



BENEMÉRITA UNIVERSIDAD  
AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS QUÍMICAS

# TESIS

## REVISIÓN DEL MÉTODO G4 HACIA EL CÁLCULO TEÓRICO DE ENTALPÍAS DE FORMACIÓN

TESIS PRESENTADA PARA OBTENER EL GRADO DE  
DOCTOR EN CIENCIAS QUÍMICAS

PRESENTA:

M.C.I. DAVID OCHOA RESENDIZ

DIRECTORES DE TESIS:

DR. JULIO MANUEL HERNÁNDEZ PÉREZ

DR. JUAN MANUEL SOLANO ALTAMIRANO

Puebla, Pue. Noviembre 2025





# BUAP

## A QUIEN CORRESPONDA:

Por este conducto la Comisión Revisora formada por el D.C. Henoc Flores Segura, D.C. Jacinto Sandoval Lira, D.C. Joaquín Barroso Flores, D.C. Juan Francisco Rivas Silva y el D.C Saúl Juan Carlos Salazar Samaniego, informan haber leído y aprobado para su impresión, la tesis denominada **“Revisión del método G4 hacia el cálculo teórico de entalpías de formación”**, que presenta el alumno **M.C.I. David Ochoa Resendiz** para obtener el grado de Doctor en Ciencias Químicas.

Se extiende la presente a los dos días del mes de septiembre del año dos mil veinticinco.

## ATENTAMENTE

D.C. Henoc Flores Segura

D.C. Jacinto Sandoval Lira

D.C. Joaquín Barroso Flores

D.C. Juan Francisco Rivas Silva

D.C. Saúl Juan Carlos Salazar Samaniego

---

*“Cuanto más precisos se hacen los cálculos, más conceptos tienden a desvanecerse en el aire.”*

*R. S. Mulliken*

# Dedicatoria

A mi esposa, la Dra. Grecia Katherine Sanchez Muñoz y a mis abuelos Teresa Resendiz Torres y Guillermo Ochoa Rodríguez

# Agradecimientos

A mi familia, por su amor, paciencia y confianza; sin su apoyo, este trabajo no habría sido posible. A la Ing. Maribel Ochoa Resendiz, al Mtro. Luis Vicario Sánchez, mi hermano Lic. Javier Ramírez Ochoa y a mi amada esposa, la Dra. Grecia Katherine Sánchez Muñoz.

A mis directores de tesis, el Dr. Julio Manuel Hernández Pérez y el Dr. Juan Manuel Solano Altamirano, por su dirección académica.

A los sinodales: el Dr. Henoc Flores Segura, el Dr. Juan Francisco Rivas Silva y el Dr. Saúl Juan Carlos Salazar Samaniego. En especial, al Dr. Jacinto Sandoval Lira, cuya influencia ha permeado en mi formación desde la licenciatura. Además, al Dr. Joaquín Barroso Flores por su valiosa asistencia técnica y sus comentarios oportunos.

A mis amigos, el Dr. Alfredo, la Dra. Cristina y el Mtro. Iván, por su disposición para escuchar y aconsejar cuando fue necesario. Finalmente, a quienes, con una pregunta oportuna o una crítica honesta, contribuyeron a mejorar este trabajo.

Este trabajo se realizó con el apoyo de los proyectos VIEP de la Benemérita Universidad Autónoma de Puebla: 100525751-VIEP2021, 100525751-VIEP2022, 100279411-VIEP2023, 100525751-VIEP2024 y 100525751-VIEP2025. Asimismo, se agradece a la Secretaría de Educación, Ciencia, Tecnología e Innovación (SEHCITI) por el apoyo otorgado mediante la beca de doctorado.

# Resumen

En el presente trabajo de tesis, se evaluó y modificó el método compuesto Gaussian-4 (G4) para calcular teóricamente las entalpías de formación de compuestos con grupos funcionales (nitro, nitrofenilos y heterociclos sustituidos con grupos nitro), estos sistemas presentan discrepancias significativas, ya que el método G4 muestra desviaciones de hasta  $15\text{--}40 \text{ kJ} \cdot \text{mol}^{-1}$ , frente a los  $3\text{--}4 \text{ kJ} \cdot \text{mol}^{-1}$  típicos en moléculas pequeñas y típicamente estudiadas.

Con el fin de mejorar la precisión en el cálculo de sus propiedades termoquímicas, se exploraron modificaciones en el método G4, incluyendo el cambio de funcional en las etapas de optimización de geometría y cálculo de frecuencias vibracionales. Sin embargo, se encontró que el conjunto de calibración y la reparametrización, específicamente el factor de escala  $F$  y el parámetro empírico  $A$ , parte del término de corrección de alto nivel (HLC, por sus siglas en inglés), ejercen una influencia mayor y no esperada, en comparación con la elección del funcional en la optimización de geometría y el cálculo de frecuencias.

Usando cálculos en Gaussian 09 y creando un conjunto de calibración, se evaluaron parámetros ajustados a diferentes conjuntos de calibración [SO67, SD43] y de prueba por medio de validación cruzada. Observamos una mejora significativa en la precisión de las entalpías de formación para nitrofenilos y heterociclos oxigenados, sulfurados y nitrogenados que contienen grupos nitro. Estos resultados evidencian que los parámetros  $F$  y  $A$  reparametrizados y adaptados a conjuntos particulares son más adecuados, proporcionando cálculos de entalpía de formación con

---

mayor concordancia teórico-experimental. Adicionalmente, se identificó que, al calcular las entalpías de formación para moléculas fuera del conjunto de calibración original de G4 [G3/05] y representativo con SO67, el método G4 tiende a perder precisión, sugiriendo que la adecuación de los parámetros depende críticamente de la información experimental y el tipo de moléculas incluidas en la calibración de los parámetros empíricos. Este estudio subraya la importancia de conjuntos de calibración específicos y el ajuste de los parámetros en este tipo de métodos compuestos, ofreciendo una metodología mejorada para el cálculo teórico de propiedades termoquímicas usando el método de atomización.

# Abstract

In the present thesis, the Gaussian-4 (G4) composite method was evaluated and modified to theoretically calculate the enthalpies of formation of compounds with functional groups (nitro, nitrophenyls, and heterocycles substituted with nitro groups). These systems exhibit significant discrepancies, as the G4 method shows deviations of up to  $15\text{--}40\text{ kJ}\cdot\text{mol}^{-1}$ , compared to the typical  $3\text{--}4\text{ kJ}\cdot\text{mol}^{-1}$  for small, commonly studied molecules.

In order to improve the accuracy of the calculated thermochemical properties, modifications to the G4 method were explored, including changing the functional in the geometry optimization and vibrational frequency calculation steps. However, it was found that the calibration set and the reparametrization—specifically the scaling factor  $F$  and the empirical parameter  $A$ , which are part of the high-level correction (HLC) term—exert a greater and unexpected influence compared to the choice of functional in geometry optimization and frequency calculations.

Using calculations in Gaussian 09 and creating a calibration set, parameters fitted to different calibration sets [SO67, SD43] and test sets were evaluated by cross-validation. We observed a significant improvement in the accuracy of the enthalpies of formation for nitrophenyls and oxygen-, sulfur-, and nitrogen-containing heterocycles bearing nitro groups. These results demonstrate that the reparameterized  $F$  and  $A$  parameters, adapted to specific datasets, are more suitable, providing enthalpy of formation calculations with greater theoretical–experimental agreement. Additionally, it was found that when calculating formation enthalpies for molecules

---

outside the original G4 [G3/05] calibration set and the representative SO67 set, the G4 method tends to lose accuracy, suggesting that the suitability of the parameters depends critically on experimental data and the type of molecules included in the empirical parameter calibration. This study underscores the importance of specific calibration sets and parameter tuning in such composite methods, offering an improved methodology for the theoretical calculation of thermochemical properties using the atomization method.

## Publicaciones relacionadas con la tesis

- D. Ochoa-Resendiz, J. M. Solano-Altamirano, J. Sandoval-Lira, Julio M. Hernández-Pérez; Specialized modification to the G4 composite method for computing formation enthalpies of nitrophenyl and nitrogen-heterocycle organic compounds. *J. Chem. Phys.* 21 July 2025; 163 (2): 024121. <https://doi.org/10.1063/5.0264222>
- F. Ramos, D. Ochoa-Resendiz, E. A. Camarillo, J. M. Hernández-Pérez, H. Flores, J. M. Solano-Altamirano; The thermochemical properties of 2- and 3-Thiophenemethanol: Eluding incomplete combustion issues through combining experimental and computational methodologies. *Thermochim. Acta.* 2025; 751: 180034. <https://doi.org/10.1016/j.tca.2025.180034>

### Participación en eventos

- Participación en modalidad de póster en la XXI Reunión Mexicana de Fisicoquímica Teórica (ICBI, UAEH, Pachuca, 9–11 noviembre 2023).
- Participación en modalidad de póster en la XX Reunión Mexicana de Fisicoquímica Teórica (UAEMor, Morelos, 17–19 noviembre 2022).
- Participación continua en ponencias en los Seminarios Internos del Posgrado en Ciencias Químicas (FCQ-BUAP), 2021–2024.

# Lista de abreviaturas

<b>Abreviatura</b>	<b>Significado</b>
AcTc	Active Thermochemical Tables (Tablas Termodinámicas Activas)
B3LYP	Funcional Becke, 3-parameter, Lee–Yang–Parr
CBS	Complete Basis Set (límite de conjunto de bases)
CCSD(T)	Coupled Cluster with Singles, Doubles and perturbative Triples
DFT	Density Functional Theory (Teoría del Funcional de la Densidad)
DMA	Desviación Media Absoluta
Full CI	Full Configuration Interaction (interacción completa de configuraciones)
G3/05	Conjunto de calibración de 270 moléculas usado en G3 y G4
G4	Método compuesto Gaussian-4
G4-nitro-cerrado-B3LYP	Variante modificada de G4 calibrada en SO110
HLC	High Level Correction (Corrección de alto nivel)
HF	Hartree–Fock
M06-2X	Funcional meta-GGA híbrido de la familia M06
MP2, MP4	Møller–Plesset perturbation theory de segundo/cuarto orden
NIST	National Institute of Standards and Technology
SD43	Subconjunto de 43 moléculas con nitro y heterociclos (Dorofeeva)
SO67	Subconjunto de 67 moléculas orgánicas del G3/05
SO110	Unión de SO67 y SD43 (110 moléculas)
ZPE	Zero Point Energy (Energía de Punto Cero)

# Índice general

<b>Lista de abreviaturas</b>	<b>XII</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Antecedentes</b>	<b>6</b>
2.1. Introducción al método compuesto G4 . . . . .	6
2.2. Controversia en nitrofenilos y heterociclos . . . . .	11
<b>3. Hipótesis y Objetivos</b>	<b>19</b>
3.1. Hipótesis . . . . .	19
3.2. Objetivo General . . . . .	19
3.3. Objetivos Específicos . . . . .	20
<b>4. Fundamento Teórico</b>	<b>21</b>

---

4.1. Métodos de la química computacional . . . . .	21
4.2. Mínimos de la SEP y frecuencias . . . . .	23
4.3. Método de Hartree-Fock . . . . .	24
4.4. Correlación electrónica . . . . .	26
4.5. Métodos post-Hartree-Fock . . . . .	27
4.5.1. Teoría de perturbaciones de Møller-Plesset (MPn) . . . . .	27
4.5.2. Interacción de configuraciones (IC) . . . . .	28
4.5.3. Cúmulos Acoplados (CC) . . . . .	29
4.6. Teoría de funcionales de la densidad . . . . .	31
4.7. Conjuntos de base . . . . .	35
4.8. Extrapolación de base infinita . . . . .	37
4.9. Métodos compuestos . . . . .	40
4.9.1. Métodos Gn y conjuntos de calibración . . . . .	41
4.9.2. Métodos compuestos CBS-n . . . . .	46
4.10. Termoquímica . . . . .	48
4.11. Entalpía de formación estándar . . . . .	49
4.11.1. Método de atomización . . . . .	50

4.11.2. Termodinámica estadística: contribuciones térmicas . . . . .	53
<b>5. Metodología</b>	<b>57</b>
5.1. Revisión del método G4 . . . . .	57
5.1.1. Pasos del cálculo G4 . . . . .	62
5.2. Implementación Computacional . . . . .	63
5.2.1. Reproducción y modificación del método G4 . . . . .	63
5.2.2. El algoritmo de parametrización . . . . .	70
5.3. Conjuntos y subconjuntos de calibración . . . . .	73
5.4. Consideraciones termoquímicas . . . . .	75
<b>6. Resultados y discusión</b>	<b>77</b>
6.1. Superficie de parametrización . . . . .	78
6.2. Optimización y cálculo en G4 . . . . .	79
6.3. Evaluación de G4 . . . . .	82
6.4. Calibración del conjunto SO110 . . . . .	83
6.5. Evaluación conjunto prueba . . . . .	88
6.5.1. Moléculas sin -NO <sub>2</sub> . . . . .	88
6.5.2. Moléculas sin grupos nitro ni heterocíclicos . . . . .	90

---

6.6. Sistemas con entalpías atípicas . . . . .	91
6.7. Comparación isodésmica . . . . .	92
<b>7. Conclusiones</b>	<b>95</b>
<b>A. Apéndice</b>	<b>97</b>

# Índice de figuras

2.1. Comparación del rendimiento del método G4 respecto a G3 en el conjunto de prueba G3/05 y subconjunto [sin H]. . . . .	8
2.2. Comparación de las desviaciones en las entalpías de formación estándar $\Delta_f H_{298.15\text{ K}}^\circ(T-Exp)$ , utilizando los métodos G3, G4, CBS-APNO y W1(RO) para compuestos seleccionados. La desviación se calcula como la diferencia entre los valores experimentales y teóricos. . . . .	14
5.1. Diagrama de flujo del programa ExtractLabfqotG4.sh para obtener los archivos *.dat.	65
5.2. Diagrama de flujo del programa LABFQOTG4, parte A. . . . .	67
5.3. Diagrama de flujo del programa LABFQOTG4, parte B. . . . .	69
5.4. Diagrama del flujo para la parametrización de $F$ y $A$ . . . . .	72
6.1. Superficie de parametrización de factor de escala $F$ y parámetro $A$ en el conjunto SO110 en función de la DMA, reportado en $\text{kJ} \cdot \text{mol}^{-1}$ . La figura 6.1a muestra la representación en 3D, mientras que la figura 6.1b muestra la representación en 2D; el punto rojo representa los valores óptimos de $F$ y $A$ con una DMA mínima. . . . .	79

6.2.  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  en función del número de electrones en el conjunto de calibración SO110. Se utilizan círculos vacíos para distinguir las moléculas que corresponden al subconjunto SO67, y círculos cerrados para las moléculas del subconjunto SD43. Las líneas discontinuas representan los límites de  $\mu \pm 2\sigma$  alrededor del promedio  $\mu$ , lo cual indica la dispersión de los datos en un intervalo de confianza del 95%. La línea discontinua central corresponde al valor promedio de todos los datos, reportado en  $\text{kJ} \cdot \text{mol}^{-1}$ . . . . . 86

A.1. Curvas de distribución normal para el conjunto [SDG-100], barrido en parámetro de factor de escala  $F$ . . . . . 97

A.2. Curvas de distribución normal para el conjunto [SDG-100], barrido en parámetro  $A$ . 98

A.3. Estimaciones de T1Diagnostic para evaluar la correlación no dinámica de el subconjunto SD43 . . . . . 98

A.4.  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  utilizando los métodos G4-nitro-cerrado-B3LYP, G4-nitro-cerrado-MP2, G4-nitro-cerrado-M062X, G4-nitro-cerrado- $\omega$ B97XD en función del número de electrones en el conjunto de calibración SO110. Se utilizan círculos vacíos para distinguir las moléculas que corresponden al subconjunto SO67, y círculos cerrados para las moléculas del subconjunto SD43. Las líneas discontinuas representan los límites de  $\mu \pm 2\sigma$  alrededor del promedio  $\mu$ , lo cual indica la dispersión de los datos en un intervalo de confianza del 95%. La línea discontinua central corresponde al valor promedio de todos los datos, reportado en  $\text{kJ} \cdot \text{mol}^{-1}$ . 99

A.5. Diferencias entre las entalpías de formación experimentales y las calculadas ( $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$ ) para un conjunto de moléculas que contienen grupos nitro, nitrofenilos o heterociclos nitrogenados (Tabla 6.5), utilizando diferentes métodos compuestos: CBS-QB3, G3, G3(MP2), G4(MP2), G4-STD, G4(MP2)-6X y G4-nitro-cerrado-B3LYP. En cada figura se muestran los valores para las 10 moléculas de prueba, junto con las líneas horizontales correspondientes a  $\mu \pm 2\sigma$ . Es importante mencionar que todos estos métodos compuestos contienen parámetros empíricos ajustados a ciertos conjuntos de calibración. . . . . 99

# Índice de cuadros

4.1. Componentes de los métodos Gn . . . . .	42
6.1. Evaluación de parámetros estadísticos y de calibración para el conjunto SO110 mediante diferentes variantes propuestas de G4. . . . .	82
6.2. Evaluación de parámetros estadísticos del método G4-STD aplicados a diferentes conjuntos. . . . .	82
6.3. Calibración y evaluación de parámetros de $F$ y $A$ para los conjuntos SO67, SD43 y SO110 con la variante G4-nitro-cerrado-B3LYP. . . . .	84
6.4. Desviaciones medias absolutas (DMA) en $\text{kJ} \cdot \text{mol}^{-1}$ al usar distintos conjuntos como <i>entrenamiento</i> (columnas) y aplicarlos a cada conjunto ( <i>aplicado a</i> , filas). . . .	85
6.5. Diferencias entre $\Delta_f H_{298.15\text{K}}^\circ$ experimental y calculado para moléculas con grupos tipo nitro, nitrofenilo o heterociclos nitrogenados, utilizando los métodos G4-STD y G4-nitro-cerrado-B3LYP. Valores en $\text{kJ} \cdot \text{mol}^{-1}$ . . . . .	89

---

6.6. Diferencias entre $\Delta_f H_{298.15K}^\circ$ experimental y calculado para un conjunto de moléculas sin grupos nitro ni heterocíclicos nitrogenados, utilizando los métodos G4-STD, G4/ReoptSO67 y G4-nitro-cerrado-B3LYP, reportados en $\text{kJ} \cdot \text{mol}^{-1}$ . . . . .	91
6.7. Comparación de $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$ obtenidas con diferentes métodos para sistemas atípicos. . . . .	92
6.8. Diferencias entre $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$ obtenidas mediante los métodos G4-STD y G4-nitro-cerrado-B3LYP. . . . .	94
A.1. Propiedades termodinámicas de moléculas, las referencias experimentales para las moléculas 1–67, y para las moléculas 68–110. . . . .	100

## Introducción

En las últimas décadas, la química computacional se ha consolidado como una subdisciplina esencial y autónoma dentro de la química moderna. Esto se debe a que los estudios teóricos permiten predecir resultados experimentales, comprender los mecanismos de reacción y, de manera práctica, optimizar condiciones experimentales. Dicha optimización se realiza mediante el cálculo de perfiles de energía libre y su relación con magnitudes observables, como las constantes de equilibrio  $K(T) = \exp[-\Delta G^\circ(T)/RT]$  y de velocidad. Estas herramientas permiten explorar de forma sistemática el efecto del solvente, la temperatura, la presión o el catalizador sobre la viabilidad de una reacción, guiando así la elección de condiciones experimentales más favorables [1–4]. Dentro de las diversas metodologías disponibles, los métodos *ab initio* son los más rigurosos, ya que se fundamentan en las ecuaciones de la mecánica cuántica sin recurrir a parámetros empíricos [5].

Desde su surgimiento a comienzos del siglo XX, la mecánica cuántica ha progresado hasta transformarse en un instrumento esencial en la química contemporánea. Sin embargo, su aplicación en cálculos termoquímicos y espectroscópicos resulta muchas veces restrictiva debido a su alto costo computacional. Como expresó P. Dirac:

"Las leyes físicas subyacentes necesarias para el tratamiento matemático de una

gran parte de la física y la totalidad de la química son completamente conocidas, y la dificultad recae sólo en el hecho de que la aplicación de estas leyes suponga resolver ecuaciones demasiado complejas" [6].

La cita anterior nos muestra la necesidad de continuar con el desarrollo de métodos aproximados que ayuden a resolver problemas químicos con buena precisión y costos de computación razonables. A lo largo del siglo XX, se lograron avances significativos para resolver estas ecuaciones numéricamente bajo ciertas aproximaciones. Podemos citar el método de Hartree-Fock (HF) [7], el de interacción de configuraciones (CI, por sus siglas en inglés) [8] y el método de cúmulos acoplados (CC, por sus siglas en inglés) [9]. Muchas de estas ideas evolucionan continuamente en la actualidad. Los métodos comúnmente utilizados pueden alcanzar una precisión de menos de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  [10], para sistemas pequeños a moderados [11]. Sin embargo, en la práctica, métodos como CI y CC utilizando un conjunto de bases considerable resultan inviables para muchos sistemas de interés.

Una alternativa a los métodos *ab initio* ha sido el desarrollo de la teoría funcional de la densidad (DFT, por sus siglas en inglés), que ofrece un buen equilibrio entre precisión y costo computacional. No obstante, los métodos de la DFT no siempre son mejores en comparación con los métodos *ab initio*. Dentro de la DFT, la "escalera de Jacob" representa una jerarquía de métodos y aproximaciones [12], pero el mejor funcional en esta jerarquía no puede alcanzar el límite del método interacción de configuraciones completa (CI) debido a la ausencia de conocimiento del funcional exacto de intercambio-correlación. Esta limitación ha llevado al desarrollo de diversas aproximaciones para el tratamiento de esta expresión. La elección del funcional adecuado depende de la experiencia y criterio del investigador, ya que cada uno tiene fortalezas y debilidades específicas según el tipo de molécula o propiedad investigada. Además, algunos funcionales incluyen parámetros empíricos ajustados a datos experimentales o cálculos *ab initio*. Un ejemplo destacado es el funcional B3LYP [13], que es quizás el más popular y ampliamente utilizado en la DFT. Este funcional combina contribuciones teóricas con parámetros empíricos para mejorar su precisión en la predicción de propiedades moleculares, energías y geometrías en equilibrio, lo que ha contribuido a su aceptación en la comunidad científica. No obstante, al usar este funcional para obtener propiedades termoquímicas como la entalpía de formación

( $\Delta_f H^\circ$ ) es insuficiente, ya que B3LYP no representa con precisión las energías electrónicas [13]. Debido a estas limitantes, la comunidad experimental de termoquímica ha impulsado el uso de métodos de estructura electrónica para determinar propiedades termoquímicas, con un costo computacional moderado y con resultados lo suficientemente confiables sin superar un cierto límite de precisión, establecido en  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  [14].

La entalpía de formación es una propiedad termodinámica fundamental que se define como el cambio de entalpía asociado a la reacción de formación de un compuesto a partir de sus elementos constituyentes más estables en sus estados estándar (habitualmente a 1 bar y 298.15 K) [15]. Esta desempeña un papel fundamental en el campo de la química al representar la energía vinculada a distintas estructuras moleculares que influyen en el comportamiento químico de las sustancias. Por consiguiente, es vital determinar esta propiedad de manera precisa para su utilización en una amplia gama de áreas que abarcan desde la síntesis hasta el diseño de reactores en ingeniería [16]. No sólo se busca obtener un valor, sino también una alta precisión en su estimación.

Para abordar esta complejidad, la precisión en el cálculo teórico de la entalpía de formación se ha convertido en un objetivo prioritario. La evolución continua de los métodos compuestos probablemente ha sido la respuesta a estas exigencias. Estos métodos emergen como una alternativa al uso de métodos tan robustos como CI y CC. Los métodos compuestos integran múltiples cálculos de bajo costo computacional para aproximar resultados de alto nivel como un Full CI utilizando diversos conjuntos de bases, representando una opción económica desde el punto de vista computacional para lograr estimaciones precisas con un costo reducido. Dentro de estos métodos, los métodos gaussianos (Gn) son fundamentales para aproximar con precisión la energía de sistemas mediante cálculos *ab initio* usando distintos niveles de precisión y conjuntos de bases. La serie incluye las versiones G1 [17], G2 [18], G3 [19] y G4 [20], cada una mejorando la precisión y reduciendo las desviaciones a menos de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  en el conjunto de calibración G2. Por ejemplo, el método G4 integra cálculos como CCSD(T), junto con métodos perturbativos como (MP2 y MP4) y ajustes adicionales como la Corrección de Alto Nivel (HLC, por sus siglas en inglés) y factor de escala en los modos normales de vibración B3LYP/6-31G(2df,p)(para ver más detalle, consulte la sección 5.1). Estos últimos serán de

crucial relevancia para esta tesis. Los métodos Gn han sido evaluados mediante conjuntos de calibración como G2/97 [21] y G3/05 [22]; estos han demostrado eficacia en la predicción de propiedades termodinámicas, aunque muestran desviaciones importantes en sistemas como metales de transición y ciertos grupos funcionales [23].

El método G4, desarrollado por Curtiss *et al.* [20], es el último de la serie Gn. Este método, basado en cálculos que buscan recuperar la mayor parte de la correlación electrónica, como CCSD(T), ha demostrado una precisión excepcional en la predicción de entalpía estándar de formación a 298.15 K en fase gaseosa ( $\Delta_f H_{298.15\text{K}}^\circ$ ) para un amplio espectro de moléculas pequeñas. No obstante, su aplicación en moléculas de mayor tamaño enfrenta retos computacionales significativos. Además, estudios recientes han evidenciado que, al aplicarlo a moléculas con ciertos grupos funcionales nitro [24,25], pueden aparecer errores sistemáticos que superan el límite de precisión química de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$ . Moléculas como nitrofenilos y heterociclos desafían las capacidades actuales de la química computacional [25], presentando complicaciones en la determinación precisa de  $\Delta_f H_{298.15\text{K}}^\circ$ . Estos compuestos, con aplicaciones que van desde explosivos hasta medicamentos y materiales avanzados [26], plantean retos particulares para los métodos computacionales, especialmente al emplear el método de atomización.

La efectividad de los métodos compuestos para predecir  $\Delta_f H_{298.15\text{K}}^\circ$  también depende del método termoquímico implementado. La metodología empleada, se basa en las energías de atomización, que consideran la diferencia energética entre una molécula y sus átomos aislados, todos en estado gaseoso. Aun así, este enfoque puede ser susceptible a una descripción inadecuada de la correlación electrónica, ya que el uso de métodos como el de interacción de configuraciones completo no es viable para muchos sistemas de interés, y en su lugar se emplean métodos aproximados que no recuperan toda la correlación electrónica. Para superar estas limitaciones, se han propuesto esquemas alternativos como las reacciones isodésmicas [27], que utilizan la cancelación de errores al comparar energías moleculares experimentales en lugar de energías atómicas, permitiendo estimaciones más precisas de  $\Delta_f H_{298.15\text{K}}^\circ$  incluso con niveles de teoría como *ab initio* o DFT.

Existen otras metodologías compuestas *ab initio* para el cálculo de entalpías de formación, pero

son computacionalmente más costosas que los Gn. Las técnicas W1–W4 [28–30], desarrolladas por Martin y *et al.*, y los métodos de extrapolación de Dixon, Feller y Peterson (FPD) [31, 31], se basan en CCSD(T) y extrapolaciones al límite de conjunto de base completo (CBS, por sus siglas en inglés), logrando una precisión de hasta  $0.41 \text{ kJ} \cdot \text{mol}^{-1}$ . No obstante, su aplicación se limita principalmente a moléculas pequeñas, lo que las hace inviables para sistemas de mayor interés. Otras metodologías relacionadas incluyen los métodos CBS de Petersson y *et al.* [32,33], los métodos multicoeficientes de Truhlar [34] y las aproximaciones compuestas consistentes en correlación (ccCA) [35] y HEAT [36,36].

En esta tesis, evaluaremos el desempeño del método G4 en la predicción de entalpías de formación para nitrofenilos y heterociclos nitrogenados. Investigaremos la modificación del método G4 en el factor de escala  $F$  y el parámetro  $A$  en el término HCL con el objetivo de mejorar la concordancia teórico-experimental para estos sistemas, que presenta desafíos significativos al emplear el método de atomización.

## Antecedentes

Antes de seleccionar una metodología de cálculo, es esencial reconocer las diversas variantes del método compuesto G4 y el estado actual de los cálculos teóricos de la entalpía de formación para nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados. Existen diversas modificaciones y variantes a los métodos compuestos que pueden modificar el valor teórico calculado de la entalpía de formación.

### **2.1. Introducción al método compuesto G4**

En este contexto, el método Gaussian-4 (G4) representa un avance significativo respecto a sus predecesores, como los métodos Gaussian-3 (G3) [19] y G3(MP2) [37].

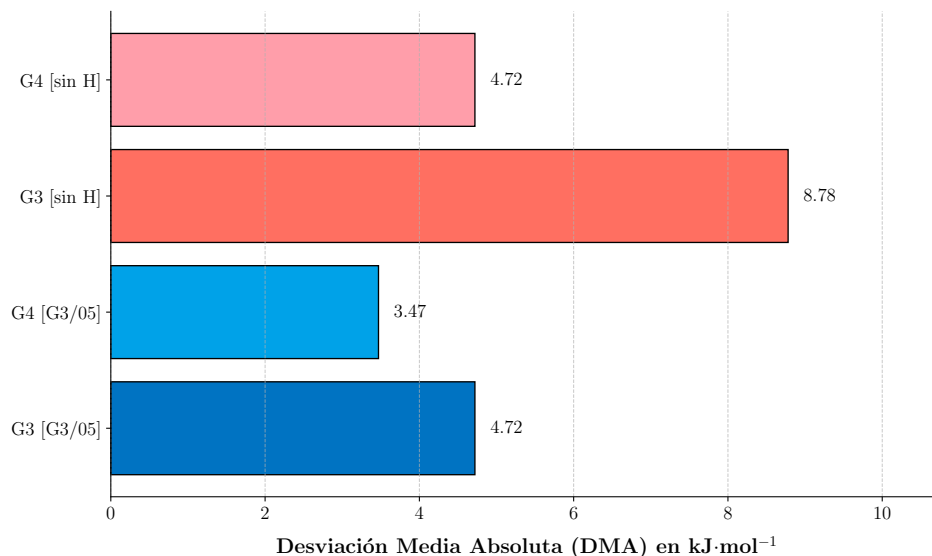
El método G4 surgió para disminuir las desviaciones entre los valores teórico-experimentales de las versiones previas. En particular, se buscó mejorar la precisión en el conjunto de calibración G3/05 [38]. Los métodos anteriores mostraban restricciones en capacidad y aplicabilidad para ciertas moléculas, lo que motivó la creación de G4. Las mejoras implementadas en G4 incluyen cambios significativos al método G3 [19]. Entre estas, se destaca la estimación del límite de energía Hartree-Fock (HF), que optimiza la energía de los cálculos a base infinita [20]. Además,

se amplió el conjunto de funciones de polarización, conocido como G3LargeXP. Este conjunto de bases es particular de Gaussian y no se encuentra publicado en bases de datos.

Otra mejora notable fue la incorporación del método CCSD(T) en lugar de QCISD(T) para el mejor cálculo de la correlación electrónica. También se usaron las geometrías y energías de punto cero calculadas con el funcional B3LYP/6-31G(2df,p), mejorando los resultados globales, especialmente para moléculas no hidrogenadas (ejemplos claros: O<sub>3</sub>, SO<sub>2</sub>, SO<sub>3</sub>, SF<sub>6</sub>) [20] donde MP2 generaba geometrías menos realistas [39]. Finalmente, la corrección de alto nivel HLC y el factor de escala  $F$  fueron reparametrizados. En particular, el factor de escala tiene un valor de 0.9854 y se utiliza para ajustar las correcciones de energía de punto cero y térmicas. Mientras la corrección a alto nivel es un parámetro empírico que se modifica para mejorar la concordancia con los datos experimentales. El factor de escala se determinó a partir de las frecuencias vibracionales experimentales calculadas en un conjunto de moléculas, basándose en el trabajo de Scott y Radom [40]. Este ajuste permite que las correcciones de energía de punto cero se alineen de manera más precisa con un conjunto de los valores experimentales.

La Figura 2.1 muestra la mejora del método G4 en comparación con G3 en el conjunto de calibración G3/05. Esta evaluación se llevó a cabo usando el conjunto de calibración G3/05 [22], que contiene 454 energías experimentales, como entalpías de formación, energías de ionización y afinidades electrónicas y protónicas. La desviación media absoluta (DMA) de G4 disminuyó significativamente en comparación con G3, pasando de  $4.72 \text{ kJ} \cdot \text{mol}^{-1}$  a  $3.47 \text{ kJ} \cdot \text{mol}^{-1}$ . Esta mejora se observa en las entalpías de formación de sistemas sin hidrógeno, donde el error bajó de  $8.78 \text{ kJ} \cdot \text{mol}^{-1}$  en G3 a  $4.72 \text{ kJ} \cdot \text{mol}^{-1}$  en G4. Ejemplos de sistemas sin hidrógeno son: O<sub>3</sub>, CO, CO<sub>2</sub>, etc.

Curtiss *et al.* [41] evaluaron la precisión del método G4 y sus variantes en el cálculo de barreras de energía para reacciones de transferencia de hidrógeno y sin hidrógeno. Compararon las geometrías optimizadas con B3LYP/6-31G(2df,p) para analizar su impacto en la capacidad predictiva de los resultados. Para esta nueva variante recalibraron nuevamente el término de corrección de alto nivel (HLC) para ajustar mejor a las barreras experimentales [41].



**Figura 2.1:** Comparación del rendimiento del método G4 respecto a G3 en el conjunto de prueba G3/05 y subconjunto [sin H].

De manera paralela, se desarrolló una variante adicional denominada G4(MP2) [42], que reemplaza cálculos de nivel MP4 por MP2; Este cambio permitió sustituir los cálculos computacionalmente costosos de la teoría de perturbaciones de tercer y cuarto orden por aproximaciones de segundo orden. La simplificación lograda facilita obtener resultados precisos para datos termoquímicos a un costo computacional reducido, especialmente en comparación con G4. En evaluaciones realizadas con el conjunto G3/05, G4(MP2) mostró una desviación media absoluta de  $4.35 \text{ kJ}\cdot\text{mol}^{-1}$ . Fue más preciso que G3 y G3(MP2), pero menos preciso que G4, que tuvo una DMA de  $3.47 \text{ kJ}\cdot\text{mol}^{-1}$ . Este estudio destacó a G4(MP2) como una opción eficiente para cálculos termodinámicos, equilibrando precisión y costo computacional.

Recientemente, Dandu *et al.* [43] modificaron el método G4(MP2) mediante la integración de correcciones específicas por átomo, dando lugar al desarrollo del método G4(MP2)A. Este nuevo enfoque alcanzó una precisión comparable al método G4 con un costo computacional aún menor. Los resultados de estas mejoras se integraron con algoritmos de aprendizaje automático para predecir energías moleculares de manera más eficiente. Las calibraciones de estas correcciones se realizaron utilizando el conjunto de datos GDB9, compuesto por 9000 moléculas. Sin embargo, se excluyeron 191 moléculas debido a desviaciones superiores a  $8.3 \text{ kJ}\cdot\text{mol}^{-1}$  en las entalpías de formación teóricas [43].

Otro avance significativo es el método G4CEP desarrollado por Silva *et al.* [44], que amplió la aplicabilidad del método G4 a sistemas con elementos más pesados. Este desarrollo implementó el uso de pseudopotenciales (compact effective potential, CEP, por sus siglas en inglés) y reparametrizó los parámetros de corrección de alto nivel (HLC) para optimizar la predicción de propiedades termoquímicas. Se llevaron a cabo ajustes específicos en los parámetros de elementos de la tercera fila (As, Se, Br), optimizando las funciones de valencia y los exponentes de las bases de cálculo para minimizar errores energéticos. Esta metodología mejoró significativamente los cálculos para moléculas con elementos de mayor número de electrones.

Mayhall *et al.* [45] realizaron una reparametrización del método G4 para optimizar su desempeño en el cálculo teórico de entalpías de formación para metales de transición. La versión modificada y denominada G4(rel) y G4(MP2, rel) incorpora correcciones relativistas escalares y un conjunto de bases ampliado para átomos de metales de transición, denominado G3LargeXP. Adicionalmente, los parámetros de la corrección de alto nivel (HLC) fueron ajustados, lo que redujo los errores en sistemas con elementos de transición del primer periodo (Sc-Zn). Además evaluaron ambos métodos mediante validación cruzada en un conjunto de prueba compuesto por 20 moléculas representativas de metales de transición. G4(MP2, rel) mostró un rendimiento superior comparado con G4 y G4(MP2), con una desviación media absoluta (DMA) de  $11.88 \text{ kJ} \cdot \text{mol}^{-1}$ , significativamente menor a la de G4(rel), que alcanzó  $17.15 \text{ kJ} \cdot \text{mol}^{-1}$ . Este desempeño se atribuye a la mayor eficiencia de las aproximaciones aditivas en la teoría de perturbaciones de Møller-Plesset. Sin embargo, ciertos compuestos, como  $\text{CrO}_3$ , presentaron desviaciones más altas, lo que resalta limitaciones específicas del modelo.

Grimme propuso las variantes SCS-MP2 y SOS-MP2 como mejoras significativas al método MP2 tradicional, ajustando selectivamente las contribuciones energéticas por espín para corregir deficiencias en la descripción de correlación electrónica, además de incluir nuevamente el término HLC en la calibración de ambos métodos. Estos métodos mostraron una precisión superior en el cálculo de entalpías de formación en moléculas grandes y sistemas químicamente complejos (DMA =  $5.02 \text{ kJ} \cdot \text{mol}^{-1}$ ), superando claramente al método MP2 y al funcional B3LYP (DMA =  $8.78 \text{ kJ} \cdot \text{mol}^{-1}$ ) evaluados en el conjunto de calibración G3/05 [46]. Esta implementación también hace uso del término HLC adaptado a ese conjunto particular para obtener esa

mejora en términos de la DMA.

Los desarrollos anteriores muestran que, a pesar de que G4 es un método compuesto que en principio es aplicable a una amplia gama de sistemas, presenta limitaciones. La elección de un método compuesto no debe basarse únicamente en la concordancia con valores experimentales, sino también en la capacidad del método para capturar las características del sistema químico. En la última década, se ha utilizado un enfoque común que consiste en promediar los resultados de distintos métodos compuestos para aumentar la precisión en la medición teórica de energías mediante atomización. No obstante, el uso del método de reacciones isodésmicas ha obtenido una mayor aceptación debido a que puede dar valores más cercanos al experimental [47].

En este contexto, Simmie y Somers [48] evaluaron la exactitud de diversos métodos compuestos, como CBS-QB3, CBS-APNO, G3, G4 y W1BD, en el cálculo de las entalpías de formación de moléculas de la forma  $C_xH_yO_z$  presentes en las Tablas Termoquímicas Activas (ATcT) [49]. Los resultados mostraron una concordancia óptima entre los valores promedio de los diferentes métodos compuestos anteriormente citados y valores experimentales, aunque se detectaron discrepancias notables en casos específicos, como el 1,3-ciclopentadieno. La integración de los métodos CBS-APNO con G3 o G4 permitió reducir errores medios y desviaciones, aproximándose a la “precisión química” sin incurrir en los elevados costos computacionales de métodos más complejos como W1BD.

Adicionalmente, Simmie y Sheahan [50] continuaron esta línea de investigación, validando bases de datos de entalpías de formación mediante métodos compuestos de costo computacional de nivel medio. Compararon sus resultados con los valores de referencia de las Tablas Termoquímicas Activas (ATcT) y detectaron discrepancias en casos como el de la hidrazina, subrayando la importancia de emplear un método robusto como G4 para evitar errores sistemáticos.

En un análisis más reciente, Simmie y Somers [51] exploraron si una combinación de métodos compuestos podía proporcionar una mayor precisión que el uso aislado de cada uno. Encontraron que la combinación de G3 y G4 con CBS-APNO mejoraba significativamente la precisión sin

aumentar los costos computacionales. Esta estrategia permitió alcanzar una desviación media inferior a  $4.183 \text{ kJ} \cdot \text{mol}^{-1}$ , consolidándose como una opción viable para aplicaciones en química computacional sin recurrir a metodologías más costosas como W1BD.

Métodos más robustos, como W1, son generalmente preferibles al emplear el método de atomización en cálculos teóricos de entalpías de formación. Aun así, surge la interrogante: ¿en qué condiciones o sistemas el método G4 es una opción aceptable frente a métodos como Wn?. Estudios como el de Rayne y Forest [52] han demostrado que G4 puede ofrecer resultados comparables o superiores al método W1BD en la estimación de entalpías estándar de formación ( $\Delta_f H_{298.15\text{K}}^\circ$ ). En un análisis de 54 compuestos, G4 mostró desviaciones absolutas medias de  $3.2 \text{ kJ} \cdot \text{mol}^{-1}$  a  $8.2 \text{ kJ} \cdot \text{mol}^{-1}$ , mientras que W1BD registró desviaciones de 5.6 a  $12.9 \text{ kJ} \cdot \text{mol}^{-1}$ . El anterior estudio muestra que G4 es una opción válida y práctica para estudios termoquímicos y es particularmente útil para sistemas sin alta complejidad conformacional.

## 2.2. Controversia en nitrofenilos y heterociclos

Los compuestos de nitrofenilo y los heterociclos sulfurados, oxigenados y nitrogenados son ejemplos de sistemas moleculares que presentan retos significativos en el cálculo de entalpías de formación. Los nitrofenilos, caracterizados por sus grupos nitro que atraen electrones, manifiestan efectos de inducción y resonancia a través de los anillos aromáticos, impactando su estructura electrónica y estabilidad. Por otro lado, los heterociclos, que contienen átomos distintos del carbono en sus anillos, suelen presentar características electrónicas y conformaciones singulares.

Un sistema molecular ilustrativo es el nitrobenzono, un compuesto aromático con un grupo funcional nitro ( $-\text{NO}_2$ ). Este compuesto ha generado debates en la comunidad científica debido a las discrepancias entre los valores de entalpía de formación obtenidos teóricamente y experimentalmente. Lebedeva *et al.* [53, 54] reportaron un valor experimental de  $67.5 \pm 1.4 \text{ kJ} \cdot \text{mol}^{-1}$ , mientras que otros estudios indican una entalpía de formación de  $62.5 \pm 0.67 \text{ kJ} \cdot \text{mol}^{-1}$  [49].

Cálculos teóricos previos, realizados por Dorofeeva y Moiseeva [55] utilizando el método G3X, arrojaron un valor de  $56 \text{ kJ} \cdot \text{mol}^{-1}$ , cuestionando así los valores experimentales más recientes. Investigaciones posteriores empleando el método de atomización con métodos compuestos como G4 y W1 obtuvieron valores de  $56.5 \text{ kJ} \cdot \text{mol}^{-1}$  y  $53.4 \text{ kJ} \cdot \text{mol}^{-1}$ , respectivamente, según Verevkin *et al.* [24]. Pese a ello, el valor experimental sugerido actualmente es el reportado por Pedley [56], de  $67.53 \pm 0.4 \text{ kJ} \cdot \text{mol}^{-1}$ .

Estas discrepancias evidencian un debate sobre el valor experimental adecuado para la entalpía de formación del nitrobenzono, dificultando la reconciliación teórico-experimental mediante el uso de métodos compuestos. Teóricamente, Verevkin y *et al.* [24] calcularon con G4 y, usando reacciones isodésmicas, reportaron un valor de entalpía de formación de  $64.8 \text{ kJ} \cdot \text{mol}^{-1}$ . Estas discrepancias reportadas evidencian una limitación crítica en la precisión de los métodos actuales, lo que compromete la confianza en los resultados teóricos y su utilidad en validar o recomendar un valor experimental.

La determinación de la entalpía de formación del nitrometano también presenta incertidumbres significativas. Los valores experimentales varían notablemente, oscilando entre  $-81.0 \pm 1.0 \text{ kJ} \cdot \text{mol}^{-1}$  (valor sugerido) [54] y  $-74.5 \pm 0.8 \text{ kJ} \cdot \text{mol}^{-1}$  [49]. Esta variabilidad resalta la complejidad en determinar con precisión la entalpía de formación del nitrometano. Los valores teóricos obtenidos mediante reacciones isodésmicas con G4 indican  $-73.2 \text{ kJ} \cdot \text{mol}^{-1}$  [24], alineándose con el valor sugerido. No obstante, el método de atomización usando G4 y W1U proporciona valores de  $-78.1 \text{ kJ} \cdot \text{mol}^{-1}$  y  $-76.5 \text{ kJ} \cdot \text{mol}^{-1}$ , respectivamente [24].

Estos ejemplos ilustran los desafíos que enfrentan los métodos compuestos, como G4, en el cálculo teórico de entalpías de formación. En el caso de los heterociclos nitrogenados, Verevkin *et al.* examinaron las propiedades termoquímicas de las diazinas (piridazina, pirimidina y pirazina) mediante mediciones experimentales y métodos compuestos. Compararon las entalpías de formación calculadas con métodos como G3, G4, CBS-APNO y W1(RO), evidenciando que G4 se distingue por su buen acuerdo teórico-experimental en la predicción de propiedades termoquímicas en heterociclos nitrogenados [57].

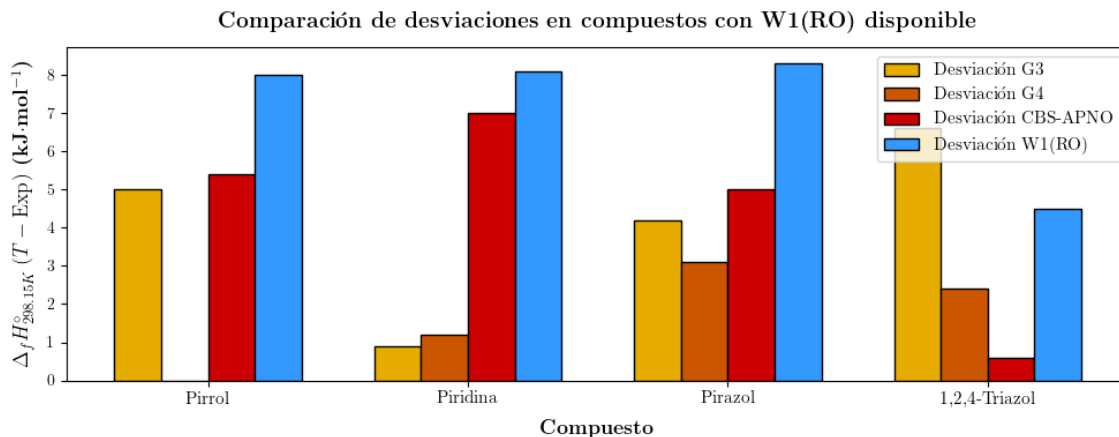
Además, Verevkin *et al.* [57] señalaron que las desviaciones sistemáticas pueden corregirse mediante un enfoque modificado de atomización, ya que se encontró una correlación entre las entalpías de atomización y las entalpías experimentales, evitando la necesidad de usar el método de reacciones isodérmicas. Los hallazgos revelaron inconsistencias en datos experimentales previos, logrando resolver variaciones de hasta 6 a 7  $\text{kJ} \cdot \text{mol}^{-1}$  en las entalpías de formación de pirazina y pirimidina. La implementación de G4 y W1(RO) permitió minimizar los errores en la predicción de entalpías a un rango de 1  $\text{kJ} \cdot \text{mol}^{-1}$  a 2  $\text{kJ} \cdot \text{mol}^{-1}$ .

La Figura 2.2 muestra las discrepancias entre las entalpías experimentales y las calculadas en compuestos como pirrol, piridina, pirazol, 1,2,4-triazol, usando diferentes métodos computados, los cuales son G3, G4, CBS-APNO y W1(RO). Según Verevkin *et al.* [24] reportó que, G4 proporciona mayor exactitud en las predicciones, especialmente en compuestos como el pirrol y la piridina. Aunque W1(RO) tiene un costo computacional más alto, también ofrece resultados aceptables, siendo útil cuando los recursos lo permiten. Aun así, en sistemas como el pirazol y el 1,2,4-triazol, G4 demostró un rendimiento superior, sugiriendo que W1(RO) no siempre supera a G4 en términos de acercarnos al valor experimental.

En este contexto, la notación  $\Delta_f H_{298.15 \text{ K}}^\circ(T - \text{Exp})$  se define como la desviación en la entalpía estándar de formación a 298.15 K, calculada como:

$$\Delta_f H_{298.15 \text{ K}}^\circ(T - \text{Exp}) = \Delta_f H_{298.15 \text{ K}}^\circ(T) - \Delta_f H_{298.15 \text{ K}}^\circ(\text{Exp}). \quad (2.1)$$

donde  $T$  representa el valor obtenido con el método teórico y Exp el valor experimental recomendado.



**Figura 2.2:** Comparación de las desviaciones en las entalpías de formación estándar  $\Delta_f H_{298.15\text{K}}^\circ(T - Exp)$ , utilizando los métodos G3, G4, CBS-APNO y W1(RO) para compuestos seleccionados. La desviación se calcula como la diferencia entre los valores experimentales y teóricos.

He *et al.* [58] también evaluaron la precisión de siete métodos compuestos (G1, G2, G3, G4, CBS-4M, CBS-Q y CBS-QB3) en el cálculo de las entalpías de formación de 63 compuestos nitrogenados. G4 fue el método más preciso, con una DMA de  $3.47\text{ kJ}\cdot\text{mol}^{-1}$  utilizando el método de atomización, y al emplear el método de reacciones isodésmicas, la DMA disminuyó a  $2.63\text{ kJ}\cdot\text{mol}^{-1}$ . La precisión de G4 mejoró aún más con el método de reacciones isodésmicas. Este método ha mostrado mejores resultados en algunos casos comparado con el de atomización. Al seleccionar cuidadosamente las reacciones, el método isodésmico puede ser tan preciso o incluso más que otros métodos, ya que compensa las deficiencias en el conjunto de bases y la correlación [59].

Estas explicaciones tienden a pasar por alto el hecho de que no siempre se ha mostrado que el método de reacciones isodesmicas dé mejores resultados que el de atomización, por ejemplo: D. Bond [60] investigó la exactitud de diferentes métodos compuestos para un conjunto de 300 moléculas orgánicas. Su estudio comparó varios métodos de alto nivel, incluidos G2, G3 y métodos basados en la teoría del funcional de la densidad. La investigación reveló que el método de atomización a menudo proporcionaba resultados más precisos que el método de reacciones isodésmicas, particularmente en moléculas pequeñas. El método compuesto G3 presentó una desviación media absoluta de  $3.0\text{ kJ}\cdot\text{mol}^{-1}$  en las entalpías de formación. Si bien el método de reacciones isodésmicas ofreció resultados consistentes para moléculas más grandes, el método

de atomización mantuvo su precisión en una gama más amplia de compuestos. Recientemente, Yang Li *et al.* [61] llevaron a cabo un estudio sobre las propiedades termodinámicas de hidrocarburos insaturados y sus radicales, comparando el método de atomización con el de reacciones isodésmicas. Aunque el método de reacciones isodésmicas demostró ser más consistente con los valores experimentales, especialmente para moléculas de mayor tamaño, el método de atomización también fue confiable, con resultados dentro de un margen de  $0.83 \text{ kJ} \cdot \text{mol}^{-1}$ . Por lo tanto, a partir de estos trabajos, no podemos concluir que el método de reacciones isodésmicas es siempre más preciso que el de atomización, y en algunos casos los resultados obtenidos con ambos métodos son prácticamente iguales. Además, la desventaja clara del método de reacciones isodésmicas es que depende fuertemente de la elección específica de las reacciones propuestas para calcular con precisión la entalpía de formación. Por lo tanto, no existe certeza de que este método resuelva adecuadamente las deficiencias del método de atomización, ni de que se aproxime al valor experimental de la entalpía por las razones correctas.

Los anteriores ejemplos muestran que los compuestos nitrofenilos y heterociclos oxigenados, sulfurados y nitrogenados, son sistemas particularmente problemáticos para el cálculo teórico de la entalpía de formación usando el método de atomización. En la literatura se observan dos hipótesis que podrían explicar la complejidad para obtener la entalpía a 298.15 K en compuestos con el grupo funcional nitro ( $-\text{NO}_2$ ). La primera hipótesis propuesta por Dorsett y White [62], sugiere que las limitaciones de los métodos compuestos, especialmente G2, para modelar moléculas con grupos nitro, se deben a la complejidad de su estructura electrónica. La presencia de grupos nitro genera efectos significativos de correlación no dinámica, atribuibles a energías similares entre los orbitales HOMO y LUMO, complicando su descripción mediante métodos de un solo determinante de Slater. Una descripción adecuada podría requerir tratamientos multi-configuracionales que incluyan orbitales ocupados y desocupados.

La segunda hipótesis se basa en una posible geometría de equilibrio inapropiada para estos sistemas. Tikhonov *et al.* [63] investigaron la formación de la nitroxolina (NXN) en fase gaseosa mediante técnicas experimentales y cálculos computacionales, identificando discrepancias entre las geometrías teóricas y experimentales. Los métodos basados en DFT predicen que la geometría de esta molécula es plana con respecto al anillo aromático y el grupo funcional  $\text{NO}_2$ ,

coincidiendo con los hallazgos experimentales, mientras que métodos *ab initio* como MP2 y CCSD(T) tienden a prever una estructura no plana.

Se ha propuesto que utilizar el funcional de densidad B3LYP en las etapas iniciales del método G4 es adecuado para asegurar una geometría de mínima energía y, consecuentemente, proporcionar resultados precisos de entalpía. Sin embargo, la investigación anterior mencionada indica que la geometría obtenida con B3LYP podría diferir de la obtenida con métodos más robustos como CCSD(T). Esta discrepancia podría resultar en que, aunque la geometría B3LYP sea la más estable, la energía calculada mediante CCSD(T) no corresponda a un mínimo, y viceversa.

Un trabajo particularmente importante para esta tesis es el de Dorofeeva *et al.* [25], quienes examinaron la eficacia del método G4 en la determinación de entalpías de formación en fase gaseosa para 122 moléculas orgánicas pequeñas y medianas, incluyendo nitrofenilos y heterocíclicos con nitrógeno, oxígeno y azufre. Este trabajo reveló que la desviación media absoluta (DMA) más alta, de  $10.7 \text{ kJ} \cdot \text{mol}^{-1}$ , se observó en los compuestos con grupos funcionales nitro ( $\text{NO}_2$ ), evidenciando una tendencia del método G4 a mostrar diferencias entre las estimaciones teóricas y experimentales de las entalpías de formación de estos compuestos entre 5 y  $15 \text{ kJ} \cdot \text{mol}^{-1}$ . Para los heterociclos nitrogenados, la DMA fue de  $6.8 \text{ kJ} \cdot \text{mol}^{-1}$ , menor en comparación con los compuestos nitro, pero aún con desviaciones significativas respecto a los valores experimentales. Los heterociclos que contienen oxígeno y azufre mostraron una DMA mayor, de  $9.1 \text{ kJ} \cdot \text{mol}^{-1}$ . Es importante destacar que, contrariamente a lo esperado, no se observó una acumulación significativa de errores sistemáticos al aumentar el tamaño molecular en los sistemas estudiados, que incluían hasta 15 átomos distintos de hidrógeno. En contraste, indicaron que las deficiencias del método G4 son más evidentes en moléculas con múltiples heteroátomos y enlaces conjugados, donde la determinación precisa de las energías de atomización es más compleja y propensa a errores [25].

Estos hallazgos subrayan la necesidad de realizar ajustes empíricos adicionales para optimizar la capacidad del método en compuestos que no están adecuadamente representados en el conjunto de calibración original.

Surgen entonces las siguientes preguntas: ¿Es posible modificar el método G4 para mejorar la energía electrónica calculada y, al mismo tiempo, aumentar la precisión en las entalpías de formación utilizando el método de atomización?. Los métodos compuestos  $G_n$  tienen parámetros internos, calibrados mediante conjuntos de datos específicos. ¿Cuál es la relevancia de los parámetros internos de G4, particularmente el factor de escala  $F$  y el término  $A$  de la corrección de alto nivel (HLC, Higher-Level Correction, por sus siglas en inglés), en el cálculo teórico de la entalpía de formación?. Para ver más detalle de estos términos y parámetros  $F$  y  $A$  y cómo actúan en el método G4, se recomienda revisar la sección 5.1.

Si se modifica el conjunto de calibración, los valores de  $F$  y  $A$  podrían requerir un reajuste para optimizar su aplicabilidad en conjuntos de prueba específicos. Esto plantea interrogantes sobre la generalización de ciertos supuestos metodológicos, como los señalados por John Pople en el desarrollo del método G1, donde señaló que: "La aplicación uniforme de la HLC sólo es razonable si el conjunto de bases utilizado, 6-311+G\*\* (2df), está equilibrado, es decir, si los errores residuales por electrón son aproximadamente constantes en una amplia gama de moléculas" [17]. En el contexto actual del desarrollo y uso de los métodos  $G_n$ , la validez de este principio podría depender de la naturaleza de los sistemas moleculares estudiados, lo que merecería investigación adicional. Un aspecto clave sería evaluar cómo los conjuntos de calibración influyen en los parámetros internos de G4, para el caso de sistemas de capa cerrada estudiar la dependencia de  $F$  y  $A$ , especialmente si su transferibilidad a sistemas químicamente distintos a los de la calibración original podría estar limitada.

Recientemente, Das *et al.* [64] realizaron un estudio sobre la precisión de diversos métodos compuestos (G4, G4(MP2), ccCA, CBS-QB3) y varias aproximaciones de DFT en la predicción de entalpías de formación ( $\Delta_f H_{298K}^\circ$ ). Presentaron un extenso conjunto de datos que incluye 1796 compuestos y utilizaron métodos probabilísticos para eliminar datos con comportamientos atípicos o posibles incertidumbres experimentales. Para verificar la exactitud de las predicciones del método G4, calcularon los errores absolutos en las entalpías de formación predichas, empleando el error medio absoluto (MUE, por sus siglas en inglés) por electrón de valencia como parámetro estadístico para su evaluación. Savin *et al.* [65] destacan las dificultades y posibles trampas del benchmarking en DFT, enfatizando que la elección de métricas estadísticas —como

el error absoluto medio (MAE) y el error porcentual medio (MAPE)— puede llevar a interpretaciones divergentes. Los autores muestran que, dependiendo de la medida estadística seleccionada, la evaluación de la precisión de un funcional puede variar significativamente. Este análisis subraya la importancia de utilizar conjuntos de datos lo suficientemente grandes y de tener en cuenta la influencia de valores atípicos, de modo que las conclusiones sobre la calidad de los métodos computacionales sean robustas. Este estudio evidenció que el método G4 es confiable para un amplio espectro de moléculas orgánicas. No obstante, a pesar de incluir sistemas como nitrofenilos y heterociclos nitrogenados, se sugiere que la transferibilidad del término de corrección de alto nivel (HLC) puede no ser aplicable a todos los sistemas, especialmente si se busca que las diferencias entre valores teóricos y experimentales no excedan  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$ .

Considerando lo anterior estas problemáticas en este tipo de compuestos han sido observadas en nuestro grupo de investigación. A través de nuestra colaboración teórico-experimental con el grupo de termoquímica de la Benemérita Universidad Autónoma de Puebla (BUAP), hemos encontrado consistentemente que, al usar el método G4 con el método de atomización, se predicen diferencias significativas, superiores a  $8 \text{ kJ} \cdot \text{mol}^{-1}$ , respecto a los valores experimentales en sistemas que contienen C, N, O y H y particularmente en moléculas con grupos funcional(es)  $-\text{NO}_2$  [25].

## Hipótesis y Objetivos

### 3.1. Hipótesis

La implementación de modificaciones específicas en los pasos 1 y 2 del método G4, junto con la reoptimización de los parámetros empíricos existentes, como el factor de escala ( $F$ ) y el parámetro de corrección de alto nivel ( $A$ ), utilizando el método de atomización, puede mejorar la precisión en los cálculos de energía electrónica y, consecuentemente, en las entalpías de formación de sistemas problemáticos como los nitrofenilos y los heterociclos.

### 3.2. Objetivo General

Analizar el efecto de modificaciones específicas en las etapas iniciales del método G4 y la reparametrización de sus parámetros empíricos, determinando los valores óptimos del factor de escala ( $F$ ) y del parámetro de corrección de alto nivel ( $A$ ) en la corrección energética  $E_{HLC}$ , sobre la precisión del cálculo teórico de entalpías de formación mediante el método de atomización, con la finalidad de reducir las desviaciones observadas en sistemas problemáticos como nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados.

### 3.3. Objetivos Específicos

- Seleccionar y construir nuevos conjuntos de calibración, incluyendo sistemas del conjunto de prueba G3/05 y moléculas problemáticas específicas como nitrofenilos y heterociclos, que permitan evaluar la precisión del método G4 modificado.
- Realizar cálculos de estructura electrónica utilizando el software Gaussian 09 para cada uno de los conjuntos de calibración definidos, obteniendo los datos energéticos necesarios para evaluar las modificaciones en el método G4.
- Determinar los valores numéricos óptimos del factor de escala ( $F$ ) y del parámetro de corrección de alto nivel ( $A$ ) en la corrección energética  $E_{HLC}$  del método G4, mediante ajustes sobre los resultados obtenidos para los distintos conjuntos de calibración.
- Cuantificar el impacto de las modificaciones introducidas en el método G4 (reparametrización y cambio del método de estructura electrónica) calculando la desviación media absoluta (DMA) respecto a valores experimentales.
- Evaluar el desempeño global del método reparametrizado G4-nitro-cerrado-B3LYP comparándolo con el método original G4-STD, utilizando datos experimentales como referencia.
- Modificar el método G4 mediante la sustitución del funcional B3LYP en los pasos 1 y 2, explorando alternativas como  $\omega$ B97XD, M06-2X o el método MP2, y cuantificar su efecto en el cálculo de entalpías de formación.

## Fundamento Teórico

### 4.1. Métodos de la química computacional

La química computacional es una rama de la química que emplea métodos matemáticos y computacionales para describir y predecir el comportamiento de sistemas químicos a nivel molecular [4,66]. Esta disciplina resulta esencial para determinar propiedades fisicoquímicas como geometrías de mínima energía, estados de transición, constantes de velocidad, energías relativas y momentos dipolares. Además, permite analizar propiedades espectroscópicas, como desplazamientos químicos, constantes de acoplamiento y frecuencias vibracionales (IR). También permite validar resultados experimentales y estudiar fenómenos que serían costosos o inviables de abordar únicamente mediante métodos experimentales. Sin embargo, aún enfrenta desafíos y limitaciones [67]. Para obtener estas propiedades, se utilizan distintos enfoques que buscan aproximar la solución de la ecuación de Schrödinger independiente del tiempo:

$$\hat{H}\Psi(X) = E\Psi(X), \quad (4.1)$$

donde  $\hat{H}$  corresponde al operador hamiltoniano que abarca tanto la energía cinética como la potencial del sistema, mientras que  $E$  representa la energía total asociada al sistema. Para resolverla en sistemas moleculares se emplea la aproximación de Born–Oppenheimer, la cual separa los movimientos electrónicos y nucleares gracias a la gran diferencia de masas entre electrones

y núcleos.

El Hamiltoniano total se expresa como:

$$\hat{H}_{tot} = \hat{T}_e + \hat{T}_N + \hat{V}_{ee} + \hat{V}_{NN} + \hat{V}_{eN}, \quad (4.2)$$

donde  $\hat{T}_e$  y  $\hat{T}_N$  son los términos cinéticos de electrones y núcleos, mientras que  $\hat{V}_{ee}$ ,  $\hat{V}_{NN}$  y  $\hat{V}_{eN}$  describen las interacciones electrón–electrón, núcleo–núcleo y electrón–núcleo.

Por lo tanto, la función de onda electrónica se factoriza como:

$$\Psi(r, R) \approx \psi_e(r; R)\chi_N(R), \quad (4.3)$$

donde  $\psi_e(r; R)$  es la función electrónica dependiente de las coordenadas nucleares fijas, y  $\chi_N(R)$  describe el movimiento nuclear. El problema electrónico se resuelve primero con núcleos inmóviles:

$$\hat{H}_e(r; R)\psi_e(r; R) = E_e(R)\psi_e(r; R), \quad (4.4)$$

y la energía electrónica  $E_e(R)$  se incorpora posteriormente al Hamiltoniano nuclear. Esta aproximación constituye el fundamento de los métodos de estructura electrónica empleados en química computacional y define la superficie de energía potencial [7, 68].

Más allá del átomo de hidrógeno, la solución analítica de esta ecuación no es factible; es por esta razón que se requieren aproximaciones donde los métodos de química computacional juegan un papel fundamental. El método de Hartree–Fock (HF) es uno de los más utilizados y sirve como punto de partida para el desarrollo de técnicas más avanzadas, conocidas como métodos post-Hartree–Fock. Entre estos destacan la teoría de perturbaciones de Møller–Plesset (MPn) y el método de cúmulos acoplados (CC), los cuales superan las limitaciones del método HF al incorporar la energía de correlación electrónica. Por otro lado, la Teoría Funcional de la Densidad (DFT) se ha consolidado como una técnica popular gracias a su balance entre precisión y costo computacional. En contraste, ninguno de estos enfoques por sí solo ofrece a la química computacional la precisión requerida para el cálculo teórico de la entalpía de formación, puesto que se exige que la diferencia entre los valores teóricos y experimentales sea menor de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  para propiedades termodinámicas. En este contexto, los métodos compuestos emergen como

una alternativa eficaz, ofreciendo predicciones precisas con un costo computacional razonable. En este capítulo, explicaremos las herramientas necesarias para utilizar los métodos de la química computacional, en particular usar métodos compuestos para estimar el cálculo teórico de la entalpía de formación.

## 4.2. Mínimos de la SEP y frecuencias

Para determinar la geometría de equilibrio y los modos normales de vibración, es necesario optimizar la geometría, minimizando la energía potencial del sistema molecular, buscando los puntos críticos de la superficie de energía potencial (SEP). La SEP es una representación multidimensional que muestra cómo cambia la energía de un sistema molecular en función de las posiciones de los átomos [59]. El vector gradiente de la energía, denotado como  $g$ , está compuesto por las derivadas parciales de la energía con respecto a cada coordenada nuclear  $R_i$ :

$$g = \left( \frac{\partial E}{\partial R_1}, \frac{\partial E}{\partial R_2}, \dots, \frac{\partial E}{\partial R_N} \right). \quad (4.5)$$

En el punto de mínimo de energía, cada componente del vector gradiente es cero:

$$\frac{\partial E}{\partial R_i} = 0 \quad \text{para todo } i. \quad (4.6)$$

De todas formas, este criterio no es suficiente para caracterizar completamente la SEP, ya que un punto crítico podría ser un mínimo o un máximo. Para distinguirlos, se utiliza el hessiano  $H$ , cuya matriz está formada por las derivadas parciales de segundo orden de la energía con respecto a las coordenadas nucleares:

$$H_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j}. \quad (4.7)$$

La matriz hessiana se representa como:

$$H = \begin{pmatrix} \frac{\partial^2 E}{\partial R_1^2} & \frac{\partial^2 E}{\partial R_1 \partial R_2} & \dots & \frac{\partial^2 E}{\partial R_1 \partial R_N} \\ \frac{\partial^2 E}{\partial R_2 \partial R_1} & \frac{\partial^2 E}{\partial R_2^2} & \dots & \frac{\partial^2 E}{\partial R_2 \partial R_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial R_N \partial R_1} & \frac{\partial^2 E}{\partial R_N \partial R_2} & \dots & \frac{\partial^2 E}{\partial R_N^2} \end{pmatrix}. \quad (4.8)$$

En una geometría correspondiente a un mínimo en la SEP, el análisis de la matriz hessiana permite determinar los modos normales de vibración. Estos modos describen los patrones específicos con los cuales las moléculas vibran y son fundamentales para la caracterización espectroscópica y el análisis detallado de propiedades moleculares [69].

Para un sistema formado por  $N$  átomos, se obtienen  $(3N - 6)$  modos normales en moléculas no lineales o  $(3N - 5)$  en moléculas lineales, excluyendo los desplazamientos globales (traslaciones y rotaciones) respecto al centro de masas. A partir de estos modos normales es posible calcular la energía de punto cero (ZPE, por sus siglas en inglés), definida como la energía mínima residual que posee una molécula incluso a  $T = 0\text{K}$ , resultado del movimiento cuántico persistente de los núcleos atómicos. Esta energía surge como consecuencia del principio de incertidumbre, que impide que los átomos estén completamente en reposo [59]. Matemáticamente, la ZPE se determina a partir de las frecuencias vibratorias armónicas  $\nu_i$ , obtenidas del análisis del hessiano, mediante la siguiente expresión:

$$ZPE = \sum_i h\nu_i, \quad (4.9)$$

donde  $h$  es la constante de Planck y  $\nu_i$  es la frecuencia vibracional del modo normal  $i$ . Esta expresión corresponde al modelo armónico y, por tanto, no incluye correcciones anarmónicas explícitas. En los métodos compuestos como G4, los efectos anarmónicos se aproximan mediante un factor de escala empírico aplicado a las frecuencias armónicas.

### 4.3. Método de Hartree-Fock

El método Hartree-Fock (HF) es uno de los enfoques fundamentales en química cuántica para aproximar la función de onda de un sistema de muchos electrones. Este método se basa en el *principio variacional*, el cual establece que la energía del estado fundamental de un sistema cuántico es siempre mayor o igual al valor esperado del hamiltoniano calculado con cualquier función de onda de prueba.

En Hartree-Fock, la función de onda aproximada se representa mediante un único determinante de Slater, el cual garantiza la antisimetría requerida por el principio de exclusión de Pauli:

$$\Psi(X_1, X_2, \dots, X_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(X_1) & \phi_1(X_2) & \cdots & \phi_1(X_N) \\ \phi_2(X_1) & \phi_2(X_2) & \cdots & \phi_2(X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(X_1) & \phi_N(X_2) & \cdots & \phi_N(X_N) \end{vmatrix}, \quad (4.10)$$

donde  $\phi_i(X)$  son los espín-orbitales ortonormales.

La minimización de la energía con respecto a estos espín-orbitales conduce a las *ecuaciones de Hartree-Fock*:

$$\hat{F}\phi_i(1) = \epsilon_i\phi_i(1), \quad (4.11)$$

donde  $\hat{F}$  es el operador de Fock y  $\epsilon_i$  la energía orbital asociada a  $\phi_i$ . Este operador se construye como

$$\hat{F}(1) = \hat{h}(1) + \sum_j^{\text{ocupados}} [\hat{J}_j(1) - \hat{K}_j(1)], \quad \hat{h}(1) = -\frac{1}{2}\nabla_1^2 - \sum_A \frac{Z_A}{r_{1A}}. \quad (4.12)$$

El término  $\hat{h}(1)$  es el operador monoeléctrico, que incluye la energía cinética del electrón y la atracción núcleo–electrón. Los operadores de Coulomb  $J_j$  y de intercambio  $K_j$  representan, respectivamente, la repulsión electrostática promedio entre el electrón en un orbital  $\phi_i$  y la densidad asociada a un orbital ocupado  $\phi_j$ , así como el efecto de intercambio derivado de la antisimetría del determinante de Slater, el cual actúa únicamente entre electrones con el mismo espín. En conjunto, esta formulación muestra que cada electrón se mueve en un *potencial efectivo promedio* generado por los demás electrones. Este potencial, descrito por el operador de Fock  $\hat{F}$ , incluye una parte local (Coulomb) y otra no local (intercambio), y depende directamente de la ocupación y forma de los orbitales moleculares  $\phi_j$ .

El procedimiento de campo autoconsistente (Self-Consistent Field, SCF) consiste en partir de un conjunto inicial de orbitales para construir el potencial promedio, resolver las ecuaciones de Fock y obtener un nuevo conjunto de orbitales. Estos, a su vez, se utilizan para reconstruir el potencial y volver a resolver las ecuaciones, en un ciclo que se repite hasta que la variación en la

energía entre iteraciones sucesivas sea menor que un umbral de convergencia preestablecido. En ese punto, se acepta que se han encontrado los orbitales solución que satisfacen:

$$\hat{F}\psi_i = \varepsilon_i\psi_i. \quad (4.13)$$

## 4.4. Correlación electrónica

Aunque el método de HF es ampliamente utilizado, presenta limitaciones significativas en su capacidad para describir la energía de correlación electrónica, ya que trata la interacción electrónica como un campo promedio, en otras palabras, sin una descripción explícita de la interacción entre electrones. Este enfoque, basado en la aproximación de Hartree-Fock (HF), no logra capturar la energía de correlación electrónica, lo que limita su capacidad para describir con precisión las interacciones electrón-electrón. Löwdin [70] introdujo el término energía de correlación electrónica para referirse a la diferencia entre la energía exacta de un sistema y la calculada con el método de Hartree-Fock:

$$E_{\text{corr}} = E_{\text{exacta}} - E_{\text{HF}}. \quad (4.14)$$

La energía de correlación electrónica se suele dividir en dos tipos:

- **Correlación dinámica:** Esta se refiere a la correlación instantánea entre los electrones. Es decir, se centra en las interacciones electrónicas que hacen que los electrones eviten acercarse demasiado, lo cual esencialmente es una interacción a corta distancia, que es precisamente el tipo de correlación en la que falla la aproximación de campo promedio [71].
- **Correlación estática o no dinámica:** Ocurre cuando un sistema no puede ser adecuadamente descrito por una única configuración electrónica. En estos casos, el estado electrónico se representa mediante una combinación lineal de determinantes de Slater casi degenerados, lo cual permite una descripción equitativa de los estados electrónicos cerca-

nos o degenerados. Este efecto es crucial en sistemas en los que se presenta degeneración orbital, como en las rupturas de enlaces  $\sigma$  y  $\pi$ , intersecciones cónicas, entre otros [72].

El enfoque de Hartree-Fock, basado en un sólo determinante de Slater, no captura la correlación electrónica. En cambio, debido a su antisimetría, respeta el principio de exclusión de Pauli, asegurando que no tengamos dos electrones con el mismo espín en la misma posición, un efecto conocido como "Hueco de Fermi", que Hartree-Fock describe correctamente. En las siguientes secciones, se describirán los métodos desarrollados para recuperar la mayor cantidad posible de correlación electrónica [73].

## 4.5. Métodos post-Hartree-Fock

### 4.5.1. Teoría de perturbaciones de Møller-Plesset (MPn)

La teoría de perturbaciones de Møller-Plesset (MPn) es un método utilizado en química computacional para incorporar efectos de correlación electrónica más allá de los obtenidos mediante el método Hartree-Fock (HF). Esta teoría se fundamenta en una expansión perturbativa de la energía electrónica total, basada en correcciones sucesivas a partir de una referencia de Hartree-Fock. El nivel más utilizado, MP2, introduce correcciones de segundo orden a la energía de Hartree-Fock, permitiendo una descripción más precisa de la interacción dinámica entre los electrones [7]. La energía total en MP2 se expresa como:

$$E_{\text{MP2}} = E_{\text{HF}} + E^{(2)}, \quad (4.15)$$

donde  $E_{\text{HF}}$  es la energía de Hartree-Fock y  $E^{(2)}$  es la corrección de segundo orden, dada por:

$$E^{(2)} = \sum_{i < j}^{\text{ocupados}} \sum_{a < b}^{\text{virtuales}} \frac{|\langle ij || ab \rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}, \quad (4.16)$$

en la cual  $i, j$  corresponden a orbitales ocupados,  $a, b$  a orbitales virtuales,  $\epsilon$  son las energías orbitales de Hartree-Fock y  $\langle ij||ab \rangle$  representa las integrales bielectrónicas antisimetrizadas (Coulomb menos intercambio).

A niveles más altos, como MP3 y MP4, se incluyen correcciones adicionales de orden superior, aumentando la precisión pero también el costo computacional. Es importante destacar que, aunque el método MP2 es eficaz para describir ciertos sistemas, tiende a sobrestimar la energía de correlación electrónica. Por su parte, MP3 suele subestimar esta energía, mientras que MP4 proporciona una aproximación más cercana al valor exacto. El método MP4, al incluir términos hasta el cuarto orden en la serie de perturbaciones, mejora la descripción de la correlación electrónica en algunos casos específicos [74].

#### 4.5.2. Interacción de configuraciones (IC)

El método de Interacción de Configuraciones (CI) parte de la idea de que la función de onda exacta de un sistema de  $N$  electrones puede representarse como una combinación lineal de todos los posibles determinantes de Slater construidos a partir de un conjunto de espín-orbitales. Este conjunto es, en principio, completo para describir el espacio multielectrónico. Bajo este esquema, es posible recuperar la correlación electrónica de manera exacta y, además, obtener información no solo del estado fundamental, sino también de estados excitados, siempre que se consideren todas las excitaciones posibles desde orbitales ocupados hacia orbitales virtuales. Este enfoque se conoce como *Full CI* [75].

En la práctica, sin embargo, la complejidad computacional de Full CI crece factorialmente con el número de electrones y orbitales, lo que restringe su aplicación a sistemas muy pequeños. Para hacer viable el método se introducen truncamientos en el número de excitaciones consideradas. Por ejemplo, en el modelo CIS solo se incluyen excitaciones simples, mientras que CISD incorpora excitaciones simples y dobles, y así sucesivamente. De forma esquemática, la función de onda CI puede expresarse como:

$$\Psi_{\text{CI}} = a_0 \Phi_{\text{HF}} + \sum_S a_S \Phi_S + \sum_D a_D \Phi_D + \sum_T a_T \Phi_T + \dots = \sum_{i=0}^N a_i \Phi_i, \quad (4.17)$$

donde  $a_0$  es el coeficiente asociado al determinante de referencia de Hartree–Fock  $\Phi_{\text{HF}}$ ,  $\Phi_S$ ,  $\Phi_D$  y  $\Phi_T$  representan determinantes excitados simples, dobles y triples, respectivamente, y los coeficientes  $a_S$ ,  $a_D$  y  $a_T$  cuantifican la contribución de cada excitación a la función de onda total. La forma compacta  $\sum_{i=0}^N a_i \Phi_i$  resume que la función de onda CI es una combinación lineal de determinantes de Slater, y que en el límite de Full CI converge a la solución exacta dentro del espacio de base elegido [75, 76].

### 4.5.3. Cúmulos Acoplados (CC)

El método Cúmulos Acoplados (por sus siglas en inglés, CC) es una técnica avanzada de estructura electrónica que describe con alta precisión la correlación electrónica dinámica en sistemas moleculares de tamaño moderado. Su principal fortaleza es la consistencia con el tamaño del sistema, es decir, que la energía calculada escala correctamente al considerar sistemas más grandes [77].

La función de onda en CC se construye aplicando un operador exponencial de excitaciones a la función de onda de referencia de Hartree–Fock:

$$|\Psi_{\text{CC}}\rangle = e^T |\Phi_0\rangle, \quad (4.18)$$

donde  $T$  es el operador de excitación total, compuesto por  $T_1$ ,  $T_2$ ,  $T_3$ , etc. Cada término  $T_n$  promueve de manera colectiva  $n$  electrones desde orbitales ocupados a orbitales virtuales en la referencia de Hartree–Fock  $|\Phi_0\rangle$ .

$$T_1 = \sum_{i,a} t_i^a a_a^\dagger a_i, \quad T_2 = \frac{1}{4} \sum_{i,j,a,b} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i,$$

donde  $a_i$  y  $a_a^\dagger$  son operadores de aniquilación y creación en orbitales ocupados ( $i, j$ ) y virtuales ( $a, b$ ), respectivamente.

En la práctica, la serie se trunca para hacer los cálculos viables. Así, CCS incluye únicamente excitaciones simples; CCSD incorpora excitaciones simples y dobles; y CCSDT añade triples de

forma completa. Una variante ampliamente utilizada es CCSD(T), donde las triples se incluyen de manera perturbativa T, lo que permite capturar los efectos más importantes de estas excitaciones sin el costo prohibitivo de CCSDT. La forma exponencial garantiza que, aun con truncamientos, se incorporen implícitamente contribuciones de orden infinito, otorgando al método gran precisión y robustez. Por estas razones, CCSD(T) es conocido como el *estándar de oro* de la química cuántica. En particular, dentro de los métodos compuestos de la familia Gaussian- $n$ , CCSD(T) constituye la etapa crucial que recupera la mayor parte de la correlación dinámica y, por tanto, resulta de vital importancia en la predicción precisa de entalpías de formación en el método G4 [11].

Un aspecto importante al aplicar métodos de tipo CCSD(T) es determinar si la aproximación de un único determinante de Slater (Hartree–Fock) resulta adecuada o si, por el contrario, el sistema exhibe un carácter multirreferencia significativo. Para este propósito, Lee y Taylor [78] propusieron el denominado diagnóstico  $T_1$ , definido como la norma euclidiana del vector de amplitudes de excitación simple ( $t_i$ ), normalizada con respecto al número de electrones correlacionados:

$$T_1 = \left( \frac{1}{n_e} \sum_{i,a} |t_i^a|^2 \right)^{1/2}, \quad (4.19)$$

donde  $t_i^a$  son las amplitudes de excitación simple que promueven un electrón desde un orbital ocupado  $i$  hacia un orbital virtual  $a$ , y  $n_e$  es el número de electrones correlacionados. Valores pequeños de  $T_1$  indican que la referencia de Hartree–Fock es dominante y que la correlación electrónica es principalmente dinámica, por lo que métodos de referencia única como CCSD(T) son confiables. En cambio, cuando  $T_1 > 0.02$  el diagnóstico señala la presencia importante de correlación no dinámica y sugiere la necesidad de emplear metodologías multirreferencia.

## 4.6. Teoría de funcionales de la densidad

La DFT, ampliamente aplicada en sistemas atómicos, moleculares y sólidos, se basa en los teoremas de Hohenberg y Kohn [79], que se resumen en:

- **Teorema 1.** Para cualquier sistema de partículas interactuantes en un potencial externo, el potencial externo es determinado unívocamente por la densidad electrónica del estado base.
- **Teorema 2.** Para cualquier potencial externo particular, la energía del estado base del sistema, que es un funcional de la densidad, es el valor mínimo absoluto de este funcional, y la densidad que minimiza el funcional es la densidad electrónica exacta del estado base.

Dado que la energía es una función de la densidad, se puede escribir en sus componentes como:

$$E_0[\rho] = T[\rho] + E_{ee}[\rho] + E_{en}[\rho] + V_{nn}, \quad (4.20)$$

en donde  $E_0[\rho]$  es la energía total del sistema,  $T[\rho]$  es la energía electrónica,  $E_{ee}[\rho]$  representa la interacción electrón-electrón, así también,  $E_{en}[\rho]$  representa la interacción electrón-núcleo y por último,  $V_{nn}$  representa la repulsión internuclear. La ecuación 4.20 se puede dividir en la parte dependiente del sistema, que es la que contempla la atracción electrón-núcleo  $E_{en}$  en la que intervienen las coordenadas específicas de los núcleos, y la independiente, con la funcional universal que se representa como  $F_{HK}$  y se define como funcional de Hohenberg-Kohn:

$$F_{HK}[\rho] = T[\rho] + E_{ee}[\rho]. \quad (4.21)$$

El conocimiento de este funcional permitiría resolver exactamente la ecuación de Schrödinger, pero al no conocerse, se deben emplear aproximaciones. El funcional  $F_{HK}$  se divide en una parte de interacción electrón-electrón, que corresponde a la interacción clásica de Coulomb entre electrones en la nube electrónica de densidad  $\rho$ , y otro término que describe el resto de interacciones no clásicas ( intercambio, correlación ) :

$$E_{ee}[\rho] = J[\rho] + E_{xc}[\rho]. \quad (4.22)$$

Por último, la energía se puede escribir como:

$$E_0[\rho] = E_{en}[\rho] + V_{nn} + J[\rho] + (T[\rho] + E_{xc}[\rho]) \quad (4.23)$$

Los términos dentro del paréntesis representan la energía cinética y el intercambio-correlación; este último no tiene forma conocida y se evalúa por distintos modelos. La energía cinética se divide en una parte para un sistema de electrones independientes y otra que se agrega  $E_{xc}$  para su consideración ahí en las formas aproximadas. Los métodos DFT se clasifican según el método utilizado para calcular el término de intercambio y correlación:

- La Aproximación de Densidad Local (LDA, por sus siglas en inglés)

Se utilizó inicialmente en la Teoría del Funcional de la Densidad (DFT) para determinar la energía de intercambio y correlación ( $E_{xc}$ ) en una posición específica  $r$ . Esta energía puede ser calculada exclusivamente a partir del valor local de la densidad electrónica ( $\rho$ ). Teóricamente, el único requisito para  $\rho$  es que tenga un valor único en cada punto. En la práctica, los funcionales basados en esta definición se derivan del análisis del gas electrónico uniforme, en el que la densidad electrónica es constante en todas las posiciones. Esto se aplica, asumiendo que el sistema se comporta como un gas ideal en cada elemento de volumen con su propia densidad local, sin requerir que la densidad sea constante en todo el sistema. Por lo tanto, podemos expresar en LDA el funcional de intercambio y correlación como:

$$E_{xc}^{LDA}[\rho_0] = \int \rho_0(r) \epsilon_{xc}(r) dr, \quad (4.24)$$

donde  $\epsilon_{xc}$  es la energía de correlación e intercambio por electrón en un gas electrónico uniforme donde, en cada punto, el valor de  $\rho_0$  no cambia.

- Aproximación de gradiente generalizado (GGA y meta-GGA, por sus siglas en inglés Generalized Gradient Approximation)

Debido a las consideraciones anteriores, la LDA presenta limitaciones significativas en su aplicabilidad a otros sistemas. Una forma evidente de mejorar el funcional de correlación es hacerlo dependiente no solo del valor local de la densidad electrónica, sino también de la variación local de esta densidad; es decir, del gradiente de la densidad electrónica. Esto se expresaría como:

$$E_{xc}^{GGA}[\rho_0] = \int f_{xc}^0(\rho(r), \nabla\rho(r)) dr. \quad (4.25)$$

Posteriormente, es posible añadir elementos adicionales dependientes de la densidad electrónica. Esto lleva a la tercera aproximación, denominada meta-GGA, donde  $E_{xc}$  ahora depende de la densidad, del gradiente y del laplaciano de la densidad electrónica, y se expresa de la siguiente forma:

$$E_{xc}^{\text{GGA}}[\rho_0] = \int f_{xc}^0(\rho(r), \nabla\rho(r), \nabla^2\rho(r)) dr. \quad (4.26)$$

En esta tesis se empleará el funcional M06-2X [80], clasificado como un meta-GGA híbrido de intercambio y correlación. Este funcional depende de tres variables: la densidad de espín ( $\rho_\sigma$ ), el gradiente reducido de la densidad de espín y la densidad de energía cinética de espín. Utiliza una doble cantidad de intercambio no localizado (2X) y está parametrizado para moléculas sin metales. Se recomienda para estudios termoquímicos, constantes de velocidad y para interacciones no covalentes [80].

- Métodos híbridos

Las aproximaciones locales (LDA) y semilocales (GGA) marcaron un avance importante en DFT, pero presentaban errores sistemáticos al describir energías de disociación, entalpías de formación y barreras de reacción [81]. Para superar estas limitaciones se introdujeron los métodos híbridos, que incorporan una fracción del intercambio exacto de Hartree-Fock junto con términos aproximados de DFT, mejorando notablemente la precisión en el cálculo de propiedades.

El funcional más famoso y utilizado en la DFT es seguramente B3LYP [82]. Este funcional híbrido, desarrollado por Becke, Lee, Yang y Parr, también será estudiado en este trabajo. Este funcional combina la energía de intercambio con la energía exacta de Hartree-Fock (HF), así como correcciones de gradiente para el intercambio y la correlación, según la ecuación:

$$E_{xc}^{\text{B3LYP}} = (1 - a_0)E_x^{\text{LSDA}} + a_0E_x^{\text{HF}} + a_x\Delta E_x^{\text{B88}} + a_cE_c^{\text{LYP}} + (1 - a_c)E_c^{\text{VWN}}, \quad (4.27)$$

donde  $E_x^{\text{LSDA}}$  es el funcional de intercambio con la aproximación de densidad local (LSDA),  $E_x^{\text{HF}}$  es el intercambio exacto de Hartree-Fock (HF),  $E_x^{\text{B88}}$  es la corrección de gradiente de Becke al intercambio,  $E_c^{\text{LYP}}$  es el funcional de correlación de Lee, Yang y Parr, y  $E_c^{\text{VWN}}$  es

la aproximación de densidad local de Vosko, Wilk y Nusair (VWN) para la correlación. Los tres parámetros empíricos conocidos que modifican estas contribuciones son  $a_0 = 0.20$ ,  $a_x = 0.72$  y  $a_c = 0.81$  [82]. No obstante, el funcional B3LYP presenta limitaciones bien documentadas, como la subestimación de barreras de reacción y una precisión insuficiente en ciertos sistemas, especialmente aquellos donde los efectos de dispersión son relevantes [13, 83, 84]. En 2015, Lu investigó cómo mejorar este funcional mediante la optimización de los parámetros empíricos que definen las proporciones de intercambio y correlación. Para ello, empleó el conjunto de calibración G2/97 como conjunto de entrenamiento y el conjunto G3/05 como conjunto de prueba. A través de una optimización global de los parámetros  $a_0$ ,  $a_x$  y  $a_c$ , desarrolló una nueva versión del funcional, denominada OpB3LYP. Esta versión optimizada mostró una mejora considerable en la precisión de los cálculos, especialmente en la predicción de energías de atomización, potenciales de ionización y entalpías de formación de moléculas medianas y grandes, superando algunas de las deficiencias conocidas del B3LYP original [85]. Además, surgen diferentes aproximaciones, en las cuales se basan en intentar describir mejor las interacciones de Van der Waals (fuerzas de dispersión), como una que puede ser considerada dentro de las mismas moléculas a estudiar la interacción con otras. Los funcionales desarrollados para este fin se han etiquetado como DFT-D (DFT con correcciones de dispersión). Por ejemplo,  $\omega$ B97XD [86] es un funcional de densidad híbrido corregido para interacciones de largo alcance que incorpora tanto correcciones de intercambio de Hartree-Fock como términos de dispersión de van der Waals. Este funcional ha sido diseñado para mejorar la precisión en cálculos de energía de interacciones no covalentes, energías de excitación de transferencia de carga de largo alcance, así como también para una mayor exactitud en termodinámica y cinética molecular. Utiliza una corrección basada en la función de error complementaria para separar los componentes de largo y corto alcance en las interacciones electrón-electrón, lo que permite una mejor representación de las interacciones de largo alcance que otros funcionales híbridos globales. Cabe mencionar que este modelo incluye correcciones de dispersión basadas en el modelo de Grimme [86, 87].

Considerando las diferentes aproximaciones funcionales que hemos citado y que usaremos en esta tesis, es importante considerar que la teoría del funcional de la densidad presenta sus limi-

taciones, dentro de las que sobresalen:

- No existe un procedimiento sistemático para mejorar el  $E_{xc}$ .
- Todos los funcionales usados constituyen el uso de diferentes aproximaciones y parámetros, algunos de ellos empíricos y otros con respecto a la densidad.
- No existe ningún funcional que sea lo suficientemente preciso para calcular la energía electrónica con suficiente calidad y obtener resultados confiables en el cálculo de propiedades termoquímicas, en particular para el cálculo teórico de entalpías de formación.

## 4.7. Conjuntos de base

En los cálculos de estructura electrónica, la forma de la función de onda y el número de integrales requeridas para describir a los electrones son cruciales. Para ello, se emplea un conjunto de base. La elección del conjunto de base depende de la naturaleza del sistema y de los tiempos de cálculo, ya que, si se utiliza un conjunto pequeño, los cálculos se realizan más rápidamente, pero con menor precisión. En estos cálculos se emplean conjuntos de funciones base construidos a partir de expresiones analíticas que reproducen la forma de las funciones orbitales atómicas. Entre ellos destacan las funciones tipo Slater (STO) y las funciones tipo gaussianas (GTO) [75].

$$\text{STO} = N \cdot r_a^{n-1} \cdot e^{-\zeta r_a} \cdot Y_{lm}(\theta_a, \phi_a), \quad (4.28)$$

donde  $Y_{lm}(\theta_a, \phi_a)$  son los armónicos esféricos y  $N$  es la constante de normalización. Cada orbital molecular se describe mediante una combinación lineal de funciones base  $\chi$ .

En el caso de las funciones gaussianas, su expresión es:

$$g_{ijk}(\mathbf{r}_b) = N x_b^i y_b^j z_b^k e^{-\alpha r_b^2}, \quad (4.29)$$

donde  $x_b$ ,  $y_b$  y  $z_b$  son las coordenadas cartesianas respecto al centro  $b$ ;  $r_b = \sqrt{x_b^2 + y_b^2 + z_b^2}$ ;  $N$  es la constante de normalización;  $\alpha$  es el exponente que controla la rapidez del decaimiento radial de la función gaussiana con la distancia  $r_b$ ; y los enteros  $i$ ,  $j$  y  $k$  indican la forma angular, determinando así el tipo de orbital atómico.

El cálculo de funciones base para moléculas poliatómicas resulta complejo y computacionalmente costoso, principalmente debido a la dificultad de evaluar las integrales bielectrónicas cuando se emplean funciones de base tipo Slater (4.28). Por ello, en sistemas de gran tamaño se opta por utilizar funciones del tipo gaussiano (GTF, por sus siglas en inglés, Gaussian Type Functions), cuyas expresiones matemáticas se muestran en la ecuación (4.29). Dichas funciones se utilizan en combinaciones lineales normalizadas de funciones gaussianas individuales. Una forma de clasificar las funciones base, según lo propuesto por Pople, es la siguiente:

- **Conjunto de base mínima:** Estos conjuntos contienen el número mínimo de funciones base necesarias para cada átomo, de modo que cada orbital atómico se representa mediante una única función STO. Para simplificar los cálculos, estas funciones STO se aproximan utilizando  $N$  funciones GTO. Las bases del tipo STO-NG son ejemplos de este enfoque; entre ellas, la más utilizada es la STO-3G, que aproxima cada función STO con tres funciones gaussianas primitivas.
- **Conjunto de base de valencia dividida:** Para aumentar el tamaño de una base, se incrementa el número de funciones por átomo. En estos conjuntos, se conserva una base mínima para los electrones internos (también conocidos como “electrones de core”), mientras que se duplican o triplican las funciones destinadas a los electrones de valencia. Esto da lugar a las bases doble zeta (3-21G y 6-31G) y triple zeta (6-311G).
- **Conjunto de base polarizada:** Aunque los conjuntos de base dividida permiten modificar el tamaño de los orbitales, no cambian su forma. Esta limitación se puede superar añadiendo orbitales con un momento angular superior al requerido para describir el estado fundamental de cada átomo. Por ejemplo, los conjuntos de base polarizados añaden funciones  $d$  a los átomos de carbono y funciones  $f$  a los metales de transición; algunos también agregan funciones  $p$  a los átomos de hidrógeno. El conjunto de base 6-31G\* in-

dica un conjunto 6-31G con funciones  $d$  añadidas a los átomos pesados, mientras que el conjunto 6-31G\*\* incluye tanto funciones  $d$  para los átomos pesados como funciones  $p$  para los átomos de hidrógeno.

- **Bases difusas:** Estas funciones, de tipo  $s$  o  $p$ , permiten que los orbitales se extiendan sobre una región amplia del espacio. Los conjuntos de base con funciones difusas son esenciales para sistemas donde los electrones se encuentran relativamente alejados del núcleo, como en moléculas con pares libres, aniones, dímeros con enlaces de hidrógeno y sistemas en estados excitados. El conjunto 6-31+G(d) incorpora funciones difusas en los átomos pesados, mientras que el 6-31++G(d) también las añade a los átomos de hidrógeno.

Dunning y *et al.* [88–91] desarrollaron otro tipo de funciones base destinadas a cálculos que incluyen la correlación electrónica, como las bases cc-pVDZ y cc-pVTZ, que se agrupan bajo el nombre cc-pVXZ, donde “X” se refiere al nivel de zeta de valencia polarizada compatible con la correlación. Las funciones difusas, indicadas por el prefijo “aug”, añaden funciones con un pequeño exponente para cada momento angular a las bases cc, y son aplicables a átomos distintos del hidrógeno.

## 4.8. Extrapolación de base infinita

Como hemos comentado previamente, una labor intrínseca del químico computacional es la selección tanto del método de estructura electrónica a utilizar, lo que determina cuánta correlación electrónica podemos recuperar, como del conjunto de base adecuado. A medida que se seleccionan conjuntos de bases más grandes y robustos, o se utilizan una mayor cantidad de funciones gaussianas, el costo computacional se incrementa significativamente. En este contexto, es bien conocido y lógico pensar que realizar cálculos utilizando un conjunto infinito de funciones gaussianas o un conjunto de base completo (CBS, por sus siglas en inglés) es inviable en la práctica y su aplicabilidad en sistemas de tamaño medio. Para superar esta limitación, se han desarrollado esquemas de extrapolación que permiten estimar el límite del conjunto de base sin necesidad de realizar cálculos extremadamente costosos. Las extrapolaciones de conjun-

tos de bases son más útiles (y también más importantes) para métodos de estructura electrónica altamente correlacionados [92]. De acuerdo con Halkier *et al.* [68], al estudiar la convergencia de la energía al límite de Hartree-Fock, se observó que los errores de energía total se ajustan mejor a una forma exponencial que a una forma de potencia, siendo la forma exponencial más adecuada para esquemas de extrapolación. Esto implica que, a través de un número finito de cálculos con conjuntos de bases cada vez más grandes, es posible extrapolar hacia el límite de conjunto de bases completo (CBS) y obtener estimaciones precisas de la energía total de sistemas moleculares [93]. En nuestros días existen diferentes esquemas de extrapolación de base al límite. Para extrapolar la energía al límite del conjunto de bases completo, se utiliza una fórmula específica. En el método G4 se utiliza el esquema de extrapolación propuesto por Jensen [92] y su fórmula es la siguiente:

$$E_X = E_{limit} + B e^{-\alpha \cdot n}, \quad (4.30)$$

donde  $E_X$  es la energía en el conjunto de base usando un número cardinal y  $E_{limit}$  corresponde a la energía al conjunto de base al límite. Los parámetros  $\alpha$  y  $B$  son valores que corresponden a un conjunto de calibración usado; detallaremos estos valores en la sección 5.1. Específicamente, los autores destacan que los conjuntos de bases como el cc-pV6Z ya proporcionan energías muy cercanas al límite CBS con errores de aproximadamente  $0.26 \text{ kJ} \cdot \text{mol}^{-1}$ , el uso de esquemas de extrapolación puede reducir aún más estos errores y, en algunos casos, evitar cálculos adicionales en conjuntos de bases más grandes.

Un aspecto importante a considerar en la extrapolación al límite de base completa (CBS) es el error de superposición de base (BSSE, por sus siglas en inglés). Este error surge cuando dos fragmentos de un sistema molecular  $A$  y  $B$  comparten funciones de base: cada fragmento está descrito con mayor flexibilidad en el complejo  $AB$  que en aislamiento, lo que conduce a una energía de interacción artificialmente estabilizada [94]. El procedimiento más común para corregirlo es el método de counterpoise (CP), propuesto por Boys y Bernardi, donde cada fragmento se calcula en la base extendida del complejo completo, incluyendo las funciones de los demás fragmentos como orbitales fantasma. En principio, al aumentar el tamaño del conjunto de base hasta el límite CBS, el BSSE tiende a desaparecer, ya que cada fragmento se describe de manera suficientemente completa sin necesidad de “prestar” funciones. No obstante, esta cancelación es efectiva únicamente cuando la extrapolación se realiza a partir de bases grandes

(por ejemplo, cc-pVQZ o cc-pV5Z).

El procedimiento más común para corregirlo es el método de *counterpoise* (CP), propuesto por Boys y Bernardi, donde cada fragmento se calcula en la base extendida del complejo completo, incluyendo las funciones de los demás fragmentos como orbitales fantasma. La corrección de BSSE para un dímero  $AB$  se expresa como:

$$E_{\text{int}}^{\text{corr}} = E_{AB}^{AB} - E_A^{AB} - E_B^{AB}. \quad (4.31)$$

donde  $E_A^{AB}$  y  $E_B^{AB}$  son las energías de los fragmentos  $A$  y  $B$  calculadas en la base completa del complejo (es decir, incluyendo las funciones de base del otro fragmento como orbitales fantasma), mientras que  $E_A^A$  y  $E_B^B$  corresponden a las energías de los fragmentos en sus propias bases.

Si se emplean bases pequeñas (cc-pVDZ, cc-pVTZ, 6-31G, 6-311G) o si se estudian sistemas donde predominan fuerzas de dispersión y enlaces débiles, el BSSE residual puede persistir, aun después de la extrapolación. En esos casos, es recomendable combinar el esquema CBS con la corrección CP para obtener energías más fiables [95–97]. Una pregunta natural y poco obvia sería pensar en que tan importante puede resultar la BSSE en interacciones intermoleculares. Aunque CBS reduce este error, si se utilizan bases pequeñas o se analizan conformaciones estabilizadas por puentes de hidrógeno o interacciones débiles, el BSSE puede sesgar las energías relativas. En estos casos, es aconsejable considerar la corrección counterpoise adaptada a fragmentos intramoleculares o utilizar conjuntos de base suficientemente grandes. Cabe señalar que, aunque el error de superposición de bases (BSSE) puede tener un efecto significativo en interacciones intramoleculares débiles, en el presente trabajo este aspecto no fue considerado de manera explícita.

Hasta el momento hemos descrito todos los conceptos e ideas importantes para poder describir con mayor detalle los métodos compuestos, en particular el método G4. A continuación, describiremos de forma general qué son los métodos compuestos y, en particular, el desarrollo de G<sub>n</sub> hasta llegar a G4.

## 4.9. Métodos compuestos

Los métodos compuestos han demostrado ser una herramienta fundamental en la predicción de propiedades termoquímicas, al combinar una serie de cálculos que utilizan diferentes conjuntos de bases y niveles teóricos. Estos métodos permiten estimar propiedades termodinámicas, como las entalpías de formación, con una desviación media absoluta cercana a  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  con respecto a los datos experimentales. En lugar de realizar un único cálculo energético, los métodos compuestos emplean varias aproximaciones secuenciales que, al combinarse, ofrecen resultados más precisos con un costo computacional reducido. Entre los métodos compuestos más conocidos, las teorías Gaussian- $n$  (G $n$ ) son especialmente destacables por su robustez y precisión, logrando desviaciones medias absolutas por debajo de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$  en la predicción de diversas propiedades termoquímicas. La versión posterior, G4, alcanza una desviación media de  $3.48 \text{ kJ} \cdot \text{mol}^{-1}$  para un conjunto de 454 valores experimentales que incluyen entalpías de formación, afinidades electrónicas y potenciales de ionización. Aunque estos métodos se discutirán con mayor detalle en la siguiente sección, es importante destacar su relevancia en estudios donde se requiere un alto nivel de precisión [20]. La serie Petersson CBS (Complete Basis Set) es otra familia de métodos compuestos que extrapola la energía de un sistema hacia el límite del conjunto de bases completo. Métodos como CBS-QB3 [32] y CBS-APNO [98] han alcanzado desviaciones medias absolutas de hasta  $4.31 \text{ kJ} \cdot \text{mol}^{-1}$  para moléculas pequeñas y medianas, lo que los convierte en herramientas eficientes para obtener resultados cercanos a los cálculos Full CI (Configuración Completa de Interacción), pero con menores requerimientos computacionales [99]. Los métodos Weizmann- $n$ , como W1 y W2, también son considerablemente precisos, logrando errores de hasta  $2.68 \text{ kJ} \cdot \text{mol}^{-1}$  en la predicción de propiedades como entalpías de formación y energías de disociación, lo que los posiciona como los más precisos para moléculas pequeñas. Estos métodos no incluyen parámetros empíricos, lo que garantiza su aplicabilidad en una variedad de sistemas químicos. En contraste, su aplicabilidad en sistemas de tamaño medio y moderado resulta ser impráctica debido al alto costo computacional de estos métodos [28].

### 4.9.1. Métodos Gn y conjuntos de calibración

Desarrollados por Sir John Pople *et al.* [23], los métodos Gn realizan cálculos de energías moleculares mediante la combinación de métodos de correlación de alto nivel, como QCISD(T) o CCSD(T), con diferentes conjuntos de bases de tamaño moderado. Los cálculos de nivel inferior, como MP2 o MP4, se complementan con conjuntos de bases más grandes para mejorar la precisión. A lo largo del tiempo, los métodos Gn han evolucionado desde la versión G1 [17] hasta la más reciente G4 [20], mejorando en precisión gracias al refinamiento de los conjuntos de bases y los ajustes empíricos. El término HLC en los métodos Gn se emplea para ajustar las deficiencias residuales en los cálculos de energía después de aplicar correcciones basadas en la correlación electrónica y otros efectos. Estos ajustes están diseñados para reducir las diferencias entre las energías calculadas y los datos experimentales. La inclusión de esta corrección originó la creación de conjuntos de calibración para ajustar y validar estos métodos. En la siguiente sección discutiremos los conjuntos de calibración originados y usados en los métodos Gn. La precisión de los métodos Gn ha sido medida utilizando la desviación media absoluta (DMA) con respecto a referencias experimentales. Estos métodos han mejorado desde la primera versión Gaussian-1 (G1) [100] con una DMA de alrededor de  $6.69 \text{ kJ} \cdot \text{mol}^{-1}$  hasta su última versión, Gaussian-4 (G4) [20], con una DMA de  $3.47 \text{ kJ} \cdot \text{mol}^{-1}$  respecto al conjunto de moléculas denominado G2/97. Los componentes de los métodos Gn se describen en la tabla 4.1. Si se desea ver cómo se combinan estas componentes para obtener las energías de los métodos G1, G2, G3, puede revisarse en los anexos. El método G4 se encuentra en la primera parte de nuestra sección metodológica, ya que es punto de partida de nuestro trabajo.

**Tabla 4.1:** Componentes de los métodos Gn

Nivel de teoría	G1 [17,100]	G2 [18]	G3 [19]	G4 [20]
<b>Geometría</b>	MP2/6-31G(d)	MP2/6-31G(d)	MP2/6-31G(d)	B3LYP/6-31G(2df,p)
<b>Factor de escala de ZPE</b>	HF/6-31G(d) [0.893]	HF/6-31G(d) [0.893]	HF/6-31G(d) [0.893]	B3LYP/6-31G(2df,p) [0.9854]
<b>SCF</b>	6-311G(2df,p)	6-311G+(3df,2p)	6-311G+(3df,2p)	aug-cc-pVQZ
<b>Correlación de electrones de valencia</b>	MP2/6-311G(2df,p)	MP2/6-311G(2df,p)	MP2/6-311G+(3df,2p)	MP2/G3LargeXP
	MP4/6-311+G(d,p)	MP4/6-311+G(d,p)	MP4/6-31G(d)	MP4/6-31+G
	MP4/6-311G(2df,p)	MP4/6-311G(2df,p)	MP4/6-31G(2df,p)	CCSD/6-31G(d)
	QCISD/6-311G(d,p)	QCISD/6-311G(d,p)	QCISD/6-31G(d)	MP2/G3LargeXP
(T)	6-311G(d,p)	6-311G(d,p)	6-31G(d)	6-31G(d)
<b>Correlación de núcleo-valencia</b>	-	-	-	MP2/G3LargeXP
<b>Corrección relativista</b>	-	-	-	-
<b>DMA (kJ · mol<sup>-1</sup>)</b>	6.31	6.19	4.31	3.47

El método G3 evaluado en el conjunto de calibración G3/05 obtuvo una DMA de  $4.72 \text{ kJ} \cdot \text{mol}^{-1}$ . Por otro lado, en cada variante existen versiones ligeras de G3 y G4, que se desarrollaron de manera alternativa como G3(MP2) y G4(MP2) respectivamente.

### G3(MP2)

La variante G3(MP2) mantiene la misma arquitectura general de G3, pero reemplaza las contribuciones más costosas de QCISD(T) por cálculos de nivel MP2 en conjuntos de bases grandes. De esta forma, la energía final puede expresarse como:

$$E_{G3(\text{MP2})} = E_{\text{MP2}} + \Delta E_{\text{MP4}} + \Delta E_{\text{SCF}} + E_{\text{ZPVE}}^{\text{esc}} + E_{\text{HLC}}, \quad (4.32)$$

reduciendo significativamente el costo computacional a expensas de un ligero incremento en la desviación estadística. Esta variante se evaluó en el conjunto G3/05 obteniendo una DMA de  $5.81 \text{ kJ} \cdot \text{mol}^{-1}$  [37], el factor de escala usado es de 0.8929 y el parámetro  $A$  usado en HLC es 9.279 mHartrees.

### G4(MP2)

El método G4(MP2) es una variante de menor costo computacional que sustituye los cálculos de MP4 y CCSD(T) a bases grandes de G4 por niveles basados en MP2 [42]. La energía total mantiene la misma estructura general:

$$E_{G4(\text{MP2})} = E_{\text{CCSD(T)}}^{6-31G(d)} + \Delta E_{\text{MP2}}^{\text{LargeXP}} + \Delta E_{\text{HF}}^{\text{CBS}} + E_{\text{ZPVE}}^{\text{esc}} + E_{\text{HLC}} + E_{\text{SO}}. \quad (4.33)$$

Este esquema conserva la extrapolación de Hartree–Fock al límite de base completa, el cálculo de la energía de punto cero mediante frecuencias B3LYP/6-31G(2df,p) con un factor de escala de 0.9854, igual que G4 y la corrección de alto nivel (HLC) con  $A=9.741$  mHartrees.

Su desempeño se evaluó en el conjunto de calibración G3/05 con una DMA de  $4.35 \text{ kJ} \cdot \text{mol}^{-1}$ . Aunque menos preciso que G4 con una DMA de  $3.47 \text{ kJ} \cdot \text{mol}^{-1}$ , el método resulta seis a ocho veces más rápido y aún supera en precisión a G3 y G3(MP2). Por ello, G4(MP2) ofrece un balance favorable entre costo y precisión, siendo atractivo para moléculas de mayor tamaño y estudios exploratorios en termodinámica y energías de reacción.

### G4(MP2)-6X

La variante G4(MP2)-6X fue propuesta por Chan y Radom para mejorar la precisión de G4(MP2) sin incrementar el costo computacional [101]. Este esquema emplea geometrías y frecuencias BMK/6-31+G(2df,p), introduce seis factores de escalamiento que ajustan de manera separada las contribuciones de correlación y reparametriza el término de corrección de alto nivel (HLC). Los parámetros fueron optimizados sobre un conjunto de calibración ampliado (E2, 526 energías), lo que permitió mejorar la descripción de entalpías de formación, energías de reacción, barreras y complejos no covalentes.

La energía total se expresa como:

$$E_{\text{G4(MP2)-6X}} = E_{\text{HF}}^{\text{CBS}} + E_{\text{corr}}^{\text{scaled}} + E_{\text{HLC}}^{\text{reopt}} + E_{\text{ZPVE}}^{\text{esc}}(\text{BMK}) + E_{\text{SO}}, \quad (4.34)$$

donde  $E_{\text{corr}}^{\text{scaled}}$  corresponde a las contribuciones de MP2, CCSD y (T) multiplicadas por factores de escalamiento específicos. Con estas modificaciones, la DMA se reduce de  $4.4$  a  $3.6 \text{ kJ} \cdot \text{mol}^{-1}$  evaluado en el conjunto E2, alcanzando un desempeño cercano al de G4, pero con un costo comparable al de G4(MP2).

### Conjuntos de calibración usados en Gn

Según la Real Academia Española, en física calibrar se define como ajustar, con la mayor exactitud posible, las indicaciones de un instrumento de medida con respecto a un patrón de referencia [102]. Sin embargo, en el contexto de la química computacional, donde no empleamos instrumentos físicos, nuestro patrón de referencia corresponde a los valores experimentales repor-

tados. Así, los conjuntos de calibración —compuestos por moléculas con propiedades termoquímicas medidas experimentalmente— actúan como referencias para ajustar los parámetros empíricos y minimizar las discrepancias entre predicciones computacionales y realidad experimental. En este contexto, los conjuntos de calibración juegan un papel esencial en la evaluación y ajuste de los métodos Gn. Estos conjuntos contienen datos experimentales que permiten optimizar los parámetros empíricos de HLC para minimizar las desviaciones entre los cálculos teóricos y los valores experimentales. El primer conjunto de calibración desarrollado fue el G2-test set, también conocido como G2/97, el cual contiene 301 datos termoquímicos que incluyen entalpías de formación, energías de ionización, afinidades electrónicas y afinidades protónicas. Este conjunto fue fundamental para el ajuste del método G2 y se utilizó como referencia para medir la precisión de este método [21]. El conjunto G2/97 permitió la optimización de los dos parámetros empíricos de G1 y G2 para moléculas y átomos, con el fin de corregir las deficiencias en la predicción de energías. Con el desarrollo de métodos más avanzados, como G3 y G4, también se crearon nuevos conjuntos de calibración.

El conjunto de calibración G3/05 representa una expansión significativa de conjuntos previos como G2-test set (125 cambios energéticos), G2/97 (301) y G3/99 (376), incorporando un total de 454 datos experimentales. Este conjunto incluye 270 entalpías de formación, 105 energías de ionización, 63 afinidades electrónicas, 10 afinidades protónicas y 6 complejos enlazados por puentes de hidrógeno [22, 76]. Aunque los valores de referencia utilizados en G2/97 y G3/05 permanecen sin cambios, la principal diferencia radica en el aumento del tamaño y la diversidad del conjunto. La inclusión de especies más grandes y no convencionales, como moléculas con elementos de la tercera fila, permitió establecer un estándar más riguroso para validar métodos computacionales. Este conjunto fue empleado para optimizar los parámetros de HLC en G3, incluyendo ajustes diferenciados para moléculas de capa cerrada y abierta. Posteriormente, el método G4 incorporó mejoras adicionales al término  $E_{\text{HLC}}$  mediante dos nuevos parámetros, con el fin de aumentar la precisión en sistemas con electrones desapareados [20]. En conjunto, la diversidad y representatividad del G3/05 han sido clave para afinar la predicción de propiedades termoquímicas en métodos compuestos como G4. Esta estrategia también ha demostrado ser esencial en la evaluación de métodos de la teoría del funcional de la densidad (DFT). Además, han señalado que la diversidad molecular y la calidad de los valores de referencia son cruciales

para garantizar la precisión y predictibilidad en métodos computacionales avanzados [103].

#### 4.9.2. Métodos compuestos CBS-n

El concepto de extrapolación al límite de base completa (Complete Basis Set, CBS) fue introducido por Petersson y colaboradores a finales de los años ochenta. Estos autores demostraron que la energía calculada a nivel MP2 podía extrapolarse hacia el límite CBS mediante una expansión en términos de pares de orbitales naturales (PNO, Pair Natural Orbitals). A partir de esta idea se incorporó la extrapolación MP2/CBS dentro de los métodos compuestos de la química cuántica, dando origen al modelo CBS-APNO (Atomic Pair Natural Orbital) [104–106].

Posteriormente, Petersson y su grupo desarrollaron variantes de menor costo computacional: CBS-4, cuyo nivel más alto es un cálculo MP4, y CBS-Q, cuyo nivel más alto corresponde a un cálculo QCISD. Estas metodologías fueron refinadas con el tiempo hasta convertirse en las versiones CBS-4M y CBS-QB3 [32, 98].

De esta manera, los métodos CBS conforman una jerarquía de precisión y costo:

- CBS-4M: adecuado para sistemas grandes (más de 50 átomos), con una precisión moderada.
- CBS-QB3: aplicable a sistemas de tamaño medio (hasta 25 átomos), con una precisión intermedia y un balance favorable entre exactitud y costo computacional.
- CBS-APNO: el más preciso (con desviaciones en torno a  $0.5 \text{ kJ} \cdot \text{mol}^{-1}$ ), aunque limitado a elementos de la primera fila de la tabla periódica, debido al uso de bases APNO específicamente diseñadas, por lo tanto no aplicable para nuestro sistemas.

Los métodos de base completa (CBS, Complete Basis Set) fueron diseñados para reducir el error de truncamiento de la base mediante extrapolaciones al límite CBS. Sin embargo, tanto CBS-QB3 como CBS-APNO incluyen parámetros empíricos ajustados frente a conjuntos de ca-

libración estándar G2 test, lo cual es clave para alcanzar la precisión química menor de  $4.18 \text{ kJ} \cdot \text{mol}^{-1}$ . No obstante, los métodos más utilizados de esta familia corresponden a CBS-QB3 y CBS-APNO. A continuación se describirán con más detalle.

### CBS-QB3

El método CBS-QB3 combina extrapolaciones y correcciones de alto nivel con varios términos empíricos. Su energía total puede expresarse como:

$$E_{\text{CBS-QB3}} = E_{\text{UMP2}} + \Delta E_{\text{CBS(2)}} + \Delta E_{\text{UMP4}} + \Delta E_{\text{UCCSD(T)}} + \Delta E_{\text{spin}} + \Delta E_{\text{emp}} + \Delta E_{\text{ZPVE}}, \quad (4.35)$$

donde los parámetros empíricos son:

$$\Delta E_{\text{spin}} = -9.54 mE_h \cdot \langle S^2 \rangle \quad ; \quad \Delta E_{\text{emp}} = -5.79 mE_h \sum_i \left( \sum_m C_{mii} \right)^2 |S_{ii}|^2. \quad (4.36)$$

En CBS-QB3, la corrección de espín compensa errores por contaminación de  $\langle S^2 \rangle$  en sistemas abiertos, mientras que el término empírico ajusta discrepancias residuales de la correlación electrónica no capturadas por los pasos perturbativos o de acoplamiento de clúster. La energía de punto cero se escala con un factor empírico de 0.99. Todos estos parámetros fueron ajustados frente al G2/97 test set, logrando una desviación media absoluta de  $1.1 \text{ kJ} \cdot \text{mol}^{-1}$  [32].

### CBS-APNO

El método CBS-APNO representa el nivel más alto de la familia CBS. Su energía final es:

$$E_{\text{CBS-APNO}} = E_{\text{HF}}^{\text{CBS}} + \Delta E_{\text{MP2}}^{\text{CBS}} + \Delta E_{\text{MP4}}^{\text{CBS}} + \Delta E_{\text{QCISD(T)}}^{\text{CBS}} + \Delta E_{\text{emp}} + E_{\text{ZPVE}}, \quad (4.37)$$

con una única corrección empírica:

$$\Delta E_{\text{emp}} = -1.74 mE_h \sum_{i=1}^{n_b} \left( \sum_{m=1}^{N_{\text{virt}}+1} C_{mii} \right)^2 |S_{ii}|^2. \quad (4.38)$$

Este término empírico corrige errores residuales de correlación electrónica asociados con la participación de orbitales virtuales y solapamientos, reduciendo desviaciones sistemáticas en las energías aun después de aplicar las extrapolaciones CBS y QCISD(T). Este término fue calibrado con 64 moléculas de la G2 test set, reduciendo el error medio de 1.07 a 0.53 kJ · mol<sup>-1</sup> [98]. Obsérvese, al igual que con los métodos Gn, el aspecto semiempírico de los métodos CBS y en consecuencia usan los conjuntos de calibración de Gn para hacer algunos ajustes.

## 4.10. Termoquímica

La termoquímica estudia la transferencia de calor asociada a las reacciones químicas [15]. Una propiedad central es la entalpía,  $H$ , conveniente para procesos a presión constante. Formalmente,  $H$  se obtiene como una *transformada de Legendre* de la energía interna  $U(S, V)$  respecto al par conjugado  $(V, -p)$ , con el propósito de sustituir la variable natural  $V$  por su conjugada  $p$  [107]:

$$H(S, p) \equiv U(S, V) + p V. \quad (4.39)$$

Tomando diferenciales, a partir de la primera ley  $dU = T dS - p dV$  se obtiene

$$dH = dU + p dV + V dp = T dS + V dp, \quad (4.40)$$

por lo que las variables naturales de  $H$  son  $S$  y  $p$ :

$$T = \left( \frac{\partial H}{\partial S} \right)_p, \quad V = \left( \frac{\partial H}{\partial p} \right)_S. \quad (4.41)$$

En condiciones de trabajo habituales (presión constante), el calor intercambiado coincide con  $\Delta H$ . A  $T = 0$  K, la energía interna molecular puede descomponerse como

$$U_{0K} = E_{\text{elec}} + V_{nn} + \sum_i h\nu_i, \quad (4.42)$$

donde  $E_{\text{elec}}$  es la energía electrónica,  $V_{nn}$  la repulsión núcleo-núcleo y  $\sum_i h\nu_i$  es la energía de punto cero (ZPE). En tales condiciones, la entalpía satisface la identidad

$$H_{0K} = U_{0K} + p V, \quad (4.43)$$

que no es una definición independiente, sino la consecuencia directa de la transformada de Legendre en (4.39) [15, 107].

donde,  $V_{nn}$  representa la repulsión núcleo-núcleo, es decir, la energía asociada a la interacción electrostática repulsiva entre los núcleos de la molécula. Usando el modelo de gas ideal, podemos indicar que la entalpía a 0 K ( $H_{0K}$ ) es igual a la energía interna  $U_{0K}$ . Los términos de la ecuación 4.42 son los que nosotros, como químicos computacionales, podemos calcular con una alta precisión y usando relaciones termodinámicas podemos derivar entonces la entalpía de formación  $\Delta_f H^\circ$ . En muchos procesos químicos, como las reacciones químicas, se asume que el sistema opera a presión constante. El cambio de entalpía asociado a una reacción a presión constante se conoce como entalpía de reacción o calor de reacción, y se denota como  $\Delta_r H$  [15]. Este cambio de entalpía puede ser exotérmico (liberando calor) o endotérmico (absorbiendo calor), y se expresa matemáticamente como:

$$\Delta_r H^\circ = H_{\text{productos}}^\circ - H_{\text{reactivos}}^\circ \quad (4.44)$$

donde  $\Delta_r H^\circ$  representa la entalpía estándar de reacción, calculada como la diferencia entre la entalpía de los productos y la de los reactivos en sus respectivos estados estándar. Las técnicas experimentales usadas para poder calcular la entalpía de reacción  $\Delta_r H$  se conocen como calorimetría; esta técnica permite medir el flujo de calor en el proceso de la reacción química.

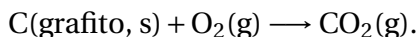
## 4.11. Entalpía de formación estándar

Un proceso esencial en el cálculo de datos termoquímicos es conocer la entalpía de formación estándar, esta es una propiedad fundamental de los compuestos, que se define como el cambio de entalpía que acompaña a la formación de 1 mol de compuesto en su estado estándar a partir de sus elementos constituyentes en sus estados estándar [108]. Este cambio de entalpía es también una entalpía de reacción,  $\Delta_r H$ , para el caso particular de la reacción de formación del compuesto a partir de sus elementos en sus estados estándar. El cambio de entalpía asociado a este proceso se conoce como entalpía de formación, denotada como  $\Delta_f H^\circ$ , donde el subíndice  $f$  indica que la sustancia se formó a partir de sus elementos químicos en su forma más estable.

La variación de entalpía de una reacción depende de la temperatura, la presión y el estado de agregación de reactivos y productos. Para hacer comparables estos valores entre distintas reacciones, se adopta el estado estándar, un conjunto de condiciones de referencia bajo el cual se tabulan la mayoría de las entalpías.

El *estado estándar* ( $^{\circ}$ ) de una sustancia es su forma pura a  $p^{\circ} = 1$  bar; la temperatura de referencia se indica explícitamente (comúnmente  $T = 298.15$  K). El cambio de entalpía estándar de reacción se denota  $\Delta_r H^{\circ}$  y corresponde al cambio de entalpía cuando todos los reactivos y productos están en sus estados estándar. La entalpía estándar de formación,  $\Delta_f H^{\circ}$ , es el cambio de entalpía al formar 1 mol de un compuesto a partir de sus elementos en sus estados estándar.

**Ejemplo.** Formación de dióxido de carbono en fase gas ( $T = 298.15$  K,  $p^{\circ} = 1$  bar):



Generalmente, los valores de  $\Delta_f H^{\circ}$  se reportan a  $T = 298.15$  K. Si un elemento existe en múltiples formas bajo condiciones estándar, la forma más estable del elemento se utiliza típicamente para la reacción de formación. Por convención,  $\Delta_f H^{\circ}$  de un elemento es cero únicamente para su forma más estable en el estado estándar (puro a  $p^{\circ} = 1$  bar y a la temperatura de referencia, usualmente 298.15 K). Así,  $\Delta_f H^{\circ}(\text{C}(\text{grafito})) = \Delta_f H^{\circ}(\text{H}_2(\text{g})) = \Delta_f H^{\circ}(\text{O}_2(\text{g})) = 0$ , mientras que otras alotropías o fases no estándar (p. ej., C(diamante)) tienen  $\Delta_f H^{\circ} \neq 0$ . Como se mencionó anteriormente, no podemos determinar el valor absoluto de la entalpía de una sustancia. Solo se pueden calcular valores relativos a una referencia arbitraria. En termodinámica, nos interesan los cambios en  $H$ . Aunque cualquier valor de  $\Delta_f H^{\circ}$  asignado arbitrariamente funcionaría como referencia, usar ceros simplifica los cálculos.

#### 4.11.1. Método de atomización

Existen tres métodos principales para calcular la entalpía de formación de un compuesto:

- Método de formación

- Método de atomización
- Método de reacciones isodésmicas

Una reacción isodésmica es un proceso químico en el cual se conserva el número de enlaces de un tipo formal dado, pero con un cambio en la relación entre ellos [27]. En otras palabras, en una reacción isodésmica, el número total y tipo de enlaces químicos (como enlaces simples, dobles, etc.) se mantiene constante entre los reactivos y los productos, aunque estos enlaces pueden estar reorganizados de manera diferente. Aunque el método de reacciones isodésmicas puede proporcionar resultados muy cercanos a los valores experimentales <sup>1</sup> no es utilizado por nuestro grupo debido a su alta sensibilidad a la selección de reacciones, lo que compromete su precisión desde una perspectiva de principios físicos, es decir, es muy sensible a la selección de las reacciones propuestas para acercarnos a un valor experimental y en muchas ocasiones no se reportan todas las reacciones propuestas. El método de formación se basa en calcular una “pseudo entalpía de formación” a partir de los átomos en sus estados más estables. No obstante, este enfoque es menos preciso comparado con el método de atomización; si se desea entrar en profundidad en el método de formación y de reacciones isodésmicas, se puede revisar la siguiente referencia [59]. Por estas razones, preferimos el método de atomización, ya que proporciona un valor único que no depende de la elección de una reacción específica. En términos de precisión, los métodos se clasifican generalmente en el siguiente orden: reacciones isodésmicas >atomización >formación [59]. Para calcular la entalpía de formación estándar a 298.15 K ( $\Delta_f H_{298.15K}^\circ$ ) de un compuesto con fórmula empírica ( $C_a H_b O_c N_d$ ), primero es necesario determinar la entalpía de formación a 0 K ( $\Delta_f H_{0K}^\circ$ ). Este cálculo se realiza mediante el método de atomización, que implica calcular el cambio de entalpía necesario para romper todos los enlaces de una molécula, descomponiéndola en átomos individuales, y predecir las entalpías de formación de los átomos a su estado estándar. La entalpía de formación a 0 K se calcula utili-

---

<sup>1</sup>Si se seleccionan las reacciones de manera cuidadosa puede ser un método preciso, pese a ello, esta elección es arbitraria, lo que podría resultar en una cancelación incorrecta de errores al hacer comparaciones con los valores experimentales.

zando la ecuación:

$$\begin{aligned} \Delta_f H_{0\text{K}}^\circ(\text{C}_a\text{H}_b\text{O}_c\text{N}_d) &= a\Delta_f H_{0\text{K}}^\circ(\text{C}(\text{s, grafito})) + \frac{b}{2}\Delta_f H_{0\text{K}}^\circ(\text{H}_2(\text{g})) \\ &+ \frac{c}{2}\Delta_f H_{0\text{K}}^\circ(\text{O}_2(\text{g})) + \frac{d}{2}\Delta_f H_{0\text{K}}^\circ(\text{N}_2(\text{g})) \\ &- \text{EA}(\text{C}_a\text{H}_b\text{O}_c\text{N}_d). \end{aligned} \quad (4.45)$$

Las entalpías de formación a 0 K de carbono, hidrógeno molecular, oxígeno molecular y nitrógeno molecular en sus estados estándar se obtienen de tablas de referencia experimentales. La energía de atomización del compuesto, denotada como  $\text{EA}(\text{C}_a\text{H}_b\text{O}_c\text{N}_d)$ , se calcula a partir de las energías de los átomos libres:

$$\begin{aligned} \text{EA}(\text{C}_a\text{H}_b\text{O}_c\text{N}_d) &= aE(\text{C}({}^3\text{P})) + bE(\text{H}({}^2\text{S})) + cE(\text{O}({}^3\text{P})) + dE(\text{N}({}^4\text{S})) \\ &- E(\text{C}_a\text{H}_b\text{O}_c\text{N}_d). \end{aligned} \quad (4.46)$$

Esta última ecuación (4.46) también es considerada una entalpía de atomización,  $\Delta_a H_{0\text{K}}^\circ$ , ya que la entalpía es igual a la energía electrónica a 0 K. Para obtener la entalpía a 298.15 K, es necesario considerar las contribuciones térmicas ( $\Delta H_T(x)$ ), que se derivan de la mecánica estadística. Estas incluyen principalmente los modos normales de vibración, además de las contribuciones rotacionales y traslacionales para moléculas no lineales, cuyos valores son ambos iguales a  $\frac{3}{2}RT$ . La contribución térmica total se calcula con la siguiente expresión, válida para moléculas no lineales:

$$\Delta H_T(x) = \underbrace{\frac{3}{2}RT}_{\text{tras}} + \underbrace{\frac{3}{2}RT}_{\text{rot}} + \underbrace{RT \sum_i \left( \frac{h\nu_i}{kT} \right) \left( \frac{\exp\left(-\frac{h\nu_i}{kT}\right)}{1 - \exp\left(-\frac{h\nu_i}{kT}\right)} \right)}_{\text{vib térmica}}. \quad (4.47)$$

La corrección térmica de la entalpía  $\Delta H_T(x)$  se descompone en contribuciones traslacional, rotacional y vibracional para una molécula en fase gas ideal:

- $\frac{3}{2}RT$  (*tras*): aporte traslacional. Proviene de los tres grados de libertad traslacionales de un gas ideal (equipartición:  $\frac{1}{2}RT$  por grado de libertad). Unidades:  $\text{kJ mol}^{-1}$ .
- $\frac{3}{2}RT$  (*rot*): aporte rotacional en el modelo de rotor rígido para moléculas no lineales (tres ejes de rotación  $\Rightarrow \frac{3}{2}RT$ ). Si la molécula es lineal, este término es  $RT$  (dos ejes).

- $RT \sum_i \frac{h\nu_i}{kT} \frac{e^{-h\nu_i/(kT)}}{1 - e^{-h\nu_i/(kT)}}$  (*vib térmica*): contribución vibracional térmica en el modelo de oscilador armónico. La suma recorre los modos vibracionales  $i = 1, \dots, 3N - 6$  (no lineal) o  $3N - 5$  (lineal). Este término *no* incluye la energía de punto cero (ZPE), pues la ZPE se contabiliza en la energía a 0 K. A bajas temperaturas y frecuencias altas, la contribución tiende a cero; a temperaturas altas, cada modo tiende a aportar  $\approx RT$ . Si se emplea un factor de escala de frecuencias  $f$ , debe aplicarse aquí:  $\nu_i \rightarrow f \nu_i$ .

Sumando  $\Delta_f H_{0K}^\circ(C_a H_b O_c N_d)$ ,  $\Delta H_T(x)$  y las entalpías de formación de los elementos a 298.15 K ( $\sum \Delta H_T(\text{elementos})$ ), el cual se obtiene usando información experimental, se obtiene la entalpía de formación del compuesto:

$$\Delta_f H_T^\circ(C_a H_b O_c N_d) = \Delta_f H_{0K}^\circ + \Delta H_T(x) + \sum \Delta H_T(\text{elementos}). \quad (4.48)$$

De esta manera, se obtiene la entalpía de formación  $\Delta_f H_T^\circ(C_a H_b O_c N_d)$  del compuesto en cuestión.

#### 4.11.2. Termodinámica estadística: contribuciones térmicas

Todos los métodos descritos anteriormente corresponden al uso de la mecánica cuántica y, en particular, de la química computacional para obtener la energía electrónica a nivel molecular, geometría y otras propiedades. No obstante, no es posible obtener o calcular propiedades como la capacidad calorífica  $C_p$ , los cambios de entalpía  $\Delta H$  y la entropía  $\Delta S$  directamente. El área de la fisicoquímica que nos permite hacer esta conexión del mundo microscópico al macroscópico es la termodinámica estadística. Con este objetivo hacemos uso de la función de partición  $Z$  [109].

$$Z(T) = \sum_i g_i e^{-\epsilon_i/k_B T}, \quad (4.49)$$

la suma se extiende a todos los posibles niveles de energía  $\epsilon_i$  y  $g_i$  es la degeneración del nivel de energía. Generalmente, uno elige el nivel de energía más bajo para que sea el 0 con el fin de que no existan energías negativas. A partir de la función de partición, es posible derivar todas las funciones termodinámicas. En este contexto, nos limitaremos a emplear las ecuaciones que

facilitan el cálculo de la diferencia relativa de entalpía, como se puede ver en la ecuación de Kirchhoff [15].

$$H(T) - H(0K) = \int_0^T C_p dT = \frac{RT^2}{Z} \frac{\partial Z}{\partial T} + RT. \quad (4.50)$$

Al usar el modelo de gas ideal, se asume que en la función de partición los movimientos traslacionales, rotacionales, vibracionales y electrónicos se encuentran desacoplados, es decir, no se ven influenciados entre sí. Además, se asume que la energía necesaria para acceder al primer estado electrónico excitado es mayor que  $kT$ , por lo tanto, se supone que no es posible acceder ni al primer estado excitado ni a ningún otro a cualquier temperatura dada [110]. Por lo tanto, podemos escribir una función de partición de la siguiente forma:

$$Z_{total} = Z_{rotacional} Z_{vibracional} Z_{traslacional} Z_{electrónica}. \quad (4.51)$$

No obstante, cuando salimos de este modelo, es decir, vamos más allá del modelo armónico, los términos  $Z_{rotacional}$  y  $Z_{vibracional}$  se encuentran acoplados. De esta forma, una descripción tan detallada al incluir términos anarmónicos incrementa el costo computacional considerablemente. Pese a ello, Jacobsen y *et al.* [111] encontraron que los cálculos de frecuencias vibracionales anarmónicas usando niveles de teoría HF, MP2, B3LYP, PBE no son ventajosos cuando se utilizan conjuntos de bases como 6-31G, 6-31+G(d,p), cc-pVTZ, ya que no ofrecen una mejora significativa en precisión que justifique su mayor costo computacional en comparación con las correcciones armónicas escaladas.

### Contribución traslacional

Para obtener la función de partición traslacional  $Z_{trans}$  de una molécula en una caja cúbica de volumen  $V_m = \frac{RT}{p}$  (volumen molar de un gas ideal a temperatura  $T$  y presión  $p$ ), se utiliza la aproximación integral. Dado que los niveles discretos de energía traslacional están muy próximos entre sí, se reemplaza la suma discreta por una integral continua para facilitar el cálculo. En su lugar, la suma se aproxima mediante una integral para obtener las siguientes ecuaciones. Esta aproximación es adecuada cuando  $m^{3/2} T^{5/2} p^{-1} \gg h^3 (2\pi)^{-3/2} k^{-5/2}$ . A presión estándar  $p = 1 \text{ bar} = 10^5 \text{ Pa} = 0.986923 \text{ atm}$ , esta condición se cumple para moléculas suficientemente pesadas,  $m \gg 11.4 T^{-5/3}$ , y para temperaturas suficientemente altas,  $T \gg 4.31 m^{3/5}$  (donde  $m$

se expresa en una). Para un gas ideal monoatómico, únicamente está presente el movimiento traslacional, ya que no existen grados de libertad vibracionales o rotacionales [110].

$$Z_{\text{trans}} = (2\pi mkT)^{3/2} h^{-3} V. \quad (4.52)$$

Además, el cambio de entalpía traslacional desde 0 K se puede expresar como:

$$[H(T) - H(0)]_{\text{trans}} = \frac{3}{2} RT. \quad (4.53)$$

### Contribución rotacional

La función de partición rotacional describe la cuantización de la rotación libre de una molécula rígida, donde el momento angular y su proyección son múltiplos enteros de  $\hbar/2\pi$ , lo cual restringe la energía rotacional a ciertos niveles discretos. Los espectros rotacionales se caracterizan por las constantes  $A$ ,  $B$  y  $C$ , donde  $A = \frac{h^2}{8\pi^2 I_A}$ , y de manera similar para  $B$  y  $C$ . Las cantidades  $I_A$ ,  $I_B$  e  $I_C$  representan los momentos de inercia principales de la molécula, siguiendo la convención  $I_A < I_B < I_C$  (o  $A > B > C$ ). Muchos programas, incluido Gaussian, reportan las constantes rotacionales cuando se les proporciona la geometría molecular [110]. Los momentos de inercia también pueden calcularse manualmente como los autovalores del tensor de inercia, cuyas componentes son, por ejemplo,  $I_{xy} = \sum m_i x_i y_i$  e  $I_{xx} = \sum m_i (y_i^2 + z_i^2)$ , donde el índice  $i$  recorre todos los átomos de la molécula y el origen de coordenadas se encuentra en el centro de masa. Las moléculas lineales (donde  $I_A = 0$ ) se describen mediante una única constante rotacional,  $B$ , y un único momento de inercia,  $I$ . Los detalles adicionales se pueden encontrar en los libros de texto de espectroscopía molecular [109, 112]. Las ecuaciones (4.55) y (4.53) corresponden a la contribución rotacional a la entalpía, que se deducen de la función de partición rotacional, para moléculas lineales y no lineales, respectivamente:

$$[H(T) - H(0)]_{\text{rot}} = RT. \quad (4.54)$$

Para una molécula no lineal, la función de partición es:

$$[H(T) - H(0)]_{\text{rot}} = \frac{3}{2} RT. \quad (4.55)$$

### Contribución vibracional

Para completar el modelo básico de rotor rígido/oscilador armónico (RR/OA), es necesario considerar las vibraciones moleculares. Como se mencionó en la discusión sobre la energía del punto cero (ZPE, ecuación 4.9), una molécula con  $N$  átomos tiene  $3N - 6$  frecuencias vibracionales, o  $3N - 5$  para el caso de moléculas lineales. Para calcular el cambio en la entalpía vibracional desde 0 K, se utiliza la siguiente expresión:

$$[H(T) - H(0)]_{\text{vib}} = RT \sum_i \left( \frac{h\nu_i}{kT} \right) \frac{e^{-\frac{h\nu_i}{kT}}}{1 - e^{-\frac{h\nu_i}{kT}}}, \quad (4.56)$$

donde  $R$  es la constante de los gases,  $T$  es la temperatura,  $h$  es la constante de Planck,  $\nu_i$  representa las frecuencias vibracionales, y  $k$  es la constante de Boltzmann [110]. Esta ecuación proporciona el cambio de entalpía debido a las vibraciones moleculares en función de la temperatura [113].

## Metodología

En este capítulo describimos la metodología para modificar el método G4. Nuestro objetivo es mejorar el cálculo teórico de la entalpía de formación a 298.15 K en moléculas problemáticas como nitrofenilos y heterociclos. Pese a ello, el método G4 es un método compuesto implementado en Gaussian, el software más popular en la química computacional para cálculos de estructura electrónica. La problemática reside en que los usuarios de este software no tienen la posibilidad de modificar el código y/o los parámetros del método G4 ni de otros métodos implementados. Esta limitación nos lleva a explorar un enfoque que permitirá reproducir y modificar el método G4 en pasos. El enfoque planteado es el siguiente: primero, reproducir el método G4 paso a paso; segundo, modificar algunos de estos pasos y reproducir los cálculos para obtener la entalpía de formación a 298.15 K; tercero, desarrollar un algoritmo iterativo para ajustar los parámetros de  $F$  y  $A$ . Por último, se muestran los conjuntos de calibración calculados que se utilizarán en esta tesis y se discutirá su influencia en el capítulo de resultados y discusiones.

### 5.1. Revisión del método G4

A continuación se presenta una descripción del método G4, incluyendo las ecuaciones y parámetros explorados que se emplean en nuestra implementación. La energía del método com-

puesto G4 se expresa mediante la siguiente ecuación:

$$\begin{aligned}
 E_0(G4) = & E[\text{CCSD(T,FC)}/6-31\text{G(d)}/\text{B3LYP}/6-31\text{G(2df,p)}] \\
 & + \Delta E(+)+\Delta E(2\text{df,p})+\Delta E(\text{G3largeXP})+\Delta E(\text{HF})+ \\
 & +E_{\text{HLC}}(A)+ZPE(F)+\Delta E(\text{SO}).
 \end{aligned}
 \tag{5.1}$$

La definición de los componentes de cada término en las ecuaciones del método G4 es la siguiente:

**Corrección de funciones difusas  $\Delta E(+)$ :** La inclusión de funciones difusas en el cálculo de la energía MP4 proporciona una corrección para describir mejor la densidad electrónica en moléculas que exhiben interacciones de largo alcance. Este término captura efectos de electrones alejados del núcleo, mejorando la precisión en el cálculo de energías para moléculas grandes o polares.

$$\begin{aligned}
 \Delta E(+) = & E[\text{MP4(FC)}/6-31+G(\text{d})/\text{B3LYP}/6-31\text{G(2df,p)}] - \\
 & E[\text{MP4(FC)}/6-31\text{G(d)}/\text{B3LYP}/6-31\text{G(2df,p)}].
 \end{aligned}
 \tag{5.2}$$

**Corrección de funciones de polarización adicional  $\Delta E(2\text{df,p})$ :**

Las funciones de polarización permiten flexibilidad en la forma de los orbitales atómicos, especialmente en átomos pesados, lo cual es crucial para una descripción precisa de la correlación electrónica. Al incluir funciones de polarización adicionales, este término mejora la capacidad del método para describir distribuciones de carga complejas y efectos de correlación.

$$\begin{aligned}
 \Delta E(2\text{df,p}) = & E[\text{MP4(FC)}/6-31\text{G(2df,p)}/\text{B3LYP}/6-31\text{G(2df,p)}] - \\
 & E[\text{MP4(FC)}/6-31\text{G(d)}/\text{B3LYP}/6-31\text{G(2df,p)}].
 \end{aligned}
 \tag{5.3}$$

**Corrección de funciones de mayor nivel de correlación  $\Delta E(\text{G3largeXP})$ :**

Primero, el conjunto de bases G3Large que fue desarrollado por Mayhal *et al.* [114], es una base de triple zeta con funciones de polarización y difusas que ha sido optimizada para describir estados electrónicos e interacciones de electrones en metales de transición. Esta base fue usada

en el desarrollo de G3, G3MP2. El conjunto de bases G3LargeXP es una extensión del conjunto de bases G3Large utilizado previamente en los métodos compuestos anteriormente citados. La extensión incluye funciones de polarización adicionales, marcadas como XP (extra polarization). No obstante, la base G3LargeXP no se encuentra liberada por lo que no es posible utilizar otros software de estructura electrónica para reproducir la energía a nivel G4. Este término corrige deficiencias residuales en la energía de correlación al usar bases más pequeñas en cálculos a nivel MP2.

$$\begin{aligned} \Delta E(\text{G3largeXP}) = & E[\text{MP2(FULL)}/\text{G3largeXP} // \text{B3LYP}/6 - 31\text{G}(2\text{df},\text{p})] - \\ & E[\text{MP2(FC)}/6 - 31\text{G}(2\text{df},\text{p}) // \text{B3LYP}/6 - 31\text{G}(2\text{df},\text{p})] - \\ & E[\text{MP2(FC)}/6 - 31 + \text{G}(\text{d}) // \text{B3LYP}/6 - 31\text{G}(2\text{df},\text{p})] + \\ & E[\text{MP2(FC)}/6 - 31\text{G}(\text{d}) // \text{B3LYP}/6 - 31\text{G}(2\text{df},\text{p})]. \end{aligned} \quad (5.4)$$

**Energía Hartree-Fock extrapolada al límite  $\Delta E(\text{HF})$ :** La extrapolación al límite de base de Hartree-Fock reduce errores sistemáticos asociados a la incompletitud del conjunto de bases. Esto se logra mediante un esquema de extrapolación de dos puntos con conjuntos de bases aumentados, esencial para obtener una energía total precisa en sistemas de moléculas grandes o pesadas.

$$\Delta E(\text{HF}) = E[\text{RHF}/\text{límite}] - E[\text{RHF}/\text{G3largeXP}]. \quad (5.5)$$

## Corrección de alto nivel (HLC)

**Corrección de alto nivel (HLC):** Las correcciones de alto nivel (*Higher Level Correction*, HLC) contrarrestan deficiencias residuales en el cálculo de la energía, según el tipo de sistema (molécula de capa cerrada, molécula de capa abierta o átomo). Estas correcciones son empíricas y se basan en el número de electrones de valencia, permitiendo un ajuste fino que mejora la precisión de las energías calculadas para diferentes tipos de sistemas moleculares y atómicos.

La expresión general del HLC está dada por:

$$E_{\text{HLC}} = \begin{cases} -An_{\text{val}} & \text{para moléculas de capa cerrada,} \\ -A'n_{\text{val}} - B(n_{\alpha} - n_{\beta}) & \text{para moléculas de capa abierta,} \\ -Cn_{\text{val}} - D(n_{\alpha} - n_{\beta}) & \text{para átomos (neutros y cargados),} \end{cases}$$

donde:

- $n_{\text{val}}$ : Número de electrones de valencia.
- $n_{\alpha}$ : Número de electrones de espín alfa.
- $n_{\beta}$ : Número de electrones de espín beta.

El término  $E_{\text{HLC}}$  depende de los parámetros  $A$ ,  $A'$ ,  $B$ ,  $C$  y  $D$ , ajustados para minimizar la Desviación Media Absoluta (DMA) respecto a datos experimentales. Además, para sistemas con un único par de electrones en la capa de valencia, se introduce un parámetro adicional  $E$ , cuya forma es:

$$E = 2.745 \text{ mHartree.}$$

Los valores ajustados para cada parámetro son:

$$A = 6.947 \text{ mHartree, } A' = 7.128 \text{ mHartree, } B = 2.441 \text{ mHartree,}$$

$$C = 7.116 \text{ mHartree, } D = 1.414 \text{ mHartree.}$$

Estas correcciones han sido optimizadas utilizando el conjunto de calibración G3/05, para el método G4 y contiene 454 energías experimentales. Dicho conjunto incluye entalpías de formación, energías de ionización, afinidades electrónicas y protónicas, así como energías de enlace de hidrógeno. Por tanto, no se utiliza únicamente la entalpía de formación como referencia para ajustar estos parámetros.

El parámetro  $A$  se aplica exclusivamente en moléculas de capa cerrada y este parámetro es de vital importancia y será explorado en esta tesis, el cual está pensado para sistemas orgánicos sin radicales, mientras que  $A'$  y  $B$  corrigen deficiencias en sistemas de capa abierta, considerando la polarización de espín en radicales. Finalmente,  $C$  y  $D$  se enfocan en átomos, ya sean neutros o cargados [20].

#### **Energía de punto cero $ZPE(F)$ :**

La energía de punto cero, calculada a partir de frecuencias vibracionales armónicas y escalada por un factor empírico de 0.9854 para el método G4, incorpora la contribución vibracional molecular al valor total de energía. Este factor  $F = 0.9854$  corrige sistemáticamente errores provenientes de usar el modelo armónico y del tratamiento incompleto de la correlación electrónica en los cálculos a nivel G4 (que emplea B3LYP/6-31G(2df,p)) [113]. El uso de esta corrección resulta indispensable para obtener valores de energía en el estado base ( $T = 0$  K) que reflejen con mayor fidelidad los datos experimentales.

$$ZPE(F) = Fh \sum_i \nu_i. \quad (5.6)$$

Los términos  $E_{HLC}(A)$  y  $ZPE(F)$  en la ecuación 5.1 están influenciados por los valores del factor de escala  $F$  y el parámetro  $A$ . En Gaussian 09,  $F$  tiene un valor de 0.9854 y  $A$  de 6.947 mHartrees, únicamente para el método G4. El uso del parámetro  $F$  es imprescindible, ya que el error habitual oscila entre un 5 y 10%, dependiendo del método de estructura electrónica y la base de datos empleada [115]. El valor de  $F = 0.9854$  se determinó para igualar las energías de punto cero en un conjunto de moléculas, utilizando datos obtenidos por Scott y Radom que son ampliamente utilizados [39, 40]. El método G4 emplea el concepto de acoplamiento espín-órbita ( $\Delta E(SO)$ ) para corregir la interacción entre el momento angular del espín y el momento angular orbital. Este efecto surge de la interacción intrínseca entre el espín y el movimiento orbital del electrón, y adquiere particular importancia en átomos pesados, donde los electrones cercanos al núcleo presentan altos valores de momento angular y están sometidos a fuertes campos eléctricos. Por ello, se incorpora un término de acoplamiento espín-órbita, ya sea mediante la

inclusión de parámetros y aproximaciones basadas en datos experimentales o cálculos teóricos, cuando dichos valores no están disponibles. En el caso de los átomos involucrados en los sistemas estudiados (C, N, O, H, S), se estima que esta corrección tiene una importancia reducida debido a la menor magnitud de la interacción entre el espín y el momento angular orbital.

### 5.1.1. Pasos del cálculo G4

La ecuación 5.1 requiere la realización de diversos cálculos de estructura electrónica. Los 8 pasos necesarios para obtener una energía a nivel G4 son los siguientes:

1. Optimización de geometría a nivel B3LYP/6-31G(2df,p): Para obtener la estructura de mínima energía de la molécula.
2. Cálculo de frecuencias a nivel B3LYP/6-31G(2df,p): Para determinar las frecuencias vibracionales y calcular la ZPE.
3. Cálculo de energía electrónica a nivel CCSD(T)/6-31G(d): Para incluir correlación electrónica de alto nivel.
4. Cálculo de energía electrónica a nivel MP4/6-31+G(d): Para añadir correcciones por funciones difusas.
5. Cálculo de energía electrónica a nivel MP4/6-31G(2df,p): Para añadir correcciones por funciones de polarización.
6. Cálculo de energía electrónica a nivel MP2/G3largeXP: Para añadir correcciones por una base de mayor tamaño.
7. Cálculo de energía electrónica a nivel HF/mod-aug-cc-pVQZ: Este cálculo proporciona una aproximación inicial del límite de la base Hartree-Fock.
8. Cálculo de energía electrónica a nivel HF/mod-aug-cc-pV5Z: Este cálculo, complementario al anterior, permite realizar una extrapolación del límite Hartree-Fock.

### Modificaciones propuestas

En búsqueda de mejorar el acuerdo teórico-experimental del método G4 para el cálculo teórico de la entalpía de formación, nuestro trabajo se centra en modificar el método de estructura electrónica utilizado en los pasos 1 y 2. Además, examinamos cómo la modificación del conjunto de calibración afecta los parámetros  $F$  y  $A$  en el cálculo teórico de la entalpía de formación. A continuación, describimos cómo implementamos estas modificaciones.

## 5.2. Implementación Computacional

Para implementar estas modificaciones, utilizamos el software Gaussian 09 en los equipos de cómputo del Laboratorio de Fisicoquímica Orgánica Teórica, dirigido por el Dr. Julio Manuel Hernández Pérez, y del Laboratorio de Modelado Matemático y Desarrollo de Software Científico, dirigido por el Dr. Juan Manuel Solano Altamirano. El código programado para ajustar y modificar parámetros fue escrito en Bash y C++. A continuación, describimos el proceso de reproducción y modificación del método G4.

### 5.2.1. Reproducción y modificación del método G4

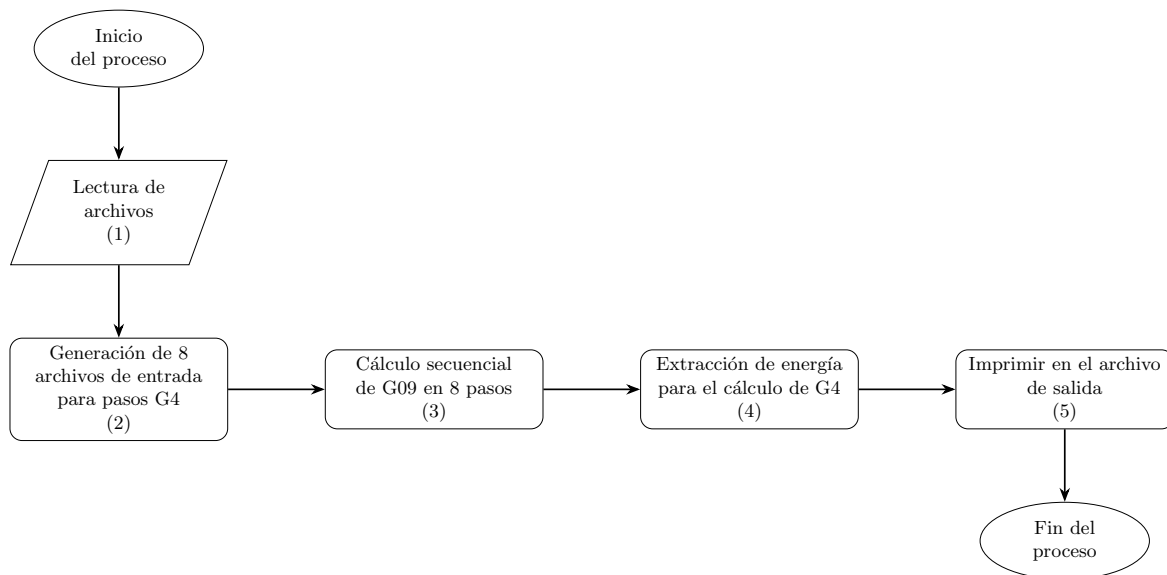
La figura 5.1 muestra los pasos generales para obtener un archivo de texto \*.dat, que será utilizado por el programa LABFQOTG4, el cual está programado en C++, para calcular la energía a nivel G4 de la ecuación 5.1, así como la entalpía de formación a 298.15 K ( $\Delta_f H_{298.15K}^\circ$ ). A continuación se detalla cada paso del proceso:

1. **Lectura de archivos:** Se realiza una lectura de archivos de texto con el formato name.txt. Dentro de este archivo, en formato de matriz-Z, se encuentra la estructura molecular de cada sistema a calcular y el nombre de la molécula (name).

2. **Generación de 8 archivos de entrada para pasos G4:** Estos archivos name.txt se utilizan para generar los archivos de entrada o inputs de Gaussian, con la información geométrica tomada de la matriz-Z contenida en cada archivo name.txt.
3. **Cálculo secuencial en Gaussian 09 en 8 pasos:**
  - Optimización de la geometría: Se hace el paso 1 del método G4, es decir, se hace una optimización de geometría a nivel B3LYP/6-31G(2df,p). Una vez que el cálculo converge a la geometría de mínima energía o más estable, esta será la nueva geometría de referencia usada para los cálculos posteriores de G4, los pasos 3-8. Por motivos previamente descritos en antecedentes y en seguimiento de nuestros objetivos, exploramos distintos métodos de estructura electrónica para la optimización de geometría y el cálculo de frecuencias, los cuales fueron (MP2, M06-2X,  $\omega$ B97XD).
  - Cálculos posteriores: Se manda calcular desde el paso 3 hasta el paso 8 del método G4. Todos los cálculos se realizan de forma secuencial en Gaussian 09, es decir, no se inicia ningún cálculo hasta que el anterior no esté terminado. Este proceso concluye una vez que se completen los 8 cálculos del método G4.
4. **Extracción de energía para el cálculo de G4:** Una vez terminados los 8 pasos de G4, se extraen los valores de energía de estos archivos de salida generados por Gaussian 09, también llamados comúnmente outputs.
5. **Imprimir en el archivo de salida:** Se recopila y organiza todos los datos adquiridos de los archivos de salida para generar un informe con terminación \*.dat. Este archivo se utiliza como entrada para el programa LABFQTOG4 que describiremos en la brevedad, para reproducir la energía a nivel G4 y posteriormente calcular la entalpía de formación a 298.15 K.

Modificando esta parte del código en la generación de los 8 archivos de entrada para los pasos de G4, es donde podemos cambiar el método de estructura electrónica usado en la geometría de partida en los cálculos de G4 y el cálculo de frecuencias en los pasos 1 y 2.

En las figuras 5.2 y 5.3 se muestra el diagrama de flujo para el cálculo teórico de la entalpía de



**Figura 5.1:** Diagrama de flujo del programa ExtractLabfqotG4.sh para obtener los archivos \*.dat.

formación a 298.15 K mediante el método G4, dividido en dos secciones para una mayor claridad. La figura 5.2 muestra las primeras etapas del cálculo, mientras que la figura 5.3 contiene las etapas finales a partir del cálculo de las contribuciones térmicas.

### Parte A:

1. **Lectura de archivo:** En este punto se da lectura al archivo con terminación \*.dat que contiene toda la información de la energía de los 8 pasos para reproducir la energía a nivel G4 usando la ecuación 5.1.
2. **Cálculo de la energía a nivel G4:** En esta etapa primero se reproduce el valor de energía a nivel G4 (véase la ecuación 5.1). Nosotros introducimos un valor definido de  $F$  y  $A$  como argumentos, es decir, nosotros podemos modificar o variar estos parámetros y en consecuencia modificará la energía de la expresión final de  $E_0(G4)$ , recordando que los valores estándar son  $F = 0.9854$  y  $A = 6.947$  mHartrees.
3. **Utilización de energía G4 átomos:** En esta etapa se da lectura al archivo \*.dat de cada molécula para identificar qué tipo de átomo contiene esta molécula, así como la este-

quiometría de cada una de ellas, para entonces utilizar la energía de los átomos a nivel G4.

4. **Calcular la entalpía de atomización a 0 K de la molécula:** En esta etapa se usa el método de atomización para obtener la entalpía de atomización de una molécula. Por ejemplo, si usamos la molécula de agua ( $\text{H}_2\text{O}$ ), entonces la expresión de atomización sería de la siguiente forma:

$$\Delta_a H_{0\text{K}}^\circ(\text{H}_2\text{O}) = \Delta E_{0\text{K}}^{\text{total}}(\text{O}(^3\text{S}) + 2\text{H}(^2\text{S})) - \Delta E_{0\text{K}}^{\text{total}}(\text{H}_2\text{O}). \quad (5.7)$$

Donde el término  $\Delta E_{0\text{K}}^{\text{total}}(\text{H}_2\text{O})$  se obtiene usando la energía a nivel G4 de la molécula de  $\text{H}_2\text{O}$  y  $\Delta E_{0\text{K}}^{\text{total}}(\text{O}(^3\text{S}) + 2\text{H}(^2\text{S}))$  corresponde a las energías electrónicas de los átomos libres que forman la molécula del agua, esta energía por motivos de nuestra tesis varía dependiendo de los valores que tome  $F$  y  $A$ .

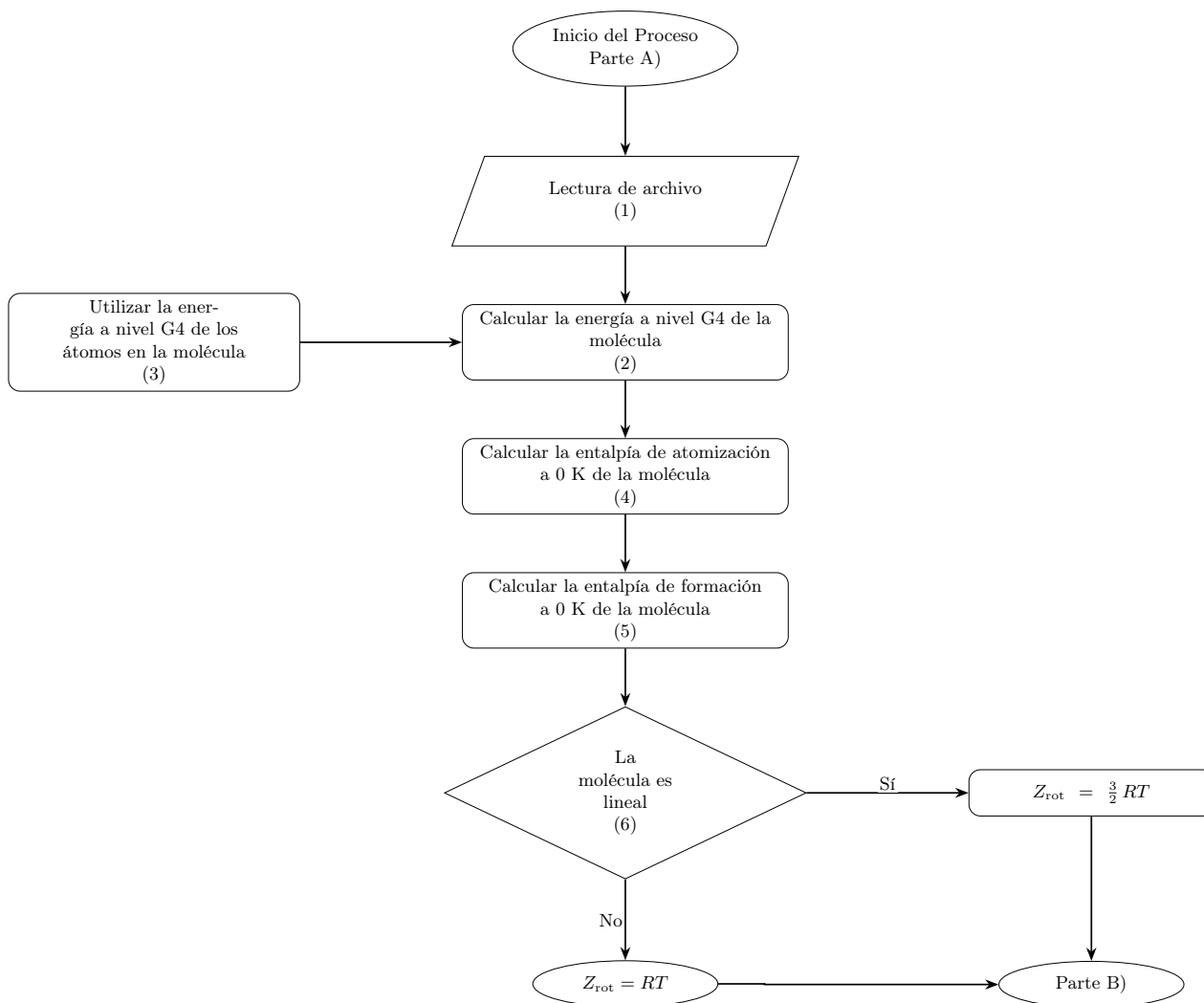
5. **Cálculo de la entalpía de formación a 0 K:** En esta etapa, una vez obtenida la entalpía de atomización, se calcula a partir de ella y de la entalpía requerida para llevar las especies constituyentes desde sus estados estándar hasta átomos gaseosos libres; los valores experimentales de referencia se toman del NIST [54]. En otras palabras, la entalpía de formación del agua a 0 K ( $\Delta_f H_{0\text{K}}^\circ(\text{H}_2\text{O})$ ) se calcula con la siguiente ecuación:

$$\Delta_f H_{0\text{K}}^\circ(\text{H}_2\text{O}) = \Delta H_{0\text{K}}^\circ(\text{H}_2(\text{g})) + \frac{c}{2} \Delta H_{0\text{K}}^\circ(\text{O}_2(\text{g})) - \Delta_a H_{0\text{K}}^\circ(\text{H}_2\text{O}). \quad (5.8)$$

En la ecuación 5.8, el valor de  $c$  representa el número de átomos de oxígeno que participan en la formación de la molécula de agua, por lo que  $c = 1$  para  $\text{H}_2\text{O}$ . Además, el término  $\Delta_a H_{0\text{K}}^\circ(\text{H}_2\text{O})$  corresponde a la entalpía de atomización del agua a 0 K, calculada con la ecuación (5.7) y representa la energía necesaria para descomponer la molécula de agua en átomos libres en estado gaseoso.

6. **Molécula lineal o no lineal:** En esta etapa es importante tomar en cuenta dos posibilidades para poder calcular las próximas estimaciones en la entalpía: la primera de ellas es estimar la linealidad de la molécula. Si la molécula es lineal, la contribución rotacional es  $RT$ ; si es no lineal, es  $\frac{3}{2}RT$ . Concluida la determinación, se procede a la siguiente etapa en

el cálculo de las contribuciones térmicas, lo cual nos permitirá llevar la entalpía de 0 K a 298.15 K o cualquier temperatura deseada.



**Figura 5.2:** Diagrama de flujo del programa LABFQOTG4, parte A.

### Parte B:

- 7. Calcular contribuciones térmicas:** En este punto se determina la suma de las contribuciones rotacionales, vibracionales, traslacionales y electrónicas de la molécula, principal-

mente estudiamos moléculas no lineales, por lo tanto la expresión general es la siguiente:

$$\Delta H_T(\text{H}_2\text{O}) = \frac{3}{2}RT + \frac{3}{2}RT + RT \sum_i \left( \frac{(F)h\nu_i}{kT} \right) \left( \frac{\exp\left(-\frac{(F)h\nu_i}{kT}\right)}{1 - \exp\left(-\frac{(F)h\nu_i}{kT}\right)} \right). \quad (5.9)$$

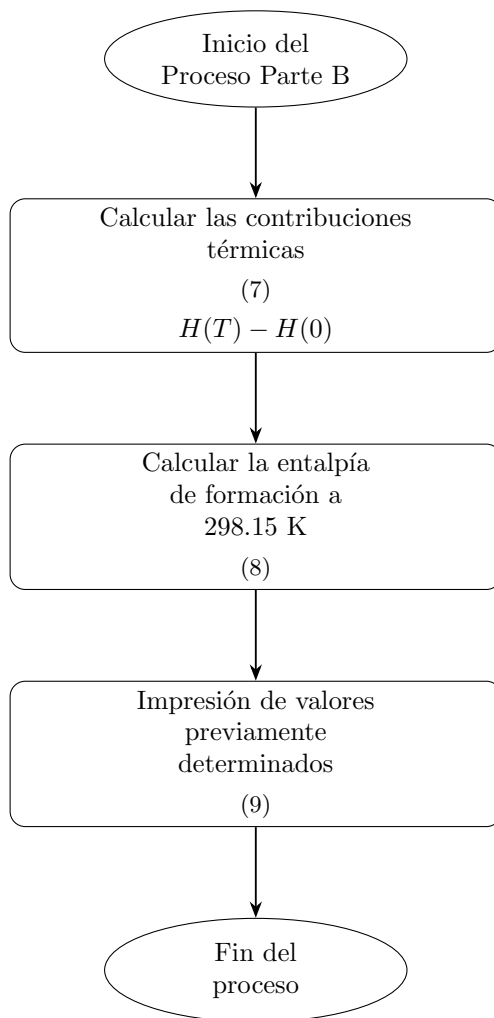
En esta expresión,  $\nu_i$  corresponde a las frecuencias de los modos normales de vibración de la molécula, las cuales se calcularon en el segundo paso del método G4 utilizando el funcional B3LYP. Además, estas frecuencias están escaladas por un factor de escala  $F$ ; por lo tanto, el factor de escala también impacta en las contribuciones térmicas.

Para una molécula no lineal, los términos de contribución rotacional corresponden a  $\frac{3}{2}RT$ , como se muestra en la ecuación anterior. Sin embargo, en el caso de una molécula lineal, el término de contribución rotacional se reduce a  $RT$  en el modelo de rotor rígido,

8. **Calcular la entalpía de formación a 298.15 K ( $\Delta_f H_{298.15\text{K}}^\circ$ ):** En este punto, se utiliza la entalpía de formación obtenida de la ecuación 5.8 y se suman las contribuciones térmicas de la ecuación 5.9. Además, se suman las entalpías de formación experimentales, restando los incrementos correspondientes para los elementos de la molécula en sus estados estándar, también usando referencias experimentales. Al final, obtenemos la ecuación 5.10 que corresponde a la entalpía de formación a  $T = 298.15 \text{ K}$ :

$$\Delta_f H_{298.15\text{K}}^\circ(\text{H}_2\text{O}) = \Delta_f H_{0\text{K}}^\circ + \Delta H_T(\text{H}_2\text{O}) + \left( 2\Delta H^\circ(\text{H}_2) + \frac{1}{2}\Delta H^\circ(\text{O}_2) \right). \quad (5.10)$$

9. **Impresión de valores previamente determinados:** Por último, se imprimen en pantalla todos los valores previamente determinados, es decir:  $E_0(\text{G4})$ ,  $\Delta_f H_{0\text{K}}^\circ(\text{H}_2\text{O})$ ,  $\Delta_f H_{298.15\text{K}}^\circ(\text{H}_2\text{O})$ .



**Figura 5.3:** Diagrama de flujo del programa LABFQOTG4, parte B.

### 5.2.2. El algoritmo de parametrización

La figura 5.4 ilustra el algoritmo empleado para la parametrización de los factores  $F$  y  $A$ , cuyo propósito es reducir al mínimo la desviación media absoluta en los cálculos de G4 en relación con la entalpía de formación teórica establecida en 298.15 K. Para llevar a cabo la parametrización, es necesario adherirse a una secuencia de etapas y a un proceso iterativo para alcanzar los valores óptimos de  $F$  y  $A$ . Se procederá a detallar cada etapa del algoritmo:

1. **Bucle sobre  $F$ :** Se ejecuta un bucle sobre el factor de escala  $F$ , variando desde 0.88 hasta 1.0 con pasos de 0.00001.
2. **Bucle sobre  $A$ :** Se ejecuta un bucle sobre el parámetro  $A$ , variando desde 4.0 hasta 9.0 mHartrees en pasos de 0.0001 mHartrees. Este parámetro empleado en el método compuesto G4 no posee un significado físico directo, pero representa un ajuste empírico desarrollado por Curtiss *et al.* [20] para mejorar la precisión en el cálculo de distintas propiedades. Este ajuste asegura una mayor concordancia entre los valores calculados y experimentales del conjunto de calibración, corrigiendo las deficiencias electrónicas no contempladas en otros pasos de G4. En nuestro trabajo, exploramos diferentes valores de  $A$  para cubrir un amplio rango de la superficie de parametrización y obtener el valor óptimo de  $A$  para ese conjunto.
3. **Lectura de archivos:** Se da lectura a los archivos necesarios para realizar los cálculos, es decir, los \*.dat que fue explicado en el diagrama 5.1. En esta parte nosotros definimos el conjunto de calibración y/o subconjuntos de moléculas a los que se les da lectura y que serán usados en este algoritmo.
4. **Cálculo de la entalpía de formación:**
  - Utilizando el procedimiento descrito en el diagrama 5.2 y 5.3 para cada combinación de valores de  $F$  y  $A$ , se calculan las entalpías de formación a 298.15 K de todas las moléculas del conjunto de calibración estudiado.
  - Los resultados se almacenan en archivos de reporte para cada valor de  $F$  y  $A$  evaluado.

5. **Lectura de la entalpía de formación experimental:** Se da lectura a la entalpía de formación experimental de cada molécula para comparar con los valores calculados y poder estimar:

$$\Delta_f H_{298.15\text{K}}^\circ(T - Exp) = \Delta_f H_{i,298.15\text{K}}^\circ(T_{F,A}) - \Delta_f H_{i,298.15\text{K}}^\circ(Exp), \quad (5.11)$$

donde la  $\Delta_f H_{i,298.15\text{K}}^\circ(T_{F,A})$  es la entalpía teórica usando el método G4 y el método de atomización que depende de un valor de  $F$  y  $A$  para una molécula  $i$  y el término  $\Delta_f H_{i,298.15\text{K}}^\circ(Exp)$  es el valor experimental de la entalpía de formación a 298.15 K de esa misma molécula.

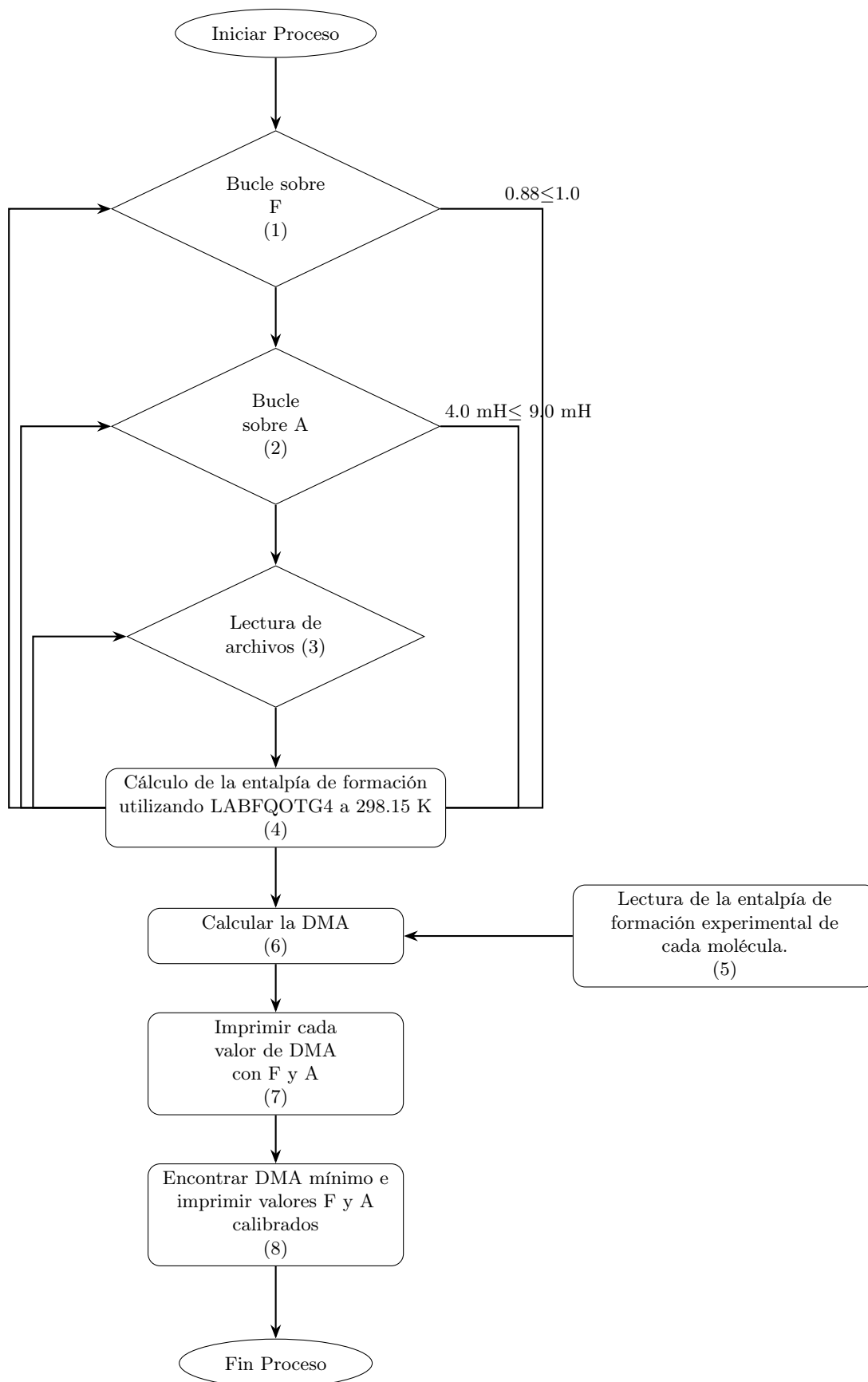
6. **Calcular la DMA:** Una vez evaluado todo el conjunto de calibración para un valor de  $F$  y  $A$ , calculamos ahora la desviación media absoluta con la siguiente ecuación:

$$DMA(F, A) = \sum_{i=1}^N \frac{|\Delta_f H_{i,298.15\text{K}}^\circ(T_{F,A}) - \Delta_f H_{i,298.15\text{K}}^\circ(Exp)|}{N}, \quad (5.12)$$

donde  $N$  es el número de moléculas que se está parametrizando para un conjunto de moléculas.

7. **Imprimir valores DMA con  $F$  y  $A$ :** Se registran e imprimen los valores de DMA correspondientes a cada par de valores  $F$  y  $A$ . Estos datos se almacenan para análisis posteriores. Este conjunto de datos nos proporciona una superficie de parametrización. Este ciclo termina al iterar en ciertos rangos todos los valores de  $F$  y  $A$ , para todas las moléculas.
8. **Encontrar DMA mínimo e imprimir valores de  $F$  y  $A$ :** Por último, evaluamos cuál es el valor de DMA más bajo y así obtenemos los valores óptimos de  $F$  y  $A$  que se encuentran reportados en el capítulo de resultados.

Para evaluar el rendimiento del método G4 en el cálculo teórico de la entalpía de formación a 298.15 K se emplea la desviación media absoluta (DMA), una métrica que indica qué tan cercanos están los valores teóricos de los experimentales; es decir, una DMA baja implica predicciones más precisas. Además, se utiliza la desviación estándar ( $\sigma$ , ver ecuación 5.13) para cuantificar la dispersión de los errores, aunque esta medida resulta más sensible a los valores atípicos. Por esta razón, se prefiere la DMA para estudios de este tipo [116].

**Figura 5.4:** Diagrama del flujo para la parametrización de  $F$  y  $A$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \Delta_f H_{i,T}^{\circ}(F, A) - \Delta_f H_{i,Exp}^{\circ} \right)^2}. \quad (5.13)$$

$$\mu = \frac{1}{N} \sum_{i=1}^N \left( \Delta_f H_{i,T}^{\circ}(F, A) - \Delta_f H_{i,Exp}^{\circ} \right). \quad (5.14)$$

Siguiendo las recomendaciones de Savin [65], el uso de la DMA como principal métrica se justifica en que esta medida ofrece una estimación más estable de la precisión, al verse menos afectada por outliers (valores atípicos), que son datos que difieren de manera significativa del resto de la población. Además reportamos el promedio de los errores ( $\mu$ ) y la desviación estándar ( $\sigma$ ) que complementa la evaluación del desempeño del método G4, especialmente al ajustar los parámetros  $F$  y  $A$  para minimizar la diferencia entre los valores teóricos y experimentales.

### 5.3. Conjuntos y subconjuntos de calibración

Para calibrar y determinar los parámetros óptimos  $F$  y  $A$  en nuestro método modificado, es fundamental seleccionar un conjunto de moléculas con información experimental confiable. El objetivo es ajustar dichos parámetros de modo que las predicciones teóricas se acerquen lo máximo posible a los datos experimentales. En este sentido, definimos tres conjuntos que abarcan distintas poblaciones de moléculas:

- **Subconjunto SO67:** Este subconjunto se conforma de 67 moléculas seleccionadas del conjunto G3/05 [21]. En esta selección se priorizaron moléculas de capa cerrada y cuya característica es un rango reducido en número de electrones. Cabe destacar que la escasa presencia de nitrofenilos y heterociclos no se debe a una exclusión deliberada, sino a que en el diseño original de los conjuntos G2 y G3/05 estas familias de compuestos estuvieron subrepresentadas. El objetivo de SO67 es, por tanto, construir un subconjunto que represente de manera consistente a G3/05 en el dominio de moléculas de capa cerrada, de modo que la parametrización resultante reproduzca un valor cercano al parámetro empírico

A empleado en el método G4. Posteriormente, la comparación con conjuntos ampliados (SD43 y SO110) permite evaluar el impacto de incorporar especies más complejas como nitrofenilos y heterociclos. Los valores experimentales correspondientes pueden consultarse en el Apéndice A, Tabla A.1.

- **Subconjunto SD43:** Este conjunto se extrajo del trabajo de Dorofeeva et al. [25], e incluye 43 moléculas que presentan desviaciones significativas en la entalpía de formación ( $\Delta_f H_{298.15K}^\circ(T-Exp)$ ) cuando se usa el método compuesto G4. Se seleccionaron aquellas moléculas que, además, no presentan más de un confórmero ni rotores impedidos, lo que facilita su cálculo y evita complicaciones adicionales. Este subconjunto abarca compuestos con un mayor número de electrones y contiene información específica sobre nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados, permitiendo evaluar la capacidad de nuestro método para corregir las deficiencias en estos sistemas, también pueden revisarse en el apéndice A, tabla A.1, los cuales corresponden de las moléculas 68 hasta 110.
- **Conjunto SO110:** Este conjunto se obtiene como la unión de SO67 y SD43 (es decir, SO110 = SO67  $\cup$  SD43). De esta manera, se engloba una amplia diversidad de compuestos orgánicos, tanto simples como aquellos con funciones particulares (por ejemplo, nitrofenilos y heterociclos). El objetivo es que SO110 sirva como un conjunto de calibración global para moléculas orgánicas de capa cerrada, permitiendo evaluar la capacidad predictiva de la metodología en un rango químico amplio.

La estrategia de parametrización se fundamenta en seleccionar inicialmente moléculas con valores experimentales definidos y, posteriormente, comparar las predicciones teóricas obtenidas con nuestros parámetros  $F$  y  $A$  optimizados. La efectividad de nuestra aproximación se sustentan en una validación cruzada entre los conjuntos de calibración y los conjuntos de prueba, conformados por moléculas que no forman parte de la calibración.

De este modo, se busca minimizar las discrepancias entre los valores teóricos y experimentales, aplicados a moléculas que presenten cierta característica química/electrónica que esperamos capturar, por ejemplo para SD43 esperamos que esta calibración al evaluarse a moléculas con

características estructurales y grupos funcionales parecido presente una mejor estimación con moléculas de este grupo. Tanto SO67, SD43, y en consecuencia SO110, incluyen solo moléculas orgánicas de capa cerrada, por lo tanto no es necesario investigar o reparametrizar otros parámetros de HLC que no actúen para moléculas de capa cerrada.

## 5.4. Consideraciones termoquímicas

La calibración de los parámetros  $F$  y  $A$  en el método G4 requiere disponer de un conjunto de calibración que abarque la diversidad molecular y cuente con referencias experimentales robustas. La selección de estas referencias constituye un desafío, ya que existen discrepancias entre bases de datos: Por ejemplo, la base NIST [54, 117] ofrece una amplia variedad de datos experimentales, aunque presenta cierta variabilidad en los valores reportados para un mismo sistema, mientras que la base AcTC [49] ofrece valores más precisos pero con menor amplitud en la cobertura química. Además, en el método de atomización resulta necesario disponer de datos experimentales relativos al cambio de entalpías de formación de 0 K a 298.15 K para los átomos en sus estados libres — información disponible en NIST pero no en AcTC —, lo que justifica nuestra preferencia por la base NIST siempre que se cuente con esta información.

Hemos optado por utilizar las referencias experimentales empleadas originalmente por Curtis y *et al.*, ya que sus valores se muestran coherentes con las predicciones del método G4 y permiten mantener una calibración cercana a la original en la medida de lo posible. Somos conscientes de que, a pesar de que algunas referencias experimentales puedan ser cuestionables como fue recomendando por B.Chan [118], en este trabajo se decide conscientemente adherirse a las mismas que se utilizaron en la calibración original de G4, evitando así introducir discrepancias adicionales.

La variabilidad y la falta de consenso en el valor recomendado para ciertos sistemas complican la elección de las referencias experimentales, especialmente cuando nuevos datos experimentales o teóricos pueden modificar dichos valores. Dado que esta información no siempre está disponible para todos los sistemas, se opta por trabajar con los conjuntos que garanticen la ma-

por coherencia posible.

En nuestro estudio, los valores de referencia para las entalpías de formación experimentales se toman del subconjunto SO67 utilizando la referencia [21], mientras que para compuestos con nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados corresponden al subconjunto SD43, se utilizan los valores reportados en [25]. Es importante destacar que se ha considerado únicamente un subconjunto del total de datos del G3/05, y que la calibración se centra exclusivamente en las entalpías de formación, sin abordar otras propiedades que también se calibraron originalmente en G4, también considerando únicamente moléculas orgánicas de capa cerrada.

En este capítulo se ha descrito la metodología seguida para modificar y mejorar la estimación teórica de la entalpía de formación a 298.15 K mediante el método G4. En el siguiente capítulo se discutirán los resultados obtenidos con las modificaciones propuestas, haciendo especial énfasis en cómo varían los parámetros  $F$  y  $A$  al emplear distintos conjuntos y subconjuntos de calibración.

## Resultados y discusión

A continuación, se presentan los resultados obtenidos en la presente tesis. Para su descripción, los resultados se dividen en tres partes. En la primera parte (Sección 6.1-2), se tratará de la superficie de parametrización para la obtención de los parámetros  $F$  y  $A$ , así como del cambio de distintos métodos de estructura electrónica de los pasos 1 y 2 del método G4 y su efecto en el cálculo teórico de entalpías de formación. En la segunda parte (Sección 6.3) se hablará del efecto de modificar los conjuntos de calibración sobre los parámetros  $F$  y  $A$  del método G4 y sus consecuencias en términos de la desviación teórico-experimental para el conjunto de calibración SO110. Finalmente, en la sección (6.4-6) se mostrará el desempeño de nuestros parámetros utilizando dos conjuntos de prueba y, además, se comparará con el promedio de las  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  usando reacciones isodésmicas para un conjunto de moléculas. Nuestro objetivo es verificar si las modificaciones implementadas al método compuesto G4 mejoran la entalpía de formación para los nitrofenilos y heterociclos sin empeorar en gran medida la precisión en otros compuestos.

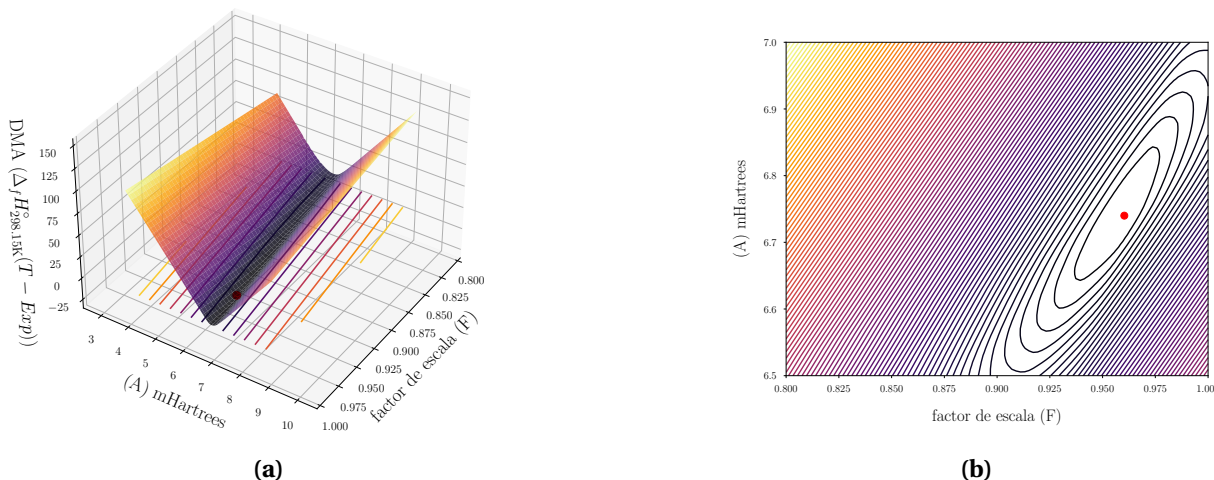
## 6.1. Superficie de parametrización

Se evaluaron distintos valores de  $F$  y  $A$ , aplicándolos a todas las moléculas del conjunto de calibración. Para cada par ( $F$  y  $A$ ) se calculó la (DMA) entre los valores teóricos y los experimentales de  $\Delta_f H_{298.15K}^\circ$ . La superficie de error obtenida (representada por los valores de la DMA en función de  $F$  y  $A$ ) depende de tres factores: el conjunto de moléculas empleado en la calibración, las referencias experimentales utilizadas y cualquier modificación en los pasos del método G4. Tanto en la Figura 6.1a como en la Figura 6.1b se observa que  $F$  y  $A$  están fuertemente acoplados, lo que impide analizarlos de forma independiente. A diferencia de lo propuesto por da Rocha y R. Custorio [119], quienes estudiaron la influencia del factor de escala en el método G3X y sugirió el uso de factores específicos para subconjuntos de G3/05.

En nuestro trabajo se busca identificar conjuntamente los valores óptimos de  $F$  y  $A$  que minimicen la DMA, parámetros ideales para el conjunto de calibración evaluado y para cualquier modificación aplicada al método compuesto.

La Figura 6.1a muestra la superficie en 3D para el conjunto SO110, donde el eje  $x$  representa los valores de  $F$ , el eje  $y$  los de  $A$  (en mHartrees) y el eje  $z$  la DMA en  $\text{kJ} \cdot \text{mol}^{-1}$ . Se observa que, al mantener  $F$  en su valor óptimo, ligeras variaciones en  $A$  (del orden de  $\pm 0.09522$  mHartrees) generan cambios significativos en la DMA (aproximadamente  $3.5 \text{ kJ} \cdot \text{mol}^{-1}$ ), evidenciando la alta sensibilidad de la corrección electrónica aportada por  $A$  frente al factor de escala  $F$ . La Figura 6.1b presenta la misma información mediante curvas de nivel en 2D, lo que facilita la visualización del mínimo global y descarta la presencia de mínimos locales, permitiendo identificar claramente los valores óptimos de  $F$  y  $A$  que minimizan la DMA.

Los análisis se complementan con los gráficos adicionales en las Figuras A.1 y A.2 del Apéndice A. En estas figuras se pueden ver las distribuciones de los errores  $\Delta_f H_{298.15K}^\circ(T - Exp)$  que se obtuvieron al cambiar de forma independiente los parámetros  $F$  y  $A$ . Estas figuras muestran cómo se modifica la dispersión de los valores teóricos respecto a los experimentales, y confirman que ambos parámetros están fuertemente acoplados, por lo que su ajuste conjunto resulta indispensable para mejorar la predicción de las entalpías de formación. En consecuencia, su



**Figura 6.1:** Superficie de parametrización de factor de escala  $F$  y parámetro  $A$  en el conjunto SO110 en función de la DMA, reportado en  $\text{kJ} \cdot \text{mol}^{-1}$ . La figura 6.1a muestra la representación en 3D, mientras que la figura 6.1b muestra la representación en 2D; el punto rojo representa los valores óptimos de  $F$  y  $A$  con una DMA mínima.

ajuste conjunto es indispensable para optimizar la predicción de  $\Delta_f H_{298.15\text{K}}^\circ$  y reducir la DMA al mínimo, logrando una mayor concordancia entre los valores teóricos y experimentales.

En la siguiente sección se analizará cómo las modificaciones en los pasos 1 y 2 del método G4, mediante la variación de los métodos de estructura electrónica, afectan la optimización de la geometría, el cálculo de frecuencias y, en consecuencia, la entalpía de formación ( $\Delta_f H_{298.15\text{K}}^\circ$ ); dado que se está estudiando un cambio en el método G4, es necesaria la reparametrización de  $F$  y  $A$ .

## 6.2. Optimización y cálculo de frecuencias en G4 con otros métodos de estructura electrónica

Para implementar la modificación de los pasos 1 y 2 del método G4, se partió del programa LabFQOTG4, escrito en C++ y desarrollado en el grupo de investigación. Este programa fue descrito en la Sección 5.2. Se utilizó la primera parte del programa para cambiar los cálculos de estructura electrónica que componen G4. Es decir, la geometría optimizada se obtiene con dis-

tintos métodos a B3LYP, para posteriormente reparametrizar los valores de  $F$  y  $A$ . Antes que nada, es necesario definir el conjunto de calibración y las referencias utilizadas, las cuales fueron descritas en las secciones 5.2.2 y 5.3. Una vez definidos nuestros conjuntos, es necesario generar nuestra base de datos enviando a calcular en Gaussian 09, la cual nos permitirá explorar distintas combinaciones de  $F$  y  $A$  y poder desarrollar todo el tratamiento estadístico. Una vez determinado el valor óptimo de los parámetros para ese conjunto dado, se desarrollan los siguientes pasos:

Primero, se obtienen los valores de  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$ , DMA,  $\mu$  y  $\sigma$  para el conjunto de calibración SO110. Este cálculo se realiza utilizando el código `calculatemadparametrization.sh`, descrito en la Sección 5.2. Dado que los valores de  $F$  y  $A$  empleados para calibrar el método G4-STD fueron ajustados a un conjunto específico de moléculas y al funcional B3LYP, no es posible utilizarlos directamente en este contexto. Por lo tanto, es necesario modificar los valores de  $F$  y  $A$ . Con la notación G4-nitro-cerrado- $\omega$ B97XD, G4-nitro-cerrado-M06-2X y G4-nitro-cerrado-MP2, nos referimos a variantes del método G4 en las que el funcional B3LYP de los pasos 1 y 2 se reemplaza por  $\omega$ B97XD, M06-2X o MP2, respectivamente, recalibrando en cada caso los parámetros  $F$  y  $A$  con el conjunto SO110. Por otro lado, cuando usamos la etiqueta G4-nitro-cerrado-B3LYP, nos referimos a mantener los pasos 1 y 2 del método G4 con B3LYP/6-31G(2df,p), pero recalibrando igualmente los parámetros  $F$  y  $A$  para el conjunto SO110. Finalmente, denominaremos como G4-STD a la variante utilizada por Gaussian 09 [120], que será nuestra referencia, para observar las diferencias en la DMA de cada modificación propuesta a G4.

En la Tabla 6.1 se muestran los resultados de los valores óptimos de  $F$  y  $A$  del proceso de parametrización y utilizando distintos métodos de estructura electrónica en los pasos 1 y 2 de G4. Además, en dicha tabla se presentan las medidas estadísticas utilizadas en nuestro análisis de las  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  usando los distintos métodos de estructura electrónica usados para calcular geometría de partida y el cálculo de frecuencias. El método G4-STD muestra una DMA relativamente alta de  $4.65 \text{ kJ} \cdot \text{mol}^{-1}$ . Esto se atribuye a los problemas del G4-STD para representar el subconjunto SD43, que presenta los mayores errores. Las razones de esto se discutirán en la Sección 6.3. Por otro lado, al comparar los parámetros estadísticos en términos de DMA, observamos que G4-nitro-cerrado-B3LYP  $\geq$  G4-nitro-cerrado-MP2  $>$  G4-nitro-cerrado- $\omega$ B97XD

>G4-nitro-cerrado-M06 – 2X en términos de disminuir la diferencia teórica-experimental. La variante G4-nitro-cerrado-MP2 presenta la necesidad de una corrección más marcada en las frecuencias vibracionales, dado que MP2 tiende a sobreestimarlas bajo el modelo armónico. Esta tendencia se refleja en el hecho de que el factor de escala óptimo obtenido  $F$  resulta más bajo que en las demás variantes, lo cual es consistente con la literatura que reporta la necesidad de aplicar una corrección mayor a los valores predichos por MP2. Por ejemplo, un estudio mostró que el nivel MP2/6-31G(d,p) no predice correctamente las frecuencias de dos modos en el espectro vibracional del fenol, presentando desviaciones de hasta  $140\text{ cm}^{-1}$  y  $220\text{ cm}^{-1}$  en los modos 14 y 4, respectivamente [121]. Como resultado, la corrección aplicada al utilizar un factor de escala en MP2 es mayor, como también se ha observado en otros trabajos [40]. Con respecto a los otros funcionales, podemos notar que las diferencias entre B3LYP y  $\omega$ B97XD son tan pequeñas que cualquiera de los dos podría usarse como geometría de referencia para los cálculos posteriores de G4.

Podemos afirmar que en el método G4 para la entalpía de formación de nitrofenilos y heterociclos nitrogenados, la corrección importante no proviene de la modificación del método de estructura electrónica o el funcional usado en la optimización de geometría y el cálculo de frecuencias. Sin embargo, estos parámetros con diferentes métodos nos permiten calcular la  $\Delta_f H_{298.15\text{K}}^\circ(\text{T})(\text{g})$  sin tener que optimizar la geometría nuevamente con el funcional B3LYP. Al notar el rendimiento superior de G4-nitro-cerrado-B3LYP, concluimos que la corrección electrónica capturada por medio de la parametrización y el conjunto de prueba es más importante y significativa de lo que se había previsto. Por esta razón, decidimos analizar la dependencia de los parámetros  $F$  y  $A$  con respecto al conjunto de calibración utilizado, con el fin de evaluar cómo la elección de referencia experimental y si el tipo de molécula influye y se refleja en una reducción en la DMA.

**Tabla 6.1:** Evaluación de parámetros estadísticos y de calibración para el conjunto SO110 mediante diferentes variantes propuestas de G4.

Método	Parámetros de calibración y medidas estadísticas de evaluación				
	Factor de Escala $F$	$A$ /(mHartree)	DMA/(kJ mol <sup>-1</sup> )	$\mu$ /(kJ mol <sup>-1</sup> )	$\sigma$ /(kJ mol <sup>-1</sup> )
G4-STD	0.9854	6.947	4.65	-2.85	5.89
G4-nitro-cerrado-B3LYP	0.95932	6.7354	3.22	0.89	4.22
G4-nitro-cerrado- $\omega$ B97XD	0.95606	6.7867	3.25	0.21	4.24
G4-nitro-cerrado-M06-2X	0.96196	6.8188	3.26	0.36	4.30
G4-nitro-cerrado-MP2	0.93312	6.7168	3.23	0.50	4.31

### 6.3. Evaluación del método G4 en diferentes conjuntos

En primer lugar, comparamos el rendimiento de G4-STD (usando sus parámetros originales) en tres conjuntos de entrenamiento: SO67, SD43 y SO110. La Tabla 6.2 muestra la desviación media absoluta (DMA), el promedio ( $\mu$ ) y la desviación estándar ( $\sigma$ ) de las diferencias entre valores teóricos y experimentales.

**Tabla 6.2:** Evaluación de parámetros estadísticos del método G4-STD aplicados a diferentes conjuntos.

Conjunto	Parámetros estadísticos		
	DMA/(kJ · mol <sup>-1</sup> )	$\mu$ /(kJ · mol <sup>-1</sup> )	$\sigma$ /(kJ · mol <sup>-1</sup> )
SO67	2.46	0.02	3.55
SD43	8.06	-7.32	6.03
SO110	4.65	-2.85	5.89

El mejor desempeño se observa en SO67, con una DMA de 2.46 kJ mol<sup>-1</sup>, valor comparable al reportado para el conjunto G3/05 (3.47 kJ mol<sup>-1</sup>) que incluía 270 entalpías de formación. Aun con solo 67 moléculas, SO67 describe de forma uniforme diversas familias orgánicas. Para SD43, la DMA de 8.06 kJ mol<sup>-1</sup> y el promedio negativo de -7.32 kJ mol<sup>-1</sup> indican errores más considerables en la estimación de la entalpía de formación. Finalmente, en SO110 (que combina SO67 y

SD43), la DMA se reduce a  $4.65 \text{ kJ mol}^{-1}$ , lo cual sugiere que las moléculas bien descritas de SO67 compensan parcialmente las desviaciones de SD43. Con ello, observamos la necesidad de una recalibración de G4-STD para describir de forma adecuada los compuestos presentes en SD43. En este conjunto también evaluamos un indicador para establecer si el problema provenía de no contemplar la correlación no dinámica por medio de un indicador llamado T1diag. Pese a ello, no encontramos una correlación directa entre mayor correlación no dinámica y mayor diferencia entre los valores teóricos y experimentales, puede observarse la gráfica de la Figura A.3 (Ver apéndice A).

## 6.4. Calibración del conjunto SO110

Para el ajuste en el conjunto SO110, evaluamos el rendimiento del método G4 utilizando parámetros de  $F$  y  $A$  calibrados en diferentes subconjuntos. Esto se hizo para observar el efecto de aplicar parámetros de conjuntos con características estructurales distintas a un conjunto específico. Tal como se describió en la Sección 5.3, exploramos el uso de parámetros de calibración obtenidos de SO67 y SD43 aplicados a SO110. El objetivo es cuantificar la efectividad de estos parámetros en sistemas que no fueron considerados durante la calibración de G4-STD.

En la Tabla 6.3 se muestran los parámetros optimizados aplicados a SO110. G4-STD presenta una DMA de  $4.65 \text{ kJ} \cdot \text{mol}^{-1}$ , un promedio  $\mu$  negativo de  $-2.85 \text{ kJ} \cdot \text{mol}^{-1}$  y una desviación estándar  $\sigma$  de  $5.89 \text{ kJ} \cdot \text{mol}^{-1}$ . Al evaluar los parámetros obtenidos del subconjunto SO67 y aplicarlos a SO110, se observa un comportamiento similar al de G4-STD, lo cual es consistente dado que SO67 es un subconjunto de las moléculas y referencias experimentales usadas en Gaussian 09 [120] para la calibración de G4-STD. La alta DMA y las desviaciones se deben a que los sistemas de SD43 no están representados adecuadamente en este ajuste, indicando que probablemente hubo una subrepresentación de ciertos compuestos en la calibración de G4-STD. Como resultado, los parámetros  $F$  y  $A$  no capturan adecuadamente características estructurales y químicas específicas para el subconjunto SD43.

Por otro lado, al usar los parámetros de SD43 y aplicarlos en SO110, se observa una mejora clara

en los valores estadísticos, aunque algunos sistemas de SO67 presentan un peor desempeño. Los mejores resultados se obtuvieron con la calibración G4-nitro-cerrado-B3LYP, que describe adecuadamente ambos subconjuntos (SO67 y SD43) sin generar parámetros excesivamente específicos. Este ajuste alcanza una DMA de  $3.22 \text{ kJ} \cdot \text{mol}^{-1}$ ,  $\mu$  de  $0.86 \text{ kJ} \cdot \text{mol}^{-1}$  y  $\sigma$  de  $4.23 \text{ kJ} \cdot \text{mol}^{-1}$ . Al comparar el factor de escala, se observa que G4-nitro-cerrado-B3LYP requiere una corrección mayor respecto a G4-STD (de 2% a 5%), lo cual compensa la ausencia del modelo anarmónico en los cálculos de entalpía de formación. Además, el parámetro  $A$  resulta menor en los sistemas de SD43, lo que refleja un efecto electrónico diferente al de SO67. En los anexos (Tabla A.1) se reportan los valores de  $\Delta_f H_{298.15\text{K}}^\circ(T - Exp)$  para SO110 obtenidos con ambos métodos.

**Tabla 6.3:** Calibración y evaluación de parámetros de  $F$  y  $A$  para los conjuntos SO67, SD43 y SO110 con la variante G4-nitro-cerrado-B3LYP.

Conjunto	Parámetros de calibración y medidas estadísticas de evaluación				
	Factor de Escala $F$	$A$ /(mHartrees)	DMA/( $\text{kJ} \cdot \text{mol}^{-1}$ )	$\mu$ /( $\text{kJ} \cdot \text{mol}^{-1}$ )	$\sigma$ /( $\text{kJ} \cdot \text{mol}^{-1}$ )
SO67	0.97915	6.9093	2.41	-0.01	3.49
SD43	0.95492	6.7123	3.70	0.62	4.92
SO110	0.95932	6.7354	3.22	0.89	4.22

En la tabla 6.4 se observa como la selección del conjunto de calibración tiene un impacto significativo en la calidad y la transferibilidad de los parámetros empíricos  $F$  y  $A$ . En particular, al emplear el conjunto SO67 como referencia de entrenamiento, se observa que las desviaciones medias absolutas (DMA) obtenidas para los conjuntos externos SD43 y SO110 aumentan notablemente ( $7.19$  y  $4.28 \text{ kJ} \cdot \text{mol}^{-1}$ , respectivamente), lo que indica una fuerte dependencia de los parámetros a las características específicas de este conjunto reducido.

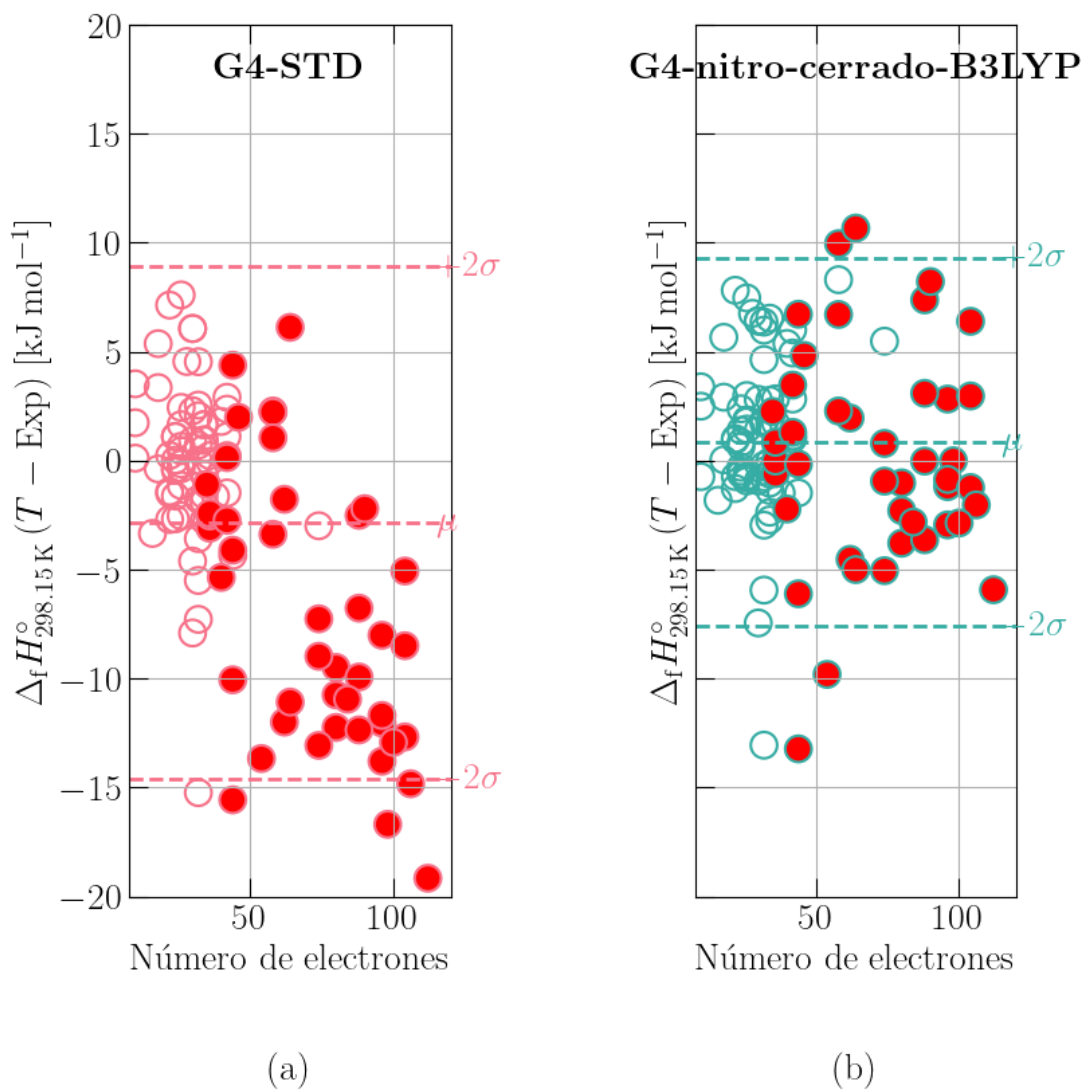
Por el contrario, la calibración de  $F$  y  $A$  utilizando los conjuntos SD43 o SO110 produce valores de DMA más consistentes, tanto en sus propios conjuntos de entrenamiento como al extrapolarse entre sí ( $3.70$  y  $3.22 \text{ kJ} \cdot \text{mol}^{-1}$ , respectivamente). Este comportamiento sugiere que los conjuntos de mayor tamaño y diversidad (SD43 y SO110) son más representativos y permiten obtener parámetros con mejor capacidad de transferencia, en contraste con SO67.

**Tabla 6.4:** Desviaciones medias absolutas (DMA) en  $\text{kJ} \cdot \text{mol}^{-1}$  al usar distintos conjuntos como *entrenamiento* (columnas) y aplicarlos a cada conjunto (*aplicado a*, filas).

Aplicado a \ Entrenamiento	SO67	SD43	SO110
SO67	2.41	2.95	2.88
SD43	7.19	3.70	3.74
SO110	4.28	3.24	3.22

En la figura 6.2 se muestran las  $\Delta_f H_{298.15\text{K}}^\circ(\text{T} - \text{Exp})(\text{g})$  evaluadas para el conjunto de moléculas SO110 utilizando los parámetros de a) G4-STD y b) G4-nitro-cerrado-B3LYP, respectivamente, en función del número de electrones. En el eje horizontal se representan las moléculas ordenadas por el número de electrones en este conjunto, mientras que en el eje vertical se muestran las  $\Delta_f H_{298.15\text{K}}^\circ(\text{T} - \text{Exp})(\text{g})$  calculadas con sus correspondientes parámetros. Las líneas de  $\pm 2\sigma$  representan el intervalo de dos veces la desviación estándar de las entalpías de formación alrededor del promedio, reflejando la dispersión global de los datos. Los resultados muestran que los parámetros utilizados en G4-nitro-cerrado-B3LYP reducen significativamente esta dispersión en comparación con los parámetros usados en G4-STD. Esto es evidente debido a la menor amplitud de las barras de error y a los valores más cercanos a 0 en las  $\Delta_f H_{298.15\text{K}}^\circ(\text{T} - \text{Exp})(\text{g})$  en ambos subconjuntos, lo que demuestra una mayor precisión en la predicción teórica de las entalpías de formación. Al evaluar los parámetros usados en el método G4-STD, observamos que el subconjunto SD43, representado por los círculos rellenos, presenta las mayores desviaciones en las entalpías de formación calculadas. No obstante, al utilizar los parámetros optimizados con G4-nitro-cerrado-B3LYP, las moléculas del subconjunto SD43 presentaron una corrección importante, disminuyendo globalmente los parámetros estadísticos evaluados. Al evaluar los parámetros por subconjuntos, se observa ligeramente un peor desempeño en SO67, que pasa de una DMA de  $2.46 \text{ kJ} \cdot \text{mol}^{-1}$  a  $2.88 \text{ kJ} \cdot \text{mol}^{-1}$ , mientras que las correcciones más importantes provienen de SD43, que pasa de  $8.06 \text{ kJ} \cdot \text{mol}^{-1}$  a  $3.74 \text{ kJ} \cdot \text{mol}^{-1}$ .

Cada conjunto de calibración puede contener moléculas con características únicas, como diferentes grupos funcionales, estructuras moleculares o tipos de enlace, que influyen en la calidad de los valores óptimos de los parámetros usados en el método G4-STD. Esta adaptación permite capturar mejor las variaciones representadas por cada conjunto de datos, mejorando la



**Figura 6.2:**  $\Delta_f H_{298.15\text{K}}^\circ(T - \text{Exp})(g)$  en función del número de electrones en el conjunto de calibración SO110. Se utilizan círculos vacíos para distinguir las moléculas que corresponden al subconjunto SO67, y círculos cerrados para las moléculas del subconjunto SD43. Las líneas discontinuas representan los límites de  $\mu \pm 2\sigma$  alrededor del promedio  $\mu$ , lo cual indica la dispersión de los datos en un intervalo de confianza del 95%. La línea discontinua central corresponde al valor promedio de todos los datos, reportado en  $\text{kJ} \cdot \text{mol}^{-1}$ .

precisión y la confiabilidad de las predicciones en la entalpía de formación. La inclusividad en el conjunto de calibración es esencial para capturar todo el espectro de diversidad química y molecular, además de la complejidad estructural de cada molécula. Esto mejora la concordancia entre los resultados teóricos y experimentales. En los anexos en la figura A.4, se muestran las dispersiones de las  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  utilizando otros métodos de estructura electrónica en los pasos 1 y 2 de G4. En general y aunque se observan algunos sistemas donde se observan diferencias entre distintos métodos, las correcciones más importantes provienen de la calibración de  $F$  y  $A$ . Asimismo, la relevancia de incluir en el conjunto de calibración suficiente diversidad química es fundamental. De no ser así, los parámetros pueden quedar sesgados hacia ciertas familias de compuestos y perder precisión en otras, como sucedía inicialmente con los nitrofenilos y los heterociclos en G4-STD. Estas buenas prácticas se discuten en el trabajo de A. Karton *et al.* [103], quienes sugieren que aumentar la diversidad en los conjuntos de calibración no solo mejora la representatividad, sino que también puede hacer que los métodos calibrados sean más robustos y aplicables a un rango más amplio de propiedades químicas y sistemas. No obstante, el aumentar el conjunto de calibración para describir toda la familia de sistemas de capa cerrada con solo 2 parámetros  $F$  y  $A$ , siempre resultara en ser insuficiente para describir todo el espacio químico. Es por esta razón que el aumentar el conjunto de calibración siempre resultara en dejar familias o conjuntos de moléculas sin una correcta descripción.

La selección del conjunto de calibración juega un papel crucial en la calidad predictiva del método compuesto G4, así como en los valores experimentales usados de referencia. Creemos que este efecto, al incluir una mayor diversidad química en el conjunto de calibración, también es aplicable a cualquier método compuesto que utilice parámetros empíricos. Por lo tanto, la precisión en el cálculo de  $\Delta_f H_{298.15K}^\circ(T)(g)$  estará fuertemente influenciada por qué tan similar sea la molécula al conjunto de calibración usado para  $F$  y  $A$  en el método G4.

## 6.5. Evaluación de la parametrización con moléculas fuera del conjunto SO110

Para explorar la aplicabilidad del método G4-nitro-cerrado-B3LYP más allá del conjunto de calibración SO110, se analizaron moléculas que no fueron incluidas en dicho conjunto, incluyendo tanto sistemas con grupos funcionales relevantes (nitro-, nitrofenilos, heterociclos nitrogenados) como compuestos sin estos grupos. Esta evaluación resulta esencial para determinar si las mejoras observadas en la calibración se traducen en una mayor capacidad predictiva general del método.

### 6.5.1. Moléculas con grupos funcionales tipo nitro, nitrofenilo o heterocíclicos

En la Tabla 6.5 se muestran las diferencias entre los valores experimentales de la entalpía estándar de formación ( $\Delta_f H_{298.15K}^\circ$ ) y los valores calculados por los métodos G4-STD y G4-nitro-cerrado-B3LYP para un conjunto representativo de compuestos que contienen grupos funcionales del tipo nitro, nitrofenilo o heterociclos nitrogenados. En términos generales, se evidencia que la reparametrización empleada en G4-nitro-cerrado-B3LYP contribuye a una notable mejora en la predicción para la mayoría de los sistemas considerados.

Por ejemplo, para compuestos como 1-(2-Nitrofenil)pirrol, N-Metilcarbazol, Fenacina, Trinitrometano, Peroxinitrato de metilo y Peroxinitrato de etilo la diferencia entre el valor calculado y el experimental se reduce de manera considerable al aplicar G4-nitro-cerrado-B3LYP, lo que indica que los parámetros ajustados permiten corregir deficiencias presentes en el método estándar. En contraste, en el caso del ácido nítrico, el error cambia de  $-3.12 \text{ kJ} \cdot \text{mol}^{-1}$  con G4-STD a  $+1.83 \text{ kJ} \cdot \text{mol}^{-1}$  con G4-nitro-cerrado-B3LYP, lo que podría interpretarse como una leve sobreestimación en la predicción; este cambio de signo sugiere además una modificación en la tendencia del método. Para los compuestos 3-Nitroanilina y Ácido peroxinitríco si bien el error también varía, se observa una inversión que indica la posible necesidad de ajustes adicionales en ese sis-

tema en particular, mientras que en el caso de la benzoxazola, la reducción del error de  $-15.86 \text{ kJ} \cdot \text{mol}^{-1}$  a  $-10.48 \text{ kJ} \cdot \text{mol}^{-1}$  reafirma la mejora en la precisión de la nueva parametrización.

En conjunto, estos resultados demuestran que G4-nitro-cerrado-B3LYP se ajusta adecuadamente a los sistemas funcionalizados, mejorando la capacidad predictiva del método para compuestos con características similares a los utilizados en la calibración. No obstante, la variabilidad observada en ciertos casos, como en el ácido nítrico y la 3-Nitroanilina, sugiere que, aunque la reparametrización mejora la precisión global, pueden existir particularidades que requieran refinamientos adicionales para lograr una concordancia óptima en todos los sistemas.

Podemos observar un efecto de transferibilidad entre el conjunto de calibración SO110 y moléculas con grupos funcionales  $-\text{NO}_2$ , heterociclos y nitrofenilos. Lo cual nos muestra que la precisión del método G4 esta fuertemente ligado al conjunto de calibración y en consecuencia a las moléculas incluidas en este proceso.

**Tabla 6.5:** Diferencias entre  $\Delta_f H_{298.15\text{K}}^\circ$  experimental y calculado para moléculas con grupos tipo nitro, nitrofenilo o heterociclos nitrogenados, utilizando los métodos G4-STD y G4-nitro-cerrado-B3LYP. Valores en  $\text{kJ} \cdot \text{mol}^{-1}$ .

Molécula	$\Delta_f H_{298.15\text{K}}^\circ(\text{Exp})(g)$	(G4-STD - Exp)	(G4-nitro-cerrado-B3LYP - Exp)
Ácido nítrico	$-134.31$ [54]	$-3.12$	$1.83$
1-(2-Nitrofenil)pirrol	$235.7 \pm 2.7$ [25]	$-13.49$	$-4.69$
N-Metilcarbazol	$199.1 \pm 0.5$ [25]	$-8.49$	$-2.84$
3-Nitroanilina	$62.5 \pm 2.8$ [122]	$-4.20$	$2.54$
Trinitrometano	$10.2 \pm 3.0$ [26]	$-19.24$	$-7.10$
Fenazina	$331.7 \pm 1.3$ [123]	$-6.38$	$1.01$
Benzoxazola	$45.14 \pm 0.52$ [124]	$-15.86$	$-10.48$
Ácido peroxinitrónico	$-52.72 \pm 1.9$ [125, 126]	$-3.13$	$3.37$
Peroxinitrato de metilo	$-43.93$ [125, 126]	$-11.39$	$-5.08$
Peroxinitrato de etilo	$-82.42$ [125, 126]	$-6.99$	$-0.87$
$\mu$	—	$-9.23$	$-2.23$
DMA	—	$9.23$	$3.98$
$\sigma$	—	$4.61$	$4.35$

Una comparación con otros métodos compuestos y las desviaciones molécula por molécula entre las entalpías de formación experimentales y las calculadas mediante los distintos métodos compuestos (CBS-QB3, G3, G3(MP2), G4(MP2), G4-STD, G4(MP2)-6X y nuestra variante G4-nitro-cerrado-B3LYP) se presenta en el Anexo (Figura A.5). En dicha figura se aprecia que, aunque todos los métodos muestran limitaciones en moléculas con grupos nitro y heterociclos, el método propuesto logra reducir sistemáticamente la dispersión (reflejada en  $\pm 2\sigma$ ) respecto a G4-STD y otras variantes, confirmando su superioridad para este tipo de sistemas.

### 6.5.2. Moléculas sin grupos nitro ni heterocíclicos

En este subconjunto se analizan compuestos que no contienen ni grupos nitro ni heterociclos nitrogenados, lo que permite evaluar la capacidad predictiva de los métodos en sistemas que no pertenecen al dominio prioritario de la reparametrización orientada a moléculas como el subconjunto SD43. La Tabla 6.6 muestra las diferencias entre los valores experimentales de la entalpía estándar de formación ( $\Delta_f H_{298.15K}^\circ$ ) y aquellos calculados mediante tres parámetros obtenidos de diferentes conjuntos de calibración evaluados en este conjunto de prueba G4-STD, G4/ReoptSO67 y G4-nitro-cerrado-B3LYP.

En términos generales, se observa que el método G4/ReoptSO67 exhibe una precisión comparable a la de G4-STD, evidenciando una estabilidad en la predicción para este tipo de compuestos. Por ejemplo, ambos métodos presentan errores moderados que se mantienen en rangos similares, lo que respalda la reparametrización basada en un subconjunto del conjunto de calibración G3/05. En contraste, G4-nitro-cerrado-B3LYP muestra un comportamiento más variable en estos sistemas. Mientras que en algunos compuestos, como el dibenzofurano, se consigue una notable reducción del error (con una diferencia casi nula respecto al valor experimental), en otros, como la acetofenona o el antraceno, el error aumenta considerablemente. Esta variabilidad se interpreta como una señal de que la optimización empleada en G4-nitro-cerrado-B3LYP está específicamente orientada a sistemas funcionalizados, de modo que su desempeño en compuestos sin dichos grupos puede resultar menos consistente.

En resumen, para moléculas que no contienen grupos nitro ni heterociclos, G4/ReoptSO67 reproduce el desempeño de G4-STD (precisión comparable). En cambio, G4-nitro-cerrado-B3LYP no ofrece una mejora sistemática fuera de su dominio de calibración y, en algunos casos, la desviación frente a los datos experimentales incluso aumenta, lo que refleja su especialización y la limitada transferibilidad a estos sistemas.

**Tabla 6.6:** Diferencias entre  $\Delta_f H_{298.15K}^\circ$  experimental y calculado para un conjunto de moléculas sin grupos nitro ni heterocíclicos nitrogenados, utilizando los métodos G4-STD, G4/ReoptSO67 y G4-nitro-cerrado-B3LYP, reportados en  $\text{kJ} \cdot \text{mol}^{-1}$ .

Molécula	$\Delta_f H_{298.15K}^\circ(\text{Exp})(g)$	(G4-STD – Exp)	(G4/ReoptSO67 – Exp)	(G4-nitro-cerrado-B3LYP – Exp)
Cloruro de hidrógeno	$-92.31 \pm 0.10$ [127]	1.81	2.10	3.58
Dimetilamina	$-19.00 \pm 2$ [54]	3.04	2.61	2.44
Acetofenona	$-86.7 \pm 1.7$ [54]	2.03	2.23	5.84
Antraceno	$223.0 \pm 0.1$ [54]	-1.42	-1.05	4.40
Dibenzofurano	$55.3 \pm 0.3$ [54]	-6.73	-6.20	-0.09
Fenantreno	$202.2 \pm 2.3$ [54]	-3.78	-3.41	2.02
Piperidina	$-47.15 \pm 0.63$ [54]	-0.01	-0.11	-0.52
Tolueno	$50.1 \pm 1.1$ [128]	0.03	-0.12	1.66
DMA	—	2.36	2.22	2.57

Lo anterior sugiere que nuestros parámetros con el método G4-nitro-cerrado-B3LYP poseen una naturaleza química particular, lo que los hace especialmente adecuados para representar estos sistemas, grupos  $-\text{NO}_2$ , nitrofenilos y heterociclos que contengan los elementos C, H, O, N, Cl, y S, los cuales no habían sido lo suficientemente representados en los conjuntos de calibración originales de G4.

## 6.6. Sistemas con estimaciones atípicas de entalpía de formación atípico

En esta sección, identificamos cinco moléculas que presentaban diferencias significativas y atípicas en el cálculo teórico de  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  con respecto al experimento, en concordancia

cia con lo reportado por Dorofeeva *et al.* [25]. Es decir, tampoco utilizando el método de reacciones isodésmicas se fue capaz de acercarnos al valor experimental. En la Tabla 6.7 mostramos los cálculos de la  $\Delta_f H_{298.15K}^\circ(T - Exp)(g)$  usando los parámetros G4-STD y utilizando nuestros mejores parámetros G4-nitro-cerrado-B3LYP. Observamos en las DMA de estos sistemas una tendencia hacia valores totalmente distintos a los experimentales.

Durante el proceso de reparametrización, notamos que estas moléculas no se veían tan influenciadas al ajustar los parámetros  $F$  y  $A$ . Esto sugiere que la corrección de  $A$  no es capaz de capturar adecuadamente ciertos efectos electrónicos presentes en estos sistemas. Debido a estas discrepancias, consideramos necesaria una revisión exhaustiva de los datos experimentales para asegurar la precisión y fiabilidad de los resultados. Por esta razón, estos sistemas fueron excluidos del conjunto de entrenamiento.

**Tabla 6.7:** Comparación de  $\Delta_f H_{298.15K}^\circ(T - Exp)(g)$  obtenidas con diferentes métodos para sistemas atípicos.

Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD - Exp	G4-nitro-cerrado-B3LYP - Exp
NN-Dinitrometanamina	$53.6 \pm 1.3$ [129]	42.3	50.7
2,3,5,6-Tetrachloronitrobenzeno	$24.3 \pm 3.1$ [130]	-29.9	-14.3
2-Metil-2H-tetrazol	$328.4 \pm 0.7$ [131]	-31.9	-27.9
2,3-Dihidroxipiridina	$-263.9 \pm 4.6$ [132]	32.4	37.7
4-Acetilpiridina	$-36.2 \pm 2.0$ [133]	12.4	17.0
DMA	—	34.10	32.65

Nota: Exp (g) representa los valores experimentales de entalpía de formación en fase gaseosa ( $\text{kJ} \cdot \text{mol}^{-1}$ ). Las columnas G4-STD - Exp y G4-nitro-cerrado-B3LYP - Exp muestran la diferencia entre los valores calculados por los métodos respectivos y los valores experimentales, indicando la precisión de cada método para los sistemas atípicos estudiados.

## 6.7. Comparación con reacciones isodésmicas

En la Tabla 6.8 se muestran los valores de  $\Delta_f H_{298.15K}^\circ(T - Exp)(g)$  usando tanto el método de reacciones isodésmicas de la Tabla 5 del trabajo de Dorofeeva *et al.* [25], como los métodos G4-STD y G4-nitro-cerrado-B3LYP. Se debe tener en cuenta que sólo seleccionamos moléculas que presentaban un único conformero. Esto nos permitió evitar la complejidad adicional de en-

contrar, estudiar y calcular la energía de todos los confórmeros. Así, evaluamos el rendimiento de G4-STD, G4-nitro-cerrado-B3LYP y el promedio de las reacciones isodésmicas reportadas. Al analizar la DMA observamos que el método G4-STD presentaba un valor elevado de  $18.63 \text{ kJ} \cdot \text{mol}^{-1}$ , indicando una alta desviación promedio respecto a los valores experimentales para estos sistemas. En contraste, al utilizar nuestros parámetros de G4-nitro-cerrado-B3LYP, la DMA se redujo a  $12.07 \text{ kJ} \cdot \text{mol}^{-1}$ . Este valor sigue siendo relativamente alto en comparación con los valores experimentales, no obstante, si comparamos con el promedio de reacciones isodésmicas, mostró una DMA de  $11.14 \text{ kJ} \cdot \text{mol}^{-1}$ . Estos resultados sugieren que nuestra parametrización es efectiva, mejorando significativamente las predicciones de G4 utilizando el método de atomización, específicamente para sistemas nitrofenilos y heterociclos. Además, contradicen la idea de que el origen del problema del cálculo de la  $\Delta_f H_{298.15\text{K}}^\circ(T - Exp)(g)$  usando el método G4 se atribuye al método de atomización, ya que, como observamos, el problema se origina en cómo se calcula la energía electrónica en el método G4 y, por lo tanto, debido a la baja representación de los nitrofenilos y heterociclos en los parámetros  $F$  y  $A$ , los resultados obtenidos se ven afectados por el conjunto de calibración.

la precisión del método G4 está fuertemente influenciada por qué tan semejante estructuralmente sea una molécula al conjunto de calibración usado para determinar los parámetros de  $F$  y  $A$ . Además, podemos asegurar que la parametrización de  $F$  y  $A$  aplicada a un conjunto de calibración SO110 funciona adecuadamente para el cálculo teórico de entalpía de formación en nitrofenilos y heterociclos. La parametrización permite que la representación y corrección de efectos de correlación electrónica de los sistemas SD43 se haga de una manera más eficiente, dado que previamente estos sistemas habían sido soslayados y, por lo tanto, la naturaleza química que se capturaba de los parámetros de G4-STD no era adecuada para representar estos sistemas tan problemáticos. Además, nos asegura que para estos sistemas tenemos un método en G4-nitro-cerrado-B3LYP lo suficientemente flexible para cubrir estos sistemas.

**Tabla 6.8:** Diferencias entre  $\Delta_f H_{298.15K}^\circ(T - Exp)(g)$  obtenidas mediante los métodos G4-STD y G4-nitro-cerrado-B3LYP.

Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD - Exp	G4/Reopt - Exp	$\mu - Exp$
Tetranitrometano	$82.4 \pm 1.7$ [129]	-16.69	0.00	2.45
N,N-Dinitrometanamina	$53.6 \pm 1.3$ [129]	42.26	50.64	47.90
2,4-Dicloronitrobenceno	$40.4 \pm 2.1$ [134]	-13.80	-2.94	1.25
2,4,6-Tricloronitrobenceno	$28.1 \pm 1.7$ [130]	-19.17	-5.92	-1.63
2,3,5,6-Tetrachloronitrobenceno	$24.3 \pm 3.1$ [130]	-29.91	-14.29	-8.50
2-Cloro-4-nitrofenol	$-139.0 \pm 2.8$ [130]	-2.51	7.38	-7.05
1-Metilimidazol	$137.8 \pm 4.0$ [135]	-15.57	-13.23	-16.15
5-Nitroindol	$136.9 \pm 2.1$ [136]	-10.95	-2.82	1.90
Benzoxazol	$45.1 \pm 0.5$ [137]	-15.82	-10.47	-8.00
DMA	—	18.63	12.07	11.14

Nota:  $Exp(g)$  representa los valores experimentales de entalpía de formación en fase gaseosa ( $\text{kJ} \cdot \text{mol}^{-1}$ ). G4-STD- $Exp$  indica la diferencia entre el valor calculado mediante el método G4 estándar y el experimental; G4/Reopt- $Exp$  muestra la diferencia entre el valor calculado mediante el método G4 con reoptimización (SO110) y el experimental;  $\mu-Exp$  se refiere a la desviación media de los valores calculados respecto a los valores experimentales usando reacciones isodésmicas.

## Conclusiones

En el presente trabajo, se han abordado los objetivos propuestos para optimizar el cálculo teórico de las entalpías de formación de sistemas desafiantes como los nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados, empleando tanto el método compuesto G4 como el método de atomización. En una etapa inicial, llevamos a cabo cálculos utilizando Gaussian 09 para un conjunto de moléculas que comprende tanto sistemas del conjunto de prueba G3/05 como estos compuestos, lo cual facilitó la construcción de nuevos conjuntos de calibración (SO67, SD43 y SO110). Posteriormente, modificamos el método G4 sustituyendo el funcional B3LYP en los pasos 1 y 2 por otros métodos de estructura electrónica como  $\omega$ B97XD, M06-2X y MP2, con el fin de evaluar su impacto en la predicción de entalpías de formación.

Los resultados muestran que estas modificaciones de funcional y método no mejoran de manera significativa la precisión. En contraste, la reparametrización de los parámetros empíricos  $F$  y  $A$  conduce a mejoras claras. Usando el conjunto extendido SO110 se obtuvieron valores óptimos de  $F = 0.95932$  y  $A = 6.7354$  mHartrees. Con estos parámetros, el método G4-nitro-cerrado-B3LYP se evaluó obteniendo una desviación media absoluta (DMA) de  $3.22 \text{ kJ} \cdot \text{mol}^{-1}$ , frente a  $4.65 \text{ kJ} \cdot \text{mol}^{-1}$  del método G4-STD en el mismo conjunto. Además, se mantuvo una precisión cercana en el conjunto SO67 ( $2.88$  vs  $2.41 \text{ kJ} \cdot \text{mol}^{-1}$ ) y se logró una drástica reducción de error en SD43 ( $3.74$  vs  $8.06 \text{ kJ} \cdot \text{mol}^{-1}$ ). Al aplicar los parámetros recalibrados, se confirmó que

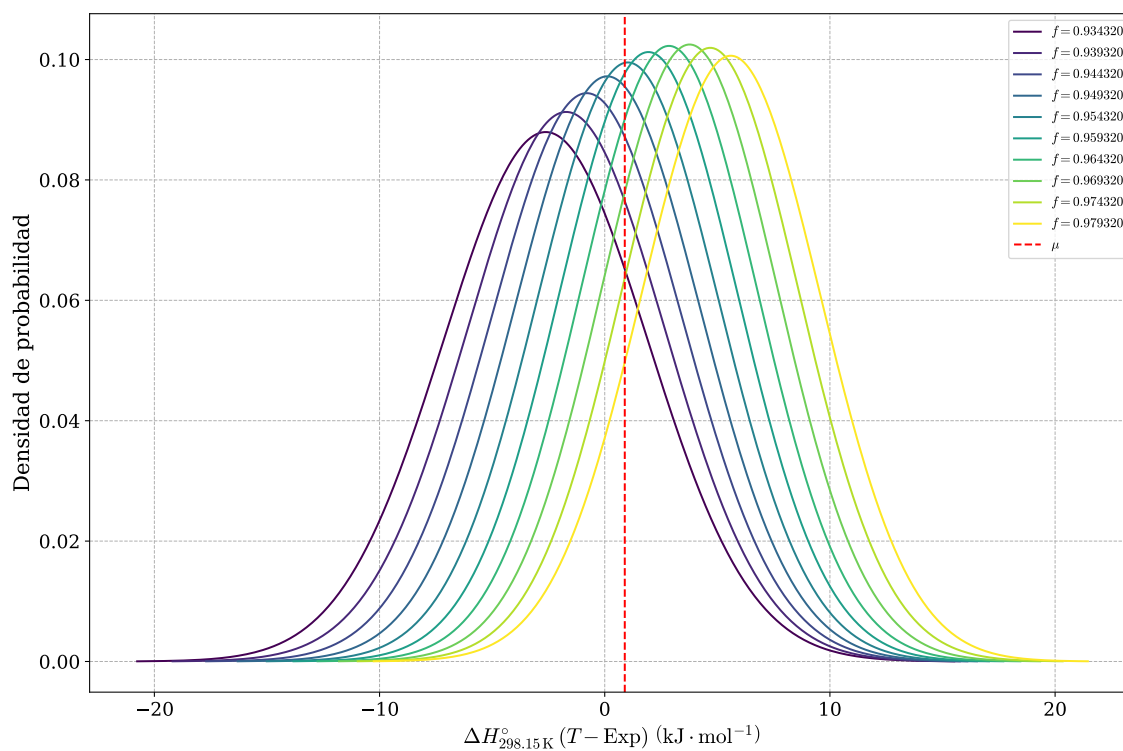
cambiar el método de optimización geométrica y cálculo de frecuencias (B3LYP,  $\omega$ B97XD, M06-2X o MP2) no afecta de manera significativa la precisión final, ya que las DMA fluctúan entre 3.22 y 3.26 kJ · mol<sup>-1</sup>.

Estos resultados evidencian que el problema no radica en el método de atomización en sí, sino en cómo se parametrizan los términos empíricos  $F$  y  $A$  a partir de conjuntos de calibración poco representativos. La mejora obtenida confirma que la deficiencia de G4 en el subconjunto SD43 se debe a la baja representación de estos sistemas en la calibración original de G4-STD.

Podemos indicar que la modificación y el uso de nuestro método G4-nitro-cerrado-B3LYP posee la capacidad de mejorar las predicciones de entalpía de formación para compuestos como nitrofenilos y heterociclos nitrogenados, sulfurados y oxigenados, así como para una amplia gama de moléculas de capa cerrada. Esta variante es aplicable únicamente a moléculas orgánicas de capa cerrada. Una solución metodológica que podría parecer evidente consistiría en ampliar de forma considerable el conjunto de calibración; sin embargo, esta estrategia sería un error, ya que podría conducir a soslayar otros subconjuntos de moléculas, como ocurrió en la calibración original de G4-STD para el conjunto SD43.

Como perspectiva de este trabajo, proponemos la implementación de modelos de machine learning adaptativos al tipo de molécula analizada. Este enfoque podría mejorar sustancialmente la descripción de los parámetros  $F$  y  $A$ , en particular del parámetro  $A$ , que no refleja únicamente la correlación electrónica dinámica, sino que también incorpora efectos derivados de las limitaciones de los conjuntos de bases reducidos, la omisión de términos de correlación de alto nivel más allá de CCSD(T) y, en ciertos casos, la correlación no dinámica. No obstante, este trabajo resulta en un estudio más allá de los alcances de esta tesis.

# Apéndice



**Figura A.1:** Curvas de distribución normal para el conjunto [SDG-100], barrido en parámetro de factor de escala  $F$ .

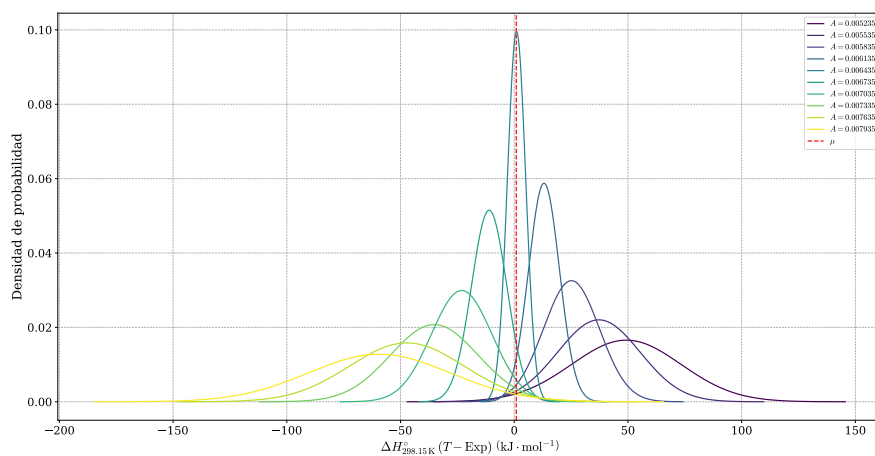


Figura A.2: Curvas de distribución normal para el conjunto [SDG-100], barrido en parámetro A.

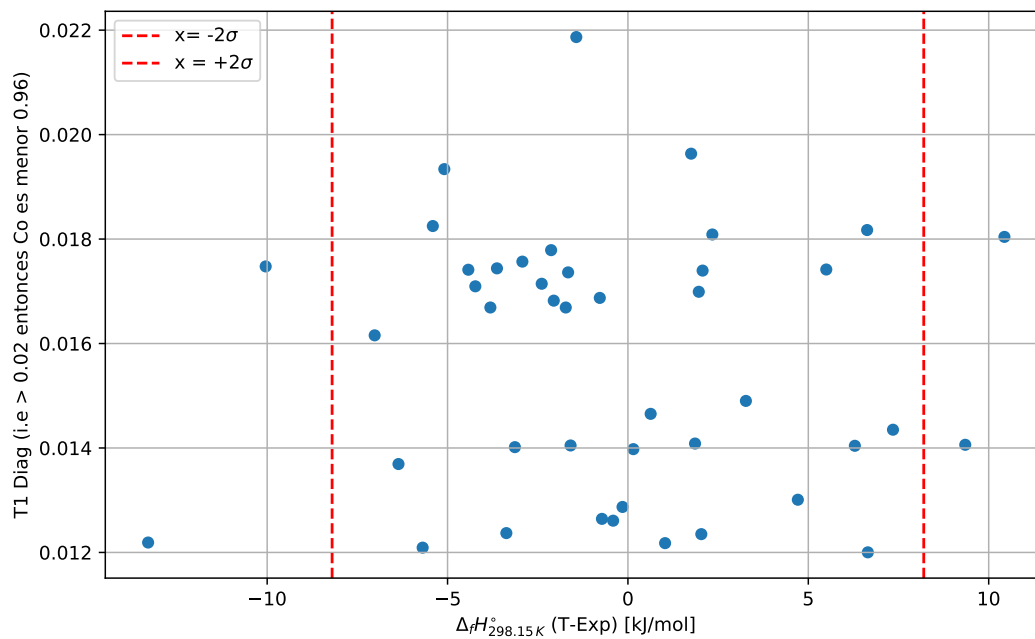
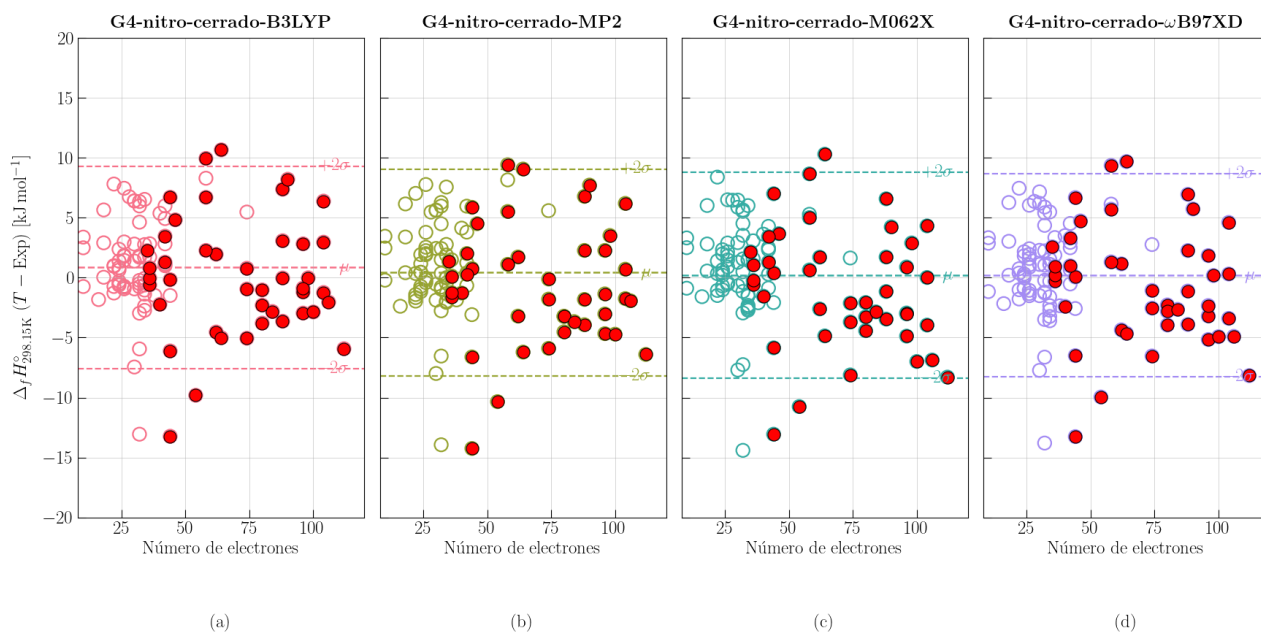
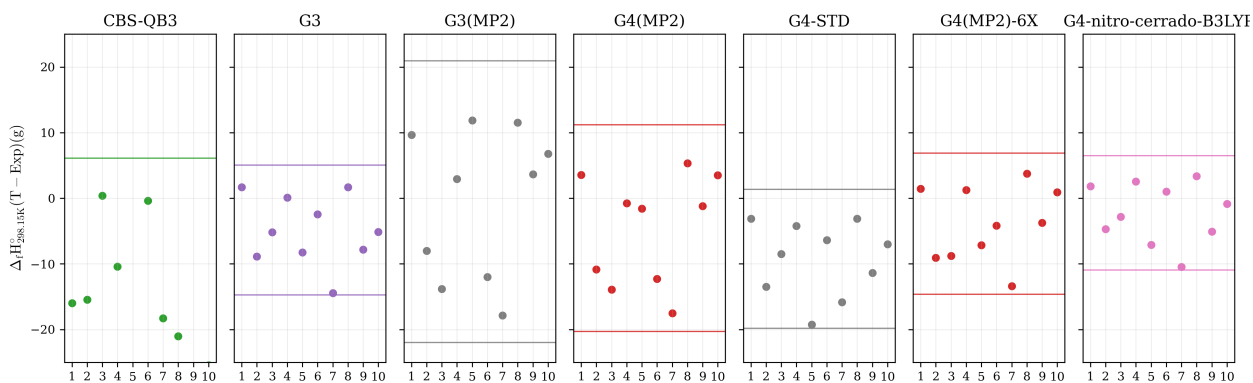


Figura A.3: Estimaciones de T1Diagnostic para evaluar la correlación no dinamica de el subconjunto SD43



**Figura A.4:**  $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$  utilizando los métodos G4-nitro-cerrado-B3LYP, G4-nitro-cerrado-MP2, G4-nitro-cerrado-M062X, G4-nitro-cerrado- $\omega$ B97XD en función del número de electrones en el conjunto de calibración SO110. Se utilizan círculos vacíos para distinguir las moléculas que corresponden al subconjunto SO67, y círculos cerrados para las moléculas del subconjunto SD43. Las líneas discontinuas representan los límites de  $\mu \pm 2\sigma$  alrededor del promedio  $\mu$ , lo cual indica la dispersión de los datos en un intervalo de confianza del 95%. La línea discontinua central corresponde al valor promedio de todos los datos, reportado en  $\text{kJ} \cdot \text{mol}^{-1}$ .



**Figura A.5:** Diferencias entre las entalpías de formación experimentales y las calculadas ( $\Delta_f H_{298.15K}^\circ(T - \text{Exp})(g)$ ) para un conjunto de moléculas que contienen grupos nitro, nitrofenilos o heterociclos nitrogenados (Tabla 6.5), utilizando diferentes métodos compuestos: CBS-QB3, G3, G3(MP2), G4(MP2), G4-STD, G4(MP2)-6X y G4-nitro-cerrado-B3LYP. En cada figura se muestran los valores para las 10 moléculas de prueba, junto con las líneas horizontales correspondientes a  $\mu \pm 2\sigma$ . Es importante mencionar que todos estos métodos compuestos contienen parámetros empíricos ajustados a ciertos conjuntos de calibración.

**Tabla A.1:** Propiedades termodinámicas de moléculas, las referencias experimentales para las moléculas 1–67, y para las moléculas 68–110.

Número	Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
1	CH <sub>4</sub> (Metano)	$-74.89 \pm 0.42$	0.11	-0.733	10
2	NH <sub>3</sub> (Amoniaco)	$-46.02 \pm 0.42$	3.52	3.34	10
3	H <sub>2</sub> O (Agua)	$-241.84 \pm 0.00$	1.74	2.5	10
4	H <sub>2</sub> CO (Formaldehído)	$-108.78 \pm 0.42$	-3.34	-1.82	16
5	H <sub>3</sub> COH (Metanol)	$-200.83 \pm 0.42$	-0.38	0.04	18
6	H <sub>2</sub> NNH <sub>2</sub> (Hidrazina)	$95.40 \pm 0.84$	5.37	5.66	18
7	CH <sub>3</sub> Cl (Cloruro de metilo)	$-82.01 \pm 0.84$	0.45	1.79	26
8	CCl <sub>4</sub> (Tetracloruro de carbono)	$-95.81 \pm 2.09$	-2.99	5.48	74
9	CH <sub>3</sub> CCH (Propino)	$184.93 \pm 0.84$	0.24	1.01	22
10	CH <sub>2</sub> CCH <sub>2</sub> (Aleno)	$190.37 \pm 1.26$	-1.5	-0.73	22
11	C <sub>3</sub> H <sub>4</sub> (Ciclopropeno)	$276.98 \pm 2.51$	7.14	7.82	22
12	CH <sub>3</sub> CHCH <sub>2</sub> (Propeno)	$20.08 \pm 0.84$	0.02	-0.30	24
13	C <sub>3</sub> H <sub>6</sub> (Ciclopropano)	$53.14 \pm 0.42$	1.1	0.64	24
14	C <sub>3</sub> H <sub>8</sub> (Propano)	$-104.60 \pm 0.42$	0.66	-0.71	26
15	CH <sub>2</sub> CHCHCH <sub>2</sub> (Butadieno)	$110.04 \pm 0.84$	0.93	1.39	30
16	C <sub>4</sub> H <sub>6</sub> (2-Butino)	$145.60 \pm 1.26$	2.29	2.81	30

*Continúa en la siguiente página*

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ (Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
17	C <sub>4</sub> H <sub>6</sub> (Metileno ciclopropano)	200.41 ± 1.67	-7.91	-7.45	30
18	C <sub>4</sub> H <sub>6</sub> (Biciclobutano)	217.15 ± 0.84	6.06	6.42	30
19	C <sub>4</sub> H <sub>6</sub> (Ciclobuteno)	156.48 ± 1.67	6.07	6.41	30
20	C <sub>4</sub> H <sub>8</sub> (Ciclobutano)	28.45 ± 0.42	-0.01	-0.75	32
21	C <sub>4</sub> H <sub>8</sub> (Isobuteno)	-16.74 ± 0.84	0.01	-0.48	32
22	C <sub>4</sub> H <sub>10</sub> (Trans-butano)	-125.52 ± 0.84	0.56	-1.03	34
23	C <sub>4</sub> H <sub>10</sub> (Isobutano)	-134.31 ± 0.84	0.93	-0.61	34
24	C <sub>5</sub> H <sub>8</sub> (Espirano)	185.35 ± 0.84	-1.63	-1.48	38
25	C <sub>6</sub> H <sub>6</sub> (Benceno)	82.42 ± 0.84	1.11	2.82	42
26	CH <sub>2</sub> F <sub>2</sub> (Difluorometano)	-450.62 ± 1.67	-0.37	2.99	26
27	CHF <sub>3</sub> (Trifluorometano)	-697.05 ± 3.35	0.96	6.54	34
28	CH <sub>2</sub> Cl <sub>2</sub> (Diclorometano)	-95.40 ± 1.26	2.36	6.00	42
29	CHCl <sub>3</sub> (Triclorometano)	-103.34 ± 1.26	2.28	8.30	58
30	CH <sub>3</sub> NH <sub>2</sub> (Metilamina)	-23.01 ± 0.42	3.35	2.91	18
31	CH <sub>3</sub> CN (Cianuro de metilo)	75.31 ± 0.84	-2.7	-1.27	22
32	CH <sub>3</sub> NO <sub>2</sub> (Nitrometano)	-74.48 ± 0.84	-3.57	-0.22	32
33	CH <sub>3</sub> ONO (Nitrito de metilo)	-66.53 ± 0.84	-1.77	1.71	32
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
34	CH <sub>3</sub> SiH <sub>3</sub> (Silano metílico)	-29.29 ± 4.18	7.61	7.46	26
35	HCOOH (Ácido fórmico)	-378.65 ± 0.42	-0.36	2.38	24
36	HCOOCH <sub>3</sub> (Formiato de metilo)	-355.64 ± 0.84	-5.5	-2.96	32
37	CH <sub>3</sub> CONH <sub>2</sub> (Acetamida)	-238.49 ± 0.84	4.54	6.36	32
38	C <sub>2</sub> H <sub>4</sub> NH (Aziridina)	126.36 ± 0.84	1.11	1.39	24
39	NCCN (Cianógeno)	306.69 ± 0.84	-2.53	1.45	26
40	(CH <sub>3</sub> ) <sub>2</sub> NH (Dimetilamina)	-18.41 ± 0.84	2.44	1.80	26
41	CH <sub>3</sub> CH <sub>2</sub> NH <sub>2</sub> (Etilamina)	-47.28 ± 0.84	-0.15	-0.80	26
42	CH <sub>2</sub> CO (Ceteno)	-47.70 ± 1.67	-1.40	0.95	22
43	C <sub>2</sub> H <sub>4</sub> O (Oxirano)	-52.72 ± 0.42	-1.63	-0.48	24
44	CH <sub>3</sub> CHO (Acetaldehído)	-166.10 ± 0.42	-0.45	0.84	24
45	HCOCOH (Glicoxal)	-212.13 ± 0.84	-4.61	-0.93	30
46	CH <sub>3</sub> CH <sub>2</sub> OH (Etanol)	-235.14 ± 0.42	1.65	1.87	26
47	CH <sub>3</sub> OCH <sub>3</sub> (Éter dimetílico)	-184.10 ± 0.42	-1.14	-0.93	26
48	C <sub>2</sub> H <sub>4</sub> S (Tiooxirano)	82.01 ± 1.26	-7.28	-5.94	32
49	(CH <sub>3</sub> ) <sub>2</sub> SO (Sulfóxido de dimetilo)	-151.46 ± 0.84	2.91	4.94	42
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ (Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
50	C <sub>2</sub> H <sub>5</sub> SH (Etanotiol)	-46.44 ± 0.42	-0.64	-0.05	34
51	CH <sub>3</sub> SCH <sub>3</sub> (Sulfuro de dimetilo)	-37.24 ± 0.84	-1.89	-1.36	34
52	CH <sub>2</sub> CHF (Fluoruro de vinilo)	-138.91 ± 1.67	-2.67	-0.61	24
53	C <sub>2</sub> H <sub>5</sub> Cl (Cloruro de etilo)	-112.13 ± 1.26	1.65	2.77	34
54	CH <sub>2</sub> CHCl (Cloruro de vinilo)	37.24 ± 1.26	-15.24	2.20	32
55	CH <sub>2</sub> CHCN (Acrilonitrilo)	180.75 ± 1.67	4.55	6.76	28
56	CH <sub>3</sub> COCH <sub>3</sub> (Acetona)	-217.15 ± 0.84	0.71	1.83	32
57	CH <sub>3</sub> COOH (Ácido acético)	-432.63 ± 1.67	2.06	4.64	32
58	CH <sub>3</sub> COF (Fluoruro de acetilo)	-442.25 ± 3.35	2.55	5.99	32
59	CH <sub>3</sub> COCl (Cloruro de acetilo)	-242.67 ± 0.84	1.8	5.38	40
60	CH <sub>3</sub> CH <sub>2</sub> CH <sub>2</sub> Cl (Cloruro de propilo)	-131.80 ± 1.26	0.23	1.13	42
61	(CH <sub>3</sub> ) <sub>2</sub> CHOH (Isopropanol)	-272.80 ± 0.42	1.65	1.71	34
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
62	C <sub>2</sub> H <sub>5</sub> OCH <sub>3</sub> (Éter metil-etílico)	-216.31 ± 0.84	-2.73	-2.72	34
63	(CH <sub>3</sub> ) <sub>3</sub> N (Trimetilamina)	-23.85 ± 0.84	-1.574	-2.367	34
64	C <sub>4</sub> H <sub>4</sub> O (Furano)	-34.73 ± 0.84	0.231	2.839	36
65	C <sub>4</sub> H <sub>4</sub> S (Tiofeno)	115.06 ± 0.84	-4.35	-1.48	44
66	C <sub>4</sub> H <sub>5</sub> N (Pirrol)	108.37 ± 0.42	-0.04	1.73	36
67	C <sub>5</sub> H <sub>5</sub> N (Piridina)	140.58 ± 0.84	-1.48	1	42
68	C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> (Nitroetano)	-102.5 ± 0.80	-5.36	-2.22	40
69	C <sub>2</sub> H <sub>4</sub> N <sub>2</sub> O <sub>4</sub> (Dinitroetano)	-96.7 ± 1.30	-12	-4.53	62
70	C(NO <sub>2</sub> ) <sub>4</sub> (Tetranitrometano)	82.4 ± 1.70	-16.69	0.00	98
71	C <sub>2</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub> (N-Nitrodimetilamina)	-7.5 ± 2.10	-1.77	1.96	62
72	C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub> (Nitrobenceno)	67.5 ± 2.10	-11.08	-5.00	64
73	C <sub>6</sub> H <sub>4</sub> ClNO <sub>2</sub> (2-Cloronitrobenceno)	62.2 ± 1.80	-9.49	-1.02	80
74	C <sub>6</sub> H <sub>4</sub> ClNO <sub>2</sub> (3-Cloronitrobenceno)	40.1 ± 1.90	-10.75	-2.29	80
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ (Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
75	C <sub>6</sub> H <sub>4</sub> ClNO <sub>2</sub> (4-Cloronitrobenceno)	39.7 ± 2.60	-12.24	-3.79	80
76	C <sub>6</sub> H <sub>3</sub> Cl <sub>2</sub> NO <sub>2</sub> (2,4-Dicloronitrobenceno)	40.4 ± 2.10	-13.80	-2.95	96
77	C <sub>6</sub> H <sub>3</sub> Cl <sub>2</sub> NO <sub>2</sub> (2,5-Dicloronitrobenceno)	40.2 ± 3.20	-12.04	-1.19	96
78	C <sub>6</sub> H <sub>3</sub> Cl <sub>2</sub> NO <sub>2</sub> (3,4-Dicloronitrobenceno)	19.7 ± 2.90	-11.69	-0.86	96
79	C <sub>6</sub> H <sub>3</sub> Cl <sub>2</sub> NO <sub>2</sub> (3,5-Dicloronitrobenceno)	12.6 ± 1.90	-8.01	2.82	96
80	C <sub>6</sub> H <sub>2</sub> Cl <sub>3</sub> NO <sub>2</sub> (2,4,6- Tricloronitrobenceno)	28.1 ± 1.70	-19.16	-5.92	112
81	C <sub>6</sub> H <sub>4</sub> ClNO <sub>3</sub> (2-Cloro-4-nitrofenol)	-139.0 ± 2.80	-2.51	7.38	88
82	C <sub>6</sub> H <sub>4</sub> ClNO <sub>3</sub> (4-Cloro-2-nitrofenol)	-149.2 ± 1.80	-6.77	3.09	88
83	C <sub>6</sub> H <sub>4</sub> ClNO <sub>3</sub> (4-Cloro-3-nitrofenol)	-108.8 ± 3.70	-9.93	-0.02	88

*Continúa en la siguiente página*

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ (Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
84	$C_6H_4Cl_2N_2O_2$ (4,5-Dicloronitroanilina)	$9.7 \pm 1.80$	-5.06	6.39	104
85	$C_6H_4Cl_2N_2O_2$ (2,5-Dicloronitroanilina)	$28.7 \pm 2.10$	-8.49	2.97	104
86	$C_6H_4Cl_2N_2O_2$ (2,6-Dicloronitroanilina)	$1.9 \pm 1.70$	-12.68	-1.23	104
87	$C_7H_7NO_4$ (2-Metoxi-5-nitrofenol)	$-267.9 \pm 2.70$	-12.36	-3.63	88
88	$C_3H_4N_2$ (Pirazol)	$179.4 \pm 0.80$	-3.1	-0.6	36
89	$C_4H_6N_2$ (1-Metilpirazol)	$156.5 \pm 2.10$	4.38	6.72	44
90	$C_3H_4N_2$ (Imidazol)	$132.9 \pm 0.60$	-2.54	-0.02	36
91	$C_4H_6N_2$ (1-Metilimidazol)	$137.8 \pm 4.00$	-15.57	-13.23	44
92	$C_2H_3N_3$ (1H-1,2,4-Triazol)	$192.7 \pm 0.80$	-2.42	0.84	36
93	$C_2H_4N_4$ (1-Metil-1H-tetrazol)	$322.9 \pm 2.00$	-10.06	-6.11	44
94	$C_2H_4N_4$ (5-Metil-1H-tetrazol)	$280.7 \pm 2.60$	-4.11	-0.16	44
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ (Exp) (g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
95	C <sub>3</sub> H <sub>6</sub> N <sub>2</sub> O (Imidazolidin-2-ona)	-176.7 ± 2.20	1.99	4.83	46
96	C <sub>3</sub> H <sub>2</sub> N <sub>2</sub> O <sub>3</sub> (Ácido parabánico)	-470.8 ± 1.20	2.23	9.96	58
97	C <sub>7</sub> H <sub>7</sub> NO (2-Acetilpiridina)	-41.3 ± 2.90	6.12	10.67	64
98	C <sub>4</sub> H <sub>4</sub> N <sub>2</sub> (Piridazina)	278.3 ± 1.30	0.14	3.47	42
99	C <sub>4</sub> H <sub>3</sub> ClN <sub>2</sub> (2-Cloropirimidina)	155.1 ± 1.60	-3.38	2.28	58
100	C <sub>4</sub> H <sub>2</sub> Cl <sub>2</sub> N <sub>2</sub> (2,4-Dicloropirimidina)	125.6 ± 2.20	-8.98	-0.94	74
101	C <sub>4</sub> H <sub>2</sub> Cl <sub>2</sub> N <sub>2</sub> (4,6-Dicloropirimidina)	126.0 ± 2.20	-13.07	-5.05	74
102	C <sub>4</sub> HCl <sub>3</sub> N <sub>2</sub> (2,4,6-Tricloropirimidina)	85.8 ± 1.30	-2.20	8.22	90
103	C <sub>4</sub> HCl <sub>4</sub> N <sub>2</sub> (2,4,5,6- Tetracloropirimidina)	86.7 ± 2.10	-14.83	-2.03	106
104	C <sub>4</sub> H <sub>3</sub> ClN <sub>2</sub> (2-Cloropirazina)	168.1 ± 1.90	1.05	6.71	58
<i>Continúa en la siguiente página</i>					

<i>Continúa desde la página anterior</i>					
Número	Molécula	$\Delta_f H_{298.15K}^\circ(Exp)(g)$	G4-STD-Exp	G4-nitro-cerrado-B3LYP-Exp	Número de electrones
105	C <sub>4</sub> H <sub>2</sub> Cl <sub>2</sub> N <sub>2</sub> (2,6-Dicloropirazina)	144.4 ± 2.70	-7.26	0.79	74
106	C <sub>3</sub> H <sub>3</sub> N <sub>3</sub> (1,3,5-Triazina)	225.9 ± 0.80	-2.74	1.3	42
107	C <sub>8</sub> H <sub>6</sub> N <sub>2</sub> O <sub>2</sub> (5-Nitroindol)	136.9 ± 2.10	-10.95	-2.82	84
108	C <sub>8</sub> H <sub>4</sub> Cl <sub>2</sub> N <sub>2</sub> (2,3-Dicloroquinoxalina)	202.9 ± 2.00	-12.93	-2.86	100
109	C <sub>3</sub> H <sub>3</sub> NOS (1,3-Oxazolidin-2-tiona)	-74.4 ± 4.60	-13.67	-9.81	54
110	C <sub>3</sub> H <sub>3</sub> NO (Oxazol)	-15.5 ± 0.50	-1.09	2.25	35

# Bibliografía

- [1] Truhlar D. G., Garrett B. C., y Klippenstein S. J., *Journal of Physical Chemistry* **100**, 12771 (1996).
- [2] Marenich A. V., Cramer C. J., y Truhlar D. G., *Journal of Physical Chemistry B* **113**, 6378 (2009).
- [3] Sol M., Proft F., y Bickelhaupt M., *ChemistryOpen* **8** **7**, 814 (2019).
- [4] Fernández I. y Cossío E., *Chemical Society reviews* **43** **14**, 4906 (2014).
- [5] Carsky P. y Urban M., *Ab initio calculations: methods and applications in chemistry*, volume 16, Springer Science & Business Media, 2012.
- [6] Dirac P. A. M., *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **123**, 714 (1929).
- [7] Szabo A. y Ostlund N., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry, Dover Publications, 1996.
- [8] David Sherrill C. y Schaefer H. F., The Configuration Interaction Method: Advances in Highly Correlated Approaches, in Löwdin P.-O., Sabin J. R., Zerner M. C., y Brändas E., editors, *Advances in Quantum Chemistry*, volume 34 of *Advances in Quantum Chemistry*, pages 143–269, Academic Press, 1999.
- [9] Bartlett R. J. y Musiał M., *Rev. Mod. Phys.* **79**, 291 (2007).

- [10] Liakos D., Sparta M., Kesharwani M. K., Martin J. M. L., y Neese F., *Journal of chemical theory and computation* **11** 4, 1525 (2015).
- [11] Trygve Helgaker W. K. y Tew D. P., *Molecular Physics* **106**, 2107 (2008).
- [12] Car R., *Nature Chemistry* **8**, 820 (2016).
- [13] Kruse H., Goerigk L., y Grimme S., *The Journal of Organic Chemistry* **77**, 10824 (2012).
- [14] Pople J. A., *Rev. Mod. Phys.* **71**, 1267 (1999).
- [15] De Paula J., *Atkins química física*, Médica Panamericana, 2000.
- [16] Cox J. y Pilcher G., *Thermochemistry of Organic and Organometallic Compounds*, Academic Press, New York, 1970.
- [17] Pople J. A., Head-Gordon M., Fox D. J., Raghavachari K., y Curtiss L. A., *The Journal of Chemical Physics* **90**, 5622 (1989).
- [18] Curtiss L. A., Raghavachari K., Trucks G. W., y Pople J. A., *The Journal of Chemical Physics* **94**, 7221 (1991).
- [19] Curtiss L. A., Raghavachari K., Redfern P. C., Rassolov V., y Pople J. A., *The Journal of Chemical Physics* **109**, 7764 (1998).
- [20] Curtiss L. A., Redfern P. C., y Raghavachari K., *The Journal of Chemical Physics* **126**, 084108 (2007).
- [21] Curtiss L. A., Raghavachari K., Redfern P. C., y Pople J. A., *The Journal of Chemical Physics* **106**, 1063 (1997).
- [22] Curtiss L. A., Redfern P. C., y Raghavachari K., *The Journal of Chemical Physics* **123**, 124107 (2005).
- [23] Curtiss L. A., Redfern P. C., y Raghavachari K., *WIREs Computational Molecular Science* **1**, 810 (2011).

- [24] Verevkin S. P., Emel'yanenko V. N., Diky V., y Dorofeeva O. V., *The Journal of Chemical Thermodynamics* **73**, 163 (2014), Special issue honoring Professor Manuel Ribeiro da Silva.
- [25] Dorofeeva O. V., Kolesnikova I. N., Marochkin I. I., y Ryzhova O. N., *Structural Chemistry* **22**, 1303 (2011).
- [26] Suntsova M. A. y Dorofeeva O. V., *Journal of Chemical & Engineering Data* **59**, 2813 (2014).
- [27] Pople J. A., *Journal of the American Chemical Society* **93**, 2896 (1971).
- [28] Martin J. M. L. y De Oliveira G., *The Journal of Chemical Physics* **111**, 1843 (1999).
- [29] Boese A. D., Oren M., Atasoylu O., Martin J. M. L., Kállay M., y Gauss J., *The Journal of Chemical Physics* **120**, 4129 (2004).
- [30] Karton A., Rabinovich E., Martin J. M. L., y Ruscic B., *The Journal of Chemical Physics* **125**, 144108 (2006).
- [31] Feller D., Peterson K. A., y Dixon D. A., *The Journal of Chemical Physics* **129**, 204105 (2008).
- [32] Montgomery J. A. J., Frisch M. J., Ochterski J. W., y Petersson G. A., *Journal of Chemical Physics* **110**, 2822 (1999).
- [33] Petersson G. A., Bennett A., Tensfeldt T. G., Al-Laham M. A., Shirley W. A., y Mantzaris J., *The Journal of Chemical Physics* **89**, 2193 (1988).
- [34] Zhao Y., Lynch B. J., y Truhlar D. G., *Phys. Chem. Chem. Phys.* **7**, 43 (2005).
- [35] DeYonker N. J., Cundari T. R., y Wilson A. K., *The Journal of Chemical Physics* **124**, 114104 (2006).
- [36] Tajti A., Szalay P. G., Császár A. G., Kállay M., Gauss J., Valeev E. F., Flowers B. A., Vázquez J., y Stanton J. F., *The Journal of Chemical Physics* **121**, 11599 (2004).
- [37] Curtiss L. A., Redfern P. C., Raghavachari K., Rassolov V., y Pople J. A., *The Journal of Chemical Physics* **110**, 4703 (1999).

- [38] Wilson B. R., DeYonker N. J., y Wilson A. K., *Journal of Computational Chemistry* **33**, 2032 (2012).
- [39] Curtiss L. A., Redfern P. C., Raghavachari K., y Pople J. A., *The Journal of Chemical Physics* **114**, 108 (2001).
- [40] Scott A. P. y Radom L., *The Journal of Physical Chemistry* **100**, 16502 (1996).
- [41] Curtiss L. A., Redfern P. C., y Raghavachari K., *Chemical Physics Letters* **499**, 168 (2010).
- [42] Curtiss L. A., Redfern P. C., y Raghavachari K., *The Journal of Chemical Physics* **127**, 124105 (2007).
- [43] Dandu N. K., Assary R. S., Redfern P. C., Ward L., Foster I., y Curtiss L. A., *The Journal of Physical Chemistry A* **126**, 4528 (2022).
- [44] Silva C. d. S., Pereira D. H., y Custodio R., *The Journal of Chemical Physics* **144**, 204118 (2016).
- [45] Mayhall N. J., Raghavachari K., Redfern P. C., y Curtiss L. A., *The Journal of Physical Chemistry A* **113**, 5170 (2009), PMID: 19341257.
- [46] Grimme S., *The Journal of Physical Chemistry A* **109**, 3067 (2005).
- [47] John M. Simmie K. P. S., *The Journal of Physical Chemistry A* **119**, 7235 (2015).
- [48] Simmie J. M. y Somers K. P., *The Journal of Physical Chemistry A* **119**, 7235 (2015).
- [49] Ruscic B., Pinzon R. E., von Laszewski G., Kodeboyina D., Burcat A., Leahy D., Montoy D., y Wagner A. F., *Journal of Physics: Conference Series* **16**, 561 (2005).
- [50] Simmie J. M. y Sheahan J. N., *The Journal of Physical Chemistry A* (2016).
- [51] Simmie J. M. y Somers K. P., *The Journal of Physical Chemistry A* **119**, 7235 (2016).
- [52] Rayne S. y Forest K., *Journal of Chemical & Engineering Data* **55**, 5359 (2010).
- [53] Lebedeva N., GY A., y KATIN Y., Standard enthalpy of formation of nitrobenzene, 1971.

- [54] National Institute of Standards and Technology, NIST Chemistry WebBook, 2024, Accessed: 2024-06-07.
- [55] Dorofeeva O. V. y Moiseeva N. F., *Russian Journal of Physical Chemistry A* **82**, 136 (2008).
- [56] Pedley J., *Thermochemical Data and Structures of Organic Compounds*, Number v. 1 in TRC data series, Taylor & Francis, 1994.
- [57] Verevkin S. P., Emel'yanenko V. N., Notario R., Roux M. V., Chickos J. S., y Liebman J. E., *The Journal of Physical Chemistry Letters* **3**, 3454 (2012).
- [58] He X., Zhang J., y Gao H., *International Journal of Quantum Chemistry* **112**, 1688 (2012).
- [59] Lewars E., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Springer Netherlands, 2010.
- [60] Bond D., *Journal of Organic Chemistry* **72**, 5555 (2007).
- [61] Li Y. y Curran H. J., *The Journal of Physical Chemistry A* **127**, 1234 (2023).
- [62] Dorsett H., White A., Science D., (Australia) T. O., Aeronautical, y Division M. R. L. A. W. S., *Overview of molecular modelling and ab initio molecular orbital methods suitable for use with energetic materials*, DSTO Aeronautical and Maritime Research Laboratory, Salisbury, S. Aust., 2000.
- [63] Tikhonov D. S., Sharapa D. I., Otyotov A. A., Solyankin P. M., Rykov A. N., Shkurinov A. P., Grikina O. E., y Khaikin L. S., *The Journal of Physical Chemistry A* **122**, 1691 (2018).
- [64] Das S. K., Chakraborty S., y Ramakrishnan R., *The Journal of Chemical Physics* **154**, 044113 (2021).
- [65] Savin A. y Johnson E. R., *Judging Density-Functional Approximations: Some Pitfalls of Statistics*, pages 81–95, Springer International Publishing, Cham, 2015.
- [66] Cuevas G. y Cortés F., *Introducción a la química computacional*, Ciencia en el nuevo milenio : monografías de las redes latinoamericanas de ciencias, Fondo de Cultura Económica, 2003.

- [67] Lombardi O., González J. C. M., y Fortin S., editors, *Philosophical Perspectives in Quantum Chemistry*, volume 447 of *Synthese Library*, Springer Cham, 1 edition, 2022, Published: 17 May 2022 (Hardcover), 18 May 2023 (Softcover), 16 May 2022 (eBook).
- [68] Halkier A., Helgaker T., Jørgensen P., Klopper W., Koch H., Olsen J., y Wilson A. K., *Chemical Physics Letters* **302**, 437 (1999).
- [69] Hinchliffe A., *Molecular Modelling for Beginners*, Wiley, 2011.
- [70] Löwdin P.-O., *Correlation Problem in Many-Electron Quantum Mechanics I. Review of Different Approaches and Discussion of Some Current Ideas*, pages 207–322, John Wiley and Sons, Ltd, 1958.
- [71] Carrasco J. P. C., *Introducción a los métodos DFT: Descifrando B3LYP sin morir en el intento (Spanish Edition)*, Independently published, Amazon.com, spanish edition, 2017, Paperback, 132 pages.
- [72] Roos B., Lindh R., Malmqvist P., Veryazov V., y Widmark P., *Multiconfigurational Quantum Chemistry*, Wiley, 2016.
- [73] Helgaker T., Jørgensen P., y Olsen J., *Molecular Electronic-Structure Theory*, Wiley, 2014.
- [74] Goodson D. Z. y Sergeev A., *Physics Letters A* **359**, 514 (2006).
- [75] Jensen F., *Introduction to Computational Chemistry*, Wiley, 2013.
- [76] Levine I., *Quantum Chemistry*, Pearson advanced chemistry series, Pearson, 2014.
- [77] Pielak L., *Ideas of Quantum Chemistry*, Elsevier Science, 2007.
- [78] Lee T. J. y Taylor P. R., *International Journal of Quantum Chemistry* **36**, 199 (1989).
- [79] Hohenberg P. y Kohn W., *Phys. Rev.* **136**, B864 (1964).
- [80] Zhao Y. y Truhlar D. G., *Theoretical Chemistry Accounts* **120**, 215 (2008).
- [81] Cohen A. J., Mori-Sánchez P., y Yang W., *Chemical Reviews* **112**, 289 (2008), “Similarly to LDA, GGA/meta-GGA functionals tend to underestimate reaction barriers ...”.

- [82] Becke A. D., *The Journal of Chemical Physics* **98**, 5648 (1993).
- [83] Sousa S. F., Fernandes P. A., y Ramos M. J., *The Journal of Physical Chemistry A* **111**, 10439 (2007).
- [84] Hirao H., *The Journal of Physical Chemistry A* **115**, 9308 (2011).
- [85] Lu L., *International Journal of Quantum Chemistry* **115**, 553 (2015).
- [86] Chai J.-D. y Head-Gordon M., *The Journal of Chemical Physics* **128**, 084106 (2008).
- [87] Grimme S., *Journal of Computational Chemistry* **27**, 1787 (2006).
- [88] Dunning J., T. H., *The Journal of Chemical Physics* **90**, 1007 (1989).
- [89] Woon D. E. y Dunning J., T. H., *The Journal of Chemical Physics* **100**, 2975 (1994).
- [90] Woon D. y Dunning T., *The Journal of Chemical Physics* **98**, 1358 (1993), Copyright: Copyright 2017 Elsevier B.V., All rights reserved.
- [91] Woon D. E. y Dunning J., T. H., *The Journal of Chemical Physics* **103**, 4572 (1995).
- [92] Jensen F., *Theoretical Chemistry Accounts* **113**, 267 (2005).
- [93] Feller D., *The Journal of Chemical Physics* **98**, 7059 (1993).
- [94] Boys S. F. y Bernardi F., *Molecular Physics* **19**, 553 (1970).
- [95] Helgaker T., Klopper W., Koch H., y Noga J., *Journal of Chemical Physics* **106**, 9639 (1997).
- [96] Halkier A., Helgaker T., Jørgensen P., Klopper W., y Olsen J., *Chemical Physics Letters* **302**, 437 (1999).
- [97] Feller D., Peterson K. A., y Dixon D. A., *Journal of Chemical Physics* **135**, 044102 (2011).
- [98] Ochterski J. W., Petersson G. A., y Jr. J. A. M., *The Journal of Chemical Physics* **104**, 2598 (1996).
- [99] Buendía-Atencio C., Gómez-Heredia A. X., Correa D., y Lorett-Velásquez V. P., *Revista Colombiana de Ciencias Químico-Farmacéuticas* **51**, 811 (2022).

- [100] Curtiss L. A., Jones C., Trucks G. W., Raghavachari K., y Pople J. A., *Journal of Chemical Physics* **93**, 2537 (1990).
- [101] Chan B., Deng J., y Radom L., *Journal of Chemical Theory and Computation* **7**, 112 (2011).
- [102] Real Academia Española, Diccionario de la lengua española: Término “calibrar”, 2024, Definición 4: “Fís. Ajustar, con la mayor exactitud posible, las indicaciones de un instrumento de medida [...]”.
- [103] Karton A. y de Oliveira M. T., *WIREs Computational Molecular Science* **15**, e1737 (2025), e1737 CMS-1118.R1.
- [104] Petersson G. A., *Complete Basis-Set Thermochemistry and Kinetics*, volume 677 of *ACS Symposium Series*, pages 237–266, American Chemical Society, 1998, 0.
- [105] Petersson G. A., Bennett A., Tensfeldt T. G., Al-Laham M. A., Shirley W. A., y Mantzaris J., *The Journal of Chemical Physics* **89**, 2193 (1988).
- [106] Petersson G. A., Complete Basis-Set Thermochemistry and Kinetics, in Lillestolen D. L. y Peterson R. A., editors, *Ab Initio Methods for Energy Surfaces, Part A, Theoretical Methods*, volume 32 of *Advances in Quantum Chemistry*, pages 445–446, Academic Press, 1998.
- [107] Callen H. B., *Thermodynamics and an Introduction to Thermostatistics*, Wiley, New York, 2nd edition, 1985.
- [108] Vatani A., Mehrpooya M., y Gharagheizi F., *International Journal of Molecular Sciences* **8**, 407 (2007).
- [109] McQuarrie D. A. y Simon J. D., *Molecular Thermodynamics*, University Science Books, Sausalito, California, 1999.
- [110] Ochterski J. W., *Thermochemistry in Gaussian*, Gaussian, Inc., 2000, Available at: <mailto:help@gaussian.com>.
- [111] Jacobsen R. L., III R. D. J., Irikura K. K., y Kacker R. N., *Journal of Chemical Theory and Computation* **9**, 951 (2013).

- [112] Brown J. y Carrington A., *Rotational Spectroscopy of Diatomic Molecules*, Cambridge Molecular Science, Cambridge University Press, 2003.
- [113] Irikura K., *Computational Thermochemistry*, 1998, American Chemical Society, Washington, DC.
- [114] Mayhall N. J., Raghavachari K., Redfern P. C., Curtiss L. A., y Rassolov V., *The Journal of Chemical Physics* **128**, 144122 (2008).
- [115] Grev R. S., Janssen C. L., y Schaefer I., H. F. *The Journal of Chemical Physics* **95**, 5128 (1991).
- [116] Rousseeuw P. J. y Croux C., *Journal of the American Statistical Association* **88**, 1273 (1993).
- [117] Chase J., M., *NIST-JANAF Thermochemical Tables, Fourth Edition*, National Institute of Standards and Technology, 1998.
- [118] Chan B., *The Journal of Physical Chemistry A* **126**, 4981 (2022), PMID: 35878062.
- [119] da Rocha N. L. y Custodio R., *Theoretical Chemistry Accounts* **139**, 38 (2020).
- [120] Gaussian Inc., Gaussian 09, Revision B.01, 2010, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox.
- [121] Michalska D., Bieńko D., Abkowicz-Bienko A., y Latajka Z., *The Journal of Physical Chemistry* **100**, 17786 (1996).

- [122] Nishiyama K., Sakiyama M., y Seki S., *Bulletin of the Chemical Society of Japan* **56**, 3171 (1983).
- [123] Chirico R. D., Kazakov A. F., y Steele W. V., *The Journal of Chemical Thermodynamics* **42**, 571 (2010).
- [124] Steele W., Chirico R., Knipmeyer S., y Nguyen A., *Journal of Chemical Thermodynamics* **24**, 449 (1992).
- [125] Buendía-Atencio C., Leyva V., y González L., *Journal of Physical Chemistry A* **114**, 9537 (2010).
- [126] Sander S. P. y Peterson M. E., *The Journal of Physical Chemistry* **88**, 1566 (1984).
- [127] Cox J. D., Wagman D. D., y Medvedev V. A., *CODATA Key Values for Thermodynamics*, Hemisphere Publishing Corp., New York, 1989.
- [128] Roux M. V., Temprado M., Chickos J. S., y Nagano Y., *Journal of Physical and Chemical Reference Data* **37**, 1855 (2008).
- [129] Miroshnichenko E., Kon'kova T., Matyushin Y., y Inozemtsev Y., *Russian Chemical Bulletin* **58**, 2015 (2009).
- [130] Ribeiro da Silva M., Lobo Ferreira A., Cabral J., Santos A., Moreno A., Galvao T., Rocha I., Fernandes P., Salgueiro S., de Moura V., Oliveira I., Coteló P., y Ribeiro M., *Journal of Chemical Thermodynamics* **41**, 984 (2009).
- [131] Kozyro A., Simirskii V., Krasulin A., Sevruck V., Kabo G., Frenkel M., Gaponik P., y Grigor'ev Y., *Russian Journal of Physical Chemistry* **64**, 348 (1990).
- [132] Morais V., Miranda M., y Matos M., *Journal of Chemical Thermodynamics* **38**, 450 (2006).
- [133] Freitas V., Oliveira L., y Ribeiro da Silva M., *Journal of Chemical Thermodynamics* **39**, 39 (2007).
- [134] Ribeiro da Silva M., Lobo Ferreira A., y Moreno A., *Journal of Chemical Thermodynamics* **41**, 904 (2009).

- 
- [135] Mo O., Yanez M., Roux M., Jimenez P., Davalos J., Ribeiro da Silva M., Ribeiro da Silva M., Matos M., Amaral L., Sanchez-Migallon A., Cabildo P., Claramunt R., Elguero J., y Liebman J., *Journal of Physical Chemistry A* **103**, 9336 (1999).
- [136] Ribeiro da Silva M. y Cabral J., *Journal of Chemical Thermodynamics* **41**, 355 (2009).
- [137] Steele W., Chirico R., Knipmeyer S., y Nguyen A., *Journal of Chemical Thermodynamics* **24**, 499 (1992).