

Proyecto III - Biofísica Molecular

Dra. Alexandra Deriabina

Emmanuel Ramirez Bañuelos

Diciembre 2023

Índice

Capítulos	Página
1. OptiADN: Optimización de Bases Nitrogenadas mediante Método Hartree-Fock	3
1.1. Método Hartree-Fock	3
1.2. El Método de Hartree-Fock y los Conjuntos de Base	3
1.3. El método HF 6-31G en Gaussian 09	4
1.4. Moléculas y compuestos pre-optimizados en RSCB y CCDC	5
1.5. Aplicación del método HF 6-31G en la tímica	6
2. Resultados	8
2.1. Ejemplo con la Guanina	9
2.2. Ángulos de torsión en una geometría optimizada.	10

1. OptiADN: Optimización de Bases Nitrogenadas mediante Método Hartree-Fock

1.1. Método Hartree-Fock

La esencia del método de Hartree-Fock (HF) radica en descomponer una ecuación de Schrödinger de N electrones en N ecuaciones de Schrödinger de un solo electrón. En el operador de un solo electrón. Una dificultad en los cálculos de HF es la necesidad de realizar una suposición inicial de las funciones de onda de un solo electrón, una para cada electrón. Otra dificultad está en resolver las ecuaciones de Schrödinger de un solo electrón. Para solventar estos problemas, es común parametrizar algunas funciones gaussianas y luego utilizar combinaciones lineales de estas funciones gaussianas preparametrizadas para describir las ondas de los electrones en átomos y moléculas. Este conjunto de funciones gaussianas se conoce como conjunto de base. El método de Hartree-Fock se simplifica en la optimización de los coeficientes de las combinaciones lineales de estas funciones gaussianas. Por ejemplo, en el cálculo de energía HF/STO-3G de la molécula H_3 , la función de onda del orbital $1s$ de cada átomo de H se expresa como una combinación de funciones gaussianas:

$$\Psi_{1s} = \sum_{i=1}^3 d_i e^{-\zeta_i r^2} = 0.154e^{-3.425r^2} + 0.535e^{-0.624r^2} + 0.445e^{-0.169r^2}$$

donde d_i , los coeficientes frente a las funciones gaussianas, y ζ_i , los parámetros exponenciales, están pre-optimizados. ¿Por qué se usan a menudo funciones gaussianas en lugar de funciones exponenciales en los conjuntos de base? La única razón es que calcular integrales de funciones gaussianas es mucho más rápido en una computadora.

1.2. El Método de Hartree-Fock y los Conjuntos de Base

Cuanto mayor sea el conjunto de base utilizado en los cálculos de HF, más flexibles y precisas pueden ser las funciones de onda de los electrones. El conjunto de base 6-31G es uno de los conjuntos de base más grandes. El "6" denota una combinación lineal de 6 funciones gaussianas utilizadas para describir cada electrón del núcleo. El "31" denota una combinación lineal de tres combinaciones lineales de funciones gaussianas para cada electrón de valencia. La "G" representa "Gaussian". Para el átomo de hidrógeno, no existen electrones del núcleo:

$$\Psi_{H1s} = A \sum_{i=1}^3 d_i e^{-\zeta_i r^2} + B \sum_{i=1}^1 e^{-\zeta_i r^2}$$

donde A y B son optimizados en el cálculo de HF.

Para el átomo Li , los electrones del núcleo y los electrones de valencia se tratan de manera diferente:

$$\Psi_{Li\ 1s} = \sum_{i=1}^6 d_i^m e^{-\zeta_i^m r^2} \quad (1)$$

$$\Psi_{Li\ 2px} = A \sum_{i=1}^3 d_i x e^{-\zeta_i r^2} + B \sum_{i=1}^1 x e^{\zeta_i' r^2}$$

1.3. El método HF 6-31G en Gaussian 09

En el programa de química cuántica computacional, Gaussian 09, podemos aplicar el método Hartree-Fock con un conjunto de base 6-31G para **optimizar** la **configuración** de las cuatro bases nitrogenadas del ADN.

Por ejemplo, para optimizar a la **timina** partiendo de las coordenadas ¹ de las posiciones atómicas del compuesto, invocamos el método introduciendo el código de la siguiente forma:

¹En Chemcraft, podemos obtener estas coordenadas desde la pestaña **COORD** a la derecha de **SOURCE**, debajo del botón "Help"

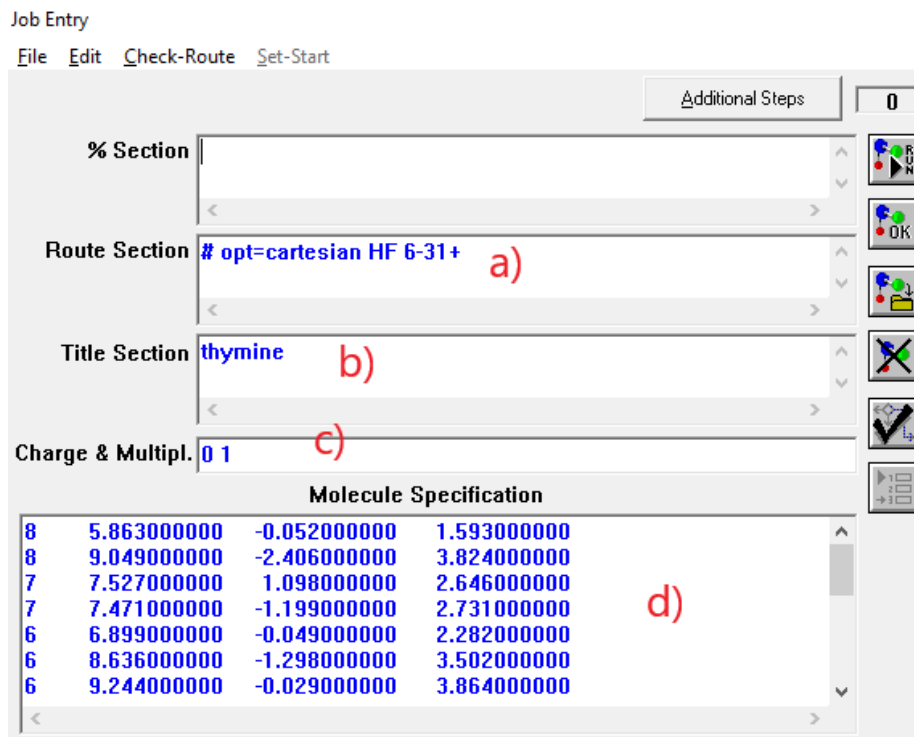


Figura 1: **Uso del método HF 6-31G** a) Se sentencia el método HF 6-31G con el comando **HF 6-31+**, el comando **opt=cartesian** declara que lo que se desea optimizar son las posiciones cartesianas del arreglo atómico de la tímína, se desea encontrar la configuración con menor energía. b) Colocamos el título del programa. c) Introducimos el tipo de carga y la multiplicidad de la misma. d) La especificación de la molécula, en este caso, se proporciona a través de las coordenadas de las posiciones atómicas.

Con ayuda de esta herramienta, encontraremos las configuraciones para cada una de las bases nitrogenadas con el mínimo de energía posible.

1.4. Moléculas y compuestos pre-optimizados en RSCB y CCDC

Para facilitar el compute del HF 6-31G, sería conveniente trabajar con moléculas o estructuras ya optimizadas o sometidas a algoritmos de simulación para definir sus parametros.

Podemos hacer uso de las bases de datos del RSCB² y el CCDC³ para hallar estructuras de las cuales podemos partir para someterlas a nuestro algoritmo.

El RSCB, usualmente aloja compuestos, proteínas y estructuras muy complejas, a pesar de que intentemos buscar las bases nitrogenadas, deberemos primero aislarlas del resto de componentes que no son de interés. Lo idóneo es comenzar con la base de datos de cristales, allí podemos encontrar compuestos más simples, en el caso de las bases nitrogenadas, usualmente vienen acompañadas de otros compuestos que son fáciles de eliminar.

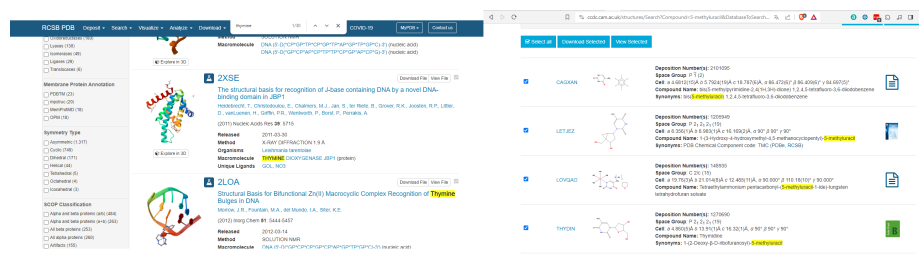


Figura 2: **A la izquierda** Los resultados del RSCB arroja proteínas y compuestos muy complicados de aislar. **A la derecha**, los resultados de la tímica en el CCDC son mas adecuados para este trabajo.

1.5. Aplicación del método HF 6-31G en la tímica

A continuación, se muestra a la tímica ya depurada en Chemcraft, evidentemente, se tuvieron que elaborar algunas correcciones, como el **añadir/quitar/cambiar** enlaces simples y dobles, así como átomos, pues el compuesto del CCDC tiene a la tímica, con una estructura química ligeramente diferente por ser parte de la estructura de un compuesto más complejo.

²Research Collaboratory for Structural Bioinformatics

³Cambridge Crystallographic Data Centre

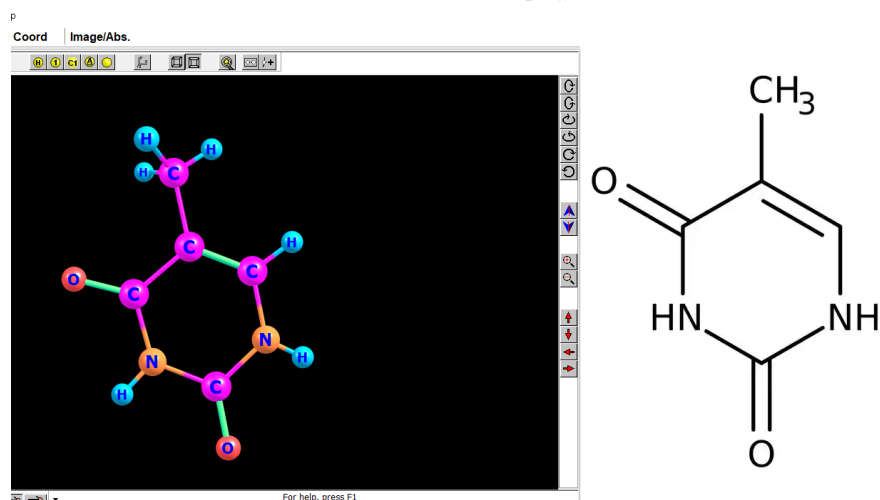


Figura 3: Timina extraída del **bis(5-metiluracilo) 1,2,4,5-tetrafluoro-3,6-diiodobenceno** del CCDC, con número de deposición: 2101095

Al igual que en el proyecto anterior con el H_2O , los archivos CIF descargados del CCDC deben abrirse en **Mercury** para salvarlos después como archivos **PDB**. Éste último es el que se abre en Chemcraft para comenzar a aislar la tímica de otros componentes.

En la pestaña **COORD**, encontraremos las coordenadas que necesitaremos para colocar en el código de optimización de Gaussian 09, las mismas las pegaremos tal como se muestra en la figura 1.

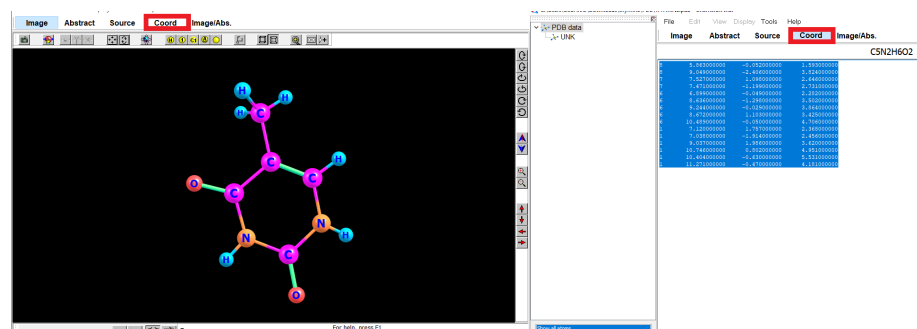


Figura 4: Posiciones atómicas.

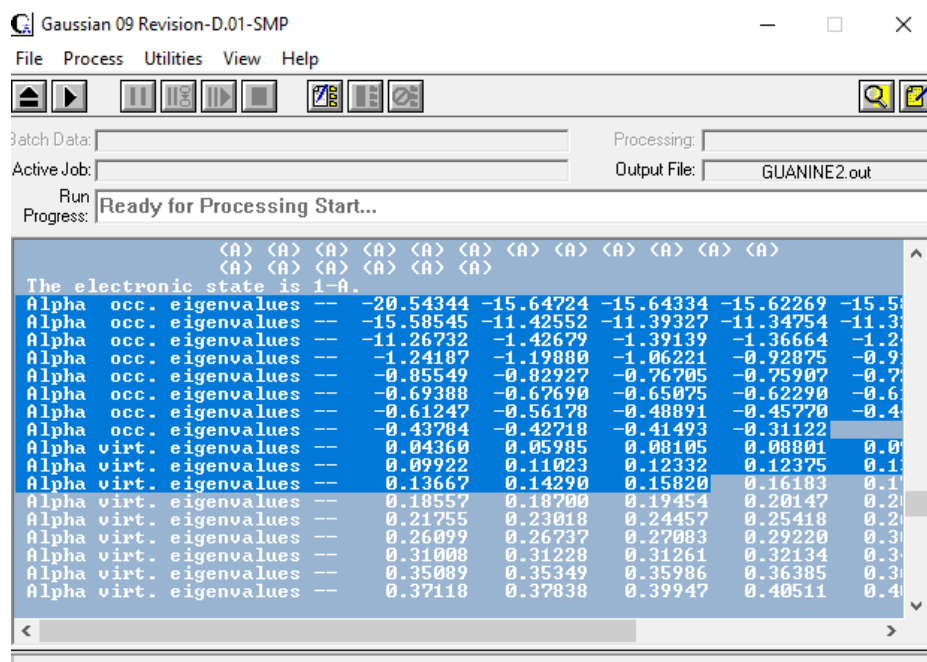


Figura 5: Finalmente, cuando hayamos colocado todas las especificaciones, el **route section**, **charge & multipl** y **nuestras coordenadas**, como se indican en la figura 1, al dar **RUN** se nos mostrará esta ventana dónde se desglozan todos los cálculos de la optimización.

2. Resultados

En nuestro **repositorio en Github** [1]⁴, alojamos cada una de las bases nitrogenadas optimizadas, las cuales se pueden encontrar en el directorio llamado: **Gaussian Optimization**.

⁴<https://github.com/Thntr/Nitrogenous-Bases>

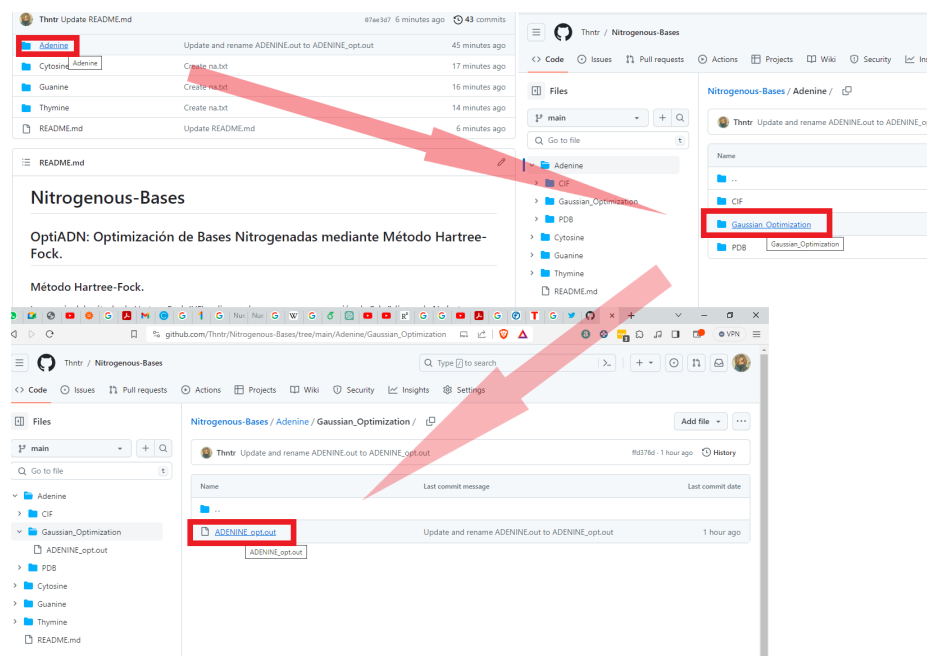


Figura 6: Navegación: Base nitrogenada → Gaussian Optimization → file.out

2.1. Ejemplo con la Guanina

A diferencia de las otras bases nitrogenadas, la **guanina** presenta una estructura de mayor complejidad por su estructura de doble-anillo y por incluir un NH y un NH_2 a diferencia de la adenina, que es la base **púrica** que más se le asemeja. Dada dicha complejidad, la simulación demora alrededor 1 hora, pero mostrando así la siguiente configuración final con el mínimo de energía que se muestra:

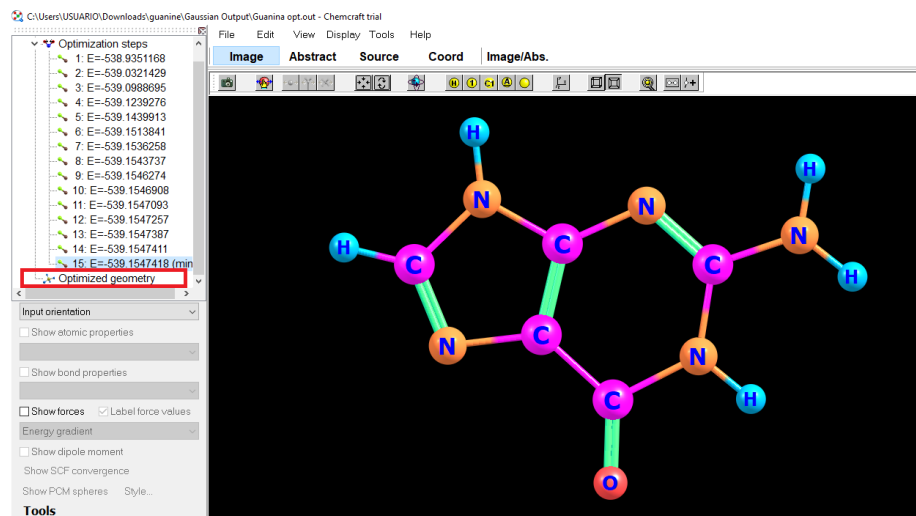


Figura 7: Con un mejor poder de procesamiento, el tiempo requerido para optimizar la geometría de la molécula fue de 15 min.

2.2. Ángulos de torsión en una geometría optimizada.

La estructura tridimensional se puede caracterizar mediante **distancias de enlace, ángulos de enlace y grupos de átomos que rotan alrededor de un enlace**. Estas rotaciones sobre un enlace central, B-C, se describen con los ángulos de torsión tomando en cuenta cuatro átomos consecutivos. El ángulo de torsión se define como el ángulo, τ , que se forma entre dos planos que contienen a los átomos A, B, C y B, C, D (figura 8). Los ángulos de torsión en las rotaciones moleculares se clasifican en distintas regiones dependiendo de su valor. Para las moléculas orgánicas la notación empleada es la propuesta por Klyne and Prelog: syn (0°), anti (180°), \pm synclinal ($\pm 60^\circ$) y \pm anticlinal ($\pm 120^\circ$). En el caso de estructuras cristalográficas la notación usualmente usada es la siguiente: cis (0°), trans (180°), + gauche g+ ($+60^\circ$) y - gauche g- (-60°). [2]

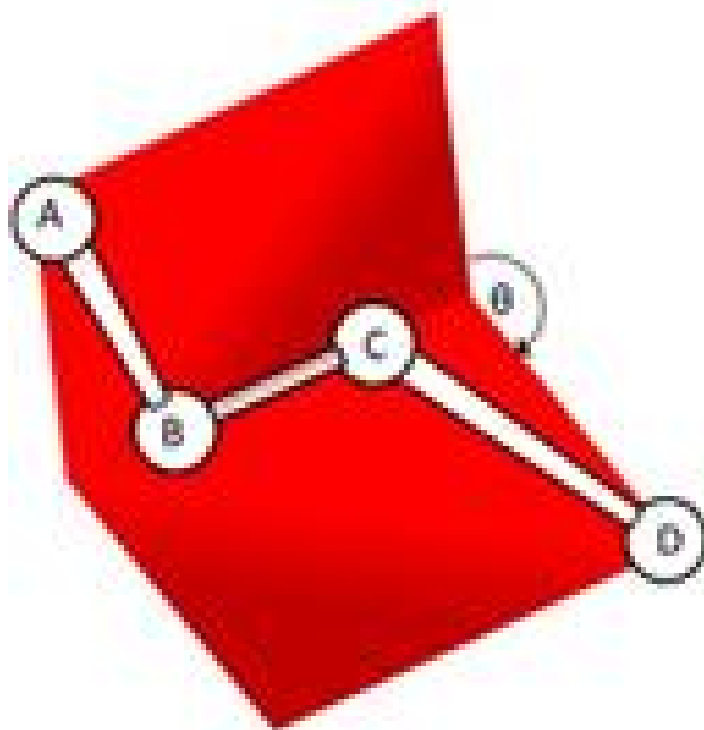


Figura 8: Definición de un ángulo de torsión.

Tomando en cuenta que los ángulos de torsión se determinan de esta forma, en nuestro modelo, podemos concluir que todas las bases nitrogenadas se encuentran en un plano en su geometría más óptima.

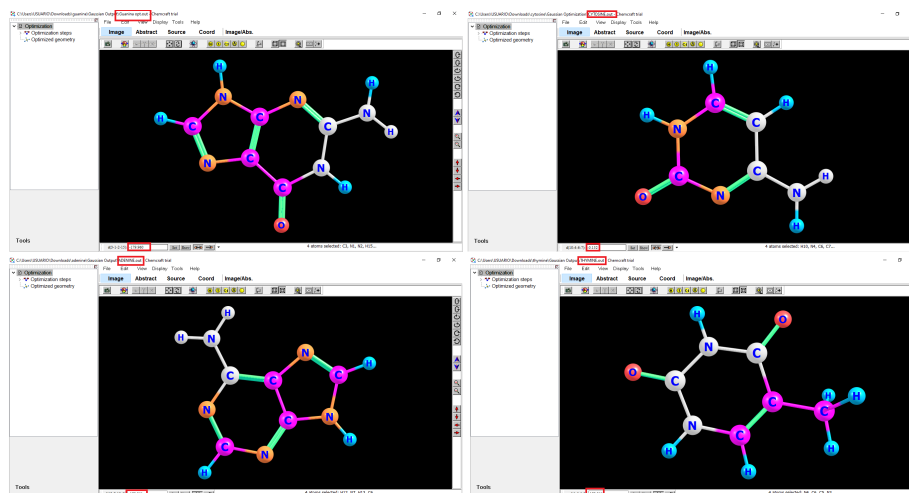


Figura 9: Cómo se puede notar, en cualquiera de las bases nitrogenadas, los ángulos de torsión al tomar cualesquiera 4 átomos seguidos, corresponden a valores muy cercanos de 0° ó 180° . Básicamente, la estructura más óptima contiene a todos sus átomos prácticamente en un plano. a) **Guanina**, b) **Citosina**, c) **Adenina**, d) **Timina**.

Referencias

- [1] <https://github.com/Thntr/Nitrogenous-Bases>
- [2] Ruiz Millán A. (2018), Comparación de los modelos mecánico moleculares y sus modificaciones para la investigación de los fragmentos de los ácidos nucleicos. BUAP
- [3] <https://www.rcsb.org/structure>
- [4] <https://www.ccdc.cam.ac.uk/structures>