



Universidad de Concepción  
Dirección de Postgrado  
Facultad de Ciencias Químicas-Programa de Doctorado en Ciencias con Mención  
en Química

**Energía libre de hidratación obtenidas por simulaciones de  
dinámica molecular utilizando cargas atómicas derivadas de  
la densidad electrónica**

Tesis para optar al grado de Doctor en Ciencias con Mención en  
Química

Maximiliano Ernesto Riquelme Troncoso  
Concepción-Chile  
Diciembre 2018

Profesor Guía: Dr. Esteban Vöhringer-Martinez  
Profesor Co-Guía: Dr. Eduardo Delgado Ramírez  
Dpto. de Físico Química, Facultad de Ciencias Químicas  
Universidad de Concepción



*«Porque Jehová da la sabiduría,  
y de su boca viene el conocimiento  
y la inteligencia.»*  
Proverbios 2:6

# Agradecimientos

En primer lugar agradezco a Dios por la oportunidad que me dio de realizar esta tesis doctoral. También quiero agradecer a mis padres por todo su esfuerzo, cariño y apoyo durante todas las etapas de mi vida.

Quisiera también agradecer a mi amada esposa Solange por todo su apoyo incondicional, por ser mi compañera de vida durante todos estos años y por todos los lindos momentos que hemos vivido como pololos, novios y esposos.

Quiero agradecer a mis profesores guías de pregrado Adelio Matamala y de postgrado Esteban Vöhringer-Martinez por ser parte importante en mi formación como científico. También quiero agradecer a mis profesores de asignatura Joel Alderete, Eduardo Delgado y Gina Pecchi por todo el conocimiento compartido durante el primer año de este doctorado.

También agradezco a mis compañeros y amigos de laboratorio: David, Alejandro, Danilo, Oscar, Aharon, Javier y Carlos por los buenos momentos de risas, almuerzos y ciencia vividos durante todos estos años. Por último, quisiera agradecer a Conicyt por la posibilidad que me dio de realizar este doctorado.

# Tabla de Contenido

<b>Lista de Figuras</b>	<b>vi</b>
<b>Lista de Tablas</b>	<b>x</b>
<b>Resumen</b>	<b>xi</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Marco teórico . . . . .	4
1.1.1. Energía libre de hidratación y coeficientes de partición . . . . .	4
1.1.2. Energía libre de hidratación y constantes de asociación/disociación . . . . .	6
1.1.3. Estado del arte . . . . .	9
1.2. Propuesta de investigación . . . . .	17
1.3. Hipótesis y objetivos . . . . .	19
<b>2. Métodos</b>	<b>21</b>
2.1. Métodos de estructura electrónica . . . . .	21
2.1.1. Aproximación Born-Oppenheimer . . . . .	21
2.1.2. Teoría del funcional de la densidad . . . . .	23
2.2. Mecánica molecular . . . . .	30
2.2.1. Parámetros enlazantes . . . . .	31
2.2.2. Parámetros no enlazantes . . . . .	33
2.3. Dinámica molecular . . . . .	34
2.4. Cargas atómicas . . . . .	37
2.4.1. Cargas atómicas de Hirshfeld . . . . .	37
2.4.2. Cargas atómicas de Hirshfeld-Iterativas . . . . .	38
2.4.3. Cargas atómicas MBIS . . . . .	39
2.4.4. Cargas atómicas dinámicas MBIS . . . . .	40
2.5. Cálculos de energía libre . . . . .	41
2.5.1. Método de Zwanzig . . . . .	42
2.5.2. Integración termodinámica . . . . .	43
2.5.3. Bennett Acceptance Ratio (BAR) . . . . .	43
2.5.4. Multistate Bennett Acceptance Ratio (MBAR) . . . . .	45

2.5.5. Ciclo termodinámico . . . . .	46
2.6. Corrección de polarización a la energía libre de hidratación . . . . .	48
2.7. Cálculo de las cargas atómicas y energía libre de hidratación . . . . .	49
2.8. Entalpía y entropía de hidratación . . . . .	52
<b>3. Resultados</b>	<b>54</b>
3.1. Energía libre de hidratación usando diferentes métodos de derivación de carga . . . . .	55
3.2. Momento dipolar molecular en disolución obtenidos desde diferentes modelos de cargas . . . . .	69
3.3. Energía libre de hidratación para una variedad de grupos funcionales	73
3.4. Energía libre de hidratación utilizando cargas dinámicas MBIS . . . . .	79
3.5. Entalpía y entropía usando cargas S-MBIS . . . . .	86
<b>4. Conclusiones</b>	<b>92</b>
<b>Bibliografía</b>	<b>95</b>
<b>A. Anexo: Energía libre de hidratación para una variedad de grupos funcionales</b>	<b>105</b>



# Lista de Figuras

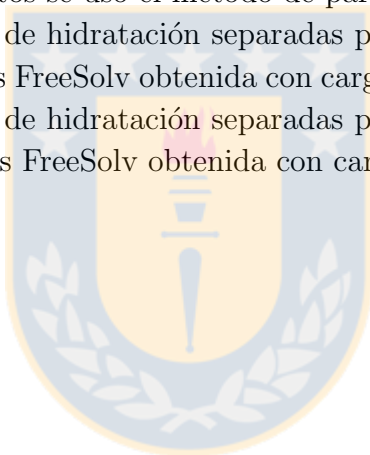
1-1.	Esquema de la absorción y deposición de contaminantes. Fuente: modificado desde Ratkova, E. L y colaboradores, 2015. <sup>1</sup> . . . . .	6
1-2.	Lipofilia. Fuente: modificado desde Ratkova, E. L y colaboradores, 2015. <sup>1</sup> . . . . .	7
1-3.	Metodologías comúnmente utilizadas en la determinación de la energía libre de solvatación. <sup>2</sup> . . . . .	10
1-4.	A) Energía libre de hidratación calculada con cargas RESP con un RMSE de 1.9 kcal/mol, B) Energía libre de hidratación calculada con cargas AM1-BCC con un RMSE de 1.3 kcal/mol, C) Energía libre de hidratación calculada con cargas Vcharge con un RMSE de 3.0 kcal/mol, D) Energía libre de hidratación calculada con cargas IPolQ-Mod con un RMSE de 1.8 kcal/mol. <sup>3</sup> . . . . .	12
1-5.	2-metilpropano. Fuente: elaboración propia. . . . .	15
1-6.	Diferentes grupos funcionales presentes en la base de datos FreeSolv. <sup>4</sup> . . . . .	18
2-1.	El método leap-frog el cual es llamado así, porque $r$ y $v$ van saltando como ranas uno sobre el otro. Fuente: elaboración propia. . . . .	36
2-2.	Esquema típico de un ciclo termodinámico. <sup>5</sup> . . . . .	46
2-3.	Cambios que ocurren a medida que vamos variando el valor de $\lambda$ . Fuente: elaboración propia. . . . .	48
3-1.	Gráfico de los valores de energías libres de hidratación calculadas versus los valores experimentales para todas las moléculas de FreeSolv consideradas, usando el campo de fuerza GAFF y las cargas atómicas obtenidas del método de partición Hirshfeld-I de la densidad de electrones moleculares BLYP / def2-TZVP en vacío (El recuadro muestra el resultado del análisis de error: error cuadrático medio (RMS), error absoluto sin signo (AUE), Kendall tau y el coeficiente de correlación Pearson R para cada conjunto de cargas.). Fuente: elaboración propia. . . . .	55

- 3-2.** Gráfico de las energías libres de hidratación calculadas versus experimentales para todas las moléculas de FreeSolv consideradas, utilizando el campo de fuerza GAFF y diferentes métodos para obtener cargas atómicas. (A) Método de estructura electrónica BLYP / def2-TZVP con modelo de solvatación SMD y método de partición Hirshfeld-I añadiendo la corrección de polarización a la energía libre; (B) el mismo método de estructura electrónica con las cargas semipolarizadas sin aplicar ninguna corrección de polarización; (C) el mismo método de estructura electrónica y el modelo de solvatación de SMD, pero usando el particionamiento iterativo a base mínima (MBIS) corrigiendo las energías libres de transferencia para la polarización y (D) usando cargas de átomo AM1-BCC en FreeSolv. Fuente: elaboración propia. . . . . 58
- 3-3.** Gráfico de las energías libres de hidratación obtenidas con cargas atómicas S-HI (azul) y las cargas atómicas S-HE (rojo) a partir de la densidad electrónica con el modelo de solvatación SMD y dos métodos diferentes de estructura electrónica: BLYP / def2-TZVP y B3LYP / def2-TZVP. Fuente: elaboración propia. . . . . 60
- 3-4.** Moléculas que presentan una desviación de la energía libre de hidratación calculada con las cargas S-MBIS del valor experimental que es más de cinco veces el error experimental. Fuente: elaboración propia. . . . . 62
- 3-5.** Distribución de los errores absolutos al valor experimental en las energías libres de hidratación obtenidas con las cargas S-HI (A), las cargas HI semipolarizadas (B), las cargas S-MBIS (C) y las cargas AM1-BCC (D). Fuente: elaboración propia. . . . . 64
- 3-6.** Energías libres de hidratación para todas las moléculas de la base de datos FreeSolv obtenidas con el campo de fuerza GAFF y las cargas MBIS incluyendo el modelo solvente SMD en el cálculo de la estructura electrónica (nivel BLYP / def2-TZVP) con la corrección de polarización (estrellas azules) y las energías libres de hidratación con el mismo campo de fuerza usando las cargas atómicas AM1-BCC frente a la referencia experimental. Fuente: elaboración propia. . . . . 65

<b>3-7.</b> Gráfico de las energías libre de hidratación experimental y calculada de todas las moléculas estudiadas en la base de datos FreeSolv obtenida como diferencia de energía libre de cálculos de estructura electrónica BLYP / def2-TZVP incluyendo corrección de energía de punto cero, rotor rígido, aproximación de oscilador armónico, las contribuciones electrostática y no electrostática (formación de cavidades, dispersión, etc.) del modelo de solvatación SMD para el estado estándar de 1 mol L <sup>-1</sup> y los mismos cálculos de estructura electrónica en vacío. Fuente: elaboración propia. . . . .	67
<b>3-8.</b> Moléculas que presentan una desviación mayor que 5 kcal / mol de la energía libre de hidratación calculada a partir del valor experimental en la figura <b>3-7</b> y su ID utilizado en FreeSolv. Fuente: elaboración propia. . . . .	68
<b>3-9.</b> Magnitud de los momentos dipolares de las moléculas en la base de datos FreeSolv calculadas para la geometría optimizada (nivel BLYP / TZVP con el modelo de disolvente SMD) directamente con el método QM frente a los momentos dipolares resultantes de las cargas atómicas obtenidas con Hirshfeld-I (A) o el método de partición MBIS (C) de la misma densidad electrónica, las cargas semipolarizadas con el método Hirshfeld-I (B) y las cargas atómicas AM1-BCC (D). Las flechas identifican valores atípicos con las estructuras que se muestran en la Figura <b>3-10</b> . Fuente: elaboración propia. . . . .	71
<b>3-10.</b> Moléculas marcadas en la Figura <b>3-9</b> que presentan una gran desviación de su momento dipolo molecular calculado con diferentes cargas atómicas. Fuente: elaboración propia. . . . .	72
<b>3-11.</b> Energías libres de hidratación separadas por grupos funcionales en la base de datos FreeSolv obtenida con cargas atómicas S-MBIS. Los valores atípicos (números) se muestran en la Figura <b>3-12</b> . Fuente: elaboración propia. . . . .	74
<b>3-12.</b> Moléculas identificadas como valores atípicos en la Figura <b>3-11</b> . Fuente: elaboración propia. . . . .	75
<b>3-13.</b> Energía libre de hidratación usando cargas D-MBIS para moléculas con bajos valores de energía libre. Fuente: elaboración propia. . . . .	80
<b>3-14.</b> Energía libre de hidratación usando cargas D-MBIS para moléculas que presentan uno o más grupos hidroxilo en su estructura. Fuente: elaboración propia. . . . .	82



- 3-15.** Energía libre de hidratación usando cargas D-MBIS para una variedad de moléculas con diferentes grupos funcionales. Fuente: elaboración propia. . . . . 83
- 3-16.** Moléculas estudiadas para la entalpía y entropía de hidratación las cuales han sido enumeradas y se informa su entalpía de hidratación experimental respectivamente en kcal/mol. Fuente: elaboración propia. 87
- 3-17.** Entalpía de hidratación usando cargas S-MBIS junto a dos disolventes explícitos y los valores reportados para las cargas AM1-CC. Fuente: elaboración propia. . . . . 89
- A-1.** Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas S-HI (para los sulfatos y fosfatos se usó el método de partición Hirshfeld-E). . . . . 105
- A-2.** Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas semipolarizadas. 106
- A-3.** Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas AM1-BCC. . 106



## Lista de Tablas

<b>3-1.</b> Comparación entre la corrección de polarización para las cargas S-MBIS y para las cargas dinámicas D-MBIS en kcal/mol. . . . .	81
<b>3-2.</b> $T\Delta S$ calculado y experimental para las 11 moléculas de la Figura <b>3-16</b> en kcal/mol. . . . .	91



# Resumen

Las simulaciones computacionales de los sistemas biomoleculares a menudo usan campos de fuerza, los cuales son combinaciones de funciones empíricas simples basadas en átomos para describir las interacciones atómicas. Aunque los campos de fuerza polarizables proporcionan una descripción más detallada de las interacciones intermoleculares, los campos de fuerza no polarizables desarrollados hace varias décadas a menudo siguen siendo preferibles debido a su bajo costo de cálculo. Las interacciones electrostáticas desempeñan un papel importante en los sistemas químicos y biomoleculares las cuales son descritas por cargas atómicas puntuales. En este trabajo se estudió el rendimiento de diferentes cargas atómicas para reproducir las energías libres de hidratación experimentales en la base de datos FreeSolv, que consiste de 613 moléculas orgánicas en combinación con el campo de fuerza GAFF. Las cargas atómicas se calcularon mediante diferentes aproximaciones de átomos en moléculas, Hirshfeld-I y Minimal Basis Iterative Stockholder (MBIS). Para tener en cuenta los efectos de la polarización, las cargas se derivaron de la densidad electrónica del soluto polarizada por un modelo de solvente implícito SMD y la energía requerida para polarizar el soluto se agregó al ciclo de energía libre. Además, para un pequeño grupo de moléculas se estudiaron las cargas dinámicas de MBIS (D-MBIS), la entalpía y entropía de hidratación.

Para los diferentes métodos de carga atómica estudiados se encontró que la mejor concordancia con los datos experimentales se observa para el método de carga atómica S-MBIS. Éste incluye la polarización de la densidad electrónica por el solvente y presenta un error cuadrático medio de  $2.0 \text{ kcal mol}^{-1}$  para las 613 moléculas orgánicas estudiadas. La mayor desviación se observó para las moléculas que contienen fósforo y las moléculas con grupos funcionales amida, éster y amina. Las cargas dinámicas dan cuenta de efectos que no describen ninguno de los demás modelos de cargas estudiados, como por ejemplo cambios conformacionales y formación de puente de hidrógeno. Estas cargas presentan problemas para compuestos alifáticos o cuando hay un gran número de grupos hidroxilos, lo cual se debe a que los parámetros de Lennard Jones del campo de fuerza fueron optimizados usando otro modelo de carga en los compuestos alifáticos y en el caso de las moléculas con gran número de hidroxilo el mayor error está asociado a la corrección de polarización que se hace

más positiva proporcional al número de hidroxilos presentes en la molécula. La entalpía de hidratación es muy dependiente del modelo de agua que se utilice, pero no depende mucho del modelo de carga utilizado ya que se estudiaron principalmente hidrocarburos, los cuales son mayormente sensibles a los parámetros no enlazantes del campo de fuerza. Pese a esto, se pudo reproducir los valores experimentales de las entalpías de hidratación con un error menor a  $1 \text{ kcal mol}^{-1}$  para la mayoría de las moléculas y con un error mayor, entre  $2\text{-}3 \text{ kcal mol}^{-1}$ , para los ciclos alcanos, lo cual se asocia a un posible error en los parámetros de Lennard Jones para este tipo de moléculas.



# 1. Introducción

A comienzos de los años 50, gracias al desarrollo de los primeros computadores digitales, nació una nueva metodología de investigación científica llamada química computacional, la cual consistía en usar computadores para resolver problemas de interés científico. Esta metodología fue en un principio una herramienta opcional a los métodos tradicionales usados en ciencia, como lo eran el método experimental y el teórico. Esta nueva metodología ejerció un gran impacto en diferentes áreas de la ciencia, siendo su principal objetivo el describir los sistemas moleculares mediante modelos y ecuaciones, las cuales son resueltas haciendo uso de computadores.

Actualmente, la química computacional es una herramienta esencial usada por teóricos y experimentalistas, debido, principalmente, a que esta herramienta permite reproducir experimentos de laboratorio y variar parámetros libremente, lo que conlleva a poder probar rangos que son imposibles de alcanzar a nivel experimental, obteniendo en algunos casos una mejor descripción microscópica del proceso de estudio.

Para estudiar un sistema molecular de manera adecuada usando la química computacional se requiere describir todas las interacciones del sistema. Idealmente se podría recurrir a modelos mecanocuánticos incluyendo los electrones en la descripción del sistema, pero en algunos casos, debido a las dimensiones de los sistemas de estudio y el tiempo de cálculo que demandan el usar modelos mecanocuánticos, no es una opción viable.

---

Los métodos clásicos son una alternativa a las metodologías cuánticas para abordar el estudio de sistemas químicos de gran tamaño en los que no se producen alteraciones en enlaces covalentes, ni cambios drásticos en la distribución electrónica. En ellos, el Hamiltoniano cuántico riguroso es reemplazado por funciones matemáticas que describen la interacción entre los átomos, lo que se conoce como campo de fuerza. Para sistemas biológicos se utilizan los siguientes campos de fuerza CHARMM,<sup>6</sup> AMBER-GAFF,<sup>7</sup> GROMOS<sup>8</sup> y OPLS.<sup>9</sup>

Un campo de fuerza esta formado por dos componentes: las funciones de energía potencial, definidas por un conjunto de ecuaciones que engloban una serie de contribuciones energéticas enlazantes y no enlazantes, y por otro lado, los parámetros empíricos usados por cada uno de los términos. Una de las características principales del campo de fuerzas es la transferencia de parámetros, de manera que haciendo la parametrización sobre un pequeño número de casos, se pueda extrapolar a un amplio espectro de moléculas reproduciendo de manera fiable los observables experimentales. Los términos enlazantes incluyen contribuciones debidas a los enlaces covalentes, ángulos de valencia y ángulos torsionales propios e impropios. Los términos no-enlazantes se definen por un término de atracción-repulsión de tipo Lennard-Jones para las fuerzas de van der Waals y un término coulombico para las interacciones electrostáticas.

En los campos de fuerza no polarizables se considera que la distribución de carga de una molécula puede ser representada utilizando cargas puntuales fraccionales normalmente centradas en los núcleos, las cuales han sido optimizadas para reproducir las propiedades electrostáticas de la molécula. La interacción electrostática entre los

átomos ( $E_{elec}$ ) queda, representada según la ley de Coulomb:

$$E_{elec} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1-1)$$

donde  $q_i$  y  $q_j$  son las cargas puntuales de cada átomo,  $r_{ij}$  la distancia entre átomos y  $\epsilon_0$  la constante dieléctrica del vacío.

La mayoría de los campos de fuerzas actuales no polarizable usan cargas derivadas de cálculos de estructura electrónica. Dependiendo del campo de fuerza utilizado las cargas atómicas varían: por ejemplo el campo de fuerza CHARMM utiliza inicialmente las cargas Merz-Kollman<sup>10,11</sup> obtenidas de cálculos mecanocuánticos a nivel MP2/6-31G\*,<sup>12</sup> el campo de fuerza AMBER99 utiliza las cargas RESP las cuales se obtienen a partir de cálculos mecanocuánticos a nivel HF/6-31G\*<sup>7,13</sup> y el campo de fuerza OPLS utiliza cargas empíricas que se han obtenido principalmente del ajuste para reproducir las propiedades de líquidos orgánicos como la entalpía de vaporización, las capacidades caloríficas y la densidad.<sup>9,14</sup> El campo de fuerza GRO-MOS utiliza cargas parciales que fueron ajustadas para reproducir las entalpías de hidratación o solvatación y los calores de vaporización en fase condensada.<sup>8</sup>

En las últimas décadas los campos de fuerza se han desarrollado y aplicado a diversos problemas químicos y biológicos, como la unión de ligandos a proteínas o ADN, el plegamiento de proteínas o la catálisis enzimática. Sin embargo, la aplicabilidad de estos campos de fuerza para predecir la función de los sistemas complejos está limitada por su capacidad para reproducir las interacciones atómicas correctamente. Por ejemplo, el proceso de unión de un ligando a una proteína está asociado con su desolvatación parcial, la eliminación de las moléculas de agua circundantes. Considerando solo este primer simple paso del proceso de unión, varios métodos computacionales y campos de fuerza han sido probados en su capacidad para reproducir la energía

libre asociada con la transferencia de una molécula o ligando de la solución acuosa a la fase gaseosa: la energía libre de hidratación  $\Delta G_{\text{Hyd}}$ .<sup>4,15-25</sup>

Para las moléculas polares, el factor dominante en las energías libres de hidratación son las interacciones electrostáticas representadas por las cargas atómicas en los campos de fuerza. Esto permite probar diferentes modelos de carga y ver si son capaces de reproducir correctamente las energías libres de hidratación, haciendo de esto una prueba exigente de la calidad del modelo de carga para así postular el uso de estos modelos al estudio de propiedades más complejas que dependan considerablemente de las interacciones electrostáticas.

## 1.1. Marco teórico

La energía libre de hidratación se define como el proceso de transferir una molécula desde una posición fija en fase gaseosa a una posición fija en la fase líquida a una temperatura y presión constante. Por convención, cuando el líquido es agua se debe hablar de energía libre de hidratación y si el líquido es cualquier otro solvente se debe hablar de energía libre de solvatación.<sup>26</sup> La energía libre de hidratación es un concepto fundamental en termodinámica, principalmente porque está relacionada intrínsecamente con varias propiedades termodinámicas importantes como los coeficientes de partición, la solubilidad y las constantes de asociación o disociación, entre otras.

### 1.1.1. Energía libre de hidratación y coeficientes de partición

El coeficiente de partición es una relación entre las concentraciones de un soluto entre dos fases en equilibrio termodinámico entre sí. Cuando se considera una



molécula de soluto entre la fase gaseosa y una disolución acuosa, el coeficiente de partición se denomina constante de la ley de Henry ( $K_H$ ):

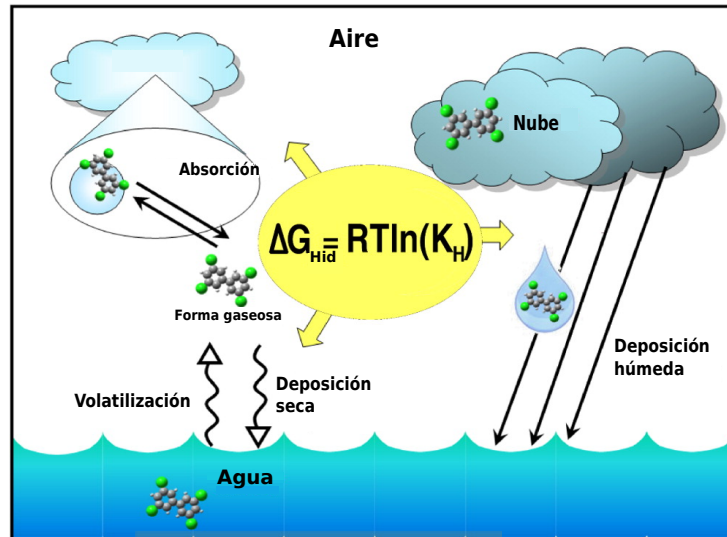
$$K_H = \frac{C_{g(i)}}{C_{ac(i)}} \quad (1-2)$$

donde  $C_{g(i)}$  y  $C_{ac(i)}$  son las concentraciones moleculares en equilibrio de las moléculas  $i$  en la fase gaseosa y acuosa, respectivamente. La relación que existe entre la energía libre de hidratación y la constante de la ley de Henry es la siguiente:

$$\Delta G_{Hid} = -RT \cdot \ln(K_H) \quad (1-3)$$

donde  $\Delta G_{Hid}$  es la energía libre de hidratación,  $R$  es la constante de los gases y  $T$  es la temperatura. Una aplicación importante de las constantes de la ley de Henry es su amplio uso en química ambiental para describir el intercambio de contaminantes orgánicos persistentes (COP) entre el aire y el agua de océanos, ríos, lagos, etc. (Figura 1-1). La contaminación producida por los COP es un problema actual e importante a nivel ambiental y ecológico, principalmente, porque estos compuestos tienen una larga persistencia, transporte y deposición a largas distancias, alta bioacumulación y todos los efectos adversos que producen sobre la biota.<sup>27-29</sup>

El logaritmo del coeficiente de partición (octanol/ agua) se relaciona con la lipofilia de grasas y lípidos en las membranas, y es un parámetro esencial para estimar la absorción, el transporte y la eliminación de drogas.<sup>30,31</sup> La lipofilia<sup>32</sup> se puede definir como el coeficiente de partición que existe entre la fase acuosa y orgánica (grasas, lípidos de membrana, etc) y es un parámetro fisicoquímico clave para comprender la distribución de moléculas en sistemas de interés farmacéutico Figura 1-2. La ecuación que relaciona el logaritmo del coeficiente de partición (octanol/ agua) con la energía libre de hidratación es la siguiente:



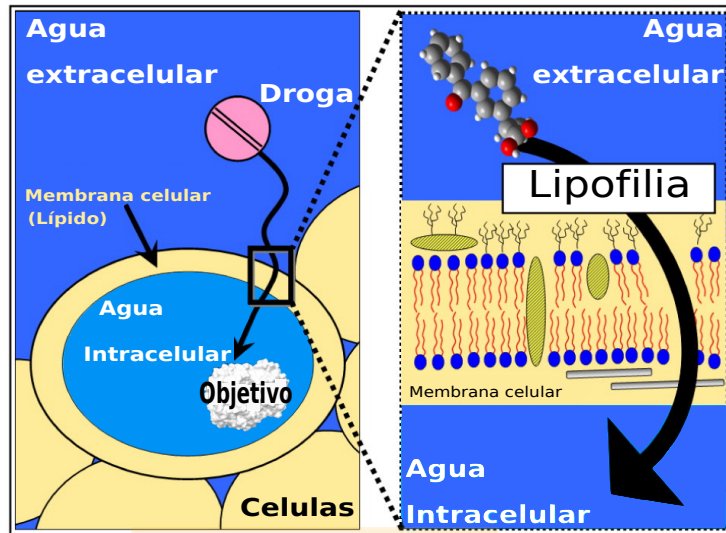
**Figura 1-1.:** Esquema de la absorción y deposición de contaminantes. Fuente: modificado desde Ratkova, E. L y colaboradores, 2015.<sup>1</sup>

$$\log P_{Oct/Agua} = \frac{-[\Delta G_{solv(Oct)} - \Delta G_{Hid}]}{RT \ln(10)} \quad (1-4)$$

donde  $\Delta G_{solv(Oct)}$  es la energía libre de solvatación en octanol. En esta ecuación  $\log P_{Oct/Agua}$  puede no corresponder exactamente al valor medido por los experimentos, debido a que la miscibilidad del agua y el n-octanol no es cero, aunque se ha aceptado ampliamente el comparar  $\log P_{Oct/Agua}$  con los valores experimentales siempre hay que tener en consideración este punto.

### 1.1.2. Energía libre de hidratación y constantes de asociación/disociación

En muchos campos de investigación como la química, bioquímica, ciencias farmacéuticas e ingeniería alimentaria, los procesos de asociación molecular son esenciales. Por ejemplo, la unión inadecuada o excesivamente fuerte de una molécula de



**Figura 1-2:** Lipofilia. Fuente: modificado desde Ratkova, E. L y colaboradores, 2015.<sup>1</sup>

fármaco a las proteínas sanguíneas puede disminuir la concentración de moléculas de fármaco libres (no unidas) en el torrente sanguíneo, lo que podría provocar una menor eficacia del fármaco frente a una enfermedad. En el proceso de unión de un ligando a una proteína la constante de asociación se denomina constante de estabilidad o constante de formación del complejo ( $K_{comp}$ ) y da cuenta de la fuerza de unión entre el ligando y la proteína. Este valor se puede determinar mediante la siguiente ecuación:

$$\ln K_{comp} = - \frac{\Delta G_r^{(g)} + \Delta G_{hid}^{LP} - (\Delta G_{hid}^L + \Delta G_{hid}^P)}{RT} \quad (1-5)$$

donde  $K_{comp}$  es la constante de formación complejo,  $\Delta G_r^{(g)}$  es la energía libre de asociación de la unión del ligando y la proteína en fase gaseosa,  $\Delta G_{hid}^{LP}$ ,  $\Delta G_{hid}^L$  y  $\Delta G_{hid}^P$  son las energías libres de hidratación del complejo proteína ligando ( $LP$ ), del ligando ( $L$ ) y de la proteína ( $P$ ), respectivamente. En general, las constantes de asociación se miden e informan como constantes de disociación ( $K_d$ ), mientras más

bajo sea el valor de  $K_d$ , mayor será la afinidad, por ejemplo, de un ligando a una proteína.

Otra área donde son importante las constante de disociación es en el estudio del comportamiento ácido-base de moléculas de interés farmacéutico, ya que muchas propiedades físicas y químicas son extremadamente sensibles a si la molécula de fármaco esta neutra o cargada, ya que esto afecta como se distribuirá dentro del organismo. Propiedades como la solubilidad acuosa, lipofilia o la permeabilidad de una membrana dependen de esta información. La acidez se puede estimar mediante la constante de disociación ácida ( $K_a$ ) de cualquier molécula. Debido a que, en general,  $K_a$  toma valores muy pequeños se suele trabajar en la práctica con el  $pK_a$  ( $pK_a = -\log K_a$ ). El  $pK_a$  se puede calcular mediante los valores de la energía libre de hidratación:

$$pK_a = \frac{\Delta G_r^{(g)} + \Delta G_{hid}^{A^-} + \Delta G_{hid}^{H^+} - \Delta G_{hid}^{HA}}{RT \ln(10)} \quad (1-6)$$

donde  $\Delta G_r^{(g)}$  es la energía libre de la reacción de disociación del ácido  $HA$  en fase gaseosa,  $\Delta G_{hid}^{A^-}$ ,  $\Delta G_{hid}^{H^+}$  y  $\Delta G_{hid}^{HA}$  son las energías libres de hidratación del anión  $A^-$ , del protón  $H^+$  y del ácido protonado  $HA$ , respectivamente,  $R$  es la constante de los gases y  $T$  es la temperatura.

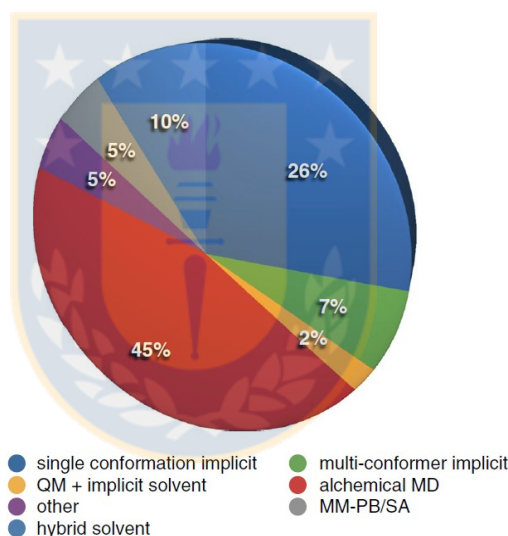
Existen dos metodologías experimentales que se utilizan comúnmente para determinar la energía libre de hidratación: los métodos directos y los métodos indirectos. La utilización del uno o del otro dependerá de la solubilidad y la volatilidad del compuesto de estudio. Los métodos directos son aplicables a moléculas con volatilidad relativamente alta ( $C_{gas}/C_{ac} > 0,1$ ) a moderada ( $0,10 > C_{gas}/C_{ac} > 0,01$ ).<sup>33</sup> Los compuestos con volatilidad menor pueden ser estudiados por métodos indirectos que calculan la energía libre de hidratación a partir de mediciones separadas de solubilidad intrínseca y presiones de vapor de compuesto puros.<sup>34</sup>

### 1.1.3. Estado del arte

La determinación de la energía libre de hidratación por simulaciones computacionales es una área que ha experimentado avances significativos en los últimos 10 años. Esto se debe principalmente a que se han implementado mejores metodologías para su determinación,<sup>35</sup> lo que ha provocado el interés de muchos investigadores alrededor del mundo. La energía libre de hidratación no es solo importante por las diferentes aplicaciones que tiene en áreas como el diseño de fármaco y química ambiental, sino que comúnmente se utiliza como parámetro para comparar campos de fuerza y estudiar posibles mejoras en ellos. La energía libre de hidratación (ELH) se utiliza principalmente, porque es una propiedad no tan compleja de calcular si se consideran solutos pequeños. El tiempo de cálculo, en general, toma un par de horas de computo y para su análisis existen diferentes metodologías simples de utilizar junto a una amplia gama de valores experimentales calculados, lo que permite poder comparar los valores computados con los valores experimentales. Además, durante el 2008 y hasta la fecha se han creado una serie de desafíos a ciegas que consisten en determinar la energía libre de hidratación de pequeñas moléculas (fragmentos de muchas drogas), energías libres de enlace y calcular coeficientes de distribución entre otros. Los desafíos fueron a ciegas, porque no se conocían o no estaban reportados los valores experimentales de dichas moléculas. Los desafíos fueron SAMPL1,<sup>36</sup> SAMPL2,<sup>37</sup> SAMPL3,<sup>38</sup> SAMPL4<sup>2</sup> y SAMPL5,<sup>39-41</sup> donde cada grupo de investigación utilizaba diferentes metodologías para determinar de la forma más precisa la energía libre de hidratación, energía libre de enlace y coeficientes de distribución de cada una de las moléculas propuestas en cada desafío. Luego, una vez terminado el desafío, se publicaban los valores experimentales para comparar qué metodologías y qué grupo de investigación se acercaban más a los valores experimentales.

Este tipo de desafíos no solo ha impulsado el interés en el cálculo de estas propiedades sino que también ha permitido que diferentes empresas farmacéuticas, como por ejemplo Merck, se interesen en determinar experimentalmente cada una de estas propiedades, aumentando así el número de moléculas que se conoce su valor de ELH experimental.

En el desafío de SAMPL4 se informaron las diferentes metodologías computacionales utilizadas por cada grupo de investigación las cuales se muestran en la Figura 1-3.



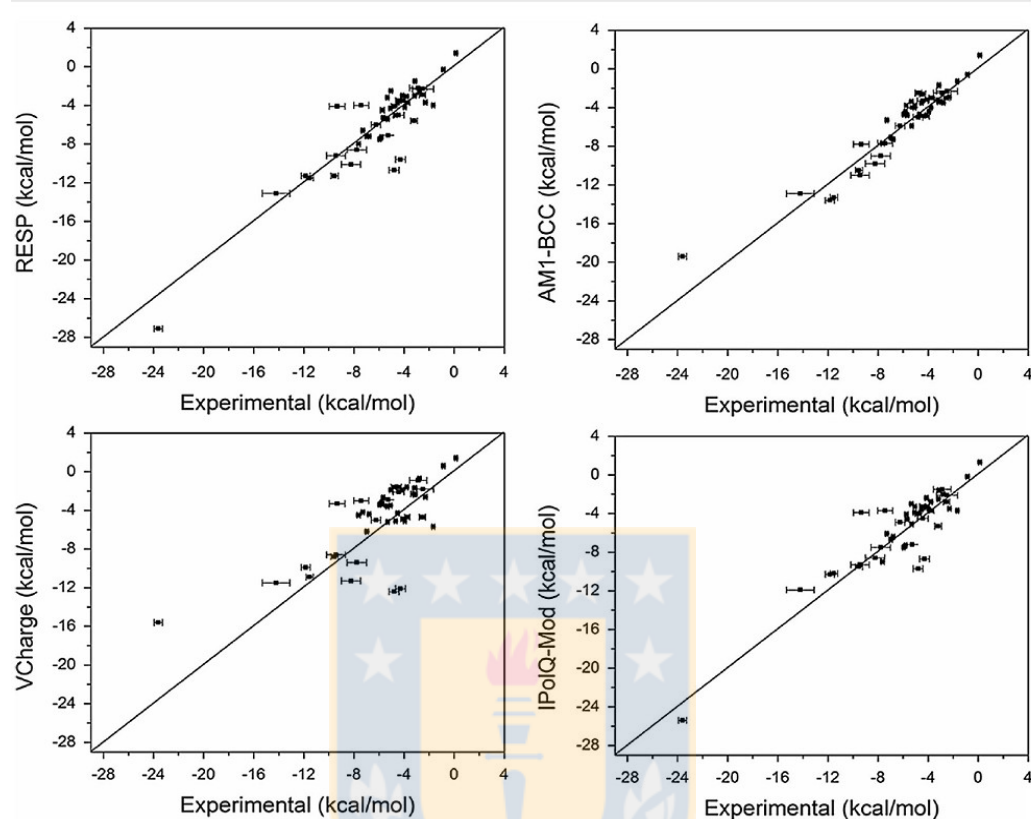
**Figura 1-3.:** Metodologías comúnmente utilizadas en la determinación de la energía libre de solvatación.<sup>2</sup>

Los mejores resultados en este estudio los obtuvieron el grupo Sandberg de la universidad de Dundee. Ellos utilizaron un enfoque a partir de mecánica cuántica con solvente implícito y correcciones de grupos funcionales.<sup>42</sup> Otro grupo que obtuvo buenos resultados fueron los científicos de software de OpenEye, ellos utilizaron un enfoque de Poisson-Boltzmann (PB) con conformaciones de baja energía en fase gas, cargas AM1-BCC y radios ZAP9.<sup>43</sup> El tercer grupo en obtener buenos resultados

fue el laboratorio de Mobley quien utilizó dinámica molecular, un solvente explícito, cargas AM1-BCC y el campo de fuerza GAFF con nuevos parámetros para el grupo hidroxilo.<sup>44</sup>

Otro estudio relevante, que está relacionado con esta temática, aplicó diferentes tipos de modelos de cargas utilizando un solvente explícito y simulaciones de dinámica molecular. Los tipos de cargas que se utilizaron fueron RESP, AM1-BCC, Vcharge y IPolQ-Mod, para las cuales se obtuvieron los gráficos mostrados en la Figura 1-4. En estos gráficos se observa que el tipo de carga atómica afecta el valor de la energía libre de hidratación que se obtiene.<sup>3</sup>

En el año 2007 se realizó el primer estudio que comparó diferentes modelos de carga atómica con el cálculo de energía libre de hidratación. Hasta esa fecha no había un estudio detallado sobre cuál modelo de carga era el más adecuado de utilizar en el cálculo de la ELH. Se compararon las cargas AM1-CM2,<sup>45</sup> AM1-BCC y RESP con diferentes metodologías, bases de cálculo y modelos de solvente explícitos. Las conclusiones que ellos obtuvieron fueron: 1) No existe un modelo de carga estándar que sea el mejor, 2) No es posible obtener errores menores a 1 kcal/mol si se comparan los resultados con los valores experimentales, 3) Usar métodos más avanzados o bases más grandes no implica obtener valores de ELH más cercanos a los experimentales y 4) Se recomiendan las cargas AM1-BCC porque es una metodología más rápida de calcular que el resto de los modelos de carga y se obtiene valores similares a, por ejemplo, las cargas RESP.<sup>46</sup> Además, no hay un consenso de que tipo de carga es mejor debido, principalmente, a que algunas cargas funcionan bien para un cierto grupo de moléculas que contienen un grupo funcional o heteroátomo en común. Esto se debe a que ciertas cargas han sido parametrizadas o incluyen correcciones para ciertos grupos funcionales de moléculas, que por lo general contienen carbono,



**Figura 1-4.:** A) Energía libre de hidratación calculada con cargas RESP con un RMSE de 1.9 kcal/mol, B) Energía libre de hidratación calculada con cargas AM1-BCC con un RMSE de 1.3 kcal/mol, C) Energía libre de hidratación calculada con cargas Vcharge con un RMSE de 3.0 kcal/mol, D) Energía libre de hidratación calculada con cargas IPolQ-Mod con un RMSE de 1.8 kcal/mol..<sup>3</sup>

hidrógeno, oxígeno y nitrógeno. Para obviar el uso de este tipo de parámetros lo ideal es utilizar cargas que se determinen a partir de primeros principios, que no incluyan correcciones y que se obtengan de una propiedad local, como por ejemplo la densidad electrónica incluyendo la polarización de ésta por el solvente.

Además de los trabajos que estudiaron el efecto del modelo de carga sobre el cálculo de la ELH existen otros estudios que abordaron otros factores, como por ejemplo el uso de disolventes implícitos basados en la ecuación de Poisson-Boltzmann, el uso



de disolventes explícitos y la variación de los parámetros para los radios atómicos de (C, H, O carbonílico, N secundario y terciario, S, F, Cl, I) utilizando moléculas con grupos funcionales como amidas, compuestos de polifluoro, haloéteres y poliésteres. Sorprendentemente los errores obtenidos en usar un solvente implícito o explícito son parecidos, sin embargo, ambos métodos tienen muchas cosas por mejorar aún, por ejemplo, ambos no consideran los efectos de polarización del soluto al pasar de la fase gaseosa a la acuosa. Los parámetros de los radios atómicos que se parametrizaron para el conjunto de moléculas estudiadas se deben actualizar en un futuro, idealmente incluyendo compuestos con otros grupos funcionales que se encuentren en moléculas similares a las drogas a medida que se disponga de datos experimentales de la ELH e incluyendo derivados de sulfonilo, derivados de fosforilo, compuestos con grupos nitro y heterociclos, por nombrar solo algunos.<sup>47</sup>

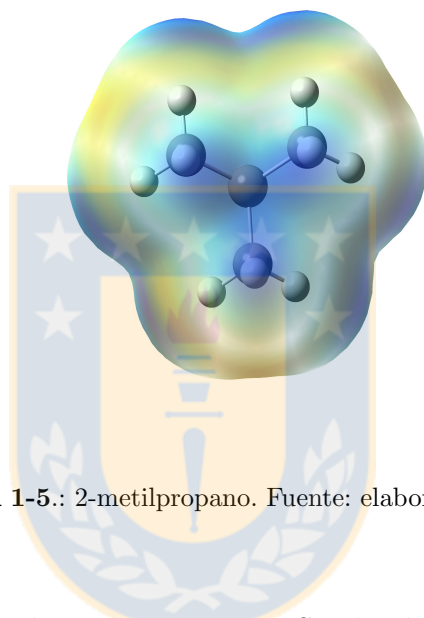
Otro factor que se ha explorado es el efecto de los cambios conformacionales en el cálculo de la energía libre de hidratación usando 504 moléculas y diferentes conformaciones, como por ejemplo la conformación más estable en fase gaseosa o la más estable en fase acuosa, todo esto usando disolvente implícito en el cálculo de la energía libre de hidratación. Las principales conclusiones fueron que el usar una sola conformación fija implica errores entre 0.23-1.96 kcal/mol dependiendo la conformación seleccionada y que los errores son el resultado de dos factores: Primero, el soluto presenta un conjunto de conformaciones en cada fase y dado que el conjunto puede cambiar con la hidratación, puede haber cambios de entropía de solutos que no están incluidos en los enfoques de conformación única. En segundo lugar, la (s) conformación (es) dominante (s) del soluto pueden diferir en las dos fases. Cuando esto sucede, ninguna conformación individual es apropiada para ambas fases y en algunos casos los errores introducidos mediante el uso de conformaciones individuales

son más grandes que los introducidos por el campo de fuerza.<sup>48</sup> Por último, existen otros estudios interesantes que se enfocaron en factores como cuáles son los mejores parámetros de simulación para el cálculo de la energía libre de hidratación,<sup>49</sup> descripción de las herramientas computacionales para el análisis de la energía libre<sup>50</sup> y códigos en python que permiten automatizar todo lo que implica el cálculo de la energía libre.<sup>51</sup>

El desafío que queda es continuar buscando metodologías que nos permitan mejorar la descripción de las interacciones atómicas en sistemas condensados y así obtener mejores valores de energías libres. El presente trabajo se enfoca en la descripción de las interacciones no enlazantes, en especial las electrostáticas descritas por cargas atómicas. Si bien es importante mencionar que las cargas atómicas no son un observable mecanocuántico, las cargas son partes del lenguaje cotidiano de la química y, aparte de las geometrías y las energías, son, ciertamente, las cantidades más comúnmente solicitadas en los cálculos quimicuánticos. La distribución de carga no solo ayuda a los químicos a estimar la estructura molecular y la estabilidad, sino que también nos habla sobre la química que pueden sufrir las moléculas y este concepto también ayuda a entender de manera más simple como interactúan las moléculas a través de las fuerzas intermoleculares.

Como ya se ha mencionado, dependiendo del campo de fuerza se utiliza un cierto tipo de carga atómica, por lo tanto es importante conocer de dónde proviene la carga que entrega el campo de fuerza y las ventajas y desventajas que conlleva el utilizarla. El campo de fuerza AMBER, por ejemplo, utiliza cargas RESP las cuales se obtienen de cálculos ab initio a nivel HF/6-31G\*. Para obtener las cargas RESP primero se deben calcular las cargas ESP,<sup>52</sup> las cuales se obtienen a partir de un ajuste de mínimos cuadrados, de tal manera que las cargas atómicas reproduzcan el

potencial electrostático molecular. La principal desventaja de las cargas ESP es que dependen de la base de cálculo que se utiliza y los átomos distantes de la superficie donde se evalúa el potencial pueden tomar cualquier valor de carga como el carbono 2 del 2-metilpropano en la Figura 1-5.



**Figura 1-5.:** 2-metilpropano. Fuente: elaboración propia.

Por esta razón se postularon las cargas RESP donde se aplica una función de penalización sobre las cargas de los átomos de carbono.<sup>53</sup> Sin embargo, la desventaja de las cargas RESP es que dependen de la conformación y la orientación de la molécula.

Recientemente se ha propuesto un tipo de cargas atómicas que se obtiene de forma iterativa a partir de una propiedad local como lo es la densidad electrónica. Estas son las cargas de Hirshfeld-I,<sup>54</sup> las cuales dan una descripción más adecuada del concepto de carga ya que se obtiene de una propiedad local y no de una propiedad externa como en el caso de las cargas RESP, las cuales reproducen el potencial electrostático. Además, se ha demostrado que las cargas de Hirshfeld-I reproducen también el potencial electrostático y no dependen de la conformación utilizada, lo que las hace muy útiles para ser utilizadas en dinámica molecular. Otro tipo de carga similar a las

cargas de Hirshfeld-I son las cargas MBIS (Minimal Basis Iterative Stockholder),<sup>55</sup> las cuales mejoran algunas deficiencias que tienen las cargas de Hirshfeld-I, como por ejemplo en el esquema de Hirshfeld-I, los proátomos son interpolaciones lineales entre densidades de iones aislados, de modo que las poblaciones AIM y las poblaciones proatómicas coincidan. En sistemas con átomos muy negativos ( $-2 < q_A < -1$ ), esto conduce a proátomos que se interpolan entre aniones aislados y dianiones, sin embargo, ningún dianión atómico aislado es estable. Cuando se estudian óxidos en los cuales el oxígeno puede estar como dianión o trianión se ha demostrado la escasa aplicabilidad de las cargas HI a este tipo de sistemas. El método MBIS no necesita densidades de iones (inestables) como entrada y, por lo tanto, uno esperaría que no sufra los mismos problemas de sobre polarización que Hirshfeld-I.<sup>56-58</sup> Las cargas HI y MBIS se derivan de la densidad electrónica en vacío y no incluye el efecto del solvente. Para enmendar este problema se puede calcular ambas cargas utilizando un modelo de solvente implícito, a las cuales desde aquí en adelante se denominarán cargas S-HI y S-MBIS. Además, existe una nueva metodología<sup>59</sup> que tiene como ventaja que considera el efecto del solvente o del entorno molecular que existe sobre el soluto a medida que transcurre la simulación de dinámica molecular y este efecto se ve reflejado sobre las cargas de Hirshfeld- I o MBIS. Esta metodología utiliza simulaciones de dinámica molecular mediante cálculos de QM/MM donde se describe el soluto mediante mecánica cuántica y el disolvente o el entorno mediante mecánica molecular. Las cargas son calculadas y reemplazadas en el campo de fuerza iterativamente hasta su convergencia. Debido a que las cargas son obtenidas de la densidad electrónica del soluto polarizada por las cargas puntuales del solvente (QM/MM) éstas incluyen el efecto del entorno molecular. A este tipo de cargas, como se derivan a partir de simulaciones de dinámica molecular, se les denomina cargas dinámicas

Hirshfeld-I (D-HI) o cargas dinámicas MBIS (D-MBIS).

El utilizar cargas que describan correctamente nuestro sistema de estudio permite describir de mejor manera las interacciones no enlazantes. Como la energía libre de hidratación es dependiente de este tipo de interacciones, el mejorar esta descripción permitirá disminuir el error en el cálculo de la energía libre y postular estas cargas para ser utilizadas dentro de algún campo de fuerza.

Inmerso dentro del concepto de la energía libre de hidratación están los términos de  $\Delta H_{hid}$  y  $\Delta S_{hid}$ , los cuales son importantes para comprender los mecanismos subyacentes de auto-ensamblaje en fenómenos tales como la formación de micelas, el reconocimiento molecular y plegamiento de proteínas.<sup>60</sup> Por lo general los cálculos de energía libre se enfocan en calcular de buena manera el valor de la energía libre de hidratación ( $\Delta G_{hid}$ ), pero no se toma en cuenta los valores de la entalpía de hidratación ( $\Delta H_{hid}$ ) y la entropía de hidratación ( $\Delta S_{hid}$ ). A partir de la termodinámica estos tres conceptos se pueden relacionar mediante la siguiente ecuación:

$$\Delta G_{hid} = \Delta H_{hid} - T\Delta S_{hid} \quad (1-7)$$

Puede suceder que por una compensación de errores el valor de energía libre de hidratación calculado sea cercano al experimental, pero esto no implica que el valor de  $\Delta H_{hid}$  y  $\Delta S_{hid}$  sean adecuados.

## 1.2. Propuesta de investigación

Como se ha analizado en la literatura, quedan muchas variables por estudiar tanto en el cálculo de la energía libre de hidratación como en el estudio de la variación de

algún parámetro dentro de un campo de fuerza. Es por esta razón que la presente tesis se enfocará en mejorar la descripción de las interacciones no enlazantes de tipo coulómbicas en campos de fuerza, utilizando las cargas atómicas de Hirshfeld-I (HI) en vacío, en disolvente implícito y las cargas dinámicas de MBIS (D-MBIS), las cuales describen de mejor manera el concepto de carga ya que se obtienen a partir de una propiedad local como lo es la densidad electrónica molecular.

Este tipo de cargas mejorará los cálculos de energía libre de hidratación sobre todo en moléculas polares, ya que en éstas las interacciones electrostáticas tienen un fuerte efecto sobre la energía libre de hidratación.

Para validar los distintos tipos de carga se calculará la energía libre de hidratación para una base de datos denominada FreeSolv,<sup>4</sup> de la cual se estudiaron 613 moléculas, las cuales en su mayoría son parte estructural de fármacos. Esta base de datos se utiliza como método de prueba para probar diferentes mejoras en los campos de fuerza ya que contiene los valores experimentales para una gama de diferentes grupos funcionales y heteroátomos (ver Fig. 1-6).

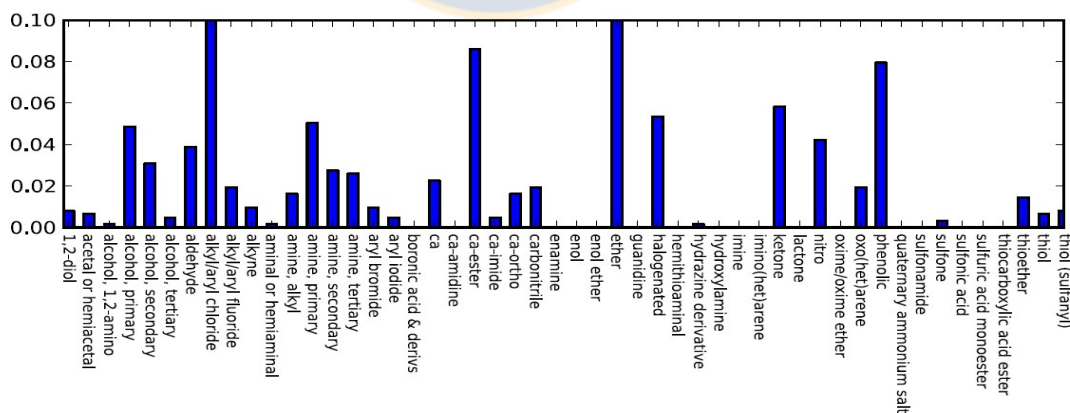


Figura 1-6.: Diferentes grupos funcionales presentes en la base de datos FreeSolv.<sup>4</sup>

Finalmente los resultados serán contrastados con los valores experimentales y los resultados reportados para simulaciones de dinámica molecular, utilizando el campo

de fuerza GAFF con cargas AM1-BCC y el modelo de agua TIP3P.<sup>61</sup>

## 1.3. Hipótesis y objetivos

### Hipótesis

Las cargas atómicas Hirshfeld-I y MBIS derivadas desde la densidad electrónica en combinación con los otros parámetros del campo de fuerza GAFF presentarán una menor desviación de los valores de energía libre de hidratación experimentales de solutos en la base de datos FreeSolv debido a una mejor descripción de las interacciones electrostáticas.

### Objetivo General

Determinar la energía libre de hidratación para moléculas de la base de datos FreeSolv utilizando las cargas S-HI y S-MBIS a 25C y 1 atm de presión.

### Objetivos específicos

1. Optimizar las geometrías de las 613 moléculas y calcular las cargas HI en vacío, semipolarizadas, S-HI y S-MBIS con el disolvente implícito SMD.
2. Simetrizar las cargas HI, S-HI, S-MBIS y con estas cargas crear los archivos de entrada de GROMACS para realizar las simulaciones de dinámica molecular.
3. Calcular la energía libre de hidratación para los 613 sistemas utilizando los diferentes modelos de carga, el programa GROMACS 5.0.4 y 20 valores de lambda.

4. Calcular para un grupo de moléculas de la base de datos FreeSolv la energía libre de hidratación utilizando las cargas dinámicas de MBIS.
5. Determinar  $\Delta H_{hid}$  y  $\Delta S_{hid}$  para un grupo de moléculas a las cuales se le conoce el valor experimental de estas propiedades.





## 2. Métodos

Para calcular la energía libre de hidratación se utiliza la química computacional la cual se encarga de usar las leyes fundamentales de la física para estudiar sistemas químicos y reacciones químicas haciendo uso de computadores. Dentro de la química computacional nuestro sistema o propiedad puede ser estimada mediante dos metodologías: los métodos de estructura electrónica o la mecánica molecular.

### 2.1. Métodos de estructura electrónica

Sí el sistema que se quiere estudiar no es tan grande se puede hacer uso de los métodos de estructura electrónica los cuales usan las leyes de la mecánica cuántica para describir los electrones y los núcleos en las moléculas resolviendo la ecuación de Schrödinger.

#### 2.1.1. Aproximación Born-Oppenheimer

La ecuación que describe la mecánica cuántica es la ecuación de Schrödinger que generalmente se usa en su versión independiente del tiempo:

$$H\psi = E\psi, \quad (2-1)$$

donde  $H$  es el operador hamiltoniano, que opera sobre la función de onda  $\psi$  y nos devuelve como valor propio la energía  $E$  del sistema. Cuando se desea resolver esta ecuación hay un término que es difícil de determinar ya que da cuenta del movimiento correlacionado de los núcleos y los electrones. Una manera de simplificar este problema es lo que se conoce como aproximación de Born-Oppenheimer<sup>62,63</sup> y es una de las bases de los métodos de química computacional. Esta aproximación considera que los núcleos son mucho más pesados que los electrones, y que, por lo tanto, se mueven mucho más lentamente que los electrones. Esta aproximación permite desacoplar estos dos movimientos y considerar las energías electrónicas para posiciones fijas de los núcleos. Esto permite omitir el término de energía cinética de los núcleos al igual que la correlación del potencial atractivo electrón-núcleo, y la repulsión entre los núcleos se considera una constante para una geometría dada. Así el hamiltoniano se la ecuación 2-1 se transforma en la siguiente expresión:

$$H_{el} = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}}. \quad (2-2)$$

De esta forma, la ecuación de Schrödinger electrónica se puede escribir como:

$$(H_{el} + V_N) \psi_{el}(r_i; R_A) = E_{el} \psi_{el}(r_i; R_A), \quad (2-3)$$

donde  $V_N$  es la energía de repulsión internuclear, y las coordenadas electrónicas  $r_i$  son variables independientes pero las coordenadas nucleares  $R_A$  son parámetros. El valor propio  $E_{el}$  de la ecuación de Schrödinger electrónica se llama la energía electrónica. Como el término  $V_N$  es una constante para una geometría dada, y las constantes añadidas al Hamiltoniano no afectan a las funciones de onda obtenidas, en la práctica se suele resolver la ecuación (2-3) sin incluir  $V_N$ , que se añade posteriormente al valor

propio obtenido para la energía.

La aproximación Born-Oppenheimer está justificada en la gran mayoría de casos, y nos permite introducir conceptos importantes como el de superficie de energía potencial, que es la superficie definida por  $E_{el}$  sobre todas las posibles coordenadas nucleares. Una vez que se conoce esta superficie, se puede tratar el movimiento nuclear desde un enfoque clásico o cuántico suponiendo que los núcleos se mueven bajo la influencia de la energía potencial electrónica. Esto permite simplificar considerablemente los cálculos requeridos para describir a un sistema químico. La aproximación Born-Oppenheimer también permite definir conceptos muy útiles como las geometrías de equilibrio o de estados de transición, que son puntos críticos de la superficie de energía potencial del sistema.

### 2.1.2. Teoría del funcional de la densidad

La teoría del funcional de la densidad del inglés (Density Functional Theory, DFT)<sup>64</sup> se basa en la idea que toda la información que puede ser obtenida por la función de onda también puede ser obtenida a partir de la densidad electrónica  $\rho(r)$ . También tiene la ventaja que solamente depende de 3 coordenadas espaciales a diferencia de la función de onda que depende de  $3n$  coordenadas espaciales donde  $n$  representa el número de electrones. Además, la densidad electrónica tiene un sentido más físico y es un observable que se puede comprender más fácilmente. Por lo tanto, la mayoría de las propiedades se pueden obtener a partir de la densidad electrónica como por ejemplo la energía.

La metodología de DFT se basa en dos teoremas que desarrollaron Hohenberg y Kohn. El primero de ellos dice que la densidad electrónica  $\rho(r)$  determina el potencial externo. Esto quiere decir que la forma de la densidad electrónica da cuenta de

las posiciones y cargas de núcleos al igual que cualquier potencial externo. Esto implica también que al integrar la densidad electrónica sobre todo el espacio se debe obtener el número de electrones del sistema en estudio, por lo tanto, el sistema está completamente determinado por la densidad electrónica. El segundo teorema dice que puede obtenerse variacionalmente la densidad que minimiza el funcional de energía y esta corresponde a la densidad del estado fundamental. Esto quiere decir que utilizando un proceso de minimización similar al usado por el método de Hartree-Fock se puede obtener la densidad electrónica exacta y por lo tanto la energía exacta del estado fundamental.

El principal problema en la teoría del funcional de la densidad es la obtención del funcional que vincula la energía con la densidad electrónica. En una primera aproximación este funcional se puede expresar mediante la siguiente expresión:

$$E[\rho] = \int v(r)\rho(r)dr + T[\rho] + V_{ee}[\rho], \quad (2-4)$$

donde  $v(r)$  es el potencial externo a los electrones,  $V_{ee}[\rho]$  es la energía de interacción electrón-electrón y  $T[\rho]$  es la energía cinética de los electrones.

### Método de Kohn-Sham

Este método fue propuesto por Kohn y Sham<sup>65</sup> el cual propone considerar un sistema de referencia ( $s$ ) de  $n$  electrones los cuales no interactúan entre sí, cuya densidad electrónica es igual a la densidad electrónica exacta del sistema interactuante. Estos electrones quedan definidos por orbitales monoeléctricos ortonormales  $\phi_i(r)$  y la función de onda queda definida por un determinante de Slater y la densidad electrónica se puede expresar por la siguiente ecuación:

$$\rho(r) = \sum_i^n |\phi_i(r)|^2, \quad (2-5)$$

y la energía cinética queda definida por:

$$T_s[\rho] = -\frac{1}{2} \sum_i^n \langle \phi_i(r) | \nabla^2 | \phi_i(r) \rangle, \quad (2-6)$$

$T_s[\rho]$  es la energía cinética para el sistema de referencia  $s$  la cual se puede considerar como una buena aproximación a la energía cinética real ( $T[\rho]$ ). Para obtener los orbitales  $\phi_i(r)$  se debe resolver la siguiente ecuación:

$$\left[ -\frac{1}{2} \nabla^2 + v_s(r) \right] \phi_i(r) = \varepsilon_i \phi_i(r), \quad (2-7)$$

en la cual  $v_s(r)$  representa al potencial externo que se introduce en el sistema de referencia que no necesariamente debe ser igual al potencial del sistema real,  $v(r)$ . Los orbitales  $\phi_i(r)$  son solo un instrumento para obtener la densidad electrónica  $\rho(r)$  y no tienen ningún significado físico. Para obtener un densidad electrónica igual a la del sistema real se intenta encontrar una expresión para el potencial externo del sistema de referencia  $v_s(r)$  mediante la siguiente partición :

$$T[\rho] + V_{ee}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho], \quad (2-8)$$

donde  $J[\rho]$  representa a la energía de interacción electrostática de los electrones definida por la siguiente expresión:

$$J[\rho] = \frac{1}{2} \int \int \frac{1}{r_{12}} \rho(r_1) \rho(r_2) dr_1 dr_2, \quad (2-9)$$

y  $E_{xc}[\rho]$  en la ecuación 2-8 es la energía de correlación-intercambio definida por la siguiente ecuación:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]), \quad (2-10)$$

El potencial externo  $v_s(r)$  llamado potencial de Kohn-Sham puede ser representado por la siguiente ecuación diferencial:

$$v_s(r) = v(r) + \frac{\delta J[\rho]}{\delta \rho} + \frac{\delta E_{xc}[\rho]}{\delta \rho} = v(r) + v_J(r) + v_{xc}(r), \quad (2-11)$$

donde  $v(r)$  es un potencial atractivo generado por los núcleos sobre los electrones,  $v_J(r)$  es igual a:

$$v_J(r) = \int \frac{\rho(r')}{|r' - r|} dr', \quad (2-12)$$

y  $v_{xc}(r)$  representa el potencial de intercambio-correlación. Esto quiere decir que agregando el potencial de Kohn-Sham (Ec. 2-11) en la ecuación 2-7 se obtiene la siguiente expresión:

$$\left[ -\frac{1}{2} \nabla^2 + v(r) + v_J(r) + v_{xc}(r) \right] \phi_i(r) = \varepsilon_i \phi_i(r), \quad (2-13)$$

Resolviendo esta ecuación se pueden obtener los orbitales  $\phi_i(r)$  que proporcionan la densidad electrónica deseada, cada uno de estos orbitales son una combinación lineal de funciones base. En la ecuación 2-13 aparecen valores que dependen de  $\rho(r)$  como  $v_J(r)$  y  $v_{xc}(r)$ , esto implica que para poder resolver esta ecuación se debe hacer de manera iterativa, al igual como sucede con el método de Hartree-Fock. Desafortunadamente, para el método de Kohn-Sham no se conoce la expresión exacta

del potencial de intercambio-correlación  $v_{xc}(r)$ , por lo tanto se hace necesario emplear potenciales aproximados y en la mayoría de los casos con parámetros ajustados. Esto implica que estos métodos pese a estar basados en principios exactos se consideran semiempíricos. Pese a todo esto, el método de Kohn-Sham es el más usado para realizar estudios de estructura electrónica, ya que se pueden realizar cálculos de sistemas relativamente complejos en un tiempo razonable y con una buena precisión.

### El funcional de correlación-intercambio

El principal problema de los métodos DFT es el potencial de correlación-intercambio como ya se ha mencionado. No existe actualmente una forma de obtener la expresión exacta para este funcional, pero si se puede buscar alternativas como por ejemplo se pueden obtener aproximaciones basándose en los soluciones de sistemas sencillos o también hacer uso de parámetros cuyo valor se calculó por ajuste a datos experimentales.

La aproximación de la densidad local

Para un sistema sencillo como un gas de electrones uniforme se puede encontrar una solución casi exacta. Para esto se define un número grande de electrones  $N$  que tiende a infinito, en un volumen  $V$ , por el cual se distribuye una carga positiva uniforme que hace al sistema neutro y la densidad  $\rho = \frac{N}{V}$ . Los resultados obtenidos son los siguientes:

$$E_{xc}^{LDA}[\rho] = \int \varepsilon_{xc}^{LDA}[\rho] \rho(r) dr, \quad (2-14)$$

$$\varepsilon_{xc}^{LDA}[\rho] = \varepsilon_x^{LDA}[\rho] + \varepsilon_c^{LDA}[\rho], \quad (2-15)$$

$$\varepsilon_x^{LDA}[\rho] = -\frac{3}{4} \left( \frac{3\rho(r)}{\pi} \right)^{1/3}, \quad (2-16)$$

donde  $\varepsilon_x^{LDA}$  representa la energía de intercambio por electrón, que genera el funcional de intercambio de Slater-Dirac.<sup>66,67</sup> Debido a la complejidad de la componente de correlación esta se ha calculado por métodos numéricos. Los autores<sup>68</sup> propusieron el siguiente funcional para la correlación:

$$\varepsilon_c^{VWN}[\rho] = A \left[ \ln \frac{x^2}{X(x)} + \frac{2b}{Q} P - \frac{bx_0}{X(x_0)} \left( \ln \frac{(x-x_0)^2}{X(x)} + \frac{2(b+2x_0)}{Q} P \right) \right], \quad (2-17)$$

$$x = \left( \frac{3}{4\pi\rho(r)} \right)^{1/6} \quad X(x) = x^2 + bx + c$$

$$Q = (4c - b^2)^{1/2} \quad P = \arctan \frac{Q}{2x + b}$$

Al combinar ambos funcionales se obtiene la aproximación de la densidad local (Local density approximation, LDA), y cada funcional depende únicamente del valor de la densidad en cada punto, por lo que se dice que son funcionales locales.

Funcionales corregidos con el gradiente

Una opción para mejorar los funcionales LDA es incluir el gradiente de la densidad electrónica, a lo cual se le denomina aproximación del gradiente (Generalised Gradient Approximations, GGA). Dentro de los funcionales GGA existe una amplia gama, donde los más destacados son :

- Funcional de intercambio de Becke:<sup>69</sup>

$$E_x^{B88}[\rho] = E_x^{LDA}[\rho] - \beta \int \rho^{4/3} \frac{x^2}{1 + 6\beta x \text{sen}^{-1} h x} dr, \quad (2-18)$$



$$x = \frac{|\nabla\rho|}{\rho^{4/3}} \quad \beta = 0,0042$$

- Funcional de correlación de Lee-Yang-Parr:<sup>70</sup>

$$E_c^{LYP}[\rho] = -a \int \frac{\rho}{1 + d\rho^{-1/3}} dr - ab \int \omega \rho^2 \left[ \frac{3}{10} (3\pi^2)^{2/3} \rho^{8/3} - \left( \frac{1}{24} + \frac{7}{72} \delta \right) |\nabla\rho|^2 \right] dr, \quad (2-19)$$

$$\omega = \frac{e^{-c\rho^{-1/3}}}{1 + d\rho^{-1/3}} \rho^{-11/3} \quad \delta = c\rho^{-1/3} + \frac{d\rho^{-1/3}}{1 + d\rho^{-1/3}}$$

$$a = 0,04918 \quad b = 0,132 \quad c = 0,2533 \quad d = 0,349$$

- Funcional híbrido de tres parámetros de Becke:<sup>71</sup>

$$\begin{aligned} E_{xc}^{B3} &= E_{xc}^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{B88} - E_x^{LDA}) + a_c(E_c - E_c^{VWN}) \\ &= a_0 E_x^{HF} + a_x E_x^{B88} + (1 - a_0 - a_x) E_x^{LDA} + a_c E_c + (1 - a_c) E_c^{VWN} \end{aligned} \quad (2-20)$$

donde  $a_0=0.20$ ,  $a_x=0.72$  y  $a_c = 0,81$ . Mediante el método de Hartree-Fock se calcula el funcional de intercambio exacto ( $E_x^{HF}$ ) usando los orbitales de Kohn-Sham. Además, para el valor de  $E_c$  se puede usar cualquier funcional de correlación, siendo uno de los más utilizado el funcional  $E_c^{LYP}$  que da como resultado el método llamado B3LYP.

Si comparamos el método de Hartree-Fock con el método de DFT ambos tie-

nen un costo computacional parecido, pero los resultados de DFT son muchos mejores e incluso comparable a los métodos Hartree-Fock. Esto implica que el usar esta metodología es una opción interesante para el estudio de cualquier propiedad, es por esta razón que en la presente tesis se usaron los métodos DFT BLYP y B3LYP para obtener las optimizaciones de las geometrías y todos los cálculos de estructura electrónica.

## 2.2. Mecánica molecular

La mecánica molecular se basa en las leyes de la mecánica clásica aplicada al estudio de sistemas moleculares en los cuales sus interacciones pueden ser representadas mediante potenciales analíticos sencillos. A esta representación se les denomina campos de fuerza los cuales están compuestos por dos componentes: el primero es un conjunto de ecuaciones matemáticas llamadas funciones de potencial las cuales son usadas para obtener la energía potencial a partir de una cierta estructura y el segundo componente es un conjunto de parámetros empíricos usados en cada una de estas ecuaciones. Existe una gran variedad de campos de fuerza, pero todos ellos hacen uso implícito de la aproximación de Born-Oppenheimer la cual se describió en la sección 2.1, la cual permite separar el movimiento nuclear del electrónico. Haciendo uso de esta aproximación se puede asumir que los núcleos se desplazan sometidos al potencial efectivo creado por la distribución electrónica. Esto hace posible solo usar las coordenadas nucleares para describir el efecto del potencial electrónico haciendo uso de una variedad de ecuaciones simples y parámetros empíricos.

Sí el tamaño del sistema que se quiera estudiar es muy grande este puede ser abordado haciendo uso de la mecánica molecular. La mecánica molecular modela las estructuras químicas mediante el uso de un campo de fuerza el cual esta caracterizado

por un conjunto de ecuaciones que definen la variación de la energía potencial de las moléculas según la posición de los átomos. Estos campos de fuerza contienen una gran cantidad de parámetros como por ejemplo: longitudes de enlace, ángulos y constantes de fuerzas, entre otros. Cuando se utiliza un campo de fuerza los electrones no son tratados explícitamente en este tipo de metodología, pero si se incluyen los efectos electrónicos mediante los parámetros del campo de fuerza.

En un campo de fuerza la energía potencial del sistema se puede separar en dos contribuciones mediante la siguiente expresión:

$$E_{sist} = E_{\text{términos enlazantes}} + E_{\text{términos no enlazantes}}, \quad (2-21)$$

el término  $E_{\text{términos enlazantes}}$  da cuenta de la energía de interacción entre átomos enlazados e incluye las contribuciones debidas a los enlaces covalentes, los ángulos de enlace y las torsiones propias e impropias. El término  $E_{\text{términos no enlazantes}}$  describe la energía de interacción entre los átomos no enlazados y se define por un término de Coulomb para las interacciones electrostáticas y un término de Lennard-Jones para las fuerzas de van der Waals.

### 2.2.1. Parámetros enlazantes

- Energía de tensión de los enlaces

Para describir la energía de tensión de los enlaces dentro de un campo de fuerza se puede representar por un potencial armónico o anarmónico. Cuando se considera un potencial anarmónico se puede usar un potencial de Morse y si se considera un potencial armónico se puede usar la ecuación de Hooke, donde cada enlace se considera como un resorte y puede ser representada por la

siguiente ecuación:

$$E_r = \sum_i^{enl.} K_i^r (r_{ij} - r_{eq})^2, \quad (2-22)$$

donde la sumatoria se hace sobre todos los enlaces que definen el sistema,  $r_{ij}$  representa la distancia de enlace,  $r_{eq}$  es la distancia de enlace de equilibrio y  $K_i^r$  es la constante de fuerza para el estiramiento de los átomos  $i$  y  $j$  y depende del tipo de enlace.

- Energía de flexión

La energía de flexión producida por dos enlaces con un átomo en común los cuales forman un ángulo puede ser representada por un oscilador armónico similar a la ecuación 2-22 mediante la siguiente ecuación:

$$E_\theta = \sum_i^{\text{áng.}} K_i^\theta (\theta_{ijk} - \theta_{eq})^2, \quad (2-23)$$

donde  $K_i^\theta$  es la constante de fuerza,  $\theta_{ijk}$  es el ángulo de enlace para una cierta geometría,  $\theta_{eq}$  es el ángulo de equilibrio formado cuando el potencial se hace mínimo y la sumatoria se realiza sobre todos los ángulos del sistema.

- Energía de torsión

Para cada ángulo en una molécula existe un plano, cuando dos ángulos tiene un enlace en común forman un ángulo de torsión  $\phi$ . Para cada torsión existe una energía asociada, por lo tanto, si consideramos la suma de todas ellas se puede determinar la energía de torsión. A diferencia de los términos energéticos anteriores la energía de torsión no puede ser representada por un potencial armónico, sino que se debe usar una serie de Fourier mediante la siguiente

expresión:

$$E_{tor} = \sum_{tor} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)], \quad (2-24)$$

donde  $V_n$  representa la barrera de torsión de cada término de la función.  $n$  es la periodicidad del ángulo,  $\phi$  el ángulo del diedro y  $\gamma$  es la fase del ángulo.

### 2.2.2. Parámetros no enlazantes

Estos parámetros dan cuenta de las interacciones de átomos que no están unidos entre si en la misma molécula o también de las interacciones entre átomos de diferentes moléculas. Estas interacciones se pueden dividir en interacciones a corta distancia que son generadas por las nubes electrónicas que rodean a cada átomo y son principalmente de tipo repulsiva y en interacciones de largo alcance las cuales se pueden subclasificar en interacciones de dispersión e interacciones electrostáticas si no se considera la polarización de la densidad electrónica. En los campos de fuerza todas estas interacciones se describen usando dos expresiones:

- La primera de ellas se denomina término de Van de Waals y es usado para evitar la fusión de átomos no unidos con cargas opuestas como también para representar interacciones atractivas a corta distancia producto de fuerzas dispersivas. Este término puede ser representado por un potencial de Lennard-Jones el cual esta compuesto por una suma de dos contribuciones; una parte repulsiva para las distancias cortas entre los átomos y una parte atractiva para distancias grandes. La expresión que representa este termino es la siguiente:

$$E_{vdw} = \sum_i \sum_{j>i} \left( \frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^6} \right), \quad (2-25)$$

donde  $A_{ij}$  y  $B_{ij}$  son constantes que dependen de cada par de átomos y  $r_{ij}$  son

las distancias entre ellos.

- La segunda de ellas se denomina término electrostático y puede ser representado por cargas atómicas interactuantes por la ley de Coulomb. Dependiendo del campo de fuerza utilizado este término puede variar principalmente, porque cada campo de fuerza utiliza un tipo diferente de carga atómica en la ecuación de Coulomb representada por la siguiente expresión:

$$E_{elec} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \quad (2-26)$$

donde  $q_i$  y  $q_j$  representan a las cargas utilizadas por el campo de fuerza para cada tipo de átomo,  $r_{ij}$  indica la distancia que existe entre el átomo  $i$  y el átomo  $j$  y  $\epsilon_0$  es la constante dieléctrica del medio.

## 2.3. Dinámica molecular

La dinámica molecular es una técnica de simulación computacional de tipo determinista de los movimientos de todas las partículas que componen el sistema. Esta basada en las leyes de la mecánica clásica, donde se integran las ecuaciones del movimiento con el fin de generar sucesivas configuraciones, permitiendo así obtener una secuencia temporal de la evolución del sistema. Esta idea fue propuesta por Laplace en 1820, donde enunció que si se conociese la posición, la velocidad y la fuerza de todos los componentes del sistema con una precisión absoluta, se podría deducir su situación en cualquier momento pasado o futuro. Si bien la dinámica molecular no aspira completamente a estas pretensiones, si asume que a partir de un conjunto de posiciones y velocidades se puede determinar la evolución temporal del sistema con una precisión que permite estudiar propiedades termodinámicas y estructurales.

El proceso inicial de toda dinámica es definir las posiciones,  $r^N$ , y las velocidades o momentos lineales,  $p^N$ , de todas las partículas del sistema. Conociendo esta configuración inicial se puede calcular la fuerza sobre cada partícula mediante la siguiente expresión:

$$F_i(r^N) = -\nabla_i E(r^N), \quad (2-27)$$

donde  $F_i$  es la fuerza que actúa sobre la partícula  $i$  y  $E(r^N)$  es la energía potencial. Esto implica que las simulaciones de dinámica molecular requieren conocer la energía potencial del sistema para poder calcular la fuerza que actúa sobre cada átomo. Este valor puede ser obtenido mediante métodos de estructura electrónica (sección 2.1) o usando mecánica molecular (sección 2.2).

La ecuación 2-27 puede ser escrita según la segunda ley de Newton la cual permite determinar la fuerza que actúa sobre cada partícula y la aceleración que esta induce mediante la siguiente expresión:

$$-\frac{\partial E}{\partial r_i} = m_i \frac{\partial^2 r_i}{\partial t^2} = m_i \ddot{r}_i = \frac{\partial p_i}{\partial t} = \dot{p}_i, \quad (2-28)$$

donde  $\ddot{r}_i$  es la aceleración que es igual a la segunda derivada de la posición de la partícula  $i$  sobre el tiempo  $t$ .

La ecuación 2-28 requiere que las ecuaciones de movimiento sean integradas numéricamente en pequeños pasos sucesivos. En cada uno de estos pasos se calculan las fuerzas de los átomos que luego se combinan con las posiciones actuales para generar las nuevas posiciones, luego, los átomos son desplazados a estas nuevas posiciones y se vuelve a comenzar el ciclo.

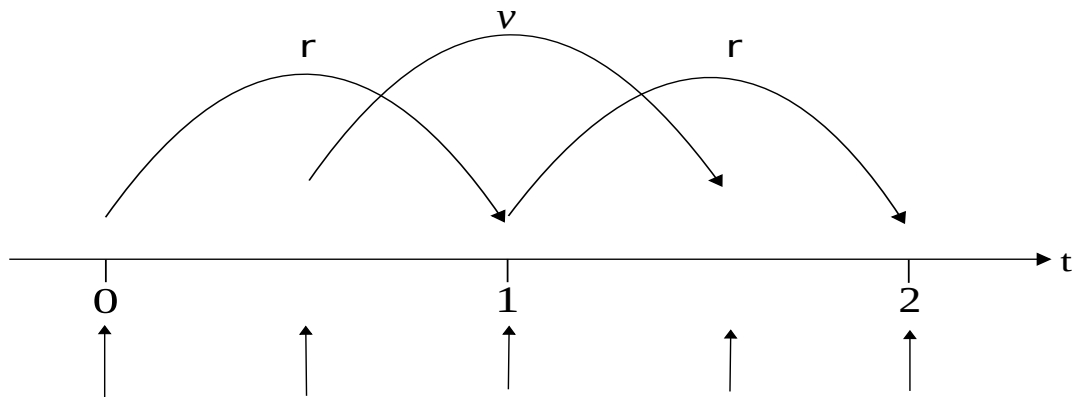
Para resolver la ecuación 2-28 se debe hacer uso de series de Taylor en las cuales se toma un intervalo de tiempo  $\Delta t$  pequeño, del orden de los femtosegundo  $fs$  y se

calculan las posiciones  $r$  y las velocidades  $v$  de las partículas que componen el sistema en diferentes instantes separados por el intervalo  $\Delta t$ . Existen diferentes métodos para resolver numéricamente las ecuaciones de Newton que han sido aproximadas a una expansión de Taylor, siendo el más usado de ellos el método leap-frog o salto de rana,<sup>72</sup> aunque además se puede usar el método de Verlet<sup>73</sup> o el método Verlet de la velocidad (velocity verlet),<sup>74</sup> entre otros.

Solo se dará una explicación más detallada del método leap-frog ya que es el que se uso en este trabajo. Este método se representa esquemáticamente en la figura 2-1 y usa las siguientes ecuaciones:

$$\begin{aligned} r(t + \Delta t) &= r(t) + v\left(t + \frac{\Delta t}{2}\right)\Delta t \\ v\left(t + \frac{\Delta t}{2}\right) &= v\left(t - \frac{\Delta t}{2}\right) + \ddot{r}(t)\Delta t \end{aligned} \quad (2-29)$$

donde la principal característica de esta metodología es que se calcula las posiciones en los instantes  $t$ ,  $t + \Delta t$ ,  $t + 2\Delta t$ , ..., y las velocidades en los instantes  $t + \frac{1}{2}\Delta t$ ,  $t + \frac{3}{2}\Delta t$ ,  $t + \frac{5}{2}\Delta t$ , etc.



**Figura 2-1.**: El método leap-frog el cual es llamado así, porque  $r$  y  $v$  van saltando como ranas uno sobre el otro. Fuente: elaboración propia.



## 2.4. Cargas atómicas

Como ya se ha mencionado existe una amplia gama de cargas atómicas, si bien es importante mencionar que las cargas atómicas no son un observable mecanocuántico, las cargas son partes del lenguaje cotidiano de la química y que pueden ser usadas en los diferentes campos de fuerzas que existen para estudiar sistemas biológicos o químicos. En todos estos campos de fuerza las cargas son utilizadas para determinar la parte electrostática del campo de fuerza representada por la ecuación de Coulomb 2-26. A continuación se discutirá un poco más en detalles los diferentes modelos de carga atómicas utilizadas en el presente trabajo desde un punto de vista teórico y práctico de como se calculan cada una de ellas.

### 2.4.1. Cargas atómicas de Hirshfeld

Las cargas atómicas de Hirshfeld se basan en el método de átomos en moléculas el cual plantea que es posible particionar la densidad electrónica para obtener como por ejemplo cargas atómicas. F. L. Hirshfeld en 1977<sup>75</sup> propuso la idea de obtener cargas atómicas a partir de la densidad electrónica de la molécula. Para obtener estas cargas hay que primero obtener la función de la densidad pro-molecular en un punto  $r$ , la cual se puede obtener mediante la siguiente expresión:

$$\rho^{pro}(r) = \sum_i \rho_i^{at}(r), \quad (2-30)$$

donde  $\rho_i^{at}(r)$  es la función de densidad del átomo aislado la cual se determina para todos los tipos de átomos que forman la molécula. Luego se define un factor de peso o contribución  $w_i(r)$  que da cuenta de cuanto aporta la densidad del pro-átomo  $\rho_i^{at}(r)$  a la densidad pro-molecular  $\rho^{pro}(r)$  mediante la siguiente expresión:

$$w_i(r) = \frac{\rho_i^{at}(r)}{\rho^{pro}(r)}, \quad (2-31)$$

Una vez obtenido este factor de peso  $w_i(r)$  se determina la densidad AIM la cual considera cuanta densidad molecular real  $\rho^{mol}(r)$  se le puede asociar a cada átomo mediante la siguiente ecuación:

$$\rho_i^{AIM}(r) = w_i(r)\rho^{mol}(r), \quad (2-32)$$

Por último para obtener la carga de Hirshfeld se usa la siguiente ecuación:

$$q_i = Z_i - \int \rho_i^{AIM}(\mathbf{r}) \, d\mathbf{r}, \quad (2-33)$$

Es conocido que las cargas de Hirshfeld presentan varios problemas como por ejemplo: las cargas son muy pequeñas y tienden a ser prácticamente cero, se usan proátomos fijos los cuales son densidades promediadas esféricamente de átomos neutros aislados, los resultados obtenidos con el método de partición de Hirshfeld dependen en gran medida de la elección de los proátomos fijos, que es esencialmente arbitrario y además estas cargas no reproducen el potencial electrostático molecular.

### 2.4.2. Cargas atómicas de Hirshfeld-Iterativas

Las deficiencias del método de Hirshfeld fue una de las motivaciones para Bultinck et al<sup>54</sup> quienes crearon un método iterativo para obtener las cargas de Hirshfeld el cual llamaron cargas atómicas de Hirshfeld-Iterativas o simplemente cargas de Hirshfeld-I (HI). Estas cargas corrigieron todos los problemas del enfoque original de Hirshfeld haciendo uso de una función de peso iterativa  $w_A^i$  la cual actualiza la

carga de cada pro-átomo hasta hacerla lo más parecida a la carga AIM.

$$w_A^i(\mathbf{r}) = \frac{\rho_A^{i-1}(\mathbf{r})}{\rho_{mol}^{i-1}(\mathbf{r})}, \quad (2-34)$$

En el esquema de Hirshfeld-I, los proátomos son interpolaciones lineales entre densidades de iones aislados, de modo que las poblaciones AIM y las poblaciones proatómicas coinciden. Las cargas de Hirshfeld-I obtenidas no son cercanas a cero, no dependen de la elección del pro-átomo, tienen una baja dependencia con la conformación utilizada y reproducen el potencial electrostático molecular en vacío.

### 2.4.3. Cargas atómicas MBIS

Las cargas MBIS del inglés Minimal Basis Iterative Stockholder fue propuesto por Toon Verstraelen<sup>55</sup> y son una variante del método de Hirshfeld. Las cargas MBIS utiliza una definición distinta para la densidad del pro-átomo  $\rho_i^0(r)$  en comparación con el método de Hirshfeld:

$$\rho_i^0(r) = \sum_{i=1}^{m_A} \rho_{Ai}^0(r), \quad (2-35)$$

con

$$\rho_{Ai}^0(r) = N_{Ai} f_{Ai} = \frac{N_{Ai}}{\sigma_{Ai}^3 8\pi} \exp\left(-\frac{|r - R_{Ai}|}{\sigma_{Ai}}\right), \quad (2-36)$$

donde  $m_A$  es el número de funciones de Slater,  $N_{Ai}$  es la población,  $\sigma_{Ai}$  es el ancho de cada capa electrónica del átomo y son parámetros variables, y  $f_{Ai}$  son funciones normalizadas de tipo Slater de la densidad. Los parámetros  $N_{Ai}$  y  $\sigma_{Ai}$  son determinados mediante un método auto consistente y la densidad pro-átomo se expande en un conjunto mínimo de funciones de densidad Slater centradas en el átomo tipo s,

cuyos parámetros se ajustan a una densidad de electrones moleculares determinada minimizando la divergencia de Kullback-Leibler (KL).

La ventaja de las cargas MBIS sobre las cargas de Hirshfeld-I es que mejoran algunas deficiencias que tienen las cargas de Hirshfeld-I. En el esquema de Hirshfeld-I, los proátomos son interpolaciones lineales entre densidades de iones aislados, de modo que las poblaciones AIM y las poblaciones proatómicas coincidan. En sistemas con átomos muy negativos ( $-2 < q_A < -1$ ), esto conduce a proátomos que se interpolan entre aniones aislados y dianiones. Sin embargo, ningún dianión atómico aislado es estable lo que impide el cálculo de su densidad electrónica. Cuando se estudian óxidos en los cuales el oxígeno puede estar como dianión o trianión se ha demostrado la escasa aplicabilidad de las cargas HI a este tipo de sistemas. El método MBIS no necesita densidades de iones (inestables) como entrada y, por lo tanto, uno esperaría que no sufra los mismos problemas que Hirshfeld-I.<sup>56-58</sup> Para considerar el efecto de disolvente sobre las cargas del soluto se calculará las cargas MBIS con el disolvente implícito SMD.

#### 2.4.4. Cargas atómicas dinámicas MBIS

Para determinar la cargas dinámicas MBIS se utilizan simulaciones de dinámica molecular y los métodos QM/MM donde el soluto es descrito mediante mecánica cuántica (QM) y el solvente se describe por mecánica molecular (MM). Así en cada paso de la simulación se realiza un cálculo de las cargas MBIS, luego se continua con la simulación hasta el siguiente paso donde nuevamente se determinan las cargas. Todo esto se repite  $n$  veces de acuerdo con el número total de pasos de la simulación, posteriormente se toma el promedio de las cargas sobre cada átomo considerando todas las cargas obtenidas durante la simulación, a este procedimiento denominamos

cargas dinámicas MBIS (D-MBIS) ya que se obtienen a partir de la simulación de dinámica molecular usando la cargas MBIS. Esta metodología para obtener cargas fue propuesta por E. Vöhringer<sup>59</sup> y deberían incluir el efecto de la polarización de la densidad electrónica de la molécula debido a las interacciones con el solvente el cual va fluctuando a medida que la simulación avanza y se espera que tengan una mejor descripción de las interacciones electrostáticas.

## 2.5. Cálculos de energía libre

Los cálculos de energía libre son importantes para estudiar diferentes propiedades como energía libre de hidratación, afinidades de unión de ligados a proteínas, coeficientes de partición, constante de asociación/disociación entre otros como se vio en la sección 1.1. Pero en general estos cálculos son difíciles de implementar correctamente, es por esta razón que en esta sección se revisaran los métodos estándares para calcular la energía libre a través de simulaciones computacionales.

La energía libre da cuenta de la estabilidad de un sistema con su entorno, ya que este valor se minimiza solo si esta en equilibrio el sistema con su alrededor. Si el sistema se encuentra en condiciones de temperatura y volumen constante se debe hablar de la energía libre de Helmholtz, y si el sistema esta a temperatura y presión constante se debe hablar de energía libre de Gibbs. En general se suele determinar la diferencia de energía libre entre dos estados diferentes como por ejemplo la energía libre de hidratación es la diferencia de energía de una molécula en fase acuosa y en el vacío, el coeficiente de partición es la diferencia entre la energía de hidratación y la energía de solvatación en algún solvente orgánico.

A continuación se detallaran los métodos que se utilizan para calcular la energía libre en aplicaciones ya sea químicas o biológicas, como también se detallaran los

pasos que se deben seguir para realizar un cálculo típico de energía libre, como realizar la simulación y el análisis de resultados.

### 2.5.1. Método de Zwanzig

El método de Zwanzig<sup>76</sup> es el más conocido para calcular la diferencia de energía libre en simulaciones computacionales. Muchas veces se le llama a este método perturbación de energía libre o promediación exponencial. Mediante este método la energía libre entre dos potenciales  $U_0(\vec{q})$  y  $U_1(\vec{q})$  sobre un espacio de coordenadas y momento  $\vec{q}$  puede ser calculada como:

$$\Delta A = -\beta^{-1} \ln \langle e^{-\beta(U_1(\vec{q}) - U_0(\vec{q}))} \rangle_0 = -\beta^{-1} \ln \langle e^{-\beta \Delta U(\vec{q})} \rangle_0, \quad (2-37)$$

donde  $\Delta U(\vec{q})$  es la diferencia entre la energía interna del estado 1 y 0,  $\Delta A$  es la diferencia de energía libre de Helmholtz entre estos dos estados si se considera un colectivo NVT,  $\beta = k_B T^{-1}$ , donde  $k_B$  es la constante de Boltzmann, T es la temperatura absoluta y el símbolo  $\langle \rangle_0$  indica un promedio sobre el colectivo del estado 0. Esta ecuación es exacta si el volumen de fase de ambos estados es el mismo. Se ha demostrado que funciona bien cuando la perturbación de 1 a 0 es pequeña, pero si el cambio es muy grande esta expresión converge muy pobremente en función del número de muestras recolectadas lo que implica que hay una muy baja superposición entre ambos estados lo que lleva a obtener valores muy pequeños de energía libre o valores errados.

### 2.5.2. Integración termodinámica

Una manera de solucionar los problemas presentados por el método de Zwanzig es usando el método de integración termodinámica. La integración termodinámica consiste en introducir un parámetro  $\lambda$  que permite interpolar entre los dos estados  $i$  y  $f$ . Por lo tanto, la diferencia de energía libre puede ser escrita mediante la siguiente expresión:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (2-38)$$

donde  $\Delta A$  es el cambio de energía libre entre el estado B ( $\lambda = 1$ ) y A ( $\lambda = 0$ ),  $U$  es la energía interna del sistema y  $\langle \rangle_{\lambda}$  indica un promedio sobre el colectivo basado en un valor del parámetro  $\lambda$ . La ventaja de esta metodología es que las perturbaciones pueden ser ajustadas por el parámetro  $\lambda$  llevando a una mejor convergencia de los valores de energía libre. Esto permite estudiar cambios significativos dentro de un sistema como es la desolvatación de moléculas pequeñas.

### 2.5.3. Bennett Acceptance Ratio (BAR)

La diferencia de la energía potencial en ambas direcciones entre dos estados puede ser usada para estimar la energía libre proporcionando una mejora sustancial si lo comparamos con los valores obtenidos por el método de Zwanzig. Esta diferencia de energía potencial para la misma configuración  $\vec{q}$  de dos estados diferentes a lo largo del camino es  $\Delta U_{ij}(\vec{q})$ . Entre las distribuciones de las diferencias de energía potencial  $\Delta U_{ji}(\vec{q})$  de las configuraciones muestreadas del estado  $i$  y  $\Delta U_{ij}(\vec{q})$  de las configuraciones muestreadas para el estado  $j$  existe una relación directa.<sup>77</sup> Esto implica, que existe un procedimiento estadísticamente óptimo que utiliza la información de ambos estados  $i$  y  $j$  para obtener una estimación de la energía libre mejorada de la

diferencia de los dos estados. La primera persona en determinar este procedimiento fue el físico Charles H. Bennett y en honor a él se le da el nombre a este método Bennett Acceptance Ratio BAR. Bennett proporciono la siguiente relación para la diferencia de energía libre:

$$\Delta A_{ij} = -\ln k_B T \frac{\Xi_i}{\Xi_j} = k_B T \ln \frac{\langle \alpha(\vec{q}) e^{[-\beta \Delta U_{ij}(\vec{q})]} \rangle_j}{\langle \alpha(\vec{q}) e^{[-\beta \Delta U_{ji}(\vec{q})]} \rangle_i} \quad (2-39)$$

Esta expresión es válida para cualquier función que  $\alpha(\vec{q}) > 0$  para todas las configuraciones  $\vec{q}$ . Bennett usó el cálculo variacional para encontrar el cambio de  $\alpha(\vec{q})$  minimizando la varianza de energía libre,<sup>78</sup> obteniendo una función implícita de  $\Delta A$  que se puede resolver fácilmente numéricamente:

$$\sum_{i=1}^{n_i} \frac{1}{1 + e^{(\ln(n_i/n_j) + \beta \Delta U_{ij} - \beta \Delta G)}} - \sum_{i=1}^{n_i} \frac{1}{1 + e^{(\ln(n_j/n_i) - \beta \Delta U_{ji} + \beta \Delta G)}} = 0 \quad (2-40)$$

donde  $n_i$  y  $n_j$  son los números de configuraciones muestreadas para cada estado. Se ha demostrado que este método es uno de los mejores que existe para obtener energía libre considerando dos estados.<sup>21</sup> Además, existen otros estudios<sup>79,80</sup> que demuestran que esta metodología es superior tanto teóricamente como numéricamente sobre el método de Zwanzig. Hacer una comparación entre la integración termodinámica y el método BAR no es correcto ya que ambos tienen bases teóricas diferentes al igual que la información que usan para estimar la energía libre, pero si es conocido que BAR tiene un mejor desempeño que el método de integración termodinámica. Esto se debe principalmente a que se requieren menos estados intermedios para BAR en comparación con IT para obtener un nivel de precisión estadística similar.



### 2.5.4. Multistate Bennett Acceptance Ratio (MBAR)

El método de MBAR es una extensión del método BAR ya que permite evaluar los datos de todos los estados y no solo de dos estados como en el caso de BAR y así predecir de mejor manera la energía libre. Esta metodología proporciona la más baja varianza en comparación con todos los métodos descritos anteriormente. En su derivación se considera un conjunto de  $K \times K$  funciones de ponderación y se busca un  $\alpha_{i,j}(\vec{q})$  que minimice la varianza durante la reponderación. Para elegir el  $\alpha_{i,j}$  óptimo se puede usar la siguiente expresión:

$$\alpha_{ij} = \frac{N_j \hat{c}_j^{-1}}{\sum_{k=1}^K N_k \hat{c}_k^{-1} \exp(-\beta U_k)} \quad (2-41)$$

Después de hacer todas las sustituciones necesarias, se puede obtener una expresión para una energía libre:

$$\hat{A}_i = -\beta^{-1} \ln \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{\exp[-\beta U_i]}{\sum_{k=1}^K N_k \exp[\beta \hat{A}_k - \beta U_k]} \quad (2-42)$$

donde  $i$  va de 1 a  $K$ , las  $A_i$  son las energías libres de cada estado,  $q_{kn}^{\vec{}}$  son las  $n$  muestra de los  $k$  estados, y  $U_i$  son los potenciales de estos  $K$  estados. Esta metodología considera todas la muestras tomadas para cada  $i$  y además proporciona las incertidumbres estadísticas de las energías libres calculadas.

Las metodologías descritas anteriormente son metodologías de equilibrio. Existe una metodología de no equilibrio donde la energía libre está relacionada con el trabajo de no equilibrio a través de la igualdad de Jarzynski's.<sup>81,82</sup> Estas metodologías presentan una alternativa para calcular la diferencia de energía libre.

### 2.5.5. Ciclo termodinámico

Para realizar el cálculo de la energía libre de hidratación se puede calcular la diferencia de energía libre entre los estados 1 y 2 de la Fig. 2-2. Esto no es posible para sistemas biomoleculares en escalas de tiempo razonables. Para determinar de manera precisa la diferencia de energía libre entre dos estados, sus integrales de espacio de fase deben tener solapamiento suficiente que en la práctica es posible solo cuando ambos estados son extremadamente similares. Cuando esto no es el caso, puede ser imposible calcular directamente la diferencia de energía libre entre estados finales 1 y 2. En tales casos, se puede calcular las diferencias de energía libre entre una serie de estados intermedios lo que permite tener suficiente superposición. Para pasar desde estado 1 al estado 2 se consideran estados intermedios que son estados típicamente artificiales, no físicos construidos para unir los estados físicos de interés. Estos estados forman parte de lo que se conoce como un ciclo termodinámico (ver 2-2 ) que une los dos estados finales de interés.

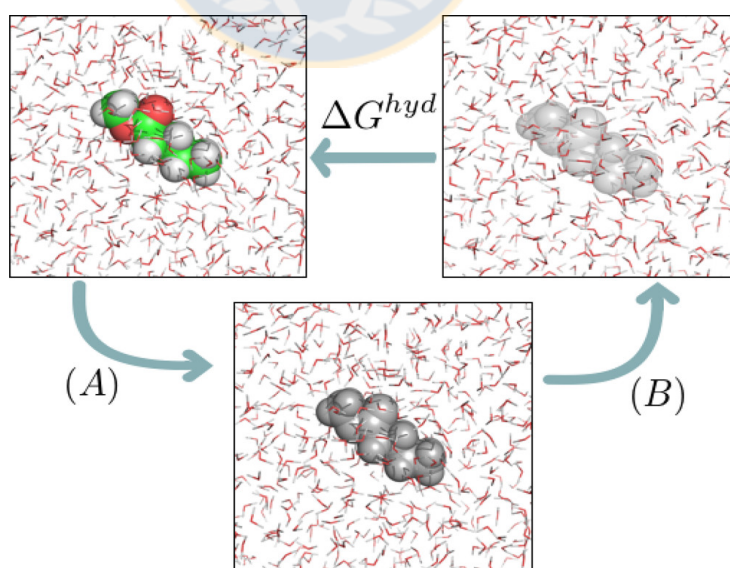


Figura 2-2.: Esquema típico de un ciclo termodinámico.<sup>5</sup>

El estado 1 de la Fig. **2-2** representa a un sistema totalmente interactuante en el que las moléculas de disolvente interactúan con la molécula de soluto a través de fuerzas tanto electrostáticas como de van der Waals, el estado 2 representa una molécula de soluto que interactúa con su entorno a través de fuerzas no polares (dispersión y repulsión) y el estado 3 es una sola molécula de soluto que no interactúa en una caja de disolvente.

Para un ciclo termodinámico la energía libre de hidratación se puede determinar mediante la siguiente expresión:

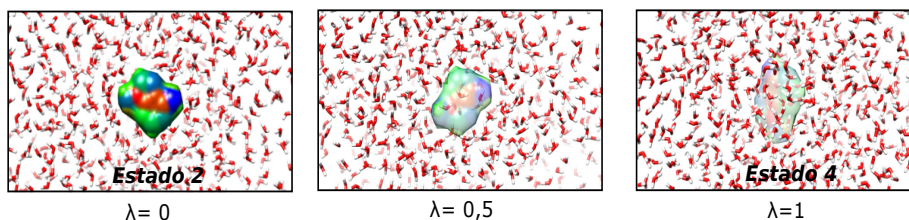
$$\Delta G_{hid} = -\Delta G_A - \Delta G_B \quad (2-43)$$

donde  $\Delta G_A$  es la energía libre que se obtiene al apagar las interacciones electrostáticas y  $\Delta G_B$  es la energía libre que se obtiene al apagar las interacciones de van der Waals.

Estas transformaciones se realizan en una serie de pasos mediante el método de integración termodinámica introduciendo un parámetro  $\lambda$  que modula la energía potencial del sistema, de modo que a medida que  $\lambda$  va de 0 a 1 se determinan las posibles transiciones de energía entre el estado inicial y el estado final de destino. Las simulaciones se ejecutan a continuación con un conjunto de diferentes valores  $\lambda$  que conectan los dos estados Figura **2-3**. En otras palabras, cada una de las dos vías de transformación se subdivide en una variedad de pasos individuales, donde cada paso implica una transición entre dos valores  $\lambda$ . El número y el espaciado de los valores de  $\lambda$  se elige para asegurar solapamiento suficiente entre los espacios conformacionales de los dos estados que se están considerando.

Estos estados intermedios, y sus correspondientes valores de  $\lambda$ , normalmente se dice que son alquímicos, ya que corresponden a estados no físicos, a menudo con un

cambio en la identidad química de las especies consideradas.



**Figura 2-3.:** Cambios que ocurren a medida que vamos variando el valor de  $\lambda$ . Fuente: elaboración propia.

## 2.6. Corrección de polarización a la energía libre de hidratación

La hidratación de una molécula incluye una polarización de la densidad electrónica del soluto. Cuando las cargas atómicas utilizadas en los cálculos de la energía libre se calculan a partir de la densidad electrónica del soluto en vacío, esta contribución de polarización no es considerada y se asume que las densidades electrónicas en la fase gaseosa y en solución son las mismas. Mediante el uso de cargas atómicas derivadas de cálculos de estructura electrónica con un modelo de solvatación implícita, se puede modelar la interacción con el solvente debido a la polarización del soluto. Pero, en el cálculo de energía libre basado en campos de fuerza no polarizables, la energía requerida para polarizar la densidad electrónica del soluto,  $E_{\text{pol},s}$  tiene que ser contabilizado:

$$E_{\text{pol},s} = \langle \Psi_{\text{solv}} | \hat{H}_{\text{vac}} | \Psi_{\text{solv}} \rangle - \langle \Psi_{\text{vac}} | \hat{H}_{\text{vac}} | \Psi_{\text{vac}} \rangle \quad (2-44)$$

El primer termino es el valor esperado del hamiltoniano del cálculo en vacío para la función de onda obtenida con el modelo de solvatación implícita y el segundo es

la energía del estado fundamental del hamiltoniano en vacío. Este enfoque polariza el soluto a través de un campo de reacción resultante del disolvente polarizado representado por un dialéctico que descuida la polarización específica por moléculas de disolvente debido, por ejemplo, a la formación de enlaces de hidrógeno.

## 2.7. Cálculo de las cargas atómicas y energía libre de hidratación

Se calcularon las energías libres de hidratación para 613 moléculas de la base de datos FreeSolv (versión 0.51) usando el campo de fuerza GAFF<sup>7</sup> y variando el parámetro de carga atómica que utiliza este campo de fuerza. Para esto se derivaron las cargas atómicas desde la densidad electrónica del soluto en vacío y utilizando el disolvente continuo SMD.<sup>83,84</sup> Las optimizaciones de la geometría y las densidades electrónicas se calcularon a nivel de teoría BLYP<sup>69,70,85</sup> y B3LYP<sup>71</sup> con la base def2-TZVP,<sup>86,87</sup> usando ORCA 3.0.2.<sup>88</sup> La mayoría de los resultados se obtuvieron con el método BLYP, mientras que los resultados de B3LYP solo se utilizaron para evaluar la influencia del intercambio exacto en las cargas parciales y las energías libres de hidratación.

Las cargas atómicas se obtuvieron a partir de la densidad electrónica con los métodos Hirshfeld-I en vacío y usando el disolvente implícito SMD a las cuales denominamos HI y S-HI respectivamente, y también se utilizó el método Minimal Basis Iterative Stockholder (MBIS) con el disolvente implícito SMD a las de ahora en adelante las llamaremos S-MBIS. Todas estas cargas se determinaron usando el programa HORTON 2.0.0.<sup>89</sup> y para un grupo seleccionado de moléculas se determinaron las cargas dinámicas MBIS (D-MBIS).

Después del cálculo de la carga atómica con cada método descrito en la sección 2.4, los átomos químicamente equivalentes, que resultan de la simetría molecular o de rotaciones libres, como por ejemplo los átomos de hidrógeno en el grupo metilo, se identificaron con el programa OpenEye python Toolkit (versión 2017.2.1) y sus cargas fueron promediadas.

La energía libre de hidratación para cada molécula se obtuvo mediante cálculo alquímicos de energía libre realizados con protocolos estándares que mantienen los errores de muestreo relativamente pequeños. Cada molécula se solvató con aproximadamente 1000 moléculas de agua y se minimizó la energía con el paquete de simulación GROMACS.<sup>90</sup> Las moléculas de agua fueron descritas por el modelo SPC/E,<sup>91</sup> que reproduce las propiedades líquidas del agua razonablemente bien manteniendo un costo computacional bajo. En el caso de las cargas S-HI, también se utilizó el modelo de agua TIP3P,<sup>92</sup> que se utilizó originalmente para calcular las energías libres de hidratación en la base de datos FreeSolv (con cargas AM1-BCC), para probar la dependencia de los resultados con el modelo de agua.

Después de minimizar la energía, todo el sistema se equilibró con el conjunto NVT, y luego con NPT a 298,15 K y 1 bar. Las simulaciones se realizaron con el programa GROMACS 5.0.4<sup>90</sup> usando una paso de tiempo de  $2fs$  en combinación con dinámicas estocásticas<sup>93</sup> ( $\tau = 2ps$ ) y el algoritmo de acoplamiento de presión Parrinello-Rahman<sup>94</sup> ( $\tau_p = 1ps$ ) que usa la compresibilidad del agua. Las interacciones electrostáticas se calcularon con el método Particle-Mesh-Ewald,<sup>95</sup> con un radio de corte de 1.2 , un orden PME de 6 y un espaciado de 0.1 nm. Las interacciones de van der Waals fueron escaladas a cero mediante una función de conmutación, que cambia el potencial a cero entre 1.0 y 1.2 . La lista de vecinos se actualizó cada 10 pasos con el esquema de corte de verlet implementado en GROMACS 5.0.4<sup>90</sup> y su

punto de corte se estableció en 1.2 . Todos los enlaces se restringieron con el algoritmo LINCS<sup>96</sup> de orden 4 y se aplicó la corrección isotrópica a la presión de energía debido a la falta de interacciones van-der-Waals.<sup>97</sup>

La energía libre de hidratación de Gibbs se calculó a partir de los sistemas solvatados equilibrados con simulaciones de dinámica molecular alquímica, como se describió anteriormente, donde las interacciones intermoleculares entre cada soluto y el disolvente se desconectaron mediante un parámetro  $\lambda$  usando potenciales de núcleo blando con valores de  $\sigma = 0,3$  y  $\alpha = 0,5$  y  $p = 1$  originalmente propuesto por Beutler *et al.*<sup>98</sup> e implementado en el paquete GROMACS 5.0.4. Primero, las interacciones electrostáticas se desacoplaron usando los valores del parámetro  $\lambda$  [0.00, 0.25, 0.50, 0.75, 1.00] seguidos por las interacciones van-der-Waals a valores de [0.00, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00].<sup>5</sup> El tiempo total de simulación por valor de  $\lambda$  fue 1 . El análisis de las simulaciones de energía libre se realizó con diferentes estimadores de energía libre implementados en la herramienta de análisis alquímico<sup>99</sup> y el método MBAR<sup>100</sup> se usó para las estimaciones de energía libre final; sus resultados fueron en general consistentes con los de otros estimadores.

Además, la energía libre de hidratación de Gibbs también se calculó utilizando un enfoque mecánico puramente cuántico junto con el modelo de solvatación SMD y el paquete ORCA 3.0.2<sup>88</sup> se consideró la diferencia de energía libre de cada molécula calculada en vacío junto con las energías libres del solvente implícito en la aproximación del rotor rígido y el oscilador armónico incluyendo la corrección de la energía del punto cero suponiendo el estado estándar de  $1 \text{ mol L}^{-1}$  como en la base de datos.

## 2.8. Entalpía y entropía de hidratación

Otra medida para comprobar algún parámetro distinto dentro de un campo de fuerza es determinar la entalpía de hidratación debido a que esta variable es muy sensible y en general no tiene una compensación por efecto entrópicos como aveces sucede con la energía libre.<sup>16,101,102</sup> Esto implica que se pueden obtener valores de energía libre de hidratación cercanos a los experimentales, pero no necesariamente se puede obtener un valor cercano para la entalpía y entropía de hidratación. Para calcular la entalpía de hidratación ( $\Delta H_{hid}$ ), se utilizó la diferencia de las energías potenciales entre una caja de agua solvatada, una caja sin soluto con la misma cantidad de moléculas de agua y una caja con el soluto en vacío:

$$\Delta H_{hid} = \langle U_{disolución} \rangle - (\langle U_{agua} \rangle + \langle U_{vacío} \rangle) \quad (2-45)$$

donde  $U_{disolución}$  es la energía interna de la disolución que contiene el soluto,  $U_{agua}$  es la energía interna de la caja que contiene el mismo número de moléculas de agua sin el soluto y  $U_{vacío}$  es la energía interna de la molécula de soluto en vacío. Cabe destacar que en todo este enfoque se ha despreciado la contribución a la entalpía del término  $P\Delta V$ , ya que para los solutos de tamaño pequeño esta contribución es mucho menor que la incertidumbre obtenida en este tipo de cálculos.<sup>16</sup> Si se consideran solutos de mayor tamaño a los estudiados en este trabajo o presiones más altas se debería considerar el termino  $P\Delta V$  ya que en esas condiciones este valor es más importante. Además, no se incluyen correcciones a los modos de vibración debido a la naturaleza cuántica de los solutos reales y las simulaciones del liquido se hicieron a la presión atmosférica y no la presión de vapor de la fase gaseosa. En general estas contribuciones se suelen omitir, pero si se desea poder comparar



correctamente con los valores experimentales se deberían considerar.

La entropía de hidratación se puede calcular mediante la siguiente ecuación:

$$\Delta S_{hid} = \frac{\Delta H_{hid} - \Delta G_{hid}}{T} \quad (2-46)$$

donde  $\Delta H_{hid}$  y  $\Delta G_{hid}$  se calculan como se describió anteriormente.

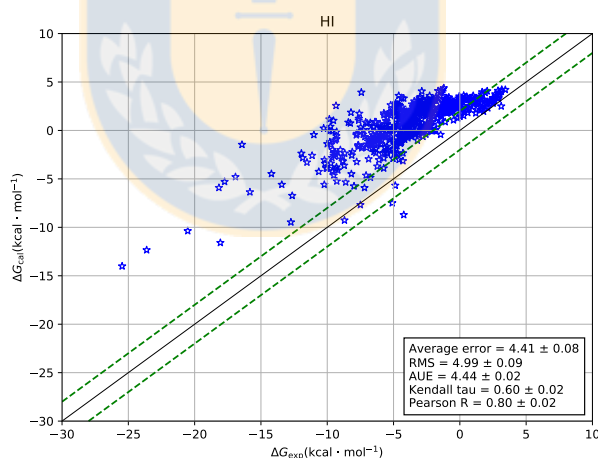


## 3. Resultados

La capacidad de las diferentes cargas atómicas en combinación con los otros parámetros del campo de fuerza GAFF, para reproducir las energías libres de hidratación experimental de la base de datos FreeSolv, fue abordada mediante cálculos de energía libre de última generación y simulaciones de dinámica molecular en solvente explícito. Varios factores fueron examinados, incluyendo el efecto que ejerce el solvente sobre la polarización de la densidad electrónica en los cálculos de estructura electrónica, la corrección de polarización a la energía libre, la influencia del método de estructura electrónica en las cargas atómicas y su efecto sobre las energías libres de hidratación, y finalmente la influencia del modelo de agua en los cálculos de energía libre. Las energías libres de hidratación fueron calculadas con cargas atómicas obtenidas por diferentes métodos de partición de la densidad electrónica como: Hirshfeld-I y Minimal Basis Iterative Stockholder (MBIS) (ver la sección Métodos) y se compararon con los valores obtenidos con las cargas atómicas AM1-BCC. Para comparar, también se obtuvieron energías libres de hidratación a partir de los cálculos de la estructura electrónica con el modelo de solvatación implícito de SMD.

### 3.1. Energía libre de hidratación usando diferentes métodos de derivación de carga

Las cargas atómicas utilizadas para calcular las energías libres de hidratación provienen de diferentes métodos de partición de la densidad electrónica, en los que la polarización de la densidad se consideró mediante el modelo de solvatación SMD. Las energías libres finales se corrigieron con la energía requerida para polarizar la densidad electrónica. Solo en el caso de las cargas atómicas semipolarizadas basadas en la aproximación de respuesta lineal y el método AM1-BCC no se incluyó la corrección de la polarización porque se descuida en este último o se tiene en cuenta en la aproximación del primero.



**Figura 3-1.:** Gráfico de los valores de energías libres de hidratación calculadas versus los valores experimentales para todas las moléculas de FreeSolv consideradas, usando el campo de fuerza GAFF y las cargas atómicas obtenidas del método de partición Hirshfeld-I de la densidad de electrones moleculares BLYP / def2-TZVP en vacío (El recuadro muestra el resultado del análisis de error: error cuadrático medio (RMS), error absoluto sin signo (AUE), Kendall tau y el coeficiente de correlación Pearson R para cada conjunto de cargas.). Fuente: elaboración propia.

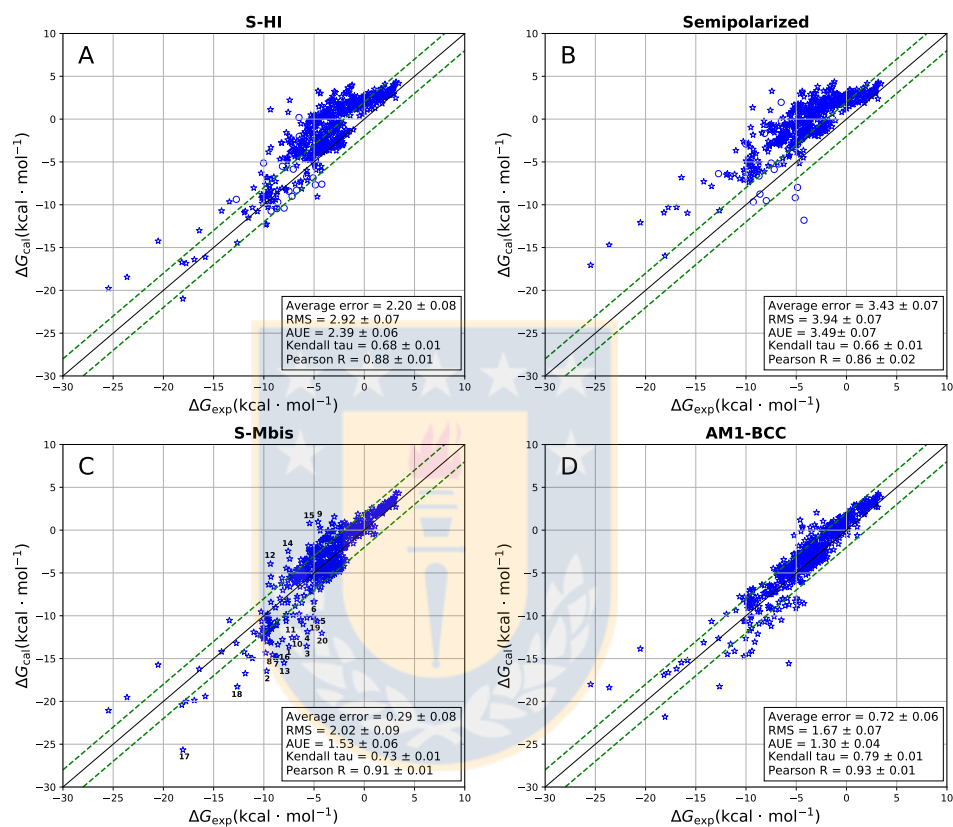
Primero se estudió el efecto de la polarización en las energías libres de hidratación

y las cargas atómicas de Hirshfeld-I calculadas a partir de la densidad electrónica molecular en el vacío. La figura **3-1** muestra el gráfico que compara las energías libres de hidratación calculadas y los valores de referencia experimentales en la base de datos FreeSolv. Este gráfico muestra una pobre correlación y un valor grande para el error cuadrático medio (RMSE) de  $5.0 \text{ kcal mol}^{-1}$ .

La figura **3-2** muestra las gráficas de todas las energías libres de hidratación calculadas con cargas atómicas que incluyen la polarización del soluto y las cargas AM1-BCC contra el valor experimental informado en la base de datos FreeSolv. En la figura **3-2 A**, se aplicó el método de partición de Hirshfeld-I a las densidades electrónicas moleculares polarizadas por el modelo de solvatación SMD usando cálculos de estructura electrónica a nivel BLYP / def2-TZVP (cargas atómicas S-HI) y se corrigió mediante la energía de polarización (para sulfatos y fosfatos se utilizó el método Hirshfeld-E y se muestran como círculos). En comparación con la Figura **3-1** se concluye que el efecto de la polarización de la densidad electrónica a través del solvente antes de la partición en cargas atómicas es esencial para mejorar las energías libres de hidratación calculadas. Agregar el modelo de solvatación implícita en el cálculo de la estructura electrónica, polariza la densidad electrónica y conduce a valores absolutos más grandes de las cargas atómicas y a energías libres más negativas, que después de la corrección de la polarización revelan mejor correlación con los valores experimentales como se muestra en la figura **3-2 A** (RMSE =  $2.9 \text{ kcal mol}^{-1}$ ). La polarización de la densidad electrónica y la energía de corrección de la polarización se vuelven más importantes para las moléculas polares, que tienen las energías libres de hidratación más negativas y las interacciones electrostáticas más fuertes con el disolvente. A pesar del éxito de usar cargas derivadas de densidades polarizadas y la corrección de polarización, la mayoría de las energías libres de hidra-

tación calculadas (Figura **3-2** A) todavía sobre estiman el valor experimental (error promedio = 2.2 kcal mol<sup>-1</sup>). Este error sistemático será analizado en detalle en los siguientes párrafos.

Para probar la sensibilidad de las cargas atómicas de Hirshfeld-I al método de estructura electrónica empleado, también se calculó las densidades electrónicas con el método computacionalmente más exigente B3LYP / def2-TZVP y el modelo de solvente implícito SMD. Con las cargas atómicas resultantes, las energías libres de hidratación se volvieron a calcular para toda la base de datos aplicando las correcciones de polarización respectivas. El cambio en el método de estructura electrónica tiene solo un efecto pequeño para moléculas con energías libres de hidratación superiores a -10 kcal mol<sup>-1</sup> y para moléculas más polares proporciona valores más negativos como se muestra en la figura **3-3**. Esto está relacionado con el error de autointeracción de los funcionales de correlación de intercambio semi-locales como BLYP, que da como resultado densidades electrónicas más homogéneas. Se ha demostrado que este error proporciona momentos dipolares más pequeños,<sup>103</sup> que también es el caso en este trabajo. Además, se observa que las cargas atómicas derivadas de las densidades BLYP son más pequeñas en comparación con sus equivalentes B3LYP. Sin embargo, la desviación absoluta sin signo de las energías libres de hidratación calculadas (método B3LYP) con respecto a las obtenidas con las cargas atómicas BLYP es de 0.7 kcal mol<sup>-1</sup> con las mayores desviaciones en el rango de 2-3 kcal mol<sup>-1</sup> (RMSE = 3.0 kcal mol<sup>-1</sup> con respecto a los valores de referencia experimentales). Para las moléculas que contienen sulfatos y fosfatos, cuyas cargas atómicas se calcularon con el método alternativo Hirshfeld-E, se observó una mayor dependencia de los métodos de estructura electrónica. Teniendo en cuenta el costo computacional del funcional híbrido B3LYP (3 veces mayor que BLYP) y el valor RMSE aumentado

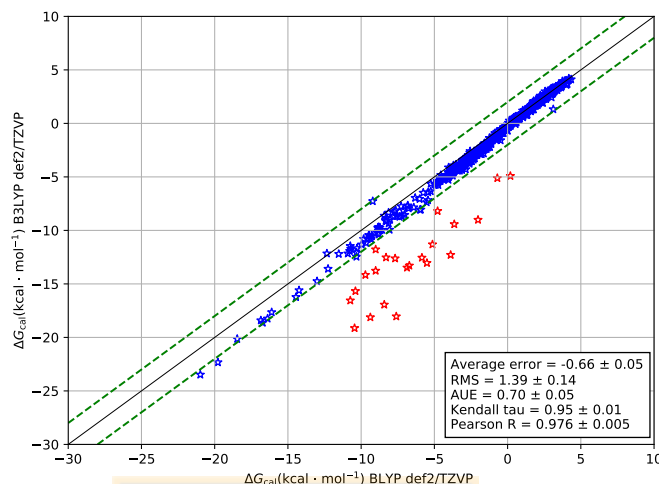


**Figura 3-2:** Gráfico de las energías libres de hidratación calculadas versus experimentales para todas las moléculas de FreeSolv consideradas, utilizando el campo de fuerza GAFF y diferentes métodos para obtener cargas atómicas. (A) Método de estructura electrónica BLYP / def2-TZVP con modelo de solvatación SMD y método de partición Hirshfeld-I añadiendo la corrección de polarización a la energía libre; (B) el mismo método de estructura electrónica con las cargas semipolarizadas sin aplicar ninguna corrección de polarización; (C) el mismo método de estructura electrónica y el modelo de solvatación de SMD, pero usando el particionamiento iterativo a base mínima (MBIS) corrigiendo las energías libres de transferencia para la polarización y (D) usando cargas de átomo AM1-BCC en FreeSolv. Fuente: elaboración propia.

con respecto a los datos de referencia experimentales, BLYP en combinación con el conjunto base empleado presenta una alternativa válida para el cálculo de un gran número de moléculas como el realizado en este estudio.

Para todos los cálculos, se consideró el modelo de agua SPC / E<sup>91</sup> debido a su mejor reproducción de las propiedades del agua líquida y al bajo costo computacional relacionado con solo tres sitios de interacción. Debido a que las energías libres de hidratación también dependen del modelo de agua utilizado en las simulaciones de dinámica molecular, se probó el efecto de cambiar el modelo de agua SPC / E por el modelo de TIP3P,<sup>92</sup> el cual se utilizó previamente en combinación con AM1 -BCC para determinar la energía libre de hidratación en la base de datos FreeSolv.<sup>5</sup> En trabajos anteriores, se seleccionó el modelo TIP3P porque a menudo se recomienda para los campos de fuerza AMBER y GAFF. La comparación de las energías libres de hidratación con los dos modelos de agua revela una desviación media cuadrática de 0.56 kcal mol<sup>-1</sup> (RMSE (modelo TIP3P) = 2.6 kcal mol<sup>-1</sup> con respecto a los valores de referencia). Por lo tanto, el modelo SPC / E fue considerado para el resto del estudio debido a su mejor reproducción de las propiedades del agua líquida al mismo costo computacional.

La figura **3-2 B** muestra los resultados de las cargas semipolarizadas de Hirshfeld-I, que es un método alternativo basado en la teoría de respuesta lineal para explicar la polarización del soluto y similar al método de carga polarizada implícitamente (IPolQ) de Cerutti. *et al.*<sup>104</sup> (ver la sección de Métodos). En lugar de corregir la energía libre de hidratación con una energía de polarización, aquí las cargas atómicas se alteran sumándole a las cargas en vacío, la mitad de la diferencia entre las cargas usando el disolvente SMD y las cargas en vacío, esto se hace para dar cuenta de la polarización. Las cargas atómicas semipolarizadas conducen a errores relativamente



**Figura 3-3.:** Gráfico de las energías libres de hidratación obtenidas con cargas atómicas S-HI (azul) y las cargas atómicas S-HE (rojo) a partir de la densidad electrónica con el modelo de solvatación SMD y dos métodos diferentes de estructura electrónica: BLYP / def2-TZVP y B3LYP / def2-TZVP. Fuente: elaboración propia.

grandes y en la mayoría de los casos se sobre estima la energía libre de hidratación, lo que se refleja en el error promedio y otras descripciones estadísticas que se muestran en el recuadro. Esto no es sorprendente porque las cargas semipolarizadas se basan en dos aproximaciones principales: (i) solo se incluyen los efectos de respuesta lineal y (ii) la expansión multipolar distribuida de las fluctuaciones de densidad inducida se trunca después de los monopolos atómicos. En contraste, la corrección de polarización en Ec. 2-44 hace menos aproximaciones: se incluyen efectos no lineales y no hay truncamiento de la expansión multipolar distribuida.

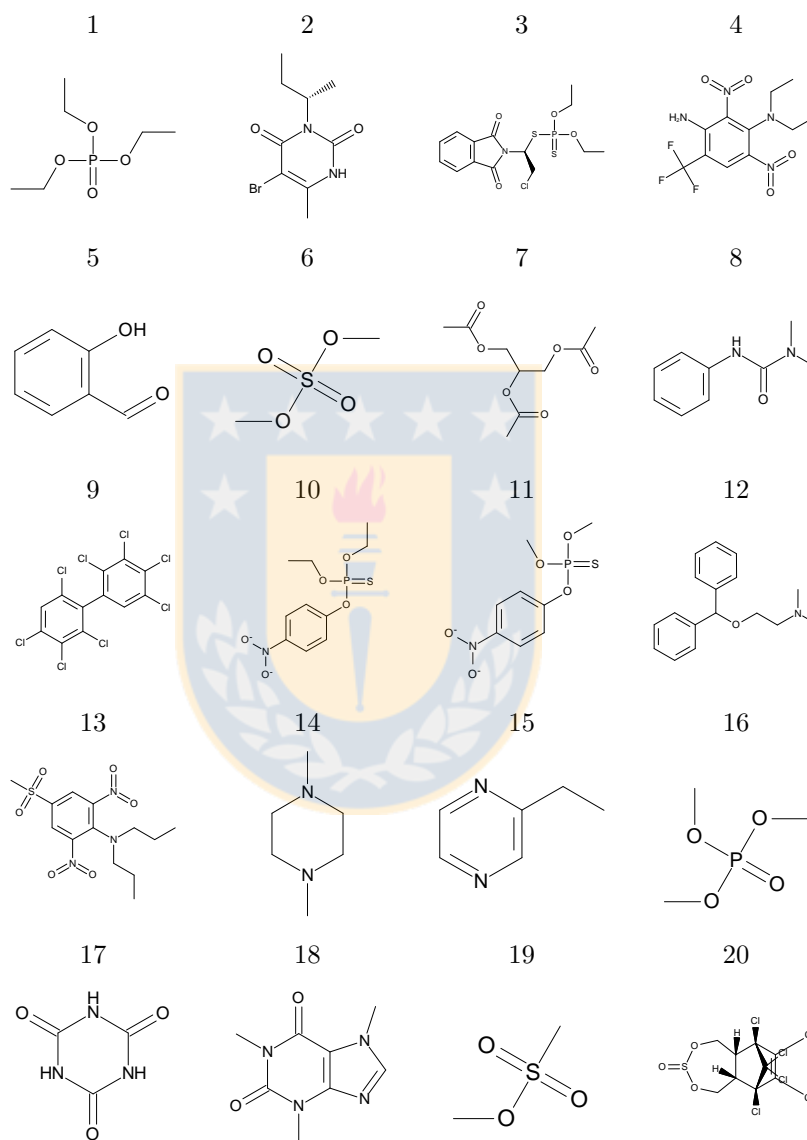
Finalmente, la Figura 3-2 C muestra las energías libres de hidratación calculadas con cargas atómicas obtenidas del método de división MBIS empleando el método de estructura electrónica BLYP / def2-TZVP, el modelo de solvatación SMD y la corrección de polarización de las energías libres. Cuando las energías libres de hidratación de este conjunto de cargas se comparan con el método de partición Hirshfeld-I en la



figura **3-2 A**, se observa una mejor correlación general con los valores experimentales, como lo demuestran los descriptores estadísticos en el recuadro. Para comparar, se agregó los resultados obtenidos con las cargas AM1-BCC presentes en la base de datos FreeSolv en la figura **3-2 D**.

En resumen, de los tres nuevos conjuntos de carga analizados (S-HI, semipolarizadas y S-MBIS), las cargas S-MBIS reproducen claramente las energías libres de hidratación mejor con un rendimiento comparable a las cargas AM1-BCC. Como se muestra en la Figura **3-1**, las cargas atómicas calculadas en vacío sin polarización por el solvente conducen a una predicción muy pobre de la energía libre de hidratación. Las cargas semipolarizadas producen mejores resultados, pero la mejora es limitada. El uso de cargas Hirshfeld-I o MBIS a partir de una densidad electrónica polarizada y la adición de la corrección de polarización mejora considerablemente la correlación con los valores experimentales. Comparando los descriptores estadísticos de las cargas S-MBIS con las cargas atómicas AM1-BCC concluimos que las cargas atómicas S-MBIS presentan errores promedio más pequeños ( $0.29$  vs  $0.79$  kcal mol<sup>-1</sup>) con valores ligeramente mayores para el error de la raíz cuadrática media (RMSE) y el error absoluto sin signo (AUE). La diferencia en los coeficientes de correlación de Kendall  $\tau$  y Pearson  $R$  es pequeña y se encuentra en el margen de error.

Estas descripciones estadísticas en la figura **3-2** se usan a menudo para validar el rendimiento de un determinado modelo; sin embargo, también es interesante analizar el número de moléculas cuyas energías libres de hidratación calculadas se encuentran dentro del rango del error experimental. Para las cargas atómicas de S-MBIS, se encontró que esto equivale a 23 %, que es igual a la fracción de moléculas para las cargas de AM1-BCC. Tomando el otro extremo, se identificó las moléculas que tenían un error de más de cinco veces el error experimental y se muestra su estructura en la

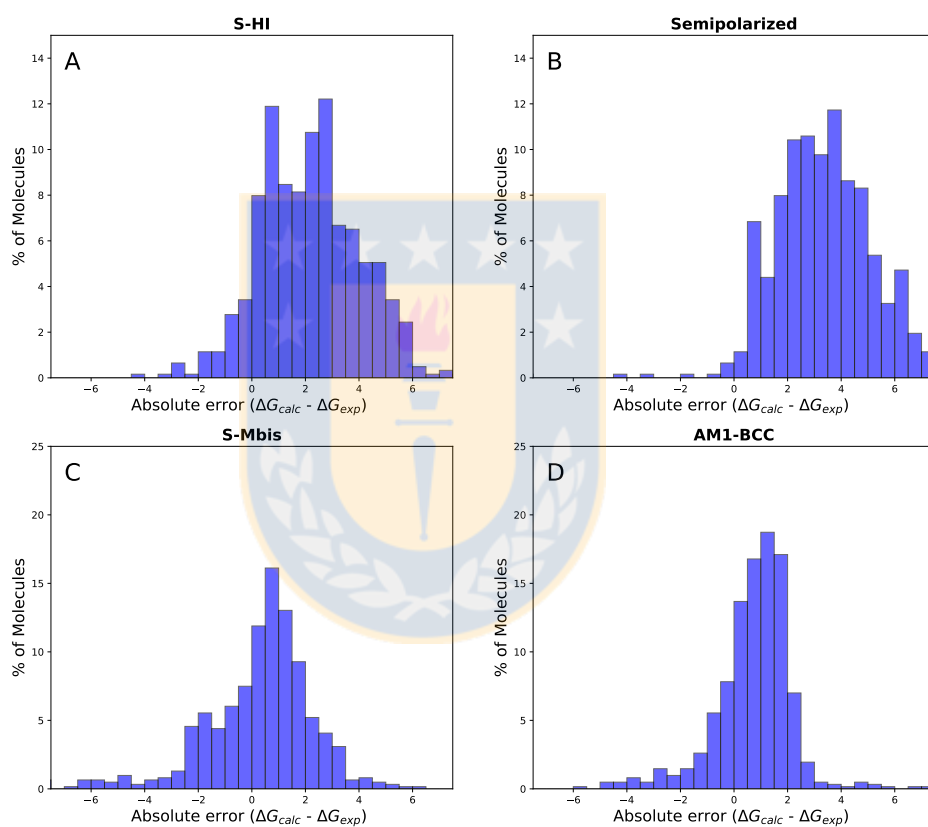


**Figura 3-4.:** Moléculas que presentan una desviación de la energía libre de hidratación calculada con las cargas S-MBIS del valor experimental que es más de cinco veces el error experimental. Fuente: elaboración propia.

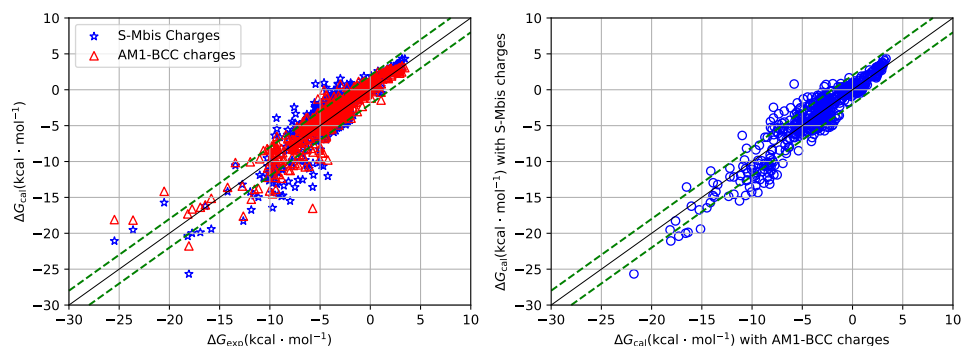
figura **3-4**. Una comparación de las diferentes moléculas muestra que la mayoría de ellas presentan átomos de fósforo o azufre y más de un grupo funcional. Suponiendo que los grupos funcionales contribuyen independientemente al error absoluto total, es de esperar que los solutos con muchos grupos funcionales tengan un error mayor.

Para analizar las desviaciones sistemáticas de las energías libres de hidratación calculadas se muestra en la figura **3-5** la distribución de los errores absolutos ( $\Delta G_{calc} - \Delta G_{exp}$ ) para cada conjunto de carga atómica. Ambos conjuntos de carga atómica que emplean las cargas Hirshfeld-I (S-HI y semipolarizado) presentan una amplia distribución con errores positivos sistemáticos resultantes de energías libres de hidratación calculadas demasiado positivas. Esto apunta a una deficiencia del método Hirshfeld-I y sus cargas atómicas derivadas cuando se usa con campos de fuerza para cálculos de energía libre. Para las cargas atómicas de S-MBIS se observa una distribución normal de descomposición rápida (valor de p de prueba de normalidad de D'Agostino y Pearson =  $3,76 \times 10^{-5}$ ) con un máximo ligeramente desplazado de cero. Para las cargas atómicas AM1-BCC también se observa una distribución normal ( $p = 2,15 \times 10^{-7}$ ), que podría originarse a partir de la parametrización de las correcciones de carga de enlace añadidas a las cargas atómicas AM1 para coincidir con las energías libres de hidratación experimental mencionadas en la referencia AM1-BCC original (páginas 1637-1638).<sup>105</sup> El cambio ligeramente mayor del máximo se relaciona principalmente con el modelo de agua SPC / E, como lo revela la comparación con las energías libres de hidratación calculadas en la base de datos FreeSolv que usa el modelo TIP3P.<sup>5</sup>

Finalmente, para comparar como se comportan las nuevas cargas atómicas, se comparan nuestros resultados con la partición de MBIS, incluida la corrección de polarización directamente a los resultados publicados anteriormente<sup>5</sup> (Figura **3-6**) que utilizó las cargas AM1-BCC junto con el modelo de agua TIP3P. La comparación



**Figura 3-5.** Distribución de los errores absolutos al valor experimental en las energías libres de hidratación obtenidas con las cargas S-HI (A), las cargas HI semipolarizadas (B), las cargas S-MBIS (C) y las cargas AM1-BCC (D). Fuente: elaboración propia.

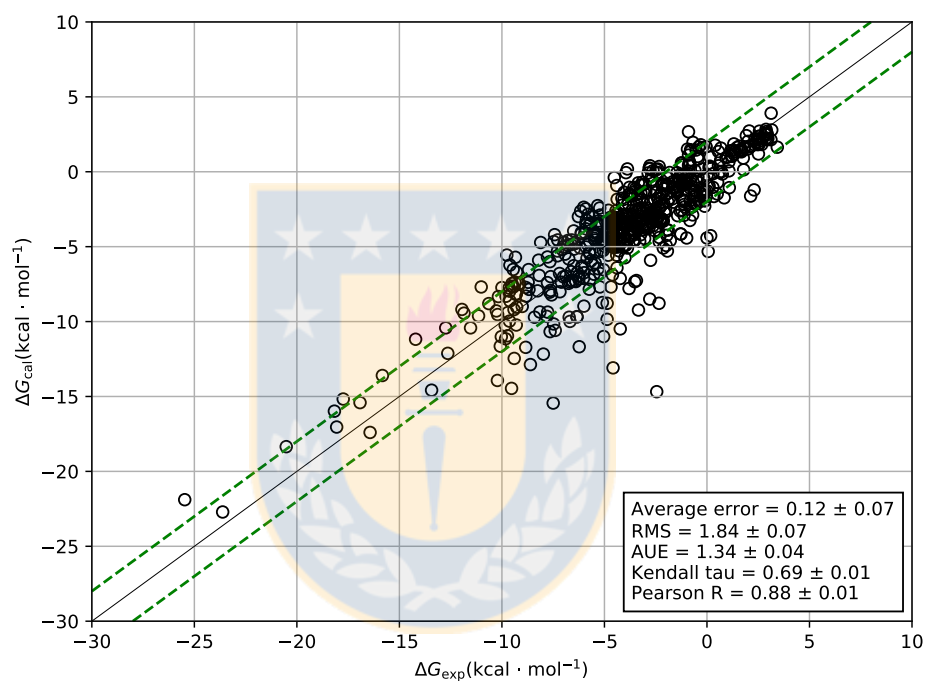


**Figura 3-6:** Energías libres de hidratación para todas las moléculas de la base de datos FreeSolv obtenidas con el campo de fuerza GAFF y las cargas MBIS incluyendo el modelo solvente SMD en el cálculo de la estructura electrónica (nivel BLYP / def2-TZVP) con la corrección de polarización (estrellas azules) y las energías libres de hidratación con el mismo campo de fuerza usando las cargas atómicas AM1-BCC frente a la referencia experimental. Fuente: elaboración propia.

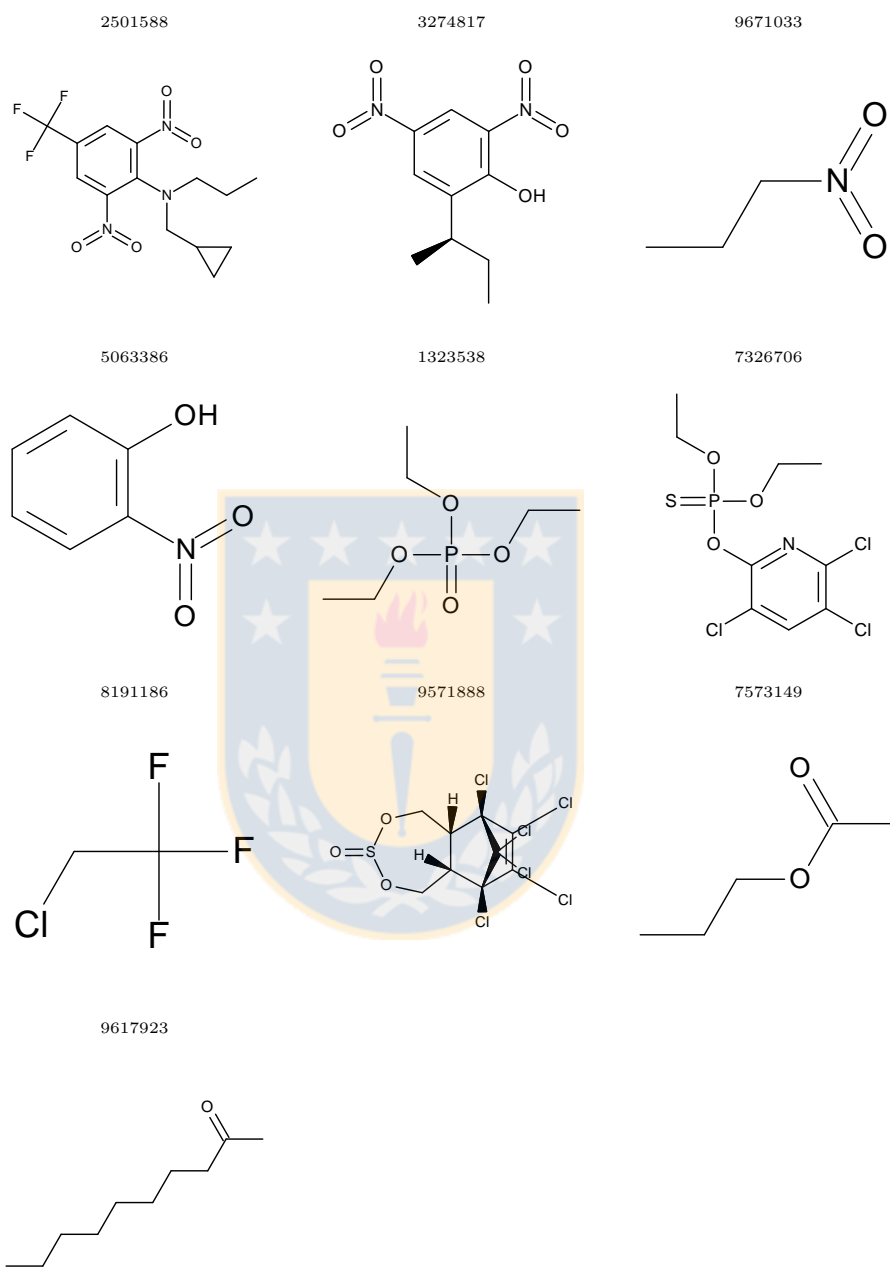
muestra que las cargas atómicas de S-MBIS proporcionan algunas mejoras para las moléculas con las energías libres de hidratación más negativas, que poseen más de un grupo hidroxilo. Para algunas de las moléculas más polares, los resultados de S-MBIS subestiman en comparación con los experimentos, mientras que los resultados de AM1-BCC en general tiene una buena correlación. Las energías libres de hidratación S-MBIS de las moléculas que contienen fósforo y azufre, con valores experimentales entre  $-10$  y  $-5$  kcal mol $^{-1}$ , se subestiman considerablemente y se discutirán en la última sección. Si este grupo de moléculas se elimina de los datos, el RMSE es 1.8 para los cargas atómicas S-MBIS en comparación con 1.3 kcal mol $^{-1}$  para los AM1-BCC en la base de datos FreeSolv.<sup>5</sup> Cuando se calcula la energía libre de hidratación con las mismas cargas atómicas AM1-BCC, pero el modelo de agua SPC / E, el RMSE aumenta a 1.7 kcal mol $^{-1}$  (ver Figura 3-2). Uno debe poner en perspectiva el buen rendimiento de las cargas AM1-BCC, ya que algunas de las correcciones de carga de enlace (BCC) en AM1-BCC se ajustaron para mejorar la reproducción de las energías libres de hidratación, mientras que tales correcciones empíricas no están

presentes en el enfoque S-MBIS.

También se calculó las energías libres de hidratación utilizando el modelo de solvatación implícito SMD que se utiliza para obtener la densidad molecular polarizada directamente de los cálculos de la estructura electrónica bajo la aproximación del oscilador armónico de rotor rígido incluyendo la corrección de energía de punto cero utilizando geometrías optimizadas tanto en vacío como con el modelo de solvatación. Las energías libres de hidratación obtenidas se muestran en el gráfico de la Figura **3-7**. Las energías libres de hidratación calculadas presentan descriptores estadísticos que se encuentran entre los valores de las cargas atómicas S-MBIS y AM1-BCC (Error promedio = 0.12, RMSE = 1.84, AUE 1.34 kcal mol<sup>-1</sup>), pero un considerablemente peor valor de Pearson R (R = 0.88) y Kendall  $\tau$  ( $\tau = 0.69$ ) además de algunos valores atípicos cuya estructura se proporciona en la Figura **3-8**. El buen rendimiento del modelo SMD puede estar sesgado: la mayoría de los parámetros para la contribución no electrostática a la energía libre de hidratación en el modelo de solvatación SMD se parametrizaron utilizando energías libres experimentales de hidratación para compuestos similares o iguales. Las moléculas para las que se observa una desviación significativa presentan grupos funcionales nitro, varios átomos de halógeno o ésteres de fosfato o tiofosfato. Una posible explicación podría ser que los valores atípicos identificados no estaban presentes en el conjunto de entrenamiento SMD.



**Figura 3-7:** Gráfico de las energías libre de hidratación experimental y calculada de todas las moléculas estudiadas en la base de datos FreeSolv obtenida como diferencia de energía libre de cálculos de estructura electrónica BLYP / def2-TZVP incluyendo corrección de energía de punto cero, rotor rígido, aproximación de oscilador armónico, las contribuciones electrostática y no electrostática (formación de cavidades, dispersión, etc.) del modelo de solvatación SMD para el estado estándar de 1 mol L $^{-1}$  y los mismos cálculos de estructura electrónica en vacío. Fuente: elaboración propia.



**Figura 3-8.:** Moléculas que presentan una desviación mayor que 5 kcal / mol de la energía libre de hidratación calculada a partir del valor experimental en la figura 3-7 y su ID utilizado en FreeSolv. Fuente: elaboración propia.

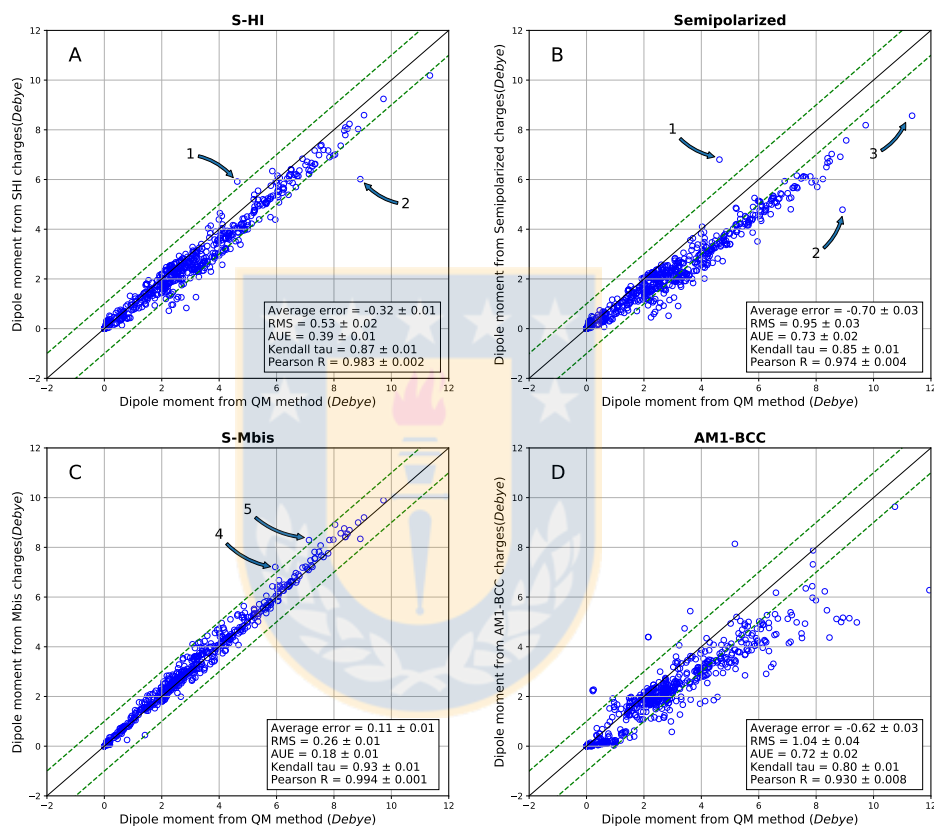


## 3.2. Momento dipolar molecular en disolución obtenidos desde diferentes modelos de cargas

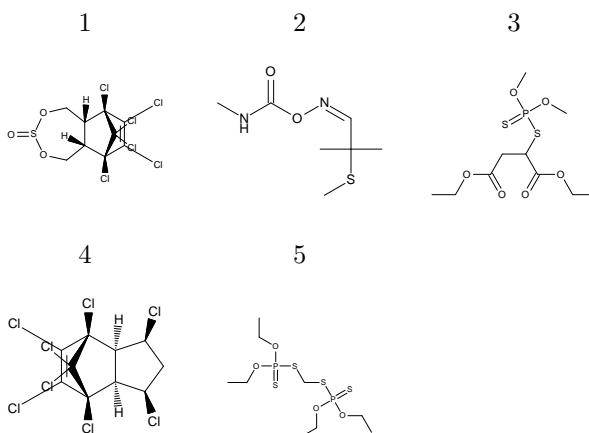
La validación primaria de las cargas atómicas parciales se basa en las energías libres de hidratación, pero también se examinó qué tan bien los conjuntos de cargas reproducen los momentos dipolares moleculares. Para una primera aproximación, las cargas atómicas deberían reproducir la magnitud del momento dipolar molecular estático del soluto en la estructura mínima en su entorno molecular. Dado que las cargas atómicas se utilizarán en simulaciones de fase condensada y no se dispone de datos experimentales sobre el momento dipolar molecular en solución acuosa, se utilizó la magnitud del momento dipolo molecular del soluto en solución acuosa calculada directamente a partir de la densidad molecular a nivel BLYP / def2-TZVP con el modelo implícito SMD como referencia. En principio, las cargas S-HI y S-MBIS se derivan de la misma densidad electrónica y, por lo tanto, deberían reproducir el momento dipolar molecular. Como se muestra en la Figura 3-9 A, las cargas S-HI muestran una buena correlación para valores pequeños, pero se observa una subestimación sistemática para valores más grandes del momento dipolo molecular como se evidencia por el valor negativo del error promedio lo cual se debe a que solo se consideraron monopolos en su determinación. Las cargas atómicas S-MBIS presentan una correlación mucho mejor, como se muestra en la Figura 3-9 C con algunos valores atípicos cuya estructura de Lewis se muestra en la Figura 3-10. Los valores atípicos son estructuras moleculares grandes con varios grupos funcionales que contienen átomos de cloro, azufre y fósforo. En los métodos de partición de Hirshfeld-I y MBIS, las cargas puntuales solo describen el término monopolo principal en la expansión multipolar atómica y se descuidan las contribuciones de dipolo atómico

al momento dipolar molecular. Una posible explicación para la desviación de las moléculas más grandes es el aumento del número de átomos cuyo momento dipolar se descuida y, por lo tanto, puede presentar un error mayor en las figuras **3-9 A** y **3-9 C**. Esto explica por qué los valores atípicos son moléculas grandes. Los valores atípicos también contienen grupos funcionales para los que se deben esperar dipolos atómicos grandes, por ejemplo, debido al  $\sigma$ -hole en halógenos más pesados o debido a grupos polares como sulfatos y fosfatos. En principio, las cargas descentradas o los multipolos centrados en el átomo pueden mejorar la representación de la densidad de electrones alrededor del átomo, pero dan como resultado un mayor costo computacional del campo de fuerza. Las cargas semipolarizadas de Hirshfeld-I no están destinadas a reproducir momentos dipolares de moléculas solvatadas. En su lugar, contienen desviaciones de la distribución de carga de equilibrio para tener en cuenta la energía de polarización. Esto es claramente visible en la Figura **3-9 B** y en los descriptores estadísticos en el recuadro. Estas cargas atómicas sistemáticamente subestiman el momento dipolar, lo cual se espera ya que solo se agrega la mitad de la diferencia a las cargas atómicas totalmente polarizadas. Curiosamente, en algunos casos se observan desviaciones muy grandes en los momentos dipolares (moléculas 1, 2 y 3 en la figura **3-10 B**). Para las moléculas 1 y 2, el método de partición y las limitaciones de las cargas atómicas discutidas anteriormente pueden contribuir considerablemente a la desviación observada ya que estos dos valores atípicos también están presentes en la figura **3-9 A** de las cargas atómicas S-HI.

Finalmente, en la Figura **3-9 D**, también se compara los momentos dipolares de las cargas AM1-BCC con los obtenidos de cálculos de estructura electrónica con solvatación implícita. Esto puede parecer una comparación injusta porque las cargas AM1-BCC se derivan del Hamiltoniano AM1 en el vacío. Sin embargo, las correc-



**Figura 3-9:** Magnitud de los momentos dipolares de las moléculas en la base de datos FreeSolv calculadas para la geometría optimizada (nivel BLYP / TZVP con el modelo de disolvente SMD) directamente con el método QM frente a los momentos dipolares resultantes de las cargas atómicas obtenidas con Hirshfeld-I (A) o el método de partición MBIS (C) de la misma densidad electrónica, las cargas semipolarizadas con el método Hirshfeld-I (B) y las cargas atómicas AM1-BCC (D). Las flechas identifican valores atípicos con las estructuras que se muestran en la Figura 3-10. Fuente: elaboración propia.



**Figura 3-10.:** Moléculas marcadas en la Figura 3-9 que presentan una gran desviación de su momento dipolo molecular calculado con diferentes cargas atómicas. Fuente: elaboración propia.

ciones de carga de enlace se agregan para tener en cuenta la polarización de las moléculas en el agua y para reproducir los valores obtenidos con el método RESP en el nivel de teoría HF / 6-31G (d). Luego se introdujeron algunas correcciones adicionales para reproducir las energías libres de hidratación. Como se evidencia a partir de los descriptores estadísticos en la Figura 3-9 D, estas cargas atómicas muestran una subestimación sistemática de los momentos dipolares con algunos valores atípicos pronunciados y una dispersión mayor. La mala correlación podría estar relacionada con correcciones de carga de enlace (BCC), que se ajustaron directamente para reproducir energías libres de hidratación sin garantizar un momento dipolar razonable.

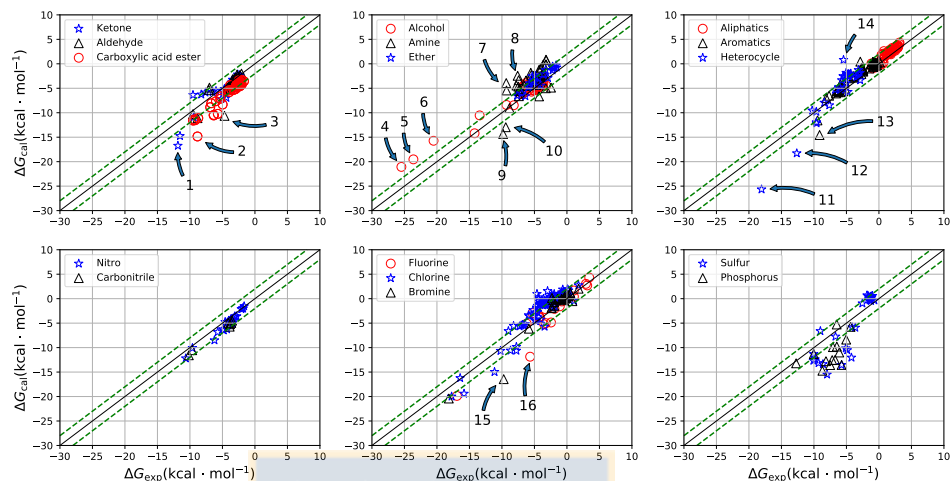
Se debe observar en este punto que el momento dipolar molecular de referencia obtenido con el funcional BLYP produce pequeños momentos dipolares en vacío<sup>103</sup> en comparación con B3LYP. Esta deficiencia del funcional BLYP se corrige parcialmente por el funcional B3LYP computacionalmente más exigente. El cálculo de los

momentos dipolares con el funcional B3LYP aumentaría la magnitud de los valores de referencia, lo que resultaría en desviaciones aún mayores para las cargas AM1-BCC. Sin embargo, para los otros tres conjuntos de cargas atómicas (S-HI, semipolarizadas y S-MBIS), las conclusiones serán similares ya que sus cargas atómicas se derivarían de la densidad electrónica molecular B3LYP la cual es más polarizada y llevaría a momentos dipolares más grandes.

En resumen, se puede concluir que la magnitud de los momentos dipolares moleculares estáticos de la estructura mínima del soluto calculada con cálculos de estructura electrónica y el modelo de solvatación SMD se reproducen mejor mediante las cargas S-MBIS seguidas por las cargas S-HI con algunas desviaciones sistemáticas. Las cargas Hirshfeld-I y semipolarizadas en algunos casos y las cargas AM1-BCC tienen un acuerdo particularmente pobre con los momentos de dipolo molecular ab-initio. Las conclusiones obtenidas para el método de partición MBIS e Hirshfeld-I concuerdan con los resultados anteriores al usar las cargas de MBIS en vacío de Verstraelen et al.<sup>55</sup> Para ser justos, ni las cargas AM1-BCC ni las Hirshfeld-I semipolarizadas deberían necesariamente reproducir los momentos dipolares particularmente bien, ya que no están diseñadas para hacerlo y podrían predecir energías libres de hidratación razonablemente bien sin producir momentos dipolares precisos.

### **3.3. Energía libre de hidratación para una variedad de grupos funcionales**

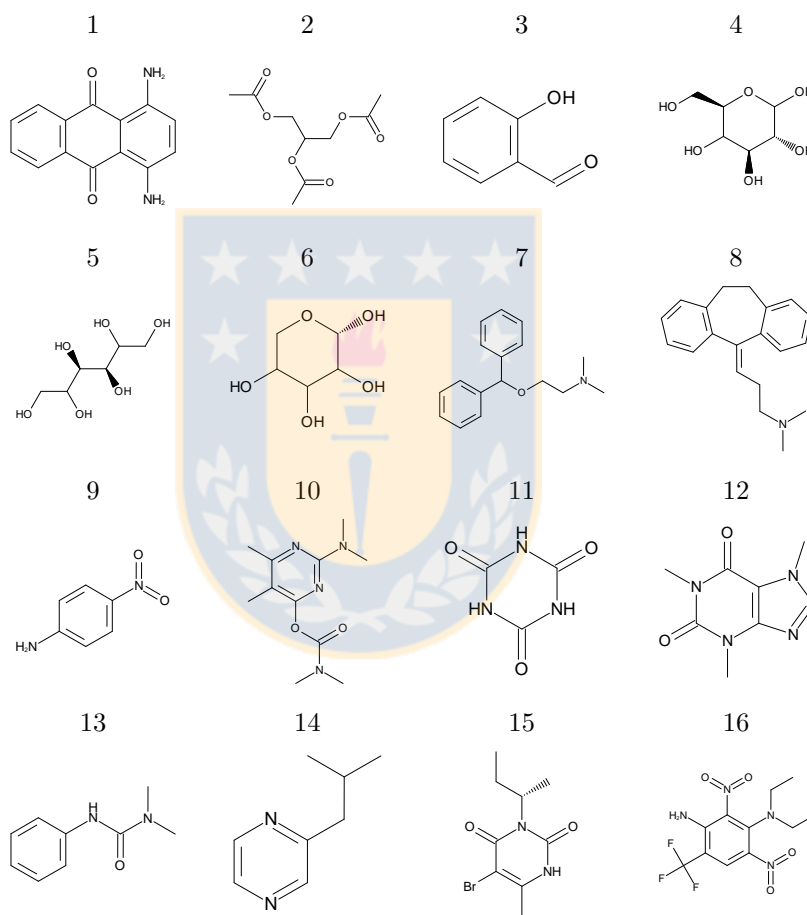
El análisis hasta ahora tomó en cuenta toda la base de datos sin examinar las tendencias con respecto a grupos funcionales específicos que son clave para impulsar el comportamiento de solvatación, por lo que se dividió el conjunto de datos



**Figura 3-11.:** Energías libres de hidratación separadas por grupos funcionales en la base de datos FreeSolv obtenida con cargas atómicas S-MBIS. Los valores atípicos (números) se muestran en la Figura 3-12. Fuente: elaboración propia.

en categorías de grupos funcionales y subconjuntos analizados. Algunas moléculas presentan más de un tipo de grupo funcional y la asignación no es inequívoca. Para este análisis, solo se consideró el conjunto de carga atómica S-MBIS ya que este se comporta mejor; para los otros conjuntos de carga atómica, los resultados se resumen en las Figuras A-1, A-2 y A-3 en el anexo A.

La figura 3-11 muestra las energías libres de hidratación calculadas para las moléculas que contienen al menos un grupo carbonilo, amina, éter o un grupo hidroxilo y los alifáticos, aromáticos y heterociclos en las tres gráficas superiores. En la parte inferior se muestran los grupos nitro y carbonitrilo seguidos por compuestos que contienen halógeno y moléculas con átomos de azufre o fósforo. Para cetonas y aldehídos, la mayoría de las energías libres de hidratación tienen errores con magnitudes inferiores a  $2 \text{ kcal mol}^{-1}$ . Las moléculas con mayores desviaciones se identificaron como 1, 2 y 3 (ver Figura 3-12) y presentan más de un grupo funcional, o grupos



**Figura 3-12.:** Moléculas identificadas como valores atípicos en la Figura 3-11. Fuente: elaboración propia.

amino como en la molécula 1. Para la molécula 3, una competencia entre un enlace de hidrógeno intramolecular y un enlace de hidrógeno con las moléculas de agua circundantes se espera.<sup>106</sup> El modelo de solvatación implícita utilizado en este estudio para derivar las cargas atómicas no explica esta competencia. Por lo tanto, sería necesario un enfoque que incluya la descripción explícita de las moléculas de agua en la derivación de las cargas atómicas y un muestreo adecuado de las conformaciones relevantes<sup>59</sup> lo cual se describirá en la sección 3.4.

Como se muestra en la figura **3-11** para ésteres de ácido carboxílico y amidas se obtienen sistemáticamente energías libres de hidratación más negativas, lo que explica el valor atípico 2 en el que el error de los tres grupos funcionales éster podría explicar este mayor error. Para los alcoholes, las mayores desviaciones se observan cuando hay más de tres grupos funcionales (moléculas 4, 5 y 6), lo que sugiere que cada grupo funcional contribuye al error de forma casi independiente, lo que da como resultado una gran desviación total. También para estas moléculas, la competencia entre enlaces de hidrógeno intra- e intermoleculares<sup>106,107</sup> desempeña un papel principal y puede alterar la estabilidad de las diversas conformaciones diferentes en solución acuosa. En las aminas, las moléculas 7 y 8, que son aminas terciarias, presentan la mayor desviación junto con la molécula 9, que también contiene un grupo nitro unido a un anillo de benceno. Este grupo nitro junto con el grupo amino en posición para formar un sistema de electrones  $\pi$  conjugado, que podría no estar bien descrito con cargas atómicas fijas centradas en el átomo. La misma molécula también se identificó como problemática utilizando las cargas atómicas AM1-BCC.<sup>4</sup> En el grupo amino, la molécula 10 también posee una gran desviación del valor experimental probablemente debido a la adición de los errores asociados con la amina terciaria, el grupo amida y el heterociclo con dos átomos de nitrógeno, lo que proporciona un error



adicional a la energía libre como se mostrará a continuación. Las energías libres de hidratación de compuestos que contienen grupos funcionales éter están todas bien reproducidas.

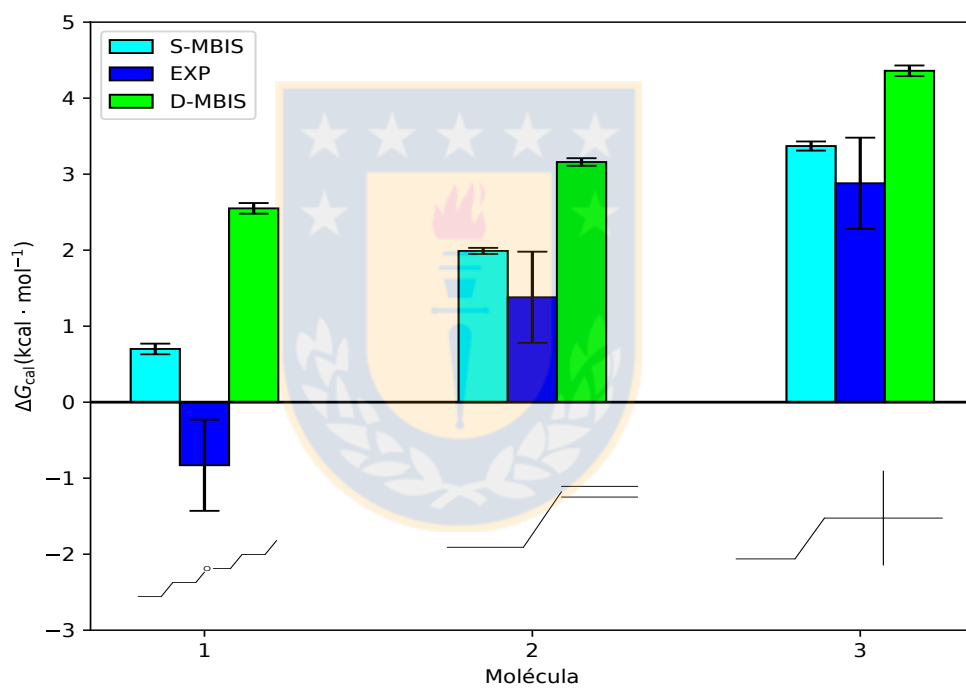
Las energías libres de hidratación calculada de compuestos alifáticos y aromáticos se encuentran principalmente dentro de la banda de error de  $2 \text{ kcal mol}^{-1}$  siendo el valor atípico 13 una molécula con una parte de su estructura similar a la urea, que puede contribuir a la desviación debida a los dos átomos de nitrógeno. En los heterociclos, se observan desviaciones pronunciadas en las moléculas 11, 12 y 14, que contienen más de un átomo de nitrógeno en el heterociclo. Para las moléculas que contienen un grupo funcional nitro o carbonitrilo, las cargas atómicas de S-MBIS proporcionan energías libres precisas, más precisas que los valores informados con las cargas AM1-BCC.<sup>5</sup> Las moléculas que contienen halógeno están más cerca del valor experimental para flúor, siendo el valor atípico 16 un compuesto con dos grupos amina, lo que podría contribuir al error absoluto. Las moléculas que contienen cloro en general presentan desviaciones más grandes, que también pueden provenir de otros grupos funcionales, la presencia de múltiples átomos de cloro en la molécula y el efecto sigma-hole. Entre los compuestos que contienen bromo, el valor atípico más pronunciado es 15, que también tiene un heterociclo con dos nitrógenos, que se sabe causan grandes errores. En la última gráfica, se muestran las energías libres de hidratación calculadas para los compuestos que contienen azufre y fósforo. La correlación con las energías libres de hidratación experimentales es mejor para compuestos que contienen azufre si el heteroátomo está unido a átomos de carbono. La desviación más grande corresponde a moléculas con grupos sulfonilo o sulfonato y fosfonato. En estos grupos, las cargas parciales atómicas son relativamente grandes en valor absoluto debido al alto estado de oxidación del azufre o átomo de fósfo-

ro. Con tales distribuciones de carga pronunciadas, las interacciones electrostáticas entre las cargas atómicas y los dipolos atómicos también se vuelven importantes, los cuales se descuidan en cualquier campo de fuerza que usa un modelo de carga puntual. Por lo tanto, no es sorprendente que esta distribución de carga pronunciada conduzca sistemáticamente a errores mayores en la energía libre de hidratación. Uno puede sesgar las cargas en estos grupos (o los parámetros de van der Waals) para compensar la ausencia de dipolos atómicos en el campo de fuerza, pero no está claro cómo se puede hacer de forma sistemática.

Este análisis general ya identifica cualitativamente algunos grupos funcionales con desviaciones más grandes, como sulfonatos, fosfonatos, aminas y heterociclos con dos átomos de nitrógeno o aminas. Sin embargo, un análisis cuantitativo no es posible con la clasificación por grupo funcional, ya que una molécula también podría presentar múltiples grupos que contribuyen al error absoluto. Para cuantificar el error debido a cada grupo funcional o incluso debido a cada elemento en un soluto, en colaboración con Prof. Verstraelen se desarrolló un modelo estadístico. Este modelo cuenta características (grupos funcionales o elementos) en cada soluto y asume que el error total entre el cálculo y el experimento es una suma de errores normales independientes debido a (i) características presentes en el soluto, (ii) una incertidumbre debida al modelo general errores (por ejemplo, en el disolvente), (iii) el error de medición experimental fijo y (iv) el error estadístico del cálculo de la energía libre. La media y la varianza de la incertidumbre asociada con cada característica y el error general se ajustaron a las diferencias entre las energías libres de S-MBIS y la hidratación experimental. La mayor desviación se observó para las moléculas que contienen fósforo y las moléculas con grupos funcionales amida, éster y amina.

### 3.4. Energía libre de hidratación utilizando cargas dinámicas MBIS

Para estudiar el efecto de las cargas dinámicas de MBIS (D-MBIS) se seleccionó un grupo de moléculas de la base de datos FreeSolv. El criterio que se usó para su selección fue principalmente moléculas que han sido reportadas que presentan cambios conformacionales o moléculas que presentaban una gran desviación de la energía libre de hidratación usando las cargas S-MBIS. Las moléculas seleccionadas se irán presentando junto a los resultados de energía libre de hidratación y han sido ordenadas por similitud en su estructura química o porque se obtiene alguna tendencia en los resultados similar. Los valores de la energía libre de hidratación han sido representados mediante histogramas para poder comparar las cargas S-MBIS, D-MBIS y los respectivos valores experimentales. El primer histograma se muestra en la figura **3-13** y considero moléculas que presentan un valor pequeño de energía libre de hidratación. En esta categoría se encontraron tres moléculas representadas por la numeración 1 al 3 en la figura **3-13**, la primera molécula es un éter con dos grandes cadenas alifáticas con el cual se obtienen mejores resultados para la energía libre de hidratación con el modelo de cargas S-MBIS en comparación con las cargas D-MBIS, lo mismo sucede para las moléculas 2 y 3 que son hidrocarburos. Si bien la primera molécula presenta un grupo éter esta tiene principalmente un carácter alifático debido a los dos sustituyentes que presenta, lo cual la hace comparable a las moléculas 2 y 3 que también son dos compuestos alifáticos. En estas tres moléculas se obtienen valores más pequeños para las cargas dinámicas en comparación con las cargas S-MBIS lo que haría que el valor de energía libre sea más positivo. Esta diferencia que existe en las cargas dinámicas está asociada con los cambios conformacionales que presentan



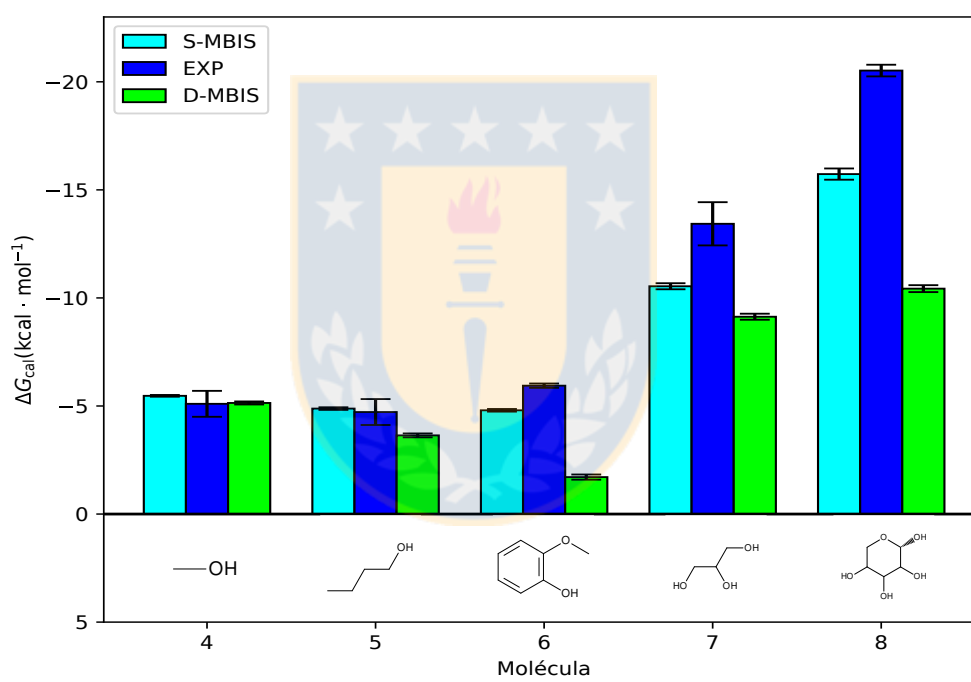
**Figura 3-13.:** Energía libre de hidratación usando cargas D-MBIS para moléculas con bajos valores de energía libre. Fuente: elaboración propia.

**Tabla 3-1.:** Comparación entre la corrección de polarización para las cargas S-MBIS y para las cargas dinámicas D-MBIS en kcal/mol.

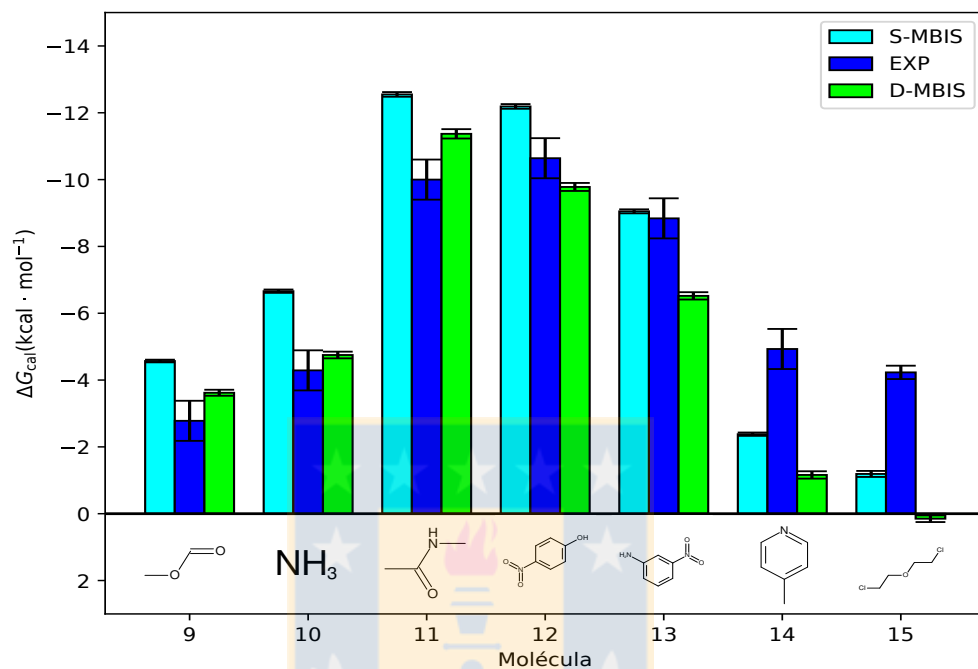
Molécula	S-MBIS	D-MBIS
1	1.02	2.98
2	0.24	1.52
3	0.05	1.22
4	1.39	3.58
5	1.40	4.05
6	1.48	4.87
7	4.15	7.86
8	3.42	10.48
9	1.73	3.25
10	1.60	4.24
11	4.60	8.38
12	4.19	6.08
13	3.31	5.26
14	2.19	3.94
15	3.21	3.49

las cadenas alifáticas a lo largo de la dinámica. Los diferentes cambios conformacionales que presentan estas moléculas, se ven reflejados también en la corrección de polarización (ver Tabla 3-1) la cual es mucho más positiva para las cargas dinámicas en comparación con las cargas S-MBIS, lo que genera que los valores finales de energía libre sean más positivos.

Otro factor importante a considerar en este tipo de moléculas que hace que ambas cargas aun estén lejos del valor experimental de la energía libre de hidratación es que en los compuestos alifáticos los factores que preponderan sobre las cargas son los parámetros de Lennard Jones en término de dispersión como también los parámetros repulsivos ( $C_{12}$ ) los cuales dependiendo como han sido parametrizados afectan considerablemente los valores de energía libre energía libre de hidratación en este tipo de moléculas.



**Figura 3-14.:** Energía libre de hidratación usando cargas D-MBIS para moléculas que presentan uno o más grupos hidroxilo en su estructura. Fuente: elaboración propia.



**Figura 3-15.:** Energía libre de hidratación usando cargas D-MBIS para una variedad de moléculas con diferentes grupos funcionales. Fuente: elaboración propia.

En la figura **3-14** se representa los resultados para las moléculas que tienen el grupo hidroxilo en su estructura. En la molécula 4 las cargas dinámicas dan un mejor valor que las cargas S-MBIS y muy cercano al valor experimental. Esto en un principio podría dar la idea de que en este tipo de compuestos las cargas dinámicas dan mejores resultados en comparación con las cargas S-MBIS, pero si se observa las moléculas 5 y 6 de la figura **3-14** los valores de la energía libre de hidratación usando las cargas D-MBIS se alejan de los valores experimentales. Esta desviación se debe principalmente en el caso de la molécula 5 al aumento en el largo de la cadena alifática la cual se demostró en los resultados de la figura **3-13** que los compuestos alifáticos dan valores más positivos usando las cargas D-MBIS.

En la molécula 6 se obtiene un mejor resultado para las cargas S-MBIS en comparación con las cargas D-MBIS. Esta diferencia se debe principalmente a que en el caso de las cargas S-MBIS se forma un puente de hidrógeno intramolecular entre el grupo eter y el grupo hidroxilo el cual no existe en el caso de las cargas dinámicas donde el grupo hidroxilo forma un puente de hidrógeno con el solvente. Esto se ve reflejado en el valor de las cargas donde el oxígeno del grupo eter toma un valor de -0.3274 para las cargas S-MBIS y -0.3174 para las cargas D-MBIS y para el oxígeno del grupo hidroxilo tiene una carga de -0.6046 para S-MBIS y -0.6256 para D-MBIS lo cual demuestra lo anteriormente descrito. En las moléculas 7 y 8 de la figura **3-14** se ha encontrado que al aumentar el número de grupos hidroxilo se obtiene una mayor corrección de polarización usando las cargas D-MBIS la cual es proporcional al número de -OH presentes en la molécula, lo que queda demostrado al comparar la corrección de polarización de ambas moléculas (ver Tabla 3-1). Esta mayor corrección de polarización se debe a que al aumentar el número de grupos hidroxilo aumenta la posibilidad de formar puentes de hidrógenos con el solvente lo que harían que la corrección de polarización sea más positiva.

En la figura **3-15** se muestran las últimas moléculas a las cuales se le calculó la energía libre de hidratación usando las cargas D-MBIS. Las primeras tres moléculas 9, 10 y 11 si bien no tienen una similitud química característica tienen un resultado similar en la estimación de la energía libre de hidratación. Para estas tres moléculas se observa en la Figura **3-15** que las cargas D-MBIS dan un mejor valor en comparación con las cargas S-MBIS. Una explicación para esto sería que en el caso de los grupos amino las cargas D-MBIS incluyen la inversión del nitrógeno lo cual se observa principalmente en la molécula 10. En las moléculas 9 y 11 las cargas D-MBIS describen mejor el valor de energía libre de hidratación que las cargas S-MBIS porque



ellas dan cuenta de la formación de puentes de hidrógeno del grupo carbonilo con las moléculas de agua explícita. Además, el factor conformacional en estas dos moléculas puede tener una importancia considerable principalmente en el grupo metilo de la amida que en el caso de las cargas dinámicas describen su rotación libre al igual que en el caso del ester que también tiene un grupo metilo de libre rotación. Las moléculas 12 y 13 tienen una similitud química y la única diferencia es la posición y el sustituyente que en el caso de la molécula 12 es un hidroxilo y en la molécula 13 es un grupo amino. En estas dos moléculas los oxígenos del grupo nitro presentan cargas mucho más negativas para las cargas S-MBIS en comparación con las cargas dinámicas lo que provoca que la energía libre de hidratación sea mucho más negativa en estas dos moléculas para este tipo de cargas. En el caso de las cargas dinámicas los oxígenos del grupo nitro tienen una carga menos negativa lo que hace que la energía libre de hidratación sea menos negativa. Además, estas dos moléculas presentan un grupo hidroxilo y un grupo amino que en el caso de las cargas dinámicas forma puente de hidrógeno con el agua generando una corrección de polarización más positiva en comparación con las cargas S-MBIS lo que también aporta a que la energía libre sea más positiva para las cargas dinámicas. En las moléculas 14 y 15 ambas cargas tienen una mala descripción siendo en ambos casos mejor las cargas S-MBIS. Anteriormente ya se había comentado el problema que presentan las moléculas que tienen pares de electrones no compartidos los cuales no pueden ser descritos con ningún tipo de carga y en el caso de algunas moléculas este factor parece ser importante para poder reproducir un buen valor de energía libre de hidratación.

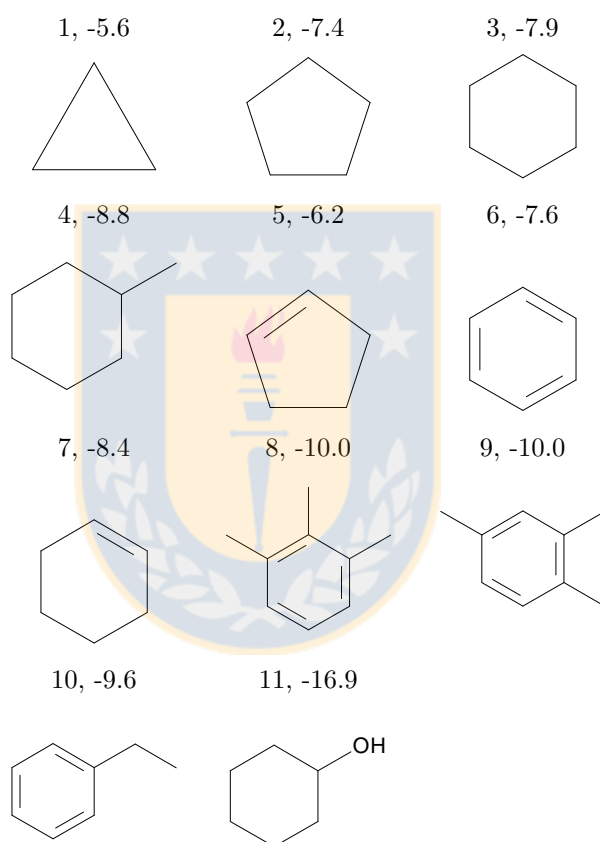
Es importante resaltar que si bien las cargas dinámicas dan cuenta de cambios conformacionales y de la formación de puentes de hidrógeno entre el soluto y el disolvente estas se calculan usando las distancias de enlace y ángulos de enlaces

provenientes del campo de fuerza los cuales pueden variar si se compara con las cargas S-MBIS donde estas son calculadas usando estos parámetros del cálculo de estructura electrónica lo cual podría ser un factor que también influye en el cálculo de la energía libre de hidratación.

### 3.5. Entalpía y entropía usando cargas S-MBIS

La entalpía y entropía de hidratación son dos valores que generalmente no se consideran en la validación de campos de fuerzas. Una de las posibles razones de esto es que no existen muchos valores experimentales de estas propiedades. Como parte final de esta tesis se estudió si además de poder reproducir los valores de energía libre de hidratación con los modelos de carga propuestos se podía reproducir la entalpía y la entropía de hidratación. Con esta idea en mente el primer problema que se tuvo es que de toda la base de datos FreeSolv la cual contiene más de 600 moléculas solo para 11 moléculas se conocen los valores de la entalpía de hidratación y para la entropía de hidratación no hay valores reportados. Por este motivo solo se calculó la entalpía de hidratación para las moléculas que tienen valores experimentales para poder comparar con los resultados obtenidos con las cargas S-MBIS.

Recientemente los creadores de la base de datos FreeSolv también calcularon la entalpía de hidratación usando la metodología descrita en la sección de métodos la cual es la misma que se utilizó en este trabajo. La principal diferencia entre el trabajo de ellos y el presentado en esta tesis es el modelo de carga utilizado que en el caso de ellos fueron las cargas AM1-BCC y en este trabajo se usan las cargas S-MBIS. Esto es una ventaja ya que además de poder comparar los resultados obtenidos con S-MBIS con los valores experimentales se pueden comparar con los resultados de Guilherme et al.<sup>5</sup>



**Figura 3-16.:** Moléculas estudiadas para la entalpía y entropía de hidratación las cuales han sido enumeradas y se informa su entalpía de hidratación experimental respectivamente en kcal/mol. Fuente: elaboración propia.

Las moléculas estudiadas se representan en la figura **3-16** donde es importante recalcar que la mayoría de las moléculas a las cuales se le calculó la entalpía de hidratación y la entropía son moléculas alifáticas y una molécula con un grupo hidroxilo. En la figura **3-17** se presentan los resultados obtenidos para las cargas S-MBIS usando agua representada por los modelos TIP3P y SPC/E junto a los valores obtenidos por Guilherme et al. En esta figura se observa que considerando el mismo modelo de carga el disolvente explícito que da un mejor resultado es el modelo de agua SPC/E. Esto se debe a que el modelo SPC/E reproduce mejor las propiedades del agua en comparación con el modelo TIP3P. Además este modelo de agua incluye una corrección de polarización promedio a la función de polarización lo que da como resultado una mejor densidad y constante de difusión.

Si se compara los dos modelos de carga S-MBIS y AM1-BCC en la mayoría de las moléculas las cargas S-MBIS dan un mejor resultado en comparación con AM1-BCC aunque es importante señalar que esta mejora es pequeña y aun se está lejos del valor experimental si se usa el modelo de agua TIP3P. Sin embargo, era de esperar que el modelo de carga no sea tan relevante en esta propiedad ya que la mayoría de las moléculas son compuestos alifáticos donde los parámetros de Lennard Jones juegan un rol más importante. En la molécula con el grupo hidroxilo si se observa una mejora con las cargas S-MBIS con respecto a las AM1-BCC.

La tendencia que se observa en los resultados es que a medida que se aumenta el número de átomos de carbono en cada uno de los ciclos el valor de entalpía se hace más negativo, por ejemplo, al comparar ciclo propano con ciclo pentano y ciclo hexano. En los ciclos con doble enlace la entalpía de hidratación es más negativa si se compara con el ciclo alcano. Y en los compuestos aromáticos si aumenta el largo de los grupos sustituyentes el valor de entalpía de hidratación se hace más negativo. En

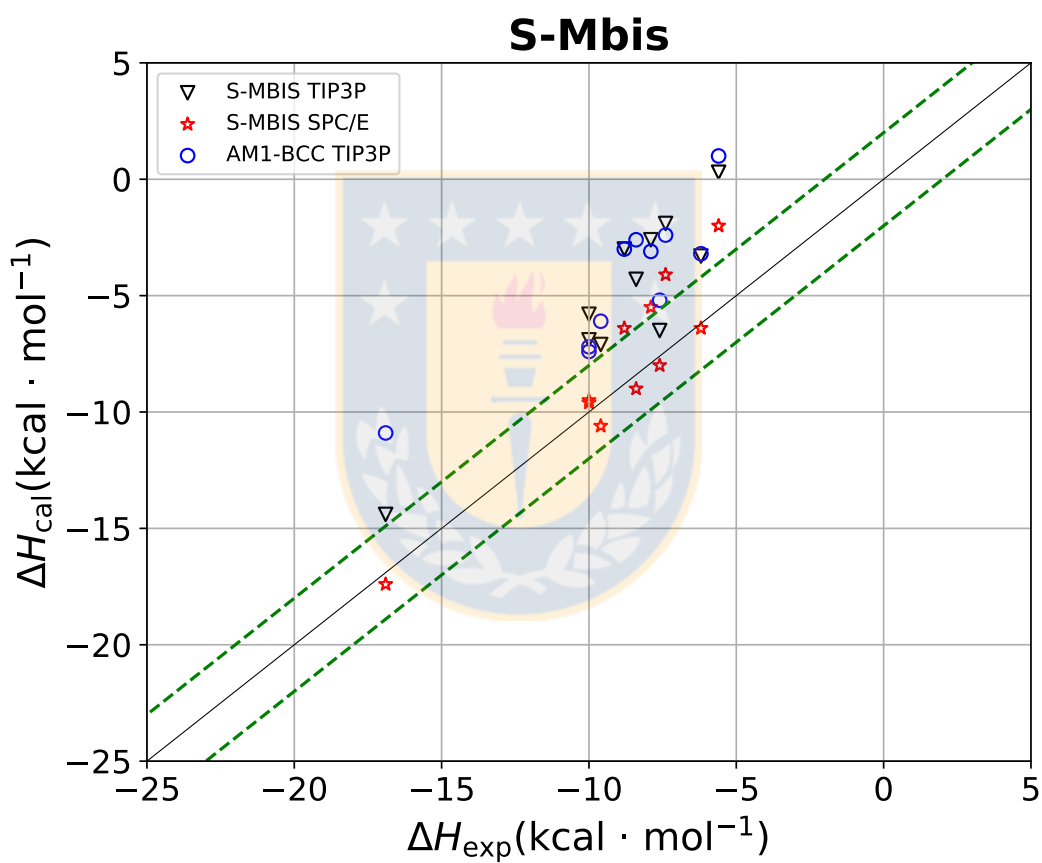


Figura 3-17.: Entalpía de hidratación usando cargas S-MBIS junto a dos disolventes explícitos y los valores reportados para las cargas AM1-CC. Fuente: elaboración propia.

la entalpía de hidratación compiten dos factores, por un lado esta la energía necesaria para crear la cavidad dentro del disolvente la cual genera un gasto de energía y es proporcional al tamaño del soluto, lo cual implica que a mayor tamaño del soluto se requieren una mayor cavidad y esto necesita una mayor energía necesaria. Por otro lado el otro factor importante en la entalpía de hidratación son las interacciones de dispersión producidas por la interacción del soluto y el disolvente. Como en todas las moléculas se obtienen valores de entalpía negativos esto implica que prioriza las interacciones dispersivas sobre la energía necesaria para generar la cavidad en el disolvente.

Las moléculas que presentan una mayor desviación sobre 2 kcal/mol para la entalpía de hidratación con el modelo de agua SPC/E son las moléculas 1-4 de la figura **3-16** donde se observa que para los ciclo alcanos existe un error que afecta solo a este tipo de moléculas. Este error sigue la siguiente tendencia  $1 > 2 > 3 > 4$ , y podría deberse a errores en los parámetros de Lennard Jones para este tipo de compuestos. Para el compuesto 1, por ejemplo, se usa un tipo de átomo para el carbono diferente a los demás compuestos siendo que en las cuatro moléculas solo existe un tipo de átomo para el carbono.

La entropía de hidratación también se hace más negativa (ver Tabla **3-2**) al aumentar el tamaño del ciclo y en el caso de los compuestos aromáticos mientras más largas sean las cadenas de los sustituyentes la entropía será más negativa. Esto se debe en gran medida al efecto hidrofóbico que producen estas moléculas sobre el disolvente, ya que las moléculas de agua se deben reordenar producto del soluto alifático. Este reordenamiento es proporcional al tamaño del soluto por eso la entropía de hidratación se hace más negativa al aumentar el tamaño del ciclo o al aumentar el tamaño de las cadenas alifáticas en los alquilbencenos. También afecta

**Tabla 3-2.:**  $T\Delta S$  calculado y experimental para las 11 moléculas de la Figura 3-16 en kcal/mol.

Molécula	$T\Delta S_{cal}$	$T\Delta S_{exp}$
1	-4.6	-6.3
2	-5.9	-8.6
3	-7.5	-9.2
4	-8.5	-10.5
5	-7.2	-6.8
6	-7.3	-6.7
7	-9.5	-8.5
8	-9.5	-8.8
9	-10.0	-9.2
10	-10.3	-8.8
11	-11.7	-11.4

a la entropía de hidratación la entropía del soluto, pero en menor medida que la entropía del disolvente.

Los resultados de la entropía de hidratación comparados con los valores experimentales se presentan en la Tabla 3-2 donde se observa una buena correlación donde la mayoría de las moléculas esta dentro del error de 2 kcal/mol para las moléculas consideradas. Es interesante que las moléculas que tienen un error mayor son solamente ciclo alcanos, aunque este error proviene principalmente del error que ya tenia el cálculo de la entalpía. Este error es mucho menor para el valor de la entropía con un valor cercano a las 2 kcal/mol, lo que implica que debe haber una cancelación de errores entre el valor de energía libre de hidratación y el valor de entalpía de hidratación que hacen que este error disminuya.

## 4. Conclusiones

En este trabajo se estudiaron diferentes modelos de cargas atómicas para reproducir las energías libres de hidratación experimental de las moléculas en la base de datos FreeSolv utilizando el campo de fuerza GAFF a 25C y 1 atm de presión. Además, se determinó la capacidad de reproducir la entalpía y entropía de hidratación para el mejor modelo de carga atómica encontrado en este trabajo.

Al comparar las cargas de Hirshfeld-I en vacío con las S-HI que usan un solvente implícito y consideran la polarización del soluto se concluye que el efecto de la polarización de la densidad electrónica a través del solvente implícito antes de la determinación de las cargas atómicas es fundamental para mejorar las energías libres de hidratación calculadas. El agregar el modelo de solvatación implícita genera valores absolutos de cargas atómicas más grandes y energías libres más negativas, las cuales al ser corregidas con la corrección de polarización producen una mejor correlación con los valores experimentales. Las cargas semipolarizadas producen mejores resultados que las cargas HI en vacío, pero una peor correlación que las cargas S-HI.

La polarización de la densidad electrónica y la energía de corrección de la polarización se vuelven más importantes para las moléculas más polares, las cuales tienen energía libre de hidratación más negativas e interacciones electrostáticas más fuertes con el solvente.

Al comparar las cargas S-HI con diferentes métodos de estructura electrónica



(BLYP / def2- TZVP y B3LYP / def2- TZVP) se concluye que hay una pequeña diferencia de los valores de energía libre de hidratación para moléculas con energía libre superiores a -10 kcal/mol y para moléculas más polares se obtiene valores más negativos con el método B3LYP / def2- TZVP. Esto se debe principalmente a que las densidades electrónicas de BLYP son más uniforme lo que implica que se obtienen valores de cargas más pequeños y energías libres de hidratación más positivas. Entre los dos métodos BLYP y B3LYP este último incluye el intercambio exacto de Hartree-Fock lo que permite obtener mejores valores de energía que con BLYP, pero tiene un costo computacional mayor. Para las moléculas que contienen sulfatos y fosfatos, cuyas cargas atómicas se calcularon con el método alternativo Hirshfeld-E, se observó una mayor dependencia de los métodos de estructura electrónica.

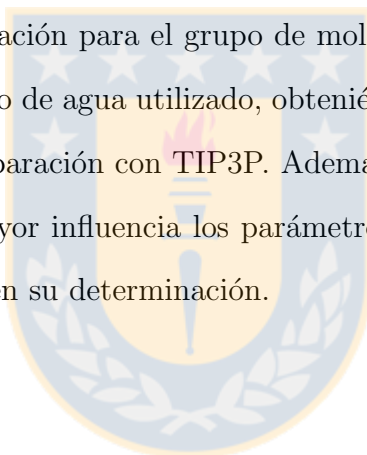
Se analizó el efecto del modelo de agua utilizado en las simulaciones de dinámica molecular variando el modelo SPC/E por el TIP3P. En general no hay una gran diferencia en los valores de la energía libre de hidratación al comparar ambos modelos, pero si hay una diferencia considerable cuando se calcula la entalpía de hidratación, donde el modelo SPC/E da un mejor resultado.

Al comparar los dos métodos de partición de la densidad electrónica (S-HI y S-MBIS), las cargas obtenidas con la partición MBIS reproducen claramente las energías libres de hidratación mejor que las cargas obtenidas con la partición HI y con un rendimiento comparable a las cargas AM1-BCC.

Los mayores errores los presentan moléculas con grupos funcionales como sulfonatos, fosfonatos, aminas y heterociclos con dos átomos de nitrógeno, donde en los dos primeros casos se debe al alto estado de oxidación del átomo de azufre y fósforo. En el caso de los demás grupos funcionales el error se asocia al no se considerar la distribución anisotrópica del par de electrones libres del nitrógeno.

Las cargas dinámicas (D-MBIS) tienen la ventaja que describen procesos como la formación de puentes de hidrógeno o si hay algún cambio conformacional lo cual en algunas moléculas mejora el cálculo de la energía libre de hidratación con este tipo de carga. Pero algunas moléculas como hidrocarburos o compuestos con anillos aromáticos presentan mayores desviaciones en el cálculo de la energía libre de hidratación, lo cual se puede deber no a un error en el modelo de carga si no a parámetros de Lennard Jones del campo de fuerza que influyen considerablemente en este tipo de moléculas.

La entalpía de hidratación para el grupo de moléculas estudiadas depende considerablemente del modelo de agua utilizado, obteniéndose un mejor resultado para el modelo SPC/E en comparación con TIP3P. Además, para el conjunto de moléculas estudiado tiene una mayor influencia los parámetros no enlazantes más que el tipo de carga que se utilice en su determinación.



# Bibliografía

- [1] Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. *Chemical Reviews* **2015**, *115*, 6312–6356.
- [2] Mobley, D.; Wymer, K.; Lim, N.; Guthrie, J. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 135–150.
- [3] Muddana, H.; Sapra, N.; Fenley, A.; Gilson, M. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 277–287.
- [4] Mobley, D.; Guthrie, J. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 711–720.
- [5] Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.
- [6] MacKerell, A. D. et al. *The Journal of Physical Chemistry B* **1998**, *102*, 3586–3616.
- [7] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- [8] Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *Journal of Computational Chemistry* **2004**, *25*, 1656–1676.

- [9] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- [10] Besler, B. H.; Merz, K. M.; Kollman, P. A. *Journal of Computational Chemistry* **1990**, *11*, 431–439.
- [11] Singh, U. C.; Kollman, P. A. *Journal of Computational Chemistry* **1984**, *5*, 129–145.
- [12] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671–690.
- [13] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.
- [14] Jorgensen, W. L.; Tirado-Rives, J. *Journal of the American Chemical Society* **1988**, *110*, 1657–1666.
- [15] Ben-Naim, A. Y. *Solvation Thermodynamics*; Springer Science & Business Media, 2013.
- [16] Hess, B.; van der Vegt, N. F. A. *J. Phys. Chem. B* **2006**, *110*, 17616–17626.
- [17] Kollman, P. *Chemical Reviews* **1993**, *93*, 2395–2417.
- [18] König, G.; Boresch, S. *J. Phys. Chem. B* **2009**, *113*, 8967–8974.
- [19] Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

- [20] Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. *J. Comput. Aided Mol. Des.* **2012**, *26*, 551–562.
- [21] Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
- [22] Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
- [23] Shivakumar, D.; Deng, Y.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 919–930.
- [24] Takahashi, H.; Omi, A.; Morita, A.; Matubayasi, N. *J. Chem. Phys.* **2012**, *136*, 214503–214503.
- [25] Wolf, M. G.; Groenhof, G. *J. Comput. Chem.* **2012**, *33*, 2225–2232.
- [26] Ben-Naim, A. Y. *Molecular Theory of Solutions*; Oxford University Press: New York, 2006.
- [27] Gong, S. L.; Huang, P.; Zhao, T. L.; Sahuvar, L.; Barrie, L. A.; Kaminski, J. W.; Li, Y. F.; Niu, T. *Atmos. Chem. Phys.* **2007**, *7*, 4001–4013.
- [28] Vallack, H. W. et al. *Environmental Toxicology and Pharmacology* **1998**, *6*, 143 – 175.
- [29] Valsaraj, K. T.; Thibodeaux, L. J. *The Journal of Physical Chemistry Letters* **2010**, *1*, 1694–1700.
- [30] Avdeef, A.; Box, K.; Comer, J.; Hibbert, C.; Tam, K. *Pharmaceutical Research* **1998**, *15*, 209–215.
- [31] Hughes, L. D.; Palmer, D. S.; Nigsch, F.; Mitchell, J. B. O. *Journal of Chemical Information and Modeling* **2008**, *48*, 220–232.

- [32] Fujita, T.; Iwasa, J.; Hansch, C. *Journal of the American Chemical Society* **1964**, *86*, 5175–5180.
- [33] Bamford, H. A.; Baker, J. E.; Poster, D. L. *Review of methods and measurements of selected hydrophobic organic contaminant aqueous solubilities, vapor pressures, and air-water partition coefficients*; 1998.
- [34] Guthrie, J. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 151–168.
- [35] Ratkova, E. L.; Palmer, D. S.; Fedorov, M. V. *Chemical Reviews* **2015**, *115*, 6312–6356.
- [36] Guthrie, J. P. *The Journal of Physical Chemistry B* **2009**, *113*, 4501–4507.
- [37] Geballe, M.; Skillman, A.; Nicholls, A.; Guthrie, J.; Taylor, P. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 259–279.
- [38] Geballe, M.; Guthrie, J. *Journal of Computer-Aided Molecular Design* **2012**, *26*, 489–496.
- [39] Bannan, C. C.; Burley, K. H.; Chiu, M.; Shirts, M. R.; Gilson, M. K.; Mobley, D. L. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 927–944.
- [40] Rustenburg, A. S.; Dancer, J.; Lin, B.; Feng, J. A.; Ortwine, D. F.; Mobley, D. L.; Chodera, J. D. *Journal of Computer-Aided Molecular Design* **2016**, *30*, 945–958.
- [41] Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. *Journal of Computer-Aided Molecular Design* **2017**, *31*, 1–19.
- [42] Sandberg, L. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 211–219.

- [43] Ellingson, B.; Geballe, M.; Wlodek, S.; Bayly, C.; Skillman, A.; Nicholls, A. *Journal of Computer-Aided Molecular Design* **2014**, *28*, 289–298.
- [44] Fennell, C. J.; Wymer, K. L.; Mobley, D. L. *The Journal of Physical Chemistry B* **2014**, *118*, 6438–6446.
- [45] Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry A* **1998**, *102*, 1820–1831.
- [46] Mobley, D. L.; Dumont, ; Chodera, J. D.; Dill, K. A. *The Journal of Physical Chemistry B* **2007**, *111*, 2242–2254, PMID: 17291029.
- [47] Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *Journal of Medicinal Chemistry* **2008**, *51*, 769–779, PMID: 18215013.
- [48] Mobley, D. L.; Dill, K. A.; Chodera, J. D. *The Journal of Physical Chemistry B* **2008**, *112*, 938–946, PMID: 18171044.
- [49] Paliwal, H.; Shirts, M. R. *Journal of Chemical Theory and Computation* **2013**, *9*, 4700–4717, PMID: 26583389.
- [50] Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. *Journal of Computer-Aided Molecular Design* **2015**, *29*, 397–411.
- [51] Klimovich, P. V.; Mobley, D. L. *Journal of Computer-Aided Molecular Design* **2015**, *29*, 1007–1014.
- [52] Cox, S. R.; Williams, D. E. *Journal of Computational Chemistry* **1981**, *2*, 304–323.

- [53] Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *The Journal of Physical Chemistry* **1993**, *97*, 10269–10280.
- [54] Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. *The Journal of Chemical Physics* **2007**, *126*, 144111.
- [55] Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. *Journal of Chemical Theory and Computation* **2016**, *12*, 3894–3912, PMID: 27385073.
- [56] Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. *Journal of Chemical Theory and Computation* **2013**, *9*, 2221–2225, PMID: 26583716.
- [57] Verstraelen, T.; Sukhomlinov, S. V.; Van Speybroeck, V.; Waroquier, M.; Smirnov, K. S. *The Journal of Physical Chemistry C* **2012**, *116*, 490–504.
- [58] Manz, T. A.; Sholl, D. S. *Journal of Chemical Theory and Computation* **2010**, *6*, 2455–2468, PMID: 26613499.
- [59] Vöhringer-Martinez, E.; Verstraelen, T.; Ayers, P. W. *The Journal of Physical Chemistry B* **2014**, *118*, 9871–9880.
- [60] Ashbaugh, H. S.; Paulaitis, M. E. *The Journal of Physical Chemistry* **1996**, *100*, 1900–1913.
- [61] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- [62] Sutcliffe, B. T.; Woolley, R. G. *The Journal of Chemical Physics* **2012**, *137*, 22A544–1–22A544–8.



- [63] Mayer, I. *Simple proofs, theorems and derivations in quantum chemistry.*; Mathematical and computational chemistry.; Springer Science+Business Media, 2003.
- [64] Parr, R.; Weitao, Y. *Density-Functional Theory of Atoms and Molecules*; International Series of Monographs on Chemistry; Oxford University Press, 1989.
- [65] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.
- [66] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385–390.
- [67] Dirac, P. A. M. *Mathematical Proceedings of the Cambridge Philosophical Society* **1930**, *26*, 376–385.
- [68] Vosko, S. H.; Wilk, L.; Nusair, M. *Canadian Journal of Physics* **1980**, *58*, 1200–1211.
- [69] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- [70] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- [71] Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- [72] Gunsteren, W. F. V.; Berendsen, H. J. C. *Molecular Simulation* **1988**, *1*, 173–185.
- [73] Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.
- [74] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *The Journal of Chemical Physics* **1982**, *76*, 637–649.
- [75] Hirshfeld, F. *Theoretica chimica acta* **1977**, *44*, 129–138.

- [76] Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- [77] Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361–2366.
- [78] Bennett, C. H. *Journal of Computational Physics* **1976**, *22*, 245 – 268.
- [79] Shirts, M. R.; Pande, V. S. *The Journal of Chemical Physics* **2005**, *122*.
- [80] Lu, N.; Singh, J. K.; Kofke, D. A. *The Journal of Chemical Physics* **2003**, *118*, 2977–2984.
- [81] Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
- [82] Jarzynski, C. *Journal of Statistical Mechanics: Theory and Experiment* **2004**, P09005.
- [83] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 2447–2464.
- [84] Curutchet, C.; Cramer, C. J.; Truhlar, D. G.; Ruiz-Lopez, M. F.; Orozco, M.; Luque, F. J. *J. Comput. Chem.* **2003**, *24*, 284–297.
- [85] Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200–206.
- [86] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- [87] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- [88] Neese, F. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 73–78.

- [89] Verstraelen, T.; Tecmer, P.; Heidar-Zadeh, F.; Boguslawski, K.; Chan, M.; Zhao, Y.; Kim, T. D.; Vandenbrande, S.; Yang, D.; González-Espinoza, C. E.; Fias, S.; Limacher, P. A.; Berrocal, D.; Malek, A.; Ayers, P. W. HORTON 2.0.0. 2015; <http://theochem.github.com/horton/>, Accessed: January 25, 2018.
- [90] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, 1-2, 19–25.
- [91] Berendsen, H.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, 91, 6269–6271.
- [92] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. *J. Chem. Phys.* **1983**, 79, 926–935.
- [93] Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, 1, 173–185.
- [94] Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, 52, 7182–7190.
- [95] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, 103, 8577–8593.
- [96] Hess, B.; Bekker, H.; Berendsen, H.; Fraaije, J. *J. Comput. Chem.* **1997**, 18, 1463–1472.
- [97] Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, 111, 13052–13063.
- [98] Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994**, 222, 529–539.
- [99] Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. *J. Comput. Aided Mol. Des.* **2015**, 29, 397–411.

- [100] Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105–124105.
- [101] Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. *Journal of Chemical Theory and Computation* **2015**, *11*, 4377–4394.
- [102] Fenley, A. T.; Muddana, H. S.; Gilson, M. K. *Proceedings of the National Academy of Sciences* **2012**, *109*, 20006–20011.
- [103] Pluta, T.; Kolaski, M.; Medved, M.; Budzák, Š. *Chem. Phys. Lett.* **2012**, *546*, 24–29.
- [104] Cerutti, D. S.; Rice, J. E.; Swope, W. C.; Case, D. A. *The Journal of Physical Chemistry B* **2013**, *117*, 2328–2338.
- [105] Jakalian, A.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.
- [106] Roseman, M. A. *J. Mol. Biol.* **1988**, *200*, 513 – 522.
- [107] König, G.; Bruckner, S.; Boresch, S. *Biophys. J.* **2013**, *104*, 453–462.

## A. Anexo: Energía libre de hidratación para una variedad de grupos funcionales

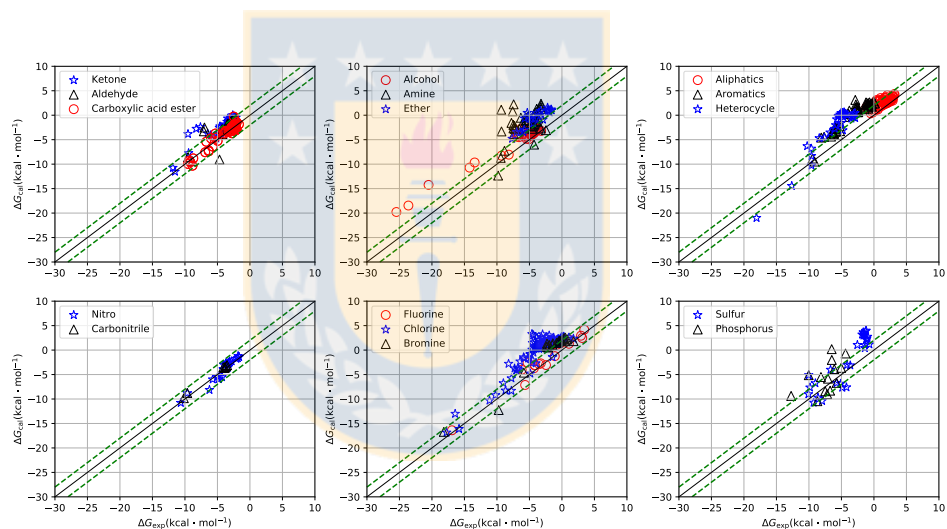


Figura A-1.: Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas S-HI (para los sulfatos y fosfatos se usó el método de partición Hirshfeld-E).

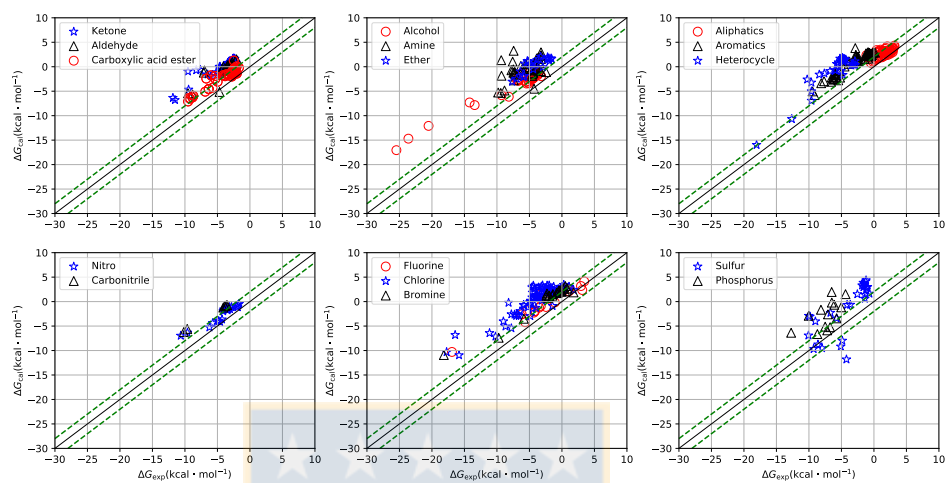


Figura A-2.: Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas semipolarizadas.

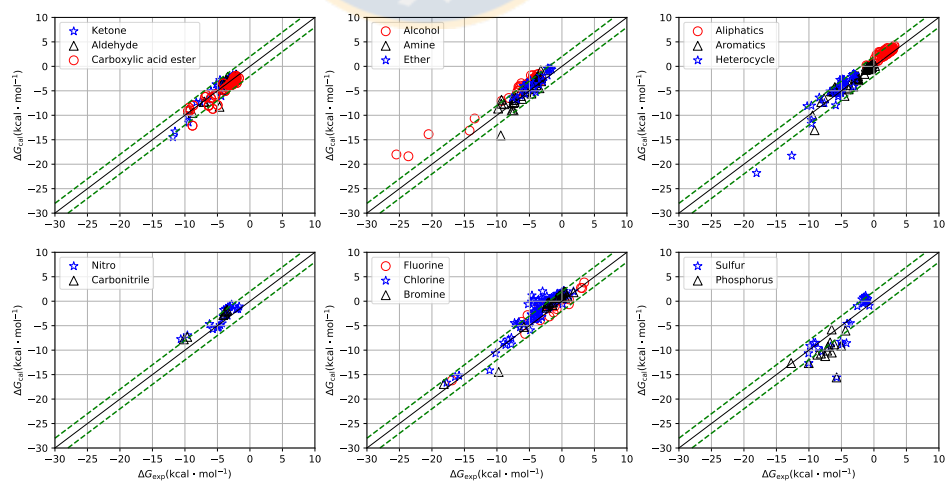


Figura A-3.: Energías libres de hidratación separadas por grupos funcionales para la base de datos FreeSolv obtenida con cargas atómicas AM1-BCC.