



Universidad de Concepción
Facultad de Ciencias Químicas
Doctorado en Ciencias con Mención en Química

**CAMPOS DE FUERZA DERIVADOS DE LA MECÁNICA
CUÁNTICA PARA LA PREDICCIÓN DE ESPECIFICIDAD Y
SELECTIVIDAD EN COMPLEJOS PROTEÍNA-LIGANDO**

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

Por: Luis Alejandro Macaya Zapata

Profesor Guía: Dr. Esteban Vöhringer-Martinez

Agosto 2025
Concepción, Chile

©Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

*"The flower that blooms in adversity
is the most rare and beautiful of all."*

The Emperor, Mulan.

Agradecimientos

En este cierre de una etapa tan significativa, quiero expresar mi más profundo agradecimiento a la Agencia Nacional de Investigación y Desarrollo (ANID), cuyo apoyo, a través de la Beca de Doctorado Nacional, hizo posible la realización de estos estudios. También agradezco a la Facultad de Ciencias Químicas por brindarme una formación rigurosa y de excelencia que me permitió crecer tanto académica como personalmente.

Agradezco con especial cariño a todos los miembros del grupo de Química Cuántica y Modelamiento Molecular (QCMM). Más allá del trabajo compartido, me brindaron compañía, apoyo y amistad durante mi paso por el laboratorio. Mi gratitud más sincera al Dr. Esteban Vöhringer-Martinez, por su guía, paciencia y confianza a lo largo de este desafiante y enriquecedor camino.

A mi querida esposa y compañera Fran, gracias por ser mi sostén constante. Junto con LeChuck y Dona, se convirtieron en parte esencial de mi vida, brindándome motivación diaria y amor incondicional para alcanzar esta meta. Les amo con todo mi corazón. A mi padre Luis, mi madre Alejandra y mi madrina Eliana, junto a mis hermanas Karina y Carolina, quienes, a pesar de mis erráticas agendas, nunca dejaron de acompañarme. Gracias por su cariño

inquebrantable. Les quiero infinitamente. A mis queridos suegros y cuñados, que a la distancia me apoyaron constantemente. A mis abuelas y abuelos, que hoy me acompañan en espíritu y a quienes llevo siempre en mi corazón. Y a mis amigos y amigas, gracias por estar presentes en los momentos buenos y por escucharme en los difíciles, recordándome siempre la importancia de la compañía y la verdadera amistad.

Este logro es tanto mío como de todos y todas ustedes.

Contenido

Lista de Figuras	VI
Lista de Tablas	XI
Resumen	XV
Abstract	XVIII
1. Introducción	1
1.1. Marco Teórico	9
1.1.1. Campos de fuerza	9
1.1.2. Esquemas de partición de la densidad electrónica . .	13
1.1.3. Modelo mecánico-estadístico de la unión proteína- ligando	15
1.1.4. Estimadores de energía libre	19
2. Hipótesis y Objetivos	26
2.1. Hipótesis	26
2.2. Objetivo general	26
2.3. Objetivos Específicos	27

3. Metodología	28
3.1. Métodos de estructura electrónica	28
3.1.1. Método de Kohn-Sham	29
3.1.2. Funcional de intercambio-correlación	30
3.1.3. Funciones base	31
3.2. Mecánica molecular	32
3.2.1. Método híbrido QM/MM	33
3.2.2. Cargas atómicas MBIS	35
3.2.3. Cargas atómicas D-MBIS	36
3.2.4. Parámetros de Lennard-Jones D-MBIS	37
3.3. Dinámica molecular	40
3.4. Cálculos de energía libre de unión absoluta	41
3.4.1. Ciclo termodinámico	42
3.4.2. Corrección por polarización	47
3.5. Descriptores estadísticos	49
4. Resultados y Discusión	52
4.1. Herramienta computacional para la derivación de parámetros no enlazantes	54
4.2. Especificidad en BRD4(1)	66
4.3. Perfil de selectividad de Bromosporina	86
4.4. Especificidad en lisozima T4 L99A/M102Q	114

5. Conclusiones	130
5.1. Proyecciones	135
Bibliografía	138
A. Anexo - Especificidad en BRD4(1)	151
B. Anexo - Perfil de selectividad de Bromosporina	158
C. Anexo - Especificidad en Lisozima T4 L99A/M102Q	165
D. Presentaciones en congresos y publicaciones	166
D.1. Congresos	166
D.2. Publicaciones	167

Lista de Figuras

1.1.	Conjunto de datos BRD4(1).	5
1.2.	Conjunto de datos Bromosporina-Bromodominios.	6
1.3.	Conjunto de datos Lisozima T4 L99A/M102Q.	7
3.1.	Esquema del ciclo termodinámico utilizado para el cálculo de la energía libre de unión absoluta.	43
3.2.	Esquema del cálculo de energía libre mediante simulaciones de no equilibrio.	46
3.3.	Esquema del ciclo termodinámico utilizado para el cálculo de energía libre de unión absoluta con corrección por polarización.	48
4.1.	Diagrama de flujo computacional utilizado para derivar los parámetros no enlazantes en ambientes moleculares explícitos.	56
4.2.	Evaluación exhaustiva de variables utilizadas en el protocolo de derivación de la herramienta computacional.	65
4.3.	Características estructurales de bromodominios.	67
4.4.	Gráfico de paridad de valores de energía libre de unión absoluta obtenidos en este estudio para el conjunto de datos BRD4(1), respecto a los publicados por Gapsys y colaboradores.	72
4.5.	Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y cargas atómicas AM1-BCC para el conjunto de datos BRD4(1).	74
4.6.	Interacciones de los ligandos 8 y 6 en el sitio de unión de BRD4(1).	78
4.7.	Ciclo esquemático de la metodología <i>Double-System/Single-box</i> (DSSB)	80

4.8. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0, sin incluir la estimación de ligandos con carga (1 y 4).	83
4.9. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 modificado, sin incluir la estimación de ligandos con carga (1 y 4).	84
4.10. Estructura cristalina resuelta para el complejo formado por bromosporina y el bromodominio BRD9.	87
4.11. Gráfico de paridad de valores de energía libre de unión absoluta obtenidos en este estudio para el conjunto de datos de bromosporina, respecto a los publicados por Gapsys y colaboradores	91
4.12. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y cargas atómicas AM1-BCC en el conjunto de datos de bromosporina.	93
4.13. Potencial electrostático de superficie de los 22 bromodominios en estudio, en un rango de valores de -5 a +5 kT/e. La orientación mostrada corresponde a una vista superior del sitio de unión de bromosporina	95
4.14. Valores de diferencia de las cargas atómicas D-MBIS entre los bromodominios seleccionados.	96
4.15. Valores de parámetros de Lennard-Jones (σ y ϵ) del campo de fuerza OFF-2.0.0 y los obtenidos en ambientes químicos de solvente y complejo proteína-ligando de los bromodominios seleccionados.	97
4.16. Valores de diferencia de las cargas atómicas D-MBIS entre los bromodominios seleccionados, usando parámetros de Lennard-Jones derivados desde el complejo y en solvente.	98
4.17. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS (cargas atómicas promedio de bromodominios y parámetros de Lennard-Jones derivados desde solvente).	100

4.18. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS (cargas atómicas por familia de bromodominios y parámetros de Lennard-Jones derivados desde solvente).	103
4.19. Estructuras cristalinas de complejos bromosporina-bromodominios evaluados.	105
4.20. Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 con parámetros no enlazantes D-MBIS (cargas atómicas por familia de bromodominios y parámetros de Lennard-Jones derivados desde solvente) usando distintas estructuras iniciales.	107
4.21. Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS, usando distintos estados conformacionales de las proteínas en el protocolo de transformaciones de no equilibrio.	109
4.22. Desglose de la rama proteica del ciclo termodinámico utilizado para los cálculos de integración termodinámica de no equilibrio	110
4.23. Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS, usando distintos estados conformacionales de las proteínas en el protocolo de transformaciones de no equilibrio, en conjunto del factor de corrección conformacional.	112
4.24. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y cargas atómicas AM1-BCC.	118
4.25. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y reemplazo de parámetros no enlazantes derivados (D-MBIS).	119
4.26. Evaluación de métricas de distancia para los distintos modos de unión de L4.	121
4.27. Evaluación de métricas estructurales para los distintos modos de unión de L6.	122

4.28. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 con restricciones de posición del ligando.	123
4.29. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y parámetros no enlazantes D-MBIS, aplicando restricciones de posición de ligando.	125
4.30. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y parámetros no enlazantes D-MBIS, usando restricciones de posición del ligando y selección de configuraciones de L8 que presentan interacciones de puente de hidrógeno con Q102.	127
A.1. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.5 y cargas atómicas AM1-BCC para el conjunto de datos BRD4(1).	153
A.2. Parámetros de Lennard-Jones de ligandos 6 y 8.	154
A.3. Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 8.	154
A.4. Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 1.	155
A.5. Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 4.	155
A.6. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 para el conjunto de datos BRD4(1), con los valores obtenidos para los ligandos 1 y 4 mediante la metodología DSSB.	156
A.7. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.5, sin incluir la estimación de ligandos con carga (1 y 4).	157
B.1. Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.7 y cargas atómicas AM1-BCC para el conjunto de datos de bromosporina.	160

B.2.	Cambio relativo porcentual de valores de parámetros de Lennard-Jones (σ y ϵ) del campo de fuerza OFF-2.0.0 y los obtenidos en ambientes químicos de solvente y complejo proteína-ligando de los bromodominios seleccionados. . . .	161
B.3.	Gráfico de paridad de valores de energía libre de unión calculados para los distintos bromodominios utilizados en la derivación de parámetros por familia con distintos conjuntos de cargas atómicas.	162
B.4.	Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 usando parámetros no enlazantes D-MBIS.	164

Lista de Tablas

4.1.	Valores de energía libre de unión experimental y calculados para el campo de fuerza OFF-2.0.0 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol ⁻¹	76
4.2.	Valores de energía libre de unión experimental y calculados para el campo de fuerza OFF-2.0.0 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol ⁻¹	104
4.3.	Valores de energía libre de unión experimental y calculados para el campo de fuerza GAFF-1.8 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol ⁻¹	129
A.1.	Conjunto de datos BRD4(1).	151
A.2.	Energía necesaria para polarizar los ligandos (1-11) en diferentes entornos moleculares para el conjunto de datos BRD4(1).152	
B.1.	Conjunto de datos Bromosporina-bromodominios.	159
B.2.	Energía necesaria para polarizar el ligando bromosporina en diferentes entornos moleculares para el conjunto de datos de bromodominios.	161
B.3.	Energía necesaria para polarizar el ligando bromosporina en diferentes entornos moleculares para el conjunto de datos de bromodominios.	162
B.4.	Conjunto de datos para complejos cristalizados de bromosporina.	163

C.1. Energía necesaria para polarizar los ligandos (L1-8) en diferentes entornos moleculares para el conjunto de datos de lisozima T4 L99A/M102Q.	165
---	-----

Resumen

La industria farmacéutica ha buscado de manera constante acelerar y abaratar el descubrimiento de nuevas moléculas activas, apoyándose crecientemente en métodos computacionales para la predicción y validación de compuestos. Entre ellos destacan la estimación de la energía libre de unión absoluta, así como la evaluación de especificidad y selectividad en complejos proteína-ligando como los más desafiantes. Para estos fines se han empleado metodologías de acoplamiento molecular, cálculos de punto final y enfoques termodinámicos basados en transformaciones alquímicas o geométricas, capaces de describir con alta exactitud los principios físico-químicos que gobiernan la formación de complejos moleculares. En particular, los métodos de perturbación alquímica y de no equilibrio han demostrado gran potencial por su exactitud, capacidad de paralelización y mejor muestreo conformacional. Sin embargo, persisten limitaciones asociadas a la adecuada descripción de interacciones intra- e intermoleculares, usualmente modeladas mediante campos de fuerza clásicos no polarizables.

En este contexto, la parametrización de moléculas pequeñas representa un desafío central, especialmente para la asignación precisa de parámetros no enlazantes en ambientes químicos diversos. Recientes avances, como los pro-

movidos por la iniciativa cooperativa y de ciencia abierta Open Force Field, han aprovechado datos de mecánica cuántica de alto nivel en la generación de mejores campos de fuerza. Por otro lado, metodologías de partición de densidad electrónica desarrollados en nuestro laboratorio, han permitido mejorar la asignación de cargas atómicas y parámetros de Lennard-Jones, obteniendo propiedades atómicas desde primeros principios.

Sobre esta base, en esta tesis se desarrolló la herramienta de código abierto `ffparaim`, destinada a la derivación automatizada de parámetros no enlazantes mediante el protocolo D-MBIS. Este enfoque incorpora el entorno molecular explícito durante la partición de la densidad electrónica, generando parámetros más representativos para campos de fuerza no polarizables.

Las simulaciones realizadas mostraron mejoras significativas en la predicción de energías libres de unión absoluta en sistemas modelo de referencia. En el conjunto de ligandos del sistema BRD4(1) se obtuvo un error cuadrático medio inferior a $1 \text{ kcal}\cdot\text{mol}^{-1}$ en ligandos neutros, mientras que la evaluación de selectividad de bromosporina en distintos bromodominios y el sistema modelo lisozima T4 L99A/M102Q dieron en manifiesto la importancia crítica de factores adicionales, como el cambio conformacional del receptor y la correcta definición de la pose inicial del ligando. Corrigiendo estos factores se logró alcanzar predicciones dentro de la exactitud química.

En conjunto, los resultados validan la utilidad de la herramienta computacional generada en esta tesis y de los parámetros no enlazantes derivados

mediante la metodología D-MBIS como una estrategia sólida para aumentar la exactitud de campos de fuerza no polarizables en la predicción de afinidad. Estos sientan las bases para el desarrollo de metodologías de simulaciones de dinámica molecular con campos de fuerza a la medida y estrategias de diseño de fármacos con un enfoque más fundamentado en principios físicos.

Abstract

The pharmaceutical industry has consistently sought to accelerate and reduce the cost of discovering new active molecules, increasingly relying on computational methods for compound prediction and validation. Among these, the estimation of absolute binding free energy, as well as the assessment of specificity and selectivity in protein–ligand complexes, stand out as most challenging. To this end, molecular docking, endpoint free energy calculations, and thermodynamic approaches based on alchemical or geometric transformations have been employed, as they are capable of accurately describing the physicochemical principles governing the formation of molecular complexes. In particular, alchemical perturbation and nonequilibrium methods have shown great potential due to their accuracy, parallelization capability and improved conformational sampling. However, limitations persist in the proper description of intra- and intermolecular interactions, which are usually modeled using non-polarizable classical force fields.

In this context, the parameterization of small molecules represents a central challenge, especially for the precise assignment of nonbonded parameters in diverse chemical environments. Recent advances, such as those promoted by the cooperative and open-science Open Force Field initiative, have leve-

arged high-level quantum mechanical data in the generation of better force fields. On the other hand, electronic density partitioning methodologies developed in our laboratory have enabled improvements in the assignment of atomic charges and Lennard-Jones parameters, deriving atomic properties from first principles.

On this basis, this thesis developed the open-source tool `ffparaim`, aimed at the automated derivation of nonbonded parameters through the D-MBIS protocol. This approach incorporates the explicit molecular environment during electronic density partitioning, generating more representative parameters for non-polarizable force fields.

The simulations performed showed significant improvements in the prediction of absolute binding free energies in reference model systems. In the BRD4(1) ligand set, a root-mean-square error below $1 \text{ kcal}\cdot\text{mol}^{-1}$ was achieved for neutral ligands, while the evaluation of Bromosporine selectivity across different bromodomains and the model system T4 Lysozyme L99A/M102Q highlighted the critical importance of additional factors, such as receptor conformational changes and the proper definition of the initial ligand pose. By correcting for these factors, predictions within chemical accuracy were achieved.

Taken together, the results validate the usefulness of the computational tool developed in this thesis and of the nonbonded parameters derived through the D-MBIS methodology as a solid strategy to enhance the accuracy

of non-polarizable force fields in affinity prediction. These findings lay the groundwork for the development of molecular dynamics simulation methodologies with bespoke force fields and drug design strategies more firmly grounded in physical principles.

1. Introducción

Dadas las exigencias de la industria farmacéutica actual (reducción de costos, rapidez de respuesta, mejorar la tasa de éxito en el desarrollo de nuevos compuestos, entre otros), se requieren métodos de diseño de fármacos más eficientes que permitan reducir los costos de las etapas tempranas de exploración de afinidad, así como aumentar la tasa de éxito en el descubrimiento y desarrollo de nuevas moléculas con actividad farmacológica¹.

Una de las aproximaciones más empleadas en las etapas tempranas es el acoplamiento molecular o *docking*², en conjunto con funciones de puntuación³, que permiten predecir modos de unión plausibles y estimar la afinidad de interacción. Esta metodología, ampliamente utilizada por su equilibrio entre precisión y eficiencia², presenta sin embargo limitaciones en su capacidad discriminativa cuando las diferencias entre compuestos son menores a un orden de magnitud en contextos de constantes de afinidad⁴.

Como respuesta a estas limitaciones, los cálculos de energía libre aplicados a complejos proteína-ligando han representado un avance significativo en el campo del diseño de fármacos asistido por computadora⁵. Estos métodos basados en principios físicos integran tanto efectos mecánico-estadísticos (como entropía conformacional, promedio sobre múltiples conformaciones o

modos de unión, y efectos del solvente) como químicos (estado de protonación, tautomería), permitiendo una descripción más precisa de las fuerzas intermoleculares que rigen la formación del complejo receptor-ligando¹. Los cálculos de energía libre de unión relativa, usados ampliamente en etapas de optimización de moléculas líder, permiten comparar diferentes ligandos respecto a uno de referencia, exigiendo que compartan un motivo estructural común^{6,7}. La energía libre se calcula en función de la transformación de fragmentos del sistema A al sistema B, conectados mediante una coordenada de transformación λ , donde $\lambda = 0$ representa el sistema A y $\lambda = 1$ el sistema B⁸. Por otro lado, los cálculos de energía libre de unión absoluta consisten en desacoplar completamente el ligando de su entorno. Este enfoque requiere además la aplicación de restricciones que impidan la difusión del ligando fuera del sitio de unión en los estados desacoplados^{9,10}. A diferencia de los cálculos relativos, los métodos absolutos no requieren una estructura común entre los ligandos lo cual permite comparaciones directas con datos experimentales¹¹, son más adecuados para priorizar modos de unión individuales¹² o comparar afinidades entre distintos blancos terapéuticos¹³. Adicionalmente, las series congénicas también pueden evaluarse mediante cálculos absolutos, permitiendo una validación cruzada entre ambos enfoques. El principal desafío del método absoluto es la mayor perturbación del sistema durante el acoplamiento/desacoplamiento completo del ligando, lo que demanda mayores tiempos de muestreo para alcanzar la convergencia del valor de la energía

libre de unión¹⁴. Para superar estas limitaciones, se han aplicado metodologías basadas en simulaciones de no equilibrio¹⁵⁻¹⁷, aprovechando su alto paralelismo computacional y manteniendo rendimientos comparables a las simulaciones de equilibrio^{14,18-21}.

Además del muestreo asociado a la convergencia de la predicción de la energía libre, otro desafío relevante es la adecuada descripción de las interacciones moleculares²². Debido al número de átomos presentes en el complejo ligando/receptor solvatado, deben emplearse campos de fuerza que representen la superficie de energía potencial de un sistema de forma fidedigna²³. Este problema es más acentuado en moléculas químicamente diversas, donde los parámetros tradicionales a menudo no capturan con precisión el espacio configuracional ni predicen correctamente propiedades de observables macroscópicos²⁴. Los campos de fuerza no polarizables para moléculas pequeñas (al igual que para proteínas, lípidos, etc.) utilizan funciones aditivas para interacciones enlazantes (enlaces, ángulos, torsiones) y no enlazantes (electrostáticas y de van der Waals)²³. Algunos ejemplos incluyen GAFF^{25,26}, CGenFF²⁷, GROMOS²⁸ y OPLS3e/4^{29,30}, que asignan parámetros en función del entorno químico local de cada átomo. Una alternativa innovadora es la propuesta de la iniciativa *Open Force Field*³¹, que introduce una tipificación basada en subestructuras químicas mediante el formato SMIRNOFF, reduciendo drásticamente el número de parámetros necesarios y ampliando la cobertura del espacio químico relevante en diseño de fármacos^{32,33}. Su ver-

sión más reciente, Open Force Field (OFF) 2.0.0 Sage³⁴, incorpora mejoras basadas en datos experimentales y cálculos de química cuántica³⁵.

Aun así, persisten desafíos como la falta de transferibilidad de parámetros en grupos químicos poco representados, la compatibilidad entre campos de fuerza y la derivación consistente de parámetros³⁶. Frente a esto, han surgido herramientas como QUBEKit³⁷ y ParaMol³⁸, que generan parámetros directamente desde cálculos de mecánica cuántica^{39,40}. Asimismo, se exploran enfoques basados en aprendizaje automático para mejorar parámetros torsionales⁴¹, asignar cargas parciales⁴²⁻⁴⁴ o emplear redes neuronales para generar potenciales más precisos⁴⁵.

En particular, los métodos de partición de densidad electrónica⁴⁶ han demostrado utilidad en la derivación de parámetros no enlazantes, como cargas atómicas y parámetros de Lennard-Jones⁴⁷. Estos métodos constan en la descomposición a subsistemas atómicos de la molécula y la evaluación de la contribución de cada uno a la densidad total⁴⁶. Algunos ejemplos de su aplicación abarcan la predicción de energías libres de solvatación⁴⁸⁻⁵⁰, coeficientes de partición^{51,52}, reproducción de propiedades termofísicas^{39,53}, energías libres de unión en sistemas anfitrión-huésped incorporando efectos de polarización del ambiente químico^{54,55} y proteína-ligando^{40,56}.

Para evaluar el desempeño de estos campos de fuerza, es esencial contar con conjuntos de prueba respaldados por datos experimentales de alta calidad^{57,58} que permiten aplicar métricas estadísticas robustas⁵⁹.

Entre ellos destacan los bromodominios, lectores epigenéticos que reconocen motivos de acetilación en lisinas y están implicados en procesos celulares como remodelación de cromatina y transcripción^{60,61}. El conjunto BRD4(1) (Figura 1.1) es un sistema de prueba ampliamente utilizado debido a su diversidad molecular y la rigidez de su sitio de unión¹².

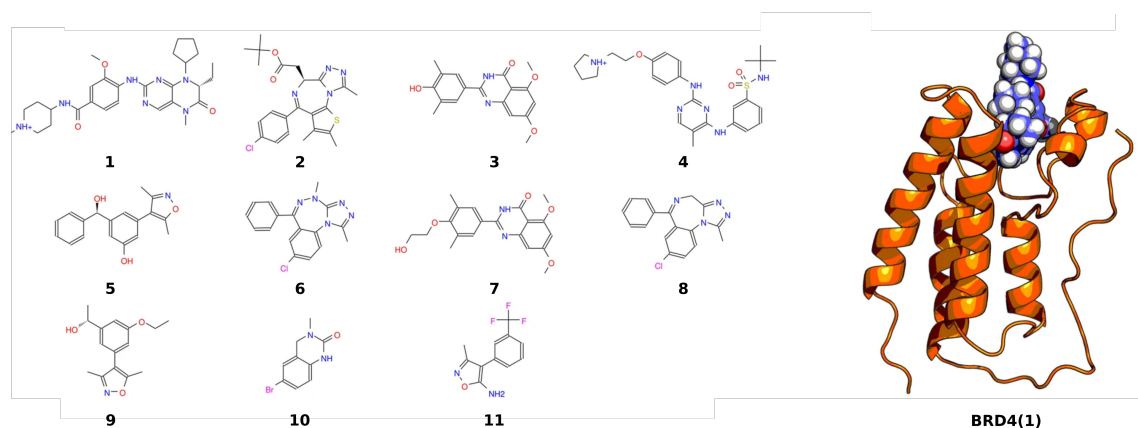


Figura 1.1: **Conjunto de datos BRD4(1)**. Imagen adaptada desde Gapsys y colaboradores¹⁴

En el diseño de fármacos no solo se debe considerar la afinidad, sino también la selectividad⁶² (preferencia por una diana específica). Estos conceptos son esenciales para evaluar y validar nuevas metodologías en etapas tempranas del desarrollo farmacológico. La bromosporina (Figura 1.2), que representa un pan-inhibidor, permite evaluar afinidad respecto a múltiples bromodominios, proporcionando una base para estudios del efecto de una molécula entre familias de estas proteínas¹³. Esto es de alta relevancia para aumentar la eficacia de potenciales desarrollos farmacéuticos, disminuir potenciales

efectos adversos por promiscuidad o dirigir el diseño racional basado en diferencias estructurales.

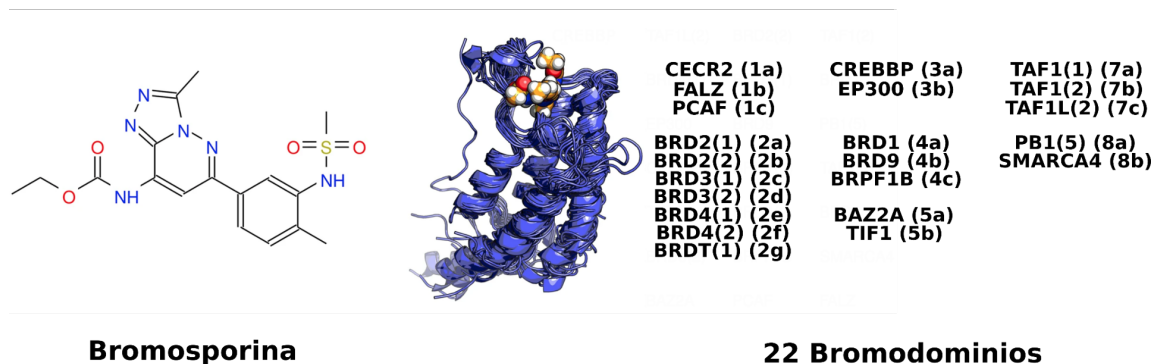


Figura 1.2: **Conjunto de datos Bromosporina-Bromodominios.** A la izquierda, se muestra una representación 2D de bromosporina. A la derecha, se muestran los 22 bromodominios y una superposición estructural de los receptores. Imagen adaptada desde Gapsys y colaboradores¹⁴

Otro conjunto relevante es la lisozima T4 L99A/M102Q⁶³ (Figura 1.3), una mutante diseñada mediante ingeniería genética, que cuenta con una cavidad polar capaz de formar enlaces de hidrógeno, lo que introduce una mayor complejidad conformacional respecto a su versión apolar L99A⁵⁷. Este sistema permite evaluar afinidad entre ligandos de menor masa molecular, estructuralmente similares pero con diferentes capacidades de unión, manteniendo la simplicidad estructural necesaria para estudios sistemáticos⁶⁴⁻⁶⁶.

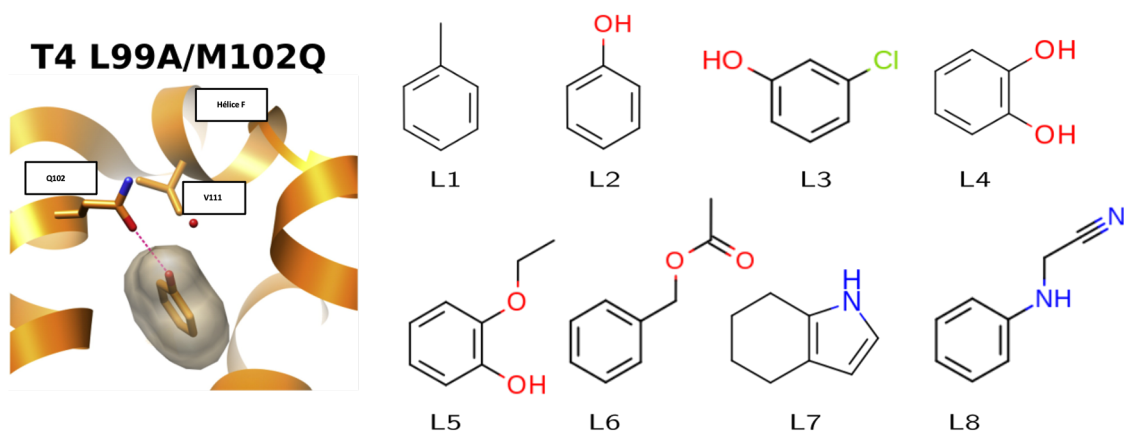


Figura 1.3: **Conjunto de datos Lisozima T4 L99A/M102Q**. Se muestra la mutación en el residuo 102 que permite la formación de enlaces de hidrógeno (en este ejemplo con fenol). Se indica la hélice F, que presenta el cambio conformacional asociado a la apertura de la cavidad, permitiendo la presencia de ligandos más voluminosos en el sitio. A la derecha, el conjunto de datos de moléculas evaluadas con afinidad reportada en literatura.

En este contexto, la presente tesis se centra en el desarrollo de una herramienta para la derivación de parámetros no enlazantes en campos de fuerza no polarizables, y en su evaluación crítica mediante metodologías de cálculo de energía libre de unión absoluta. Aprovechando los avances recientes en algoritmos de simulación y estrategias de parametrización molecular, esta investigación busca profundizar en cómo variables como la exploración conformacional, la calidad de los campos de fuerza y la naturaleza química del sistema afectan la exactitud y precisión de las predicciones físico-químicas, especialmente en escenarios donde la selectividad y especificidad resultan determinantes. A través del uso de sistemas modelo con datos experimentales de alta calidad, se propone validar enfoques alternativos de parametrización de

ligandos y establecer criterios sólidos para orientar la elección metodológica según los objetivos del estudio. De este modo, esta tesis aspira a ser no solo un ejercicio técnico de evaluación computacional, sino también una contribución significativa al desarrollo de herramientas más precisas, reproducibles y aplicables en las etapas iniciales del descubrimiento racional de fármacos.

1.1. Marco Teórico

El presente marco teórico describe los fundamentos físico-químicos que sustentan la hipótesis de trabajo, centrada en el uso de campos de fuerza clásicos, esquemas de partición de la densidad electrónica y métodos de estimación de energía libre. Estas herramientas se aplican a la predicción de propiedades termodinámicas relevantes en el contexto del descubrimiento de fármacos, como la afinidad de unión proteína-ligando.

1.1.1. Campos de fuerza

En simulaciones moleculares clásicas, las interacciones entre átomos se describen mediante funciones matemáticas conocidas como campos de fuerza, que permiten calcular la energía potencial total del sistema. Estos modelos se construyen combinando términos que representan interacciones atómicas de tipo covalente y no covalente.

La energía total del sistema se expresa típicamente como:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}} \quad (1.1)$$

donde E_{bonded} incluye los términos de enlaces, ángulos y torsiones, mientras que $E_{\text{non-bonded}}$ considera las interacciones de van der Waals y electrostáticas.

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{improper}} \quad (1.2)$$

$$E_{\text{non-bonded}} = E_{\text{elec}} + E_{\text{vdW}} \quad (1.3)$$

Los términos covalentes modelan interacciones entre átomos conectados directamente a través de enlaces. Estos se representan generalmente mediante funciones armónicas o periódicas:

- **Energía de enlace:**

$$E_{\text{bond}} = \sum_{\text{bonds}} k_r (r - r_0)^2 \quad (1.4)$$

donde k_r es la constante de fuerza del enlace y r_0 es la distancia de equilibrio.

- **Energía angular:**

$$E_{\text{angle}} = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \quad (1.5)$$

donde k_θ es la constante de fuerza angular y θ_0 el ángulo de equilibrio.

- **Energía de torsión:**

$$E_{\text{dihedral}} = \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \quad (1.6)$$

donde V_n es la barrera rotacional, n la periodicidad y γ el desfase.

- **Términos impropios:**

$$E_{\text{improper}} = \sum_{\text{impropers}} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (1.7)$$

empleados para mantener geometrías específicas, como la planaridad de anillos aromáticos o la estructura tetraédrica de centros quirales.

Las interacciones no covalentes representan fuerzas entre átomos que no están directamente conectados, descritas por fuerzas de van der Waals y electrostáticas.

- **Interacciones de van der Waals:**

Estas se modelan comúnmente mediante el potencial de Lennard-Jones:

$$E_{\text{vdW}} = \sum_{i < j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1.8)$$

donde ϵ_{ij} es la profundidad del pozo potencial y σ_{ij} la distancia a la que la energía entre los átomos i y j es cero. El término $\frac{1}{r^{12}}$ modela las

interacciones a corta distancia (repulsión de Pauli) y el término $\frac{1}{r^6}$ las interacciones de atracción (dispersión de London).

- **Interacciones electrostáticas:**

Modeladas mediante la ley de Coulomb:

$$E_{\text{elec}} = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1.9)$$

donde q_i y q_j son cargas parciales y r_{ij} la distancia entre los átomos. Dado su carácter de largo alcance, requieren técnicas especializadas como las sumas de Ewald o el método PME^{67,68} con el fin de disminuir el número de interacciones a calcular.

A pesar de su utilidad, los campos de fuerza clásicos presentan limitaciones al modelar sistemas con alta polarizabilidad, dado que suponen cargas atómicas fijas. En sistemas reales, la redistribución electrónica inducida por el entorno puede tener un papel importante. Esto ha motivado el desarrollo de campos de fuerza polarizables, que incluyen explícitamente términos que permiten la variación de las cargas atómicas en respuesta al ambiente. Algunos ejemplos destacados son AMOEBA⁶⁹, CHARMM-Drude⁷⁰ y SIBFA⁷¹. Estos modelos ofrecen una mayor precisión en la predicción de propiedades físico-químicas, aunque a costa de una mayor complejidad computacional y una cantidad significativa de parámetros adicionales.

La precisión de las simulaciones moleculares depende en gran medida de la calidad del campo de fuerza empleado, particularmente en la descripción de las interacciones no covalentes. Esto motiva el uso de esquemas de partición de la densidad electrónica, que abordamos en la siguiente subsección.

1.1.2. Esquemas de partición de la densidad electrónica

Los átomos constituyen los bloques fundamentales en la construcción de las moléculas y, de una forma u otra, las propiedades moleculares están directamente relacionadas con la identidad de sus átomos constituyentes. Sin embargo, no existe una definición universalmente aceptada de lo que representa un átomo dentro de una molécula⁷². Esta ambigüedad ha dado lugar a una proliferación de métodos para descomponer las moléculas en subsistemas atómicos, conocida como teoría de átomos-en-moléculas (AIM, por sus siglas en inglés).

Estos esquemas pueden clasificarse según el tipo de partición que realizan, ya sea dividiendo la función de onda^{73,74} o dividiendo descriptores moleculares en el espacio real, como la densidad electrónica⁷⁵.

Para estos últimos, su objetivo principal es minimizar una medida de disimilitud D entre las densidades electrónicas descritas para los átomos en la molécula ($\rho_A(\mathbf{r})$) y un conjunto de densidades pro-atómicas de referencia ($\rho_A^0(\mathbf{r})$):

$$\sum_{A=1}^{N_a} D[\rho_A, \rho_A^0] \quad (1.10)$$

Esta minimización se realiza bajo la restricción de que la suma de las densidades electrónicas atómicas debe reconstruir exactamente la densidad electrónica molecular ($\rho_{\text{mol}}(\mathbf{r})$):

$$\sum_{A=1}^{N_a} \rho_A(\mathbf{r}) = \rho_{\text{mol}}(\mathbf{r}) \quad (1.11)$$

Adicionalmente, se impone que la densidad electrónica de la pro-molécula⁷⁵, definida como la suma de las densidades pro-atómicas de referencia,

$$\rho_{\text{pro}}(\mathbf{r}) = \sum_{A=1}^{N_a} \rho_A^0(\mathbf{r}) \quad (1.12)$$

pueda optimizarse para aproximarse lo más posible a la densidad electrónica de la molécula completa.

A partir de este marco general, es posible imponer restricciones adicionales según el enfoque específico adoptado^{47,76}. Así, los distintos métodos de partición se diferencian principalmente en la elección de las densidades pro-atómicas de referencia, la definición de la medida de disimilitud y las restricciones aplicadas durante el procedimiento de minimización⁴⁶.

Los diferentes esquemas de partición permiten asignar cargas parciales

consistentes con la densidad electrónica molecular, facilitando una descripción de las interacciones electrostáticas a un menor costo computacional. Esta asignación de cargas es especialmente importante en estudios donde la energía libre asociada a la interacción proteína-ligando, juega un papel central. En la próxima subsección se presenta el marco mecánico-estadístico que permite cuantificar estas interacciones a nivel termodinámico.

1.1.3. Modelo mecánico-estadístico de la unión proteína-ligando

En esencia, la unión de un ligando a una proteína puede describirse como un equilibrio químico entre los estados no unidos de cada entidad y el complejo formado, representado por la constante de disociación K_d :

$$K_d = \frac{a_P a_L}{a_{PL}} = \frac{\gamma_P \gamma_L}{\gamma_{PL}} \cdot \frac{[P][L]}{[PL]} \cdot \frac{1}{C^\circ} \quad (1.13)$$

Aquí, a_{PL} denota la actividad del complejo proteína-ligando, mientras que a_P y a_L son las actividades de la proteína y el ligando libres, respectivamente. Estas actividades se expresan como el producto entre los coeficientes de actividad γ y las concentraciones, ajustadas por una concentración estándar C° . La energía libre de unión estándar $\Delta G_{\text{bind}}^\circ$ se relaciona con la constante de disociación a una temperatura dada mediante:

$$\Delta G_{\text{bind}}^{\circ} = RT \ln K_d = -RT \ln K_a \quad (1.14)$$

Sin embargo, esta ley de acción de masas no se aplica directamente en simulaciones moleculares típicas, donde usualmente se modela una sola copia de la proteína y el ligando en una caja de solvatación pequeña⁷⁷. En su lugar, una alternativa consiste en estimar directamente la constante de equilibrio a partir de simulaciones que obtengan muestras de múltiples eventos de formación y disociación del complejo. De esta forma, se puede calcular la probabilidad de encontrar el sistema en el estado unido, $P(PL)$, o no unido, $P(P+L)$.

Bajo el supuesto de que el cambio de volumen asociado a la unión es insignificante (como suele ocurrir a 1 atm en medios acuosos), la energía libre de Gibbs ΔG_{bind} se aproxima a la energía libre de Helmholtz ΔA_{bind} .

$$\Delta G_{\text{bind}} \simeq \Delta A_{\text{bind}} = -RT \left[\ln \left(\frac{P(PL)}{P(P+L)} \right) + \ln (C^{\circ} N_{\text{Av}} V) \right] \quad (1.15)$$

El término adicional que involucra la constante de Avogadro N_{Av} y el volumen de la caja V corrige el cálculo para obtener una energía libre a concentración estándar.

Definiendo Γ_{bound} y Γ_{unbound} como los subconjuntos del espacio de configuraciones \vec{q} que representan los estados unido y no unido del sistema, es

posible expresar la probabilidad de una configuración particular \vec{q} que sigue una distribución de Boltzmann:

$$P(\vec{q}) = \frac{e^{-\beta U(\vec{q})}}{\int_{\Gamma} e^{-\beta U(\vec{q})} d\vec{q}} \quad (1.16)$$

donde $\beta = \frac{1}{k_B T}$, $U(\vec{q})$ es la energía potencial de la configuración, y la integral se realiza sobre todo el espacio de configuraciones accesibles Γ . Bajo la hipótesis de ergodicidad, la fracción de configuraciones que se encuentran en el estado unido converge a:

$$P(PL) = \int_{\Gamma_{\text{bound}}} P(\vec{q}) d\vec{q} = \frac{\int_{\Gamma_{\text{bound}}} e^{-\beta U(\vec{q})} d\vec{q}}{\int_{\Gamma} e^{-\beta U(\vec{q})} d\vec{q}} \quad (1.17)$$

Una expresión análoga puede construirse para $P(P+L)$, y la proporción de ambas define la relación entre las regiones configuracionales correspondientes:

$$\frac{P(PL)}{P(P+L)} = \frac{\int_{\Gamma_{\text{bound}}} e^{-\beta U(\vec{q})} d\vec{q}}{\int_{\Gamma_{\text{unbound}}} e^{-\beta U(\vec{q})} d\vec{q}} = \frac{Z(PL)}{Z(P+L)} \quad (1.18)$$

donde Z representa la función de partición configuracional, cuya evaluación precisa constituye uno de los principales desafíos en el cálculo de la energía libre de unión.

Esta función puede reescribirse separando explícitamente los grados de libertad globales (traslaciones y rotaciones) de los internos (vibracionales y

conformacionales)⁹:

$$\begin{aligned}
Z &= \int dx_1 \cdots dx_{3N} e^{-\beta U(x_1, \dots, x_{3N})} \\
&= \int dX dY dZ d\xi_1 d\xi_2 d\xi_3 dx'_1 \cdots dx'_{3N-6} |\mathbf{J}| e^{-\beta U(x'_1, \dots, x'_{3N-6})} \\
&= 8\pi^2 V \int dx'_1 \cdots dx'_{3N-6} |\mathbf{J}| e^{-\beta U(x'_1, \dots, x'_{3N-6})} \\
&= 8\pi^2 V \tilde{Z}
\end{aligned} \tag{1.19}$$

En esta expresión, las coordenadas x_1, \dots, x_{3N} corresponden a las posiciones cartesianas de las N partículas del sistema. Estas pueden descomponerse en: X, Y, Z , que representan el centro de masa (traslación); ξ_1, ξ_2, ξ_3 , que definen su orientación (rotación); y x'_1, \dots, x'_{3N-6} , que representan los grados de libertad internos (vibraciones y conformaciones). El jacobiano $|\mathbf{J}|$ garantiza la conservación del volumen al cambiar de variables. El volumen V y el factor $8\pi^2$ provienen de la integración sobre traslaciones y rotaciones, respectivamente. Finalmente, \tilde{Z} representa la función de partición restringida a los grados de libertad internos, que son los que capturan la dinámica estructural del sistema.

La separación explícita de estos grados de libertad reduce significativamente la dimensionalidad del problema y permite enfocar el análisis en los componentes relevantes para la unión, como en la aplicación del algoritmo de enlace virtual Boresch-Karplus y la corrección analítica mediante restric-

ciones intermoleculares⁹. Esta descomposición es fundamental para métodos como las transformaciones alquímicas, que requieren una descripción precisa de los estados unidos y no unidos en el espacio conformacional^{78,79}.

Este marco estadístico proporciona la base teórica para entender y calcular la energía libre de unión entre un ligando y una proteína, lo cual es esencial en el diseño racional de fármacos. No obstante, dado que la evaluación directa de las funciones de partición es complejo computacionalmente, es necesario recurrir a métodos aproximados que permitan estimar diferencias de energía libre entre estados relevantes. En la siguiente subsección se abordan los métodos que hacen posible estas estimaciones.

1.1.4. Estimadores de energía libre

Como se presentó en la sección anterior, la energía libre de unión puede expresarse como una diferencia logarítmica entre funciones de partición configuracionales de los estados unido y no unido. Sin embargo, la evaluación directa de estas integrales multidimensionales resulta intratable para sistemas biomoleculares, debido al enorme número de grados de libertad involucrados. En consecuencia, se han desarrollado métodos aproximados basados en mecánica estadística que permiten estimar diferencias de energía libre a partir de trayectorias generadas por simulaciones.

En el extremo superior del espectro de métodos basados en principios físicos se encuentran los enfoques de perturbación alquímica⁸⁰, que permiten

obtener estimaciones rigurosas de energía libre mediante la transformación gradual de un sistema en otro, siguiendo un camino termodinámico definido. Uno de los primeros métodos propuestos fue la relación de Zwanzig⁸¹, que permite evaluar directamente la diferencia de energía libre entre dos estados a partir de una expresión exacta:

$$\Delta A = -\beta^{-1} \ln \left\langle e^{-\beta(U_1(\vec{q}) - U_0(\vec{q}))} \right\rangle_0 = -\beta^{-1} \ln \left\langle e^{-\beta \Delta U(\vec{q})} \right\rangle_0 \quad (1.20)$$

Esta ecuación se basa en el promedio exponencial de la diferencia de energía potencial entre dos estados, evaluado sobre una distribución generada en el estado de referencia. A pesar de ser una expresión exacta, presenta una eficiencia computacional limitada, ya que puede requerir una gran cantidad de muestras para converger adecuadamente, especialmente cuando existe poco solapamiento entre los espacios de fase de los estados involucrados⁸². Sin embargo, este método resulta apropiado cuando el espacio de fase del estado final es un subconjunto del estado inicial, como en el caso de la inserción de una molécula rígida en una fase condensada⁸³.

Otro método importante es la integración termodinámica (TI)⁸⁴, que estima la energía libre a partir del promedio de la derivada de la energía potencial respecto a una coordenada alquímica λ , la cual conecta de forma continua ambos estados:

$$\Delta A = \int_0^1 \left\langle \frac{\partial U(\lambda, \vec{q})}{\partial \lambda} \right\rangle_{\lambda} d\lambda \quad (1.21)$$

Este enfoque requiere calcular explícitamente la derivada de la energía potencial con respecto a λ , lo cual representa una limitación práctica, ya que no todos los paquetes de simulación implementan esta funcionalidad de forma nativa. Además, dicha derivada puede presentar inestabilidades numéricas, especialmente en los extremos del intervalo de integración, donde los potenciales pueden provocar solapamientos de átomos al acoplar o desacoplar interacciones, si no se utilizan potenciales suavizados⁸⁵. A pesar de estos desafíos, TI es uno de los métodos más precisos cuando se emplean funciones de interacción adecuadas, como los potenciales *soft-core*⁸⁶.

La relación de aceptación de Bennett (BAR)⁸⁷ fue uno de los primeros métodos capaces de estimar diferencias de energía libre utilizando información bidireccional, proveniente de múltiples estados. BAR se basa en comparar la diferencia de energía potencial $\Delta U_{ij}(\vec{q})$ entre dos estados i y j evaluada en configuraciones comunes. Dado que se considera la misma configuración para ambos estados, se puede establecer una relación exacta entre las distribuciones de dichas diferencias de energía⁸⁸:

$$\sum_{i=1}^{n_i} \frac{1}{1 + \exp \left[\ln \left(\frac{n_i}{n_j} \right) + \beta \Delta U_{ij}^{(i)} - \beta \Delta A \right]} = \sum_{j=1}^{n_j} \frac{1}{1 + \exp \left[\ln \left(\frac{n_j}{n_i} \right) - \beta \Delta U_{ji}^{(j)} + \beta \Delta A \right]} \quad (1.22)$$

Debido a que esta relación es una función exacta de las distribuciones de energía, permite aplicar métodos de optimización para minimizar la varianza de la estimación de energía libre y mejorar su precisión⁸².

Una extensión de este enfoque es el método BAR multiestado (MBAR)⁸⁹, que generaliza la estimación a múltiples estados intermedios. MBAR permite calcular la energía libre entre cualquier par de estados utilizando todas las muestras disponibles, lo cual mejora sustancialmente la eficiencia estadística. La energía libre estimada para el estado i , denotada como \hat{A}_i , se obtiene mediante la siguiente expresión:

$$\hat{A}_i = -\beta^{-1} \sum_{j=1}^K \sum_{n=1}^{N_j} \frac{e^{-\beta U_i(\vec{q}_{j,n})}}{\sum_{k=1}^K N_k e^{-\beta [U_k(\vec{q}_{j,n}) - \hat{A}_k]}} \quad (1.23)$$

Donde K es el número total de estados, N_j es el número de muestras no correlacionadas obtenidas del estado j , y $U_k(\vec{q}_{j,n})$ es la energía potencial evaluada en la configuración n de ese estado. Cabe destacar que sólo las diferencias entre energías libres ($\hat{A}_i - \hat{A}_j$) tienen significado físico. MBAR representa el estimador de mínima varianza para la energía libre dada la información de las muestras y proporciona estimaciones de incertidumbre altamente confiables en función del número de muestras disponibles por estado⁸⁹.

Los métodos anteriormente mencionados se basan en la estimación de la energía libre desde simulaciones en equilibrio, lo cual requiere que para cada valor de λ , el sistema debe encontrarse en equilibrio y el espacio con-

figuracional debe muestrearse completamente. Una estrategia distinta a las anteriormente mencionadas, involucra la medición de trabajo realizado en transiciones de no equilibrio⁹⁰⁻⁹². El trabajo de Jarzynski involucró la derivación de una identidad que relaciona el promedio exponencial del trabajo durante estas transiciones con la diferencia de energía libre en colectivos canónicos⁹³, donde luego Cuendet demostró que esto se cumplía también en el colectivo isotérmico-isobárico⁹⁴:

$$e^{-\beta\Delta G_{AB}} = \langle e^{-\beta W} \rangle \quad (1.24)$$

La igualdad de Jarzynski requiere que las transiciones comiencen desde configuraciones obtenidas a partir de colectivos en equilibrio. Estas transiciones, son evaluadas a intervalos de tiempo (τ) muy cortos. En base a lo anterior, los valores de trabajo entre dos estados pueden obtenerse desde la relación:

$$W[\tau] = \int_{\lambda(0)}^{\lambda(\tau)} \frac{\partial H}{\partial \lambda} d\lambda \quad (1.25)$$

donde a diferencia de TI, se integran los valores instantáneos de $\frac{\partial H(x(\lambda), \lambda)}{\partial \lambda}$. El trabajo resultante contiene tanto una contribución de la diferencia de energía libre como del trabajo disipado a lo largo de la transición irreversible. Su convergencia depende en gran medida de la superposición del espacio de fase entre los estados finales (similar a los métodos de equilibrio mencionados

anteriormente), ya que la precisión de las diferencias de energía libre calculadas con la fórmula de Jarzynski depende de eventos raros, donde se disipa poco trabajo⁹⁵. Respecto a los estimadores, una relación general derivada por Crooks que involucra las distribuciones de trabajo desde transiciones en ambos sentidos, dio paso al teorema de fluctuación de Crooks^{90,96,97}:

$$\frac{P_f(W)}{P_r(-W)} = e^{\beta(W-\Delta G)} \quad (1.26)$$

Para los casos en que la superposición entre las distribuciones de trabajo es grande, la diferencia de energía libre se puede estimar directamente:

$$\ln\left(\frac{P_f(W)}{P_r(-W)}\right) = \beta W - \Delta\hat{G} \quad (1.27)$$

Dado que un buen solapamiento se alcanza en regiones cercanas al equilibrio y sólo los valores de trabajo en la intersección de las distribuciones son utilizadas para la estimación, la expresión anterior no puede ser libremente aplicada⁹⁵. En base a estas problemáticas, Goette y Grubmüller mostraron una aproximación gaussiana para la estimación, definido como intersección Gaussiana de Crooks⁹⁸:

$$\Delta\hat{G} = \frac{\frac{\langle W_f \rangle_{nf}}{\sigma_f^2} - \frac{\langle W_r \rangle_{nr}}{\sigma_r^2} \pm \sqrt{\frac{1}{\sigma_f^2 \sigma_r^2} \left(\langle W_f \rangle_{nf} + \langle W_r \rangle_{nr} \right)^2 + 2 \left(\frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2} \right) \ln \left(\frac{\sigma_r^2}{\sigma_f^2} \right)}}{\frac{1}{\sigma_f^2} - \frac{1}{\sigma_r^2}} \quad (1.28)$$

El anteriormente mencionado estimador BAR también puede ser aplicado en simulaciones de no equilibrio, donde Shirts y colaboradores⁹⁹ proveen la siguiente expresión:

$$\sum_{i=1}^{n_f} \frac{1}{1 + \exp \left[\ln \left(\frac{n_f}{n_r} \right) + \beta (W_i - \hat{\Delta G}) \right]} = \sum_{j=1}^{n_r} \frac{1}{1 + \exp \left[\ln \left(\frac{n_r}{n_f} \right) - \beta (W_j - \hat{\Delta G}) \right]} \quad (1.29)$$

Esta formulación se basa en un criterio de máxima verosimilitud que equilibra la información estadística proveniente de ambos conjuntos de trayectorias. Su ventaja principal es que aprovecha eficientemente los datos de ambos conjuntos, minimizando la varianza del estimador bajo condiciones generales.

En conjunto, los métodos de perturbación química representan una de las estrategias más rigurosas y fundamentadas para la estimación de diferencias de energía libre en sistemas biomoleculares. A pesar de su demanda computacional y la necesidad de una cuidadosa planificación en la definición de los estados intermedios, su capacidad para capturar de manera precisa los efectos termodinámicos asociados a cambios moleculares los convierte en herramientas esenciales en el diseño racional de fármacos y en la evaluación cuantitativa de afinidades de unión.

2. Hipótesis y Objetivos

2.1. Hipótesis

Parámetros no enlazantes de campos de fuerza no polarizables, derivados desde la partición de la densidad electrónica en ambientes biomoleculares, mejoran la predicción de energía libre en complejos proteína-ligando de referencia mediante cálculos de energía libre de unión absoluta.

2.2. Objetivo general

Desarrollar una herramienta computacional para la obtención de parámetros no enlazantes derivados desde la partición de la densidad electrónica de moléculas pequeñas en ambientes biomoleculares y evaluar su capacidad de predicción de energía libre en complejos proteína-ligando de referencia mediante cálculos de energía libre de unión absoluta.

2.3. Objetivos Específicos

1. Crear una herramienta computacional para la derivación de parámetros no enlazantes desde la partición de la densidad electrónica de moléculas pequeñas en ambientes biomoleculares.
2. Evaluar el efecto sobre la especificidad de unión de distintos ligandos respecto a un receptor para la proteína 4 con bromodominio (BRD4(1))
3. Evaluar el efecto sobre la selectividad de unión de un ligando frente a distintos receptores en los sistemas bromosporina-bromodominios.
4. Evaluar el efecto sobre la especificidad de unión de distintos ligandos respecto a un receptor con distintos estados conformacionales apo y holo en el sistema lisozima T4 L99A/M102Q.

3. Metodología

En esta sección se muestran las principales metodologías empleadas en la investigación. Se desarrollarán aspectos generales de los métodos de estructura electrónica, mecánica molecular y su uso en simulaciones de dinámica molecular, en conjunto a la derivación de parámetros no enlazantes. Se realiza una breve descripción de la estimación de energía libre de unión en mediante transformaciones de no equilibrio, junto a la corrección por polarización utilizada en estas estimaciones.

3.1. Métodos de estructura electrónica

Los métodos de estructura electrónica se basan en la mecánica cuántica para describir explícitamente el comportamiento de electrones en sistemas moleculares, resolviendo la ecuación de Schrödinger independiente del tiempo:

$$\hat{H}\psi_i = E_i\psi_i \quad (3.1)$$

Bajo la aproximación de Born-Oppenheimer¹⁰⁰, se asume que los núcleos

permanecen fijos, lo que permite separar el movimiento electrónico y reducir la complejidad del problema al considerar la repulsión núcleo-núcleo como una constante. El Hamiltoniano electrónico resultante, expresado en unidades atómicas, incluye términos de energía cinética de los electrones, interacción electrón-núcleo y repulsión electrón-electrón:

$$\hat{H}_{elec} = -\sum \frac{1}{2} \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i<j} \frac{1}{r_{ij}} \quad (3.2)$$

La ecuación electrónica de Schrödinger sólo puede resolverse exactamente para sistemas de un electrón. En sistemas polielectrónicos se requieren métodos aproximados para obtener soluciones numéricas.

Uno de los enfoques más usados es la teoría del funcional de la densidad (DFT, por sus siglas en inglés)¹⁰¹, que describe el sistema mediante la densidad electrónica $\rho(\vec{r})$ en lugar de la función de onda. Según los teoremas de Hohenberg y Kohn¹⁰², la energía del estado fundamental es un funcional único de esta densidad, y su minimización proporciona la densidad exacta del sistema. DFT ofrece una alternativa computacionalmente eficiente para el estudio de propiedades electrónicas en sistemas complejos.

3.1.1. Método de Kohn-Sham

El método de Kohn-Sham¹⁰³ reformula DFT mediante un sistema ficticio de electrones no interactuantes que reproduce la densidad electrónica del

sistema real. La energía total se descompone en:

$$E[\rho] = T_s[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (3.3)$$

donde $T_s[\rho]$ es la energía cinética del sistema no interactuante, $V_{ne}[\rho]$ la interacción núcleo-electrón, $J[\rho]$ la repulsión clásica entre electrones, y $E_{xc}[\rho]$ representa los efectos de intercambio y correlación.

La densidad electrónica se construye a partir de orbitales monoeléctricos $\phi_i(\vec{r})$, soluciones de la ecuación de Kohn-Sham:

$$\left[-\frac{1}{2}\nabla^2 + v_s(\vec{r}) \right] \phi_i(\vec{r}) = \varepsilon_i \phi_i(\vec{r}) \quad (3.4)$$

El potencial efectivo $v_s(\vec{r})$ incluye el potencial externo, el de Hartree y el potencial de intercambio-correlación $v_{xc}([\rho]; \vec{r})$, derivado funcional de $E_{xc}[\rho]$. Como $E_{xc}[\rho]$ no se conoce exactamente, se recurre a aproximaciones que permiten obtener soluciones numéricas con buena precisión.

3.1.2. Funcional de intercambio-correlación

En la teoría del funcional de la densidad, es crucial representar adecuadamente los efectos de intercambio y correlación entre electrones para obtener resultados precisos. Como no se conoce la forma exacta de estos efectos, se utilizan aproximaciones. Una de las más comunes en química computacional es el funcional híbrido B3LYP (Becke, 3 parámetros, Lee-Yang-Parr)¹⁰⁴,

valorado por su buen equilibrio entre precisión y costo computacional.

B3LYP combina el intercambio exacto de Hartree-Fock con aproximaciones empíricas. Su energía de intercambio-correlación se calcula mediante:

$$E_{XC}^{B3LYP} = E_{XC}^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_X(E_X^{B88} - E_X^{LDA}) + a_C(E_C^{LYP} - E_C^{VWN}) \quad (3.5)$$

donde los distintos términos representan contribuciones de intercambio y correlación bajo diferentes aproximaciones, y los coeficientes $a_0 = 0.20$, $a_X = 0.72$ y $a_C = 0.81$ se ajustan para mejorar la concordancia con datos experimentales.

B3LYP se ha convertido en un estándar en estudios de química cuántica gracias a su capacidad para predecir propiedades como energías de enlace, estructuras moleculares y características electrónicas con buena precisión¹⁰⁵. Por esta razón, será el funcional utilizado en esta tesis.

3.1.3. Funciones base

Las funciones base son un componente esencial en DFT, ya que permiten representar los orbitales atómicos y moleculares mediante combinaciones lineales de funciones centradas en los núcleos. Entre las múltiples familias de funciones base disponibles, la serie def2, desarrollada por Weigend y Ahlrichs, se destaca por su equilibrio entre precisión y eficiencia computacio-

nal¹⁰⁶.

La función base def2-TZVP (*Triple-Zeta Valence with Polarization*) emplea tres funciones tipo zeta para describir los electrones de valencia, junto con funciones de polarización que mejoran la descripción angular de la densidad electrónica. Esta base está diseñada para ser utilizada tanto con representaciones de todos los electrones, como con pseudopotenciales efectivos, en especial para elementos con número atómico $Z \geq 19$ ¹⁰⁷.

Comparada con otras funciones base de calidad triple-zeta, def2-TZVP logra una precisión similar con menos funciones gaussianas, lo que se traduce en una mejor eficiencia computacional. Por esta razón, es ampliamente usada en estudios de optimización geométrica, cálculos de energías electrónicas, análisis vibracionales y propiedades moleculares^{108,109}.

3.2. Mecánica molecular

La mecánica molecular se basa en las leyes de la mecánica clásica aplicadas al estudio de sistemas moleculares. Esta disciplina aprovecha la aproximación de Born-Oppenheimer para describir sistemas atómicos mediante potenciales físicos simplificados, que representan las interacciones entre los núcleos atómicos. Con base en lo expuesto previamente en el marco teórico sobre campos de fuerza, esta sección se centrará en el método híbrido QM/MM y en la derivación de parámetros no enlazantes utilizados en cam-

pos de fuerza modificados, los cuales serán evaluados posteriormente.

3.2.1. Método híbrido QM/MM

El método híbrido QM/MM (*Quantum Mechanics/Molecular Mechanics*) es una estrategia ampliamente utilizada para estudiar sistemas moleculares grandes, donde una descripción puramente cuántica sería computacionalmente costosa o inabordable. Este enfoque permite tratar con mayor precisión una región del sistema mediante mecánica cuántica (QM), mientras que el resto del sistema se modela mediante mecánica molecular (MM)^{110–112}.

El sistema se divide conceptualmente en dos regiones: la región cuántica (QM), que generalmente incluye el ligando y los residuos directamente implicados en procesos químicos clave, y la región clásica (MM), que representa el entorno restante, como la proteína, el solvente y los iones. La energía total del sistema en un esquema aditivo simple puede expresarse como:

$$E_{\text{QM/MM}} = E_{\text{QM}} + E_{\text{MM}} + E_{\text{int}} \quad (3.6)$$

donde E_{QM} es la energía de la región cuántica, E_{MM} la energía de la región clásica y E_{int} representa las interacciones entre ambas regiones.

En su forma más común, el término de interacción E_{int} se modela mediante interacciones electrostáticas y de van der Waals clásicas:

$$E_{\text{int}} = E_{\text{elec}}^{\text{QM/MM}} + E_{\text{LJ}}^{\text{QM/MM}} \quad (3.7)$$

donde $E_{\text{elec}}^{\text{QM/MM}}$ representa la interacción electrostática entre los átomos de la región QM y las cargas puntuales de la región MM, y $E_{\text{LJ}}^{\text{QM/MM}}$ incluye los términos de dispersión y repulsión tipo Lennard-Jones.

Existen distintos esquemas de acoplamiento (*embeddings*) entre las regiones QM y MM, los cuales determinan cómo se incluye la influencia de la región MM en los cálculos cuánticos. En el *embedding* mecánico, las regiones QM y MM interactúan únicamente a través de fuerzas clásicas, sin que el campo electrostático de la región MM influya directamente en la región QM. En el *embedding* electrostático, las cargas atómicas de la región MM se incluyen explícitamente en el hamiltoniano QM, permitiendo que la densidad electrónica se polarice en respuesta al entorno clásico¹¹¹. Finalmente, el *embedding* polarizable incorpora además el efecto del campo QM sobre el sistema MM, mediante dipolos inducidos o modelos de polarizabilidad, asociado a un mayor costo computacional¹¹⁰.

El uso del enfoque QM/MM es particularmente relevante en sistemas biomoleculares complejos, aprovechando la precisión de la descripción mecano-cuántica y la eficiencia computacional de la mecánica molecular.

3.2.2. Cargas atómicas MBIS

Las cargas MBIS (*Minimal Basis Iterative Stockholder*) fueron propuestas por Verstraelen y colaboradores⁴⁷. Este método define la densidad proatómica $\rho_A^0(\mathbf{r})$ como una suma de funciones tipo Slater:

$$\rho_A^0(\mathbf{r}) = \sum_{i=1}^{m_A} \rho_{A_i}^0(\mathbf{r}) \quad (3.8)$$

donde m_A es el número de capas del átomo A , y $\rho_{A_i}^0(\mathbf{r})$ está dada por:

$$\rho_{A_i}^0(\mathbf{r}) = N_{A_i} f_{A_i}(\mathbf{r}) = \frac{N_{A_i}}{8\pi\sigma_{A_i}^3} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{A_i}}\right) \quad (3.9)$$

En esta expresión, N_{A_i} representa la población electrónica de la capa i , σ_{A_i} su ancho, y $f_{A_i}(\mathbf{r})$ es una función de tipo Slater normalizada.

El ajuste iterativo de los parámetros N_{A_i} y σ_{A_i} se lleva a cabo mediante la minimización de la divergencia de Kullback–Leibler (KL) entre la densidad electrónica total del sistema, $\rho(\mathbf{r})$, y la densidad reconstruida a partir de las contribuciones pro-atómicas $\rho_A^0(\mathbf{r})$. En cada iteración, se asigna una fracción de la densidad total a cada átomo A mediante un esquema de partición tipo stockholder:

$$\rho_A(\mathbf{r}) = \rho(\mathbf{r}) \frac{\rho_A^0(\mathbf{r})}{\sum_B \rho_B^0(\mathbf{r})}, \quad (3.10)$$

y se optimizan los parámetros de las funciones Slater del proátomo A para

minimizar la siguiente función objetivo:

$$D_{\text{KL}} = \sum_A \int \rho_A(\mathbf{r}) \ln \left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \right) d\mathbf{r}, \quad (3.11)$$

la cual corresponde a la divergencia KL entre la densidad asignada $\rho_A(\mathbf{r})$ y la densidad modelo $\rho_A^0(\mathbf{r})$. Este procedimiento se repite iterativamente hasta alcanzar la convergencia, es decir, cuando las distribuciones pro-atómicas dejan de modificarse significativamente.

Una vez obtenida la densidad electrónica atómica $\rho_A(\mathbf{r})$, la población electrónica N_A y la carga atómica q_A del átomo A , con carga nuclear Z_A , se calculan como:

$$q_A = Z_A - N_A = Z_A - \int \rho_A(\mathbf{r}) d\mathbf{r} \quad (3.12)$$

3.2.3. Cargas atómicas D-MBIS

Los métodos tradicionales de asignación de cargas atómicas no consideran la polarización de la densidad electrónica del ligando inducida por su entorno molecular, ya sea al encontrarse unido al receptor o en solución. Para incorporar este efecto, en nuestro laboratorio se ha desarrollado una metodología que combina simulaciones de dinámica molecular con cálculos híbridos QM/MM^{54,55,113,114}.

La polarización se cuantifica a través de las variaciones en las cargas

MBIS originadas por las fluctuaciones del entorno molecular. Esto da lugar a las llamadas cargas atómicas dinámicas MBIS (D-MBIS), las cuales capturan un promedio efectivo de la polarización inducida por el ambiente y permiten una descripción más fidedigna de las interacciones electrostáticas.

El cálculo de las cargas D-MBIS se basa en realizar simulaciones de dinámica molecular acopladas a cálculos periódicos QM/MM. A lo largo de la trayectoria simulada, cada cierto número de pasos se efectúa un cálculo QM/MM para obtener la densidad electrónica del ligando, que luego se particiona mediante el método MBIS para extraer las cargas atómicas correspondientes. Este proceso se repite un total de k veces, y las cargas finales de cada átomo se obtienen como el promedio sobre todas las configuraciones muestreadas durante la simulación.

Finalmente, estas cargas promediadas se incorporan al campo de fuerza y se reinicia el proceso iterativamente hasta alcanzar la convergencia de las cargas D-MBIS.

3.2.4. Parámetros de Lennard-Jones D-MBIS

Otro componente fundamental en las interacciones no covalentes está relacionado con los efectos de dispersión y repulsión/intercambio, los cuales se modelan mediante el potencial de Lennard-Jones. Mohebifar y colaboradores¹¹⁵ demostraron que los coeficientes de dispersión atómicos C_6 incluidos en varios campos de fuerza están sistemáticamente sobreestimados en

comparación con los derivados mediante cálculos de mecánica cuántica. Esta discrepancia ha motivado el desarrollo de nuevos esquemas para su determinación. Castillo-Orellana y colaboradores¹¹⁶ demostraron que parámetros de Lennard-Jones derivados mediante el método MBIS, pueden describir correctamente interacciones de dispersión con un error medio de $-2 \text{ kJ}\cdot\text{mol}^{-1}$. Esto valida la aplicación de esta metodología para su uso en la herramienta de derivación de parámetros no enlazantes.

El volumen atómico V_A se define como el tercer momento radial de la densidad electrónica del átomo:

$$V_A = \int |\mathbf{r} - \mathbf{R}_A|^3 \rho_A(\mathbf{r}) d\mathbf{r} \quad (3.13)$$

La formalización de Tkatchenko-Scheffler¹¹⁷ permite calcular el coeficiente de dispersión $C_{6,A}$ de un átomo A en una molécula, reescalando el valor correspondiente al átomo aislado $C_{6,A}^{\text{free}}$ según su volumen efectivo:

$$C_{6,A} = \left(\frac{V_A}{V_A^{\text{free}}} \right)^2 C_{6,A}^{\text{free}} \quad (3.14)$$

Asimismo, los radios de van der Waals ($R_{\text{vdW},A}$) pueden estimarse utilizando la relación empírica propuesta por Fedorov y Tkatchenko¹¹⁸:

$$R_{\text{vdW},A} = 2.54 \alpha_A^{1/7} \quad (3.15)$$

donde la polarizabilidad estática atómica α_A también se escala en función del volumen atómico:

$$\alpha_A = \left(\frac{V_A}{V_A^{\text{free}}} \right) \alpha_A^{\text{free}} \quad (3.16)$$

A partir de estos parámetros es posible calcular los parámetros de Lennard-Jones σ_A y ϵ_A , que determinan el tamaño efectivo del átomo y la profundidad del pozo de potencial, respectivamente:

$$\sigma_A = \frac{r_{\text{mín},A}}{2^{1/6}} = \frac{2R_{\text{vdW},A}}{2^{1/6}} \quad (3.17)$$

$$\epsilon_A = \frac{C_{6,A}}{4\sigma_A^6} \quad (3.18)$$

Para garantizar la compatibilidad con el campo de fuerza empleado, se conserva el tipo de átomo (en GAFF) o el descriptor SMIRKS (en OFF) del ligando. Los nuevos parámetros de Lennard-Jones D-MBIS se obtienen promediando los valores correspondientes a todos los átomos que comparten el mismo tipo o descriptor, de acuerdo al proceso iterativo mencionado anteriormente para las cargas atómicas D-MBIS.

3.3. Dinámica molecular

La dinámica molecular es una técnica de simulación basada en la mecánica clásica que permite estudiar la evolución temporal de un sistema molecular mediante la integración numérica de las ecuaciones de movimiento de Newton. El sistema se describe en el espacio de fase, un espacio de $6N$ dimensiones donde N es el número de partículas, y su evolución temporal genera una trayectoria que representa las posiciones y momentos de cada partícula a lo largo del tiempo.

La ecuación fundamental que rige el movimiento es:

$$\vec{F}_i = m_i \ddot{\vec{r}}_i = m_i \frac{d^2 \vec{r}_i}{dt^2} \quad (3.19)$$

donde \vec{F}_i es la fuerza neta sobre el átomo i , m_i su masa, y $\ddot{\vec{r}}_i$ su aceleración.

Esta fuerza se obtiene a partir del gradiente de la energía potencial V :

$$\vec{F}_i = -\nabla_i V(\vec{r}_1, \dots, \vec{r}_N) \quad (3.20)$$

La energía potencial se calcula mediante un campo de fuerza, y junto con las condiciones iniciales del sistema (posiciones y velocidades), permite predecir su evolución. Las velocidades iniciales se asignan conforme a una distribución de Maxwell-Boltzmann para una temperatura definida.

Las simulaciones pueden realizarse bajo diferentes colectivos estadísticos. En el colectivo microcanónico (NVE) se conservan la energía, el volumen y el número de partículas, mientras que en el colectivo canónico (NVT) o isotérmico-isobárico (NPT) se mantienen constantes la temperatura o la temperatura y presión, respectivamente. Para lograr esto, se acoplan algoritmos de control conocidos como termostatos y barostatos.

El termostato de Berendsen¹¹⁹ reescala las velocidades para ajustar la temperatura hacia un valor objetivo, aunque no reproduce correctamente las fluctuaciones del colectivo canónico. Alternativamente, el termostato estocástico de Bussi y colaboradores¹²⁰ introduce ruido térmico controlado y garantiza la correcta distribución de energía cinética.

Para el control de presión, el barostato de Berendsen aplica un reescalamiento de la celda simulada hacia una presión objetivo, mientras que enfoques más rigurosos como los de Parrinello-Rahman¹²¹ o Martyna-Tobias-Klein¹²² permiten una descripción más precisa del colectivo NPT.

Los detalles específicos de cada simulación y los parámetros utilizados se describen en la sección de resultados.

3.4. Cálculos de energía libre de unión absoluta

Los cálculos de energía libre son fundamentales para estudiar propiedades termodinámicas como la afinidad de unión de ligandos a proteínas, las

energías libres de solvatación, la estabilidad de mutaciones y las constantes de asociación/disociación, entre otras. En esta sección, se revisan los principales requisitos para calcular la energía libre de unión absoluta mediante métodos de no equilibrio, incluyendo las correcciones asociadas a efectos de polarización. Esta estimación permite obtener perfiles de especificidad y selectividad, los cuales son esenciales para evaluar la capacidad predictiva de los métodos computacionales en comparación con datos experimentales.

3.4.1. Ciclo termodinámico

La estimación directa de la energía libre de unión absoluta mediante simulaciones clásicas de dinámica molecular no es factible en la práctica. Esto se debe a que los eventos de unión y desunión son procesos raros, que ocurren en escalas de tiempo prolongadas, frecuentemente fuera del alcance de las simulaciones convencionales. Además, alcanzar un muestreo estadísticamente confiable de todos los microestados relevantes requiere tiempos de simulación prohibitivos para sistemas biomoleculares.

Para superar estas limitaciones, se recurre al uso de un ciclo termodinámico que conecta los estados de unión y no unión mediante transformaciones alquímicas. Estas rutas no físicas permiten mejorar el solapamiento entre las distribuciones del espacio de fase y obtener estimaciones más robustas de la energía libre de unión (Figura 3.1).

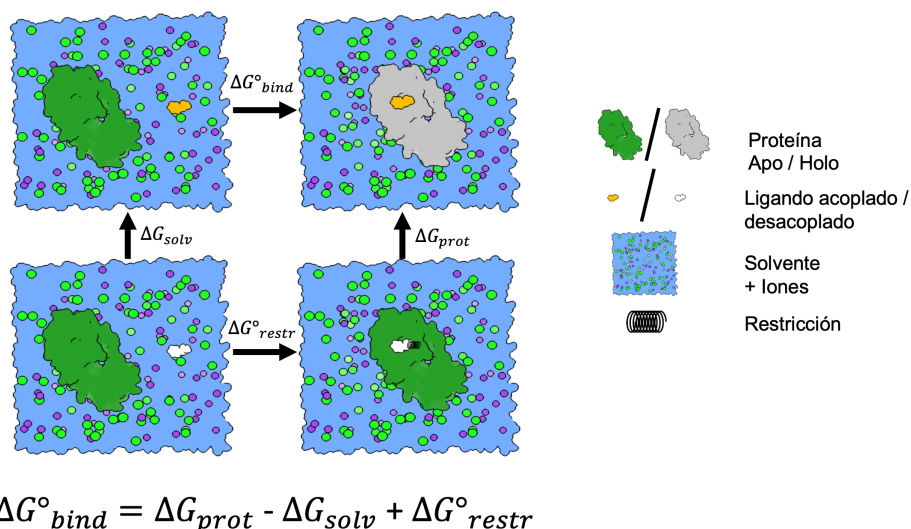


Figura 3.1: **Esquema del ciclo termodinámico utilizado para el cálculo de la energía libre de unión absoluta.** La transformación horizontal superior representa el proceso físico de asociación del ligando con la proteína en solución. Las transformaciones verticales corresponden a procesos alquímicos de acoplamiento del ligando tanto en solución como en el complejo. La transformación horizontal inferior incorpora una corrección analítica que cuantifica el costo energético de restringir el ligando al sitio de unión durante el desacoplamiento. La energía libre de unión se calcula como la diferencia entre las energías libres de desacoplamiento, sumada a la corrección de restricción.

Durante las transformaciones alquímicas, se emplean restricciones para mantener el ligando en las proximidades del sitio de unión, lo que permite aplicar la corrección del estado estándar. Estas restricciones también limitan el espacio conformacional que debe integrarse, acelerando la convergencia. Entre las más utilizadas se encuentran las restricciones armónicas, *flat-bottom* y Borech-Karplus.

La restricción tipo Borech-Karplus⁹ es particularmente adecuada para limitar los grados de libertad intermoleculares entre el receptor y el ligando

durante las transformaciones alquímicas. Esta estrategia reduce el número de orientaciones accesibles del ligando en los estados desacoplados, mejorando la convergencia estadística del cálculo.

La restricción se define mediante una combinación de una distancia, dos ángulos y tres torsiones, construidas a partir de tres átomos del receptor (\mathbf{r}_1 , \mathbf{r}_2 , \mathbf{r}_3) y tres del ligando (\mathbf{l}_1 , \mathbf{l}_2 , \mathbf{l}_3). La energía asociada a esta restricción se expresa como:

$$E_{\text{restr}} = \lambda_{\text{restr}} \left[\frac{K_r}{2} (r - r_0)^2 + \frac{K_{\theta_1}}{2} (\theta_1 - \theta_{1,0})^2 + \frac{K_{\theta_2}}{2} (\theta_2 - \theta_{2,0})^2 + \frac{K_{\phi_1}}{2} \text{hav}(\phi_1 - \phi_{1,0})^2 + \frac{K_{\phi_2}}{2} \text{hav}(\phi_2 - \phi_{2,0})^2 + \frac{K_{\phi_3}}{2} \text{hav}(\phi_3 - \phi_{3,0})^2 \right] \quad (3.21)$$

donde $\text{hav}(x) = \frac{1 - \cos(x)}{2}$ es la función de Haversine, empleada para definir la energía asociada a las torsiones. Las constantes K determinan la rigidez de cada restricción, mientras que los valores con subíndice 0 representan las posiciones de equilibrio. Finalmente, λ_{restr} es un parámetro de escala que permite modular globalmente la restricción.

Una vez aplicadas las restricciones tipo Boresch-Karplus, es posible calcular una corrección analítica al estado estándar, necesaria para convertir la energía libre de desacoplamiento bajo restricciones a una energía de unión en condiciones estándar (1 mol/L). Esta corrección compensa el sesgo introducido por limitar el volumen accesible al ligando durante el desacoplamiento.

El valor de esta corrección, $\Delta G_{\text{restr}}^{\circ}$, se calcula a partir de la integral sobre el volumen restringido del espacio de fases asociado a los seis grados de libertad intermoleculares (una distancia, dos ángulos y tres torsiones):

$$\Delta G_{\text{restr}}^{\circ} = -k_B T \ln \left(\frac{Z_{\text{restr}}}{V^{\circ}} \right) \quad (3.22)$$

donde Z_{restr} es la partición configuracional restringida, que depende de las constantes de fuerza y los valores de equilibrio definidos en la ecuación 3.21. $V^{\circ} = 1660 \text{ \AA}^3$ corresponde al volumen estándar asociado a una concentración de 1 mol/L.

En esta tesis se utilizó la metodología de doble desacoplamiento mediante integración termodinámica de no equilibrio¹⁴, en la que el ligando es desacoplado progresivamente de su entorno tanto en solución como en el complejo. El procedimiento se ilustra para el estado de unión en la Figura 3.2.

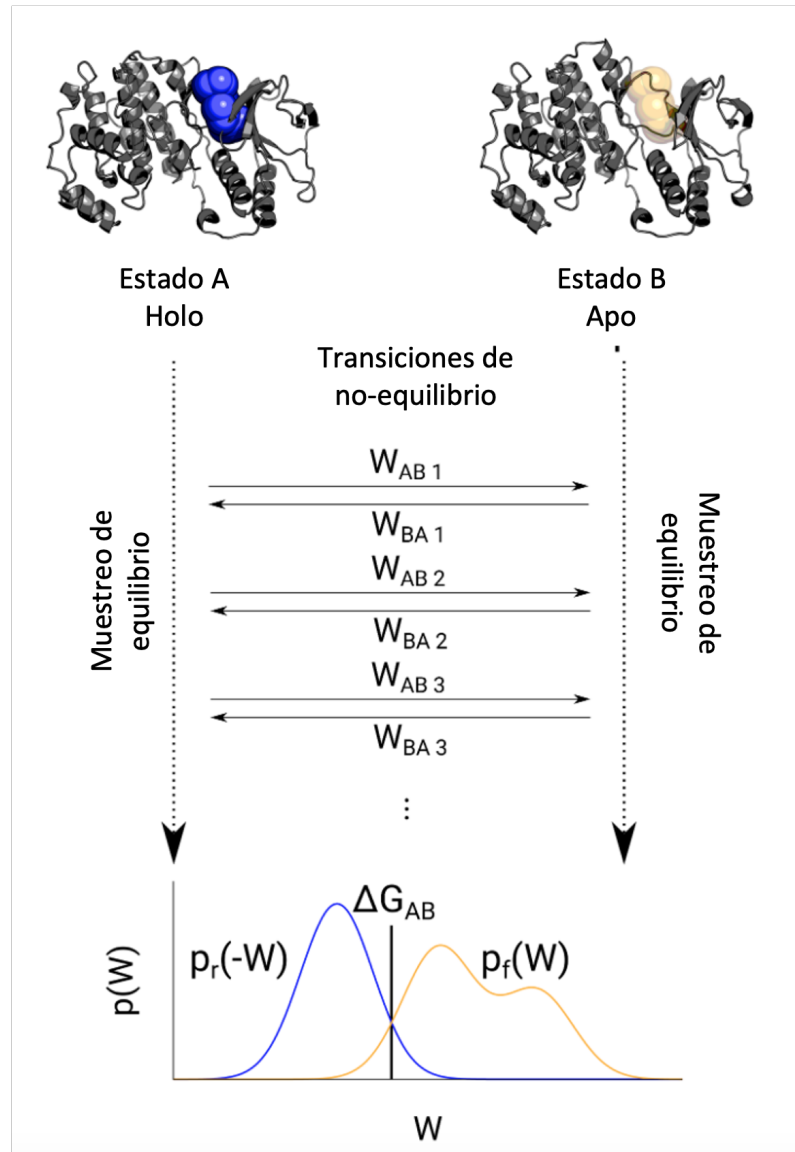


Figura 3.2: **Esquema del cálculo de energía libre mediante trayectorias de no equilibrio.** A partir de configuraciones de equilibrio de los estados holo (A) y apo (B), se generan múltiples transformaciones rápidas entre ambos estados. Las trayectorias $A \rightarrow B$ producen trabajos W_{AB} , mientras que las $B \rightarrow A$ generan W_{BA} . La energía libre ΔG_{AB} se determina mediante la superposición de las distribuciones $p_f(W)$ y $p_r(-W)$, usando el teorema de fluctuación de Crooks.

Estas transformaciones se realizan desacoplando las interacciones no en-

lazantes del ligando con el sistema, mientras se conservan sus interacciones intramoleculares. Las simulaciones se ejecutan en escalas de tiempo cortas (ps–ns) desde configuraciones iniciales obtenidas desde trayectorias de equilibrio. A partir del trabajo acumulado se obtienen las distribuciones necesarias para aplicar el teorema de Crooks (Ecuación 1.26) mediante el estimador bidireccional BAR en contextos de no equilibrio (Ecuación 1.29).

Así, la energía libre de unión estándar queda expresada como:

$$\Delta G^{\circ}_{\text{bind}} = \Delta G_{\text{prot}} - \Delta G_{\text{solv}} + \Delta G^{\circ}_{\text{restr}} \quad (3.23)$$

3.4.2. Corrección por polarización

Para considerar adecuadamente la influencia del entorno en la energía libre de unión, se incluyó la polarización electrónica del ligando en solución y en el sitio de unión. Esta polarización conlleva un costo energético positivo que se resta de la energía libre estimada en cada entorno. El esquema de doble desacoplamiento empleado fue ajustado para incorporar esta corrección, como se muestra en la Figura 3.3.

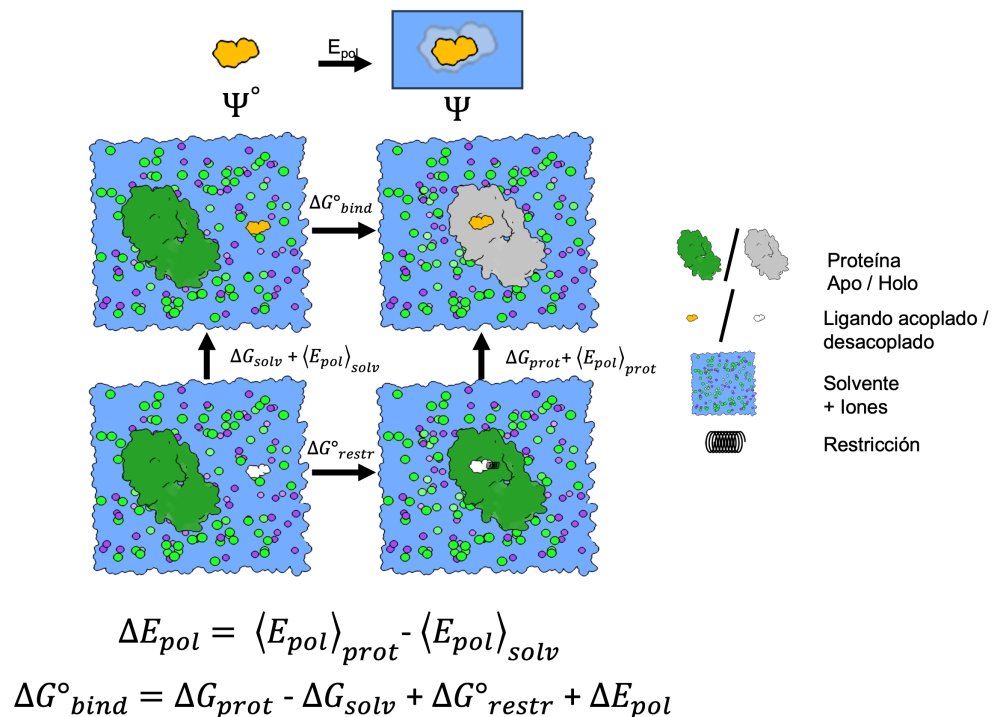


Figura 3.3: **Esquema del ciclo termodinámico para el cálculo de la energía libre de unión absoluta con corrección por polarización.** El ciclo incorpora una corrección energética en las transformaciones verticales, asociada a la deformación de la densidad electrónica del ligando al pasar del vacío al medio solvente o al sitio de unión.

Dado que las energías de polarización en ambos entornos no son equivalentes, su diferencia se incluye como una corrección a la energía libre de unión. Esta energía se estimó mediante cálculos electrónicos en vacío, utilizando la función de onda polarizada del ligando:

$$E_{pol} = \langle \Psi | \hat{H}^{\circ} | \Psi \rangle - \langle \Psi^{\circ} | \hat{H}^{\circ} | \Psi^{\circ} \rangle \quad (3.24)$$

Las energías promedio $\langle E_{\text{pol}} \rangle_{\text{prot}}$ y $\langle E_{\text{pol}} \rangle_{\text{solv}}$ se obtuvieron a partir de configuraciones representativas en cada entorno. La energía libre de unión corregida se expresa como:

$$\Delta G^{\circ}_{\text{bind}} = \Delta G_{\text{prot}} - \Delta G_{\text{solv}} + \Delta G^{\circ}_{\text{restr}} + \Delta E_{\text{pol}} \quad (3.25)$$

donde $\Delta E_{\text{pol}} = \langle E_{\text{pol}} \rangle_{\text{prot}} - \langle E_{\text{pol}} \rangle_{\text{solv}}$.

3.5. Descriptores estadísticos

Para evaluar la calidad de las estimaciones obtenidas en esta tesis, se emplearon varios descriptores estadísticos comunes en el análisis de errores y validación de modelos. Entre ellos se encuentran el error cuadrático medio (RMSE), el error absoluto medio (MUE) y los coeficientes de correlación de Pearson y Kendall.

El **error cuadrático medio** (Root Mean Square Error, RMSE) mide la desviación promedio cuadrática entre los valores predichos y los observados. Es especialmente sensible a errores grandes y se define como:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{obs}})^2} \quad (3.26)$$

El **error absoluto medio** (Mean Unsigned Error, MUE) calcula la media de los errores absolutos, proporcionando una medida más robusta ante valores

atípicos en comparación con el RMSE:

$$\text{MUE} = \frac{1}{N} \sum_{i=1}^N |y_i^{\text{pred}} - y_i^{\text{obs}}| \quad (3.27)$$

Ambos indicadores son ampliamente utilizados en química computacional y aprendizaje automático para cuantificar la precisión de modelos de predicción¹²³.

Además, se emplearon dos coeficientes de correlación no dimensionales para evaluar la relación entre los valores predichos y los experimentales:

- El **coeficiente de correlación de Pearson** (r) mide el grado de dependencia lineal entre dos variables. Su valor varía entre -1 y 1, donde 1 indica una correlación lineal perfecta:

$$r = \frac{\sum_{i=1}^N (y_i^{\text{pred}} - \bar{y}^{\text{pred}})(y_i^{\text{obs}} - \bar{y}^{\text{obs}})}{\sqrt{\sum_{i=1}^N (y_i^{\text{pred}} - \bar{y}^{\text{pred}})^2} \sqrt{\sum_{i=1}^N (y_i^{\text{obs}} - \bar{y}^{\text{obs}})^2}} \quad (3.28)$$

- El **coeficiente de correlación de Kendall** (τ) es una medida no paramétrica de asociación basada en el orden de los datos. Es útil cuando la relación entre variables no es estrictamente lineal¹²⁴. Se calcula a partir del número de pares concordantes y discordantes:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}N(N-1)} \quad (3.29)$$

donde n_c y n_d representan los pares concordantes y discordantes, respectivamente.

El uso combinado de estos descriptores permite una evaluación integral de la precisión y la validez estadística de los métodos de estimación y los distintos campos de fuerza empleados.

4. Resultados y Discusión

En la presente investigación, se buscó evaluar el efecto del cambio de parámetros no enlazantes sobre campos de fuerza no polarizables en la estimación de la energía libre de unión absoluta. Estos nuevos parámetros derivados consideran el efecto de las fluctuaciones de la densidad electrónica molecular en el proceso de unión no covalente. En adición al costo energético de polarizar la densidad electrónica, se propone la mejora en las predicciones de energía libre de unión en los sistemas proteína-ligando de referencia utilizados en esta tesis.

A continuación, se muestran los resultados más relevantes de la tesis. En primera instancia se discute la herramienta computacional desarrollada para la derivación de parámetros no enlazantes. Posteriormente, se reportan y analizan los resultados obtenidos para las energías libres de unión absoluta en los sistemas BRD4(1), bromosporina-bromodominios y lizosima T4 L99A/M102Q. Estos consideran diferentes modificaciones de parámetros no enlazantes en los distintos campos de fuerza utilizados, incluyendo efectos derivados del costo energético de polarización de la densidad electrónica molecular. Los resultados fueron comparados respecto a datos experimentales de afinidad y contrastados mediante los descriptores estadísticos: error cuadrá-

tico medio (RMSE), error absoluto medio (MUE), coeficiente de correlación de Pearson (R) y coeficiente de Kendall (τ).

4.1. Herramienta computacional para la derivación de parámetros no enlazantes

A la fecha, persisten desafíos relacionados con la descripción precisa de las interacciones intra e intermoleculares en sistemas biomoleculares. En simulaciones de dinámica molecular aplicadas a macromoléculas, estas interacciones se modelan mediante campos de fuerza clásicos, los cuales definen la energía potencial del sistema a partir de una suma de términos aditivos²³. En el caso de campos de fuerza no polarizables, dichos términos se clasifican convenientemente en dos grupos: interacciones enlazantes y no enlazantes.

Los parámetros enlazantes, los cuales incluyen constantes de fuerza y valores de equilibrio para enlaces, ángulos, torsiones y ángulos impropios, describen la geometría interna de la molécula. En cambio, los parámetros no enlazantes representan interacciones electrostáticas y de dispersión/repulsión, que se cuantifican mediante la asignación de cargas atómicas usadas en el potencial de Coulomb y coeficientes adecuados para el potencial de Lennard-Jones. Estos dos últimos juegan un papel crucial en fenómenos como el reconocimiento molecular¹²⁵.

En particular, los métodos basados en la partición de la densidad electrónica^{46,76} han demostrado ser eficaces para derivar tanto cargas atómicas como parámetros de dispersión y repulsión⁴⁷. Estas metodologías consisten en

descomponer la densidad electrónica molecular en contribuciones atómicas, permitiendo una asignación más física de los parámetros. Sus aplicaciones abarcan una amplia gama de propiedades moleculares, incluyendo la predicción de energías libres de solvatación⁴⁸⁻⁵⁰, coeficientes de partición^{51,52}, reproducción de densidades en fase líquida^{39,53,114}, así como la estimación de energías libres de unión en sistemas anfitrión-huésped^{54,55} y complejos proteína-ligando^{40,56}.

Con el objetivo de mejorar la descripción de interacciones no enlazantes en contextos moleculares específicos, desarrollamos una herramienta computacional para su derivación automatizada. Esta implementación, estructurada como un paquete de Python 3¹²⁶, sigue el flujo de trabajo presentado en la Figura 4.1.

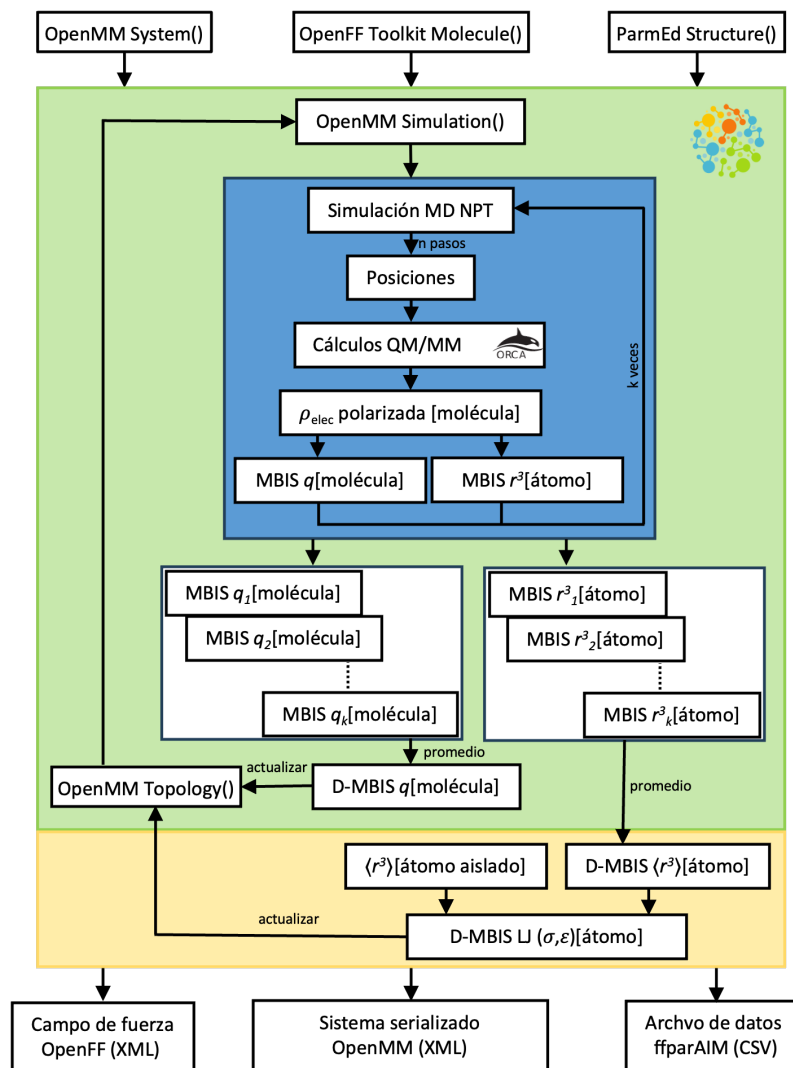


Figura 4.1: **Diagrama de flujo computacional utilizado para derivar los parámetros no enlazantes en ambientes moleculares explícitos.** El protocolo considera la generación de trayectorias mediante muestreo configuracional desde simulaciones de dinámica molecular clásica, cálculos QM/MM para la partición de densidad electrónica y la actualización iterativa de cargas atómicas y parámetros de dispersión/repulsión compatibles con campos de fuerza tipo SMIRNOFF.

El código fuente está disponible en GitHub: <https://github.com/QCMM/ffparaim>. El paquete es compatible con la estructura de datos de cam-

pos de fuerza tipo SMIRNOFF como Open Force Field (OFF)³¹ y se apoya en bibliotecas del ecosistema QCDevs (como `iodata`¹²⁷, `grid`¹²⁸, `denspart`, `gbasis`¹²⁹) así como en el paquete ParmEd.

Esta herramienta permite obtener cargas atómicas y parámetros de dispersión/repulsión (distancia del potencial igual a cero y profundidad del pozo potencial) adaptados al entorno químico de moléculas pequeñas, de forma completamente automatizada. El protocolo de derivación se basa en el campo de fuerza OFF-2.0.0 Sage³⁵, manteniendo sin modificaciones los parámetros enlazantes, aunque la herramienta admite otros campos de fuerza para moléculas pequeñas. La modificación se limita a las cargas atómicas y parámetros de Lennard-Jones, definidos mediante patrones SMIRKS^{32,33,130}.

La implementación permite ejecutar un protocolo predefinido de derivación (denominado D-MBIS^{54,55}) o utilizar sus componentes de forma modular para definir flujos de trabajo personalizados. El protocolo D-MBIS se basa en la obtención iterativa de parámetros promedios a partir de particiones de densidad electrónica obtenidas mediante cálculos QM/MM¹¹¹ desde configuraciones extraídas de simulaciones clásicas de dinámica molecular. La densidad electrónica se calcula con el nivel de teoría especificado por el usuario, dentro de las capacidades del software de química cuántica ORCA¹³¹.

Esta metodología, que considera explícitamente el entorno molecular, representa una mejora respecto a aproximaciones previas empleadas por el grupo^{48,51,52,54,55}. Las trayectorias de dinámica molecular son generadas me-

dante OpenMM¹³², mientras que los cálculos de estructura electrónica se realizan con ORCA v5.0.3¹³¹. Los requisitos mínimos para ejecutar el protocolo incluyen: un archivo de coordenadas en formato PDB, una cadena SMILES isomérica del ligando, la carga neta del ligando, y la selección del ligando en formato máscara de AMBER²⁶. Opcionalmente, puede definirse también la selección atómica del receptor para derivar parámetros en el contexto de complejos ligando-receptor, lo que permite aplicar restricciones que evitan la disociación del ligando.

Entre las variables internas del protocolo se encuentran: el número de actualizaciones iterativas, el tiempo de muestreo total para la trayectoria de dinámica molecular, el número total de cálculos QM/MM, el método de estructura electrónica (por defecto B3LYP), y la función base (por defecto def2-TZVP). La clase principal FFparAIM contiene dos métodos principales: `prepare()` y `run()`. El primero procesa el archivo PDB y la cadena SMILES, generando los objetos necesarios para iniciar la derivación. El segundo ejecuta el protocolo y genera los archivos de salida con los nuevos parámetros para su uso en simulaciones de dinámica molecular. Además, se estima un término de corrección energética E_{pol} , que representa el costo de polarización de la densidad electrónica en el entorno químico.

Este protocolo ha sido aplicado exitosamente a sistemas anfitrión-huésped^{54,55}, complejos proteína-ligando¹¹³, y en la predicción de propiedades termo-físicas¹¹⁴. A continuación, se presentan ejemplos de ejecución

del protocolo para una molécula en disolución acuosa, un complejo ligando-receptor y un flujo de trabajo personalizado.

Derivación de parámetros en disolución acuosa.

```
1 from ffparaim import FFparAIM
2
3 # PDB coordinates file.
4 pdb_file = 'solvent.pdb'
5 # SMILES string.
6 smiles = 'Cc1ccccc1'
7 # OpenMM XML force field files for chemical environment.
8 ff = ['amber14/tip3p.xml']
9
10 # Create ffparaim object.
11 nb_params = FFparAIM(qm_charge=0, # Total charge of
    ↪ the molecule to parametrize.
12                        ligand_selection=':MOL', # AMBER mask
    ↪ selection for the molecule.
13                        n_updates=5, # Number of
    ↪ parameter's updates.
14                        sampling_time=5, # Sampling time
    ↪ in nanoseconds.
15                        total_qm_calculations=25) # Number of QM/MM
    ↪ calculations.
16
17 # Prepare input objects.
18 molecule, system_structure, system = nb_params.prepare(smiles,
19                                                         pdb_file,
20                                                         ff_env=ff)
21 # Run protocol.
22 nb_params.run(molecule,
23               system_structure,
24               system,
25               off=True,
26               charges=True,
27               lj=True)
```

Derivación de parámetros en complejo proteína-ligando.

```
1 from ffparaim import FFparAIM
2
3 # PDB coordinates file.
4 pdb_file = 'complex.pdb'
5 # SMILES string.
6 smiles = 'Cc1ccccc1'
7 # Restraints dict.
8 restraint = {'harmonic': {'ligand': [1, 3, 6], 'receptor': [1621,
   ↪ 1751, 1843]}}
9 # OpenMM XML force field files for chemical environment.
10 ff =
   ↪ ['/home/lmacaya/tesis/t4-199a-m102q/ffparaim/amber99sbildn_mod.xml',
11 'amber14/tip3p.xml']
12 # Create ffparaim object.
13 nb_params = FFparAIM(qm_charge=0, # Total charge of the
   ↪ molecule to parametrize.
14                          ligand_selection=':MOL', # AMBER mask
   ↪ selection for the molecule.
15                          n_updates=5, # Number of
   ↪ parameter's updates.
16                          sampling_time=10, # Sampling time in
   ↪ nanoseconds.
17                          total_qm_calculations=25, # Number of QM/MM
   ↪ calculations.
18                          method='B3LYP', # DFT functional.
19                          basis='def2-TZVP') # Basis set.
20 # Prepare input objects.
21 molecule, system_structure, system = nb_params.prepare(smiles,
22                                                         pdb_file,
23                                                         ff_env=ff)
24 # Run protocol.
25 nb_params.run(molecule,
26               system_structure,
27               system,
```

```
28         restraint_dict=restraint,  
29         off=True,  
30         charges=True,  
31         lj=True)
```

Derivación de parámetros personalizada.

```
1 import os
2 import sys
3
4 from ffparaim import FFparAIM
5 from ffparaim.mdtools import define_molecule
6
7 from openmm import openmm
8 from openmm import app
9 from openmm import unit
10 from parmed import load_file
11
12 # SMILES string.
13 smiles = 'Cc1ccccc1'
14 # PDB coordinates file for ligand.
15 lig_pdb_file = os.path.abspath('MOL.pdb')
16 # Restraints dict.
17 restraint = {'harmonic': {'ligand': [2605, 2608, 2613], 'receptor':
    ↪ [1608, 1738, 1830]}}
18 # Use the OpenFF Toolkit to generate a chemical topology.
19 molecule = define_molecule(smiles, lig_pdb_file=lig_pdb_file)
20 # Create ParmEd Structure object.
21 prmtop = load_file('complex.prmtop', xyz='complex.inpcrd')
22 # Fix non-positive periodicity for OpenMM 7.7.
23 for dih in prmtop.dihedrals:
24     try:
25         for dihtype in dih.type:
26             if dihtype.per == 0:
27                 assert dihtype.phi_k == 0
28                 dihtype.per = 1
29     except TypeError:
30         if dih.type.per == 0:
31             assert dih.type.phi_k == 0
32             dih.type.per = 1
33 # Create OpenMM System.
```

```

34 system_structure = prmtop
35 system = prmtop.createSystem(nonbondedMethod=app.PME,
36                             nonbondedCutoff=1 * unit.nanometer,
37                             constraints=app.HBonds)
38 system.addForce(openmm.MonteCarloBarostat(1 * unit.bar, 298.15 *
↪   unit.kelvin))
39 # Create ffparaim object.
40 nb_params = FFparAIM(qm_charge=0,
41                     ligand_selection=':MOL',
42                     n_updates=5,
43                     sampling_time=5,
44                     total_qm_calculations=25)
45 # Run protocol.
46 nb_params.run(molecule,
47               system_structure,
48               system,
49               charges=True,
50               lj=True)
51

```

Como ejemplo, en la Figura 4.2 se muestra la evaluación exhaustiva de variables de la herramienta `ffparaim` en función de la energía de polarización. Para esto se usó el ligando L6 del conjunto de datos lisozima T4 L99A/M102Q, en el contexto de la derivación de parámetros en complejo. Adicionalmente, se muestran los tiempos de ejecución para cada una de las pruebas realizadas.

Para la molécula puesta a prueba, compuesta de 21 átomos, el protocolo por defecto (5 actualizaciones del campo de fuerza, 5 ns de muestreo y 25 cálculos QM/MM) toma 11 horas la obtención de parámetros no enlazantes en solvente y 17 horas en el complejo proteico. En base al escalado lineal del

tiempo de ejecución, asociado directamente al número de cálculos QM/MM involucrados en el protocolo de derivación, como proyección se busca la paralelización de estos cálculos, que llevarían a un tiempo de respuesta aún menor que el que se tiene actualmente.

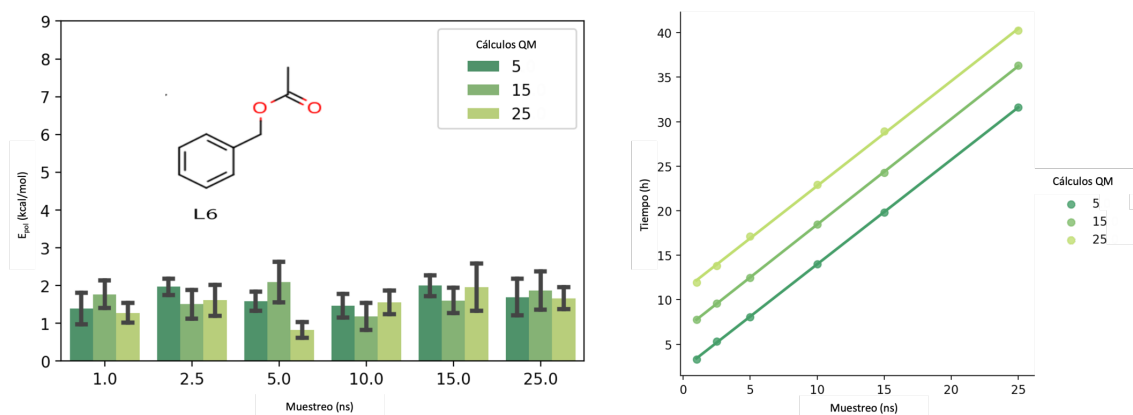


Figura 4.2: **Evaluación exhaustiva de variables utilizadas en el protocolo de derivación de la herramienta computacional.** A la izquierda se muestran los valores de energía de polarización obtenidos para las distintas variables utilizadas, mientras que a la derecha se presentan los tiempos totales de cada ejecución del protocolo con esas variables. Para cada ejecución, se mantuvo constante el número de actualizaciones del campo de fuerza (5).

4.2. Especificidad en BRD4(1)

Tras la generación de la herramienta de derivación de parámetros no enlazantes para campos de fuerza no polarizables, se procedió a evaluar su efecto en la predicción de la energía libre de unión absoluta sobre el conjunto de datos de especificidad. Los bromodominios son lectores de marcas epigenéticas que reconocen específicamente motivos de acetilación de la ϵ -N-3-lisina^{133,134}, y se han identificado en 46 proteínas nucleares y citoplasmáticas¹³⁵. Estas modificaciones postraduccionales se encuentran frecuentemente en complejos macromoleculares involucrados en la remodelación de la cromatina, la reparación del ADN y el control del ciclo celular, especialmente en histonas¹³⁶. Se ha propuesto que la acetilación de histonas favorece la activación transcripcional, y niveles anómalos de acetilación se han vinculado con desregulación en la expresión génica en contextos de cáncer e inflamación¹³⁷.

A nivel estructural, el reconocimiento del motivo acetyl-lisina por parte de los bromodominios está mediado por una cavidad formada por cuatro hélices canónicas (αZ , αA , αB y αC), dos bucles de conexión (ZA y BC), y moléculas de agua estructuralmente conservadas (Figura 4.3). Este sistema ha sido abordado en múltiples publicaciones que analizan aspectos como energía libre de unión^{12,138–141}, predicción de entalpía de unión^{140,142} y contribuciones entrópicas¹⁴⁰, además de efectos conformacionales tanto del ligando como de

la proteína^{142,143}, y el papel de redes de agua estructurales¹⁴⁴.

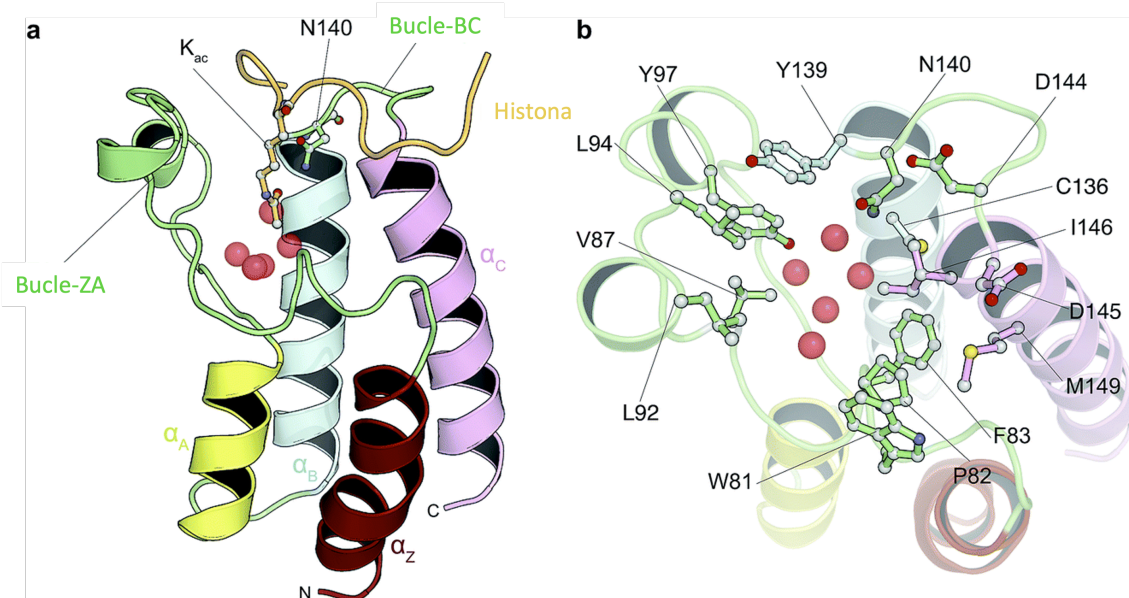


Figura 4.3: **Características estructurales de bromodominios.** A la izquierda, se muestra la estructura de BRD4(1) en complejo con un péptido acetilado derivado de una histona. A la derecha, se indican los residuos claves del sitio de unión, en conjunto a moléculas de agua estructurales. Imagen adaptada desde Aldeghi y colaboradores¹²

En este contexto, el conjunto de prueba BRD4(1) posee alta relevancia dentro de los sistemas de evaluación proteína-ligando (Figura 1.1), dado su rango de valores experimentales que permite una validación cuantitativa de predicciones computacionales, la diversidad química de los ligandos empleados (a nivel de andamiajes moleculares) y un sitio de unión relativamente rígido, lo que reduce la dependencia de un muestreo conformacional extensivo¹².

Los archivos requeridos por GROMACS (.gro y .top) para los complejos

proteína-ligando y los ligandos en solvente correspondientes a las 11 moléculas estudiadas fueron obtenidos desde el repositorio de Zenodo (<https://zenodo.org/records/57131>), utilizados originalmente en el trabajo de Aldeghi y colaboradores¹². Los valores experimentales de energía libre de unión también fueron extraídos de dicha publicación. La Tabla A.1 resume los datos experimentales relevantes asociados a este sistema.

Las simulaciones fueron ejecutadas con el paquete GROMACS 2022.4¹⁴⁵. Se empleó el campo de fuerza Amber ff99SB-ILDN para la proteína^{146,147}, junto con el modelo de agua TIP3P¹⁴⁸ y los parámetros de iones monovalentes de Joung y Cheatham¹⁴⁹. Los parámetros de los ligandos fueron generados usando GAFF-1.5²⁶ y OFF-2.0.0^{34,35}, en ambos casos con cargas AM1-BCC^{150,151}. Estos se utilizaron para validar la metodología de estimación de energía libre de unión absoluta, así como para analizar el efecto de modificaciones en parámetros no enlazantes.

Las simulaciones de producción se realizaron a temperatura y presión constantes (298.15 K, 1 bar), empleando dinámica de Langevin para el termostato¹⁵² y el barostato de Parrinello-Rahman para el control de presión¹²¹. Se utilizó un integrador de dinámica estocástica con una constante de fricción de 1 ps^{-1} y un paso de integración de 2 fs. Los enlaces con átomos de hidrógeno se restringieron mediante el algoritmo LINCS¹⁵³ con un orden de restricción de 6.

Las interacciones electrostáticas de largo alcance se trataron mediante el

método PME^{67,68}, con interpolación cúbica, una tolerancia relativa de 10^{-4} , espaciado de malla de 0.1 nm y un umbral de corte de 1.0 nm. Las interacciones de van der Waals se suavizaron mediante un potencial de *switching* entre 0.9 y 1.0 nm. Tanto las interacciones electrostáticas como de van der Waals del ligando sometido al cambio químico, se trataron con funciones de suavizado propuestas por Gapsys y colaboradores¹⁵⁴, con el fin de mejorar la convergencia en los extremos de la transformación.

Debido a la diversidad estructural de los ligandos, se utilizaron diferentes valores de corte para la lista de vecinos (`rlist`) en las simulaciones de complejo proteína-ligando y en solución. Por ejemplo: ligando 1 (2.2 nm, 2.1 nm), ligando 4 (2.2 nm, 1.9 nm) y ligando 7 (1.8 nm en ambos casos). Para el resto de los compuestos, se usó un valor de corte estándar de 1.6 nm. Esto se realizó con el fin de evitar errores en el tratamiento de interacciones intermoleculares, dados por torsiones en la molécula desacoplada que superen el valor de corte asignado automáticamente por GROMACS al inicio de la simulación.

La energía libre de unión absoluta se estimó utilizando el ciclo termodinámico propuesto por Aldeghi y colaboradores¹² (Figura 3.1), que descompone el proceso en tres contribuciones: la energía libre de solvatación del ligando (ΔG_{solv}), la del ligando en complejo con la proteína (ΔG_{prot}), y una corrección analítica debido a la restricción posicional del ligando desacoplado en el sitio de unión ($\Delta G_{\text{restr}}^{\circ}$). Esta última se incorporó mediante las restricciones

tipo Boresch-Karplus⁹.

Adicionalmente, se aplicó una corrección semi-analítica para efectos electrostáticos de tamaño finito en sistemas cargados ($\Delta\Delta G_{\text{elec corr}}$), siguiendo la metodología de Rocklin y colaboradores¹⁵⁵. La expresión completa para el cálculo de la energía libre de unión absoluta queda resumida en la Ecuación 4.1:

$$\Delta G^{\circ}_{\text{bind}} = \Delta G_{\text{prot}} - \Delta G_{\text{solv}} + \Delta G^{\circ}_{\text{restr}} + \Delta\Delta G_{\text{elec corr}} \quad (4.1)$$

Con el objetivo de validar el protocolo de simulación, se siguió la metodología propuesta por Gapsys y colaboradores¹⁴. Se generaron tres réplicas independientes para cada ligando, tanto en solución como en complejo con la proteína, utilizando el campo de fuerza GAFF-1.5.

Cada sistema fue sometido a una etapa inicial de minimización energética (10,000 pasos mediante el algoritmo *steepest descent*), seguida por un equilibrio en el colectivo NVT durante 0.5 ns. Posteriormente, se realizaron simulaciones de producción en el colectivo NPT durante 10 ns para ambos estados (ligando en solución y en complejo).

Para las transformaciones de no equilibrio, se llevaron a cabo 96 transiciones de 0.5 ns cada una para estimar ΔG_{solv} y ΔG_{prot} . Las configuraciones iniciales para estas transiciones se extrajeron cada 100 ps de la trayectoria de producción, descartando los primeros 0.4 ns como parte del equilibrio inicial.

La estimación de la energía libre de unión se realizó mediante el estimador de Bennett (*Bennett Acceptance Ratio*, BAR)⁹⁹. La incertidumbre se calculó mediante *bootstrapping* con 100 muestras independientes, utilizando la herramienta pmx¹⁵⁶. Los valores reportados corresponden al promedio de las tres réplicas, junto con su error estándar. La calidad de las predicciones se evaluó mediante los descriptores estadísticos: error medio absoluto (MUE), raíz del error cuadrático medio (RMSE), coeficiente de correlación lineal de Pearson (r) y correlación ordinal de Kendall (τ).

Con el fin de evaluar la reproducibilidad y consistencia del protocolo para el cálculo de energía libre de unión, se realizó una comparación directa con los valores reportados por Gapsys y colaboradores¹⁴, mediante un gráfico de paridad (Figura 4.4). Esta comparación mostró una excelente correlación y una desviación mínima con respecto a los valores de referencia, validando así la correcta implementación del protocolo de simulación y su aplicabilidad en estudios posteriores con otros campos de fuerza. La Figura A.1 presenta los resultados obtenidos con GAFF-1.5 y cargas AM1-BCC.

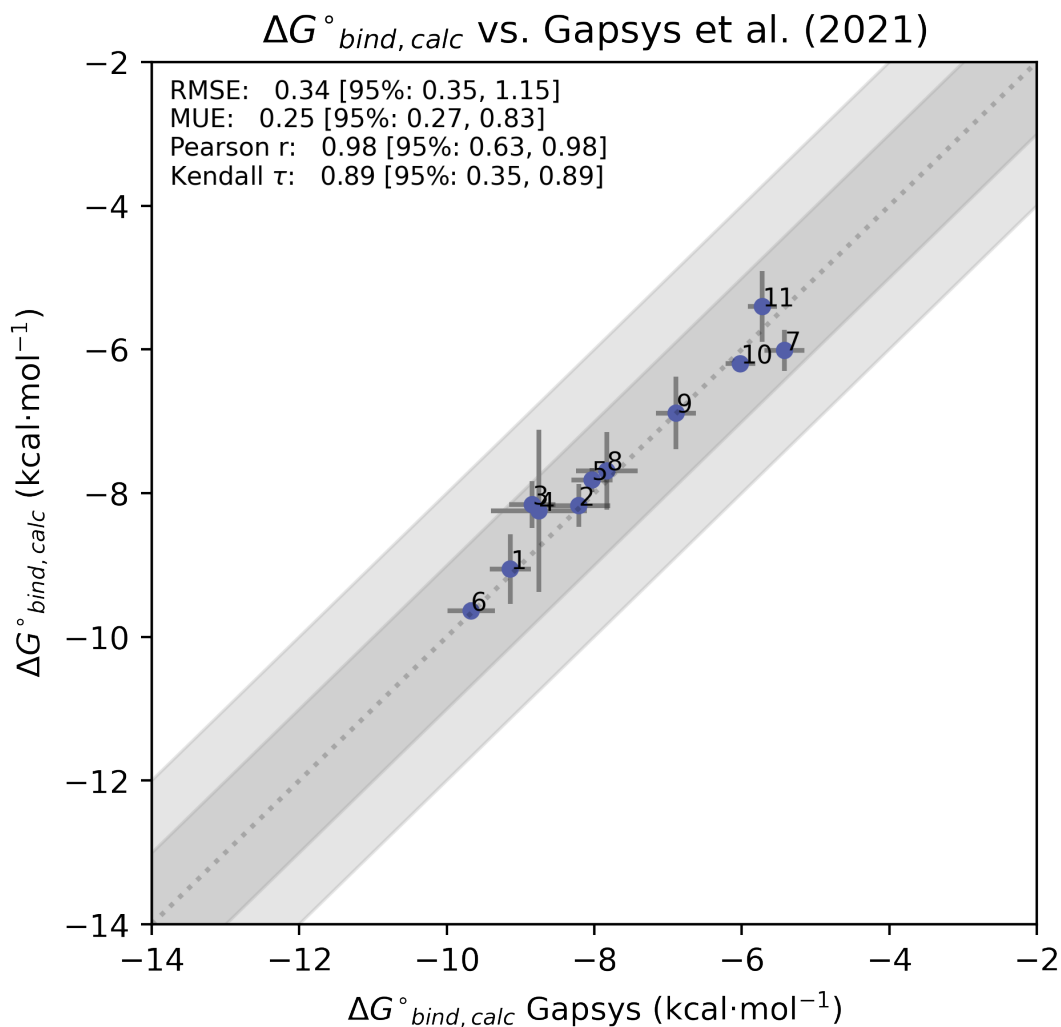


Figura 4.4: Gráfico de paridad de valores de energía libre de unión absoluta obtenidos en este estudio para el conjunto de datos BRD4(1), respecto a los publicados por Gapsys y colaboradores.

Una vez validado el protocolo, se procedió a evaluar el campo de fuerza OFF-2.0.0 en su versión original, manteniendo sin modificaciones los parámetros no enlazantes y utilizando cargas AM1-BCC. La Figura 4.5 y la ta-

bla 4.1 muestra los resultados obtenidos. En este caso, 8 de los 11 ligandos presentaron diferencias menores a $1 \text{ kcal}\cdot\text{mol}^{-1}$ con respecto a los datos experimentales, mientras que los compuestos 2, 4 y 6 exhibieron desviaciones superiores a $2 \text{ kcal}\cdot\text{mol}^{-1}$.

Los descriptores estadísticos reflejan un buen comportamiento predictivo: $\text{RMSE} = 1.45 \text{ kcal}\cdot\text{mol}^{-1}$, $\text{MUE} = 1.15 \text{ kcal}\cdot\text{mol}^{-1}$, $r = 0.83$ y $\tau = 0.64$. Esto sugiere que el modelo mantiene una capacidad predictiva razonable en términos de tendencia general, con errores absolutos dentro de los límites considerados como aceptables en predicciones de energía libre de unión absoluta.

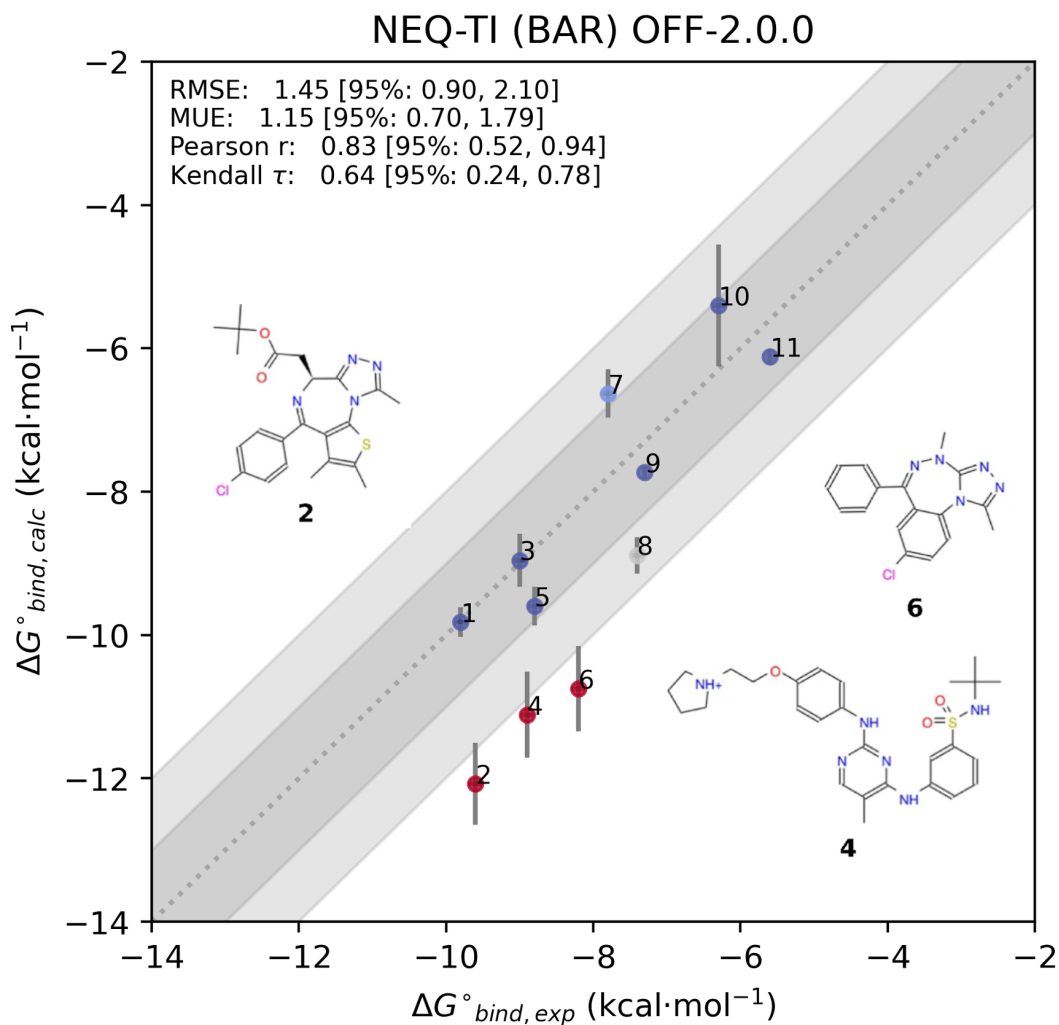


Figura 4.5: **Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y cargas atómicas AM1-BCC para el conjunto de datos BRD4(1).**

En base a estudios previos que validan el uso de parámetros derivados de la partición de la densidad electrónica molecular en simulaciones de energía libre de unión^{54,55,113}, se procedió a modificar los parámetros no enlazantes

del campo de fuerza OFF-2.0.0. Esta modificación incluyó tanto cargas atómicas como parámetros de Lennard-Jones, los cuales fueron derivados mediante el protocolo D-MBIS utilizando el programa `ffparaim` desarrollado en esta tesis.

Dado que la polarización del ligando varía entre los entornos de solvación y de unión, se incorporó una corrección asociada al costo energético de polarizar la densidad electrónica en cada ambiente (Tabla A.2). Esta corrección, descrita en la Ecuación 3.24, fue añadida al ciclo termodinámico (Figura 3.3) e incluida en la ecuación general de estimación de energía libre absoluta (Ecuación 4.1).

El protocolo de derivación incluyó cinco ciclos de refinamiento de los parámetros no enlazantes y un total de 10 cálculos QM/MM a nivel de teoría B3LYP/def2-TZVP por ciclo, utilizando estructuras extraídas de simulaciones de 2 ns en solvente y en complejo.

La tabla 4.1 presenta los resultados obtenidos tras la modificación del campo de fuerza. Como aproximación, se utilizaron los parámetros de Lennard-Jones derivados en solvente, ya que su variación respecto a los obtenidos en complejo fue inferior al 8 % en estudios previos con bromodominios (Figura B.2). Al igual que en el caso de OFF-2.0.0 original, se observaron diferencias menores a $1 \text{ kcal}\cdot\text{mol}^{-1}$ para 8 de los 11 ligandos, mientras que los compuestos 1, 4 y 8 presentaron desviaciones superiores a $2 \text{ kcal}\cdot\text{mol}^{-1}$.

Tabla 4.1: Valores de energía libre de unión experimental y calculados para el campo de fuerza OFF-2.0.0 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol⁻¹

Sistema	$\Delta G_{bind, exp}$	$\Delta G_{bind, calc}^{OFF-2.0.0}$	$\Delta G_{bind, calc}^{OFF-2.0.0 D-MBIS(q, LJ_{solv})}$
1	-9.8 ± 0.1	-9.8 ± 0.2 ¹ (-6.5 ± 0.2) ²	-5.1 ± 1.3 ¹ (-1.6 ± 1.4) ²
2	-9.6 ± 0.1	-12.1 ± 0.6	-10.0 ± 0.9
3	-9.0 ± 0.1	-9.0 ± 0.4	-9.6 ± 0.9
4	-8.9 ± 0.1	-11.1 ± 0.6 ¹ (-8.8 ± 0.1) ²	-20.3 ± 1.5 ¹ (2.0 ± 1.4) ²
5	-8.8 ± 0.1	-9.6 ± 0.3	-9.3 ± 0.8
6	-8.2 ± 0.1	-10.8 ± 0.6	-8.1 ± 0.9
7	-7.8 ± 0.1	-6.6 ± 0.3	-8.5 ± 0.6
8	-7.4 ± 0.1	-8.9 ± 0.3	-5.3 ± 1.0
9	-7.3	-7.7 ± 0.1	-6.6 ± 0.8
10	-6.3 ± 0.1	-5.4 ± 0.9	-5.0 ± 0.6
11	-5.6	-6.1 ± 0.1	-6.0 ± 0.6

¹ Valores obtenidos aplicando esquema de corrección por efectos electrostáticos en sistemas finitos.

² Valores obtenidos aplicando la metodología DSSB.

Para dar una posible explicación al efecto del cambio de parámetros no enlazantes en la predicción de afinidad, la Figura 4.6 muestra un diagrama de interacciones 2D del ligando 8. Este muestra una interacción de puente de hidrógeno con N140, aminoácido clave en el reconocimiento de histonas. Adicionalmente, se observan interacciones de tipo contacto hidrofóbico de los anillos aromáticos con P82, L92 y I146. Los dos primeros aminoácidos lo hacen directamente con el anillo halogenado.

La subestimación del ligando 8 se podría asociar a una reducción significativa en la profundidad del potencial de Lennard-Jones (ϵ) para los átomos

de nitrógeno (n20) y cloro (n24), como se muestra en la Figura A.2. Esto podría debilitar las interacciones atractivas en el complejo. Adicionalmente, el cambio en valores absolutos de las cargas atómicas del anillo nitrogenado que interacciona con N140 (Figura A.3) podría favorecer el estado solvatado del ligando y en consecuencia, reducir la afinidad predicha. Esto debido a que los parámetros no enlazantes asociados a la contribución de la rama de solvente en el ciclo termodinámico se obtienen desde un modelo de agua ya polarizado (TIP3P), mientras que el entorno proteico podría requerir un ajuste en sus parámetros dada la presencia del ligando en el sitio de unión. Esto debido a que las cargas atómicas en el campo de fuerza de la proteína son obtenidas mediante cálculos en vacío, sin polarización explícita. El uso de campos de fuerza posteriores como ff14sb¹⁵⁷ (compatible con TIP3P) o ff15ipq¹⁵⁸ (con cargas polarizadas implícitamente) podrían ser una potencial solución a este fenómeno.

Una tendencia similar se observó en el ligando 6, el cual comparte el mismo andamiaje que el ligando 8, pero con un cambio en el anillo de siete miembros (Figura 4.6). En este caso, la modificación de los parámetros permitió una mejor concordancia con el experimento, destacando la sensibilidad de la predicción a cambios locales en la distribución electrónica.

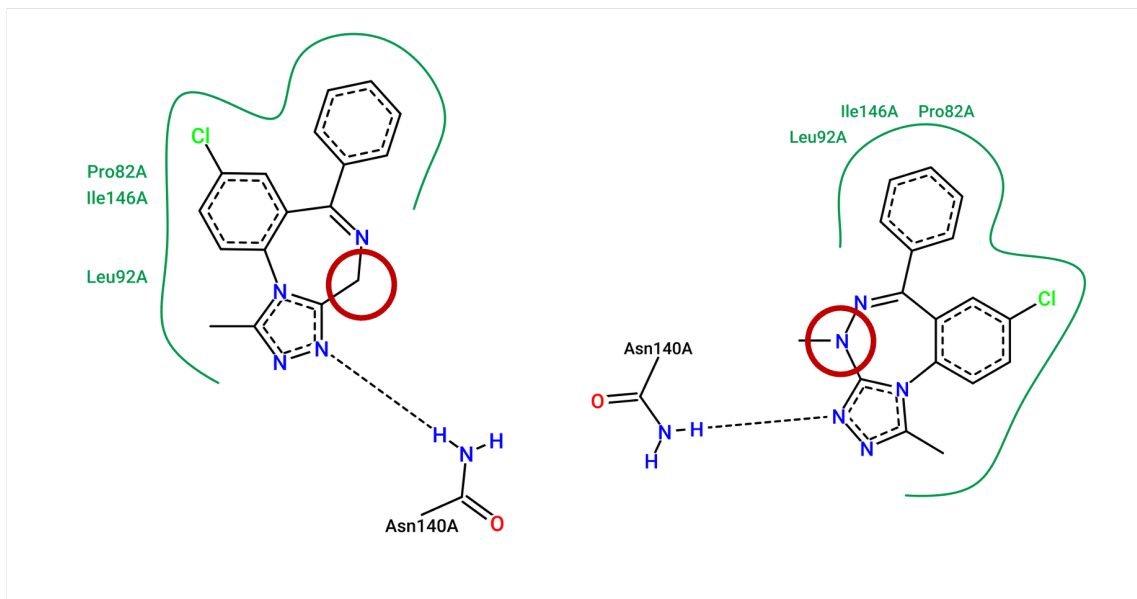


Figura 4.6: **Interacciones de los ligandos 8 y 6 en el sitio de unión de BRD4(1).** A la izquierda se muestran el diagrama 2D para las interacciones del ligando 8 en el sitio de unión de BRD4(1) y a la derecha para el ligando 6. El círculo rojo encierra la modificación al anillo de siete miembros que diferencia a ambos ligandos. La línea discontinua negra y la continua verde representa interacciones de tipo puente de hidrógeno y contactos hidrofóbicos respectivamente.

Por otra parte, desviaciones más pronunciadas fueron observadas en los ligandos cargados (ligandos 1 y 4), cuyos valores predichos fueron -5.1 ± 1.3 y $-20.3 \pm 1.5 \text{ kcal}\cdot\text{mol}^{-1}$ respectivamente.

Estas desviaciones han sido observadas en estudios previos de cálculos de energía libre que involucran cambios de carga en sistemas de caja finita, los cuales generan artefactos tanto a nivel de interacciones de largo alcance como del muestreo en simulaciones clásicas¹⁵⁹. Una posible explicación a la desviación observada con las cargas D-MBIS y la corrección electrostática aplicada, es el tamaño de la caja de simulación y el efecto de la carga neta

en las interacciones de largo alcance entre copias periódicas. Comparaciones entre las distribuciones de cargas parciales muestran diferencias sustanciales entre ambos métodos (Figuras A.4 y A.5), lo que podría amplificar artefactos de caja finita o efectos periódicos en simulaciones con carga neta¹⁵⁵.

Para mitigar el efecto de las desviaciones en la estimación de energía libre de unión al (des)aparecer una carga aplicando el método PME, se decidió implementar la metodología conocida como *Double-System/Single-Box* (DSSB)⁹⁵ (Figura 4.7). Este enfoque permite una estimación más rigurosa de la energía libre de unión en ligandos con carga neta, al evitar los artefactos asociados al uso de cajas de simulación independientes para los estados acoplado y desacoplado.

La idea principal del método DSSB consiste en construir una única caja de simulación que contenga simultáneamente ambos estados del ciclo termodinámico. El ligando en complejo con la proteína es transformado del estado acoplado al desacoplado, mientras que el ligando en solución sufre la transformación inversa, de desacoplado a acoplado. Este enfoque garantiza que la carga neta del sistema se mantenga en cero durante todo el proceso químico, eliminando la necesidad de correcciones electrostáticas posteriores. Para evitar interacciones no deseadas entre ambos ligandos, se imponen restricciones de posición y se mantiene una separación mínima entre ellos.

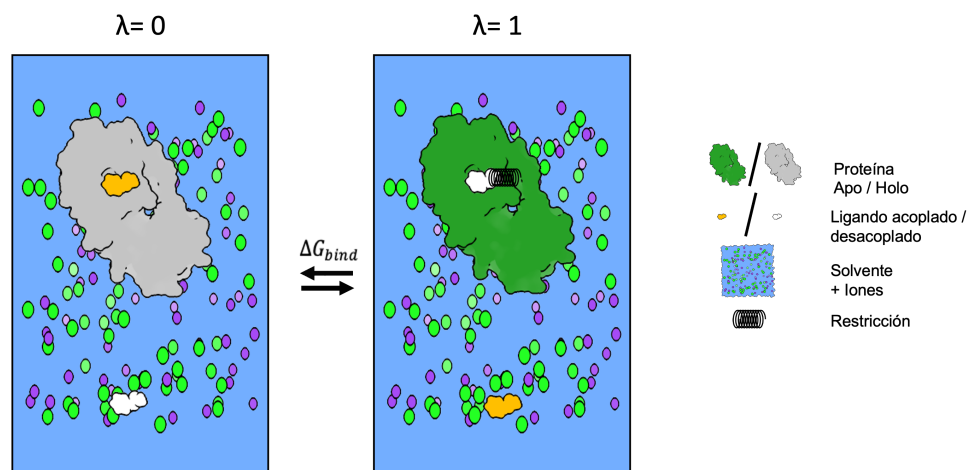


Figura 4.7: Ciclo esquemático de la metodología *Double-System/Single-box* (DSSB).

Los sistemas solvatados y los archivos necesarios para las simulaciones DSSB fueron generados mediante el script `setup_abfe.py` de `pmx`¹⁵⁶, utilizando los argumentos `-doublebox` y `-longest_axis`. Esto permitió ubicar los dos ligandos a lo largo del eje de mayor longitud de la caja, separados por al menos 2.5 nm entre sí y 1.5 nm respecto a las paredes de la caja, en condiciones de neutralidad y concentración iónica de 0.15 M.

Se mantuvieron todos los parámetros de simulación descritos anteriormente, con la excepción de restricciones de posición para el ligando en la simulación de producción y el desacople de las interacciones intramoleculares en el ligando en las transformaciones alquímicas. La energía libre de unión absoluta se estimó mediante la ecuación 4.2, que incluye la contribución dada por la transformación alquímica del ligando en el sitio de unión

del receptor (ΔG_{bind}) y el término de restricción analítico correspondiente al ligando en complejo (ΔG_{restr}).

$$\Delta G^{\circ}_{\text{bind}} = \Delta G_{\text{bind}} + \Delta G^{\circ}_{\text{restr}} \quad (4.2)$$

La tabla 4.1 muestra en paréntesis los resultados obtenidos con OFF-2.0.0 sin modificar para los ligandos 1 y 4. El efecto del cambio metodológico arrojó una disminución en la estimación para el ligando 1, pasando de -9.8 ± 0.2 kcal·mol⁻¹ a -6.5 ± 0.2 kcal·mol⁻¹ y el paso de -11.1 ± 0.6 a -8.8 ± 0.1 kcal·mol⁻¹ para ligando 4. Respecto a la comparación con los valores experimentales, fue desfavorable en el ligando 1, pero beneficioso en el ligando 4, acercándose al valor reportado en literatura.

Se repitió el análisis para el campo de fuerza OFF-2.0.0 modificado con parámetros D-MBIS (Tabla 4.1). En este caso, el cambio fue desfavorable para ambos ligandos, con una considerable subestimación en ambos (-1.6 ± 1.3 kcal·mol⁻¹ y 2.0 ± 1.4 kcal·mol⁻¹ para el ligando 1 y 4 respectivamente).

Estos resultados indican que si bien la metodología DSSB corrige parcialmente los problemas de simulación con ligandos cargados en el campo de fuerza sin modificar, aún persisten desafíos técnicos importantes a evaluar. Entre ellos se encuentran: la separación entre los ligandos en la caja de simulación, el tratamiento de las interacciones intramoleculares en las perturbaciones alquímicas y la mejora del protocolo implementado en los es-

quemadas de integración termodinámica de no equilibrio. Particularmente, el efecto de desacople de interacciones intramoleculares en el protocolo DSSB para el campo de fuerza con parámetros no enlazantes D-MBIS muestra una desviación considerable. La omisión de estas interacciones podría conllevar a una potencial sobrerrepulsión entre pares 1-4 interactuantes, basado en las torsiones del campo de fuerza OFF-2.0.0, las cuales no fueron modificadas y presentan una estrecha relación con los parámetros no enlazantes¹⁶⁰. Todos estos aspectos deberán ser abordados en estudios futuros.

Dada la limitada capacidad predictiva observada en los ligandos cargados (1 y 4) al utilizar parámetros derivados mediante la metodología D-MBIS y la incertidumbre metodológica asociada a su tratamiento electrostático, se decidió realizar un análisis complementario excluyendo estas dos moléculas del conjunto BRD4(1). Esta exclusión permite evaluar el desempeño de los distintos campos de fuerza bajo condiciones más controladas, eliminando las contribuciones dominadas por artefactos de simulación o correcciones electrostáticas inapropiadas en ligandos con carga neta.

La Figura 4.8 muestra los resultados obtenidos para el campo de fuerza OFF-2.0.0 original, excluyendo los ligandos con carga. En este caso, las métricas de desempeño mejoraron notablemente: RMSE = 1.42 kcal·mol⁻¹, MUE = 1.15 kcal·mol⁻¹, $r = 0.84$ y $\tau = 0.67$. Estos valores son comparables a los obtenidos con GAFF-1.5 (Figura A.7), indicando que el campo OFF-2.0.0 entrega predicciones competitivas para ligandos neutros.

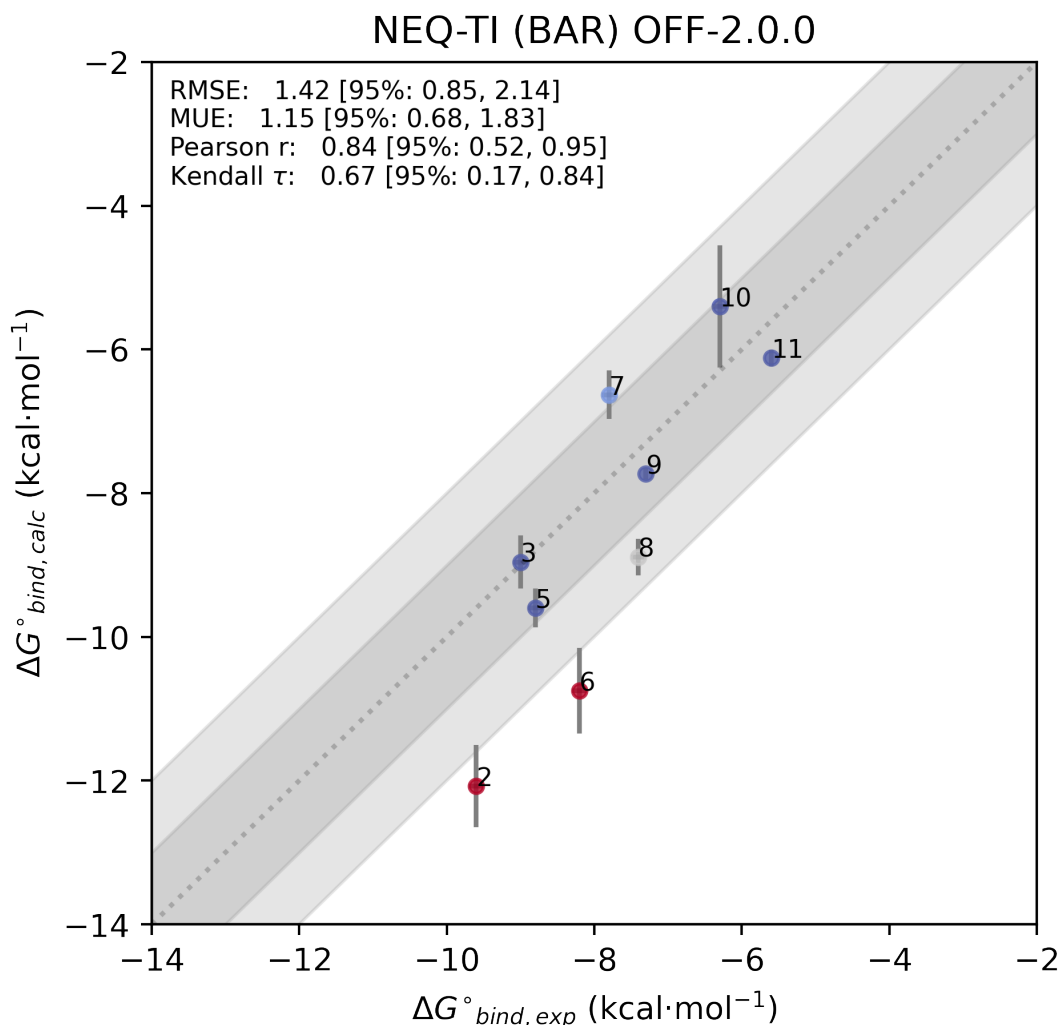


Figura 4.8: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0, sin incluir la estimación de ligandos con carga (1 y 4).

Al aplicar la misma exclusión al campo de fuerza OFF-2.0.0 modificado con parámetros D-MBIS, se obtuvo el mejor desempeño global de todos los modelos evaluados (Figura 4.9). Las métricas asociadas fueron: RMSE = 0.94 kcal·mol⁻¹, MUE = 0.76 kcal·mol⁻¹, $r = 0.89$ y $\tau = 0.78$. Estas cifras

indican no solo una mejora en la precisión absoluta, sino también una mejor concordancia con el orden jerárquico de afinidades experimentales, lo cual es clave para aplicaciones en diseño racional de fármacos.

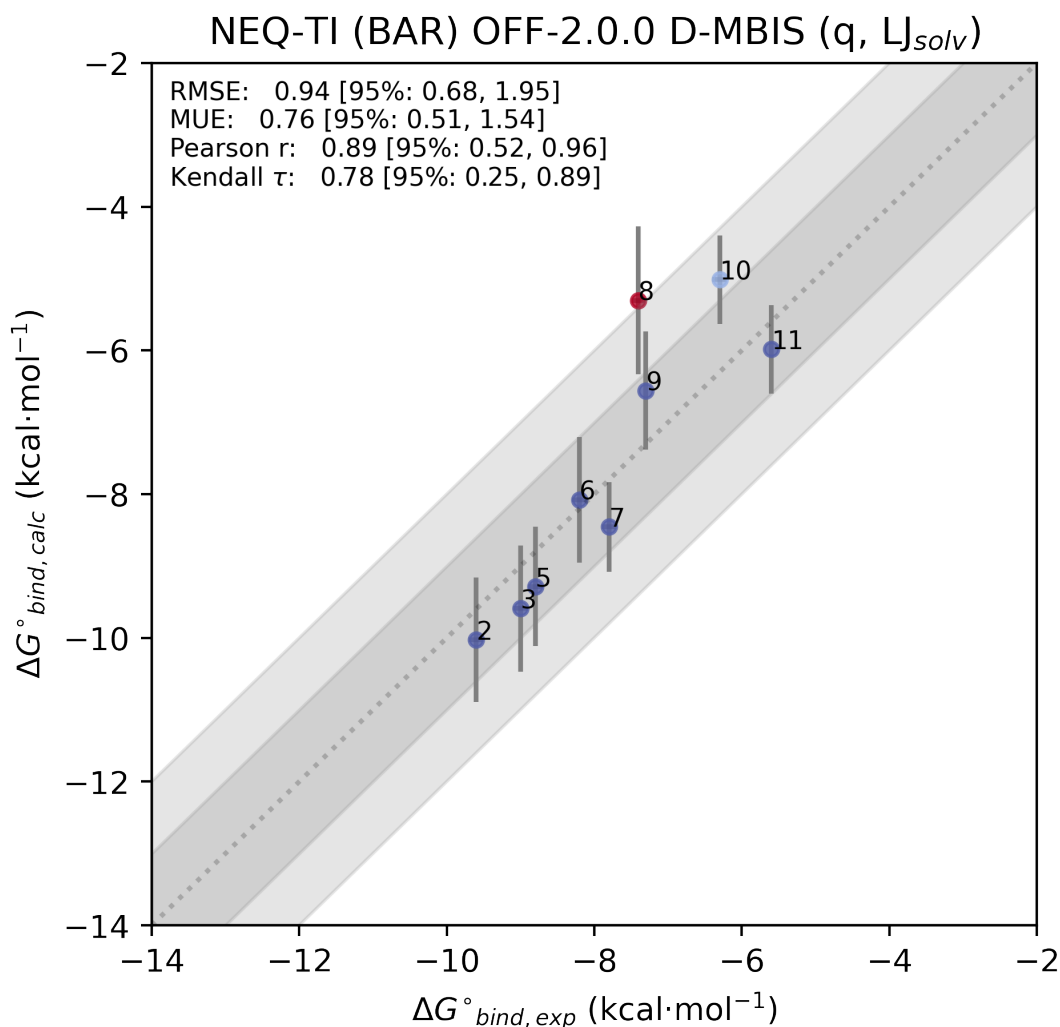


Figura 4.9: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 modificado, sin incluir la estimación de ligandos con carga (1 y 4).

En conjunto, estos resultados posicionan a la versión modificada de OFF-2.0.0 con parámetros no enlazantes derivados desde D-MBIS como la mejor alternativa para la predicción de energía libre de unión absoluta en ligandos neutros del conjunto BRD4(1), superando tanto a GAFF-1.5 como a la versión original de OFF-2.0.0, alcanzando errores absolutos en la estimación menores a $1 \text{ kcal}\cdot\text{mol}^{-1}$. Sin embargo, para ligandos cargados, persisten desafíos metodológicos significativos, lo cual abre una línea de trabajo relevante para futuras investigaciones.

4.3. Perfil de selectividad de Bromosporina

Bajo la misma línea de bromodominios, se buscó evaluar la afinidad de una molécula frente a un conjunto de receptores proteicos y generar un perfil de selectividad *in silico*. En 2016, Picaud y colaboradores interesados en blancos terapéuticos para la terapia contra el cáncer, desarrollaron la molécula bromosporina. Esta molécula (Figura 1.2) es un potente inhibidor promiscuo de la familia BRD humana, con afinidad nanomolar para múltiples bromodominios y una fuerte actividad antiproliferativa¹⁶¹. Estructuras resueltas para el complejo de bromosporina y algunos bromodominios^{161–165} (Figura 4.10), muestran una orientación mimética al motivo acetil-lisina dentro de la cavidad, promoviendo interacciones con la cadena lateral de una asparagina evolutivamente conservada en la familia BRD (N100 en BRD9). El grupo sulfonamida presenta contactos con los residuos del bucle ZA (principalmente I53 en BRD9), estabilizando aún más la interacción sin desplazar ninguna de las moléculas de agua estructuralmente conservadas en el sitio de unión.

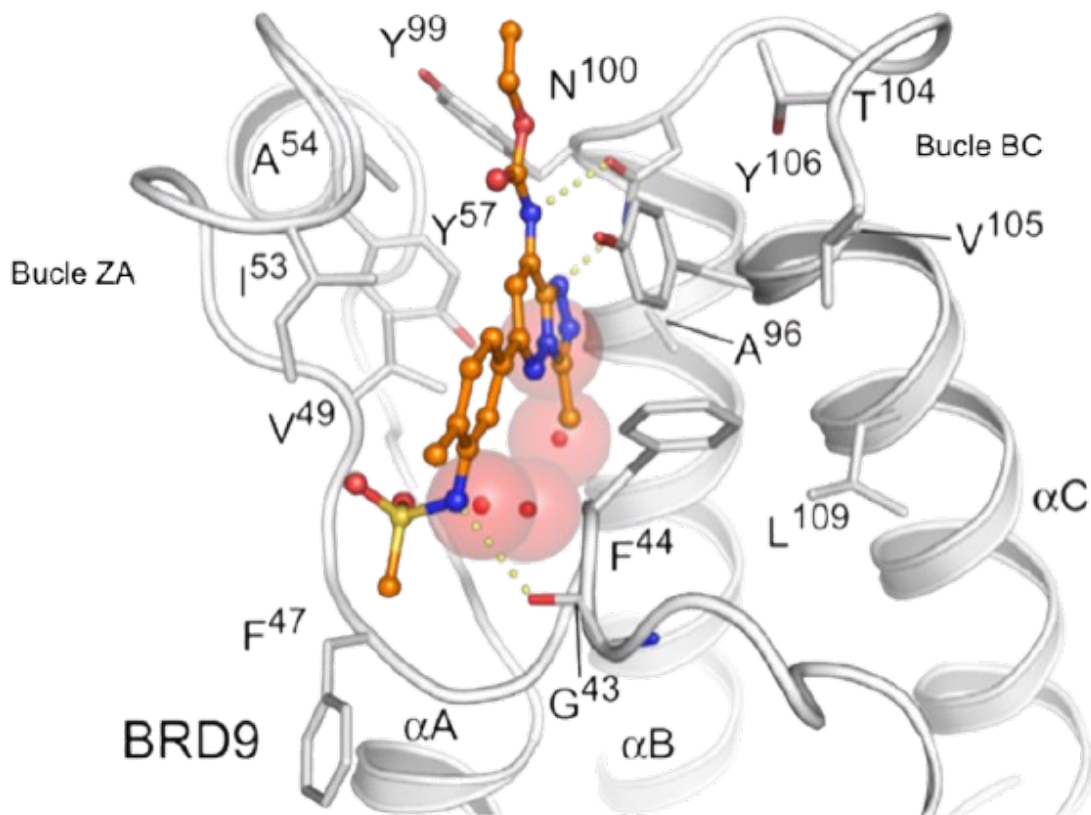


Figura 4.10: **Estructura cristalina resuelta para el complejo formado por bromosporina y el bromodominio BRD9.** Se muestran las principales interacciones del ligando bromosporina y el sitio de unión de BRD9. Adaptado de Picaud y colaboradores¹⁶¹

Basado en la evidencia existente, Aldeghi y colaboradores generaron un conjunto de datos que abarca a 7 familias de estos receptores, con un total de 22 complejos sintéticos de bromodominios y bromosporina, para los cuales se cuenta con valores de afinidad experimental¹³. Cabe destacar que las estructuras de los complejos proteína-ligando del conjunto de datos fueron modeladas mediante alineamiento estructural y posterior relajación del sistema mediante simulaciones de dinámica molecular, asumiendo una pose

conservada para todos los bromodominios. Esto fue posible usando como referencia la pose con mayor afinidad basado en cálculos de energía libre de unión, usando distintos *clusters* generados mediante *docking* sobre la estructura cristalina de BRD4(1) en estado apo (PDB ID: 2oss).

Similar al conjunto de datos de BRD4(1), los archivos requeridos por GROMACS (.gro y .top) de los complejos proteína-ligando y de bromosporina en solvente para los 22 bromodominios estudiados, fueron obtenidos desde la información suplementaria de la publicación de Aldeghi y colaboradores¹² (https://pubs.acs.org/doi/suppl/10.1021/jacs.6b11467/suppl_file/ja6b11467_si_002.zip). Los valores experimentales fueron extraídos desde esta misma publicación. En la tabla B.1 se presenta un resumen del conjunto de datos utilizado.

Las simulaciones fueron ejecutadas con GROMACS 2022.4¹⁴⁵. Se utilizaron los mismos campos de fuerza (proteína, iones) y modelo de agua que el conjunto de datos BRD4(1) (Sección 4.2). Los parámetros de los ligandos analizados fueron obtenidos desde GAFF-1.7²⁶ y OFF-2.0.0^{34,35} con cargas AM1-BCC^{150,151} para la validación de la metodología de estimación de energía libre de unión absoluta y la derivación de parámetros no enlazantes respectivamente. Los valores de parámetros no enlazantes, fueron modificados en el archivo .top, tras la aplicación del protocolo de derivación con *ffparaim* y así permitir la evaluación comparativa respecto al campo de fuerza original. Las simulaciones de producción fueron idénticas al conjunto de datos

BRD4(1) (Sección 4.2), a excepción del umbral de corte para la electrostática de largo alcance (1.2 nm) y el tratamiento de interacciones de Van der Waals, con un potencial ajustado con umbral de corte a 1.2 nm. Dada las dimensiones del ligando en cuestión, se utilizó una definición del valor de corte para la lista de vecinos de 1.8 nm.

Para el cálculo de energía libre de unión absoluta, se aplicó el mismo ciclo termodinámico esbozado por Aldeghi y colaboradores¹² (Ecuación 3.23).

Igualmente que en el caso anterior, se procedió la validación del protocolo de simulación basado en el trabajo de Gapsys y colaboradores¹⁴. Se generaron 3 réplicas independientes para el ligando en solvente y el sistema proteína-ligando con el campo de fuerza GAFF-1.7. Para todos los sistemas, se aplicó una primera etapa de minimización energética (10000 pasos del algoritmo *steepest descent*) con posteriores 0.5 ns de equilibrio en colectivo NVT. Las simulaciones de producción en colectivo NPT se extendieron por 15 ns para ambos estados (ligando en solvente y en complejo). En las transformaciones de no equilibrio, se ejecutaron 100 transiciones (0.5 ns cada transición) para calcular ΔG_{solv} y ΔG_{prot} usando configuraciones extraídas cada 100 ps desde la trayectoria de producción. De esta última, se descartaron los primeros 5 ns considerados como equilibrio del sistema en este colectivo. Al igual que el caso anterior, la estimación de energía libre de unión fue obtenida a través del método BAR con *bootstrapping* de 100 muestras usando pmx. Se consideró el valor reportado como el promedio de las réplicas y el error

estándar respectivo. Se usaron los mismos descriptores estadísticos (MUE, RMSE, r de Pearson, τ de Kendall) para la evaluación del error y correlación.

A modo de validación del protocolo a utilizar, se procedió a realizar la evaluación comparativa respecto a los resultados obtenidos por Gapsys y colaboradores. En la Figura 4.11 se presenta un gráfico de paridad de los resultados obtenidos para los distintos bromodominios. El error asociado al modelo (RMSE = $1.21 \text{ kcal}\cdot\text{mol}^{-1}$, MUE= $1.06 \text{ kcal}\cdot\text{mol}^{-1}$) indica una leve divergencia entre los resultados obtenidos, observado más claramente para los receptores BRD3(1) (2c) y PB1(5) (8a), con una desviación mayor de $\sim 2 \text{ kcal}\cdot\text{mol}^{-1}$. Esta diferencia puede ser atribuída al uso de estructuras iniciales predichas mediante *docking*, el comportamiento caótico de las simulaciones de dinámica molecular, posibles falta de equilibración de los complejos modelados y el número de réplicas utilizadas para reportar estos valores (3 vs 5 réplicas). Sin embargo, la correlación de Pearson ($r=0.81$) y Kendall ($\tau = 0.63$) sugieren un ajuste razonable de los datos a un modelo lineal, asumiendo un error basal en la predicción de $\sim 1 \text{ kcal}\cdot\text{mol}^{-1}$.

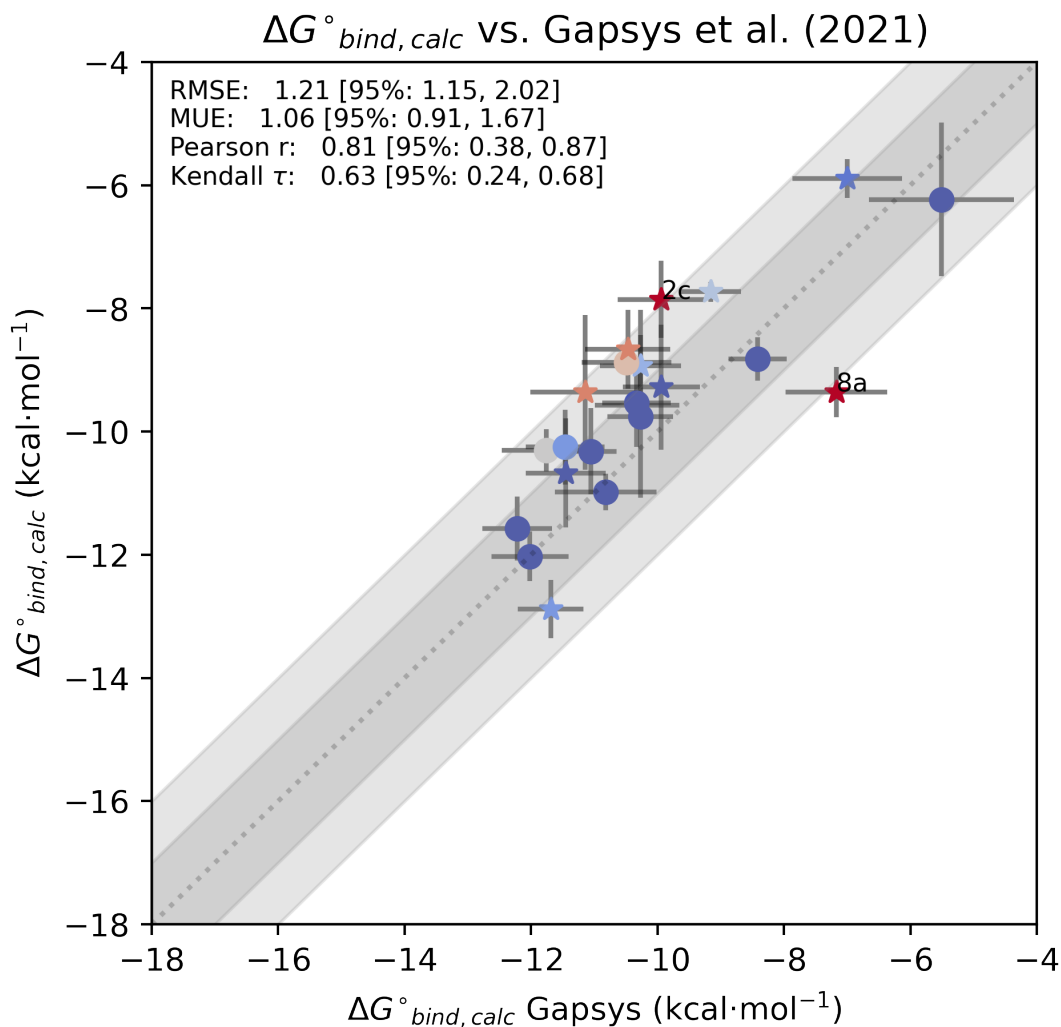


Figura 4.11: **Gráfico de paridad de valores de energía libre de unión absoluta obtenidos en este estudio para el conjunto de datos de bromosporina, respecto a los publicados por Gapsys y colaboradores.** Se incluye la etiqueta para los bromodominios con una desviación mayor a 2 kcal·mol⁻¹

En la Figura B.1 se presentan los resultados obtenidos para la estimación de energía libre de unión absoluta con el campo de fuerza GAFF-1.7 y cargas

atómicas AM1-BCC. Se obtuvo diferencias en los valores predichos respecto al experimental mayores a $2 \text{ kcal}\cdot\text{mol}^{-1}$ para 8 de los 22 bromodominios estudiados, con una sobreestimación de la afinidad para estos receptores. Los descriptores estadísticos de error ($\text{RMSE} = 2.11 \text{ kcal}\cdot\text{mol}^{-1}$, $\text{MUE} = 1.70 \text{ kcal}\cdot\text{mol}^{-1}$) se encuentran en rangos aceptables respecto a estudios anteriores^{13,14}. Se observa una correlación lineal y ordinal positiva, con valores de 0.31 y 0.19 respectivamente.

Bajo las consideraciones anteriores, se evaluó el efecto del cambio en el campo de fuerza del ligando. La Figura 4.12 exhibe los resultados obtenidos para la estimación de energía libre de unión absoluta con el campo de fuerza OFF-2.0.0 y cargas atómicas AM1-BCC. Estos muestran una diferencia menor a $2 \text{ kcal}\cdot\text{mol}^{-1}$ respecto al valor experimental para 9 de los 22 bromodominios en estudio. Los descriptores estadísticos de error ($\text{RMSE} = 2.65 \text{ kcal}\cdot\text{mol}^{-1}$, $\text{MUE} = 2.36 \text{ kcal}\cdot\text{mol}^{-1}$) muestran un deterioro en la predicción al cambio de campo de fuerza, manteniendo una baja correlación lineal y limitada discriminación en selectividad con valores de $r=0.38$ y $\tau=0.40$.

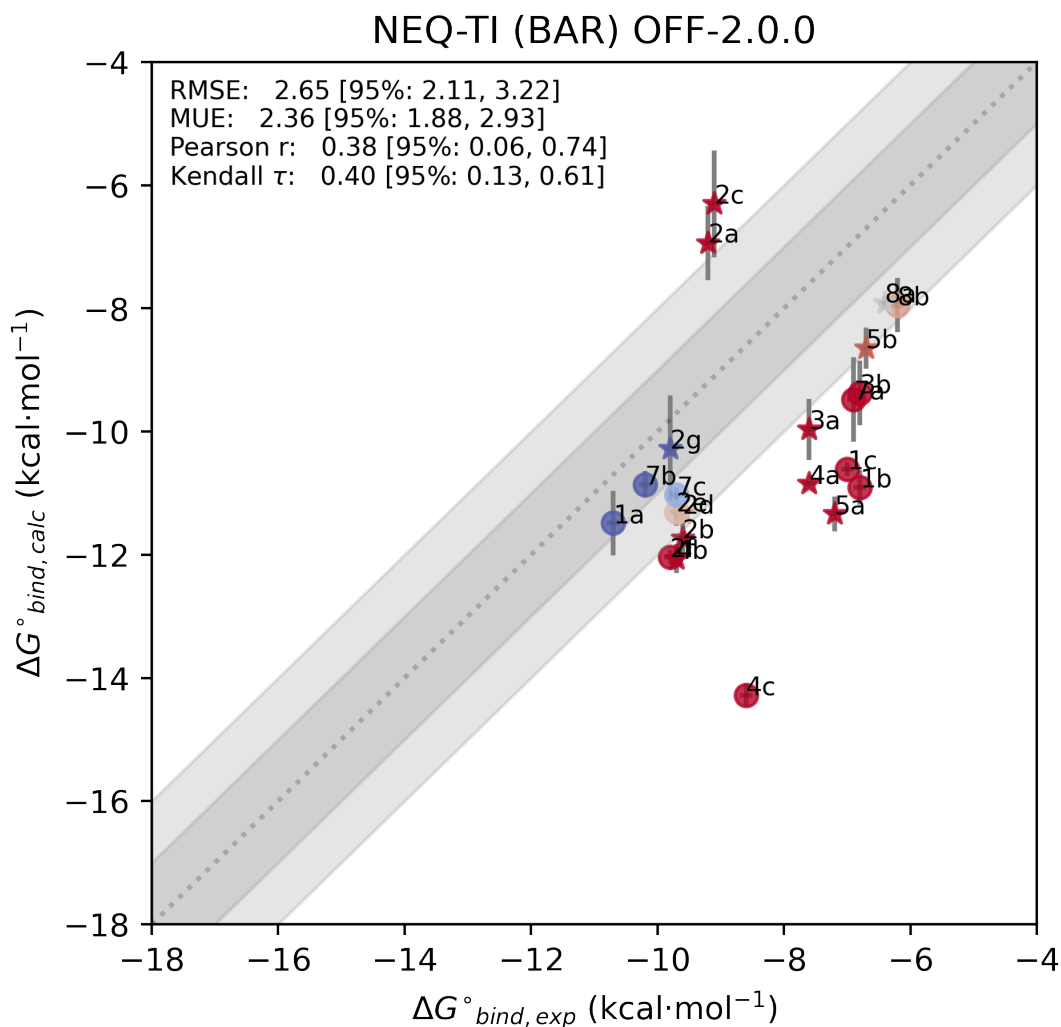


Figura 4.12: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y cargas atómicas AM1-BCC en el conjunto de datos de bromosporina.

Con la motivación de estudiar el efecto de parámetros no enlazantes, se procedió a la modificación de parámetros no enlazantes usando el protocolo de `ffparaim`. El protocolo involucró 5 actualizaciones de los paráme-

tros no enlazantes, con un total de 25 cálculos QM/MM a nivel de teoría B3LYP/def2-TZVP usando configuraciones extraídas de simulaciones de 5 ns en solvente y 10 ns en el complejo proteína-ligando. La corrección por polarización fue incluida en el ciclo termodinámico al igual que para el sistema anterior (Anexo: Tabla B.2 y B.3).

Con el propósito de reducir el costo computacional de la derivación de parámetros exhaustiva para bromosporina en los 22 bromodominios, en primera instancia se decidió evaluar de forma cualitativa el potencial electrostático de superficie de cada receptor (Figura 4.13) y seleccionar 3 de ellos que abarcaran los extremos en distribución de carga (positiva, negativa y neutra) para las cavidades de unión. Para su obtención se utilizó el *plugin* de PyMOL *APBS Electrostatics*, usando parámetros por defecto para la asignación de cargas parciales, el cálculo del mapa electrostático y su posterior visualización en la superficie proteica. De acuerdo a esto, se seleccionaron los bromodominios EP300, FALZ y TAF1(2) respectivamente y se procedió a la derivación de parámetros no enlazantes.

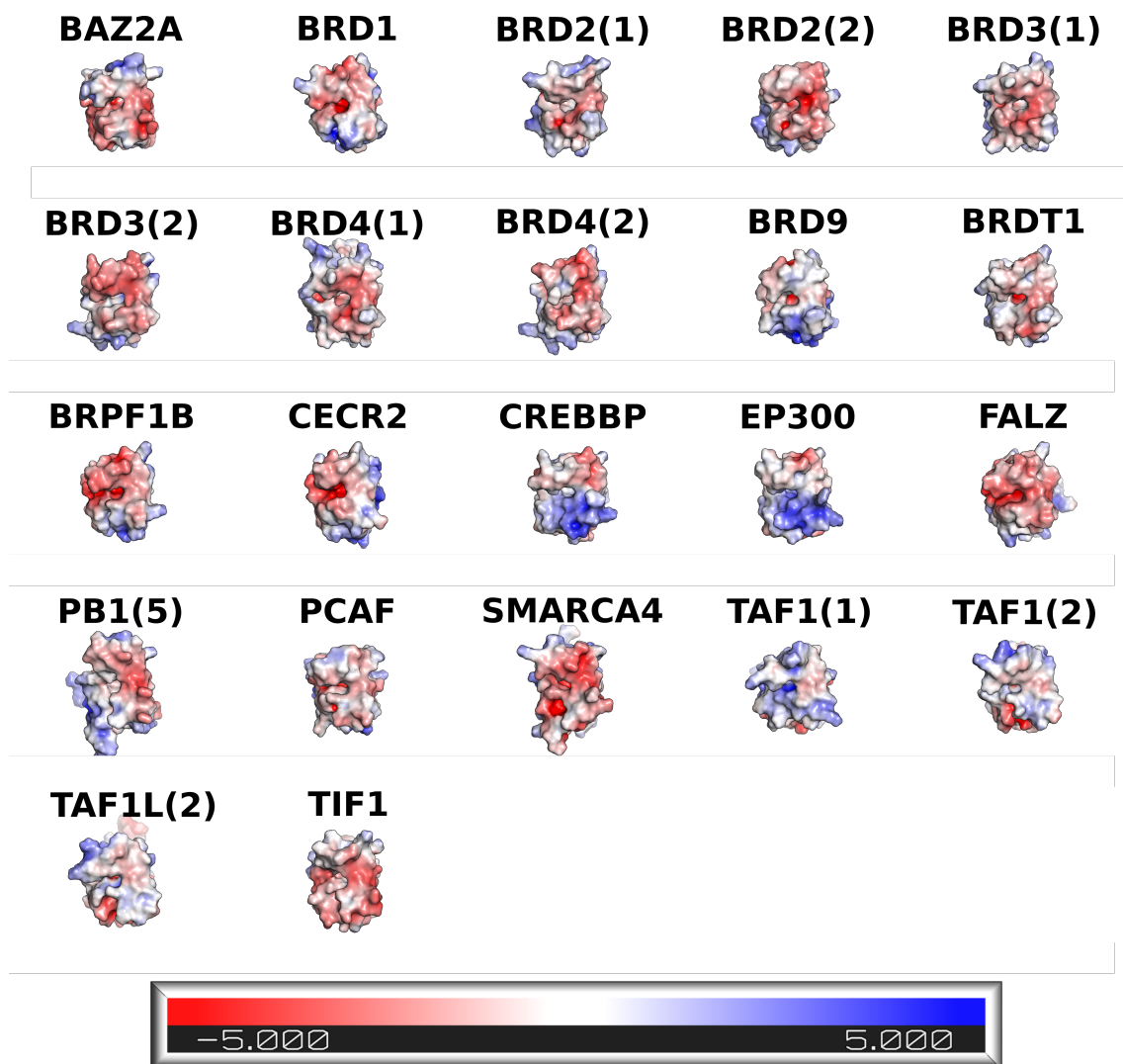


Figura 4.13: Potencial electrostático de superficie de los 22 bromodominios en estudio, en un rango de valores de -5 a +5 kT/e. La orientación mostrada corresponde a una vista superior del sitio de unión de bromosporina

Una comparación entre los valores de cargas atómicas obtenidas para cada receptor se muestra en la Figura 4.14. Para la mayoría de los átomos se observan fluctuaciones menores a 0.04 e. En base a este resultado, se propuso como primera aproximación el uso de cargas atómicas promedio para la

contribución del complejo en el ciclo termodinámico, usando los receptores seleccionados previamente como referencias.

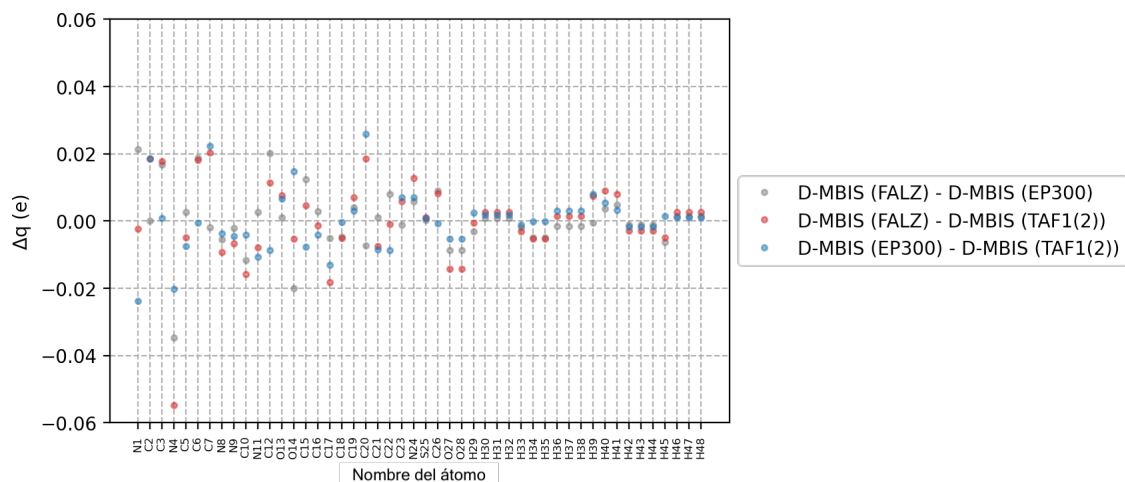


Figura 4.14: Valores de diferencia de las cargas atómicas D-MBIS entre los brodom dominios seleccionados.

Respecto a parámetros de Lennard-Jones, la Figura 4.15 muestra la comparación para los valores de σ y ϵ entre el campo de fuerza OFF-2.0.0 y los obtenidos desde el protocolo de derivación. La mayor desviación en el radio de Van der Waals (σ) se observa para el átomo de SMIRKS id n11, correspondiente al hidrógeno unido a nitrógeno. Por otra parte, las mayores desviaciones en el valor para la profundidad del potencial (ϵ) se observan para los átomos de SMIRKS id 17, 18, 20 (grupo carbamoyl) y n21 (azufre en grupo sulfonamida). El análisis de los valores obtenidos para bromosporina entre parámetros D-MBIS en complejo con la proteína y en solvente, muestran diferencias potencialmente despreciables, lo que permitió utilizar los pa-

rámetros de Lennard-Jones derivados desde solvente como una aproximación a la descripción de interacciones de dispersión y repulsión/intercambio y facilitando la implementación del desacople de interacciones inter e intramoleculares en el ciclo termodinámico, similar a lo realizado para los campos de fuerza utilizados anteriormente.

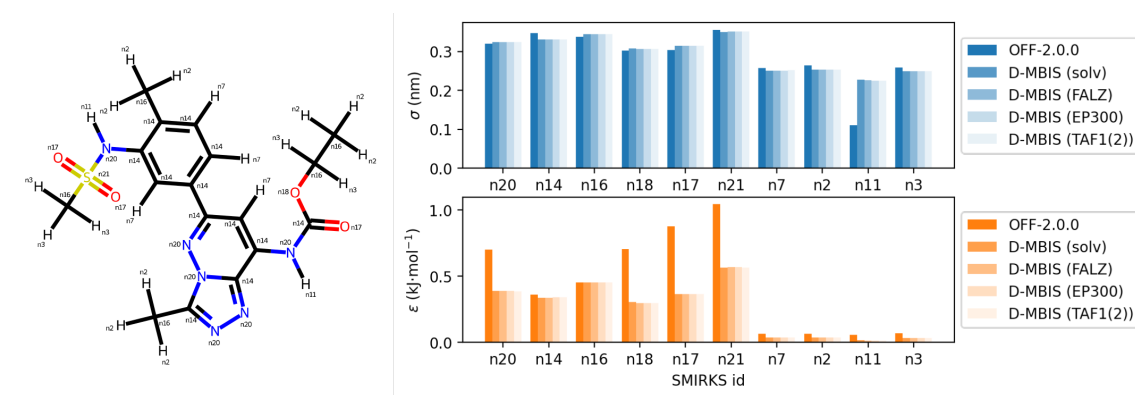


Figura 4.15: Valores de parámetros de Lennard-Jones (σ y ϵ) del campo de fuerza OFF-2.0.0 y los obtenidos en ambientes químicos de solvente y complejo proteína-ligando de los bromodominios seleccionados. A la izquierda se muestra una representación 2D de bromosporina, con los átomos etiquetados de acuerdo a la asignación de id desde patrones SMIRKS.

De forma tal de validar la aproximación a nivel de parámetros de Lennard-Jones, se volvió a ejecutar el protocolo de derivación, usando los valores de σ y ϵ obtenidos desde solvente y generando un nuevo set de cargas atómicas para los 3 receptores seleccionados anteriormente. La diferencia entre los valores para los distintos pares de receptores se muestran en la Figura 4.16. Se observa la misma tendencia en la diferencia de cargas atómicas, donde sólo un átomo (N4) presenta una desviación mayor a 0.04 e. Evaluaciones

preliminares del ambiente químico de este átomo, muestra diferentes patrones de solvatación para las configuraciones utilizadas en la partición de densidad electrónica, lo cual podría explicar las fluctuaciones de las cargas atómicas generadas.

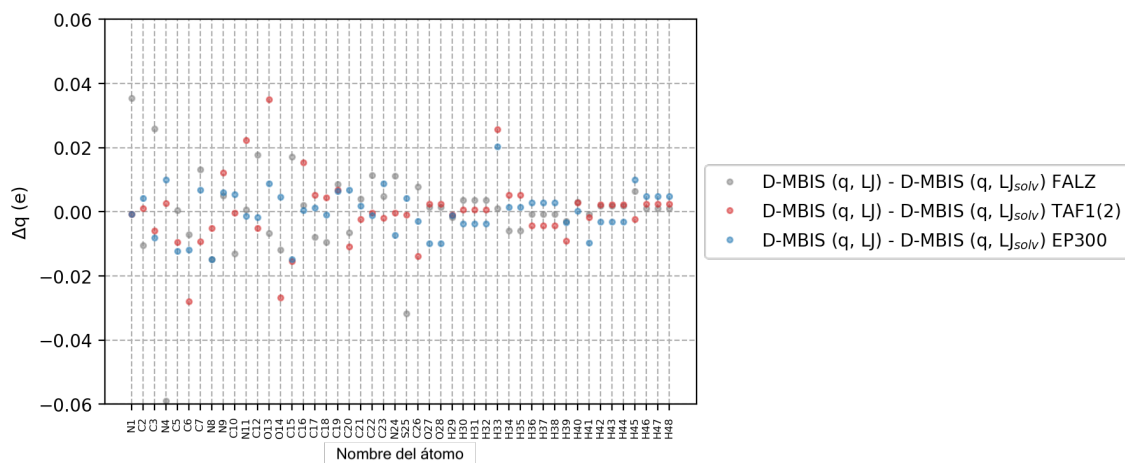


Figura 4.16: Valores de diferencia de las cargas atómicas D-MBIS entre los bromodominios seleccionados, usando parámetros de Lennard-Jones derivados desde el complejo y en solvente.

Con lo concluido respecto a la derivación de parámetros, se procedió a evaluar el efecto de modificar los parámetros de cargas atómicas y parámetros de Lennard-Jones con las aproximaciones anteriormente propuestas, incluyendo un promedio de la energía de polarización obtenida desde los receptores usados en la derivación. En la Figura 4.17 se presenta la comparación entre los valores de energía libre de unión calculados para los 22 bromodominios y su valor experimental reportado. Estos muestran una diferencia menor a $2 \text{ kcal}\cdot\text{mol}^{-1}$ respecto al valor experimental para 3 de los 22 bromodominios

en estudio. A diferencia de los campos de fuerza evaluados anteriormente, se presenta una sobrestimación notoria de la energía libre de unión para la mayoría de los receptores evaluados. Los descriptores estadísticos de error (RMSE = 4.56 kcal·mol⁻¹, MUE = 4.03 kcal·mol⁻¹) muestran un deterioro importante en la estimación a causa del cambio de parámetros no enlazantes en el campo de fuerza, manteniendo una baja correlación lineal y limitada discriminación jerárquica con valores de $r = 0.36$ y $\tau = 0.27$.

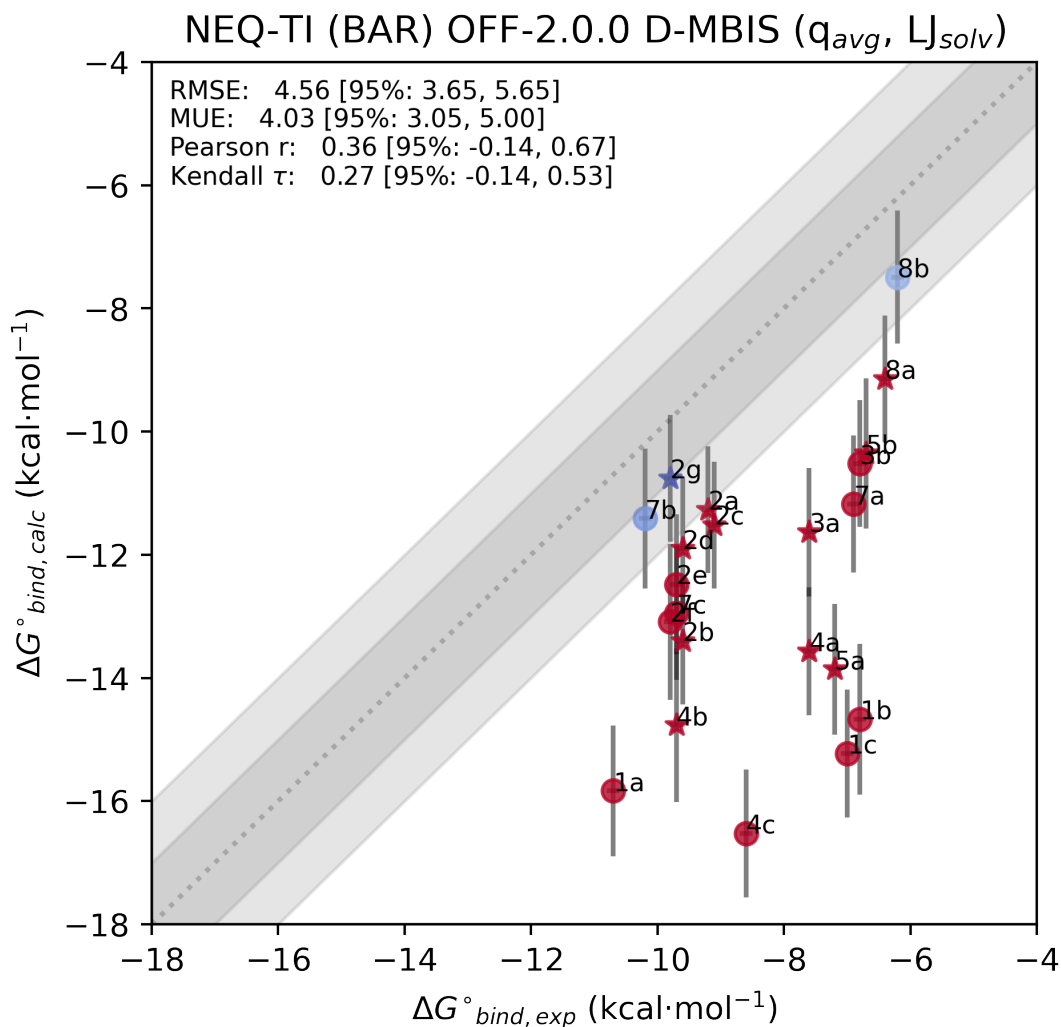


Figura 4.17: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS (cargas atómicas promedio de bromodominios y parámetros de Lennard-Jones derivados desde solvente).

En base a la inexistente mejora a nivel de métricas estadísticas, se consideró una aproximación con sustento mayormente biológico para la deri-

vación de parámetros no enlazantes asociados al complejo proteína-ligando. Esta consideró la relación evolutiva entre bromodominios, asociado a la pertenencia de cada uno a cierta familia de proteínas. De acuerdo a esto, se tomó un representante de cada una de forma arbitraria (Familia I: FALZ; Familia II: BRD4(1); Familia III: EP300; Familia IV: BRD1, Familia V: TIF1; Familia VII: TAF1(2); Familia VIII: PB1(5)), procediendo a las cargas atómicas y energías de polarización de bromosporina en los correspondientes ambientes químicos. Estos valores de cargas y de energía de polarización, fueron usados en reemplazo de las cargas atómicas y energía de polarización promedio. En la Figura B.3 se muestra el efecto puntual del cambio de cargas atómicas para estos bromodominios. Para el caso de TAF1(2) y PB1(5) se observa una mejora en la estimación de afinidad, mientras que para EP300 y TIF1 la estimación empeora respecto a las cargas promedio. Para los restantes bromodominios a los cuales se aplicó la derivación de parámetros específica (FALZ, BRD4(1), BRD1) no se observan cambios destacables en la predicción.

A modo general, en la Figura 4.18 se presenta un gráfico de paridad tras realizar las estimaciones de energía libre de unión con esta nueva aproximación. Estos muestran una diferencia menor a $2 \text{ kcal}\cdot\text{mol}^{-1}$ respecto al valor experimental para 3 de los 22 bromodominios en estudio. A diferencia del uso de cargas atómicas promedio y el campo de fuerza OFF-2.0.0 sin modificaciones, se observa un perfil más complejo sesgado a la sobreestimación de la energía libre de unión. Se destacan CECR2 (1a), FALZ (1b), PCAF

(1c) y BRPF1B (4c), pertenecientes a la familia I y IV respectivamente, los cuales presentan una notoria sobreestimación de afinidad en comparación a los campos de fuerza sin modificaciones a los parámetros no enlazantes. Los descriptores estadísticos de error (RMSE = $4.59 \text{ kcal}\cdot\text{mol}^{-1}$, MUE = $3.95 \text{ kcal}\cdot\text{mol}^{-1}$) no muestran mejoras en la estimación de energía libre de unión respecto al uso de la energía de polarización y de cargas atómicas promedio, y una disminución de la correlación lineal y ordinal, con valores de $r = 0.17$ y $\tau = -0.01$. Ninguna de las aproximaciones realizadas en la derivación de parámetros, logra mejorar la predicción realizada por el campo de fuerza OFF-2.0.0 sin modificaciones a los parámetros de Lennard-Jones en conjunto al uso de cargas atómicas AM1-BCC.

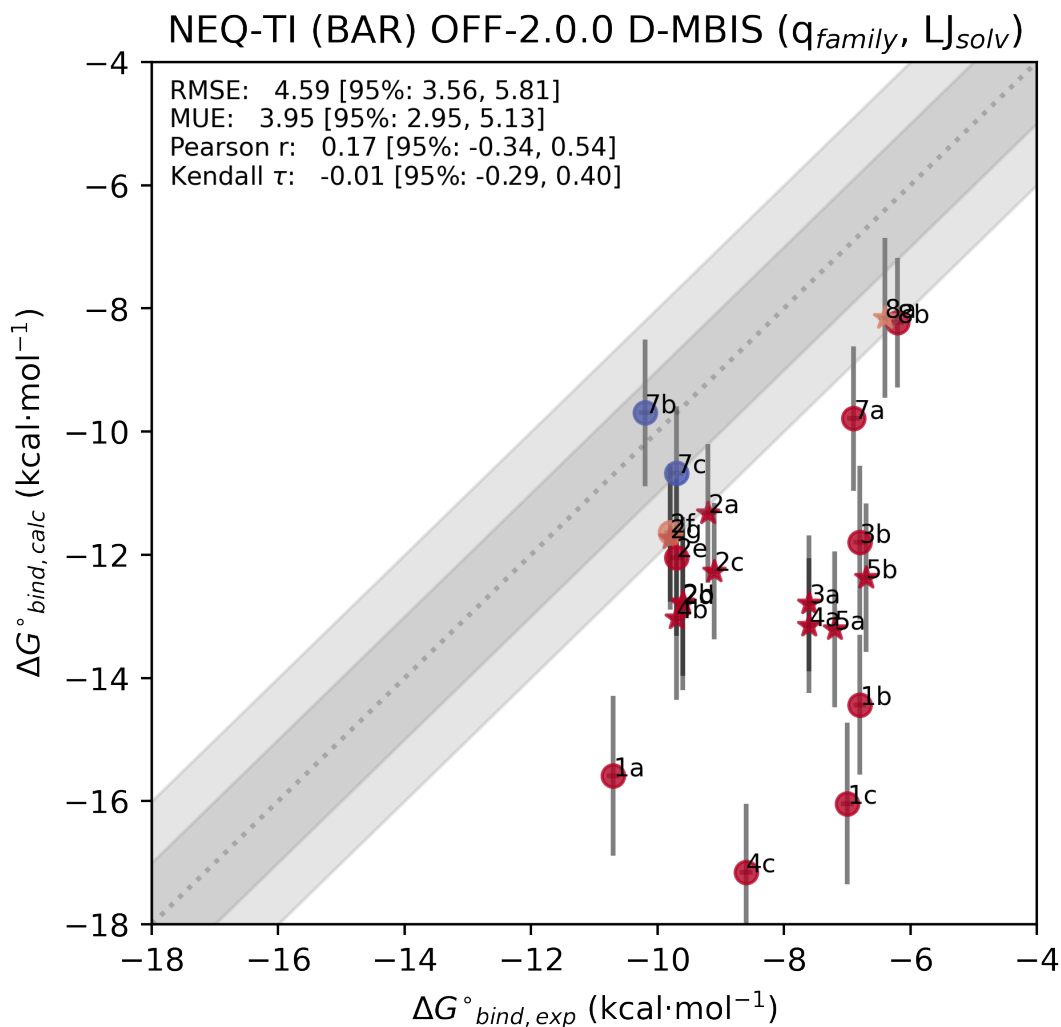


Figura 4.18: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS (cargas atómicas por familia de bromodominios y parámetros de Lennard-Jones derivados desde solvente).

A modo de resumen, se incluye una tabla con los valores obtenidos con estas aproximaciones.

Tabla 4.2: Valores de energía libre de unión experimental y calculados para el campo de fuerza OFF-2.0.0 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol⁻¹.

Sistema	ID	$\Delta G_{\text{bind, exp}}^{\circ}$	$\Delta G_{\text{bind, calc}}^{\circ}$ OFF-2.0.0	$\Delta G_{\text{bind, calc}}^{\circ}$ OFF-2.0.0	$\Delta G_{\text{bind, calc}}^{\circ}$ OFF-2.0.0
				D-MBIS (q_{avg} , LJ_{solv})	D-MBIS (q_{family} , LJ_{solv})
CECR2	1a	-10.7 ± 0.1	-11.5 ± 0.5	-15.8 ± 1.1	-15.6 ± 1.3
FALZ ¹	1b	-6.8 ± 0.1	-10.9 ± 0.3	-14.7 ± 1.2	-14.4 ± 1.1
PCAF	1c	-7.0 ± 0.1	-10.6 ± 0.1	-15.2 ± 1.0	-16.0 ± 1.3
BRD2(1)	2a	-9.2 ± 0.1	-6.9 ± 0.6	-11.3 ± 1.0	-11.3 ± 1.1
BRD2(2)	2b	-9.6 ± 0.1	-9.7 ± 0.2	-13.4 ± 1.0	-12.8 ± 1.2
BRD3(1)	2c	-9.1 ± 0.1	-11.7 ± 0.3	-11.5 ± 1.0	-12.3 ± 1.1
BRD3(2)	2d	-9.6 ± 0.1	-11.3 ± 0.1	-11.9 ± 1.2	-12.8 ± 1.4
BRD4(1) ¹	2e	-9.7 ± 0.1	-11.3 ± 0.2	-12.5 ± 1.1	-12.0 ± 1.3
BRD4(2)	2f	-9.8 ± 0.1	-12.0 ± 0.1	-13.1 ± 1.3	-11.6 ± 1.1
BRDT(1)	2g	-9.8 ± 0.1	-10.3 ± 0.9	-10.8 ± 1.0	-11.7 ± 1.2
CREBBP	3a	-7.6 ± 0.1	-10.0 ± 0.5	-11.6 ± 1.0	-12.8 ± 1.1
EP300 ¹	3b	-6.8 ± 0.1	-9.4 ± 0.5	-10.5 ± 1.0	-11.8 ± 1.3
BRD1 ¹	4a	-7.6 ± 0.1	-10.8 ± 0.1	-13.6 ± 1.0	-13.2 ± 1.1
BRD9	4b	-9.7 ± 0.1	-12.1 ± 0.2	-14.8 ± 1.2	-13.0 ± 1.3
BRPF1B	4c	-8.6 ± 0.1	-14.3 ± 0.2	-16.5 ± 1.0	-17.2 ± 1.1
BAZ2A	5a	-7.2 ± 0.1	-11.3 ± 0.3	-13.9 ± 1.1	-13.2 ± 1.3
TIF1 ¹	5b	-6.7 ± 0.1	-8.6 ± 0.3	-10.4 ± 1.2	-12.4 ± 1.2
TAF1(1)	7a	-6.9 ± 0.1	-9.5 ± 0.7	-11.2 ± 1.1	-9.8 ± 1.2
TAF1(2) ¹	7b	-10.2 ± 0.1	-10.9 ± 0.2	-11.4 ± 1.1	-9.7 ± 1.2
TAF1L(2)	7c	-9.7 ± 0.1	-11.0 ± 0.1	-12.9 ± 1.1	-10.7 ± 1.1
PB1(5) ¹	8a	-6.4 ± 0.1	-7.9 ± 0.1	-9.1 ± 1.0	-8.2 ± 1.3
SMARCA4	8b	-6.2 ± 0.1	-7.9 ± 0.4	-7.5 ± 1.1	-8.2 ± 1.1

¹ Bromodominios usados en la derivación de parámetros en complejo proteína-ligando.

De acuerdo a los resultados anteriores y a las aproximaciones realizados, la modificación de parámetros no enlazantes para el campo de fuerza OFF-2.0.0 no mejoran la predicción del perfil de selectividad de bromosporina en el conjunto de datos de bromodominios evaluado. Dado este fenómeno, se

consideraron otros factores que podrían influir en la baja capacidad discriminativa observada en los distintos campos de fuerza explorados, tanto a nivel estructural y del protocolo de integración termodinámica de no equilibrio.

Para descartar desviaciones provocadas por el método de predicción de la pose utilizada para la simulación y posterior estimación de energía libre de unión, se seleccionaron estructuras de bromodominios en complejo con bromosporina resueltas experimentalmente (Tabla B.4). Estos bromodominios seleccionados además debían contar con la solución cristalográfica del dominio en estado apo. En la Figura 4.19 se muestran los bromodominios seleccionados para la estimación.



Familia II: BRD2(1), BRD2(2), BRD3(2), BRD4(1)

Familia IV: BRD9, BRPF1B

Familia VII: TAF1(2), TAF1L(2)

Figura 4.19: Estructuras cristalinas de complejos bromosporina-bromodominios evaluados.

En primer lugar se evaluó el cambio de estructura inicial al aplicar el campo de fuerza OFF-2.0.0 modificado, usando cargas atómicas derivadas

por familia. La Figura 4.20 muestra los resultados obtenidos para la comparación de los bromodominios seleccionados. Estos muestran resultados poco conclusivos, con una sobre estimación para todos los casos evaluados, donde en uno de los ocho bromodominios en estudios se observa una mejora con el cambio al complejo cristalizado (BPRF1B). Para los demás no se observan cambios significativos, ya que se encuentran dentro del error asociado a la predicción. Sin embargo, para reducir la incertidumbre dada por la modelación de la pose, se procedió a continuar con las estructuras cristalinas como punto de partida para posteriores evaluaciones.

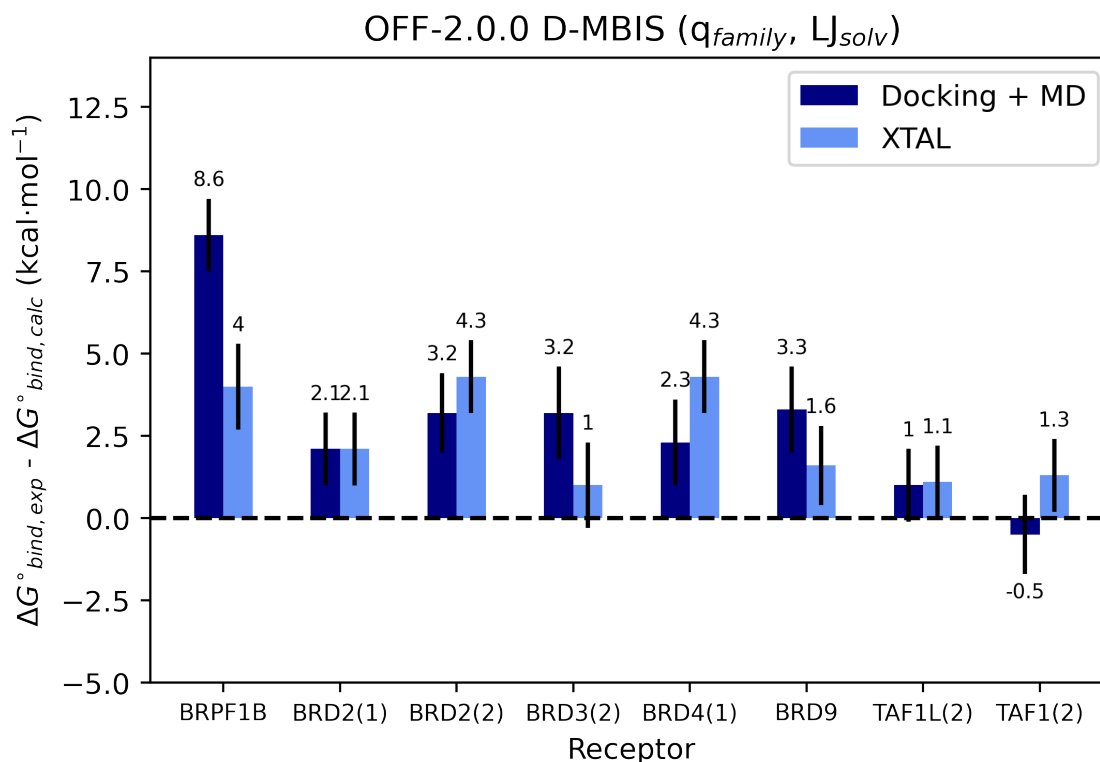


Figura 4.20: **Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 con parámetros no enlazantes D-MBIS (cargas atómicas por familia de bromodominios y parámetros de Lennard-Jones derivados desde solvente) usando distintas estructuras iniciales.**

Posteriormente, se evaluó si con solo el cambio de cargas atómicas en bromosporina derivadas específicamente en cada bromodominio, era suficiente para mejorar la estimación de la energía libre de unión absoluta. La Figura B.4 muestra la comparativa entre campos de fuerza modificados, manteniendo la estructura cristalina como referencia para los estados finales de la transformación alquímica. El cambio de cargas atómicas no fue suficiente para generar un impacto significativo en la mejora de la estimación, por lo que

existe un efecto no considerado dentro del cambio alquímico, el cual podría estar asociado al bromodominio y el proceso de exploración conformacional de este.

Con esto en consideración, se aplicó el protocolo de transformación alquímica de no equilibrio usando las estructuras holo del complejo bromosporina-bromodominio para la exploración conformacional del ligando acoplado y la estructura apo del bromodominio correspondiente para la simulación de equilibrio y extracción de configuraciones correspondientes al ligando desacoplado. La Figura 4.21 muestra el efecto de incorporar la estructura apo en el protocolo de estimación de energía libre de unión. No se observa una mejora sustancial en la predicción de afinidad en los ocho sistemas, manteniéndose la sobreestimación del valor calculado en todos los casos. Esto podría indicar la importancia de la incorporación de efectos conformacionales asociados a la unión de la proteína. Sin embargo, la exploración del espacio configuracional queda limitado al tiempo de simulación empleado por la metodología y la suposición de que las poblaciones de cada conformación sean representativamente muestreadas a tiempos finitos.

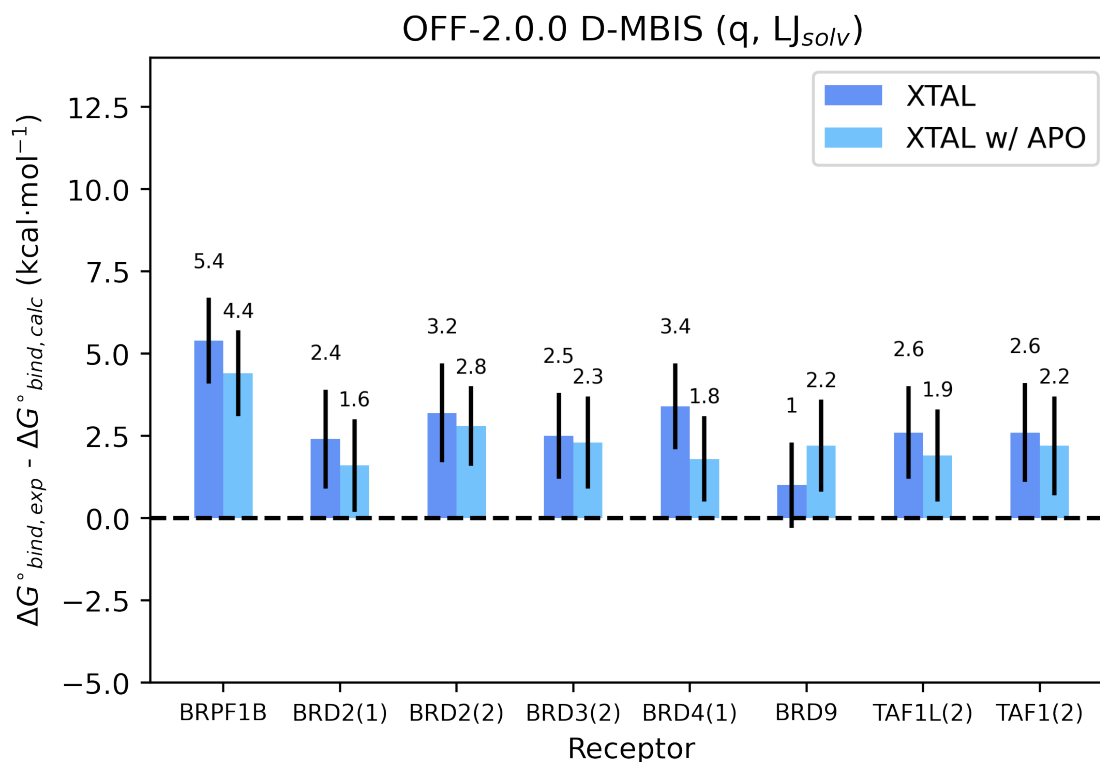


Figura 4.21: **Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS, usando distintos estados conformacionales de las proteínas en el protocolo de transformaciones de no equilibrio.**

Bajo la premisa de que existen limitaciones metodológicas dadas por el empleo de trayectorias de equilibrio cortas para la rama proteica del ciclo termodinámico, la Figura 4.22 muestra el desglose de la contribución asociada a ΔG_{prot} . Esta se compone de dos subcontribuciones, donde ΔG_{alch} asociada al proceso no físico de acoplar las interacciones intermoleculares con el receptor, mientras que ΔG_{conf} procede de la energía asociada al proceso de cambio (o selección) conformacional del receptor requerido para el reconocimiento

de ligando. Es esta última subcontribución que al ser subestimada, lleva finalmente a una sobreestimación global de la rama proteica, con directo efecto al valor predicho de energía libre de unión.

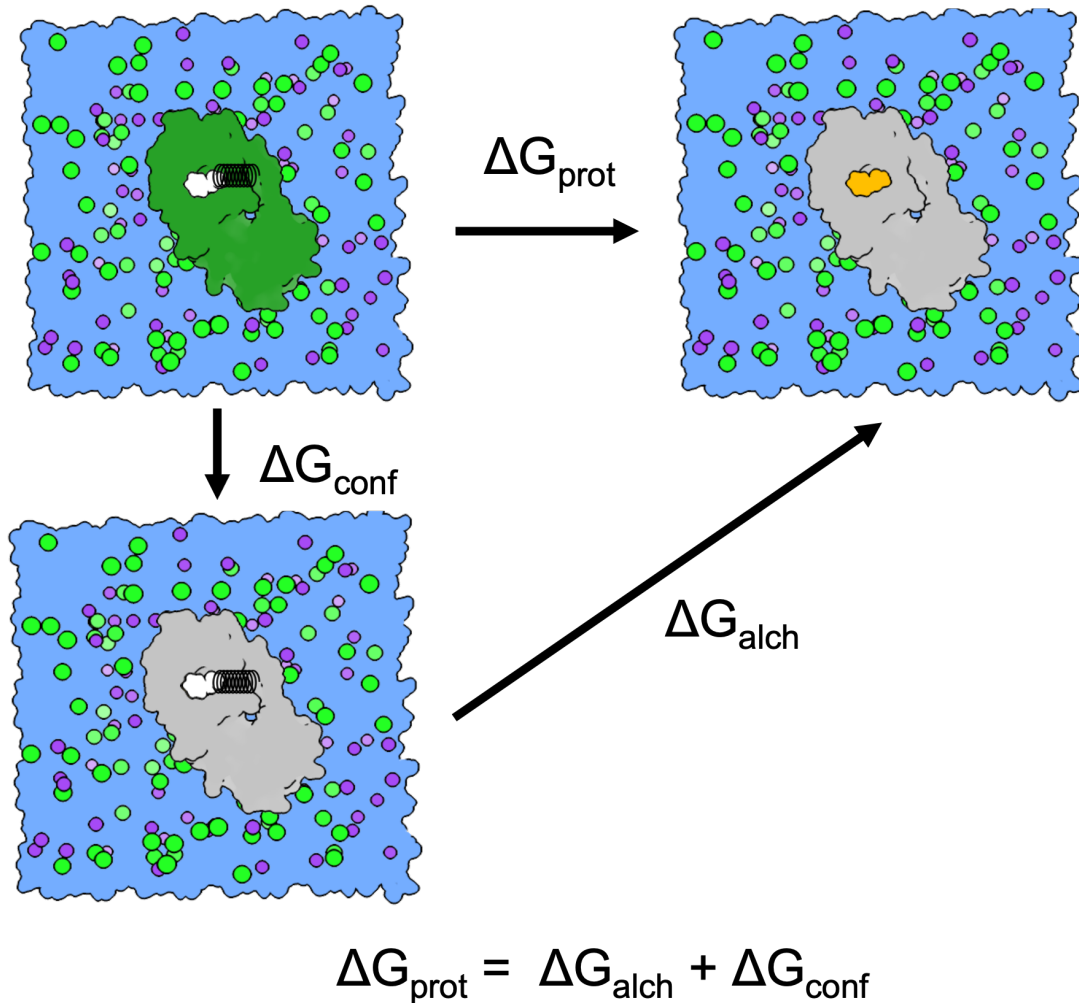


Figura 4.22: Desglose de la rama proteica del ciclo termodinámico utilizado para los cálculos de integración termodinámica de no equilibrio.

En el estudio de Bhati y colaboradores¹⁶⁶, se propone el cálculo de un factor de corrección energético asociado al efecto conformacional de la pro-

teína en la transformación alquímica. Incluido en este estudio se encuentra el conjunto de datos de bromosporina, el cual estima $\Delta G_{\text{conf}} = 2.31 \text{ kcal}\cdot\text{mol}^{-1}$ para las mismas familias de bromodominios. En la Figura 4.23 muestra los resultados de la aplicación de este factor de corrección a los resultados obtenidos usando las estructuras resueltas del complejo proteína-ligando y la estructura apo de los bromodominios. Para este subconjunto de datos, se obtuvo los mejores resultados para la comparación de valores respecto a datos experimentales, con diferencias menores a $1 \text{ kcal}\cdot\text{mol}^{-1}$ para 7 de 8 complejos considerando la estructura apo.

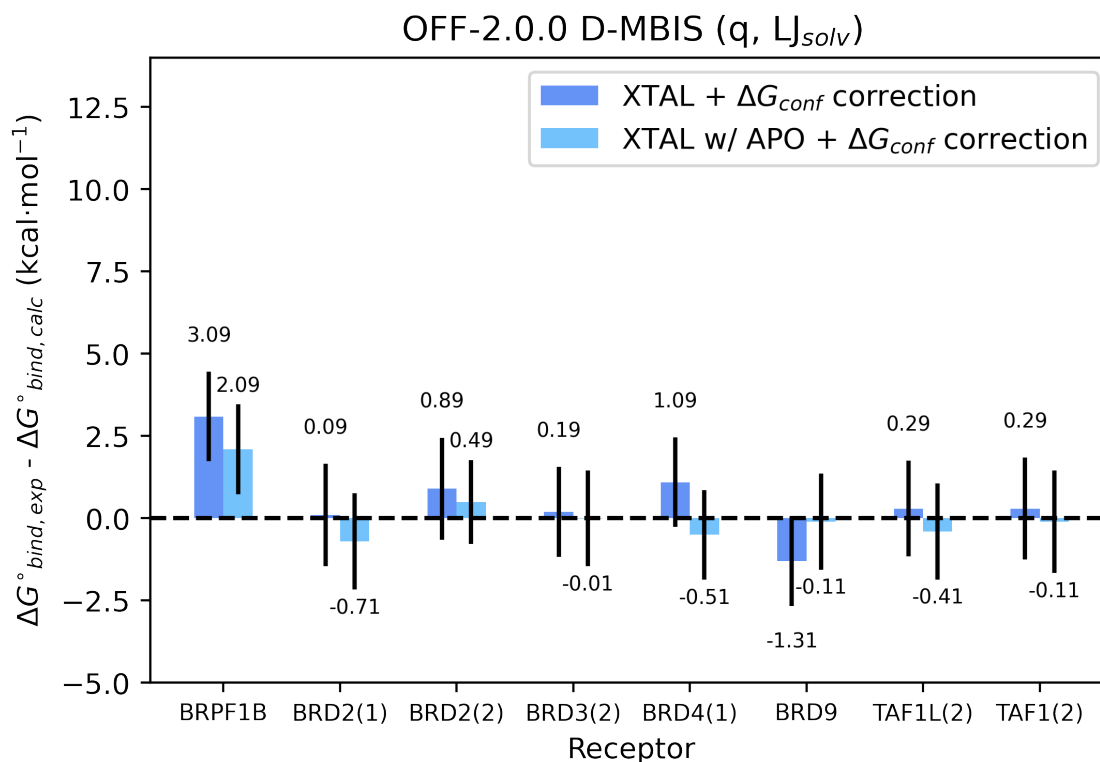


Figura 4.23: **Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 y parámetros no enlazantes D-MBIS, usando distintos estados conformacionales de las proteínas en el protocolo de transformaciones de no equilibrio, en conjunto del factor de corrección conformacional.**

Los resultados demuestran que las estimaciones de energía libre de unión absoluta para la generación de perfiles de selectividad requieren información estructural del complejo proteína-ligando de alta calidad. Esto involucra tanto la correcta definición de pose del ligando y su sitio de unión, como también una exploración eficiente de las conformaciones representativas de la proteína. De esta forma, es posible que la contribución dada por la rama proteica, represente de forma más fidedigna el promedio de microestados en el colec-

tivo evaluado y promueva una correcta estimación de esta variable macroscópica.

4.4. Especificidad en lisozima T4 L99A/M102Q

En base a los resultados anteriores, existe la necesidad preponderante de una correcta descripción del complejo proteína-ligando para una mejor estimación de la energía libre de unión absoluta, lo cual se sostiene bajo el uso de estructuras resueltas experimentalmente. La elección del sistema lisozima T4 L99A/M102Q se basó en la disponibilidad de estructuras cristalinas, las cuales describen correctamente la mayor complejidad del fenómeno de unión del ligando. Esto es producto de un sutil cambio conformacional en el estado holo, donde el movimiento de la hélice F (Figura 1.3), que permite el acceso de ligandos más voluminosos al sitio de unión. Esta variante presenta la misma cavidad hidrófoba pequeña ($\sim 150 \text{ \AA}^3$) descrita para la mutante L99A, pero incorpora una polaridad significativa y el potencial de formar un enlace de hidrógeno, introducido por la sustitución M102Q⁶³. Al igual que en la cavidad apolar de L99A, este sitio une moléculas pequeñas aromáticas⁵⁷, aunque presenta cierta selectividad hacia ligandos polares. Esta selectividad se ve modulada por un equilibrio entre la solvatación del ligando y la necesidad de una orientación precisa para establecer un enlace de hidrógeno con Q102 (Figura 1.3). Este cambio introduce restricciones adicionales, ausentes en la cavidad apolar, sin comprometer la simplicidad general del sitio de unión⁶³.

Al igual que el sistema original, el modelo L99A/M102Q es apto para estudios prospectivos, con ensayos experimentales y condiciones de cristalización bien establecidas, lo que lo convierte en un conjunto de pruebas ideal para evaluar nuevas metodologías^{64–66}. Para este sistema se consideraron ocho ligandos que se unen a la cavidad polar de la mutante, descritos por Boyce y colaboradores⁶³ en conjunto con Wei y colaboradores⁶⁵ (Figura 1.3).

Las estructuras cristalográficas resueltas de los estados apo (PDB ID: 1lgu) y holo (PDB IDs: 1li2, 1xep, 3hu8, 3huk, 3hua, 2rbn y 1li3) fueron utilizadas como coordenadas iniciales. Para la molécula de tolueno (L1), se empleó la estructura 1li2, modificando el grupo hidroxilo por un metilo mediante PyMOL. A cada ligando del conjunto se le añadieron hidrógenos utilizando la función addH de PyMOL. El estado de protonación del receptor se predijo para pH 6.8 y fuerzas iónicas correspondientes (~ 0.5 M de NaCl para L1-L3 y ~ 0.22 M de KCl para L4-L8), mediante el servidor H++ (<http://newbiophysics.cs.vt.edu/H++/>). Los complejos ligando-proteína se generaron a partir de la conformación apo utilizando el comando align de PyMOL para superponer las estructuras con respecto a los complejos en estado holo.

Los archivos requeridos para GROMACS (.gro y .top) se generaron con pdb2gmx, considerando la doble protonación de H31 (HISH) predicha por el servidor. La construcción de los sistemas solvatados y los archivos necesarios

para el protocolo de simulación se realizó mediante el *script* `setup_abfe.py` de `pmx`¹⁵⁶. Los valores experimentales de energía libre de unión fueron extraídos del repositorio: https://github.com/MobleyLab/benchmarksets/tree/andrea-t4lysozyme/input_files/t4-199a-m102q, según la revisión de Mobley y colaboradores¹⁶⁷.

Todas las simulaciones se realizaron con GROMACS 2021.2¹⁴⁵, utilizando los mismos campos de fuerza (proteína e iones) y modelo de agua empleados para BRD4(1) y bromosporina. Inicialmente, los parámetros de los ligandos se generaron con GAFF-1.8²⁶ y cargas AM1-BCC^{150,151}, debido a la existencia de datos previos en el laboratorio. Posteriormente, los parámetros no enlazantes fueron derivados con `ffparaaim` para su evaluación comparativa.

La minimización energética y equilibrio NVT fueron realizados de la misma forma que los sistemas anteriormente evaluados. Las simulaciones de producción se realizaron en el colectivo NPT (283.15 K, 1.01325 bar), utilizando dinámica de Langevin¹⁵² y el barostato de Parrinello-Rahman¹²¹. Se empleó un integrador estocástico con una constante de fricción de 1 ps^{-1} y un paso de integración de 2 fs. Los enlaces con hidrógenos se restringieron mediante LINCS¹⁵³ (orden de restricción 6). La electrostática de largo alcance se trató mediante PME^{67,68}, con tolerancia relativa de 10^{-6} , grilla de 0.1 nm y corte a 1.0 nm. Las interacciones de van der Waals se modelaron mediante un potencial de desplazamiento con corte a 1.2 nm. Ambas interacciones se sua-

vizaron en los extremos de los estados alquímicos con el esquema de Beutler y colaboradores⁸⁶.

Se aplicó el ciclo termodinámico propuesto por Aldeghi et al.¹² para estimar la energía libre de unión absoluta. Se generaron tres réplicas independientes para la solvatación y el complejo proteína-ligando en GAFF-1.8, y cinco réplicas para el campo de fuerza reparametrizado (D-MBIS). Las simulaciones de producción en agua y complejo se realizaron por 15 ns y 45 ns, respectivamente. Las transformaciones de no equilibrio incluyeron 100 transiciones (0.5 ns) para ΔG_{solv} y 400 para ΔG_{prot} , usando configuraciones extraídas cada 100 ps desde la simulación de producción, descartando los primeros 5 ns de esta última como equilibrio. El acoplamiento de los ligandos se inició desde la estructura apo y el desacoplamiento desde la estructura holo. La estimación de energía libre se realizó con BAR, usando *bootstrapping* con 100 muestras en pmx.

Se consideró el valor reportado como el promedio de las réplicas y el error estándar respectivo. Se usaron los mismos descriptores estadísticos (MUE, RMSE, r de Pearson, τ de Kendall) para la evaluación del error y correlación.

Los resultados obtenidos se muestran en las Figuras 4.24 y 4.25. En el caso del campo de fuerza modificado con parámetros no enlazantes D-MBIS, se incluyó la corrección por polarización (Tabla C.1). No se observaron diferencias relevantes en los errores absolutos (RMSE y MUE), ambos por debajo de $2 \text{ kcal}\cdot\text{mol}^{-1}$. Sin embargo, las correlaciones lineales ($r \sim 0$) y ordinales

fueron nulas.

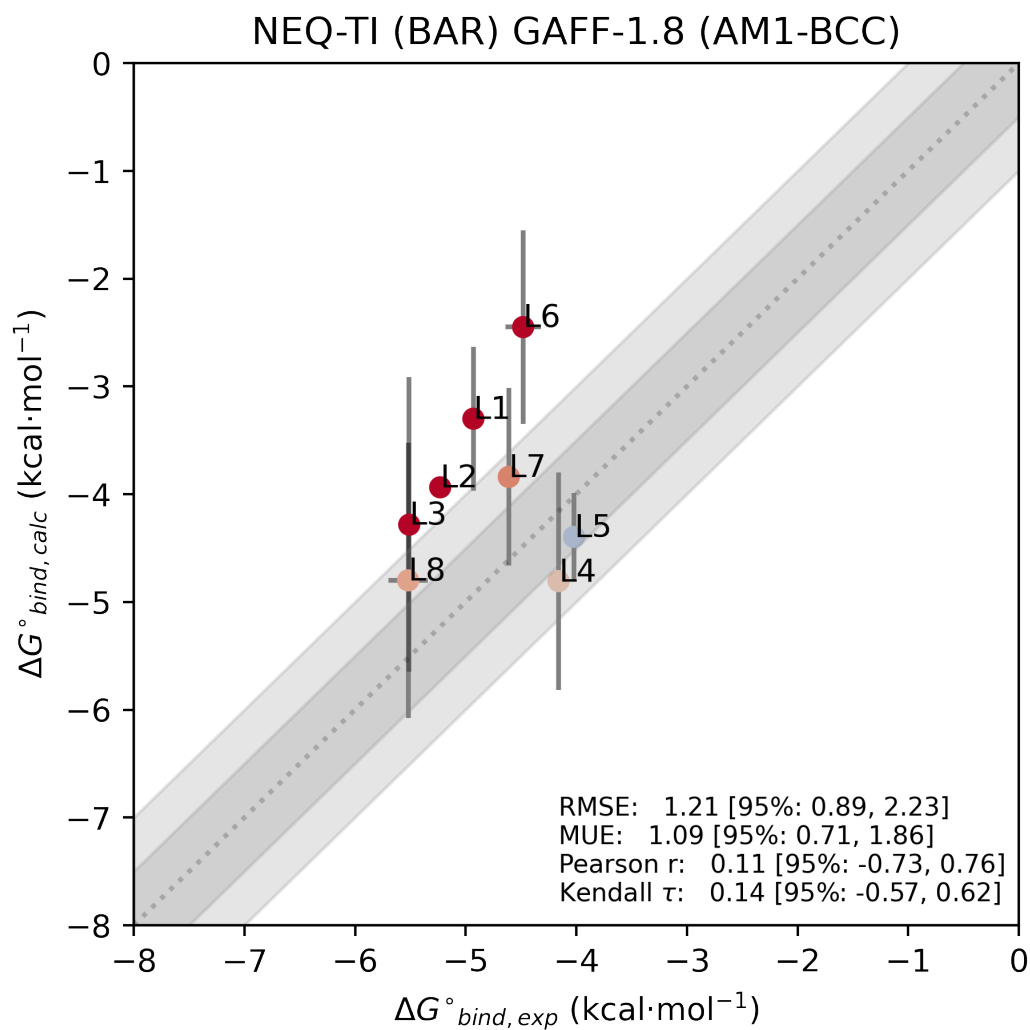


Figura 4.24: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y cargas atómicas AM1-BCC.

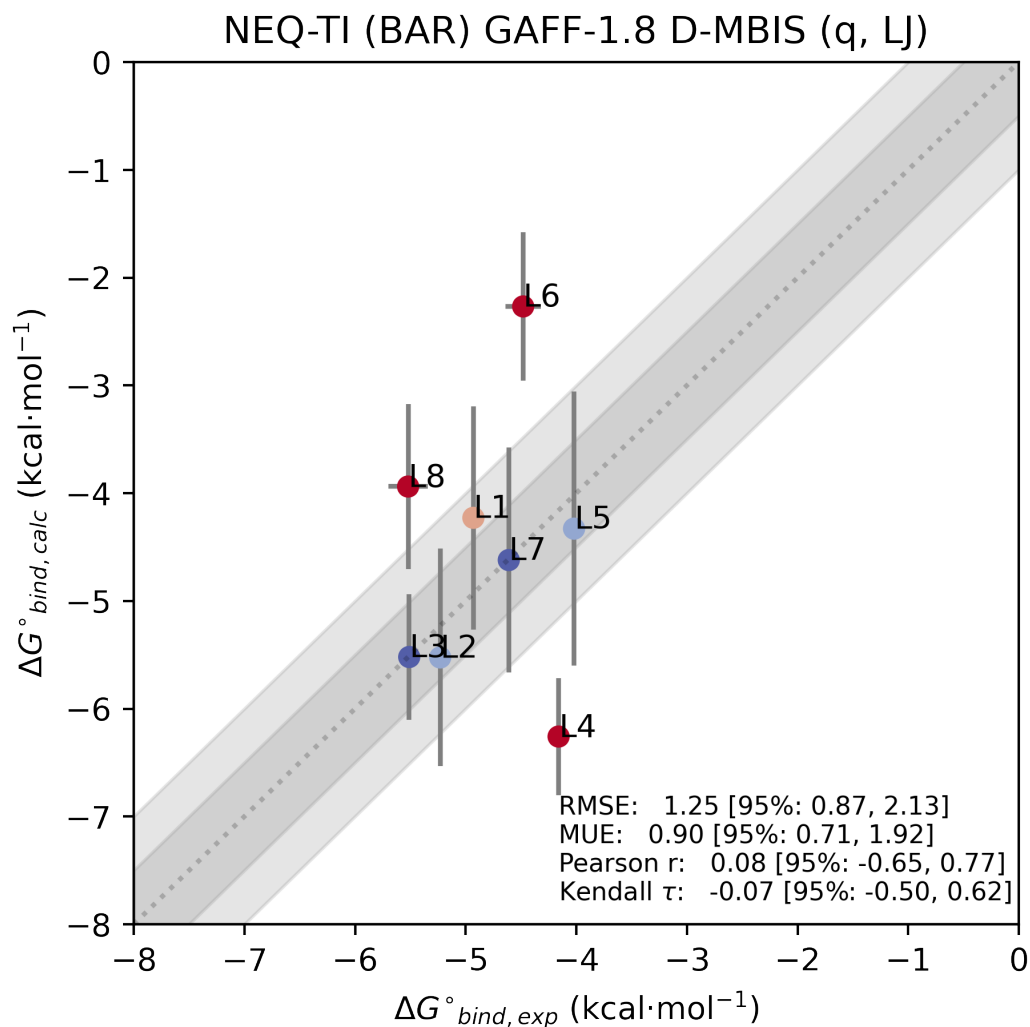


Figura 4.25: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y reemplazo de parámetros no enlazantes derivados (D-MBIS).

Se identificaron errores sistemáticos: subestimación de la energía de L6 y sobreestimación de L4 en el campo de fuerza modificado. Una posible explicación a la desviación observada es la existencia de múltiples modos de

unión, propuestos desde las estructuras cristalográficas (Figuras 4.26 y 4.27). En el tiempo de equilibrio de 45 ns utilizado para este sistema, se observó interconversión entre modos (Figura 4.26), lo que afecta la convergencia del valor estimado de afinidad. Para abordar esto, se aplicaron restricciones de posición a los átomos pesados de L4 y L6, favoreciendo localizaciones independientes. Las configuraciones se extrajeron bajo un Hamiltoniano modificado, pero la contribución energética se cancela por simetría en los ciclos. Finalmente, el valor de energía libre de unión reportado viene desde la contribución efectiva de ambos modos de unión, mediante el uso de la descomposición del espacio configuracional⁷⁸.

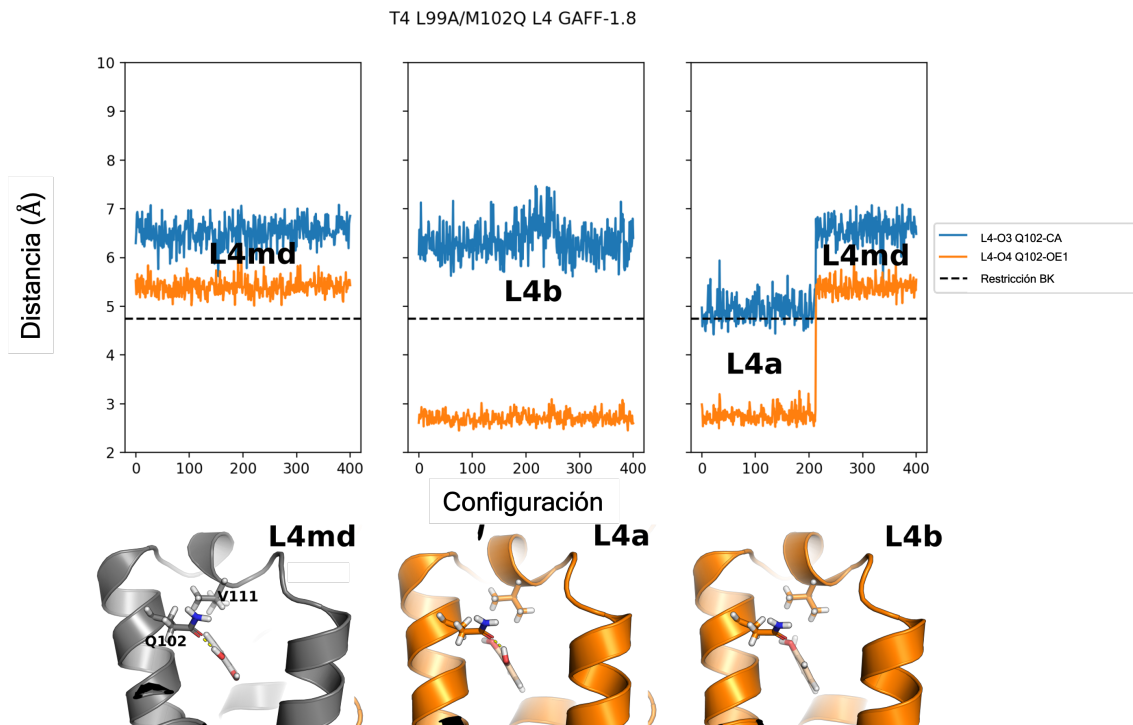


Figura 4.26: **Evaluación de métricas de distancia para los distintos modos de unión de L4.** Se muestran 3 réplicas para la simulación de producción usando el campo de fuerza GAFF-1.8. Usando dos descriptores de distancia, es posible evidenciar 3 modos de unión, la estructura representativa del *cluster* de mayor población de las trayectorias agregadas (L4md) y los reportados reportados estructuralmente (L4a y l4b). Se incluye la distancia utilizada para la restricción Boresech-Karplus, correspondiente a L4a.

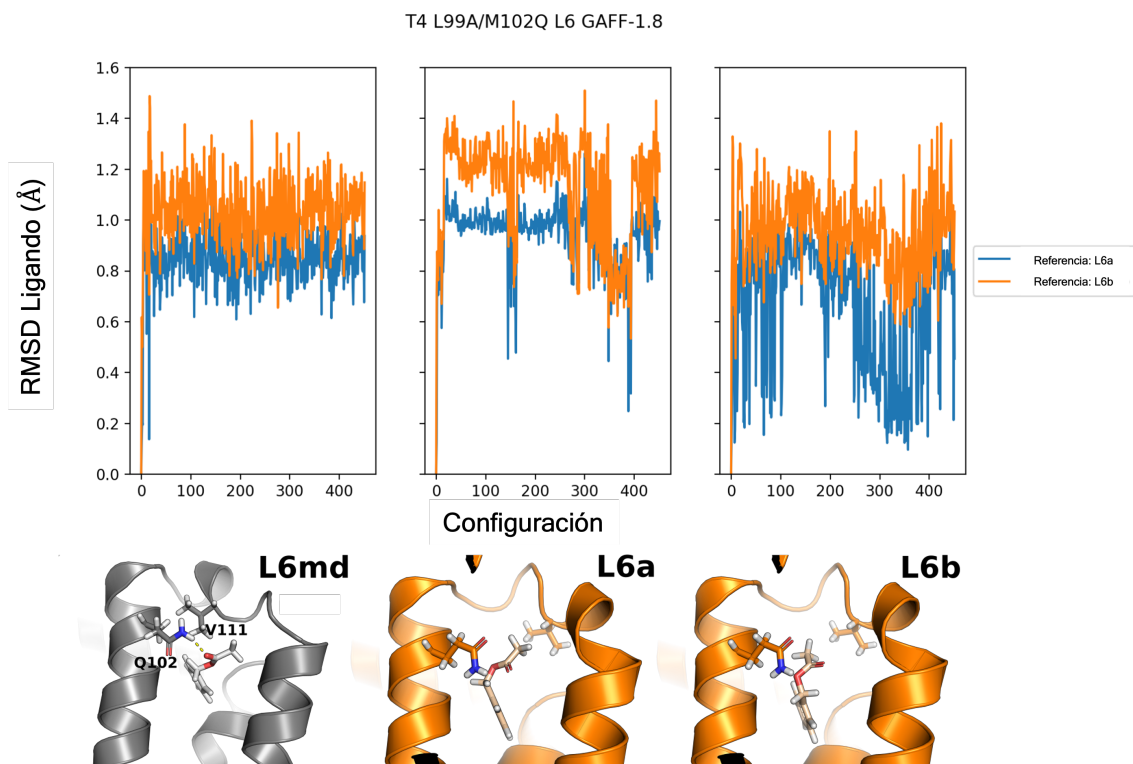


Figura 4.27: **Evaluación de métricas estructurales para los distintos modos de unión de L6.** Se muestran 3 réplicas para la simulación de producción usando el campo de fuerza GAFF-1.8. Se usó la raíz de la desviación cuadrática media (RMSD) del ligando tras remover grados de libertad roto-traslacionales mediante superposición con estructuras cristalográficas de referencia (L6a, L6b) como descriptor estructural. Adicionalmente, se incluye la estructura de mayor representatividad de las trayectorias agregadas, extraída desde el cluster de mayor población.

En la Figura 4.28 se muestran los valores previamente obtenidos y el reemplazo de los valores tras las corrección mediante la aplicación de la restricción para el campo de fuerza GAFF-1.8. En cuanto al error de predicción, se observan mejoras relevantes respecto a no aplicar las restricciones de posición para los ligandos L4 y L6 (valor de RMSE disminuye de 1.21 a 1.02 kcal·mol⁻¹ y MUE de 1.09 a 0.90 kcal·mol⁻¹), pero manteniéndose muy cer-

cano a la exactitud química ($1 \text{ kcal}\cdot\text{mol}^{-1}$). La correlación lineal y ordinal mejora sutilmente, pero no permite una correcta discriminación entre moléculas.

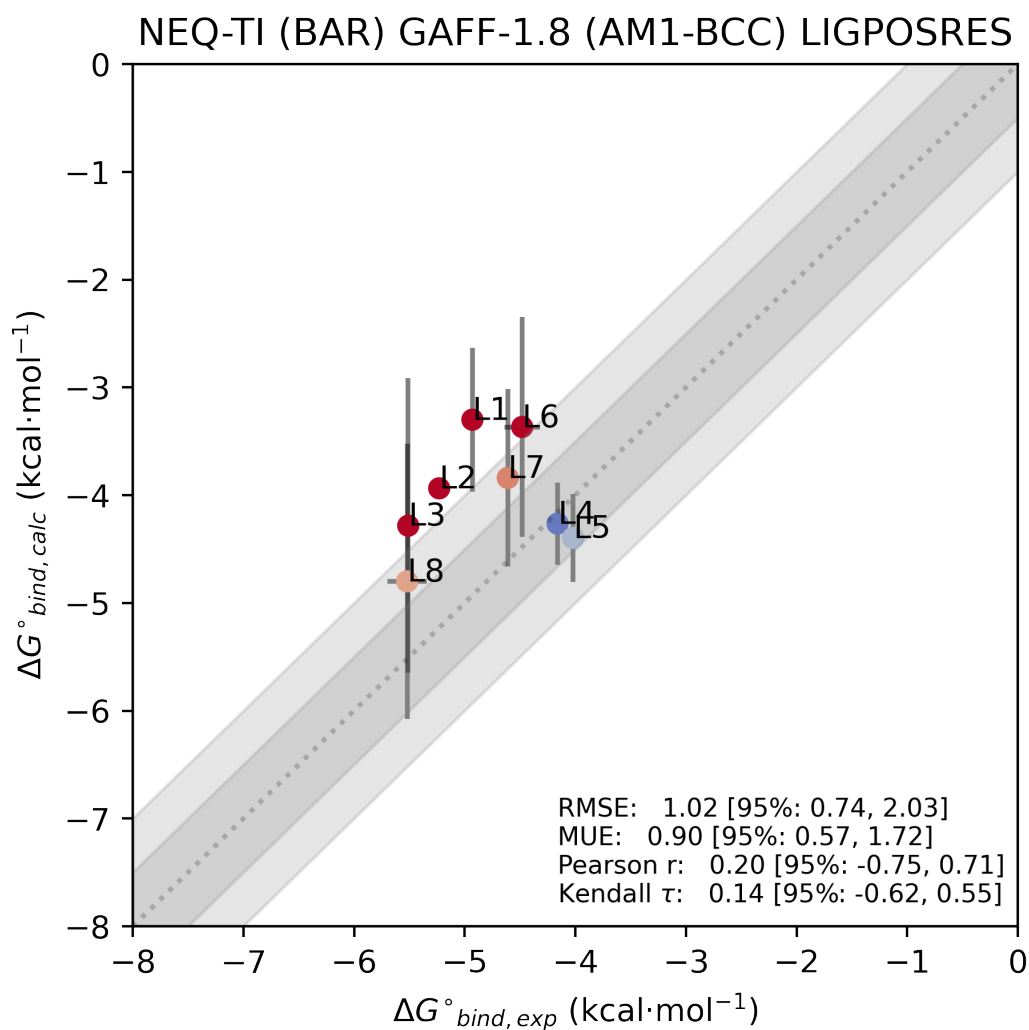


Figura 4.28: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 con restricciones de posición del ligando.

Resultados similares se observaron en el campo de fuerza modificado

(Figura 4.29): RMSE de 1.25 a 0.69 kcal·mol⁻¹ y MUE de 0.90 a 0.49 kcal·mol⁻¹, dentro de la exactitud química.

Este resultado apoya la premisa de que la aplicación de parámetros a la medida bajo contextos moleculares explícitos, mejora la descripción del sistema a nivel de interacciones intermoleculares. Al igual que el caso del campo de fuerza sin modificación, la correlación lineal ($r = 0.41$) y ordinal ($\tau = 0.14$) presenta una mejora, pero limitada capacidad discriminatoria.

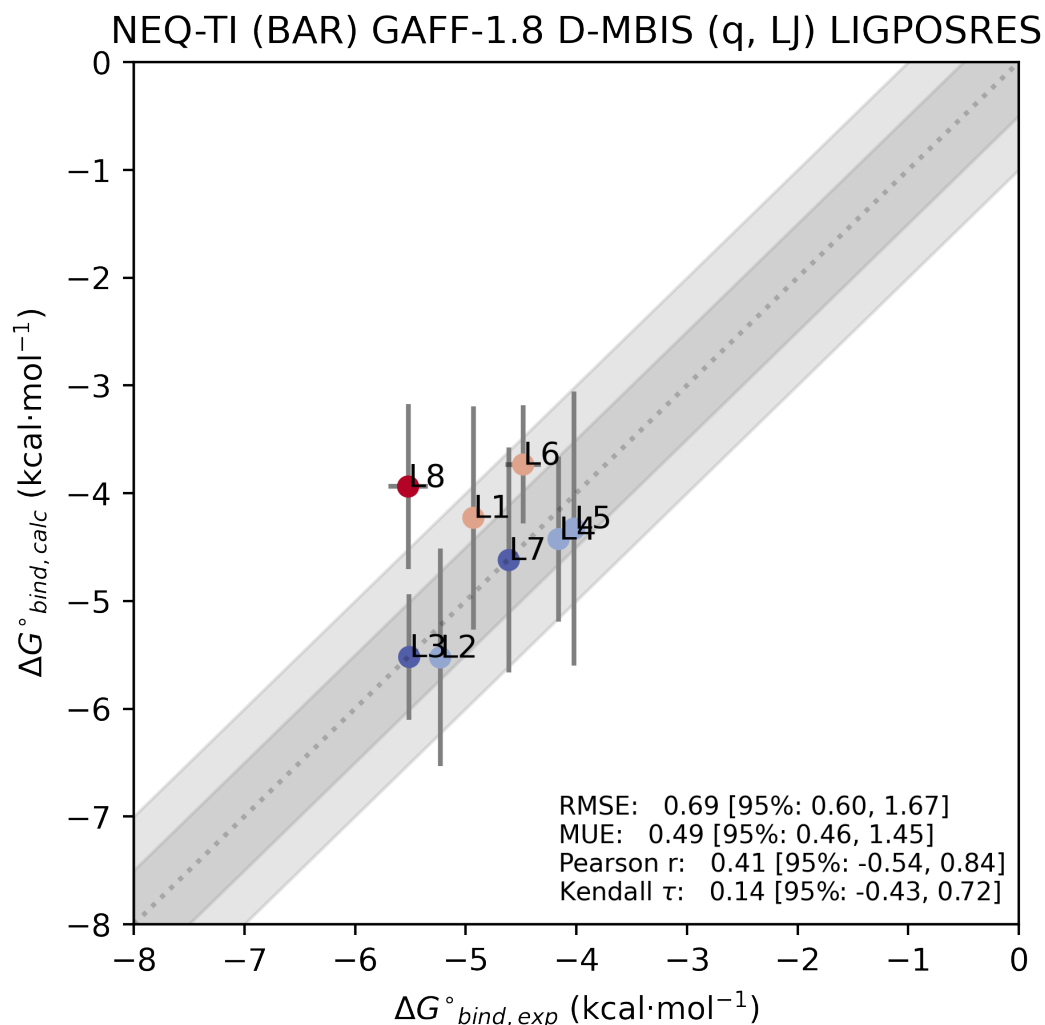


Figura 4.29: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y parámetros no enlazantes D-MBIS, aplicando restricciones de posición de ligando.

Finalmente, se evaluó el ligando L8, donde se observó una orientación del grupo amino predominante en la trayectoria de simulación. Esta presenta una dirección opuesta a la referencia cristalográfica, impidiendo la formación

del enlace de hidrógeno con Q102 (Figura 4.30). Se seleccionaron configuraciones con dicha interacción usando la función hbond de GROMACS y posteriormente se aplicó la estimación mediante BAR, observándose mejoras sustanciales en los descriptores de error (RMSE = 0.48 kcal·mol⁻¹, MUE = 0.38 kcal·mol⁻¹) y correlación ($r = 0.68$, $\tau = 0.43$).

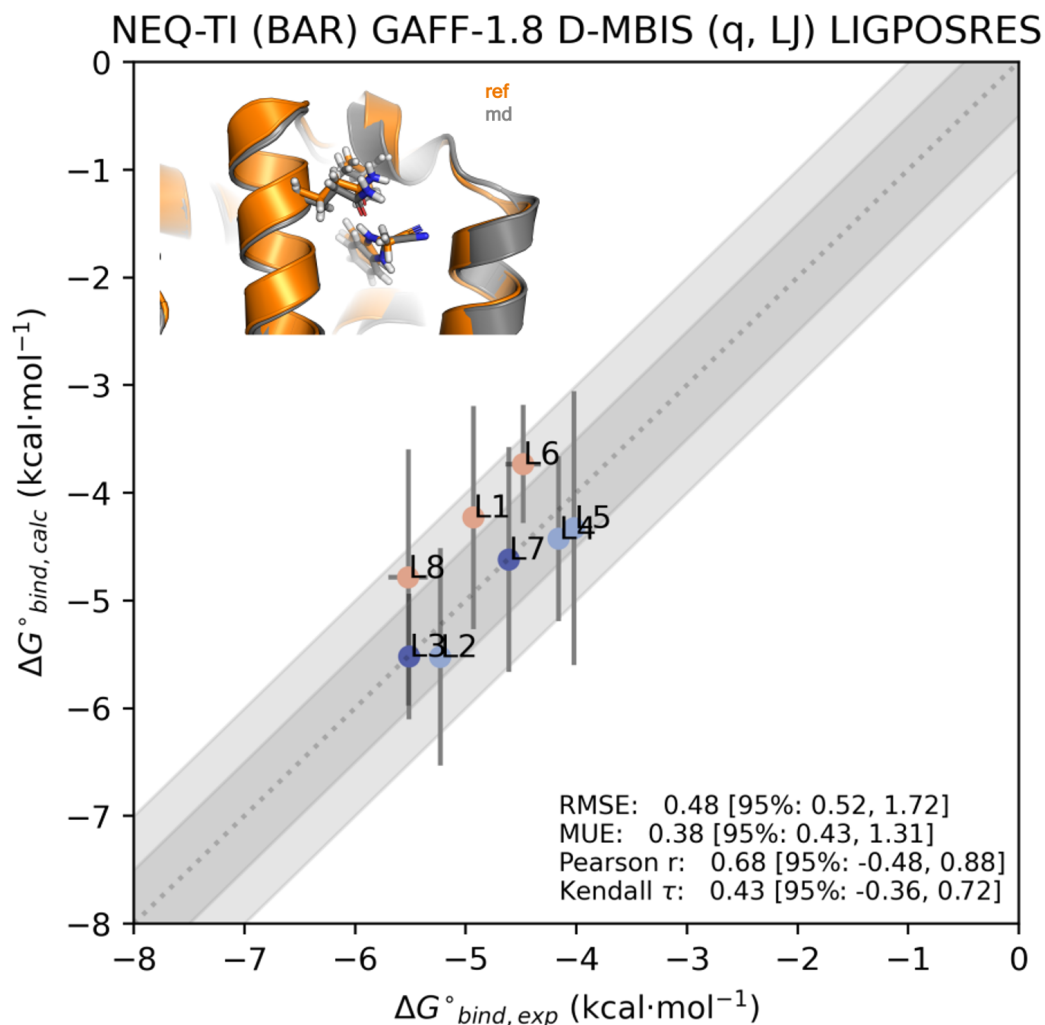


Figura 4.30: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.8 y parámetros no enlazantes D-MBIS, usando restricciones de posición del ligando y selección de configuraciones de L8 que presentan interacciones de puente de hidrógeno con Q102. En la imagen interior, se muestra una superposición de la estructura representativa del *cluster* mayormente poblado de las trayectorias agregadas (md) y la estructura cristalográfica (ref).

Considerando las notorias mejoras que otorgan las aproximaciones utili-

zadas, el conjunto presenta un desafío particular debido a su estrecho rango de energías libres experimentales ($\sim 2 \text{ kcal}\cdot\text{mol}^{-1}$)⁵⁹, lo que involucra un mayor desafío respecto a la capacidad predictiva y discriminatoria de los modelos.

La Tabla 4.3 muestra los valores experimentales del conjunto de datos lisozima T4 L99A/M102Q, los valores calculados para GAFF-1.8 original y el campo de fuerza modificado con parámetros D-MBIS corregido por polarización. De acuerdo a los resultados, el campo de fuerza modificado se presenta como la mejor opción de modelo predictivo de energía libre de unión absoluta en este sistema.

Tabla 4.3: Valores de energía libre de unión experimental y calculados para el campo de fuerza GAFF-1.8 original y con parámetros no enlazantes modificados mediante la metodología D-MBIS. Los valores se expresan en kcal·mol⁻¹

Sistema	$\Delta G_{\text{bind, exp}}$	$\Delta G_{\text{bind, calc}}^{\text{GAFF-1.8}}$	$\Delta G_{\text{bind, calc}}^{\text{GAFF-1.8 D-MBIS (q, LJ)}}$
L1	-4.9	-3.3 ± 0.7	-4.2 ± 1.0
L2	-5.23	-3.9 ± 0.2	-5.8 ± 1.0
L3	-5.51	-4.3 ± 1.4	-5.4 ± 0.6
L4 ¹	-4.16 ± 0.03	-4.3 ± 0.4	-4.4 ± 0.8
L5	-4.02 ± 0.03	-4.4 ± 0.4	-4.3 ± 1.3
L6 ¹	-4.5 ± 0.2	-3.4 ± 1.0	-4.0 ± 0.4
L7	-4.6 ± 0.1	-3.8 ± 0.8	-4.6 ± 1.0
L8	-5.5 ± 0.2	-4.8 ± 1.3	-4.8 ± 1.2 ²

¹ Valores obtenidos aplicando restricciones de posición al ligando.

² Valor obtenido tras selección de configuraciones que presentan enlace de hidrógeno con Q102.

5. Conclusiones

El objetivo principal de esta tesis fue evaluar la capacidad predictiva de la energía libre de unión absoluta mediante la modificación de parámetros no enlazantes en campos de fuerza no polarizables, utilizando conjuntos de datos de referencia ampliamente reportados en la literatura. Dichos parámetros fueron obtenidos a partir de la partición de la densidad electrónica polarizada, calculada mediante métodos de química cuántica, y posteriormente incorporados en los campos de fuerza empleados en simulaciones de dinámica molecular.

Como primer requisito para el desarrollo de este trabajo, se implementó una herramienta computacional de código abierto, denominada `ffparaim`, orientada a la derivación automatizada de parámetros no enlazantes en moléculas pequeñas. Esta implementación mejora la representación de las interacciones electrostáticas y de dispersión/repulsión en simulaciones de dinámica molecular, a través de una asignación de cargas atómicas y parámetros de Lennard-Jones que consideran el entorno molecular, basada en la inclusión explícita de la polarización de la densidad electrónica.

La herramienta se fundamenta en el protocolo D-MBIS, el cual combina cálculos QM/MM sobre configuraciones extraídas de trayectorias de dinámi-

ca molecular clásica con particiones de la densidad electrónica que permiten asignar propiedades atómicas en el contexto molecular. La incorporación explícita del entorno durante la derivación representa una mejora sustancial frente a aproximaciones tradicionales, en las que estos parámetros se estiman en vacío o mediante métodos semiempíricos.

Desde el punto de vista técnico, `ffparaim` fue concebida como un paquete de Python modular, flexible y totalmente compatible con el ecosistema Open Force Field. Permite el tratamiento de múltiples configuraciones del entorno químico y su aplicación tanto en disolución como en complejos proteína-ligando, mediante la definición opcional de restricciones geométricas. La herramienta automatiza todo el flujo de trabajo, desde el preprocesamiento de estructuras hasta la generación de archivos de parámetros listos para simulación.

Las pruebas de validación incluyeron sistemas en disolución acuosa y complejos proteína-ligando, en los cuales se obtuvieron parámetros consistentes con el entorno químico modelado. Además, se consideró un término de corrección de polarización E_{pol} en el ciclo termodinámico para la estimación de afinidades que cuantifica el costo energético asociado a la adaptación de la densidad electrónica. Se realizó también un análisis sistemático del impacto de las variables del protocolo sobre la convergencia de E_{pol} y los tiempos de ejecución, demostrando la viabilidad computacional del enfoque.

Con esta herramienta se evaluó la efectividad de los nuevos parámetros no

enlazantes en distintos campos de fuerza, aplicados a los conjuntos de datos BRD4(1), bromosporina y lisozima T4 L99A/M102Q para la estimación de energías libres de unión absoluta. En cada conjunto, bajo condiciones específicas, las modificaciones derivadas del método D-MBIS mostraron la mejor concordancia con los datos experimentales, validando su uso en la predicción de afinidad en contextos de especificidad y selectividad, y confirmando la hipótesis planteada en esta tesis.

En el caso del conjunto BRD4(1), el campo de fuerza OFF-2.0.0 con parámetros no enlazantes D-MBIS, junto con la corrección energética por polarización, proporcionó una excelente estimación de energías libres de unión absoluta. Se obtuvieron errores menores a $1 \text{ kcal}\cdot\text{mol}^{-1}$ y correlaciones competitivas en modelos de predicción de ligandos neutros. Para ligandos cargados, sin embargo, se identificaron deficiencias en la predicción tanto mediante la corrección analítica como con el método DSSB, proponiéndose posibles causas a ser abordadas en estudios posteriores.

En el análisis del perfil de selectividad de bromosporina con el conjunto de datos generado por Aldeghi y colaboradores, se observó una deficiencia predictiva generalizada en todos los campos de fuerza evaluados, incluyendo las aproximaciones con parámetros no enlazantes D-MBIS, con una tendencia a sobre-estimar las afinidades sobre los distintos receptores. La sustitución de estructuras iniciales obtenidas desde *docking* y dinámica molecular por estructuras experimentales permitió mitigar el posible sesgo asociado a la pose

del ligando, pero el fenómeno de sobre-estimación de la afinidad no fue corregido para el total del subconjunto de bromodominios evaluado. La inclusión de un factor de corrección vinculado al proceso de cambio conformacional del receptor en la unión entregó los resultados más consistentes con los datos experimentales, revelando una contribución a la afinidad no considerada mediante el protocolo de no equilibrio utilizado.

Dada la relevancia de contar con estructuras iniciales correctas y de considerar el efecto conformacional en el reconocimiento molecular y la formación del complejo proteína-ligando, el conjunto de referencia lisozima T4 L99A/M102Q constituyó un sistema ideal para evaluar de manera controlada el impacto de la modificación de parámetros no enlazantes en un campo de fuerza no polarizable. Su elección se justificó por tratarse de un sistema ampliamente estudiado, con datos experimentales curados, incluyendo estructuras cristalinas de alta calidad y poses alternativas de ligandos. No obstante, la complejidad intrínseca de este conjunto radica en que ligandos de bajo peso molecular son propensos a modos de unión alternativos y que el rango de afinidades experimentales es relativamente estrecho ($\sim 1.5 \text{ kcal}\cdot\text{mol}^{-1}$). Bajo esta premisa, la aplicación de restricciones de posición en algunos ligandos y considerar los distintos modos unión reportados en las estructuras cristalinas permitió obtener la mejor concordancia con los datos experimentales usando los nuevos parámetros no-enlazantes D-MBIS, con errores dentro del rango de la exactitud química.

En conjunto, este trabajo constituye una contribución significativa a la parametrización de moléculas en simulaciones de dinámica molecular con campos de fuerza no polarizables, y establece bases sólidas para futuros desarrollos en campos de fuerza adaptativos y metodologías de diseño de fármacos con un enfoque más fundamentado en principios físicos.

5.1. Proyecciones

El desarrollo presentado en esta tesis abre diversas líneas de trabajo que permitirán ampliar y consolidar la aplicabilidad del enfoque propuesto en el contexto de simulaciones de dinámica molecular para la predicción de energías libres de unión. Entre las principales proyecciones destacan:

- **Mejoras en la descripción de la densidad electrónica basada en campos de fuerza no polarizables.** La incorporación de sitios virtuales permitiría representar de manera más representativa la anisotropía en la distribución electrónica, particularmente en la descripción de pares de electrones no enlazantes⁵³. Este avance facilitaría una caracterización más precisa de las interacciones direccionales, como enlaces de hidrógeno, enlaces halógeno o interacciones mediadas por dipolos.
- **Optimización conjunta de parámetros enlazantes y no enlazantes.** Tras la modificación de los parámetros de Lennard-Jones y cargas atómicas mediante D-MBIS, resulta necesario abordar la optimización de las barreras de energía potencial asociadas a la rotación de ángulos diedros, con el fin de mantener la coherencia interna del campo de fuerza y evitar conformaciones no físicas generadas por sobre o sub-estimación en las energías de átomos separados por tres enlaces covalentes.
- **Eficiencia computacional.** La paralelización de cálculos QM/MM a

partir de configuraciones muestreadas en simulaciones de dinámica molecular, como modificación a la herramienta *ffparaim*, permitiría reducir significativamente los tiempos de derivación de parámetros, facilitando la aplicación rutinaria de esta metodología en estudios de gran escala.

- **Exploración de otros esquemas de partición electrónica.** Métodos alternativos, como las cargas atómicas derivadas mediante *Constrained Iterative Hirshfeld*¹⁶⁸, podrían evaluarse como estrategia complementaria para evaluación de conjuntos de datos de series congenéricas, manteniendo fijos los parámetros del andamiaje molecular y sólo modificar la sustitución de grupos funcionales. De esta manera, la adición de nuevas metodologías de partición⁷⁶ podrían mejorar la robustez y reproducibilidad de los parámetros derivados.
- **Evaluación de ligandos cargados.** La extensión del protocolo a ligandos con carga neta requiere incorporar de manera explícita las interacciones intramoleculares en el contexto de transformaciones alquímicas, con el fin de superar las limitaciones observadas en los estudios realizados con el conjunto BRD4(1).
- **Consideración de efectos conformacionales de la proteína.** En el caso del conjunto de bromosporina, las limitaciones metodológicas actuales impidieron capturar en su totalidad la contribución conformacional del

receptor a la energía libre de unión. Se propone el uso de metodologías alternativas, como simulaciones de muestreo mejorado¹⁶⁹ o protocolos híbridos basados en aprendizaje profundo^{170,171}, para incorporar adecuadamente este efecto.

En conjunto, estas proyecciones apuntan a consolidar la utilidad de metodologías basadas en la partición de densidades electrónicas y en la integración de efectos de polarización del ambiente molecular, avanzando hacia campos de fuerza más adaptativos y predicciones de afinidad con mayor exactitud y transferibilidad.

Bibliografía

- [1] Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Current Opinion in Structural Biology* **2011**, *21*, 150–160.
- [2] Caballero, J. *Expert Opinion on Drug Discovery* **2021**, *16*, 625–645.
- [3] Li, J.; Fu, A.; Zhang, L. *Interdisciplinary Sciences: Computational Life Sciences* **2019**, *11*, 320–328.
- [4] Pinzi, L.; Rastelli, G. *International Journal of Molecular Sciences* **2019**, *20*, 4331.
- [5] King, E.; Aitchison, E.; Li, H.; Luo, R. *Frontiers in Molecular Biosciences* **2021**, *8*, 712085.
- [6] Gapsys, V.; De Groot, B. L. *eLife* **2020**, *9*, e57589.
- [7] Cournia, Z.; Allen, B.; Sherman, W. *Journal of Chemical Information and Modeling* **2017**, *57*, 2911–2937.
- [8] Shirts, M. R.; Mobley, D. L.; Chodera, J. D. *Annual Reports in Computational Chemistry*; Elsevier, 2007; Vol. 3; pp 41–59.
- [9] Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. *The Journal of Physical Chemistry B* **2003**, *107*, 9535–9551.
- [10] Boresch, S. *Journal of Chemical Information and Modeling* **2024**, *64*, 3605–3609.
- [11] Deng, Y.; Roux, B. *The Journal of Physical Chemistry B* **2009**, *113*, 2234–2246.
- [12] Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. *Chemical Science* **2016**, *7*, 207–218.

- [13] Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. *Journal of the American Chemical Society* **2017**, *139*, 946–957.
- [14] Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. *Communications Chemistry* **2021**, *4*, 61.
- [15] Macchiagodena, M.; Pagliai, M.; Karrenbrock, M.; Guarnieri, G.; Iannone, F.; Procacci, P. *Journal of Chemical Theory and Computation* **2020**, *16*, 7160–7172.
- [16] Procacci, P.; Macchiagodena, M.; Pagliai, M.; Guarnieri, G.; Iannone, F. *Chemical Communications* **2020**, *56*, 8854–8856.
- [17] Procacci, P.; Guarnieri, G. *Journal of Computer-Aided Molecular Design* **2021**, *35*, 37–47.
- [18] Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. *Chemical Science* **2020**, *11*, 1140–1152.
- [19] Baumann, H. M.; Gapsys, V.; de Groot, B. L.; Mobley, D. L. *The Journal of Physical Chemistry B* **2021**, *125*, 4241–4261.
- [20] Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. *Chemical Science* **2021**, *12*, 13958–13971.
- [21] Khalak, Y.; Tresadern, G.; Hahn, D. F.; De Groot, B. L.; Gapsys, V. *Journal of Chemical Theory and Computation* **2022**, *18*, 6259–6270.
- [22] Hahn, D. F.; Gapsys, V.; De Groot, B. L.; Mobley, D. L.; Tresadern, G. *Journal of Chemical Information and Modeling* **2024**, *64*, 5063–5076.
- [23] Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*, 2nd ed.; Computational science series 1; Academic Press: San Diego, 2002.
- [24] Yildirim, A.; Ghahremanpour, M. M.; van der Spoel, D. *Physical Review Research* **2020**, *2*, 033277.
- [25] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

- [26] Case, D. A.; Cerutti, D.; Cheateham, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N. *University of California, San Francisco* **2016**,
- [27] 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* **2009**, NA–NA.
- [28] Horta, B. A. C.; Merz, P. T.; Fuchs, P. F. J.; Dolenc, J.; Riniker, S.; Hünenberger, P. H. *Journal of Chemical Theory and Computation* **2016**, *12*, 3825–3850.
- [29] Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. *Journal of Chemical Theory and Computation* **2019**, *15*, 1863–1874.
- [30] Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. *Journal of Chemical Theory and Computation* **2021**, *17*, 4291–4300.
- [31] Qiu, Y. et al. *Journal of Chemical Theory and Computation* **2021**, *17*, 6262–6280.
- [32] Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. *Journal of Chemical Theory and Computation* **2018**, *14*, 6076–6092.
- [33] Mobley, D. L.; Bannan, C.; Wagner, J.; Rizzi, A.; Lim, N. M.; Henry, M. openforcefield/smirnoff99Frosst: Version 1.1.0. 2019; <https://zenodo.org/record/596535>.
- [34] Wagner, J.; Thompson, M.; Dotson, D.; Hyejang; SimonBoothroyd; Rodríguez-Guerra, J. openforcefield/openff-forcefields: Version 2.0.0 "Sage". 2021; <https://zenodo.org/record/5214478>.
- [35] Boothroyd, S. et al. *Journal of Chemical Theory and Computation* **2023**, *19*, 3251–3275.

- [36] He, X.; Walker, B.; Man, V. H.; Ren, P.; Wang, J. *Current Opinion in Structural Biology* **2022**, *72*, 187–193.
- [37] Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. *Journal of Chemical Information and Modeling* **2019**, *59*, 1366–1381.
- [38] Morado, J.; Mortenson, P. N.; Verdonk, M. L.; Ward, R. A.; Essex, J. W.; Skylaris, C.-K. *Journal of Chemical Information and Modeling* **2021**, *61*, 2026–2047.
- [39] Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. *Journal of Chemical Theory and Computation* **2016**, *12*, 2312–2323.
- [40] Cole, D. J.; Cabeza de Vaca, I.; Jorgensen, W. L. *MedChemComm* **2019**, *10*, 1116–1120.
- [41] Galvelis, R.; Doerr, S.; Damas, J. M.; Harvey, M. J.; De Fabritiis, G. *Journal of Chemical Information and Modeling* **2019**, *59*, 3485–3493.
- [42] Bleiziffer, P.; Schaller, K.; Riniker, S. *Journal of Chemical Information and Modeling* **2018**, *58*, 579–590.
- [43] Lehner, M. T.; Katzberger, P.; Maeder, N.; Schiebroek, C. C.; Teetz, J.; Landrum, G. A.; Riniker, S. *Journal of Chemical Information and Modeling* **2023**, *63*, 6014–6028.
- [44] Wang, Y.; Pulido, I.; Takaba, K.; Kaminow, B.; Scheen, J.; Wang, L.; Chodera, J. D. *The Journal of Physical Chemistry A* **2024**, *128*, 4160–4167, arXiv:2302.06758 [physics].
- [45] Wang, J.; Cao, D.; Tang, C.; Xu, L.; He, Q.; Yang, B.; Chen, X.; Sun, H.; Hou, T. *Briefings in Bioinformatics* **2021**, *22*, bbaa183.
- [46] Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. *The Journal of Physical Chemistry A* **2018**, *122*, 4219–4245.
- [47] 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.

- [48] Riquelme, M.; Lara, A.; Mobley, D. L.; Verstraelen, T.; Matamala, A. R.; Vöhringer-Martinez, E. *Journal of Chemical Information and Modeling* **2018**, *58*, 1779–1797.
- [49] Kelly, B. D.; Smith, W. R. *Journal of Chemical Theory and Computation* **2020**, *16*, 1146–1161.
- [50] Kelly, B. D.; Smith, W. R. *ACS Omega* **2020**, *5*, 17170–17181.
- [51] Riquelme, M.; Vöhringer-Martinez, E. *Journal of Computer-Aided Molecular Design* **2020**, *34*, 327–334.
- [52] Lara, A.; Riquelme, M.; Vöhringer-Martinez, E. *Journal of Computational Chemistry* **2018**, *39*, 1728–1737.
- [53] Ringrose, C.; Horton, J. T.; Wang, L.-P.; Cole, D. J. *Physical Chemistry Chemical Physics* **2022**, *24*, 17014–17027.
- [54] González, D.; Macaya, L.; Vöhringer-Martinez, E. *Journal of Chemical Information and Modeling* **2021**, *61*, 4462–4474.
- [55] González, D.; Macaya, L.; Castillo-Orellana, C.; Verstraelen, T.; Vogt-Geisse, S.; Vöhringer-Martinez, E. *Journal of Chemical Information and Modeling* **2022**, *62*, 4162–4174.
- [56] Horton, J. T.; Allen, A. E. A.; Cole, D. J. *Chemical Communications* **2020**, *56*, 932–935.
- [57] Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shochet, B. K.; Dill, K. A. *Journal of Molecular Biology* **2007**, *371*, 1118–1134.
- [58] Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L. *Communications Chemistry* **2023**, *6*, 222.
- [59] Hahn, D. F.; Bayly, C. I.; Boby, M. L.; Bruce Macdonald, H. E.; Chodera, J. D.; Gapsys, V.; Mey, A. S. J. S.; Mobley, D. L.; Benito, L. P.; Schindler, C. E. M.; Tresadern, G.; Warren, G. L. *Living Journal of Computational Molecular Science* **2022**, *4*.

- [60] Owen, D. J. *The EMBO Journal* **2000**, *19*, 6141–6149.
- [61] Zeng, L.; Zhou, M.-M. *FEBS Letters* **2002**, *513*, 124–128.
- [62] Mencher, S. K.; Wang, L. G. *BMC Clinical Pharmacology* **2005**, *5*, 3.
- [63] Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. *Journal of Molecular Biology* **2009**, *394*, 747–763.
- [64] Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W.; Shoichet, B. K. *Journal of Molecular Biology* **2002**, *322*, 339–355.
- [65] Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. *Journal of Molecular Biology* **2004**, *337*, 1161–1182.
- [66] Graves, A. P.; Brenk, R.; Shoichet, B. K. *Journal of Medicinal Chemistry* **2005**, *48*, 3714–3728.
- [67] Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.
- [68] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.
- [69] Ren, P.; Wu, C.; Ponder, J. W. *Journal of Chemical Theory and Computation* **2011**, *7*, 3143–3161.
- [70] Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. *Chemical Reviews* **2016**, *116*, 4983–5013.
- [71] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *Journal of Chemical Theory and Computation* **2007**, *3*, 1960–1986.
- [72] Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. *The Journal of Physical Chemistry A* **2005**, *109*, 3957–3959.
- [73] Mulliken, R. S. *The Journal of Chemical Physics* **1955**, *23*, 1833–1840.
- [74] Löwdin, P.-O. *The Journal of Chemical Physics* **1950**, *18*, 365–375.
- [75] Hirshfeld, F. L. *Theoretica Chimica Acta* **1977**, *44*, 129–138.

- [76] Heidar-Zadeh, F.; Castillo-Orellana, C.; Van Zyl, M.; Pujal, L.; Vers-traelen, T.; Bultinck, P.; Vöhringer-Martinez, E.; Ayers, P. W. *Journal of Chemical Theory and Computation* **2024**, *20*, 9939–9953.
- [77] De Jong, D. H.; Schäfer, L. V.; De Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C.; Grubmüller, H. *Journal of Computational Chemistry* **2011**, *32*, 1919–1928.
- [78] Mobley, D. L.; Chodera, J. D.; Dill, K. A. *The Journal of Chemical Physics* **2006**, *125*, 084902.
- [79] Chipot, C. *Annual Review of Biophysics* **2023**, *52*, 113–138.
- [80] Van Gunsteren, W.; Daura, X.; Mark, A. *Helvetica Chimica Acta* **2002**, *85*, 3113–3129.
- [81] Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.
- [82] Shirts, M. R.; Pande, V. S. *The Journal of Chemical Physics* **2005**, *122*, 144107.
- [83] Lu, N.; Singh, J. K.; Kofke, D. A. *The Journal of Chemical Physics* **2003**, *118*, 2977–2984.
- [84] Kirkwood, J. G. *The Journal of Chemical Physics* **1935**, *3*, 300–313.
- [85] Steinbrecher, T.; Joung, I.; Case, D. A. *Journal of Computational Chemistry* **2011**, *32*, 3253–3263.
- [86] Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chemical Physics Letters* **1994**, *222*, 529–539.
- [87] Bennett, C. H. *Journal of Computational Physics* **1976**, *22*, 245–268.
- [88] Crooks, G. E. *Physical Review E* **2000**, *61*, 2361–2366.
- [89] Shirts, M. R.; Chodera, J. D. *The Journal of Chemical Physics* **2008**, *129*, 124105.
- [90] Chelli, R.; Marsili, S.; Barducci, A.; Procacci, P. *Physical Review E* **2007**, *75*, 050101.

- [91] Pohorille, A.; Jarzynski, C.; Chipot, C. *The Journal of Physical Chemistry B* **2010**, *114*, 10235–10253.
- [92] Procacci, P.; Cardelli, C. *Journal of Chemical Theory and Computation* **2014**, *10*, 2813–2823.
- [93] Jarzynski, C. *Physical Review Letters* **1997**, *78*, 2690–2693.
- [94] Cuendet, M. A. *The Journal of Chemical Physics* **2006**, *125*, 144109.
- [95] Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H. In *Molecular Modeling of Proteins*; Kukol, A., Ed.; Springer New York: New York, NY, 2015; Vol. 1215; pp 173–209, Series Title: Methods in Molecular Biology.
- [96] Crooks, G. E. *Journal of Statistical Physics* **1998**, *90*, 1481–1487.
- [97] Crooks, G. E. *Physical Review E* **1999**, *60*, 2721–2726.
- [98] Goette, M.; Grubmüller, H. *Journal of Computational Chemistry* **2009**, *30*, 447–456.
- [99] Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Physical Review Letters* **2003**, *91*, 140601.
- [100] Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *389*, 457–484.
- [101] Parr, R. G.; Weitao, Y. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press, 1995.
- [102] Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864–B871.
- [103] Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133–A1138.
- [104] Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648–5652.
- [105] Bursch, M.; Mewes, J.; Hansen, A.; Grimme, S. *Angewandte Chemie International Edition* **2022**, *61*, e202205735.
- [106] Weigend, F.; Ahlrichs, R. *Physical Chemistry Chemical Physics* **2005**, *7*, 3297.

- [107] Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *The Journal of Chemical Physics* **2003**, *119*, 11113–11123.
- [108] Rappoport, D.; Furche, F. *The Journal of Chemical Physics* **2010**, *133*, 134105.
- [109] Yuan, Y.; Wang, F. *The Journal of Chemical Physics* **2021**, *155*, 094104.
- [110] Lin, H.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2007**, *117*, 185.
- [111] Senn, H. M.; Thiel, W. *Angewandte Chemie International Edition* **2009**, *48*, 1198–1229.
- [112] Clemente, C. M.; Capece, L.; Martí, M. A. *Journal of Chemical Information and Modeling* **2023**, *63*, 2609–2627.
- [113] Macaya, L.; González, D.; Vöhringer-Martinez, E. *Journal of Chemical Information and Modeling* **2024**, *64*, 3269–3277.
- [114] Pulido, J.; Macaya, L.; Vöhringer-Martinez, E. *Journal of Chemical & Engineering Data* **2024**, acs.jced.4c00144.
- [115] Mohebifar, M.; Johnson, E. R.; Rowley, C. N. *Journal of Chemical Theory and Computation* **2017**, *13*, 6146–6157.
- [116] Castillo-Orellana, C.; Heidar-Zadeh, F.; Vöhringer-Martinez, E. *Journal of Chemical Theory and Computation* **2025**, *21*, 2043–2054, Publisher: American Chemical Society (ACS).
- [117] Tkatchenko, A.; Scheffler, M. *Physical Review Letters* **2009**, *102*, 073005.
- [118] Fedorov, D. V.; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. *Physical Review Letters* **2018**, *121*, 183401.
- [119] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.
- [120] Bussi, G.; Donadio, D.; Parrinello, M. *The Journal of Chemical Physics* **2007**, *126*, 014101.

- [121] Parrinello, M.; Rahman, A. *Journal of Applied Physics* **1981**, *52*, 7182–7190.
- [122] Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Molecular Physics* **1996**, *87*, 1117–1157.
- [123] Pernot, P. *The Journal of Chemical Physics* **2022**, *156*, 114109.
- [124] Kendall, M. G. *Biometrika* **1938**, *30*, 81–93.
- [125] King, E.; Qi, R.; Li, H.; Luo, R.; Aitchison, E. *Journal of Chemical Theory and Computation* **2021**, *17*, 2541–2555.
- [126] Van Rossum, Guido; Drake, F. L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, 2009.
- [127] Verstraelen, T. et al. *Journal of Computational Chemistry* **2021**, *42*, 458–464.
- [128] Tehrani, A.; Yang, X. D.; Martínez-González, M.; Pujal, L.; Hernández-Esparza, R.; Chan, M.; Vöhringer-Martinez, E.; Verstraelen, T.; Ayers, P. W.; Heidar-Zadeh, F. *The Journal of Chemical Physics* **2024**, *160*, 172503.
- [129] Kim, T. D. et al. *The Journal of Chemical Physics* **2024**, *161*, 042503.
- [130] Zanette, C.; Bannan, C. C.; Bayly, C. I.; Fass, J.; Gilson, M. K.; Shirts, M. R.; Chodera, J. D.; Mobley, D. L. *Journal of Chemical Theory and Computation* **2019**, *15*, 402–423.
- [131] Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. *The Journal of Chemical Physics* **2020**, *152*, 224108.
- [132] Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLOS Computational Biology* **2017**, *13*, e1005659.
- [133] Muller, S.; Filippakopoulos, P.; Knapp, S. *Expert Reviews in Molecular Medicine* **2011**, *13*, e29.

- [134] Filippakopoulos, P.; Knapp, S. *Nature Reviews Drug Discovery* **2014**, *13*, 337–356.
- [135] Ntranos, A.; Casaccia, P. *Neuroscience Letters* **2016**, *625*, 4–10.
- [136] Filippakopoulos, P.; Knapp, S. *FEBS Letters* **2012**, *586*, 2692–2704.
- [137] Jung, M.; Gelato, K. A.; Fernández-Montalván, A.; Siegel, S.; Haendler, B. *Epigenomics* **2015**, *7*, 487–501.
- [138] Huggins, D. J. *Journal of Chemical Theory and Computation* **2022**, acs.jctc.1c01208.
- [139] Silva, A. F.; Guest, E. E.; Falcone, B. N.; Pickett, S. D.; Rogers, D. M.; Hirst, J. D. *Molecular Physics* **2023**, *121*, e2124201.
- [140] Summa, C. M.; Langford, D. P.; Dinshaw, S. H.; Webb, J.; Rick, S. W. *Journal of Chemical Theory and Computation* **2024**, *20*, 2812–2819.
- [141] Guest, E. E.; Cervantes, L. F.; Pickett, S. D.; Brooks, C. L.; Hirst, J. D. *Journal of Chemical Information and Modeling* **2022**, *62*, 1458–1470.
- [142] Çınaroğlu, S. S.; Biggin, P. C. *Chemical Science* **2023**, *14*, 6792–6805.
- [143] Procacci, P. *Journal of Chemical Theory and Computation* **2023**, *19*, 8942–8954.
- [144] Barros, E. P.; Ries, B.; Champion, C.; Rieder, S. R.; Riniker, S. *Journal of Chemical Information and Modeling* **2023**, *63*, 1794–1805.
- [145] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1-2*, 19–25.
- [146] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- [147] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.
- [148] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

- [149] Joung, I. S.; Cheatham, T. E. *The Journal of Physical Chemistry B* **2008**, *112*, 9020–9041.
- [150] Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2000**, *21*, 132–146.
- [151] Jakalian, A.; Jack, D. B.; Bayly, C. I. *Journal of Computational Chemistry* **2002**, *23*, 1623–1641.
- [152] Leimkuhler, B.; Matthews, C. *The Journal of Chemical Physics* **2013**, *138*, 174102.
- [153] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.
- [154] Gapsys, V.; Seeliger, D.; de Groot, B. L. *Journal of Chemical Theory and Computation* **2012**, *8*, 2373–2382.
- [155] Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. *The Journal of Chemical Physics* **2013**, *139*, 184103.
- [156] Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. *Journal of Computational Chemistry* **2015**, *36*, 348–354.
- [157] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.
- [158] Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. *Journal of Chemical Theory and Computation* **2016**, *12*, 3926–3947.
- [159] Petrov, D.; Perthold, J. W.; Oostenbrink, C.; De Groot, B. L.; Gapsys, V. *Journal of Chemical Theory and Computation* **2024**, *20*, 914–925.
- [160] Boulanger, E.; Huang, L.; Rupakheti, C.; MacKerell, A. D.; Roux, B. *Journal of Chemical Theory and Computation* **2018**, *14*, 3121–3131.
- [161] Picaud, S. et al. *Science Advances* **2016**, *2*, e1600760.

- [162] Zhu, J.; Caffisch, A. *Journal of Medicinal Chemistry* **2016**, *59*, 5555–5561.
- [163] Karim, R. M.; Chan, A.; Zhu, J.-Y.; Schönbrunn, E. *Journal of Medicinal Chemistry* **2020**, *63*, 3227–3237.
- [164] Karim, R. M.; Bikowitz, M. J.; Chan, A.; Zhu, J.-Y.; Grassie, D.; Becker, A.; Berndt, N.; Gunawan, S.; Lawrence, N. J.; Schönbrunn, E. *Journal of Medicinal Chemistry* **2021**, *64*, 15772–15786.
- [165] Karim, R. M.; Yang, L.; Chen, L.; Bikowitz, M. J.; Lu, J.; Grassie, D.; Shultz, Z. P.; Lopchuk, J. M.; Chen, J.; Schönbrunn, E. *Journal of Medicinal Chemistry* **2022**, *65*, 4182–4200.
- [166] Bhati, A. P.; Wan, S.; Coveney, P. V. *Journal of Chemical Theory and Computation* **2025**, *21*, 440–462.
- [167] Mobley, D. L.; Gilson, M. K. *Annual Review of Biophysics* **2017**, *46*, 531–558.
- [168] Pujal, L.; Van Zyl, M.; Vöhringer-Martinez, E.; Verstraelen, T.; Bultinck, P.; Ayers, P. W.; Heidar-Zadeh, F. *The Journal of Chemical Physics* **2022**, *156*, 194109.
- [169] Plazinski, W.; Lutsyk, V.; Plazinska, A. *Journal of Chemical Theory and Computation* **2024**, *20*, 2273–2283.
- [170] Charron, N. E. et al. *Nature Chemistry* **2025**, *17*, 1284–1292.
- [171] Lewis, S. et al. *Science* **2025**, eadv9817.

A. Anexo - Especificidad en BRD4(1)

Tabla A.1: Conjunto de datos BRD4(1).

Sistema	PDB ID	Resolución (Å)
1	4ogi	1.73
2	3mxf	1.60
3	4mr3	1.68
4	4ogj	1.65
5	4j0r	1.72
6	3u5l	1.39
7	4mr4	1.66
8	3u5j	1.60
9	3svg	1.68
10	4hbv	1.63
11	Modelo (3svg)	-

Tabla A.2: **Energía necesaria para polarizar los ligandos (1-11) en diferentes entornos moleculares para el conjunto de datos BRD4(1).** Las simulaciones se realizaron utilizando el campo de fuerza OFF-2.0.0, variando los parámetros no enlazantes del campo de fuerza. Los valores de $\langle E_{\text{pol}} \rangle_{\text{prot}}$ y $\langle E_{\text{pol}} \rangle_{\text{solv}}$ se calcularon utilizando los parámetros de LJ y las cargas atómicas D-MBIS derivadas para el ligando en estado unido (proteína) y no unido (solvente). Los valores se expresan en $\text{kcal}\cdot\text{mol}^{-1}$.

Sistema	$\langle E_{\text{pol}} \rangle_{\text{prot}}$	$\langle E_{\text{pol}} \rangle_{\text{solv}}$
1	17.7 ± 0.9	21.6 ± 0.9
2	8.4 ± 0.4	12.3 ± 0.7
3	9.3 ± 0.5	11.9 ± 0.5
4	10.0 ± 0.5	20.3 ± 1.2
5	6.9 ± 0.5	10.2 ± 0.5
6	7.1 ± 0.3	9.8 ± 0.6
7	9.0 ± 0.3	13.5 ± 0.5
8	7.9 ± 0.4	11.5 ± 0.9
9	5.7 ± 0.3	8.3 ± 0.7
10	4.1 ± 0.3	7.7 ± 0.5
11	4.6 ± 0.4	5.7 ± 0.5

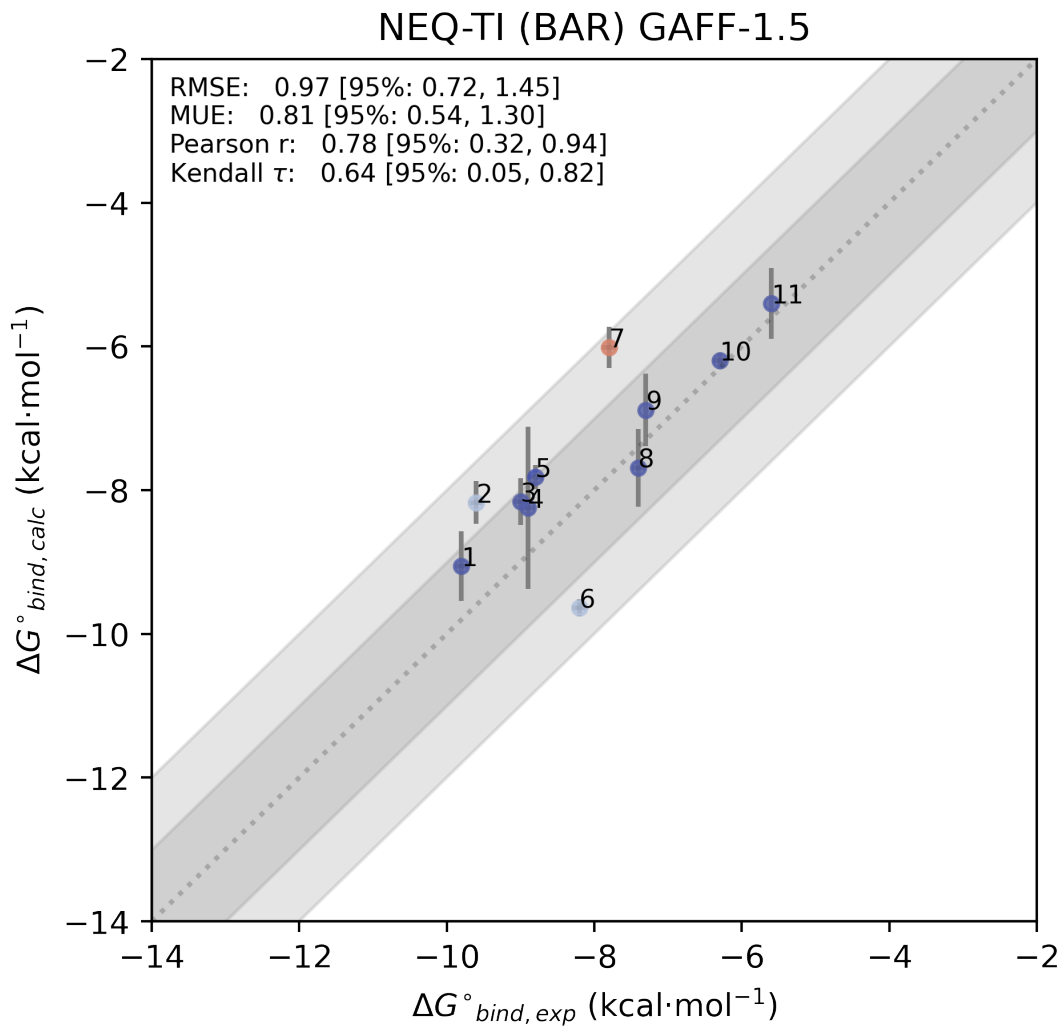


Figura A.1: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.5 y cargas atómicas AM1-BCC para el conjunto de datos BRD4(1).

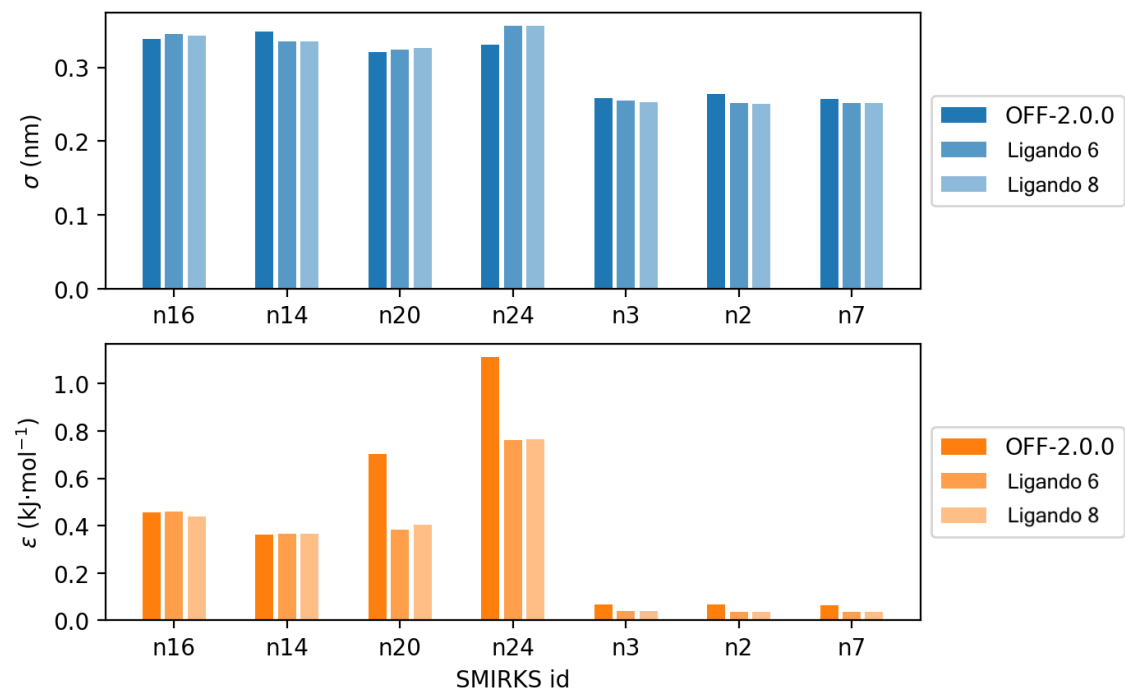


Figura A.2: Parámetros de Lennard-Jones de ligandos 6 y 8.

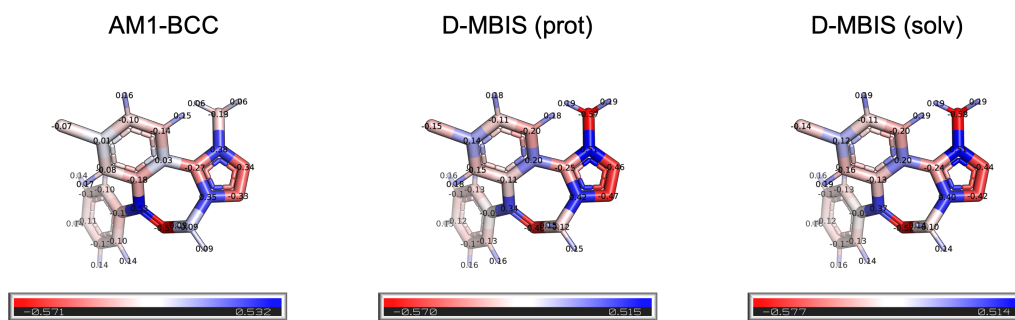


Figura A.3: Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 8.

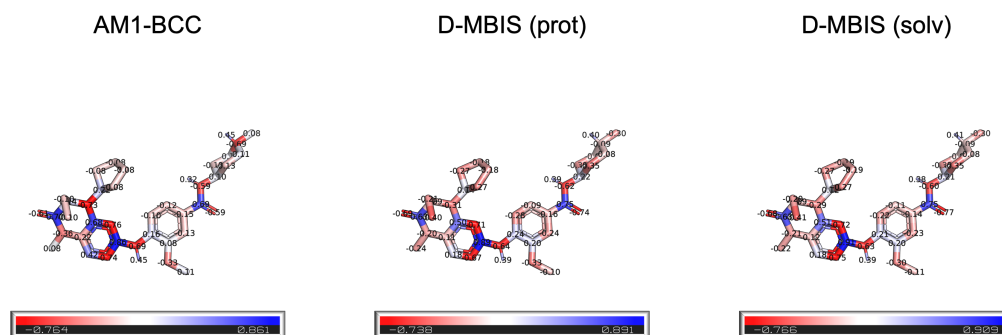


Figura A.4: Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 1.

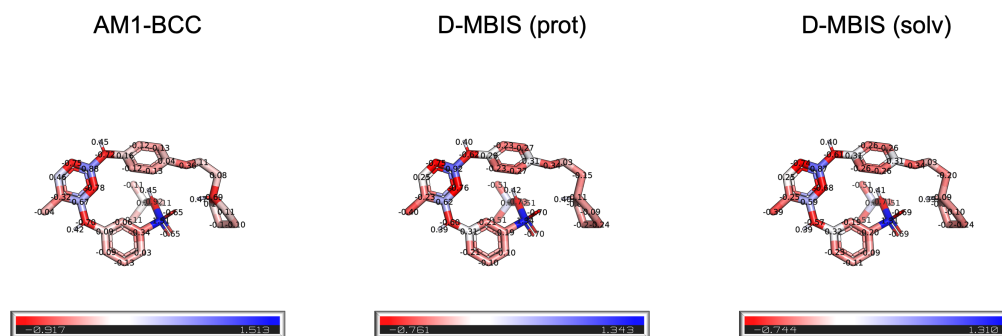


Figura A.5: Cargas atómicas AM1-BCC y D-MBIS derivados desde los distintos ambientes moleculares para el ligando 4.

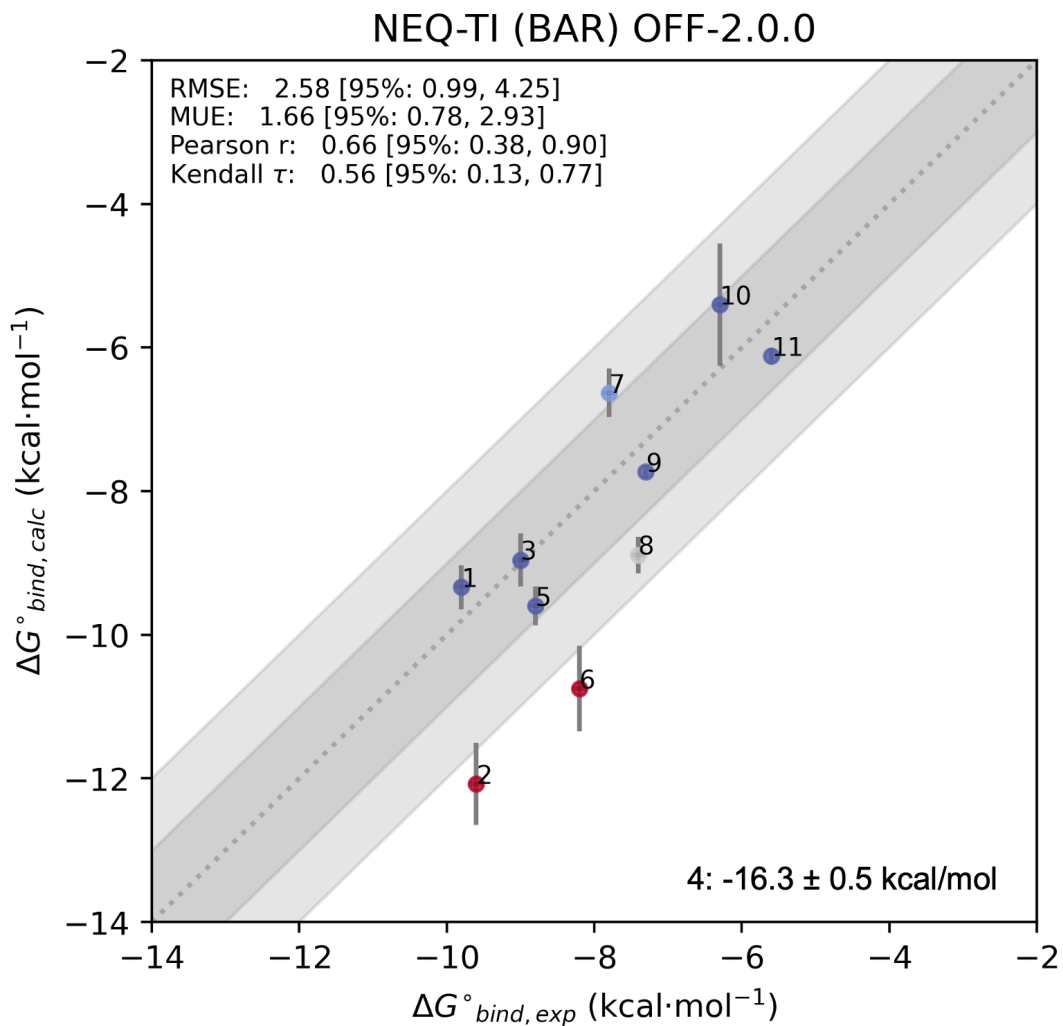


Figura A.6: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza OFF-2.0.0 para el conjunto de datos BRD4(1), con los valores obtenidos para los ligandos 1 y 4 mediante la metodología DSSB.

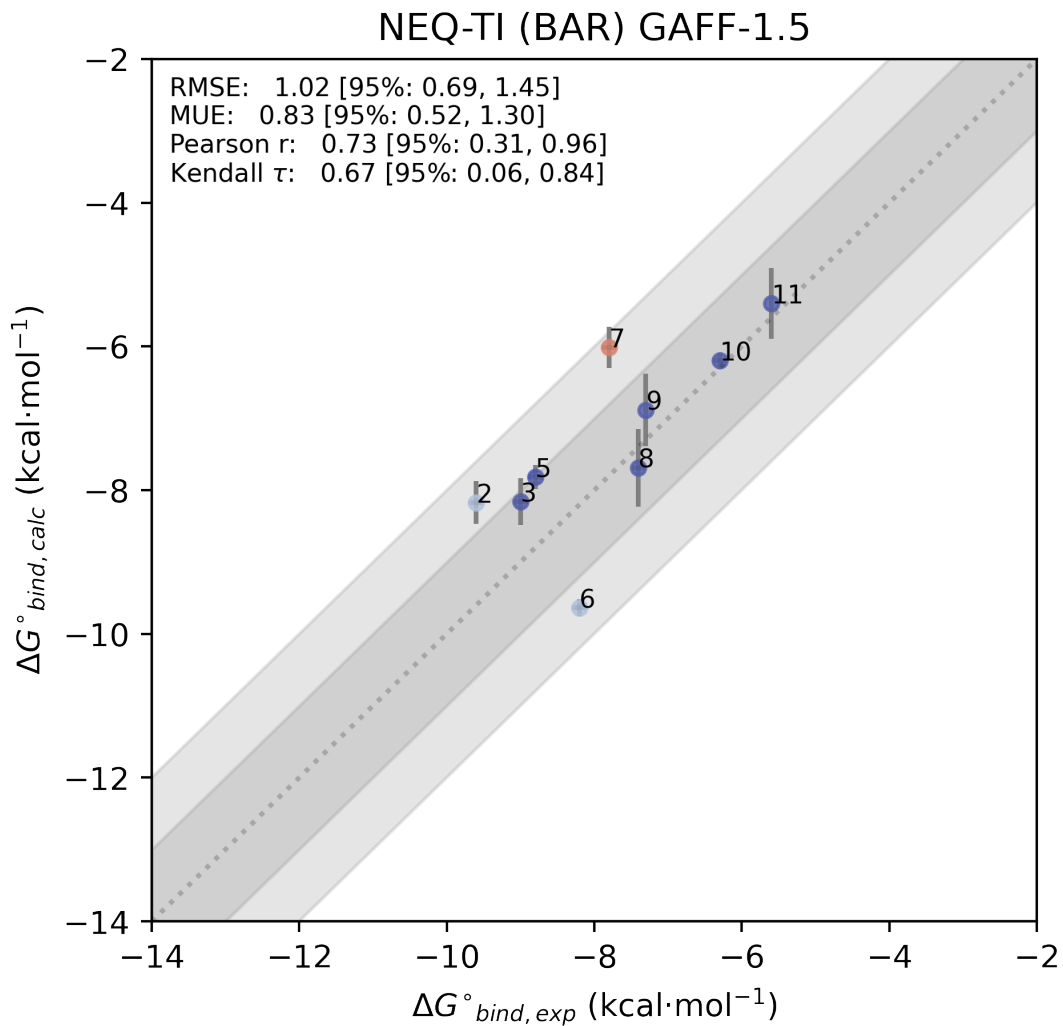


Figura A.7: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.5, sin incluir la estimación de ligandos con carga (1 y 4).

B. Anexo - Perfil de selectividad de Bromosporina

Tabla B.1: Conjunto de datos Bromosporina-bromodominios.

Sistema	Familia	Estructura ¹	Identificador	PDB ID
CECR2	I	apo	1a	3nxb
FALZ	I	apo	1b	3uv2
PCAF	I	apo	1c	3gg3
BRD2(1)	II	holo	2a	4alg
BRD2(2)	II	holo	2b	4mr5
BRD3(1)	II	holo	2c	3s91
BRD3(2)	II	holo	2d	3s92
BRD4(1)	II	apo	2e	2oss
BRD4(2)	II	apo	2f	2ouo
BRDT(1)	II	holo	2g	4kcx
CREBBP	III	holo	3a	4nyx
EP300	III	apo	3b	3i3j
BRD1	IV	holo	4a	5ame
BRD9	IV	holo	4b	4xy8
BRPF1B	IV	apo	4c	4lc2
BAZ2A	V	holo	5a	4qbm
TIF1	V	holo	5b	4ybm
TAF1(1)	VII	apo	7a	3uv5
TAF1(2)	VII	apo	7b	3uv4
TAF1L(2)	VII	apo	7c	3hnh
PB1(5)	VIII	holo	8a	3mb4
SMARCA4	VIII	apo	8b	2grc

¹ Estado conformacional usado por Aldeghi y colaboradores para la generación del complejo *in silico*.

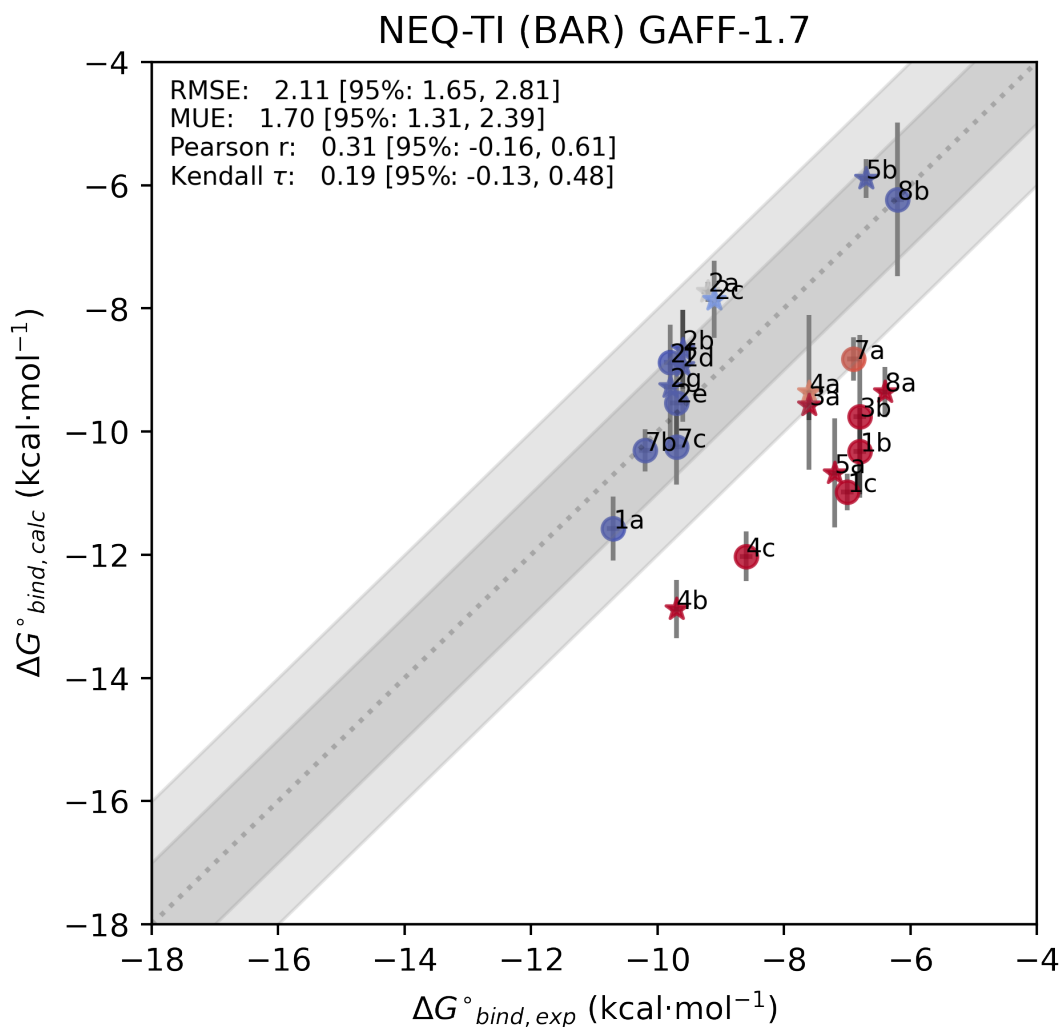


Figura B.1: Gráfico de paridad de valores de energía libre de unión obtenidos con el campo de fuerza GAFF-1.7 y cargas atómicas AM1-BCC para el conjunto de datos de bromosporina. Se presentan los distintos estados conformacionales (apo: círculo, holo: estrella) usados como estructura inicial en la generación de los complejos *in silico* por Aldeghi y colaboradores.

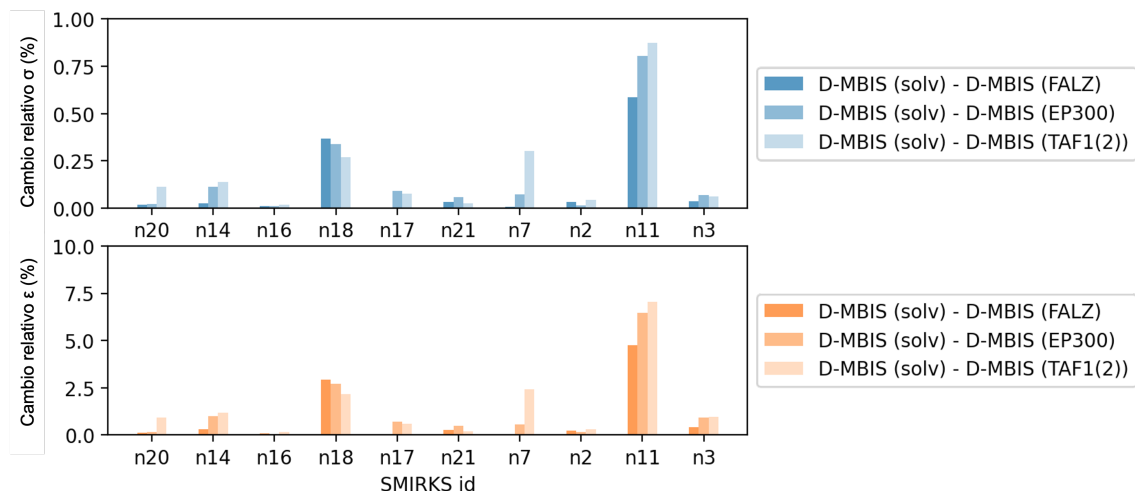


Figura B.2: Cambio relativo porcentual de valores de parámetros de Lennard-Jones (σ y ϵ) del campo de fuerza OFF-2.0.0 y los obtenidos en ambientes químicos de solvente y complejo proteína-ligando de los bromodominios seleccionados.

Tabla B.2: Energía necesaria para polarizar el ligando bromosporina en diferentes entornos moleculares para el conjunto de datos de bromodominios. Las simulaciones se realizaron utilizando el campo de fuerza OFF-2.0.0, variando los parámetros no enlazantes del campo de fuerza. Los valores de $\langle E_{\text{pol}} \rangle_{\text{prot}}$ y $\langle E_{\text{pol}} \rangle_{\text{solv}}$ se calcularon utilizando los parámetros de LJ y las cargas atómicas D-MBIS derivadas para el ligando en estado unido (proteína). Los valores se expresan en $\text{kcal}\cdot\text{mol}^{-1}$.

Sistema	$\langle E_{\text{pol}} \rangle_{\text{prot}}$
FALZ	13.0 ± 0.4
BRD4(1)	12.0 ± 0.4
EP300	11.8 ± 0.4
BRD4(1)	12.0 ± 0.4
BRPF1B	14.6 ± 0.4
TIF1	12.6 ± 0.4
TAF1(2)	13.2 ± 0.4
PB1(5)	13.6 ± 0.3

Tabla B.3: **Energía necesaria para polarizar el ligando bromosporina en diferentes entornos moleculares para el conjunto de datos de bromodominios.** Las simulaciones se realizaron utilizando el campo de fuerza OFF-2.0.0, variando los parámetros no enlazantes del campo de fuerza. Los valores de $\langle E_{\text{pol}} \rangle_{\text{prot}}$ y $\langle E_{\text{pol}} \rangle_{\text{solv}}$ se calcularon utilizando los parámetros de LJ y las cargas atómicas D-MBIS derivadas para el ligando en estado no unido (solvente). Los valores se expresan en $\text{kcal}\cdot\text{mol}^{-1}$.

Sistema	$\langle E_{\text{pol}} \rangle_{\text{solv}}$
Solvente	22.0 ± 1.0

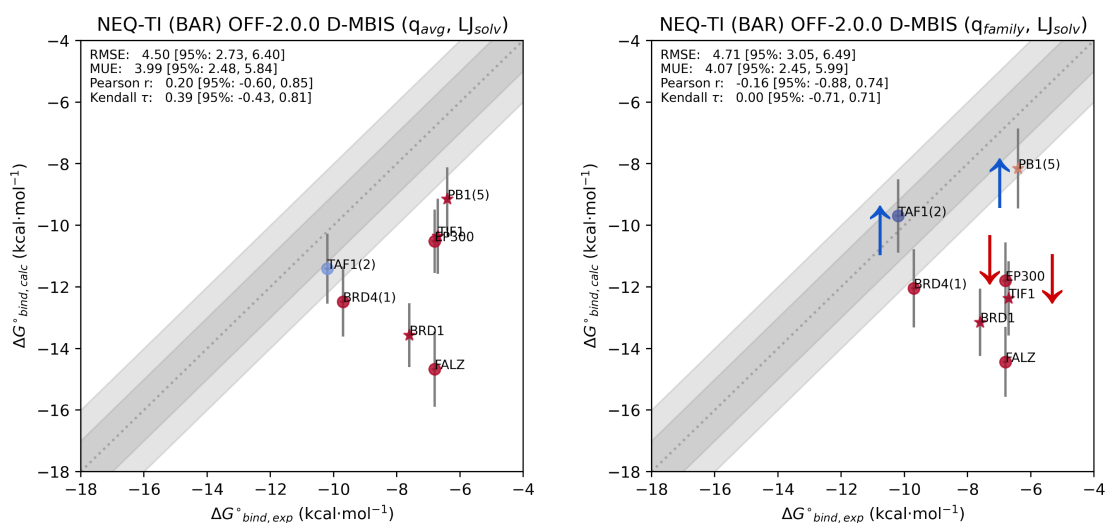


Figura B.3: **Gráfico de paridad de valores de energía libre de unión calculados para los distintos bromodominios utilizados en la derivación de parámetros por familia usando distintos conjuntos de cargas atómicas.** A la izquierda se muestran los resultados obtenidos con cargas atómicas promedio, mientras que a la derecha los valores obtenidos usando cargas atómicas específicas por bromodominio. Las flechas azules y rojas indican una mejora y deterioro en la estimación comparado al valor experimental respectivamente.

Tabla B.4: **Conjunto de datos para complejos cristalizados de bromosporina**

Sistema	PDB ID	Resolución (Å)
BRD2(1)	7lah	1.60
BRD2(2)	7l6d	1.55
BRD3(2)	7lb4	2.00
BRD4(1)	5igk	1.70
BRD9	5igm	1.60
BRPF1B	5c7n	1.75
TAF1(2)	7k1p	2.45
TAF1L(2)	5igl	2.10

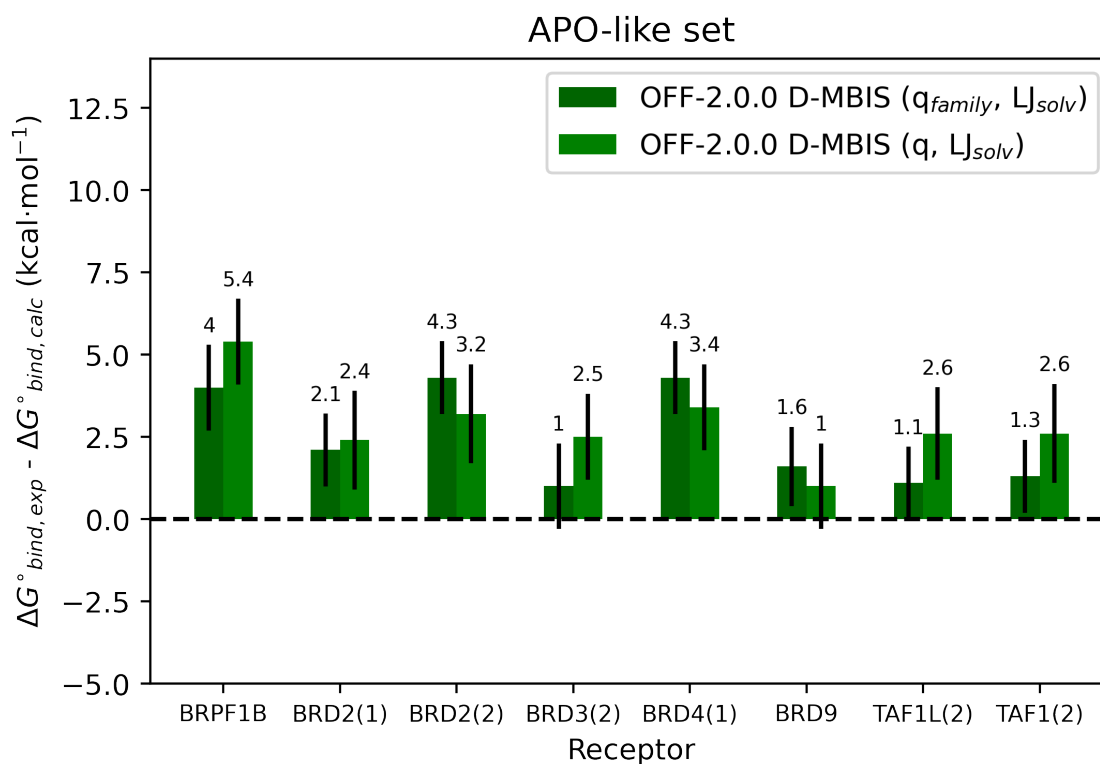


Figura B.4: **Diferencia entre valores de energía libre de unión experimental y obtenidos con el campo de fuerza OFF-2.0.0 usando parámetros no enlazantes D-MBIS.** Los parámetros no enlazantes corresponden a cargas atómicas por familia de bromodominios y cargas atómicas derivadas por bromodominio, ambos con parámetros de Lennard-Jones derivados desde solvente.

C. Anexo - Especificidad en Lisozi- ma T4 L99A/M102Q

Tabla C.1: **Energía necesaria para polarizar los ligandos (L1-8) en diferentes entornos moleculares para el conjunto de datos de lisozima T4 L99A/M102Q.** Las simulaciones se realizaron utilizando el campo de fuerza GAFF-1.8, variando los parámetros no enlazantes del campo de fuerza. Los valores de $\langle E_{\text{pol}} \rangle_{\text{prot}}$ y $\langle E_{\text{pol}} \rangle_{\text{solv}}$ se calcularon utilizando los parámetros de LJ y las cargas atómicas D-MBIS derivadas para el ligando en estado unido (proteína) y no unido (solvente). Los valores se expresan en $\text{kcal}\cdot\text{mol}^{-1}$.

Sistema	$\langle E_{\text{pol}} \rangle_{\text{prot}}$	$\langle E_{\text{pol}} \rangle_{\text{solv}}$
L1	0.7 ± 0.3	2.4 ± 0.8
L2	1.3 ± 0.3	3.5 ± 0.8
L3	1.5 ± 0.4	3.4 ± 0.9
L4	1.6 ± 0.8	4.3 ± 1.2
L5	1.1 ± 0.3	3.7 ± 1.1
L6	1.1 ± 0.2	5.4 ± 1.8
L7	1.2 ± 0.3	3.6 ± 1.2
L8	1.2 ± 0.3	5.6 ± 1.4

D. Presentaciones en congresos y publicaciones

D.1. Congresos

1. QCMM-2022, Concepción, Chile. Póster. *ffparAIM: Python workflow for non-covalent force field parameters derivation from MBIS partitioning of the polarized electron density and its applications.*
2. 12th Workshop in Computational Chemistry and Molecular Spectroscopy, Concón, Chile. Presentación oral. *ffparAIM: Python workflow for non-covalent force field parameters derivation and its applications.*
3. CUP XXIII, Santa Fe, Estados Unidos. Póster. *ffparAIM: Python workflow for non-covalent force field parameters derivation and its applications in binding affinity prediction.*
4. 3ra Reunión Anual de la Sociedad Chilena de Bioinformática, Concepción, Chile. Póster. *What is the effect of bespoke force fields in the selectivity profile of bromosporine?*

D.2. Publicaciones

1. **Nonbonded Force Field Parameters from Minimal Basis Iterative Stockholder Partitioning of the Molecular Electron Density Improve CB7 Host–Guest Affinity Predictions**

González, D., Macaya, L., Castillo-Orellana, C., Verstraelen, T., Vogt-Geisse, S. and Vohringer-Martinez, E. *Journal of Chemical Information and Modeling*, 2022, vol. 62, no 17, p. 4162-4174. DOI: [10.1021/acs.jcim.2c00316](https://doi.org/10.1021/acs.jcim.2c00316)

2. **Nonbonded Force Field Parameters from MBIS Partitioning of the Molecular Electron Density Improve Binding Affinity Predictions of the T4-Lysozyme Double Mutant**

Macaya, L., González, D., and Vohringer-Martinez, E. *Journal of Chemical Information and Modeling*, 2024, vol. 64, no 8, p. 3269-3277. DOI: [10.1021/acs.jcim.3c01912](https://doi.org/10.1021/acs.jcim.3c01912)

3. **Nonbonded Force Field Parameters from MBIS Partitioning of the Molecular Electron Density Improve Thermophysical Properties Prediction of Organic Liquids**

Pulido, J., Macaya, L., González, D., and Vohringer-Martinez, E. *Journal of Chemical and Engineering Data*, 2024, vol. 69, no 9, p. 2917-2926. DOI: [10.1021/acs.jced.4c00144](https://doi.org/10.1021/acs.jced.4c00144)