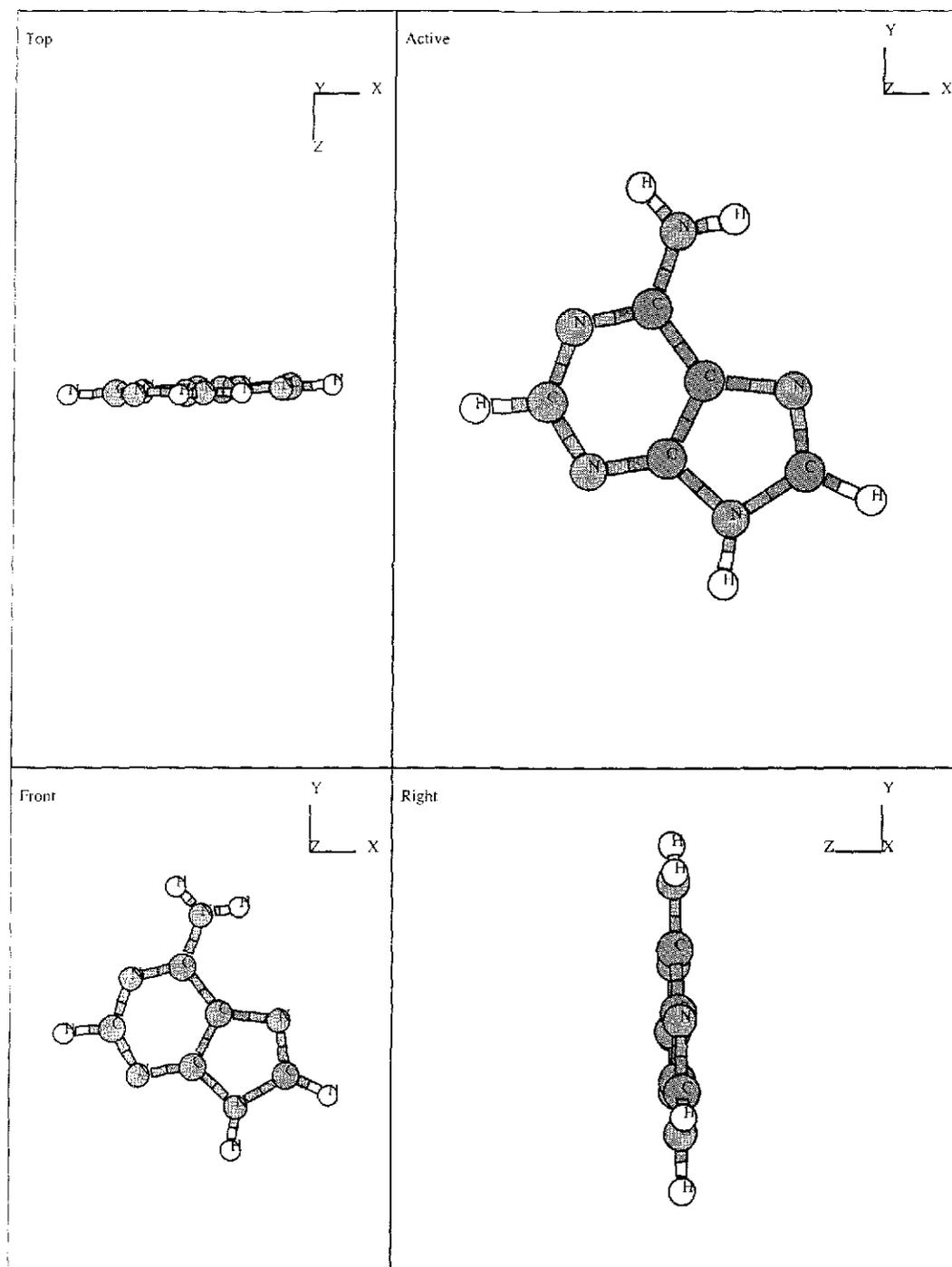


Figura 4.5: Geometría resultante de la figura 3.15 correspondiente a **A1**, después de optimización local usando cálculos de Hartree-Fock.

TESIS CON
FALLA DE ORIGEN

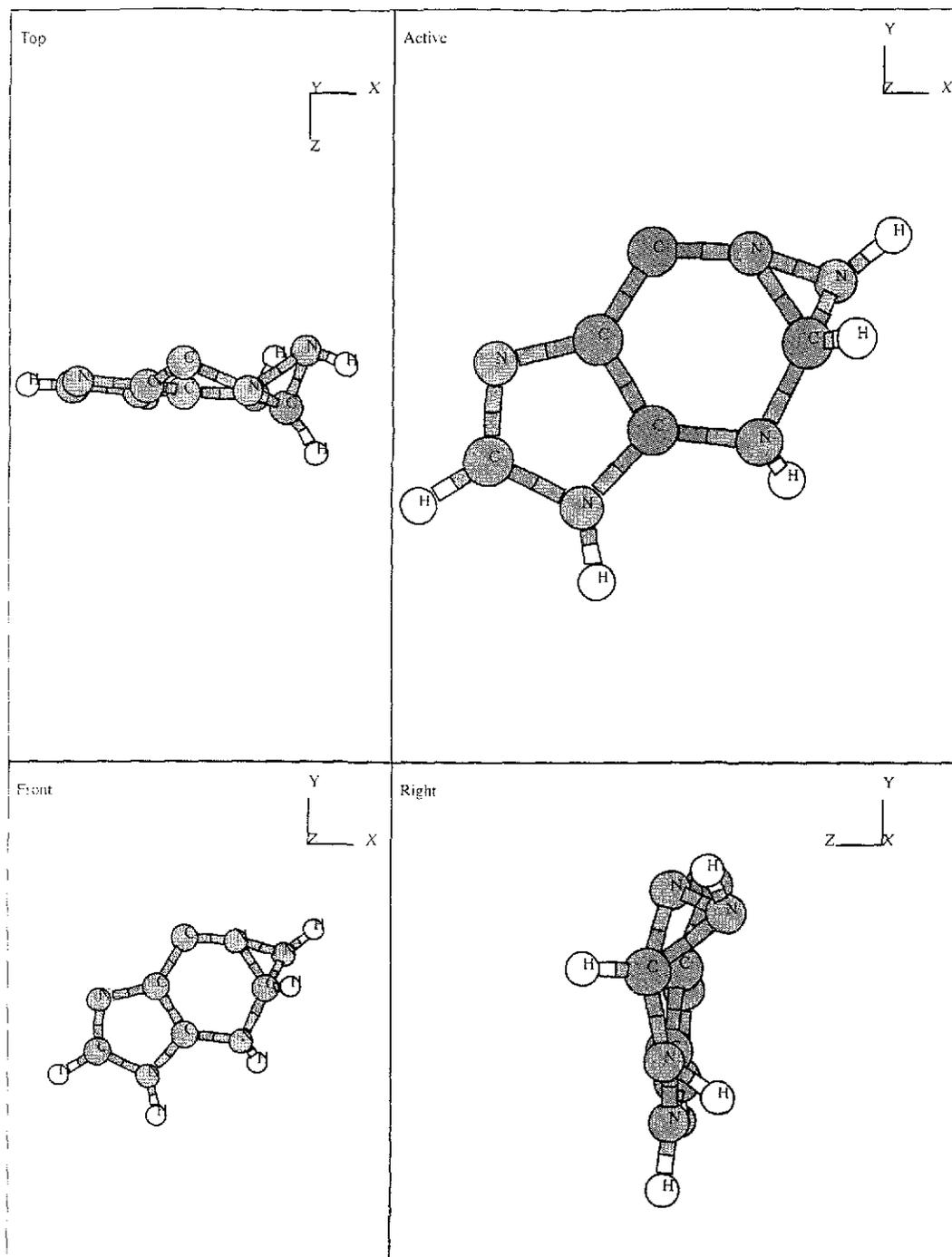


Figura 4.6: Geometría resultante de la figura 3.16 correspondiente a **A2**, después de optimización local usando cálculos de Hartree-Fock.

TESIS CON
FALLA DE ORIGEN

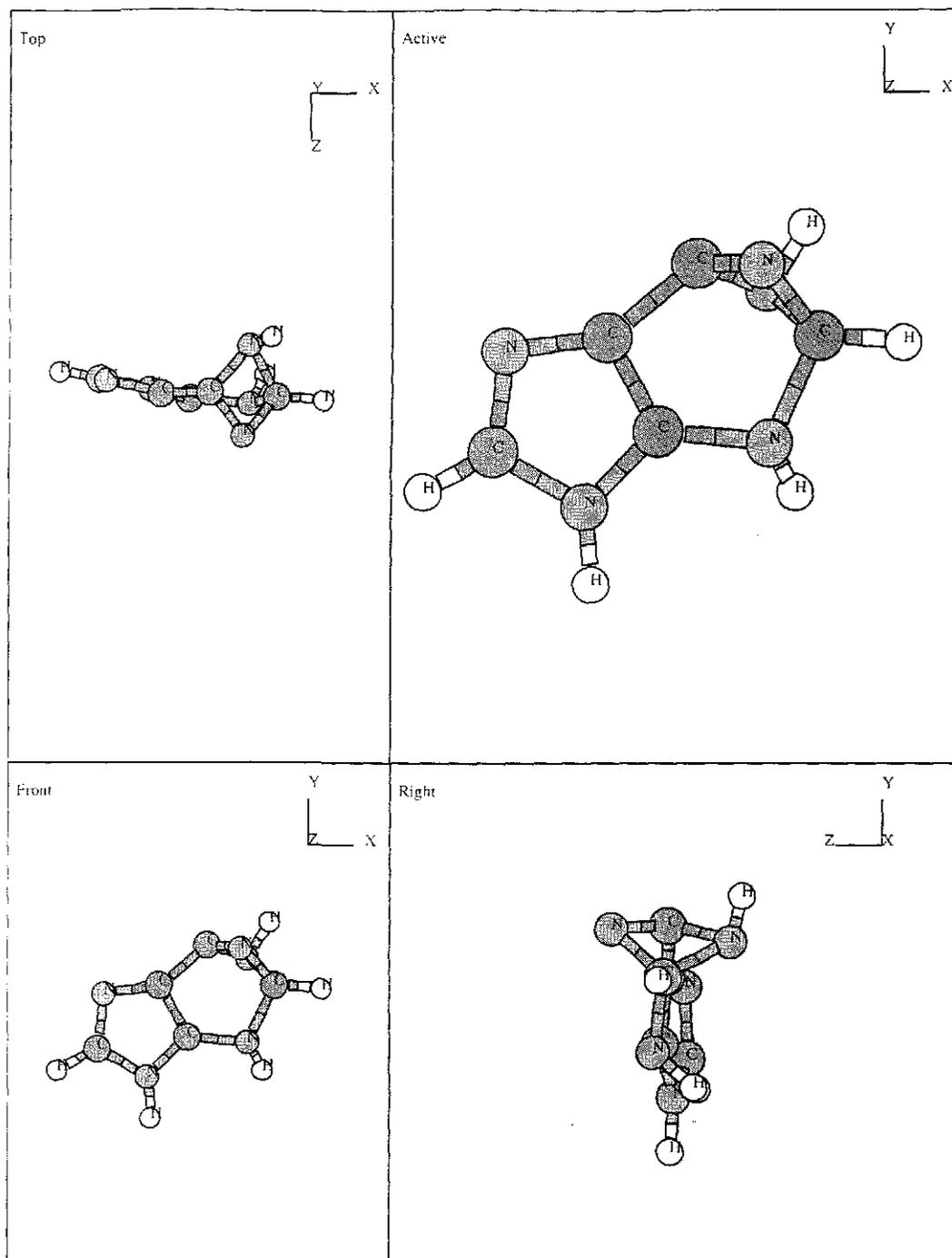


Figura 4.7: Geometría resultante de la figura 3.17 correspondiente a **A3**, después de optimización local usando cálculos de Hartree-Fock.

TESIS CON
FALLA DE ORIGEN

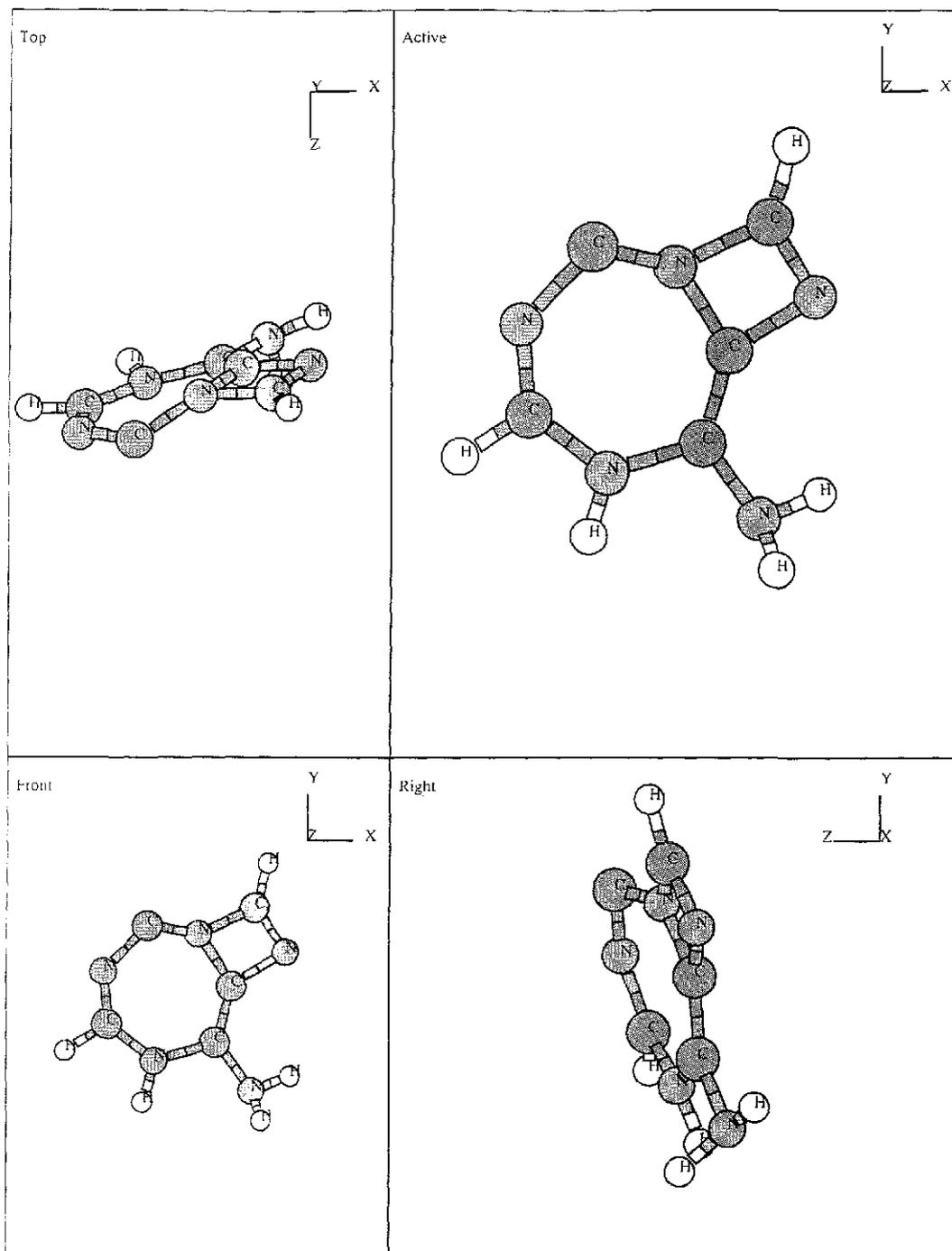


Figura 4.8: Geometría resultante de la figura 3.18 correspondiente a A4, después de optimización local usando cálculos de Hartree-Fock.

TESIS CON
FALLA DE ORIGEN

Isómeros de adenina utilizando aproximación MP2.

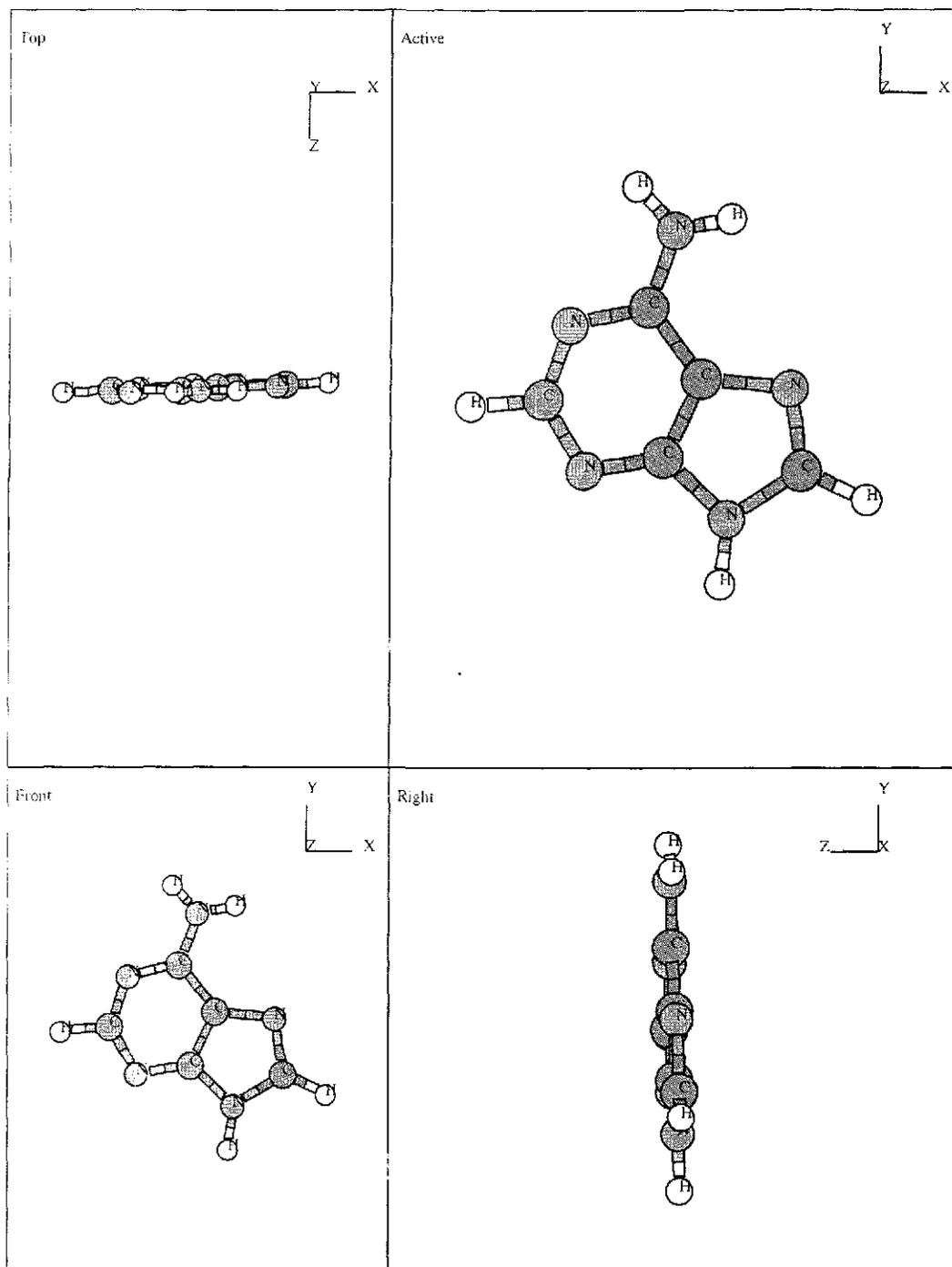


Figura 4 9: Geometría resultante de la figura 4.5 (A1), después de optimización local usando cálculos MP2.

TESIS CON
FALLA DE ORIGEN

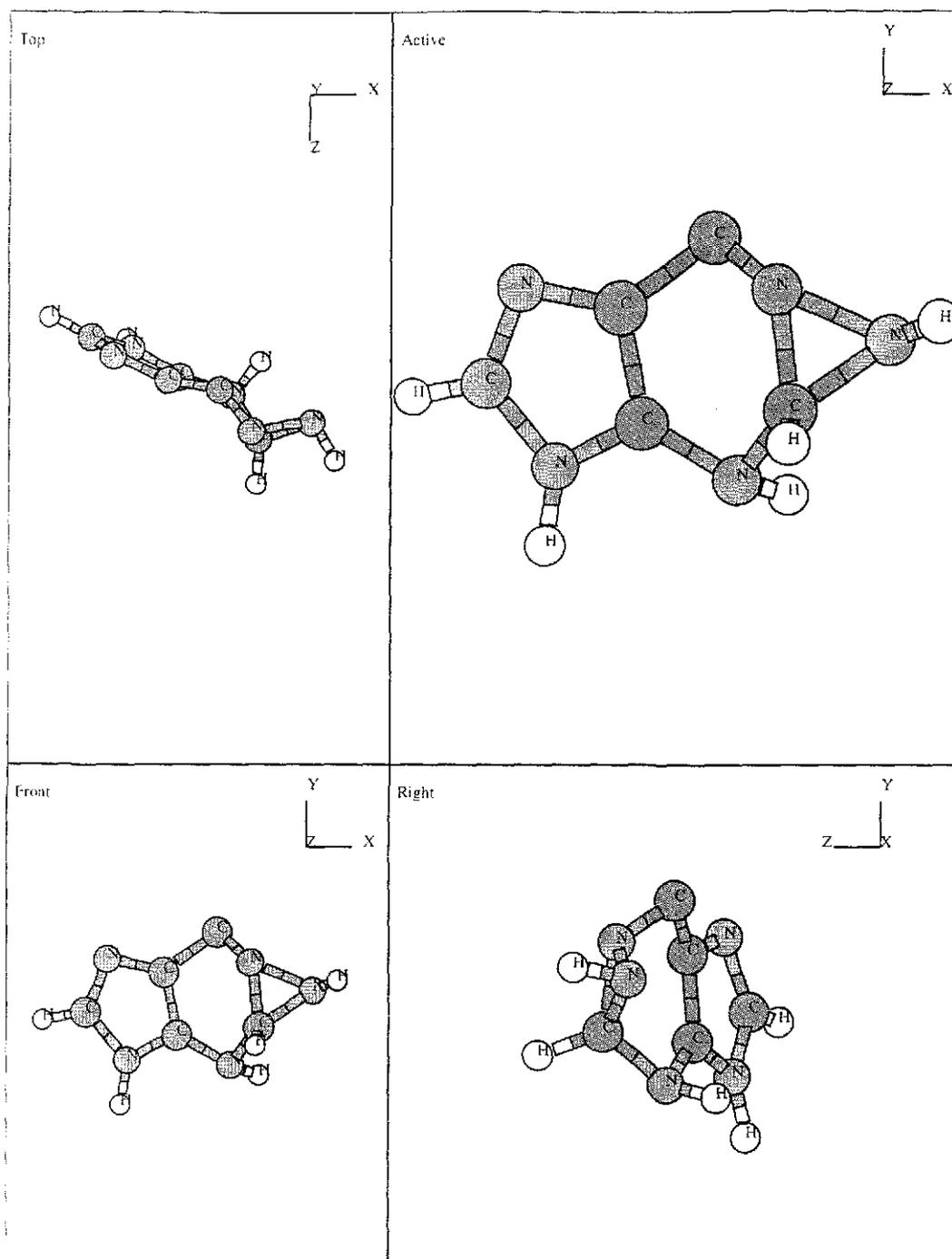


Figura 4.10: Geometría resultante de la figura 4.6 (A2), después de optimización local usando cálculos MP2

TESIS CON
FALLA DE ORIGEN

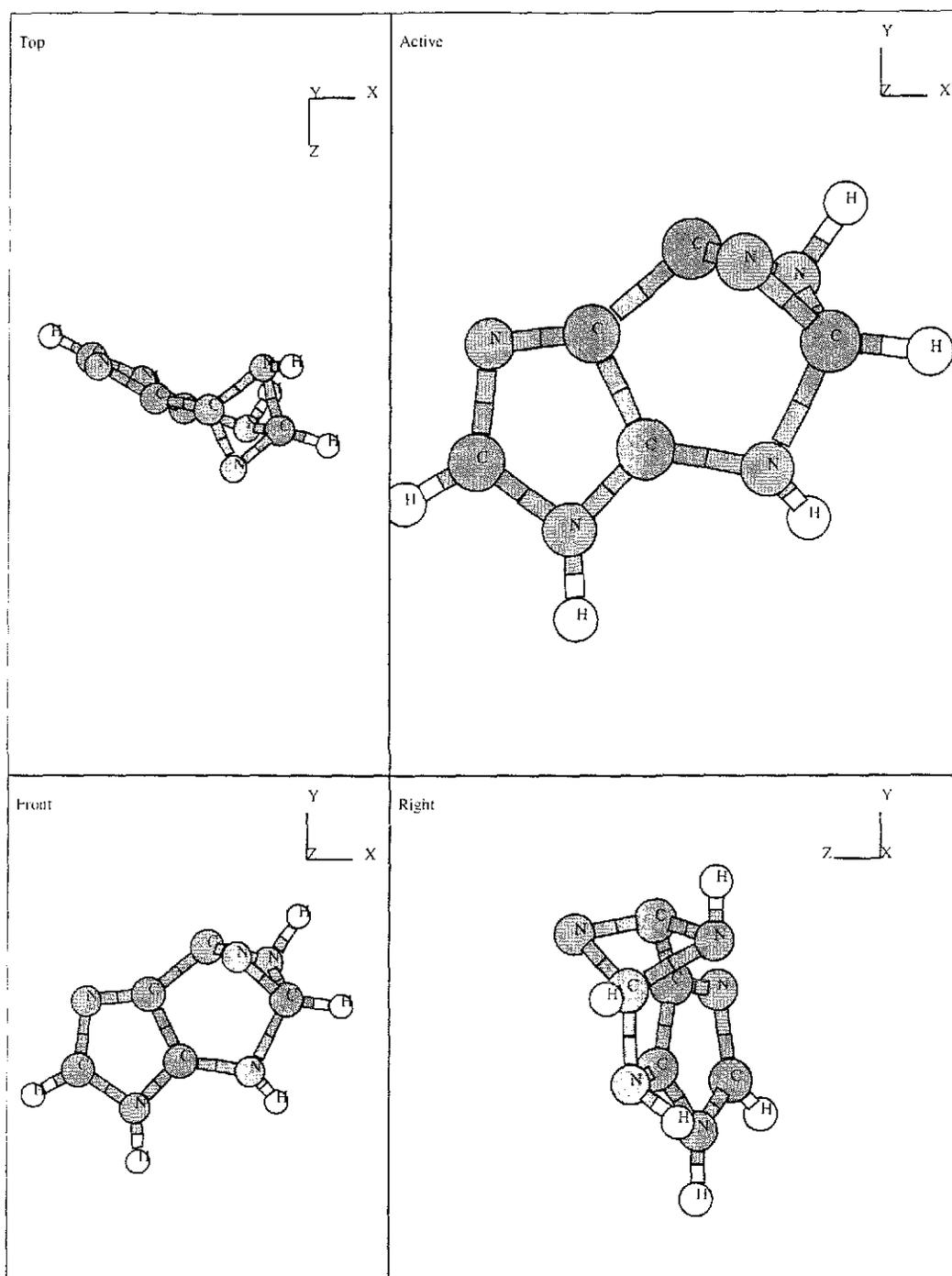


Figura 4.11: Geometría resultante de la figura 4.7 (A3), después de optimización local usando cálculos MP2.

TESIS CON
FOLIO DE ORIGEN

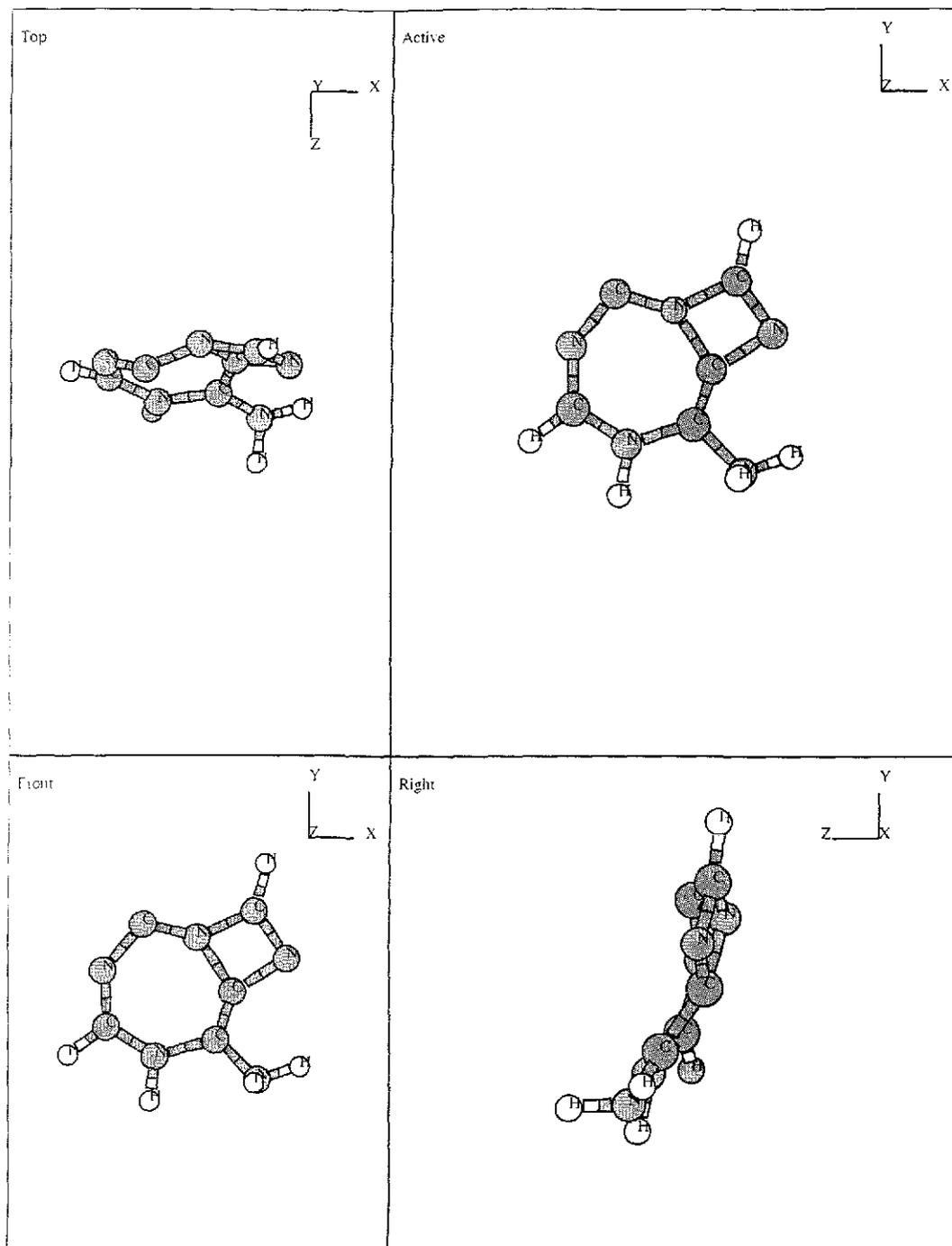


Figura 4.12: Geometría resultante de la figura 4.8 (A4), después de optimización local usando cálculos de MP2.

4.8 Resultados para guanina

Para la base guanina los resultados que se obtuvieron fueron los siguientes:

Tabla 4.3. Energías totales obtenidas con métodos Hartree-Fock y MP2 para las diferentes conformaciones de la **guanina** encontradas con el potencial y el algoritmo genético.

Núm. de isómero	Potencial kcal/mol	Hartree-Fock Hartrees	MP2 Hartrees
G1	-650.353	-539.394	-540.993
G2	-505.954	-539.394	-540.996
G3	-486.429	-539.342	-540.941
G4	-380.099	-539.362	-540.965
G*	-368.564	-539.124	-540.761
G5	-367.476	-539.139	-540.746
G6	-327.967	-539.201	-540.759
G7	-189.637	-539.147	-540.742
G8	-----	-539.218	-540.850
G9	-----	-539.232	-540.841

Tabla 4.4. Comparación de las diferencias en la energía total entre la configuración natural de la **guanina** (isómero G1) y los isómeros de más alta energía listados en las Figs. G1-G9 para el potencial modelo, Hartree-Fock y cálculos MP2

Núm. de isómero	Ener. pot. kcal/mol	Ener. HF kcal/mol	Ener. MP2 kcal/mol
G1	0.0	0.0	0.0
G2	144.399	0.0	-1.882*
G3	163.924	32.630	32.630
G4	270.254	20.080	17.570
G*	281.789	169.427	145.582
G5	282.877	160.015	154.995
G6	322.386	121.109	146.837
G7	460.716	154.995	157.505
G8	-----	110.442	89.988
G9	-----	101.657	95.381

**A diferencia del método HF, los métodos MP tienen la ventaja de ser consistentes (tienen una conducta adecuada a grandes separaciones, donde la energía es simplemente la suma de las energías atómicas). Sin embargo, los métodos MP presentan las limitaciones de no ser variacionales y no es inusual que la corrección de energía es mayor en valor absoluto que la energía de correlación y así puede algunas veces dar energías menores a la energía verdadera.*

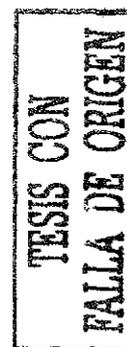
Los isómeros G1-G4 son en realidad tautómeros en donde uno o varios átomos de hidrógeno cambian de posición, también se puede observar que las estructuras G5-G7 son estructuras deformadas en las que ya no aparecen el doble anillo característico.

En el caso de los isómeros G8 y G9, no tenemos su valor de energía potencial ya que estos isómeros fueron obtenidos al variar nosotros la estructura de un isómero de adenina. Las conformaciones G8 y G9 eran originalmente mínimos en la superficie de potencial de la adenina que no habían convergido con cálculos cuánticos Hartree-Fock, pero nos parecieron interesantes ya que son similares espacialmente a dos conformaciones las cuales habían convergido en el trabajo de licenciatura con 4 optimizaciones diferentes (conform. 2 y 3 en figura 1 de [2]). Cambiando la estructura química de la adenina (le agregamos un átomo de oxígeno: esto es, pasamos de $C_5N_5H_5$ a $C_5N_5OH_5$) con ayuda del programa Xmol [138], obtuvimos dos moléculas de guanina, las cuales al optimizar cuánticamente si obtuvimos isómeros estables.

Hicimos lo anterior ya que estos dos isómeros de la guanina tienen las zonas de unión de los puentes de hidrógeno muy similares a la adenina, lo cual podría tener implicaciones muy importantes.

La superficie del potencial es diferente a la superficie de Born-Oppenheimer, El isómero G* es un mínimo local del potencial de interacción, al aplicarle optimización cuántica obtuvimos que era un punto de transición; esto es, la primera derivada de la energía total es igual a cero pero presentó frecuencias normales negativas, por lo tanto no es un isómero estable.

Un resultado muy interesante es que se encontraron dos configuraciones



que dieron energía más baja que la natural con MP2. Viéndolos se puede observar que son tautómeros en los que un hidrógeno cambió de lugar; aunque ya se sabía la existencia de tautómeros en estos tipos de moléculas, comprobar esto experimentalmente sería algo novedoso.

Isómeros de guanina utilizando aproximación Hartree-Fock.

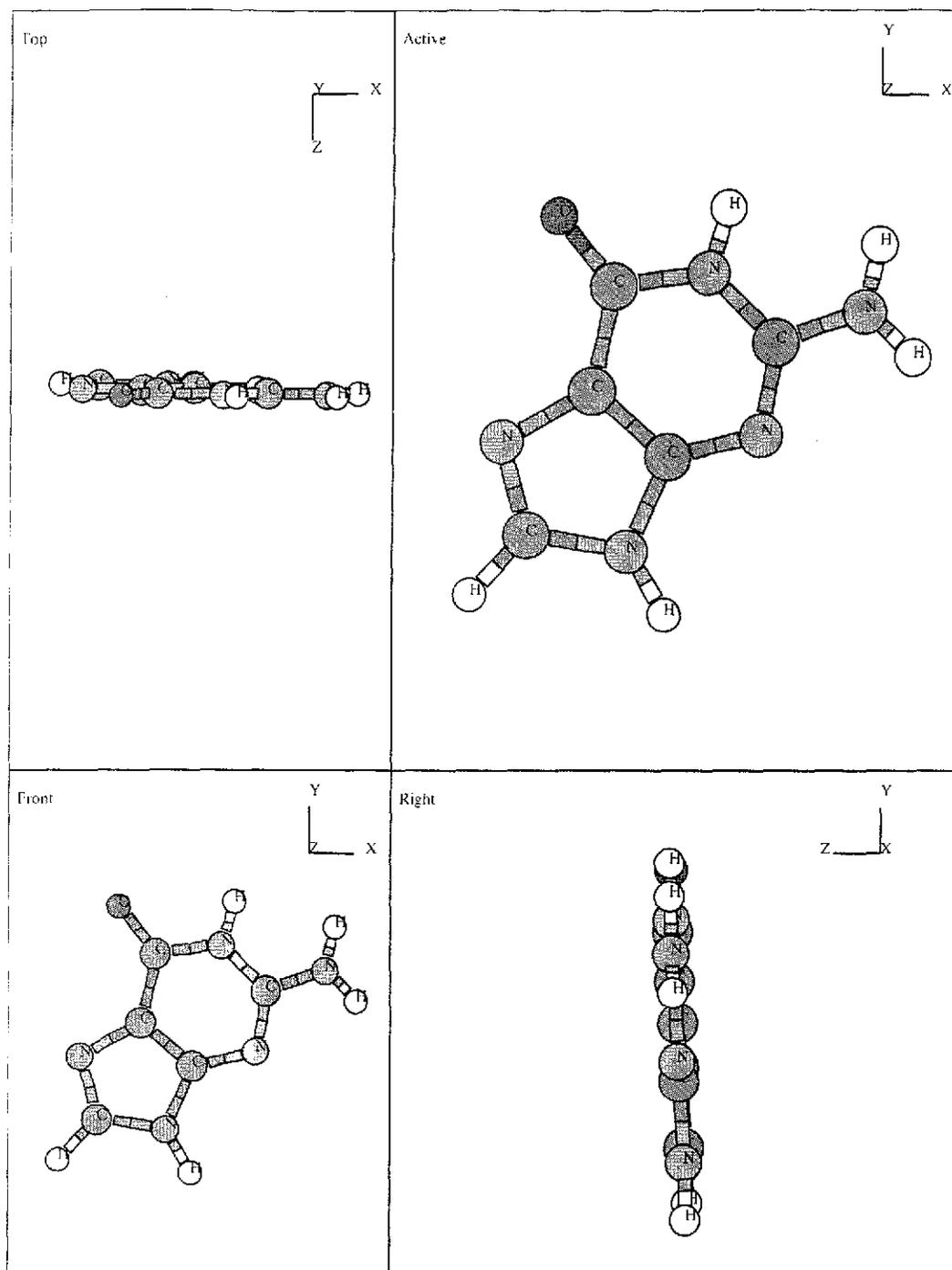


Figura 4.13: Geometría resultante de la figura 3.19 correspondiente a G1, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

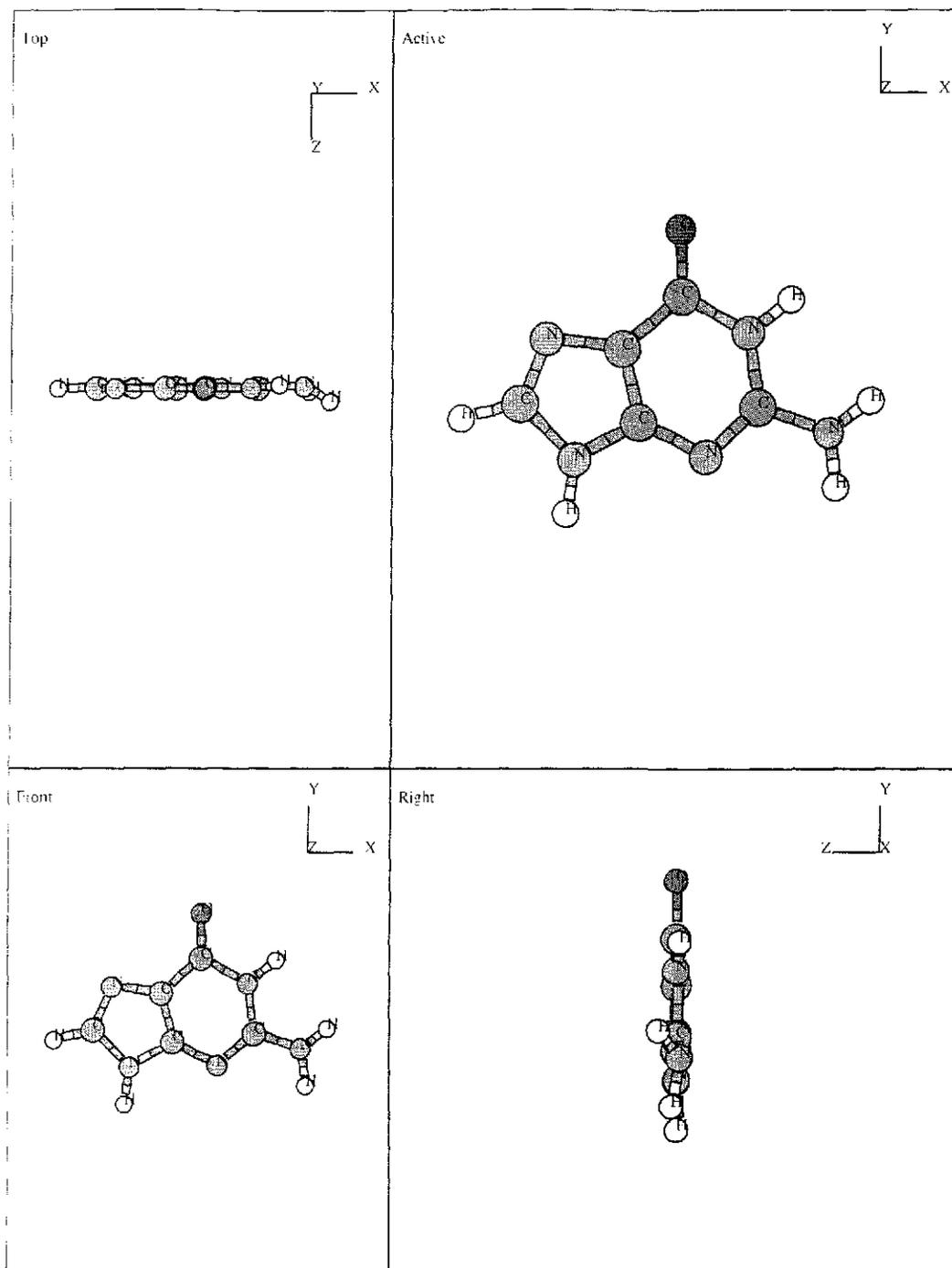


Figura 4.14: Geometría resultante de la figura 3.20 correspondiente a **G2**, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

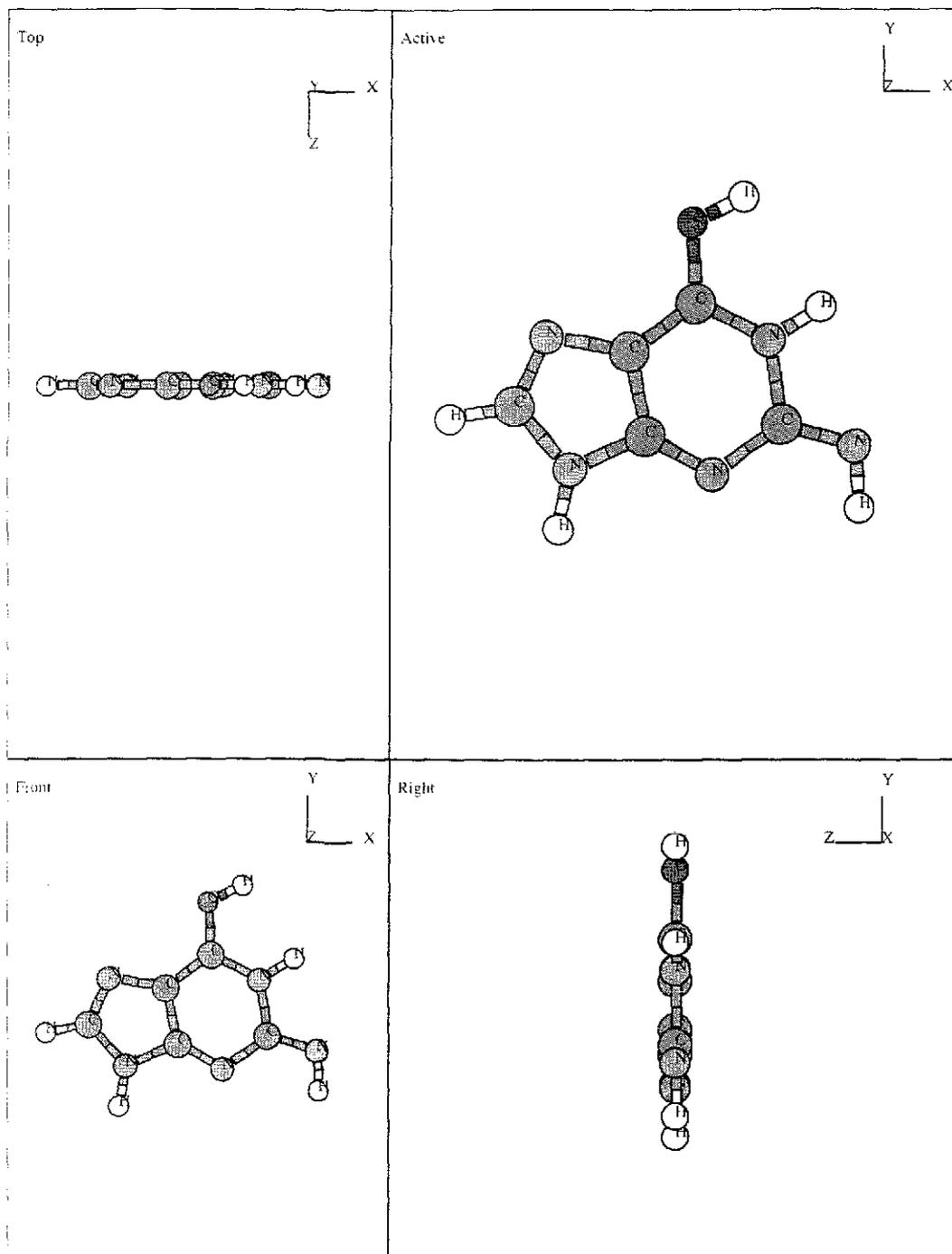


Figura 4.15: Geometría resultante de la figura 3.21 correspondiente a **G3**, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

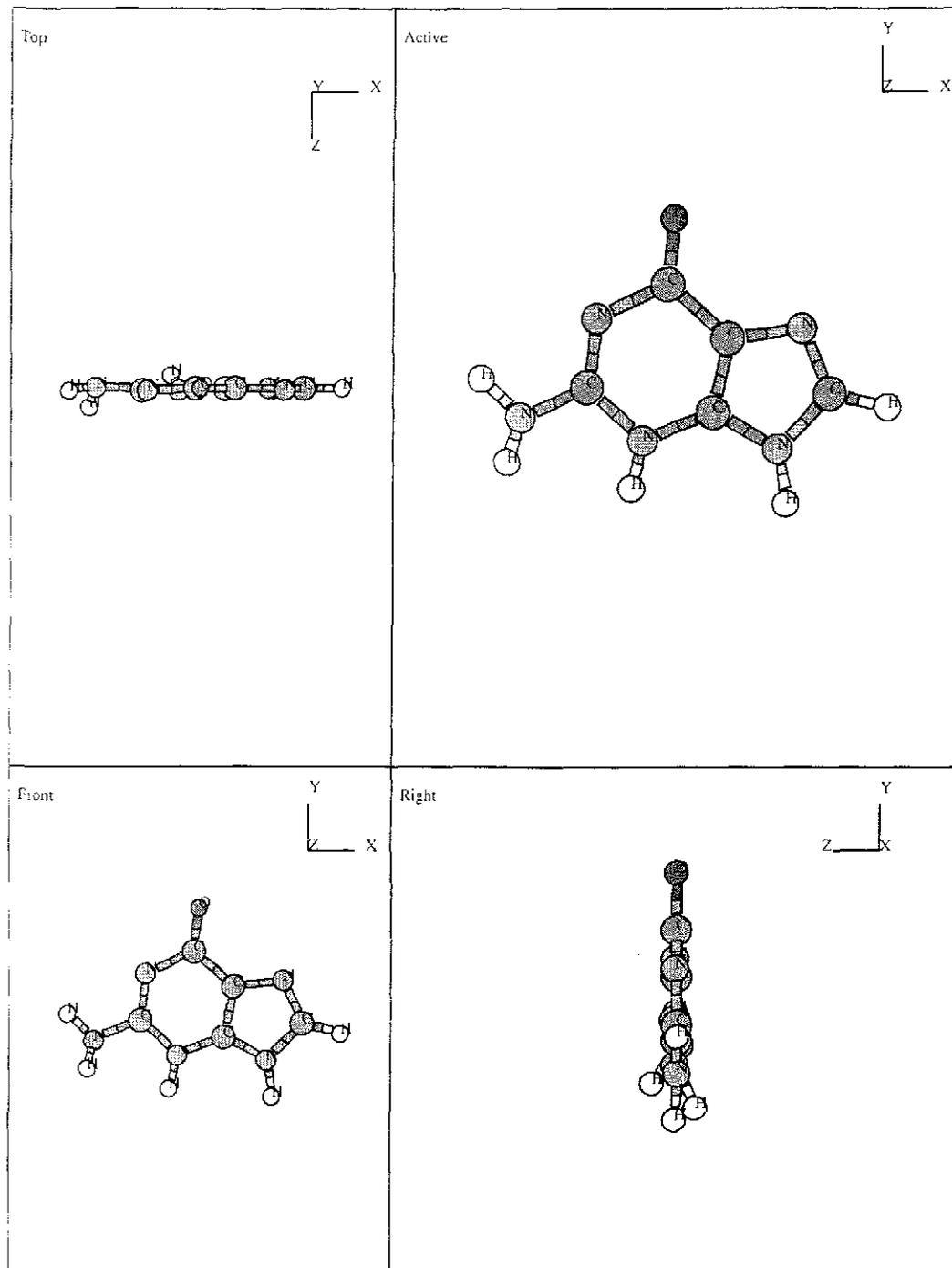


Figura 4.16: Geometría resultante de la figura 3.22 correspondiente a **G4**, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

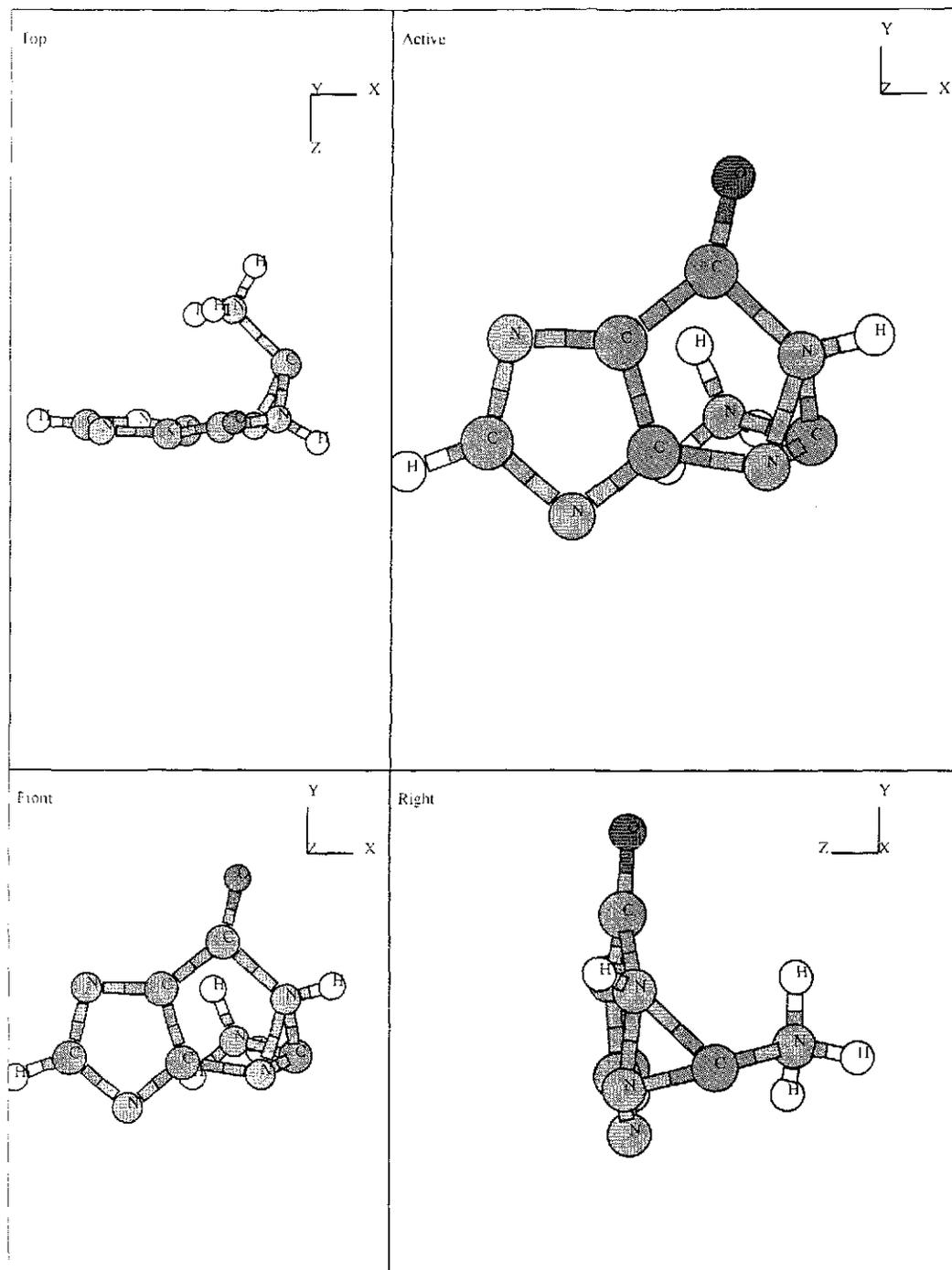


Figura 4.17: Geometría resultante de la figura 3.23 correspondiente a G^* , después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

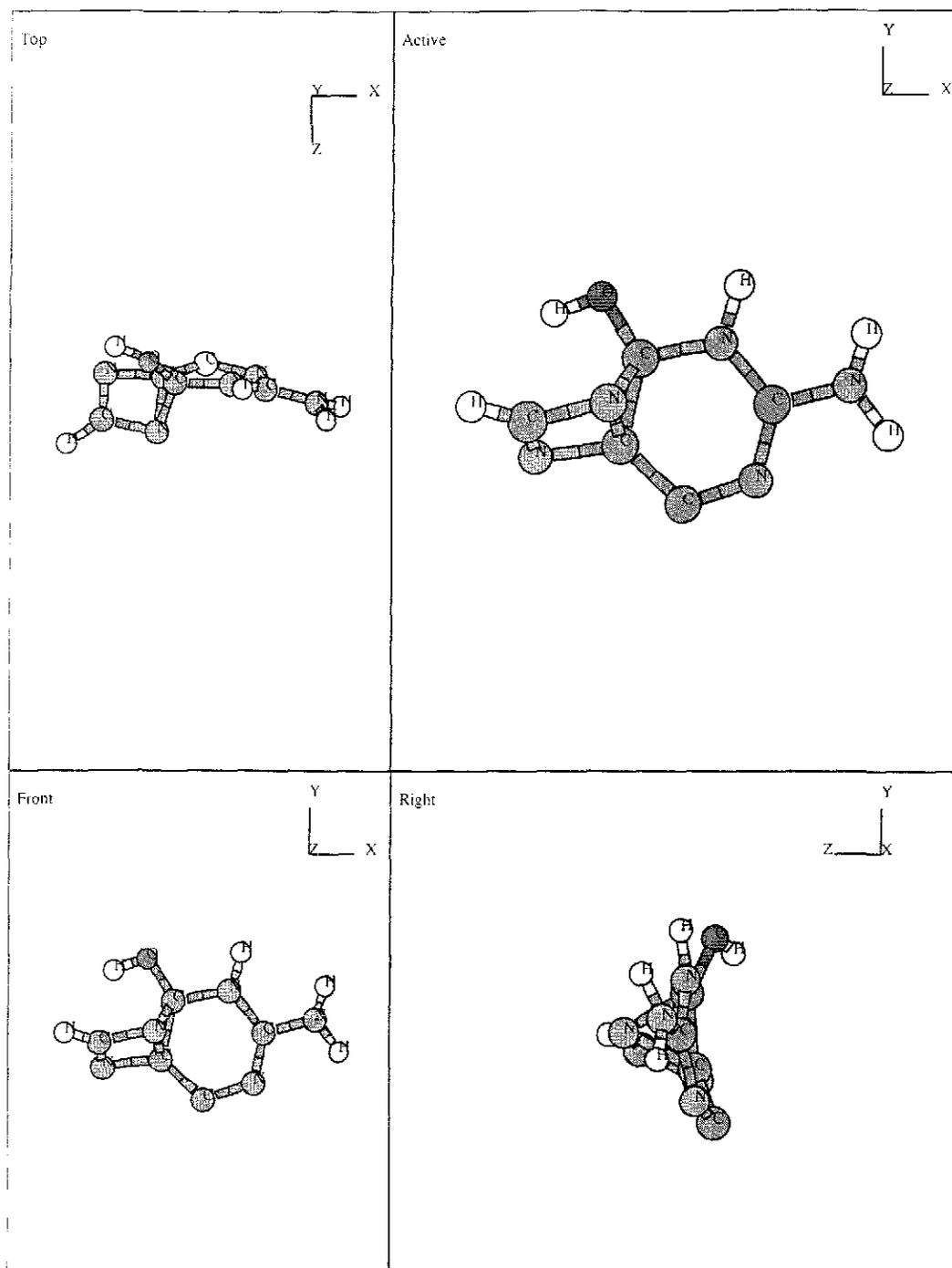


Figura 4 18: Geometría resultante de la figura 3 24 correspondiente a G5, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

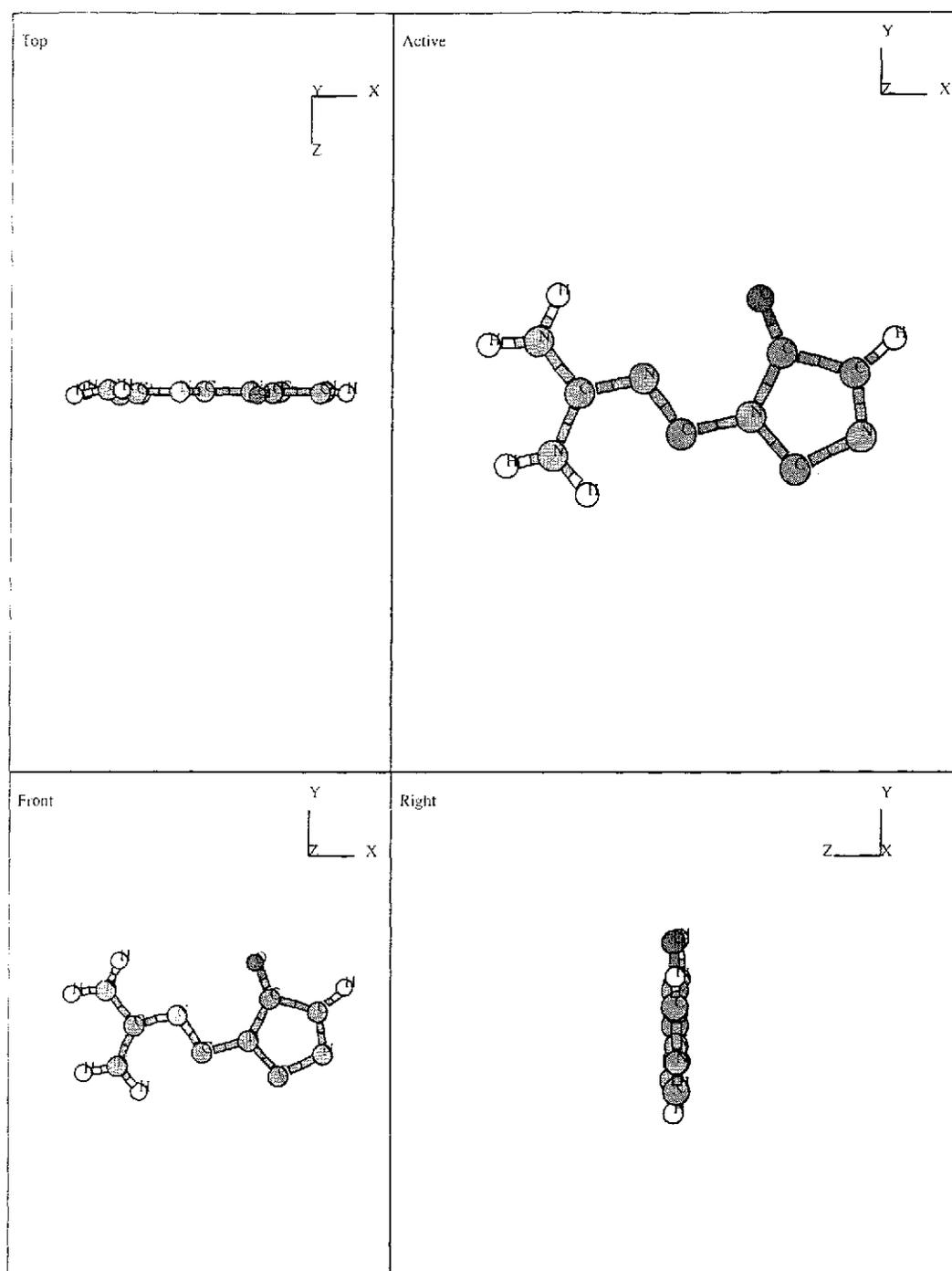


Figura 4.19: Geometría resultante de la figura 3.25 correspondiente a **G6**, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

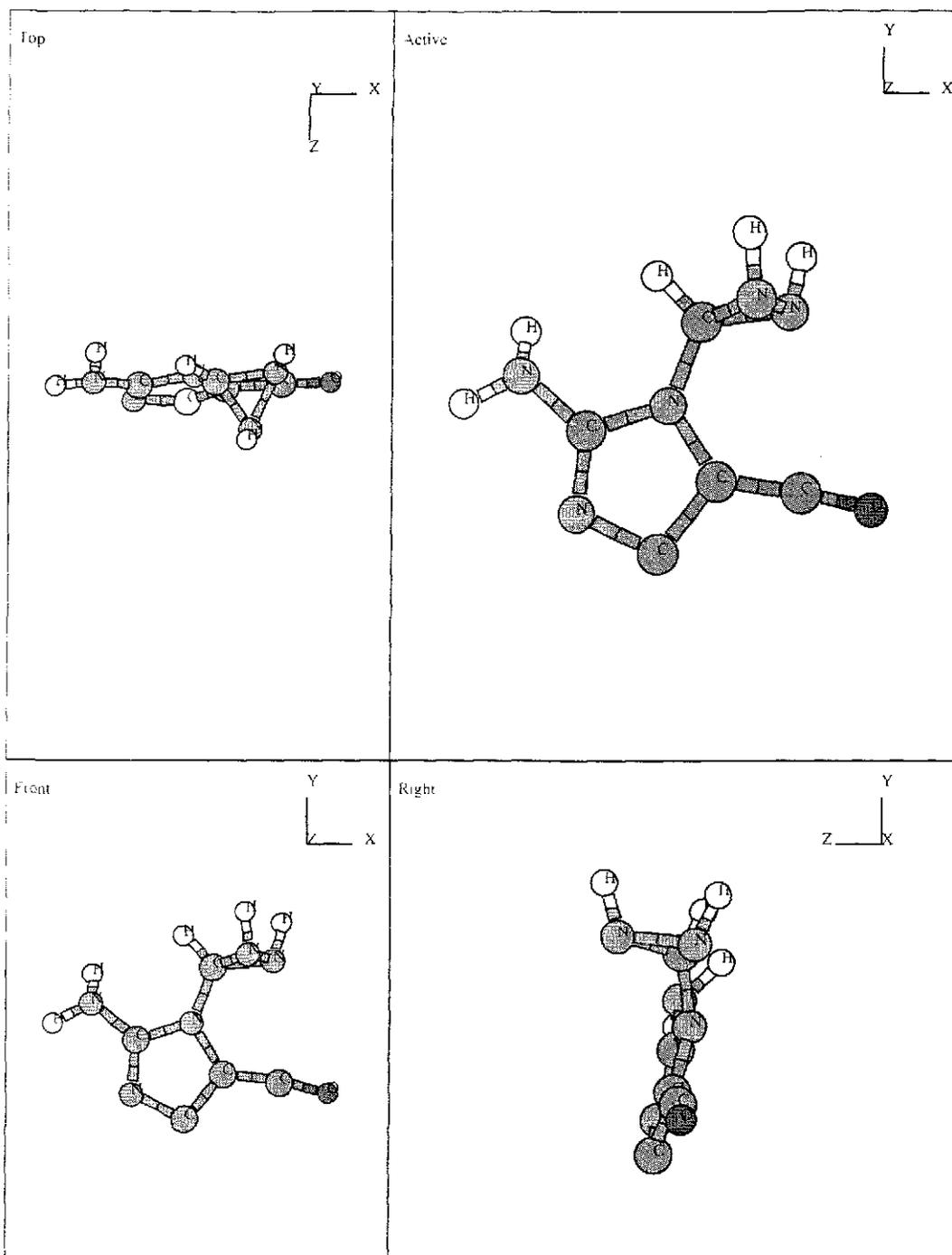


Figura 4 20: Geometría resultante de la figura 3 26 correspondiente a **G7**, después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

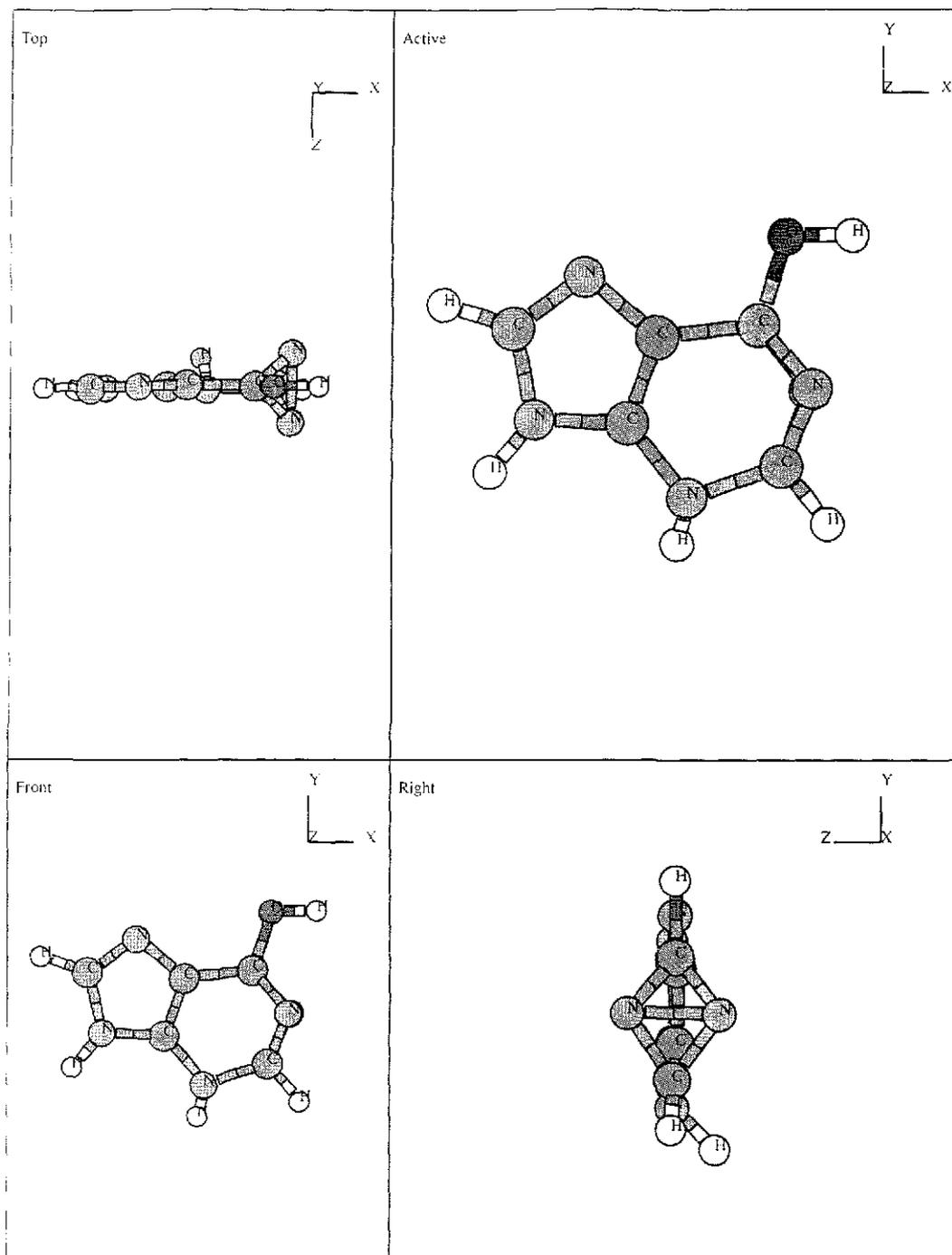


Figura 4 21: Geometría resultante al variar la estructura de un isómero de adenina. isómero G8.

TESIS CON
FALLA DE ORIGEN

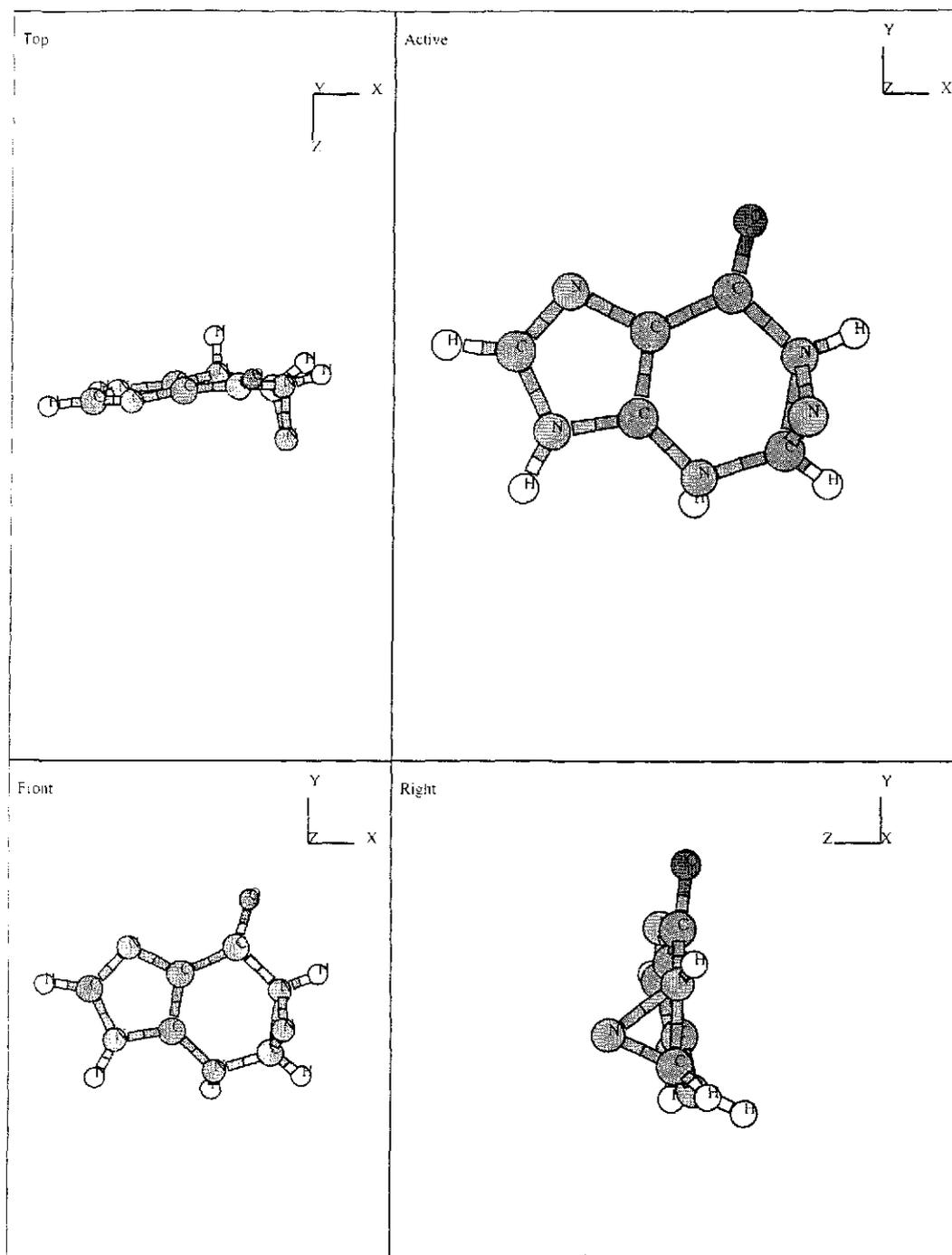


Figura 4.22: Geometría resultante al variar la estructura de un isómero de adenina. isómero **G9**.

Isómeros de guanina utilizando aproximación MP2.

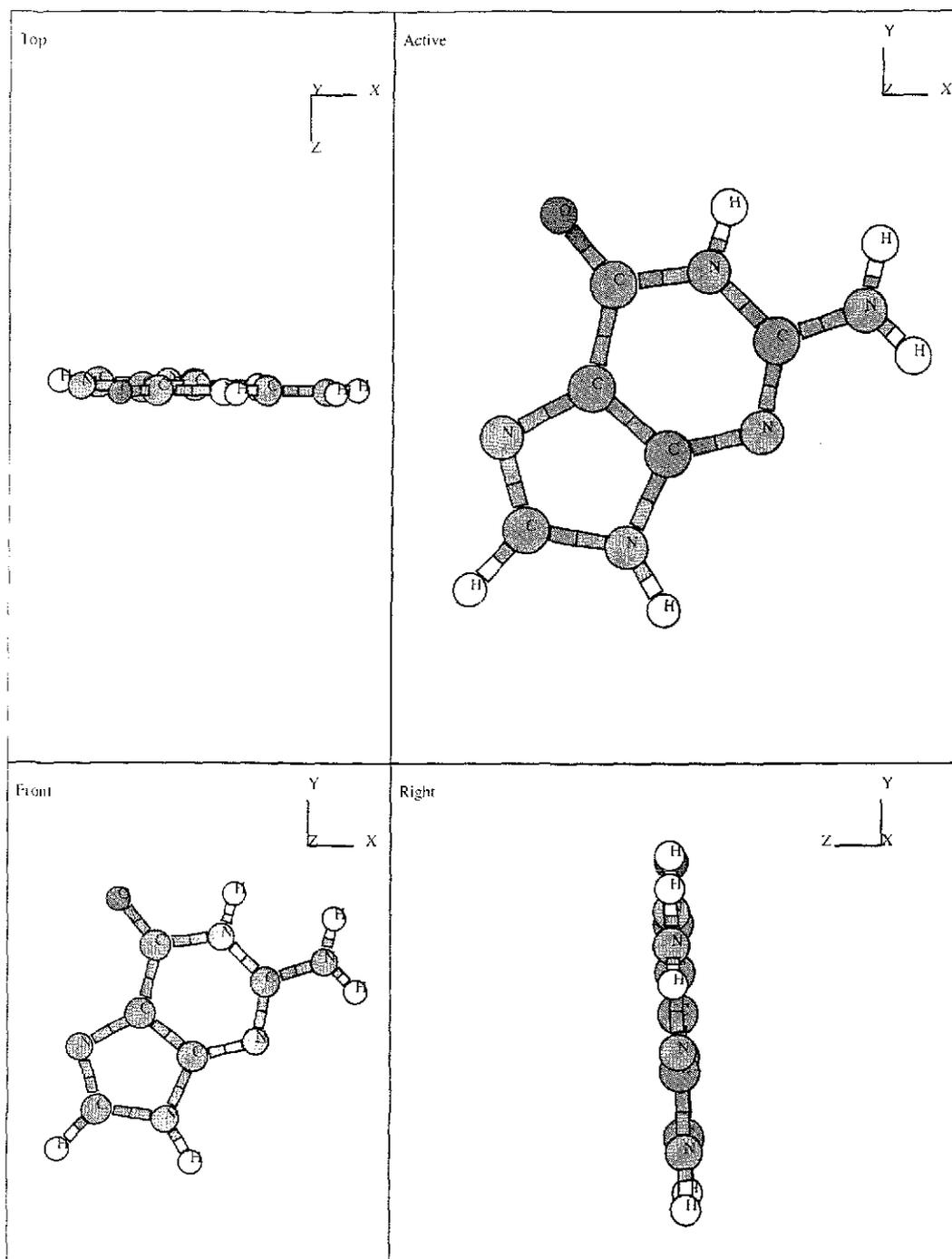


Figura 4.23: Geometría resultante de la figura 4.13 (G1), después de optimización local usando cálculos MP2.

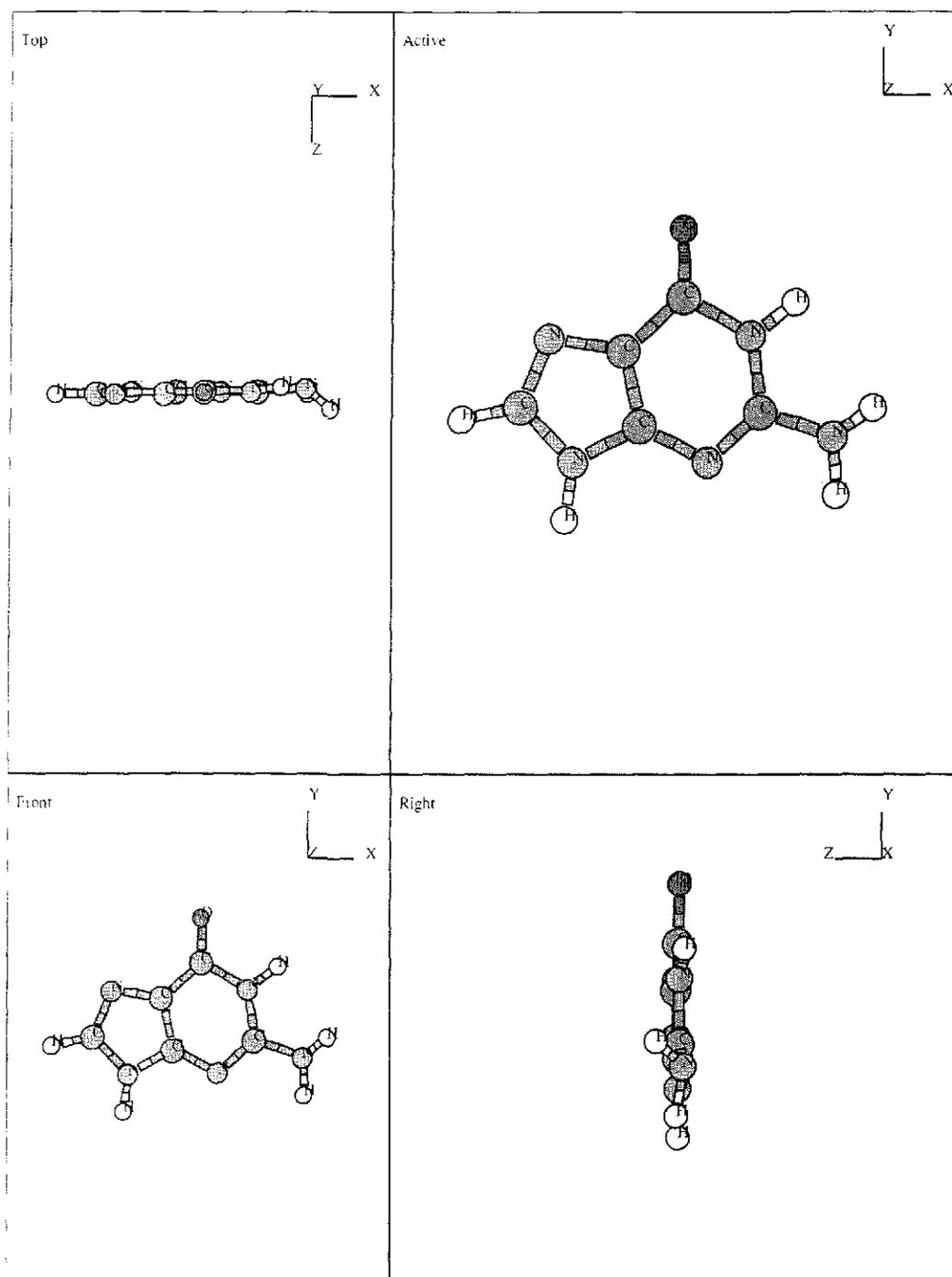


Figura 4.24: Geometría resultante de la figura 4.14 (G2), después de optimización local usando cálculos MP2

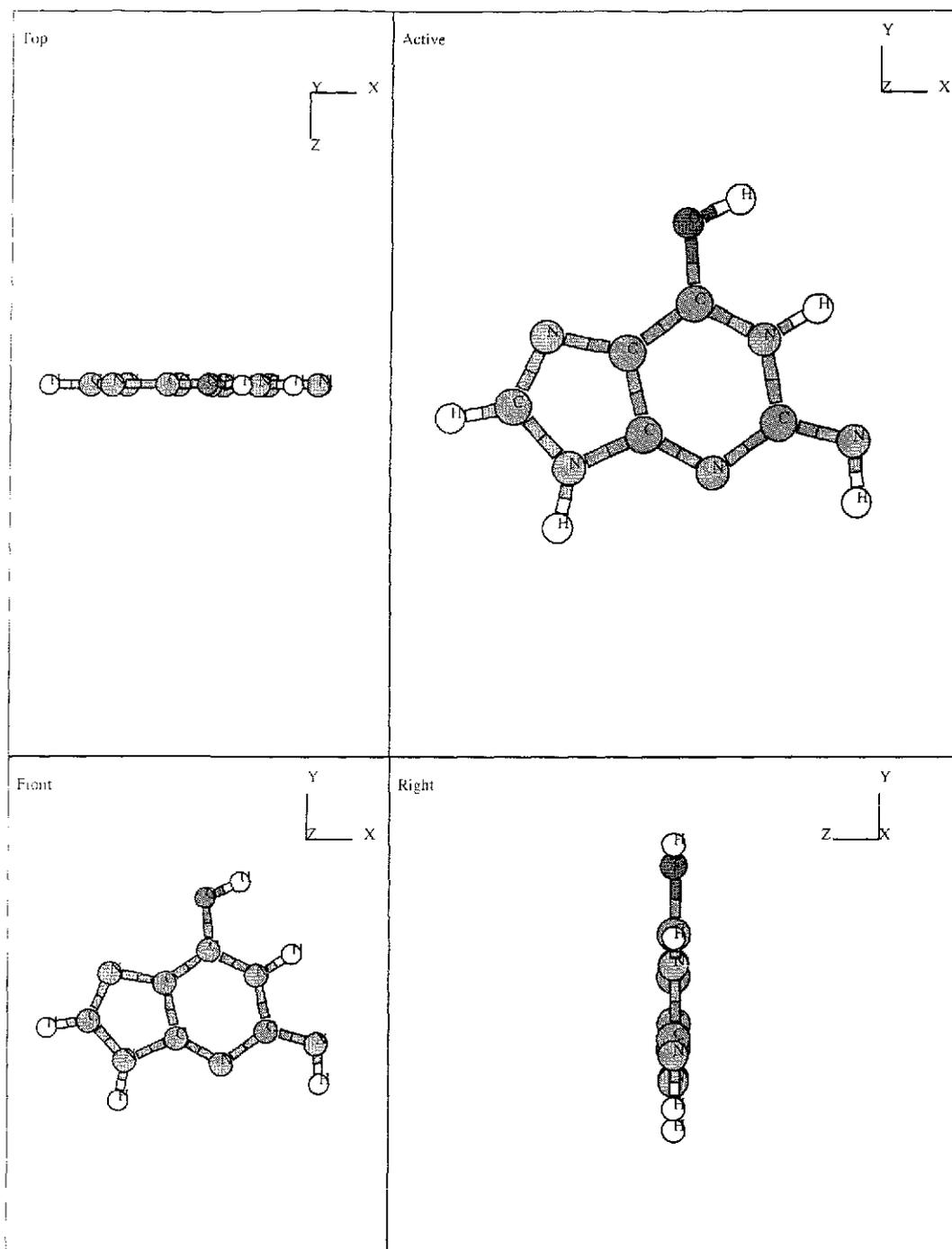


Figura 4 25: Geometría resultante de la figura 4.15 (**G3**), después de optimización local usando cálculos MP2.

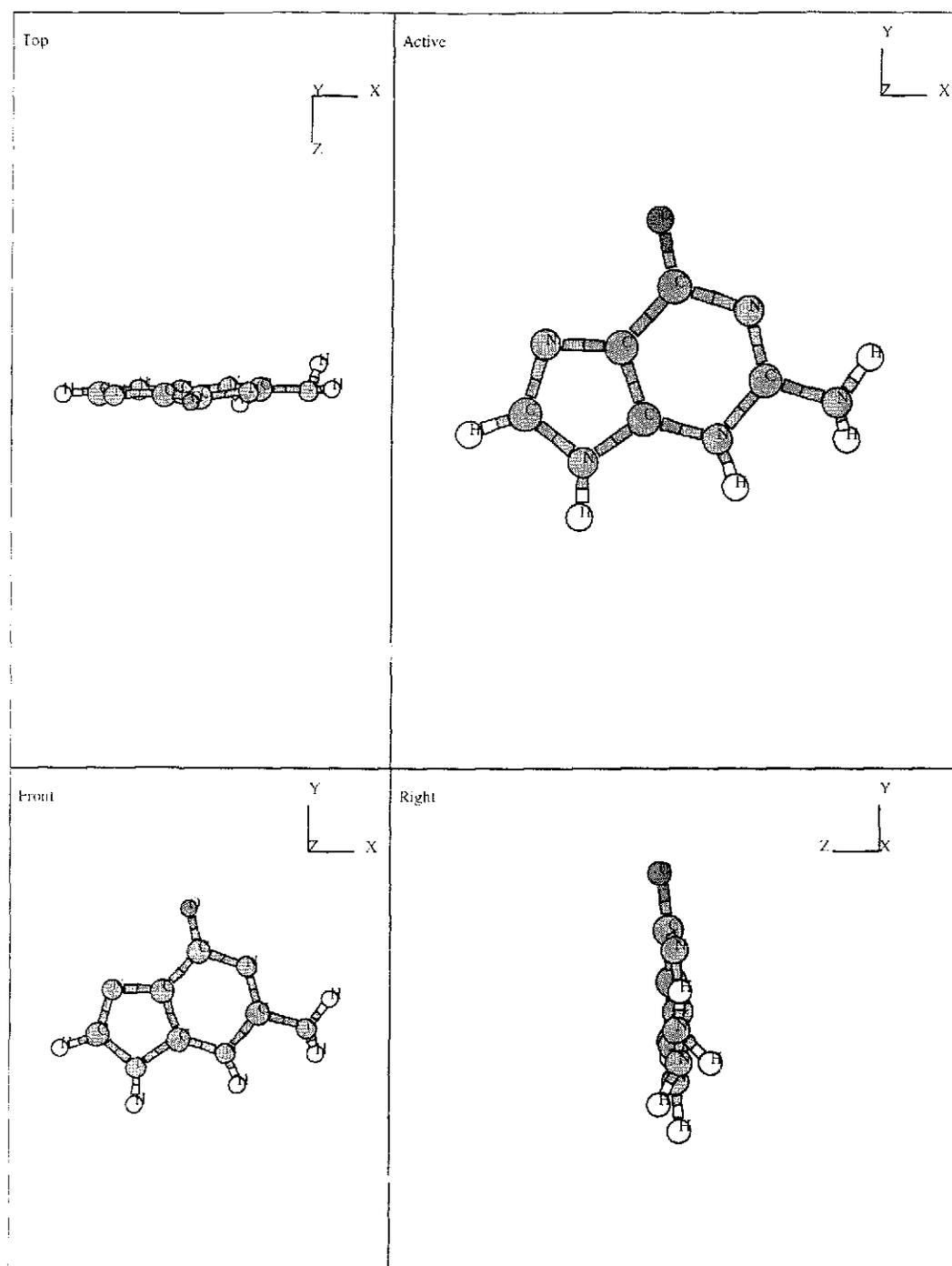


Figura 4.26: Geometría resultante de la figura 4.16 (G4), después de optimización local usando cálculos MP2.

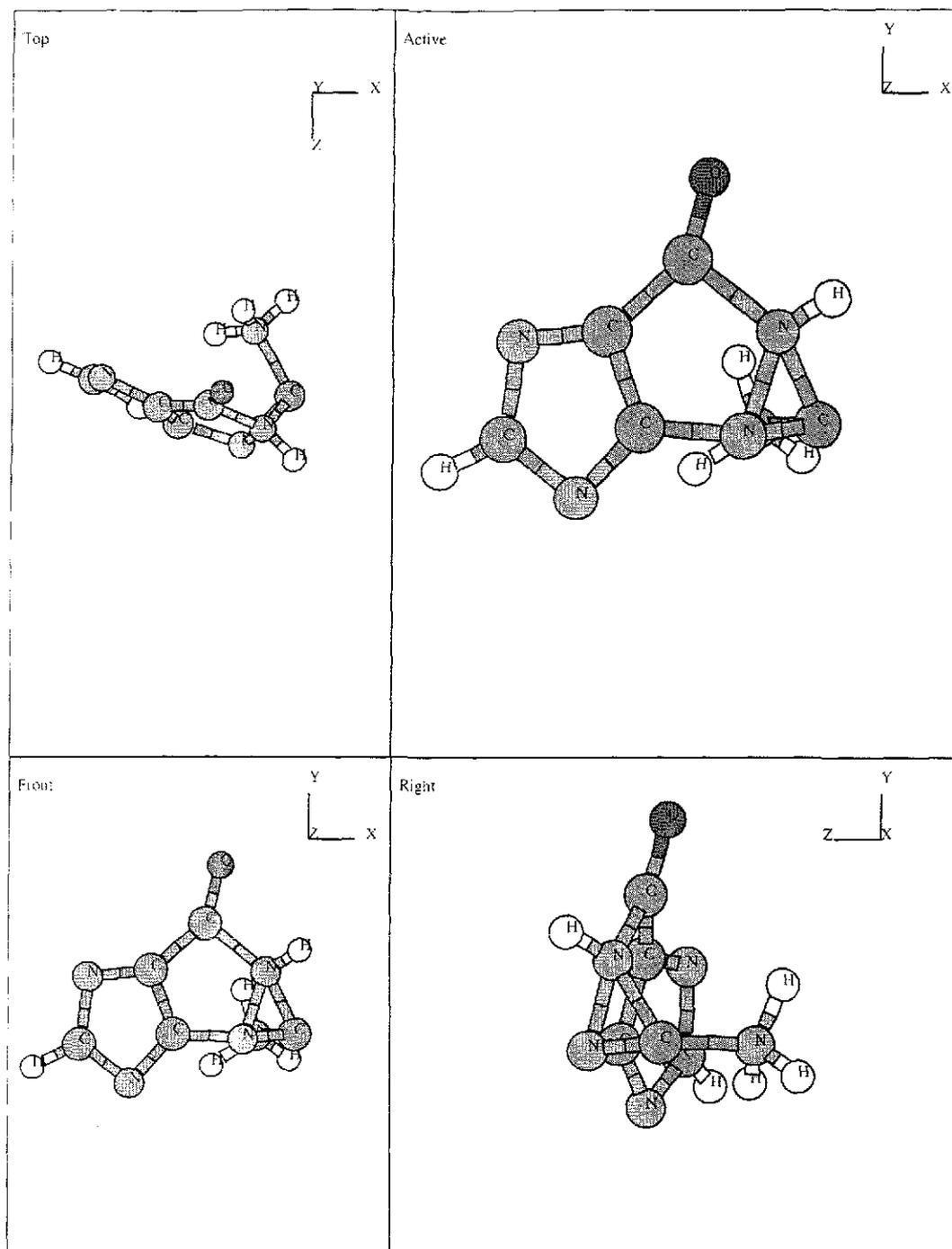


Figura 4 27: Geometría resultante de la figura 4.17 (G^*), después de optimización local usando cálculos MP2

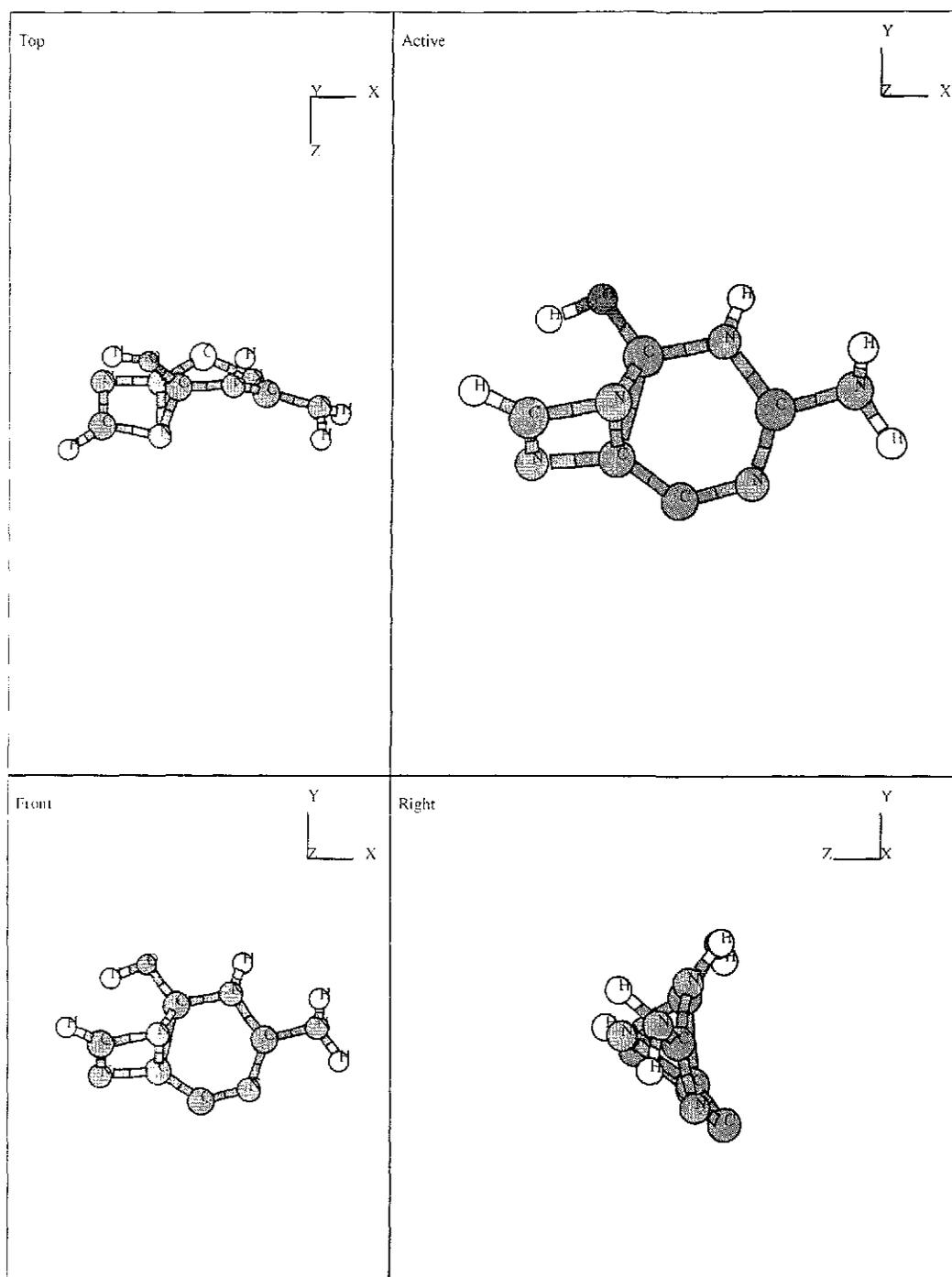


Figura 4.28: Geometría resultante de la figura 4.18 (G5), después de optimización local usando cálculos MP2

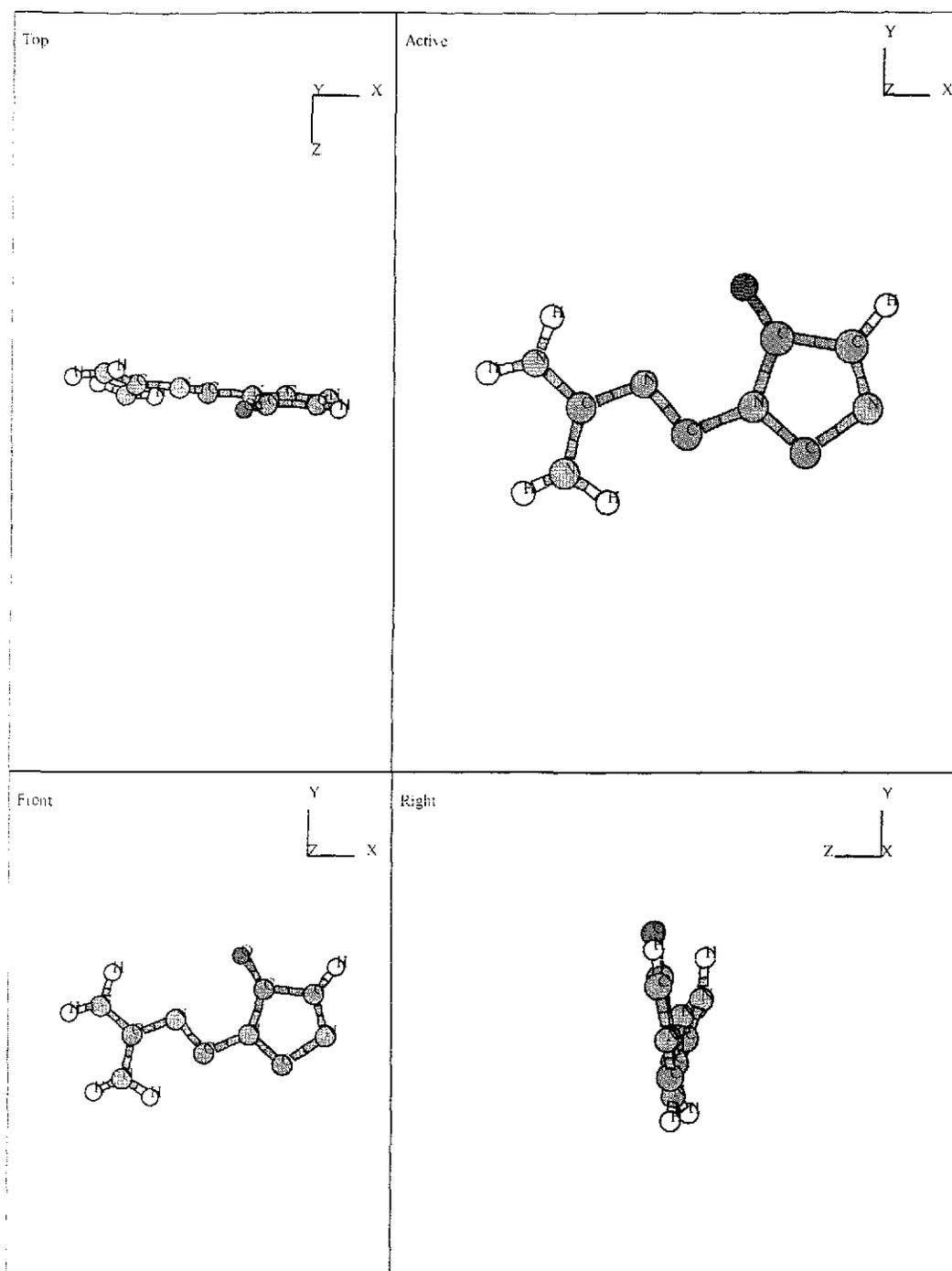


Figura 4.29: Geometría resultante de la figura 4.19 (G6), después de optimización local usando cálculos MP2.

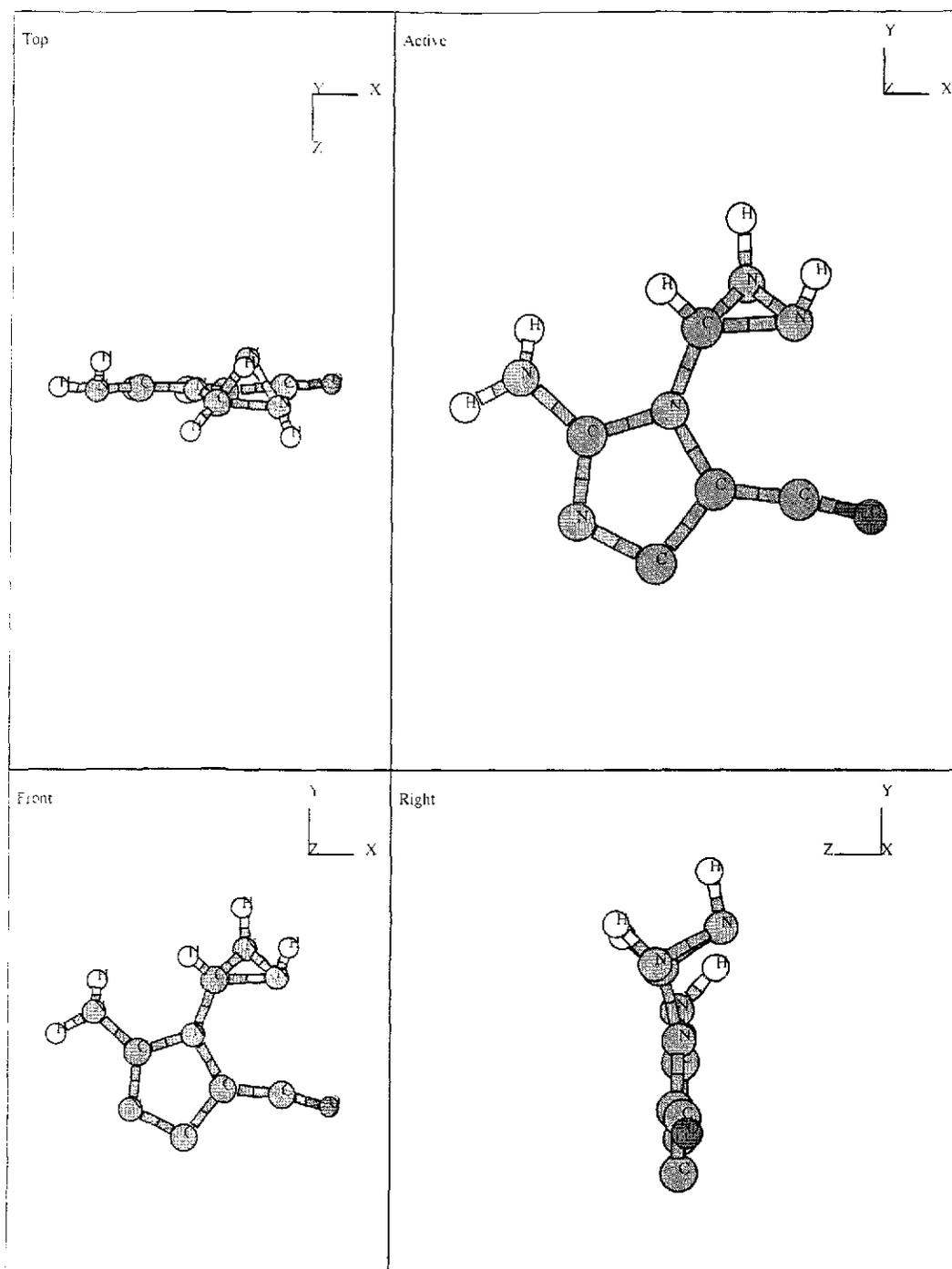


Figura 4.30: Geometría resultante de la figura 4.20 (**G7**), después de optimización local usando cálculos MP2.

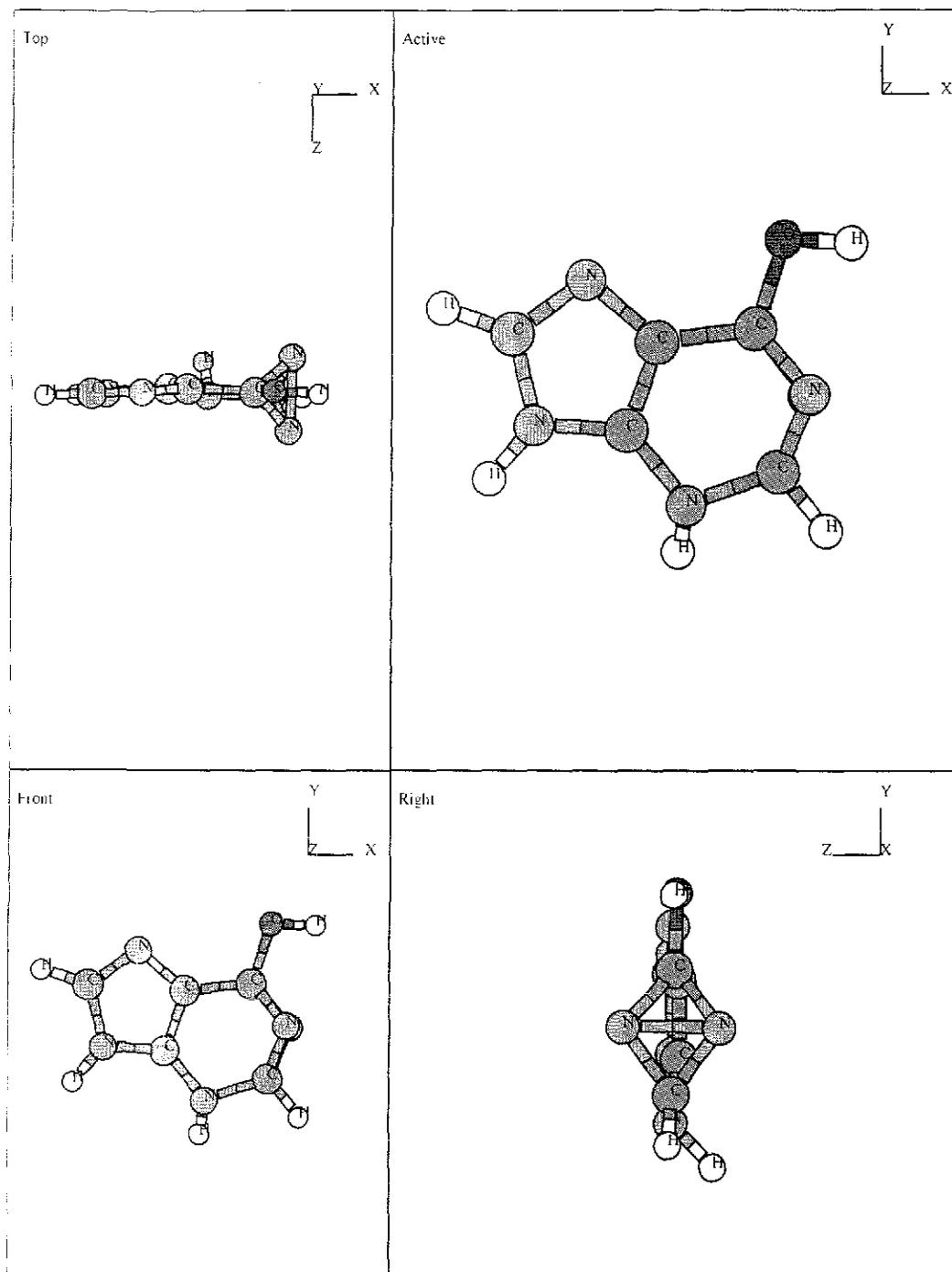


Figura 4.31: Geometría resultante de la figura 4.21 (**G8**), después de optimización local usando cálculos MP2.

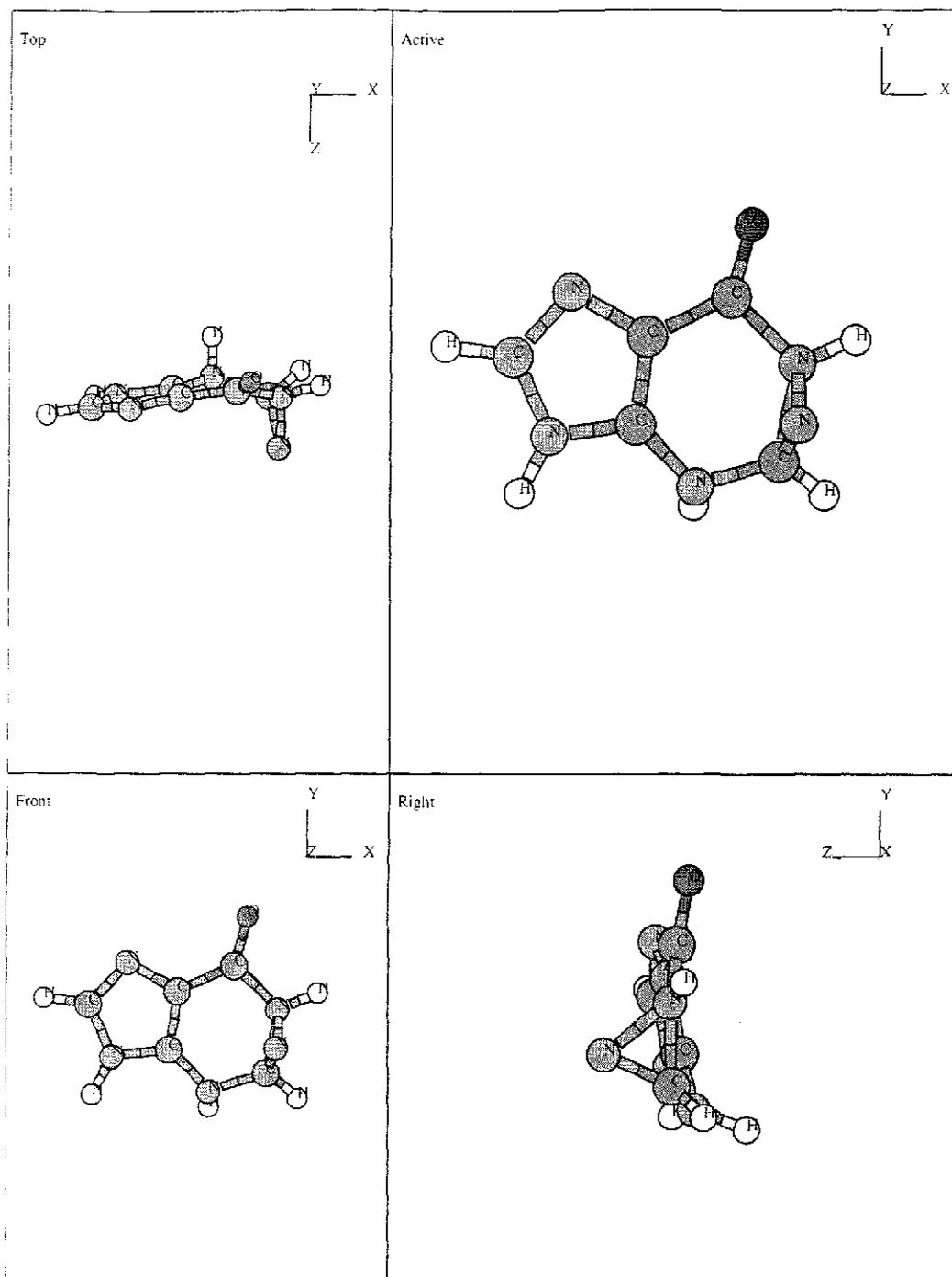


Figura 4.32: Geometría resultante de la figura 4.22 (**G9**), después de optimización local usando cálculos MP2.

4.9 Resultados para citosina

Aunque para citosina no se encontraron isómeros estables sino únicamente tautómeros, a continuación se muestran los valores de energía para estas configuraciones utilizando el potencial, cálculos cuánticos y MP2:

Para la molécula citosina los resultados son:

Tabla 4.5 Energías totales obtenidas con métodos Hartree-Fock y MP2 para las diferentes conformaciones de la **citosina** encontradas con el potencial y el algoritmo genético.

Núm. de isómero	Potencial kcal/mol	Hartree-Fock Hartrees	MP2 Hartrees
C1	-685.298	-392.615	-393.677
C2	-585.517	-392.604	-393.749
C3	-561.758	-392.615	-393.760
C4	-556.664	-392.615	-393.760

Tabla 4.6 Comparación de las diferencias en la energía total entre la configuración natural de la **citosina** (isómero C1) y los isómeros de más alta energía listados en las figs. C1-C4 para el potencial modelo, Hartree-Fock y cálculos MP2.

Núm. de isómero	Ener pot kcal/mol	Ener HF kcal/mol	Ener. MP2 kcal/mol
C1	0.0	0.0	0.0
C2	99.781	6.903	-45.181
C3	123.540	0.0	-52.083
C4	128.634	0.0	-52.083

Es interesante notar que los tautómeros encontrados tienen dentro de la aproximación de MP2 una energía menor que la configuración natural, cosa que no sucede con los cálculos HF ni con el potencial.

Isómeros de citosina utilizando aproximación Hartree-Fock.

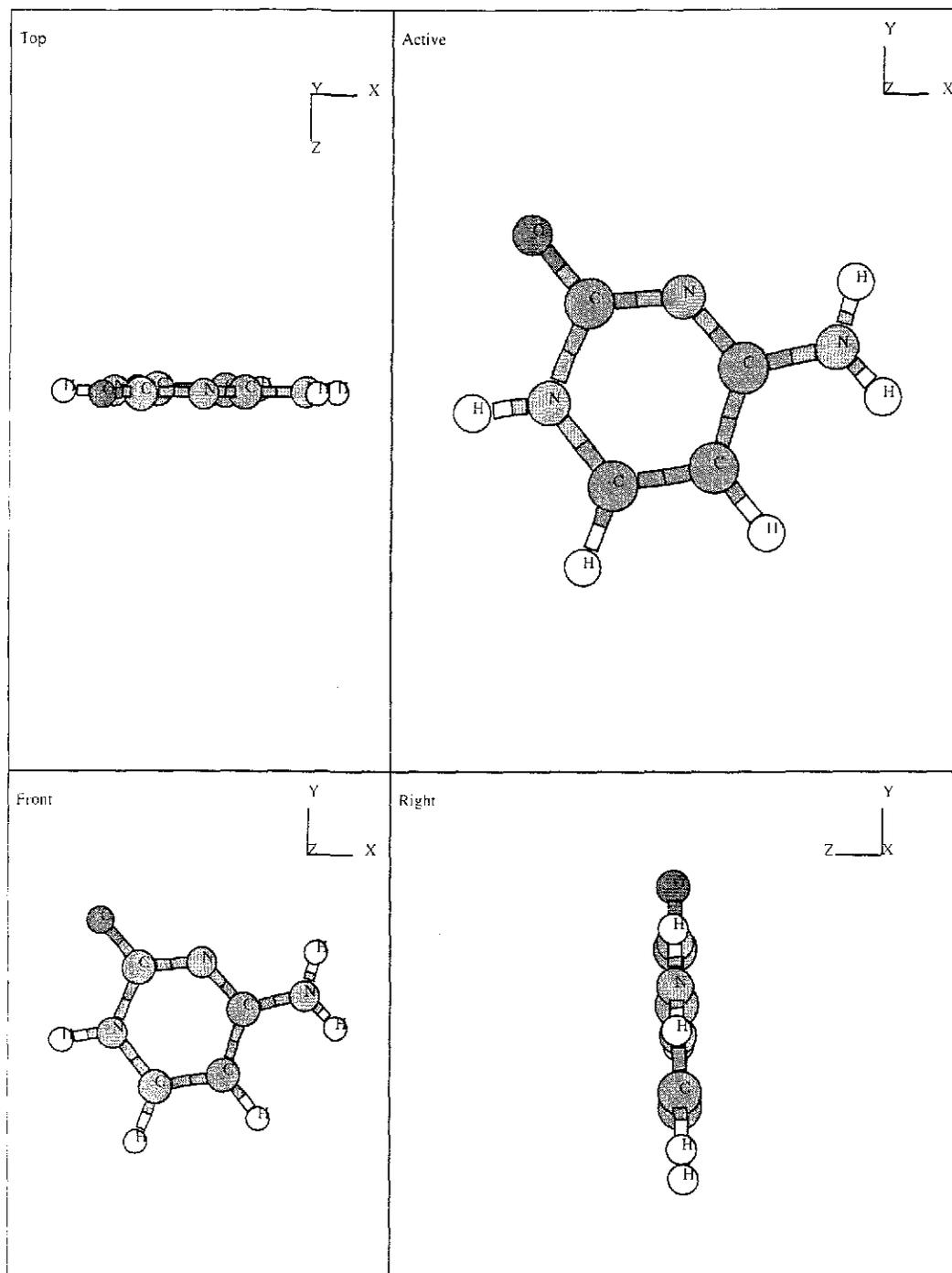


Figura 4.33: Geometría resultante de la figura 3.27 después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

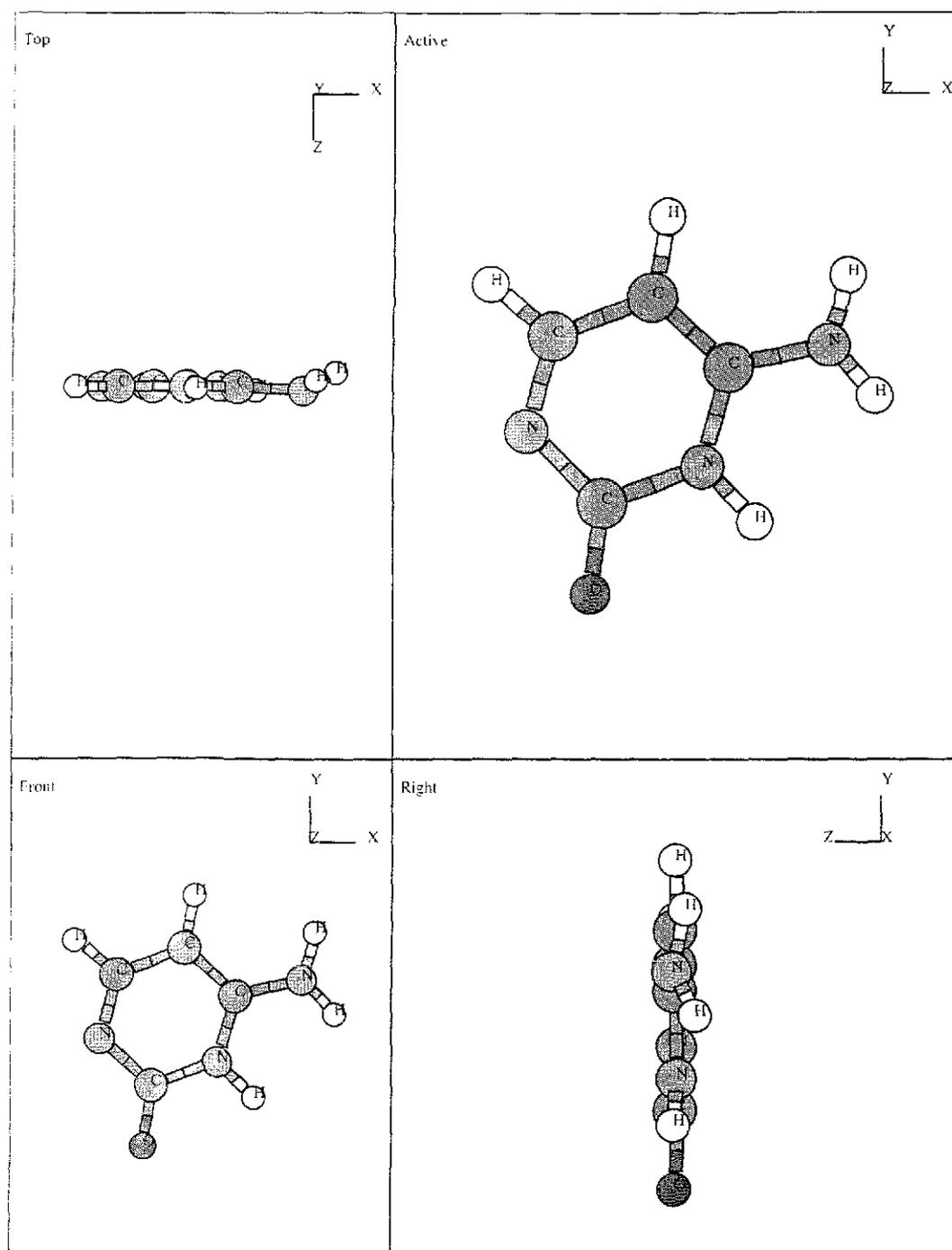


Figura 4.34: Geometría resultante de la figura 3 28 después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

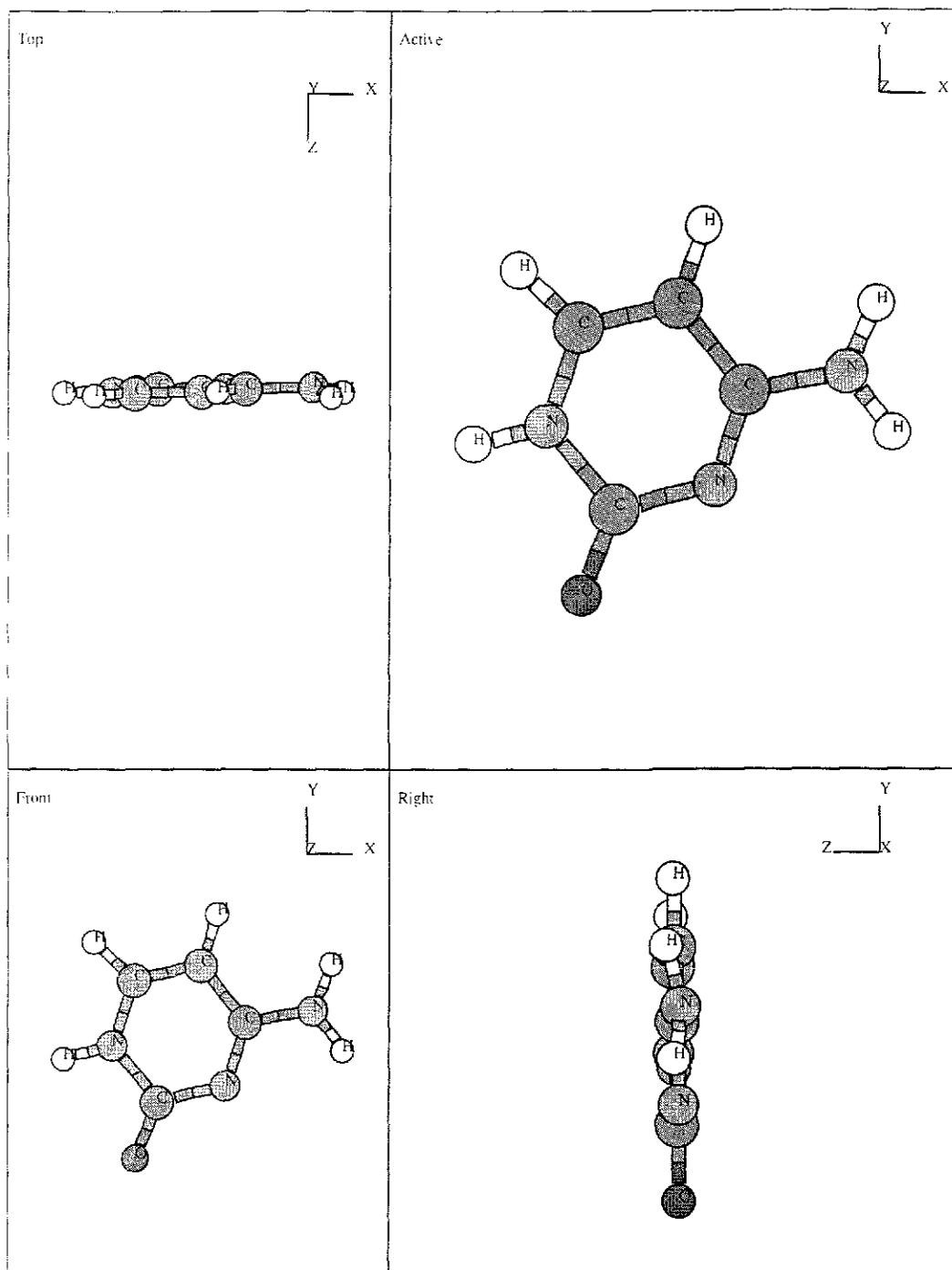


Figura 4.35: Geometría resultante de la figura 3.29 después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

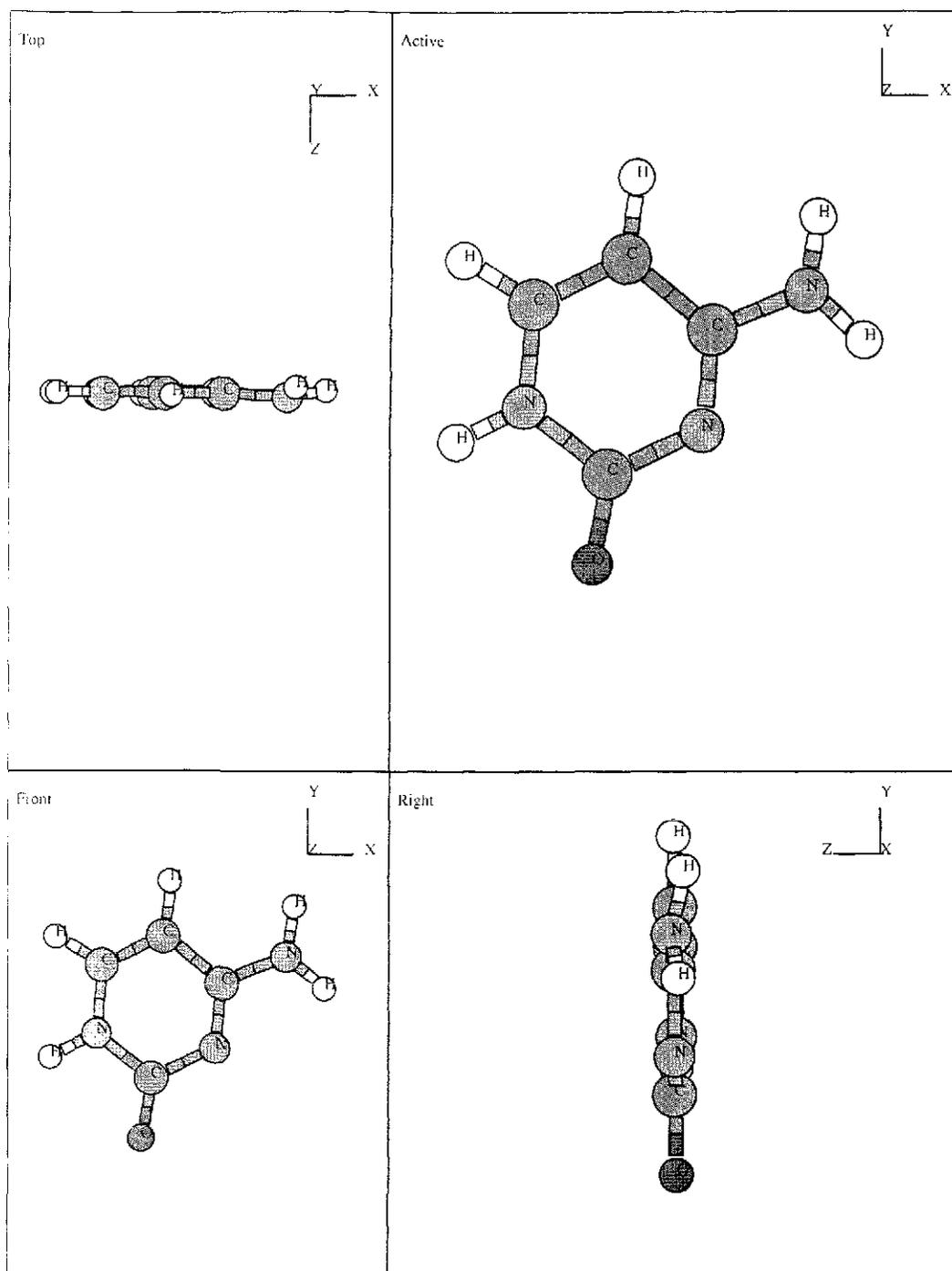


Figura 4.36: Geometría resultante de la figura 3.30 después de optimización local usando cálculos de Hartree-Fock

TESIS CON
FALLA DE ORIGEN

Isómeros de citosina utilizando aproximación MP2.

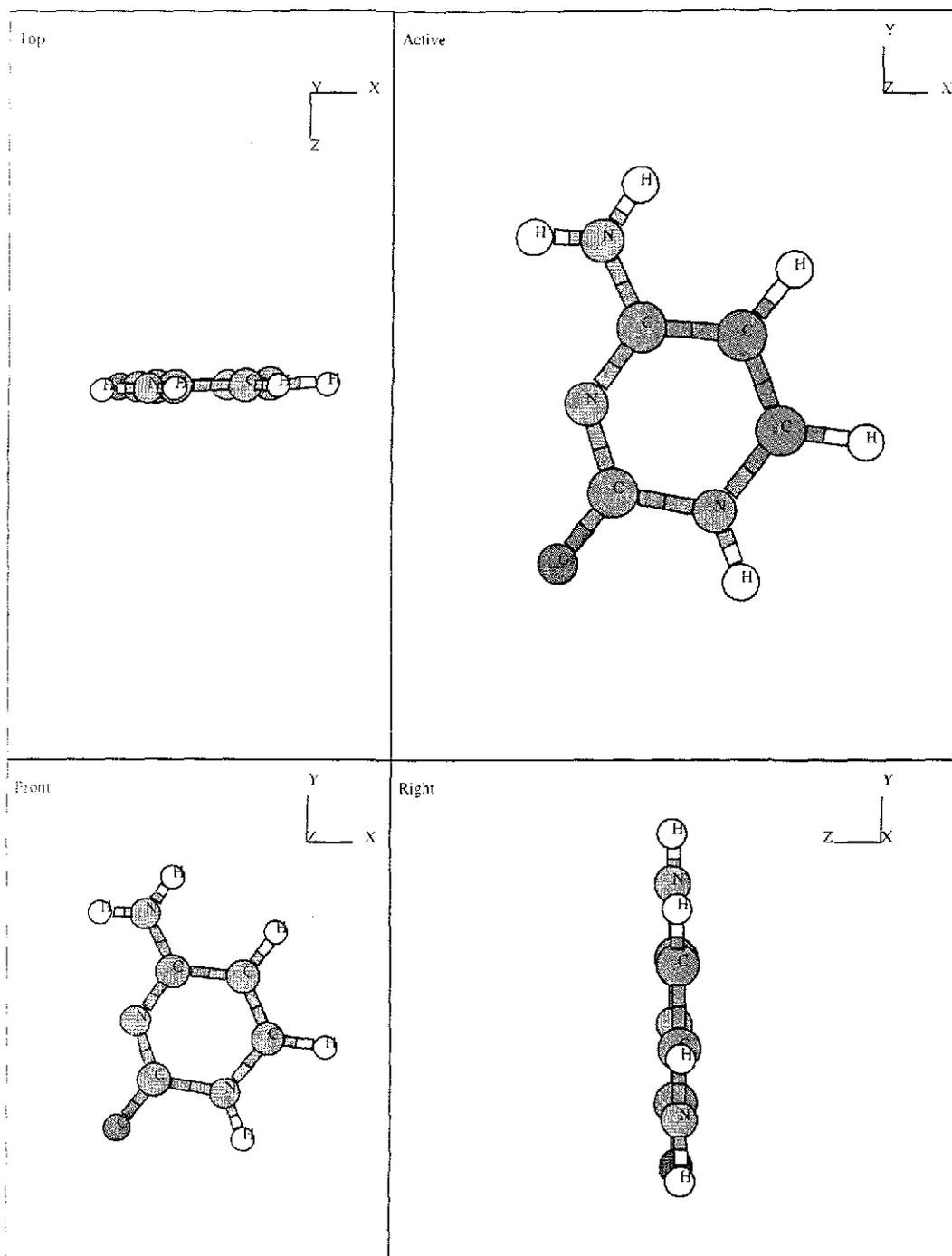


Figura 4.37: Geometría resultante de la figura 4.33 después de optimización local usando cálculos MP2

TESIS CON
FALLA DE ORIGEN

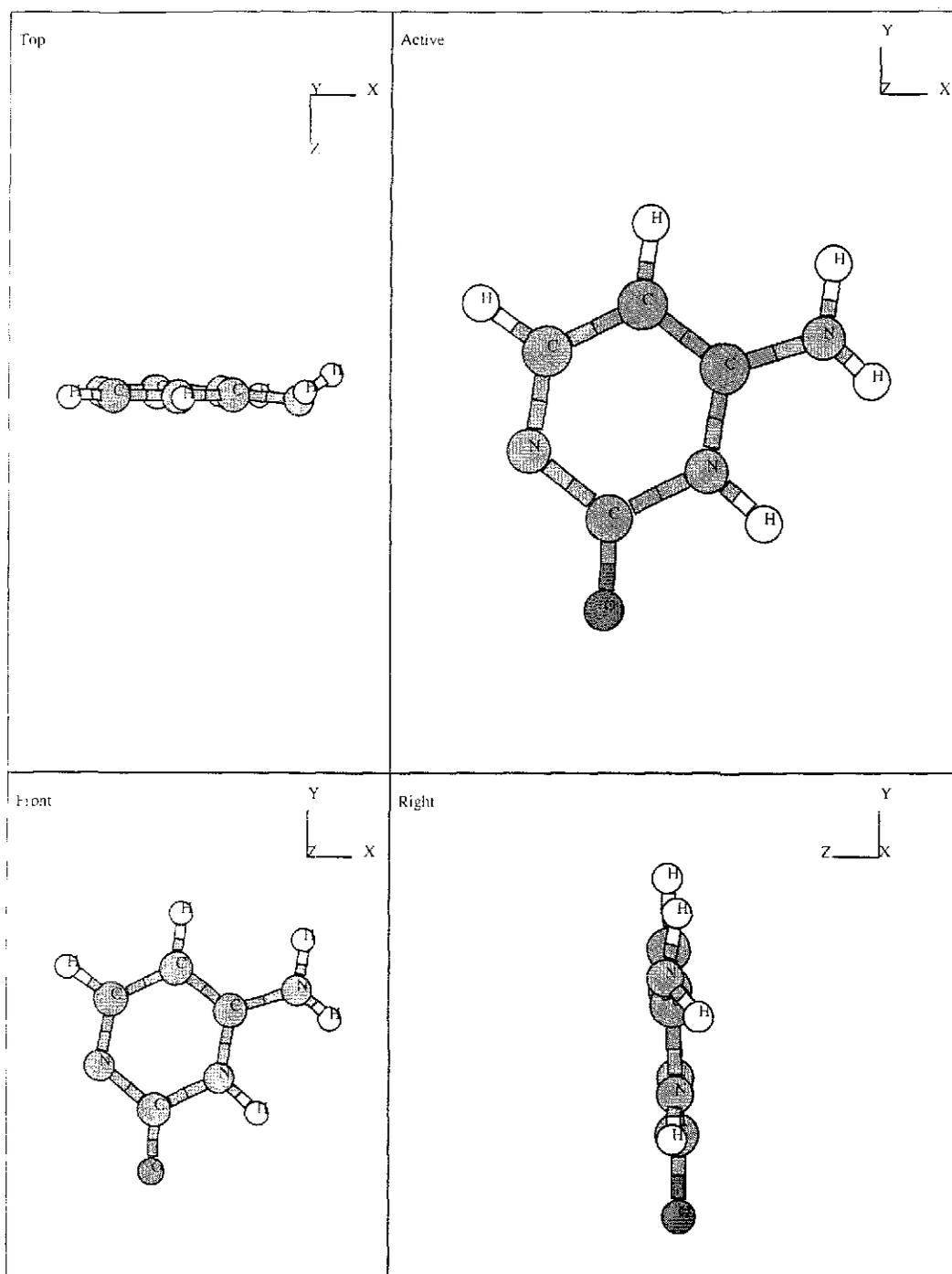


Figura 4.38 Geometría resultante de la figura 4.34 después de optimización local usando cálculos MP2

TESIS CON
FALLA DE ORIGEN

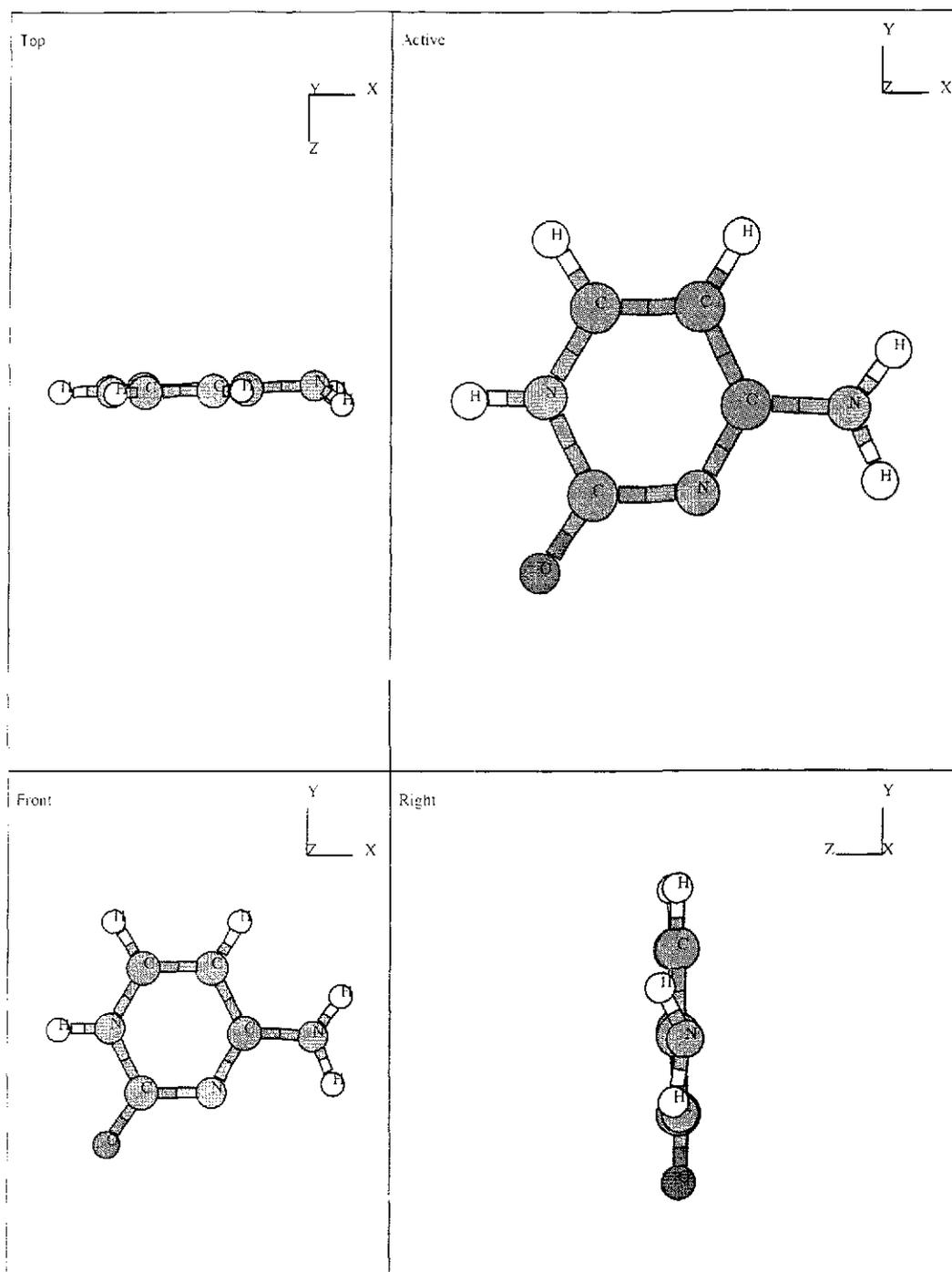


Figura 4.39: Geometría resultante de la figura 4.35 después de optimización local usando cálculos MP2

TESIS CON
FALLA DE ORIGEN

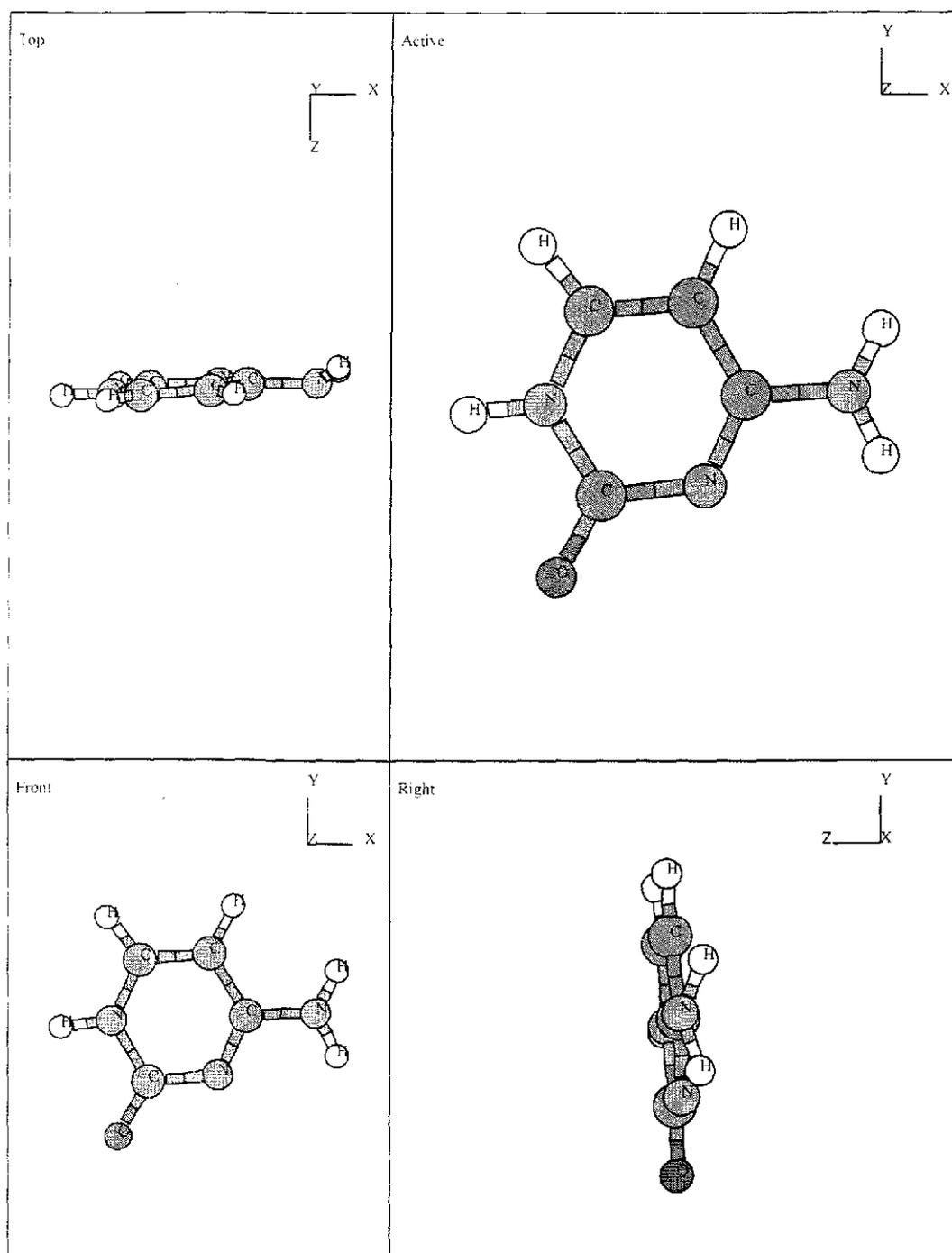


Figura 4.40: Geometría resultante de la figura 4.36 después de optimización local usando cálculos MP2

4.10 Resultados para timina

Para la molécula timina solo se encontró una configuración estable con refinamiento cuántico, correspondiente a la natural. los resultados son:

Tabla 4.7. Energías totales obtenidas con métodos Hartree-Fock y MP2 para las diferentes conformaciones de la **timina** encontradas con el potencial y el algoritmo genético

Núm. de isómero	Potencial kcal/mol	Hartree-Fock Hartrees	MP2 Hartrees
T1	-476.136	-451.510	-452.663

Es aquí necesario el recalcar que el potencial fué ajustado utilizando datos experimentales obtenidos de las configuraciones “naturales”. Si bien la diversidad en número y tipos de compuestos fué grande para su formulación, es probable que este potencial sea adecuado para conformaciones naturales o aquellas que se asemejen a estas. Las configuraciones que se alejen mucho de las naturales es probable que no sean bien descritas por el potencial. Esto limita a que nos concentraremos en aquellas conformaciones interesantes que no se alejen demasiado de la natural. En las siguientes figuras mostramos las configuraciones optimizadas con cálculos cuánticos que tuvieron todas las frecuencias positivas; se utilizó la energía obtenida del potencial para identificarlas con las tablas anteriores

Excepto para la adenina [2], estos nuevos isómeros no se habían encontrado antes, aunque ya se había encontrado que existen diversos tautómeros para las bases biológicas

Capítulo 5

CONSECUENCIAS Y CONCLUSIONES

5.1 Posibles consecuencias de la existencia de nuevos isómeros

En la figura 5.1 mostramos un posible caso de como podría ocurrir una sustitución de nucleótido. Un tipo de error que podríamos llamar “error tipográfico”. En este caso mostramos a un isómero de adenina optimizado con cálculos cuánticos distinto de la configuración natural unido a la base citosina; para hacer la unión nos ayudamos del programa Xmol[138] tratando de que pudiera haber un posible enlace por puentes de hidrógeno. La condición para que ocurra un enlace de hidrógeno es que las distancias entre un hidrógeno u oxígeno, o también entre un hidrógeno y nitrógeno es que debe ser menor que 2.4 Å. Posteriormente la configuración que obtuvimos la optimizamos con cálculos cuánticos y se investigó las frecuencias normales, las cuales dieron valores positivos, es decir, una configuración estable.

Es concebible que durante la replicación de las cadenas de nucleótidos del ADN, cuando estas se separan, un agente externo podría hacer que se reordenen los átomos del nucleótido y se originen puentes de hidrógeno distintos; que permiten el enlace del nucleótido equivocado en la nueva cadena que va a sintetizarse. Por ejemplo, en lugar de que el isómero de adenina se enlace con timina, debido a su nueva conformación espacial se enlaza a citosina. Cuando

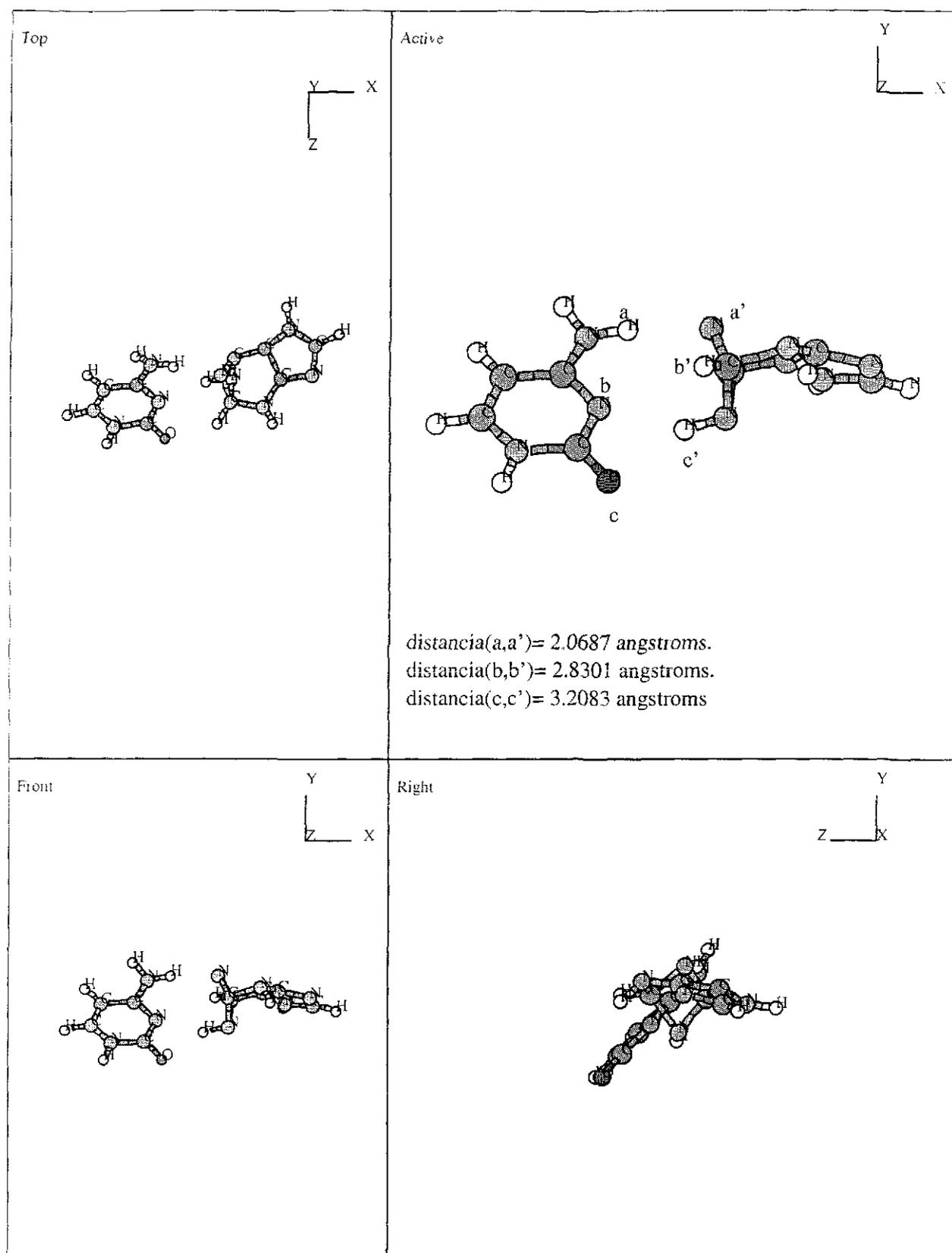


Figura 5 1: Configuración estable con optimización local cuántica MP2 de la unión de un isómero de la adenina unido a la citosina.

algo así ocurre en la replicación siguiente, la cadena con la citosina se aparea con guanina y la cadena central original se aparea con timina. Así surgiría una molécula de ADN en la siguiente generación en la que guanina-citosina sustituiría a adenina-timina figura 5.2. el comprobar experimentalmente esto sería en verdad muy interesante. Este tipo de error podría ocurrir por efectos

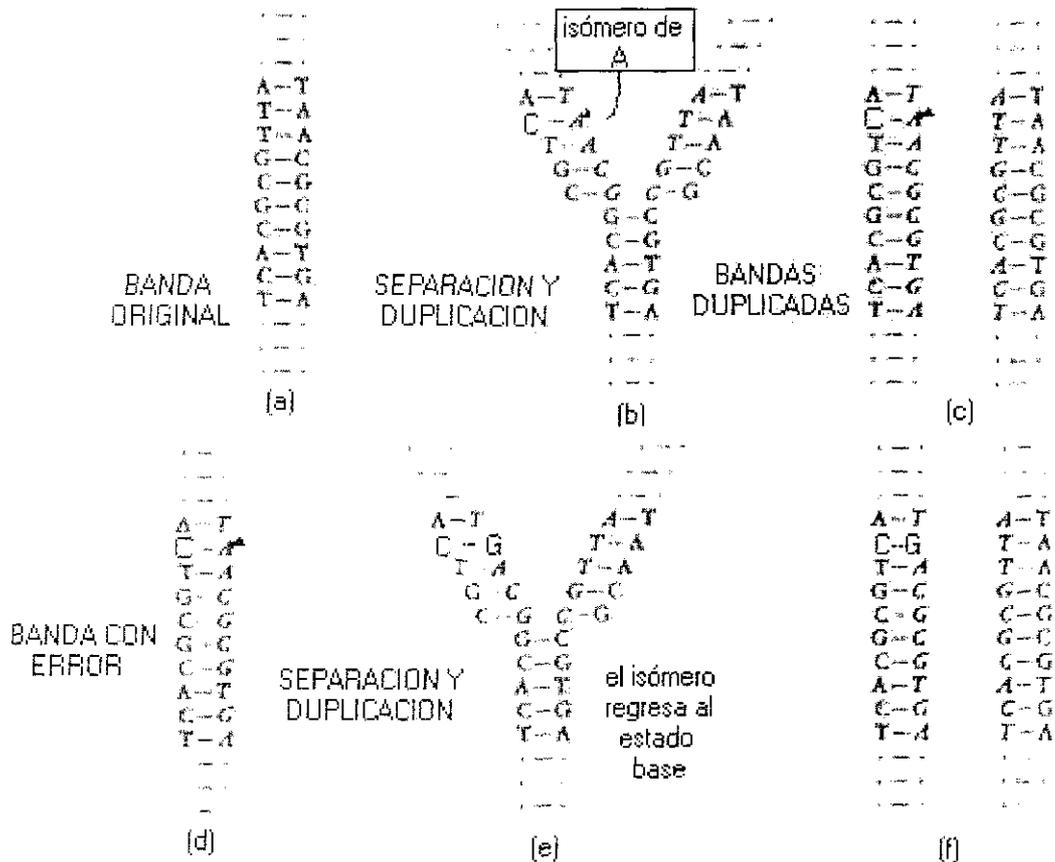


Figura 5.2: Se muestra como el isómero de la adenina altera el código original

no naturales. Por ejemplo, hay pruebas de que pueden surgir mutaciones en un ADN de cadena doble por exposición a los rayos X, gamma, ultravioleta, cósmicos y a algunos agentes químicos como el ácido nitroso; tales agentes alteran tanto a la secuencia de bases (quitando una base) como a la base misma (figura 5.3). Así, las mutaciones pueden considerarse como resultado, por lo menos en algunos casos, de cambios en la secuencia de nucleótidos del ADN. La radiación ionizante incluye a las longitudes de onda más cortas de

TESIS CON
FALLA DE ORIGEN

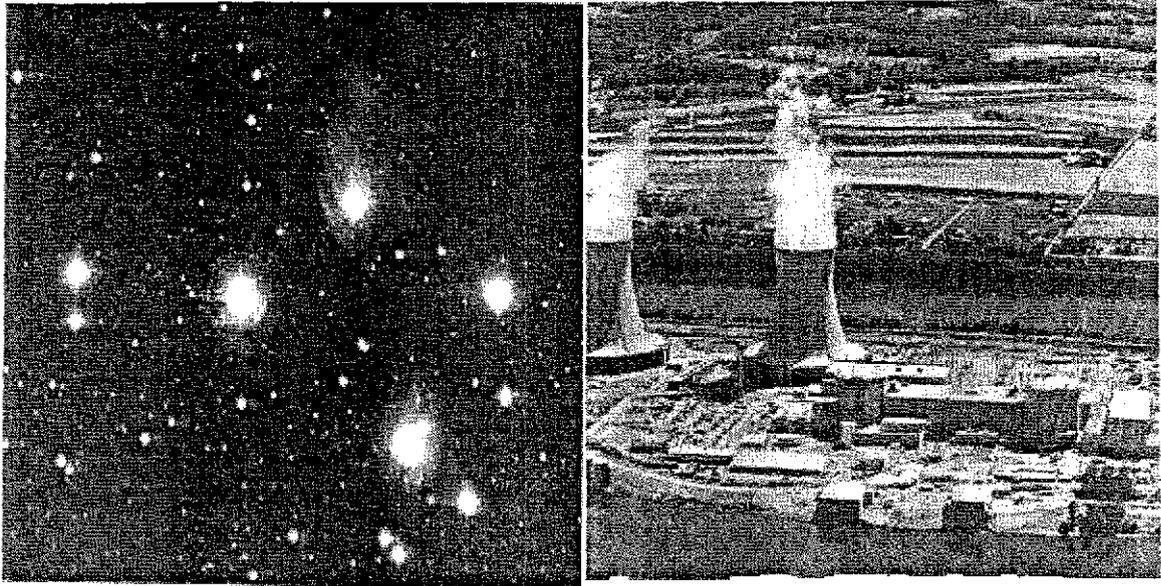
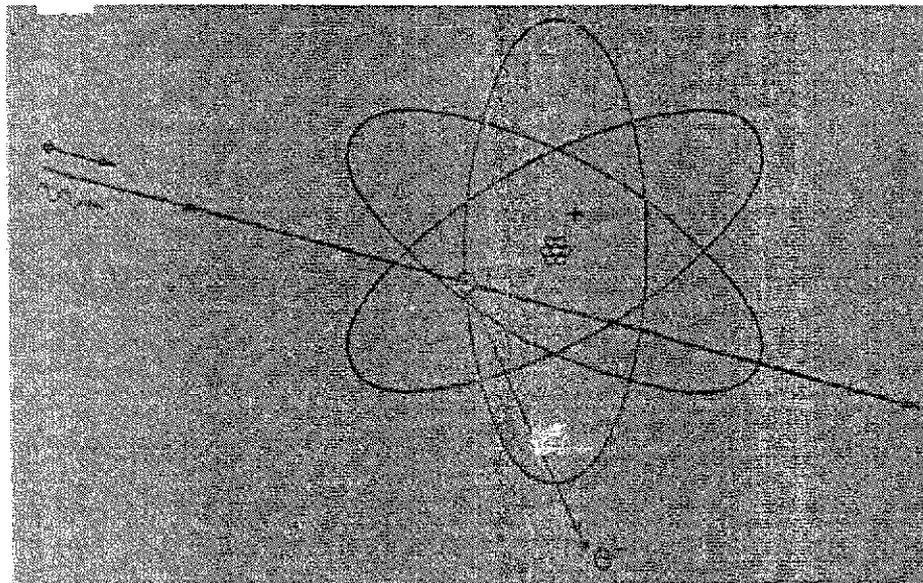
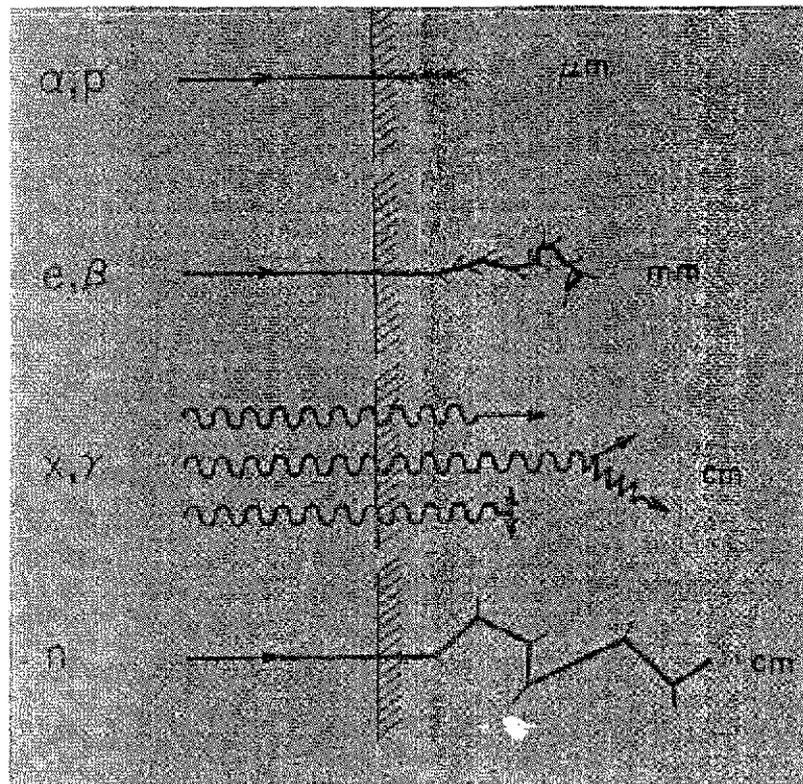


Figura 5.3: La radiación cósmica que llega del espacio así como fuentes generadas por el hombre pueden originar mutaciones y cambios en la estructura de las bases.

la energía radiante, y, por ello, las de mayor energía. Estas ondas de energía pueden sacar los electrones de los átomos de las sustancias a través de las cuales pasan (figura 5.4). La pérdida de un electrón crea un átomo cargado, un ión, y éste es sumamente reactivo químicamente. Mediante la actividad química de los iones, las moléculas de una célula viva pueden alterarse: las proteínas, especialmente las enzimas, pueden quedar inactivas, y las bases del ADN se pueden sustituir, suprimir y reordenar. La pérdida de muchas bases en el ADN deteriora la capacidad de la célula para sintetizar proteínas y produce errores en la codificación. Las implicaciones de esto tienen largo alcance, y los efectos pueden llegar a ser mortales. De las tablas que muestran la diferencia en energía potencial entre el primer isómero y el estado base vemos que para la adenina tiene un valor de 110.304 kcal/mol y para la guanina el valor corresponde a 144.399 kcal/mol. Si tomamos un valor medio de 120 kcal/mol = 5.206 eV, esta energía corresponde a fotones con frecuencia igual a $E/h = (5.206\text{eV})(1.602 \times 10^{-19} \frac{\text{J}}{\text{eV}})/(6.6256 \times 10^{-34} \text{J} \cdot \text{s}) = 1.258 \times 10^{15} \text{ Hz}$, un fotón con longitud de onda igual a $7.949 \times 10^{-8} \text{ m}$, correspondiente a la región del espectro electromagnético entre rayos ultravioleta y rayos X.

TESIS CON
FALTA DE ORIGEN



TESIS CON
FALLA DE ORIGEN

Figura 5.4: Diversos tipos de radiación producen efectos dañinos para los seres vivos, algunas de estas radiaciones pueden ser ionizantes

Estas nuevas conformaciones también podrían conducir a situaciones que no se presentan en la forma natural. Por ejemplo, en la molécula del ATP la energía por el desprendimiento de sus grupos fosfato no se localiza exclusivamente en el enlace; está distribuida en toda la molécula, pero se libera precisamente al romper el enlace. El caso de la liberación de energía por la hidrólisis de los dos últimos fosfatos puede deberse, por una parte, a que el ATP o el ADP son polielectrólitos, moléculas con alta repulsión electrostática (4 o 3 cargas negativas, respectivamente); por lo tanto, para sostenerse unidos como tales requieren una cantidad de energía interna elevada, la cual depende de su estructura espacial. Al hidrolizarse la molécula esta energía se libera, disminuyendo así la repulsión electrostática.

5.2 Efectos en mutaciones

Estos nuevos isómeros de las bases traerían trascendentales consecuencias como mutaciones, se denomina mutación a la aparición accidental de un carácter nuevo que se transmite a la descendencia. Las mutaciones silenciosas no ejercen efectos visibles en la proteína debido a que el cambio en el codón produce un aminoácido similar o químicamente relacionado. Las mutaciones puntuales dan origen a proteínas de la misma longitud que las proteínas normales. Estas proteínas mutadas son funcionalmente normales o no tienen actividad, según el tipo de aminoácido sustituido y de su localización en la cadena polipeptídica. Las mutaciones sin sentido, provocadas por una conversión de un codón que especifica un aminoácido en un codón terminal, provoca la formación de una proteína truncada, que suele no funcionar (figura 5.5).

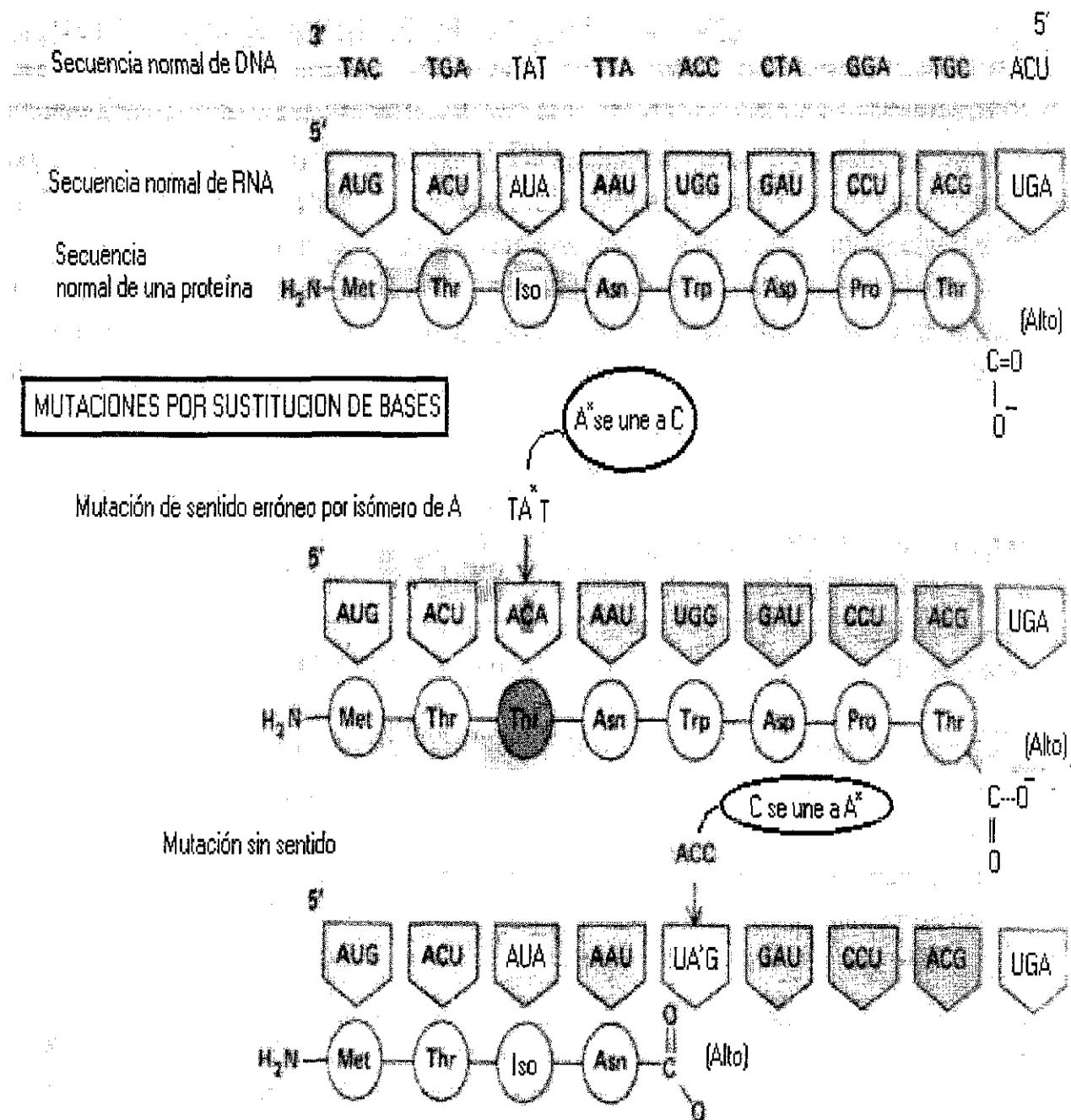


Figura 5.5: Efectos de las mutaciones que afectan a un solo par de bases por nuevos isómeros. Los aminoácidos en las figura corresponden a Met:Metionina, Thr:Treonina, Iso:Isoleucina, Asn:Asparagina, Trp:Triptófano, Asp:Asparagina, Pro:Prolina (ver figura 5.7).

TESIS CON
FALTA DE ORIGEN

MUTACIONES POR IRRECONOCIMIENTO DE ISÓMEROS DE BASES CON CONSECUENCIAS TALES COMO UN CORRIMIENTO DE LA PAUTA DE LECTURA CROMOSÓMICA

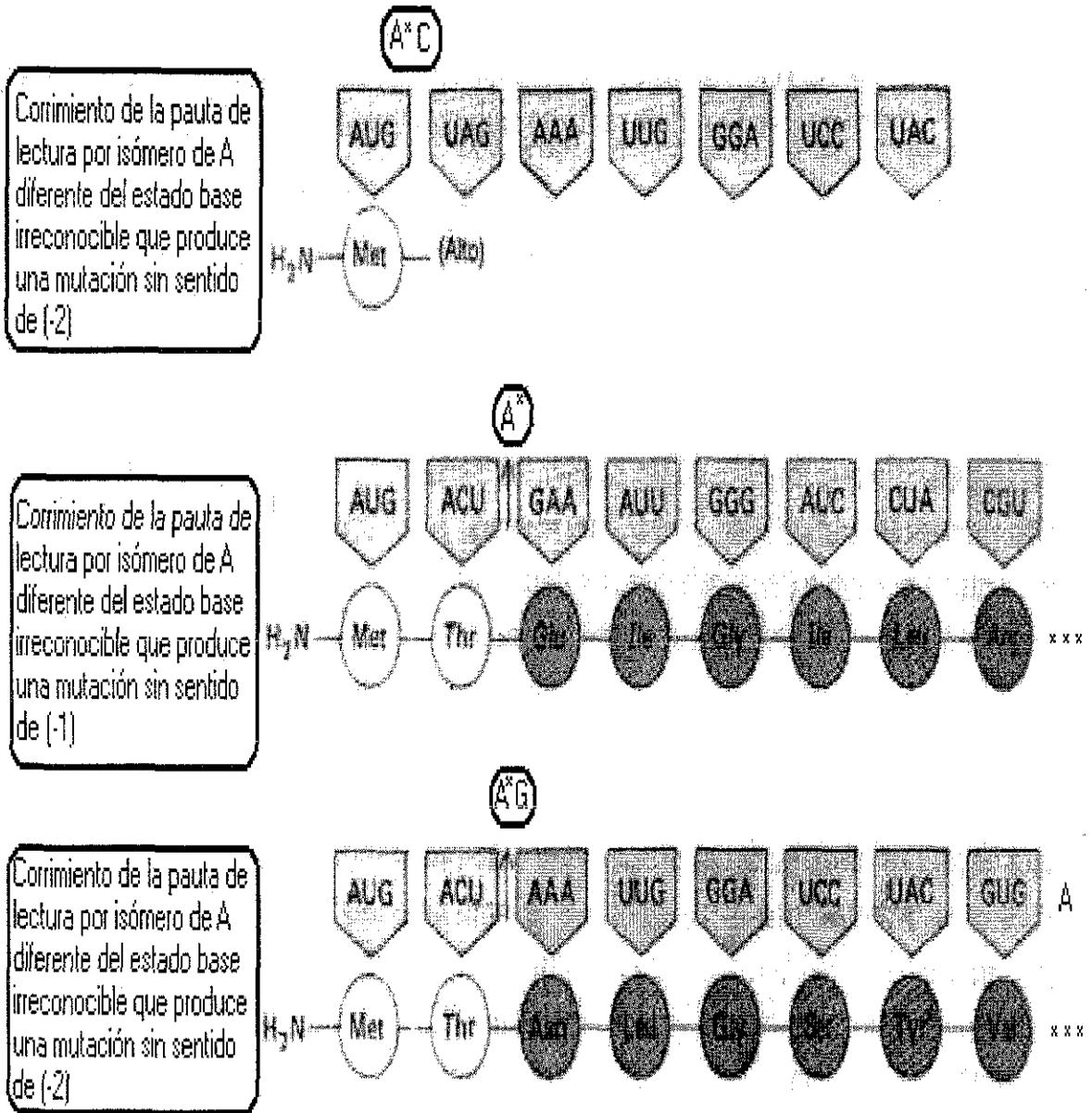


Figura 5.6: Mutaciones por irreconocimiento de isómeros de las bases. Met:Metionina, Thr:Treonina, Glu: Acido glutámico, Ile: Isoleucina, Gly: Glicina, IIs: Isoleucina, Leu: Isoleucina, Arg: Arginina, Leu: Leucina, Gly: Glicina, Ser: Serina, Tyr: Tirosina, Val: Valina (ver figura 5.7).

TESIS CON FALTA DE ORIGEN

		Segunda letra en el triplete de RNAm				Tercera letra en el triplete de RNAm						
		C		A		G						
U		serina		tirosina codón de terminación		cisteína codón de terminación triptofano						
C		prolina		histidina glutamina		arginina						
A		treonina		asparagina lisina		serina arginina						
G		alanina		ácido aspártico ácido glutámico		glicina						
U	UUU } fenilalanina UUC } UUA } leucina UUG }	UCU } UCC } UCA } UCG }	UAU } tirosina UAC } UAA } codón de UAG } terminación	UGU } cisteína UGC } UGA } codón de UGG } terminación triptofano	U	U	C	C	A	A	G	G
C	CUU } leucina CUC } CUA } CUG }	CCU } CCC } CCA } CCG }	CAU } histidina CAC } CAA } glutamina CAG }	CGU } CGC } CGA } CGG }	U	U	C	C	A	A	G	G
A	AUU } isoleucina AUC } AUA } metionina AUG }	ACU } ACC } ACA } ACG }	AAU } asparagina AAC } AAA } lisina AAG }	AGU } serina AGC } AGA } arginina AGG }	U	U	C	C	A	A	G	G
G	GUU } valina GUC } GUA } GUG }	GCU } GCC } GCA } GCG }	GAU } ácido aspártico GAC } GAA } ácido glutámico GAG }	GGU } GGC } GGA } GGG }	U	U	C	C	A	A	G	G

Figura 5.7: Secuencia de los nucleótidos en los tripletes de codones del mRNA que especifican un aminoácido.

TESIS CON
FALLA DE ORIGEN

Las mutaciones cromosómicas en la estructura que resultan de la delación o inserción de una o dos de las bases por irreconocimiento de nuevos isómeros, tienen efectos más drásticos; hacen que las secuencias de bases que sigue a la mutación cambie a una nueva franja de lectura, alterando la estructura y función de la proteína. Las mutaciones cromosómicas también producen un codón de terminación después de la mutación, que tiene el mismo efecto que una mutación sin sentido creada por una sustitución de bases (figura 5.6).

Una investigación futura es la factibilidad de estos pares de bases en soluciones acuosas, ya que la influencia de la hidratación sobre el equilibrio isomérico de las bases es un hecho experimental conocido [146]. El caso de los tautómeros keto-enol de la guanina es muy interesante ya que los tautómeros keto y enol existen en aproximadamente iguales concentraciones cuando estos están aislados en una matriz de gas inerte [147] [148], sin embargo, en soluciones acuosas la relación de concentración keto-enol es del orden de 10^4 a 10^5 [150], a pesar de la casi igual estabilidad intrínseca de las dos conformaciones, se ha encontrado que esto es debido a las diferencias en energía de hidratación de -5.6 y -6.8 kcal/mol entre los tautómeros keto y enol [146]. Estos resultados proveen una evidencia del cambio de conformación de una base en una solución acuosa. Así como los errores tipográficos rara vez mejoran un mensaje impreso sensato, los cambios al azar en la secuencia de las bases rara vez son favorables. Muchas mutaciones son letales, y por el mismo caso no se perpetúan. Si no lo son, su naturaleza favorable o desfavorable depende del medio ambiente y de la constitución genética en que el gen funcione (figuras 5.8 y 5.9).

Si se comprobara la existencia de estas configuraciones, resultados interesantes se presentarían al reaccionar con sustancias que actúan sobre estas [154], por ejemplo, el ácido nitroso reacciona sobre la adenina y la transforma en hipoxantina; esta última en un apareamiento, actúa como si fuera guanina y la siguiente vez que se aparca, se puede acoplar a una citosina, dislocando así por completo el sistema normal. La fluorodesoxiuridina altera la base al grado de que no puede entrar al ADN. Un grupo de colorantes

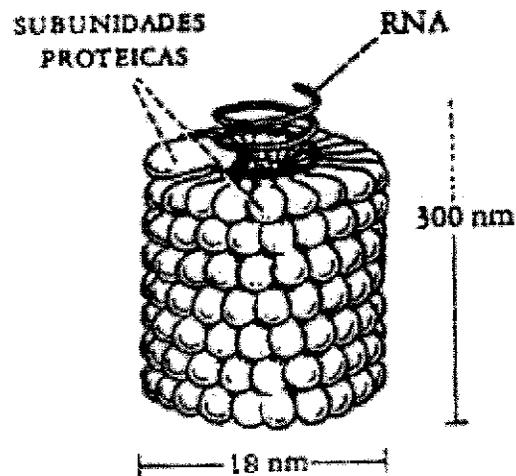


Figura 5.8: En seres simples como algunos virus una simple pérdida en la secuencia de la bases es fatal.

como el naranja de acridina, el bromuro de etidio y otros, producen mutaciones y pueden ser cancerígenos, se intercalan entre dos bases del ADN y las separan físicamente, con lo cual, durante la duplicación del ADN, se favorece la inclusión de una nueva base “complementaria” al colorante.

TESIS CON
FALLA DE ORIGEN

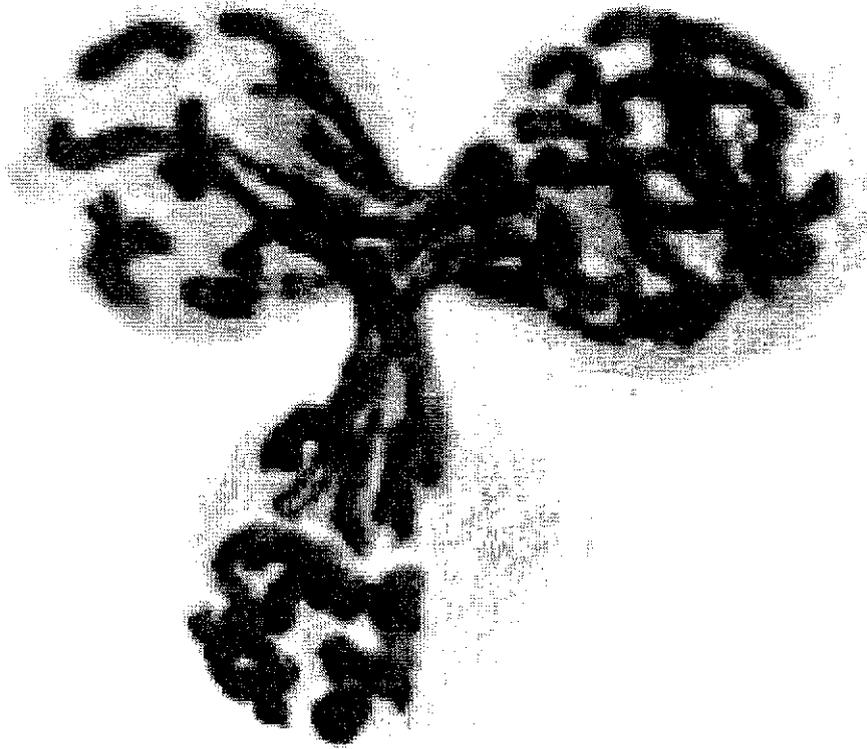


Figura 5 9: La célula arriba mostrada, procedente de una antera de un lirio, fué irradiada artificialmente con rayos X, con objeto a inducirle a que se dividiera en tres, y no en dos, células hijas. Tales mutantes no persisten; tan sólo cuando el cambio representa una ventaja para el nuevo ser -lo que indudablemente ocurre muy raras veces en una generación, pero, en cambio, con frecuencia en el curso de millones de años-, la selección natural puede intervenir con el fin de intentar el desarrollo de un organismo más eficiente.

5.3 Mutación + selección = evolución

La evolución es una modificación adaptativa que tiene lugar en el curso de muchas generaciones y de largos periodos. Cuando ocurren nuevos nacimientos por reproducción sexual, surge la variedad en los organismos, debido a la redistribución del ADN hereditario en los cromosomas (proceso que es controlado por las propias moléculas de ADN).

Parce ser que las mutaciones son el fenómeno promotor de la evolución, sobre el que actúa la selección natural en las poblaciones de cada especie: la

TESIS CON
FALLA DE ORIGEN

existencia de mutantes hace que no todos los individuos de una población sean iguales; los dotados con un conjunto de genes más favorable a la vida en un determinado medio serán los que más se reproducirán; su éxito aumentará la adecuación de la especie a dicho medio, si bien posteriores cambios de este (cambios climáticos, llegada de otras especies) producirán una nueva selección de individuos (figura 5.10).



Figura 5.10: En las áridas llanuras del oeste de Madagascar, un árbol, estrechamente emparentado con el baobab, debe sobrevivir durante los largos períodos de sequía, comprendidos entre las estaciones lluviosas. Para ello, ha desarrollado troncos voluminosos de madera ligera y esponjosa, en la que se almacena el agua. No muestra hojas durante las épocas de sequía, con el fin de reducir al máximo la evaporación .

TESIS CON
FALLA DE ORIGEN

5.4 Conclusiones

Hemos realizado una búsqueda de isómeros estables de las cuatro bases del ADN; adenina, guanina, citosina y timina. Con estas 4 moléculas se codifica toda la información de la actividad celular, el mecanismo de herencia y también son componentes de otras moléculas importantes en los seres vivos.

Primero buscamos los mínimos en la superficie de un potencial semiempírico, propuesto para las bases nucleótidas y aminoácidos. La búsqueda se llevó a cabo utilizando un algoritmo genético, el cual es un método de optimización eficiente y global. Se calculó el Hessiano y solo aquellas conformaciones que presentaron eigenvalores positivos fueron tomadas en cuenta posteriormente.

Con esta búsqueda global un gran número de configuraciones de mínima energía fueron encontrados, tanto planares como no planares. Se hicieron 13000 corridas del algoritmo genético para cada base, las frecuencias encontradas de las configuraciones mínimas en estas corridas son tabuladas y relacionadas con el ancho de la cuenca de atracción. Para las cuatro bases la conformación con más baja energía y que se encontró un mayor número de veces correspondió al estado natural ya conocido.

Aquellas configuraciones similares a la configuración natural y de baja energía, así como otras que parecían ser factibles su existencia fueron posteriormente optimizadas con aproximaciones Hartree-Fock y MP2 utilizando el programa Gaussian98. Se seleccionaron más de 10 conformaciones por base.

Se encontraron diversos isómeros estables con relajación cuántica para las bases de doble anillo, adenina y guanina. Para los isómeros de un solo anillo durante la búsqueda con algoritmos genéticos se encontraron diversos isómeros distintos de los naturales, pero todos convergieron a la conformación natural después de que se les aplicó métodos de optimización cuánticos. Para confirmar que las configuraciones fueran mínimas se calcularon las frecuencias de sus modos normales de vibración y sólo aquellas conformaciones que tuvieran todas sus frecuencias positivas fueron presentadas.

Se mostró una conformación que muestra una posible unión entre un isómero de la base adenina con la molécula citosina, la conformación pre-

sentada podría tener implicaciones importantes en efectos de mutación y de transcripción. además de efectos en otras moléculas de la célula que utilizan las bases, la configuración presentada es estable. El comprobar estos resultados experimentalmente sería muy interesante. Con este trabajo continuamos con una investigación empezada en licenciatura acerca de la posible existencia de isómeros de las bases del ADN

Capítulo 6

APÉNDICES

6.1 Apéndice A: Glosario breve de términos.

- *calor de formación*: Cambio de energía cuando un mol de una sustancia se forma a partir de sus elementos, el término se utiliza frecuentemente para referirse a la energía interna (o sea, la energía cinética y potencial total de las partículas).

- *celda unitaria*: Se dice del grupo mas pequeño de átomos, iones o moléculas que al repetirse a intervalos regulares en 3 dimensiones, producen una red de un sistema cristalino. Hay 7 tipos básicos de celdas unitarias que forman los 7 sistemas cristalinos.

- *conjunto base en Gaussian98*: La primera elección en un cálculo *ab-initio* es el llamado nivel de teoría, nosotros realizamos primeros cálculos HF y esto junto con la elección del conjunto base de orbitales escogido define el nivel de teoría. Así, una elección de un conjunto base *STO/6-31G** junto con HF significa un nivel de teoría en el lenguaje Gaussian como *HF/6-31G**.

- *energía de conformación*: Es bien conocido que la estructura de anillo de moléculas de tamaño medio requiere el doblamiento del ángulo de enlace entre 3 átomos A-B-C y la torsión de un enlace entre 2 átomos de sus geometrías de mínima energía. la energía de conformación es una función de estas distorsiones.

- *fuerzas de Van der Waals*: Fuerza de atracción intermolecular, considerablemente más débil que los enlaces químicos, que se origina de interacciones

electrostáticas entre moléculas (las energías son generalmente menores a 1 J/mol). La interacción de Van der Waals tiene contribuciones de 3 efectos: dipolo permanente, interacciones dipolares que se encuentran en moléculas polares; interacciones dipolares de dipolo inducido, donde un dipolo origina una ligera separación de carga en los enlaces que tienen alta polaridad y fuerzas dispersión que resultan de la polaridad temporal que se origina de la distribución asimétrica de electrones alrededor del núcleo.

- *hidrólisis*: La unidad de energía con la que trabaja Gaussian98 es el Hartree. El programa itera hasta que los orbitales y las energías son consistentes. El programa calcula la energía cinética, la energía potencial one-electron y la energía “total”, incluyendo a la repulsión nuclear.

- *hidrólisis*: reacción entre un compuesto y agua.

- *genotipo, fenotipo*: La característica física (aspecto) de un ser vivo se llama fenotipo (del griego “phainein” = aparecer) y se refiere a la forma en que se expresa una característica; esto es, a lo que puede verse, los ojos azules y cafés son diferentes genotipos. Además del fenotipo, tenemos también un genotipo, esto es, una serie particular de instrucciones genéticas heredadas de los padres.

- *método de Newton-Raphson*: En el método multidimensional de Newton se comienza de un punto $\mathbf{x}^{(1)}$ en la superficie de energía potencial, y se encuentra un nuevo punto $\mathbf{x}^{(2)}$ dado por

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - H^{-1}\mathbf{g}$$

donde H^{-1} es la inversa de la matriz hessiana, calculada en el punto actual $\mathbf{x}^{(1)}$ y \mathbf{g} es el vector gradiente, calculado también en el punto $\mathbf{x}^{(1)}$. El proceso se repite hasta que se encuentra el extremo.

- *método de optimización de Berny*: El procedimiento principal de optimización de geometría en Gaussian98 está basado en un programa escrito por H.B.(“Berny”) Schlegel. El método es sofisticado, pero en esencia utiliza el gradiente analítico y la matriz hessiana, la matriz G^{-1} es constantemente actualizada durante la optimización

- *potencial semiempírico*: El potencial AMBER es definido como semiempírico porque sus parámetros son derivados en parte de datos experimentales y en parte de cálculos *ab initio*. Sin embargo, el significado aceptado de semiempírico se refiere a métodos donde no todas las integrales necesarias para resolver la ecuación de Schroedinger son calculadas; algunas de ellas son despreciadas o estimadas en base a parámetros empíricos.

- *reducción del número de grados de libertad*: La manera más fácil de estudiar el movimiento de un cuerpo rígido es utilizar un sistema de referencia fijo al centro de masas y compuesto por los ejes principales de inercia, así 6 grados de libertad valdrían cero, si la molécula fuera lineal necesitaríamos 2 ejes para fijar la orientación con lo que 5 grados de libertad valdrían cero.

- *solvatación*: Atracción que ejerce un ión en solución hacia las moléculas de solvente. en el agua, por ejemplo, un ión positivo está rodeado por moléculas de agua las cuales tienden a asociarse alrededor del ión por la atracción entre la carga positiva del ión y la parte negativa de la molécula polar del agua.

- *superficie de Born-Oppenheimer*: Born y Oppenheimer (1927) demostraron que, en una muy buena aproximación, los movimientos de los núcleos y los electrones pueden ser separados. En la aproximación de Born-Oppenheimer se consideran a los núcleos como si estuvieran fijos para el propósito de calcular la energía y función de onda electrónica.

- *tautómero*: se dice generalmente cuando en una molécula ocurre la migración de un átomo de hidrógeno

6.2 Apéndice B: descripción del ajuste por mínimos cuadrados para la obtención de los parámetros del potencial utilizado

Los parámetros de las funciones de energía fueron variados iterativamente para que gradualmente se reduzca la suma de los cuadrados de las diferencias entre las cantidades experimentales y las calculadas, hasta que esta alcance un mínimo. El procedimiento desarrollado fue: Sea

$$\mathbf{x} = \mathbf{x}_m \quad (a1)$$

vector cuyos componentes son los parámetros de las diversas funciones de energía que componen la energía conformacional total. Sea

$$\Delta \mathbf{y} = \mathbf{y}_{\text{calc}} - \mathbf{y}_{\text{exptl}} \quad (a2)$$

Un vector cuyos componentes son las diferencias entre los valores calculados teóricamente y los observados experimentalmente. Este vector diferencia incluye los términos:

Exceso de energía de conformación, coordenadas de configuración de equilibrio y frecuencias de vibración de todas las moléculas bajo consideración:

$$\Delta \mathbf{y} = \Delta(\Delta H^{(n)}), \Delta \mathbf{s}_0^{(n)}, \Delta \nu_k^{(n)} \quad (a3)$$

Donde el índice (n) denota las diversas moléculas empleadas. Las conformaciones de equilibrio son representadas aquí por coordenadas internas \mathbf{s}_0 , para asegurar una correspondencia uno a uno entre las coordenadas experimentales y las calculadas. Sin embargo, todos los cálculos de las diversas propiedades moleculares fueron desarrollados en coordenadas cartesianas por conveniencia y para ahorrar tiempo de CPU. Los resultados finales fueron transformados entonces a coordenadas internas. Para pequeños cambios, hay una relación lineal entre $\Delta \mathbf{y}$ y $\Delta \mathbf{x}$:

$$\Delta \mathbf{y}(\mathbf{x} + \delta \mathbf{x}) = \Delta \mathbf{y}(\mathbf{x}) + \mathbf{Z} \delta \mathbf{x} + \dots \quad (a4)$$

donde \mathbf{Z} es la matriz con elementos

$$Z_{lm} = \frac{\partial y_l}{\partial x_m} \quad (a5)$$

con valores

$$\frac{\partial \Delta H^{(n)}}{\partial x_m} \quad (a6)$$

$$\frac{\partial s_0^{(n)}}{\partial x_m} \quad (a7)$$

$$\frac{\partial \nu^{(n)}}{\partial x_m} \quad (a8)$$

se mejora $\delta \mathbf{x}$ si se reduce el valor de

$$[(\Delta \mathbf{y}^T) \mathbf{P}] [\mathbf{P} \Delta \mathbf{y}] \quad (a9)$$

\mathbf{P} es una matriz de factores de peso, se minimiza la expresión anterior sujeta a la condición de que $\delta \mathbf{x}$ es pequeño o equivalentemente que

$$(\delta \mathbf{x}^T)(\delta \mathbf{x}) \quad (a10)$$

sea una constante pequeña. $\delta \mathbf{x}$ es una solución de

$$(\partial / \partial \delta x_m)(\Delta \mathbf{y}^T \mathbf{P}^2 \Delta \mathbf{y} + \xi \delta \mathbf{x}^T \delta \mathbf{x}) = 0 \quad (a11)$$

donde ξ es un multiplicador de Lagrange. Las ecuaciones (a4) y (a11) dan

$$\mathbf{Z}^T \mathbf{P}^2 (\Delta \mathbf{y}(\mathbf{x}) + \mathbf{Z} \delta \mathbf{x}) + \xi \delta \mathbf{x} = \mathbf{0} \quad (a12)$$

entonces

$$\delta \mathbf{x} = -(\mathbf{Z}^T \mathbf{P}^2 \mathbf{Z} + \xi \mathbf{I})^{-1} \mathbf{Z}^T \mathbf{P}^2 \Delta \mathbf{y}(\mathbf{x}) \quad (a13)$$

da el requerido mejoramiento lineal de \mathbf{x} , con tal de que ξ sea suficientemente grande, correspondiendo a valores suficientemente pequeños de $\delta \mathbf{x}^T \delta \mathbf{x}$. Este procedimiento es iterativo,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta \mathbf{x}^{(k)} \quad (a14)$$

hasta que \mathbf{x} converja a valores óptimos de los parámetros de la energía.

6.2.1 Cálculo de los elementos de matriz Z_{lm}

Los renglones de la matriz $Z_{lm} = \partial y_l / \partial x_m$ son de tres clases, $\partial \Delta H^{(n)} / \partial x_m$, $\partial \mathbf{s}_0^{(n)} / \partial \mathbf{x}_m$, y $\partial \nu_\alpha^{(n)} / \partial x_m$. Ellos son calculados como sigue:

(1) La derivada del exceso de entalpía es:

$$\partial \Delta H^{(n)} / \partial x_m = (\partial \Delta V^{(n)} / \partial x_m) + (\partial \Delta H_{vrt}^{(n)} / \partial x_m) \quad (a15)$$

El primer término se deriva directamente de la energía potencial como función explícita de \mathbf{x} . El segundo término se ha encontrado que es relativamente insensible a cambios en \mathbf{x} y es por lo tanto despreciable. Notemos, sin embargo, que ΔH_{vrt} sigue siendo parte del exceso de entalpía calculada en los componentes de el vector $\Delta \mathbf{y}$.

(2) Las derivadas de las coordenadas de equilibrio con respecto a x_m son primero obtenidas en coordenadas cartesianas, esto es, $\partial \mathbf{r}_0 / \partial x_m$, y entonces transformadas a coordenadas internas por la matriz de transformación D.

$$\delta \mathbf{s} = D \delta \mathbf{r}, \quad D_{\alpha\beta} = \partial \mathbf{s}_\alpha / \partial \mathbf{r}_\beta \quad (a16)$$

entonces

$$\partial \mathbf{s}_0 / \partial x_m = D \partial \mathbf{r}_0 / \partial x_m \quad (a17)$$

Las variables de equilibrio \mathbf{r}_0 no están dadas como funciones explícitas de \mathbf{x} , si no que están dadas por la ecuación $\partial V(\mathbf{r} : \mathbf{x}) / \partial r_\alpha = 0$. Para obtener $\partial \mathbf{r}_0 / \partial x_m$ consideremos su definición

$$\frac{\partial \mathbf{r}_0}{\partial x_m} = \lim_{\delta x_m \rightarrow 0} \frac{\mathbf{r}_0(\mathbf{x} + \delta \mathbf{x}_m) - \mathbf{r}_0(\mathbf{x})}{\delta x_m} \quad (a18)$$

donde $\delta \mathbf{x}_m$ es un vector $\delta \mathbf{x}$ con δx_m como la única componente distinta de cero, $\mathbf{r}_0(\mathbf{x} + \delta \mathbf{x}_m)$ son las coordenadas de equilibrio pertenecientes a $\mathbf{x} + \delta \mathbf{x}_m$, y $\mathbf{r}_0(\mathbf{x})$ son las coordenadas de equilibrio pertenecientes a \mathbf{x} . Ahora consideremos la solución de $\partial V(\mathbf{r}, \mathbf{x} + \delta \mathbf{x}_m) / \partial r_\alpha = 0$ por la ec. (a4),

$$\mathbf{r}_0(\mathbf{x} + \delta \mathbf{x}_m) = \mathbf{r}(\mathbf{x} + \delta \mathbf{x}_m) - \mathbb{F}^{-1}(\mathbf{r}, \mathbf{x} + \delta \mathbf{x}_m) \nabla V(\mathbf{r}, \mathbf{x} + \delta \mathbf{x}_m) \quad (a19)$$

Si escogemos $\mathbf{r}(\mathbf{x} + \delta \mathbf{x}_m)$ de tal forma que

$$\mathbf{r}(\mathbf{x} + \delta \mathbf{x}_m) = \mathbf{r}_0(\mathbf{x}) \quad (a20)$$

De las dos ecuaciones anteriores se sigue que:

$$\frac{\partial \mathbf{r}_0}{\partial x_m} = \lim_{\delta x_m \rightarrow 0} \frac{-F^{-1}[\mathbf{r}_0(\mathbf{x}); \mathbf{x} + \delta \mathbf{x}_m] \nabla V[\mathbf{r}_0(\mathbf{x}); \mathbf{x} + \delta \mathbf{x}_m]}{\delta x_m} \quad (a21)$$

este límite es igual a

$$-F^{-1}(\mathbf{r}_0; \mathbf{x}) \partial \nabla V(\mathbf{r}_0; \mathbf{x}) / \partial x_m \quad (a22)$$

Las expresiones F^{-1} y F han sido ya usadas en la derivación de \mathbf{r}_0 y ν_α , y $\partial \nabla V / \partial x_m$ se deriva de las expresiones analíticas de el gradiente ∇V como función explícita de \mathbf{x}

(3) La ecuación secular para las vibraciones de modos normales es:

$$\mathcal{F} \delta \mathbf{q} = \lambda \mathbf{q} \quad (a23)$$

donde $\delta \mathbf{q} = \mathbf{M}^{1/2} \delta \mathbf{r}$, y $\mathcal{F} = M^{1/2} F M^{1/2}$. Entre los $3m$ eigenvalores (donde m es el número de átomos) $3m - 6$ son $\lambda_\alpha = (2\pi\nu_\alpha)^2$, ν_α siendo las frecuencias normales. Los restantes 6 eigenvalores son cero. Considerando las frecuencias de modos normales como funciones de las constantes de fuerza y usando el método de mínimos cuadrados para obtener los valores óptimos de las constantes de fuerza. De la ecuación anterior se obtiene que:

$$\frac{\partial \lambda_\alpha}{\partial x_m} = \delta \mathbf{q}_\alpha^T \left(\frac{\partial \mathcal{F}}{\partial \mathbf{x}_m} \right) \delta \mathbf{q}_\alpha \quad (a24)$$

donde \mathcal{F} no depende de las coordenadas de equilibrio. En el presente caso, sin embargo, \mathcal{F} es una función no lineal tanto de los parámetros de energía \mathbf{x} como de las coordenadas de equilibrio \mathbf{r}_0 , las cuales son funciones no lineales de \mathbf{x} :

$$\frac{\partial \lambda_\alpha}{\partial x_m} = \delta \mathbf{q}_\alpha^T \left[\left(\frac{\partial \mathcal{F}}{\partial x_m} \right)_{\mathbf{r}=\mathbf{r}_0} + \sum_i \frac{\partial \mathcal{F}}{\partial r_i} \frac{\partial r_i}{\partial x_m} \right] \delta \mathbf{q}_\alpha \quad (a25)$$

La evaluación de la tercera derivada $\partial^3 / \partial r_\alpha \partial r_\beta \partial r_\gamma$ en esta expresión ocuparía mucho tiempo de computo. En vez de esta, se uso la siguiente expresión

$$\frac{\partial \lambda_\alpha}{\partial x_m} = \delta \mathbf{q}_\alpha^T \left(\lim_{\delta x_m \rightarrow 0} \frac{(\mathcal{F}[\mathbf{x} + \delta \mathbf{x}_m; \mathbf{r}_0(\mathbf{x} + \delta \mathbf{x}_m)] - f(\mathbf{x}, \mathbf{r}_0))}{(\delta x_m)} \right) \delta \mathbf{q}_\alpha \quad (a26)$$

donde $\mathbf{r}_0(\mathbf{x} + \delta \mathbf{x}_m)$ es obtenida de las ecuaciones (a19)-(a22), y la ecuación última fué evaluada numéricamente usando la expresión analítica para \mathcal{F} .

6.2.2 Breve historia del desarrollo del potencial

Esta técnica de mínimos cuadrados ya ha sido la base para desarrollar diversas funciones de energía [107], cada función tenía sus propios parámetros para ajustarse a los mas diversos datos experimentales, entre los cuales se incluyen frecuencias vibracionales, geometrías de equilibrio, energías de conformación y términos energéticos de vibración-traslación-rotación entre otros. A. Warshel y S. Lifson fueron los primeros en aplicar este método de mínimos cuadrados en un conjunto de funciones de energía para los alcanos [108, 109]. a este conjunto de funciones de energía se le dió el nombre de campo de fuerza consistente, el término “consistente” describe el modo en que se construyó esta función de energía, es decir, por una variación iterativa de los parámetros hasta alcanzar la consistencia con los diferentes datos experimentales.

Este campo de fuerza en función de las coordenadas internas y de las distancias interatómicas fué extendido por A. Warshel y S. Lifson a la fase cristalina de varios hidrocarburos [108]; esta extensión abarcaba la evaluación con las funciones del potencial de los parámetros de la celda unitaria cristalina (componentes vectoriales de la celda unitaria), vibraciones de red, energías de vibración y calores de sublimación de las moléculas n-hexano y n-octano. Con esta extensión, se mejoró el campo de fuerza al abarcar interacciones tanto intra- como inter-moleculares.

Posteriormente A.T. Hagler, E. Huler y S. Lifson transfirieron este potencial para desarrollar una función de energía para péptidos y proteínas, este campo de fuerza incluía la descripción de puentes de hidrógeno [109]. Utilizando la idea original de Lifson y Warshel de ajuste por mínimos cuadrados, recientemente Weiner *et al*[110] desarrollaron un campo de fuerza para ácidos nucleicos y proteínas, esta función de energía fué la primera que se desarrolló para ácidos nucleicos y reproducía adecuadamente las estructuras, energías y frecuencias vibracionales de las bases del ADN y proteínas. Este campo de fuerza representaba los grupos CH , CH_2 y CH_3 como si se trataran de una sola identidad esférica. esta representación conducía a ajustes más pobres con los experimentos. Finalmente, para solucionar estos problemas Wein-

er *et al*[111] desarrollaron la función de energía presentada aquí, la cual ya incluye a todos los átomos.

6.3 Apéndice C: Métodos de optimización cuánticos

La jerarquía en los modelos de química cuántica se muestra a continuación:

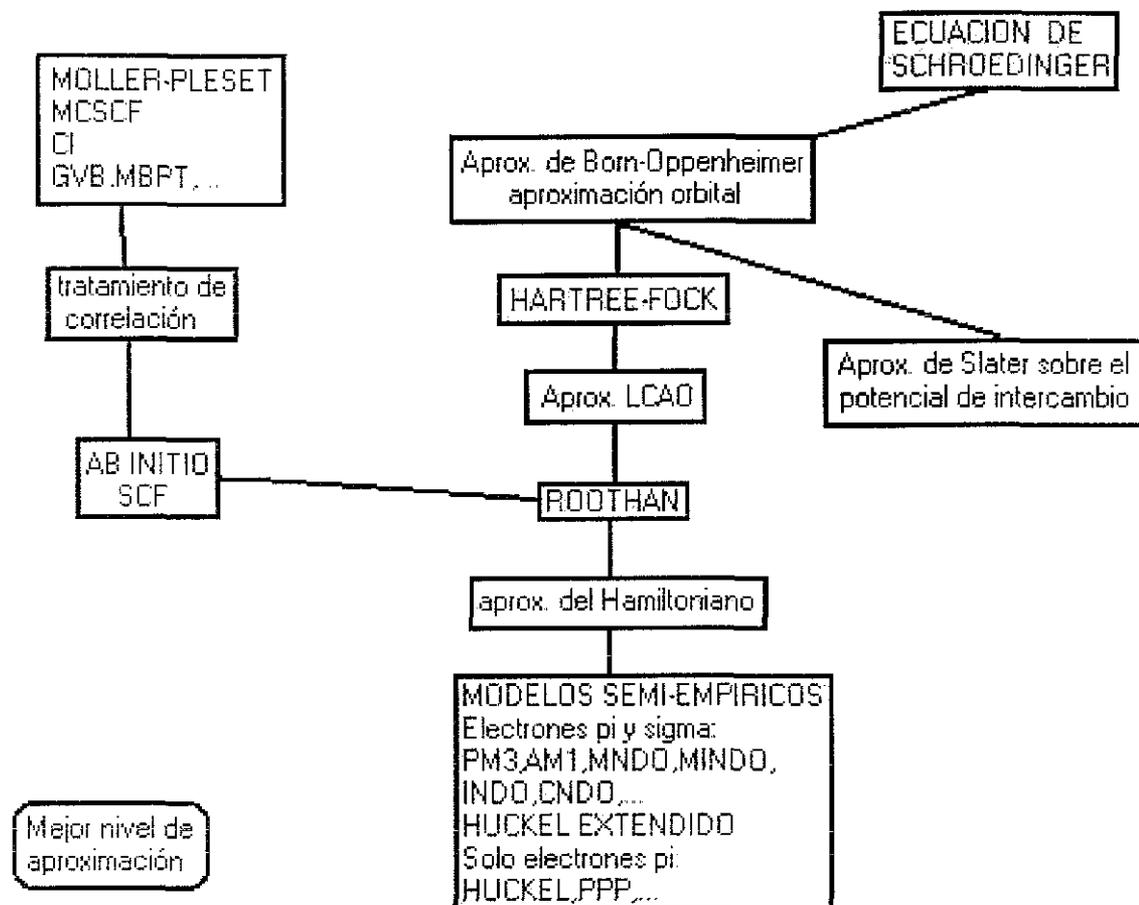


Figura 6.1: Hay 3 principales grupos: *Ab Initio* Hartree-Fock y procedimientos post-SCF, modelos semi-empíricos y *density-functional*.

6.3.1 Introducción

Para un sistema compuesto de $2M$ electrones y NUC núcleos, la ecuación de schrodinger independiente del tiempo es:

$$H\Psi = E\Psi$$

TESIS CON
FALLA DE ORIGEN

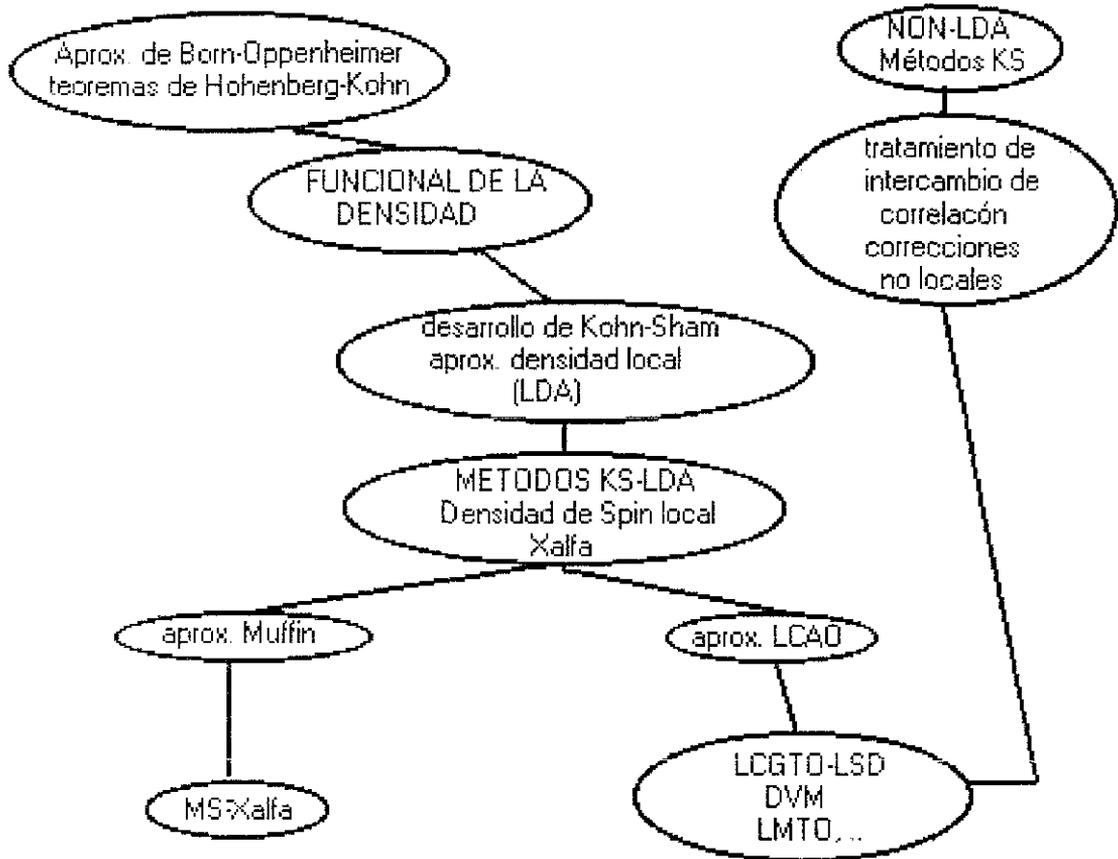


Figura 6.2: Procedimientos en base a la densidad atómica (*density-functional*).

la expresión de cada término es:

$$T_{el} = -\frac{\hbar^2}{8\pi^2m} \sum_{i=1}^{2M} \nabla_i^2$$

$$T_n = -\frac{\hbar^2}{8\pi^2} \sum_{\nu=1}^{NUC} \frac{1}{M_\nu} \nabla_\nu^2$$

$$V_{el,n} = -\sum_{i=1}^{2M} \sum_{\nu=1}^{NUC} \frac{Z_\nu e^2}{r_{i\nu}}$$

$$V_{el,el} = -\frac{1}{2} \sum_{i=1}^{2M} \sum_{j \neq i}^{2M} \frac{e^2}{r_{ij}}$$

$$V_{n,n} = -\frac{1}{2} \sum_{\delta=1}^{NUC} \sum_{\nu \neq \delta}^{NUC} \frac{Z_\nu Z_\delta e^2}{R_{\nu\delta}}$$

TESIS CON
FALLA DE ORIGEN

esta ecuación no tiene solución exacta, la primera aproximación utiliza el método de Born-Oppenheimer, el cual consiste en la separación de los movimientos nucleares y electrónicos

$$\Psi(r, R) = \Psi_{el}(r)\Psi_n(R)$$

La función de onda electrónica depende paramétricamente de las coordenadas de los núcleos y es la solución de la ecuación de Schrodinger electrónica:

$$H_{el}\Psi_{el}(r) = \epsilon\Psi_{el}(r)$$

Donde:

$$H_{el} = T_{el} + V_{el,n} + V_{el,el}$$

6.3.2 Campos Autoconsistentes

La esencia del modelo de Hartree-Fock (HF) o de Campo Auto Consistente (Self Consistent Field SCF) es resolver la ecuación de Schrodinger para un simple electrón moviéndose en un potencial donde los movimientos complicados de los restantes electrones son “promediados”. Estas interacciones no son ciertamente tomadas como cero, pero el modelo obviamente pierde los detalles mas finos causados por las interacciones instantáneas electrón-electrón. Esta idea física por lo tanto es simple y puede ser relacionada con la densidad electrónica. El significado físico de la función de densidad $\rho_1(\mathbf{x})$; es que el producto $\rho_1(\mathbf{x})d\tau ds$ nos da la oportunidad de encontrar cualquier electrón el el elemento de volumen espacio-spin $d\tau ds$ con los otros electrones en cualquier lugar y con cualquier spin. $P_1(\mathbf{x})d\tau$ da la correspondiente probabilidad de encontrar cualquier electrón con cualquier spin en el volumen espacial $d\tau$. Para conseguir esto hay diversos modos en que podemos proceder. El tradicional es buscar una ecuación de eigenvalores para los orbitales HF. Otro modo es concentrarse en la densidad electrónica misma y buscar métodos para encontrar $P_1(\mathbf{r})$ directo, sin recurrir a la función de onda. La vasta mayoría de las moléculas conocidas son orgánicas, sin simetría, la mayoría de las cuales tienen estados bases electrónicos singletes, los cuales pueden ser

escritos como configuraciones orbitales moleculares $\psi_A^2\psi_B^2\dots\psi_M^2$. En este caso se habla de M orbitales moleculares ocupados, y el número de electrones es $2M$, ya que en cada orbital molecular esta localizado un electron con spin α y otro con spin β . En el modelo de Hartree original, la función de onda para varios electrones es escrita como un producto de funciones de onda de un electrón

$$\Psi_{el}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{2M}) = \psi_A(\mathbf{r}_1)\alpha(s_1)\psi_A(\mathbf{r}_2)\beta(s_2)\dots\psi_M(\mathbf{r}_{2M})\alpha(s_{2M})$$

Este es el modelo de Hartree, debido a la “indistinguibilidad”, sabemos que el mas pequeño bloque lógico posible para tal sistema es:

$$\Psi_{el}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{2M}) = \begin{bmatrix} \psi_A(\mathbf{r}_1)\alpha(s_1) & \psi_A(\mathbf{r}_1)\beta(s_1)\dots & \psi_M(\mathbf{r}_1)\beta(s_1) \\ \psi_A(\mathbf{r}_{2M})\alpha(s_{2M}) & \psi_A(\mathbf{r}_{2M})\beta(s_{2M})\dots & \psi_M(\mathbf{r}_{2M})\beta(s_{2M}) \end{bmatrix}$$

donde hemos ocupado doblemente cada orbital SCF y se ha tomado en cuenta la antisimetría al escribir un determinante de Slater. Esto es conocido como el modelo de Hartree-Fock. Los terminos SCF y HF son intercambiables. Asumamos que el estado electrónico particular es el de más baja energía para la simetría dada y que las ψ 's son ortonormales. La primera asunción es vital, la segunda hace el álgebra un poco más fácil. Tradicionalmente lo que se tiene que hacer es minimizar la energía variacional

$$\epsilon_{el} = \frac{\int \Psi_{el} H_{el} \Psi_{el} d\tau}{\int \Psi_{el}^2 d\tau}$$

La integración es sobre las coordenadas de todos los electrones, y todas las Ψ 's se asumen que son cantidades reales en vez de complejas. La intención de la teoría de HF es encontrar la mejor forma posible de la función de onda electrónica, la cual es tomada como un producto de funciones de onda de un electrón. Cada electrón se mueve en un potencial promedio debido al núcleo y a los restantes electrones. El primer paso es trabajar la expresión de energía formal en términos de operadores de uno y dos electrones y de orbitales HF ψ_R . Para una molécula poliatómica con $2M$ electrones y NUC núcleos, el hamiltoniano electrónico es una suma de términos que representan:

◦La energía cinética de cada uno de los $2M$ electrones (los cuales se denotan por el subíndice i)

o La energía potencial mutua de cada núcleo (NUC en total, con subíndice α) con cada electrón.

o La energía potencial mutua de cada par de electrones.

En símbolos esto es:

$$H_{el} = -\frac{\hbar^2}{8\pi^2m} \sum_{i=1}^{2M} \nabla_i^2 - \frac{e^2}{4\pi\epsilon_0} \sum_{i=1}^{2M} \left(\sum_{\alpha=1}^{NUC} \frac{Z_\alpha}{R_{\alpha,i}} \right) + \frac{e^2}{4\pi\epsilon_0} \sum_{i=1}^{2M-1} \left(\sum_{j=i+1}^{2M} \frac{1}{r_{ij}} \right)$$

Se han agrupado los operadores por una razón. Normalmente se escribe el Hamiltoniano electrónico como una suma de operadores de uno y dos electrones.

$$H_{el} = \sum_{i=1}^{2M} h(\mathbf{r}_i) + \sum_{i=1}^{2M-1} \left[\sum_{j=i+1}^{2M} g(\mathbf{r}_i, \mathbf{r}_j) \right]$$

Ellos están definidos como:

$$h(\mathbf{r}_i) = -\frac{\hbar^2}{8\pi^2m} \nabla_i^2 - \frac{e^2}{4\pi\epsilon_0} \left(\sum_{\alpha=1}^{NUC} \frac{Z_\alpha}{R_{\alpha,i}} \right)$$

y

$$g(\mathbf{r}_i, \mathbf{r}_j) = \frac{e^2}{4\pi\epsilon_0} \frac{1}{r_{ij}}$$

La expresión de la energía puede ahora ser encontrada utilizando las reglas de Slater Condon Shortley. Estas se pueden encontrar en todos los textos clásicos tales como EWK. La idea es que la expresión de la energía la cual involucra las coordenadas de todos los electrones puede ser reducida a una suma mucho mas simple de términos involucrando las coordenadas de uno o dos electrones. La energía electrónica puede ser escrita términos de las ψ 's y los operadores de uno y dos electrones como

$$\begin{aligned} \epsilon_{el} = & 2 \sum_{R=A}^M \int \psi_R(\mathbf{r}_1) h(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau_1 + \\ & \sum_{R=A}^M \sum_{S=A}^M (2 \int \int \psi_R(\mathbf{r}_1) \psi_R(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \psi_S(\mathbf{r}_2) \psi_S(\mathbf{r}_2) d\tau_1 d\tau_2 - \\ & \int \int \psi_R(\mathbf{r}_1) \psi_S(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \psi_R(\mathbf{r}_2) \psi_S(\mathbf{r}_2) d\tau_1 d\tau_2 \end{aligned}$$

El primero de los términos de dos electrones tiene una significancia física clásica. Para ver esto, recordemos que $-e\psi_R^2(\mathbf{r}_1)$ es la densidad de carga

asociada con el electrón etiquetado como 1 y $-e\psi_S^2(\mathbf{r}_2)$ es la densidad de carga para el electrón 2. La integral es por lo tanto la energía potencial electrostática mutua del par de electrones. El segundo término de dos electrones es un poquito más difícil de entender; formalmente, este representa la energía potencial electrostática mutua del traslape de la distribución de carga $-e\psi_R(\mathbf{r}_1)\psi_S(\mathbf{r}_1)$ debido al electrón 1 con un término idéntico $-e\psi_R(\mathbf{r}_2)\psi_S(\mathbf{r}_2)$ debido al electrón 2. Ahora intentaremos encontrar el mínimo de esta cantidad variando la forma de los valores de ψ sujetos a la restricción de que ellos permanecen ortonormales.

6.3.3 El procedimiento LCAO

Lo que normalmente se hace para moléculas es usar el modelo LCAO donde cada orbital molecular puede ser descrito como una combinación lineal de n orbitales atómicos $\chi_1, \chi_2, \dots, \chi_n$. Esta formulación de las ecuaciones de HF, donde se usa la aproximación LCAO, se acredita a Roothaan (1951a). Cada electrón con vector de posición \mathbf{r} es descrito por un orbital LCAO escrito como una combinación lineal de las χ 's, esto es

$$\psi_n(\mathbf{r}) = a_1\chi_1(\mathbf{r}) + a_2\chi_2(\mathbf{r}) + \dots + a_n\chi_n(\mathbf{r})$$

Ahora trataremos de obtener una expresión para la energía variacional en términos de los coeficientes LCAO, los operadores para uno y dos electrones y las χ 's.

En el caso de un orbital molecular simplemente ocupado la densidad de carga $P_1(\mathbf{r})$ y la matriz de carga \mathbf{P}_1 se calculan como:

$$P_1(\mathbf{r}) = \begin{pmatrix} \chi_1(\mathbf{r}) & \chi_2(\mathbf{r}) & \dots & \chi_n(\mathbf{r}) \end{pmatrix} \begin{pmatrix} a_1^2 & a_1a_2 & \dots & a_1a_n \\ a_2a_1 & a_2^2 & \dots & a_2a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_na_1 & a_na_2 & \dots & a_n^2 \end{pmatrix} \begin{pmatrix} \chi_1(\mathbf{r}) \\ \chi_2(\mathbf{r}) \\ \vdots \\ \chi_n(\mathbf{r}) \end{pmatrix}$$

donde la matriz cuadrada en este caso es \mathbf{P}_1 . En el caso de M orbitales doblemente ocupados, los elementos de la matriz \mathbf{P}_1 están dados por

$$P_{ij} = 2 \sum_{l=a}^{i=m} r_l r_j$$

Un modo fácil de recordar esto es juntar los coeficientes LCAO para todos los M orbitales doblemente ocupados en una matriz \mathbf{U} de dimensiones $n * M$

$$\mathbf{U} = \begin{pmatrix} a_1 & b_1 & \dots & m_1 \\ a_2 & b_2 & \dots & m_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_n & b_n & \dots & m_n \end{pmatrix}$$

Entonces $\mathbf{P}_1 = 2\mathbf{U}\mathbf{U}^T$, donde \mathbf{U}^T es la transpuesta de \mathbf{U} .

$$\mathbf{U}^T = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ \vdots & \vdots & \ddots & \vdots \\ m_1 & m_2 & \dots & m_n \end{pmatrix}$$

Formemos una matriz \mathbf{S} que contenga los traslapes de los orbitales atómicos. Esta matriz \mathbf{S} de dimensiones $n * n$ contiene las integrales de traslape de las funciones bases $\int \chi_i(\mathbf{r})\chi_j(\mathbf{r})d\mathbf{r}$.

$$\begin{pmatrix} S_{1,1} & S_{1,2} & \dots & S_{1,n} \\ S_{2,1} & S_{2,2} & \dots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & S_{n,2} & \dots & S_{n,n} \end{pmatrix}$$

Los orbitales LCAO son usualmente ortogonales y este hecho puede ser resumido en un resultado matricial. Algo de manipulación demuestra que $\mathbf{U}^T\mathbf{S}\mathbf{U}$ es una matriz unitaria (con n renglones y n columnas) y también que

$$\mathbf{P}_1\mathbf{S}\mathbf{P}_1 = 4\mathbf{P}_1$$

Ahora usemos el principio variacional para buscar los mejores valores posibles de los coeficientes LCAO. Recordemos que para una molécula estamos trabajando dentro de la aproximación de Born-Oppenheimer y así los núcleos están clavados en sus posiciones en el espacio. Lo que vamos a hacer es calcular la integral variacional electrónica ϵ_{el} como se dijo y entonces optimizar esto (poniendo su primera derivada a cero). La optimización así involucra calcular el cambio de primer orden en la energía con respecto a todas las variables, y entonces poner este cambio de primer orden igual a cero y resolver

las ecuaciones resultantes. Uno podría trabajar con los coeficientes LCAO, pero otro método es trabajar directamente con la matriz densidad de carga \mathbf{P}_1 . Forcemos el requerimiento de que los LCAO tienen que ser ortonormales al hacer que

$$\mathbf{P}_1 \mathbf{S} \mathbf{P}_1 = 4 \mathbf{P}_1$$

Si juntamos todas las integrales que involucran los orbitales atómicos

$$\chi_1(\mathbf{r}), \chi_2(\mathbf{r}), \dots, \chi_n(\mathbf{r})$$

así como operadores de un electrón en una matriz \mathbf{h}_1 $n \times n$ tal que

$$(\mathbf{h}_1)_{ij} = \int \chi_i(\mathbf{r}) h(\mathbf{r}) \chi_j(\mathbf{r}) d\tau$$

Entonces los operadores de un electrón hacen una contribución

$$\sum_{i=1}^n \sum_{j=1}^n (\mathbf{h}_1)_{ij} (\mathbf{P}_1)_{ji}$$

a la energía variacional. Esta doble suma puede ser escrita un poco mas clara como la traza del producto matricial $\mathbf{P}_1 \mathbf{h}_1$. La contribución de los operadores de dos electrones a la energía puede ser escrita como $\frac{1}{2} \text{traza}(\mathbf{P}_1 \mathbf{G})$, donde los elementos de la matriz \mathbf{G} depende de los elementos de la matriz \mathbf{P}_1 en un modo complicado:

$$G_{ij} = \sum \sum P_{1,kl} \left(\int \int \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 - \right. \\ \left. \frac{1}{2} \int \int \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \right)$$

La energía electrónica total está finalmente dada por

$$\epsilon_{el} = \text{traza}(\mathbf{h}_1 \mathbf{P}_1) + \frac{1}{2} \text{traza}(\mathbf{P}_1 \mathbf{G})$$

El siguiente paso en el cálculo variacional es examinar como la energía cambia cuando la densidad electrónica cambia. Haciendo $\mathbf{P}_1 \rightarrow \mathbf{P}_1 + \delta \mathbf{P}_1$ y después de algo de manipulación matemática se encuentra que el cambio de primer orden en la energía es:

$$\delta \epsilon_{el} = \text{traza}(\mathbf{h}_1 \delta \mathbf{P}_1) + \text{traza}(\delta \mathbf{P}_1 \mathbf{G})$$

Notemos que el factor $\frac{1}{2}$ ha desaparecido de la expresión de la energía, esto porque la misma matriz \mathbf{G} depende de la densidad electrónica. Se puede escribir la ecuación $\delta\epsilon_{el}$ en términos de la matriz hamiltoniana de Hartree-Fock

$$\mathbf{h}^F = \mathbf{h}_1 + \mathbf{G}$$

Como la expresión siguiente

$$\delta\epsilon_{el} = \text{traza}(\delta\mathbf{P}_1 \mathbf{h}^F)$$

Los elementos de la matriz HF son

$$h_{ij}^F = \int \chi_i(\mathbf{r}_1) h(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau_1 + \\ \sum \sum P_{1,kl} \left[\int \int \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \right] - \\ \frac{1}{2} \int \int \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) g(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2$$

Deseamos encontrar un cambio $\delta\mathbf{P}_1$ que haga la variación de energía $\delta\epsilon_{el}$ igual a cero. \mathbf{P}_1 contiene información acerca del número de electrones, y esta también satisface la condición de que

$$\mathbf{P}_1 \mathbf{S} \mathbf{P}_1 = 4\mathbf{P}_1$$

La cual es lo mismo que decir que los orbitales LCAO son ortonormales. Substituyendo el nuevo \mathbf{P}_1 en esta ecuación y después de una pequeña manipulación se puede demostrar que, en el mínimo de energía

$$\mathbf{h}^F \mathbf{P}_1 = \mathbf{P}_1 \mathbf{h}^F$$

Esto demuestra que, cuando hemos encontrado la matriz de densidad electrónica correcta \mathbf{P}_1 y se ha calculado el hamiltoniano de Hartree-Fock correcto a partir de esta, las dos matrices satisfacen la condición de que $\mathbf{h}^F \mathbf{P}_1 = \mathbf{P}_1 \mathbf{h}^F$

Se dice entonces que ellas conmutan. Esto no nos ayuda inmediatamente a encontrar la matriz de densidad electrónica (recordemos que \mathbf{h}^F está definida

en términos de \mathbf{P}_1) pero en cambio nos dice la condición para que sea un mínimo en la energía variacional. Lo que se hace a menudo es:

- Tener una suposición de la densidad electrónica \mathbf{P}_1
- Calcular \mathbf{h}^F a partir de esta \mathbf{P}_1
- Checar la condición para que sea un mínimo
- Mejorar de alguna manera \mathbf{P}_1 , hasta que esta condición sea cumplida.

Una formulación alternativa del problema de campo autoconsistente LCAO en términos de estos coeficientes es la debida a Roothaan. Esta trabaja en términos de los eigenvalores y eigenvectores de la matriz hamiltoniana de Hartree-Fock. En términos de un orbital típico LCAO que describe un electrón con posición \mathbf{r}

$$\psi_1(\mathbf{r}) = a_1\chi_1(\mathbf{r}) + a_2\chi_2(\mathbf{r}) + \dots + a_n\chi_n(\mathbf{r})$$

Entonces en el mínimo de energía los coeficientes a_i forman un eigenvector de \mathbf{h}^F , y los eigenvectores de la energía mas baja corresponden a los orbitales moleculares LCAO ocupados. Así, si escribimos los coeficientes LCAO como un vector columna

$$\mathbf{a}^T = (a_1, a_2, \dots, a_n)$$

Entonces

$$\mathbf{h}^F \mathbf{a} = \epsilon_4 \mathbf{S} \mathbf{a}$$

donde ϵ_4 es llamada la energía orbital. El procedimiento de Roothaan involucra encontrar la matriz \mathbf{h}^F , la cual depende de los coeficientes orbitales LCAO. El procedimiento es casi igual ◦ Tener una suposición de los coeficientes orbitales LCAO

- Calcular \mathbf{h}^F a partir de estos coeficientes
- Checar la condición de que sea un mínimo
- Mejorar estos coeficientes LCAO de alguna manera, hasta que esta condición se cumpla

6.3.4 La energía electrónica

Supongamos que tenemos una molécula con $2M$ electrones, restringida a ocupar M orbitales LCAO, y que los orbitales SCF tienen energías $\epsilon_A, \epsilon_B, \dots, \epsilon_M$. Los coeficientes LCAO son agrupados en los vectores columna $\mathbf{a}, \mathbf{b}, \dots, \mathbf{m}$. La energía electrónica está dada por:

$$\epsilon_{el} = \text{traza}(\mathbf{h}_1 \mathbf{P}_1) + \frac{1}{2} \text{traza}(\mathbf{P}_1 \mathbf{G})$$

Y la suma de las energías orbitales está dada por

$$\epsilon_{orb} = 2(\epsilon_A + \epsilon_B + \dots + \epsilon_M)$$

Como

$$\mathbf{h}^F \mathbf{a} = \epsilon_A \mathbf{S} \mathbf{a}$$

Entonces

$$\mathbf{a}^T \mathbf{h}^F \mathbf{a} = \epsilon_A \mathbf{a}^T \mathbf{S} \mathbf{a} = \epsilon_A$$

Y con un poco de manipulación se encuentra que

$$\epsilon_{el} = \epsilon_{orb} - \frac{1}{2} \text{traza}(\mathbf{P}_1 \mathbf{G})$$

Así las energías orbitales no se suman a la energía electrónica total.

Bibliografía

- [1] I. W. Sherman, V. G. Sherman. *Biología, una Perspectiva Humana*, tercera edición, 1990. Mc Graw Hill México.
- [2] J. Vega, K. Michaelian, L. Hernández, M. R. Beltrán, I. L. Garzón, "Isomers of Adenine", *Journal of Molecular Structure, THEOCHEM*. 493 (1999) 275-285.
- [3] D.H. Andrews *Phys. Rev* , 36:1930; 544-554
- [4] N L Allinger *Advances in Physical Organic Chemistry*, V. Gold and J. Bethell (eds.), Academic Press, London 13: 1976; 1-82.
- [5] E. Osawa and H. Musso *Angew. Chem. Int. Ed. Engl* , 22:1983; 1-12
- [6] J. Lomas *L'Actualit Chimique*, 1986; 7-22.
- [7] S. Wilson *Chemistry by Computer*, Plenum Press, New York, 1986; 85-110
- [8] C. Altona and D H. Faber *Top. Curr. Chem.*, 45: 1974; 1-38.
- [9] O. Ermer *Struct. Bonding*, 27: 1976; 161-211.
- [10] S.R. Niketic and K. Rasmussen *The Consistent Force Field. A Documentation*, Springer-verlag, Berlin. 1977.
- [11] O. Ermer *Aspecte von Kraftfeldrechnungen*, Wolfgang Baur, Munich, 1981

- [12] U. Burkert and N.L. Allinger *Molecular Mechanics* American Chemical Society, Washington, DC, 1982.
- [13] N.L. Allinger *J. Am. Chem. Soc.*, 99:1977; 8127-8134.
- [14] N.L. Allinger and J.H. Lii *J. Comput. Chem.*, 8: 1987; 1146-1153.
- [15] N.L. Allinger, H.J. Geise, W. Pyckhout, L.A. Paquette and J.C. Gallucci *J. Am. Chem. Soc.*, 111: 1989; 1106-1114.
- [16] N.L. Allinger, Y.H. Yuh and J.H. Lii *J. Am. Chem. Soc.*, 111: 1989; 8551-8565
- [17] J.H. Lii and N.L. Allinger *J. Am. Chem. Soc.*, 111: 1989; 8566-8575.
- [18] J.H. Lii and N.L. Allinger *J. Am. Chem. Soc.*, 111: 1989; 8576-8582.
- [19] A. Warshel and S. Lifson *J. Chem. Phys.*, 53: 1970; 582-594
- [20] S. Lifson and A. Warshel *ibid.*, 49: 1968; 5116-5129.
- [21] N.L. Allinger, *Molecular mechanics*, International School of Crystallography: Static and Dynamic Implications of Precise Structural Information. Erice, Italy, 1985: 149-164.
- [22] W. Klyne and V. Prelog *Experientia*, 16: 1960; 521.
- [23] R.A. Buckingham *Proc. Roy. Soc. London A*, 168: 1938; 264-283
- [24] T.L. Hill *J. Chem. Phys.*, 16:1948; 399-404.
- [25] D.E. Williams *J. Chem. Phys.*, 43: 1965; 4424-4426.
- [26] U. Burkert *Tetrahedron*, 33: 1977; 2237-2242.
- [27] N.L. Allinger and D.Y. Chung *J. Am. Chem. Soc.*, 98:1976; 6798-6803.
- [28] D.H. Wertz and N.L. Allinger *Tetrahedron*, 30: 1974; 1579-1586.
- [29] S. Fitzwater and L.S. Bartell *J. Am. Chem. Soc.*, 98: 1976; 5107-5115.

TESIS CON
FALLA DE ORIGEN

- [30] R.H. Boyd J. Am. Chem. Soc., 97: 1975; 5353-5357.
- [31] A. Bondi J. Phys. Chem., 68: 1964; 441-451.
- [32] S.J. Wiener, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P.K. Weiner J. Am. Chem. Soc., 106: 1984; 765-784.
- [33] J.M. Leonard and W.P. Ashman J. Comp. Chem., 11: 199; 952-957.
- [34] A.J. Hopfinger and R.A. Pearlstein J. Comp. Chem., 5: 1984; 486-499.
- [35] J.L.M. Dillen J. Comput. Chem., 13: 1992; 257-267.
- [36] J.L.M. Dillen J. Comput. Chem., 11: 1990; 1125-1138.
- [37] K. Palmo, L.O. Pietila and S. Krimm J. Comput. Chem., 12:1991; 385-390.
- [38] K. Palmo, L.O. Pietila and S. Krimm J. Comput. Chem., 13: 1992; 1142-1150.
- [39] U.C. Singh and P.A. Kollman J. Comput. Chem., 5: 1984; 129-145.
- [40] J. Mullay J. Comput. Chem., 9:1988; 399-405.
- [41] W.J. Mortier, S.K. Ghosh and S. Shankar J. Am. Chem. Soc., 108: 1986; 4315-4320.
- [42] J. Mullay J. Am. Chem. Soc., 108: 1986; 1770-1776.
- [43] W.J. Mortier, K. Van Genechten and J. Gasteiger J. Am. Chem. Soc., 107:1985:829-835.
- [44] J. Gasteiger and M. Marsili Tetrahedron, 36: 1980; 3219-3288.
- [45] R.J. Abraham and P. Smith J. Comput. Chem., 9: 1988; 288-297.
- [46] L.G. Hammarstrom, T. Liljefors and J. Gasteiger J. Comput. Chem., 9: 1988. 424-440.

TESIS CON
FALLA DE ORIGEN

- [47] R.J. Abraham and G.H. Grant *J. Comput. Chem.*, 9: 1988; 244-256.
- [48] L. Dosen-Micovic, D. Jeremie and N.L. Allinger *J. Am. Chem. Soc.*, 105: 1983: 1716-1722 and 1723-1733.
- [49] N.L. Allinger, S.H.M. Chang D.H. Glaser and H. Honig *Isr. J. Chem.*, 20:1980: 51-56.
- [50] R.F. McGuire, F.A. Momany and H.A. Scheraga *J. Phys. Chem.*, 76:1972: 375-393
- [51] H.A. Scheraga *Adv Phys. Org. Chem.*, 6:1968; 103-184.
- [52] R A Scott and H.A Scheraga *J. Chem. Phys.*, 45: 1966; 2091-2101.
- [53] E.R. Lippincott and R. Schroeder *Chem. Phys.*, 23: 1955; 1099-1106
- [54] R Schroeder and E R. Lippincott *Phys. Chem.*, 61: 1957; 921-928.
- [55] N.L. Allinger, R.A. Kok and M.R. Imam *J. Comput. Chem.*, 9:1988; 591-595.
- [56] M K. Gilson and B. Honig *J. Comput-Aided Mol. Des.*, 5: 1991: 5-20.
- [57] A. Warshel and A. Lippicirella *J. Am. Chem. Soc* , 103: 1981: 4664-4673.
- [58] N.L. Allinger and J T. Sprague *J. Am. Chem. Soc.*, 95: 1973; 3893-3907.
- [59] J.T. Sprague, J.C. Tai, Y.H. Yuh and N L. Allinger *J. Comput. Chem.*, 8: 1987: 581-603
- [60] J.C. Tai, J.H. Lii and N.L. Allinger *J. Comput. Chem.*, 10: 1989; 635-647.
- [61] M. Froimowitz *J. Comput. Chem.*, 12: 199 1; 1 1 29-1133.
- [62] D.L. DeTar, S. Binzet and P. Darba *J. Org. Chem* , 50.- 1985: 2826-2836

TESIS CON
FALLA DE ORIGEN

- [63] W.H. Press, B.P. Flannery, S.A. Teutolsky and W.T. Vetterling Numerical Recipes, Cambridge University Press, Cambridge (USA), 1989; p 274-334
- [64] J. Labanowski, J. Motoc, C.B. Naylor, D. Mayer and R.A. Dammkoehler Quant. Struct -Act. Relat., 5: 1986; 138-152.
- [65] K.B. Wiberg J. Am. Chem. Soc., 87:1965; 1070-1078.
- [66] E.M. Engler, J.D. Andose and P. von R. Schleyer J. Am. Chem. Soc., 95: 1973; 8005-8025.
- [67] I.F. Stanton and D.E. Bernholdt J. Comput. Chem., 11: 1990, 53
- [68] K. Zimmermann J. Comput. Chem., 12:1991, 310-319.
- [69] H.A. Scheraga Peptides. Proceedings of the Fifth American Peptides Symposium, M. Goodman and J. Meienhofer eds. J. Wiley, New York, 1977.
- [70] M. Kontovianni, A.J. Hoffman and J.P. Bowen J. Comput. Chem., 13: 1992 57-65
- [71] E.L. Stewart and J.P. Bowen J. Comput. Chem., 13:1992; 1125-1137
- [72] L. Carballeira, R.A. Mosquera and M.A. Rios J. Comput. Chem., 9:1988; 851-860
- [73] L. Carballeira, R.A. Mosquera and M.A. Rios J. Comput. Chem., 10: 1989; 911-920.
- [74] S.A. Vasquez, M.A. Rios and L. Carballeira J. Comput. Chem., 12: 1991; 872-879.
- [75] I.S. Fernandez, M.A. Rios and L. Carballeira J. Comput. Chem., 12:1991 78-90.

TESIS CON
FALLA DE ORIGEN

- [76] I. Brocker, R.W. Hoffmann and K.N. Houk J. Am. Chem. Soc., 113: 1991: 5006-5017.
- [77] S. Giorgas and T.H. Lane J. Comput. Chem., 9: 1988; 25-39.
- [78] R.J. Abraham and G.H. Grant J. Comput. Chem., 9: 1988; 709-718.
- [79] S. Profeta Jr, R.J. Unwalla and F.K. Cartledge J. Comput. Chem., 10: 1989: 99-103.
- [80] J.L. Nicholas, A.J. Hopfinger, F.R. Trouw and L.E. Iton J. Am. Chem. Soc. 113: 1991; 4792-4800.
- [81] T. Liljefors, J.C. Tai, S. Li and N.L. Allinger J. Comput. Chem., 8: 1987: 1051-1056.
- [82] C.J.M. Huige, A.M.F. Hezemans and K. Rasmussen J. Comput. Chem., 8: 1987; 204-225.
- [83] M. Saunders and R.M. Jarret J. Comput. Chem., 7: 1986; 578-588
- [84] P.K. Weiner and P.A. Kollman J. Comput. Chem., 2: 1981; 287-303.
- [85] W.L. Jorgensen and J. Tirado-Rives J. Am. Chem. Soc. 110., 1988: 1657-1666.
- [86] W.L. Jorgensen, S.G. Wierschke and W.L. Jorgensen J. Am. Chem. Soc., 113:1991;2810-2819.
- [87] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impex and M.L. Klein J. Chem. Phys., 79: 1983; 926.
- [88] R.A. Momany, R.F. McGuire, A.W. Burgess and H.A. Scheraga J. Phys. Chem., 79: 1975;2361-2381.
- [89] B.R. Geliti and M. Karplus Biochemistry, 18: 1979; 1256-1268.
- [90] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus J. Comput. Chem., 4: 1983; 187-217.