



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

**TRANSFORMACIONES ALQUÍMICAS DE NO-EQUILIBRIO
APLICADAS A LA INGENIERÍA DE PROTEÍNAS**

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

Por: Carlos Andrés Castillo Orellana

Profesor Guía: Dr. Esteban Vöhringer-Martinez

Marzo 2025
Concepción, Chile

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

*"Do not go where the path may lead,
go instead where there is no path
and leave a trail"*

Ralph Waldo Emerson

Agradecimientos

En primer lugar, agradezco a la Agencia Nacional de Investigación y Desarrollo (ANID) por el apoyo financiero a mi proyecto doctoral a través de la Beca de Doctorado Nacional, así como a la Facultad de Ciencias Químicas por proporcionarme la formación académica necesaria para llevar a cabo esta tesis.

Asimismo, quiero expresar mi gratitud a cada uno de los miembros del grupo de Química Cuántica y Modelamiento Molecular (QCMM) por sus valiosos consejos y sugerencias, así como por brindarme un entorno científico de primer nivel que ha sido fundamental en mi desarrollo. En especial, agradezco al Dr. Esteban Vöhringer-Martinez, quien, como mi profesor guía tanto en pregrado como en posgrado, me ha brindado un apoyo incondicional, permitiéndome crecer profesionalmente y acompañándome en este camino que comenzó hace más de ocho años.

Adicionalmente, quiero agradecer a la Dra. Farnaz Heidar-Zadeh, por ser mi mentora en mi pasantía doctoral en la Universidad de Queens, en Canadá, siendo una segunda guía en este trabajo de tesis. En este contexto quiero agradecer a los miembros de su grupo de investigación, por sus consejos y sugerencias durante mi estadía. Además, quiero agradecer a la agencia de

asuntos globales de Canadá, quienes me otorgaron la beca *Emergin Leaders in the Americas Program* (ELAP), financiando los gastos asociados a mi pasantía.

Finalmente, quiero agradecer a mi familia y seres queridos, cuyo apoyo incondicional y contención han sido un pilar fundamental en mi vida. A ellos dedico especialmente este trabajo.

Contenido

Lista de Figuras	VI
Lista de Tablas	VIII
Resumen	XI
Abstract	XIV
1. Introducción	1
1.1. Ingeniería de proteínas	1
1.2. Metodologías experimentales en la ingeniería de proteínas .	2
1.3. Diseño asistido por computadora	4
1.3.1. métodos computacionales basados en principios estadísticos	6
1.3.2. métodos computacionales híbridos	7
1.3.3. métodos computacionales basados en principios físicoquímicos	8
1.3.4. métodos computacionales basados en mecánica cuántica	10

1.3.5. Transformaciones alquímicas de no-equilibrio como potencial metodología para la evaluación de variantes enzimáticas	12
1.4. Hipótesis y Objetivos	18
2. Metodología	20
2.1. Mecánica cuántica	21
2.1.1. Método usados en química cuántica	22
2.2. Mecánica molecular	29
2.2.1. Interacciones no-covalentes	30
2.2.2. Interacciones covalentes	31
2.2.3. Parametrización de un campo de fuerza	33
2.3. Metodos de simulación multiescala mecánica cuántica/mecánica molecular (QM/MM)	39
2.4. Dinámica molecular	42
2.5. Cálculos de energía libre mediante transformaciones alquímicas	45
2.5.1. Estimadores de energía libre en transformaciones al- químicas	49
2.5.2. Transformaciones de No-equilibrio	54
2.6. Cálculos de energía libre de activación $\Delta G^\ddagger$	59
2.6.1. Método de la cuerda adaptativa	61

2.7.	Calculos de energía libre de activación relativa mediante transformaciones alquímicas	65
2.7.1.	Derivación de un campo de fuerza para estados de transición	67
3.	Resultados y Discusión	74
3.1.	Validación de las transformaciones alquímicas de no-equilibrio en la enzima Crotonil-CoA Carboxilasa/Reductasa CCR . .	76
3.2.	Derivación de parámetros no enlazantes para aminoácidos mediante la teoría de átomos-en-moléculas (AIM)	93
3.2.1.	Interacciones electrostáticas	95
3.2.2.	Interacciones de dispersión	102
3.2.3.	Interacciones de repulsión	103
3.2.4.	Interacciones dipéptido-dipéptido	107
3.3.	Evaluación del efecto de nuevos parámetros no-covalentes sobre la estimación de $\Delta\Delta G^\ddagger$	111
3.4.	Validación de la metodología en un segundo sistema enzimático: Dihidrofolato Reductasa (DHFR)	115
3.5.	Obtención del camino de reacción para la enzima Old Yellow Enzyme (OYE)	123
3.6.	Aplicación de las transformaciones alquímicas para la búsqueda de nuevas variantes de OYE Y187F	131

4. Conclusiones	139
4.1. Proyecciones	145
Bibliografía	148
A. Análisis de los valores de energía y parámetros no enlazantes para las interacciones entre residuos aminoacídicos	162
B. Corrección al perfil de energía libre usando el método DFTB3	169
C. Parámetros de Simulación	171
D. Análisis Conformacional del sistema OYE	172
E. Resultados Experimentales	175
F. Presentaciones en congresos y publicaciones	176
F.1. Congresos	176
F.2. Publicaciones	177
F.3. Pasantías y Reconocimientos	178

Lista de Figuras

1.1. Perfil de reacción y ciclo Termodinámico para la estimación de la energía libre de activación	13
2.1. Métodos de estimación de energía libre basados en transformaciones alquímicas	47
3.1. Flujo de trabajo para la estimación de $\Delta\Delta G^\ddagger$	82
3.2. Perfil de energía libre para la reacción catalizada por CCR .	83
3.3. Distancia donador-aceptor(Å) del ET de H365N.	89
3.4. Conjunto de dímeros de residuos aminoacídicos utilizados en el análisis de parámetros no covalentes	96
3.5. Errores absolutos en la energía de interacción electrostática clásica	98
3.6. Gráfico de paridad de la energía de interacción electrostática clásica, comparado a los valores de referencia de ALMO-EDA2	101
3.7. Gráfico de paridad de las energías de interacción de dispersión, comparando los valores ALMO-EDA2 con la energía de dispersión clásica calculada utilizando parámetros AIM .	104
3.8. Gráfico de paridad de la energía de interacción de repulsión por intercambio	108
3.9. Interacciones dipéptido-dipéptido	110
3.10. Sitio activo de DHFR	116
3.11. Perfil de energía libre de reacción del sistema DHFR	120
3.12. Mecanismo de reacción de la variante Y187F de <i>Old Yellow Enzyme</i>	124
3.13. Perfil de energía libre de reacción para la mutante Y187F de OYE	126
3.14. Variables colectivas asociadas al mecanismo de reacción de la mutante Y187F de la enzima OYE	128

3.15. Perfil de energía libre de reacción corregido, para la transferencia de hidruro en la mutante Y187F de la enzima OYE . .	130
3.16. Residuos de OYE Y187F a mutar para la búsqueda de variantes catalíticamente más activas	132
3.17. Ciclo termodinámico para la estimación de la estabilidad del complejo de Michaelis-Menten	134
A.1. Efecto de la corrección por penetración de carga en sistemas cargados de igual signo	163
A.2. Efecto de la corrección por penetración de carga en sistemas cargados de signo opuesto	164
A.3. Efecto de la corrección por penetración de carga en sistemas aromático-aromático e hidrofílico-hidrofílico	165
A.4. Efecto de la corrección por penetración de carga en sistemas aromáticos e hidrofílicos	166
A.5. Distribución de los coeficientes C_6 para los átomos de carbono e hidrógeno	167
A.6. Distribución de los coeficientes radios de VdW para los átomos de carbono e hidrógeno	168
B.1. Corrección al perfil de energía libre de DHFR	170
D.1. Evolución del RMSD de la proteína OYE en el complejo de Michaelis-Menten, para diferentes mutantes a lo largo del tiempo	173
D.2. Evolución del RMSD de la proteína OYE en el estado de transición, para diferentes mutantes a lo largo del tiempo	174
E.1. Actividad aldólica de distintas variantes de OYE	175

Lista de Tablas

3.1.	Diferencia de energía libre de activación experimental en kJ/mol de las mutantes: N81L, F170A, E171A y H365N	84
3.2.	Análisis de la robustez en la estimación de $\Delta\Delta G^\ddagger$ sobre el sistema F170A	85
3.3.	Estimación del cambio en la energía libre de activación para diferentes mutantes en CCR	87
3.4.	Análisis comparativo de las contribuciones energéticas obtenidas por ALMO-EDA2 y SAPT	106
3.5.	Estimación del cambio en la energía libre de activación en mutantes de CCR con nuevos parámetros no covalentes. . . .	113
3.6.	Cambio en la energía libre de activación para diferentes mutantes de DHFR. Los términos q y LJ representan reparametrización de cargas atómicas y parámetros de Lennard-Jonnes respectivamente. La descripción de las interacciones enlazantes en el estado de transición y reactante son modeladas derivando nuevos parámetros enlazantes [sem]. La nomenclatura [aa] indica una reparametrización del aminoácido mutado. Los errores de valores de referencia QM/MM no fueron reportados	123
3.7.	Estimación del cambio en la energía libre de activación en mutantes de OYE	136

Resumen

La ingeniería de proteínas es una disciplina que busca diseñar, modificar y optimizar proteínas con funciones específicas. En este campo, la ingeniería de enzimas es una de las áreas más importantes debido a la relevancia de estos biocatalizadores en la biotecnología, la industria y la medicina. Experimentalmente, estas nuevas enzimas suelen ser obtenidas a través de la evolución dirigida, realizando múltiples ensayos de prueba y error. Debido al alto costo monetario y capital humano que implica realizar tales búsquedas aleatorias, es que nace el diseño racional de enzimas. Estos enfoques utilizan métodos computacionales basados en principios fisicoquímicos para postular variantes con propiedades específicas. Estas propiedades pueden clasificarse en termodinámicas y cinéticas, donde estas últimas definen la velocidad y el mecanismo por el cual ocurre una reacción, requiriendo información acerca de la ruptura y formación de enlaces. Debido a esto, los métodos computacionales utilizados deben considerar explícitamente los grados de libertad electrónicos, lo que limita las opciones a enfoques como cálculos de mecánica cuántica o simulaciones multiestados QM/MM. El alto costo computacional de estas técnicas limita el número de variantes a estudiar a sólo unos pocos ejemplares.

En esta tesis se desarrolló una nueva metodología basada en transformaciones alquímicas de no-equilibrio, para obtener el cambio en la energía libre de activación $\Delta\Delta G^\ddagger$, producto de mutaciones. Este nuevo método es al menos un orden de magnitud más rápido que las metodologías QM/MM actualmente establecidas, debido a que basa sus fundamentos en simulaciones de mecánica molecular.

En las transformaciones alquímicas es necesario describir las interacciones del estado de transición a través de la mecánica molecular. Para lograr esto, se propuso un esquema de derivación de parámetros enlazantes basado en la matriz Hessiana del estado de transición y una modificación del método de Seminario. Los términos no enlazantes por su parte fueron derivados utilizando la densidad electrónica y el esquema de partición MBIS, basado en la teoría de átomos-en-moléculas (AIM).

Las transformaciones alquímicas en conjunto con la derivación de campos de fuerza fueron validadas en dos sistemas enzimáticos: Crotonil-CoA Carboxilasa/Reductasa (CCR) y Dihidrofolato Reductasa (DHFR). Los resultados mostraron que las transformaciones alquímicas logran reproducir tanto valores experimentales, como sus tendencias, con errores cercanos a la precisión química (1kcal/mol). Además, se evaluó la modificación de los parámetros no enlazantes de aminoácidos, logrando una mejora de hasta 3 kJ/mol en DHFR.

Una vez validada la metodología se aplicó a una enzima diseñada artificialmente para catalizar una reacción aldólica: la variante de *Old Yellow Enzy-*

me(OYE) Y187F. El mecanismo de esta nueva reacción aldólica en OYE se determinó mediante simulaciones multiestado QM/MM, utilizando el método de la cuerda adaptativa. Los resultados mostraron que la coenzima de mononucleótido de flavina (FMN) transfiere un hidruro al sustrato 2-ciclohexenona, generando un enolato. Este intermediario a su vez ataca a una molécula de formaldehído, para formar la 2-metoxiciclohexenona posterior a su protonación. El camino de mínima energía libre obtenido por el método de la cuerda reveló que el paso limitante de la reacción era la transferencia de hidruro.

Identificado el paso limitante y caracterizado el estado de transición, se realizaron transformaciones alquímicas de no-equilibrio en OYE Y187F, para encontrar variantes catalíticamente más activas. Los resultados predigieron al menos 2 variantes que reducen la energía libre de activación, de las cuales, la doble mutante Y187F/T130W viene a ser la variante más prometedora con un $\Delta\Delta G^\ddagger$ de -8.8 kJ/mol.

En resumen, esta tesis logró demostrar que las transformaciones alquímicas de no-equilibrio, permiten predecir variaciones de la constante de velocidad de reacción catalítica de enzimas con una precisión similar a la de métodos basados en primeros principios, pero al menos, a un décimo de su costo computacional. Esto abre la posibilidad de realizar un análisis a gran escala, lo que facilitaría la búsqueda de nuevas variantes enzimáticas en el campo de la ingeniería de proteínas.

Abstract

Protein engineering is a discipline that designs, modifies, and optimizes proteins for specific functions. Within this field, enzyme engineering is one of the most important areas due to the relevance of these biocatalysts in biotechnology, industry, and medicine. Experimentally, new enzymes are typically obtained through directed evolution, performing multiple trial-and-error essays. Given the human and financial resources required for such random processes, rational enzyme design has emerged. This approach uses computational methods and physical principles to propose variants with specific properties. Since kinetic properties require knowledge of the reaction mechanism, computational methods are limited to first-principles-based approaches, such as quantum mechanics (QM) calculations or multi-state QM/MM simulations. However, their high computational cost restricts the number of variants to be studied to only a few.

This thesis proposes a new methodology based on non-equilibrium alchemical transformations to obtain the change in activation free energy, $\Delta\Delta G^\ddagger$, induced by mutations. This new method is at least one order of magnitude faster than the currently well-established QM/MM methodologies, as it relies on purely classical molecular mechanics.

The alchemical transformations involve describing molecular interactions of transition states with molecular mechanics. To achieve this, a bonding parameter derivation scheme is proposed based on the Hessian matrix of the transition state and the modified Seminario method. Non-bonded interaction parameters are derived using the electron density of substrates and the MBIS partitioning scheme, based on the Atoms-in-Molecules (AIM) theory.

The alchemical transformations and force field parameterization were validated in two enzymatic systems: Crotonyl-CoA Carboxylase/Reductase (CCR) and Dihydrofolate Reductase (DHFR). The results show that the alchemical transformations successfully reproduce both experimental values and trends, with errors close to chemical accuracy (1 kcal/mol). Additionally, modifications to the non-bonded parameters of mutating amino acids in the alchemical transformations are evaluated, improving predictions by up to 3 kJ/mol in the DHFR system.

Once the methodology was validated, it was applied to an artificially designed enzyme: the Y187F variant of Old Yellow Enzyme (OYE). Since this artificial variant catalyzes a new-to-nature aldol reaction in OYE, multi-state QM/MM simulations were performed using the adaptive string method to disclose the reaction mechanism. The results show that a flavin mononucleotide (FMN) coenzyme transfers a hydride to 2-cyclohexenone, generating an enolate. This intermediate subsequently attacks formaldehyde molecule leading to 2-methoxycyclohexenone after protonation. The minimum free

energy path obtained using the string method indicates that the rate-limiting step of the reaction is the hydride transfer.

After identifying the rate-limiting step and characterizing the transition state, non-equilibrium alchemical transformations were performed on OYE Y187F to find catalytically more active variants. The results predict at least two variants that reduce the activation free energy. The Y187F/T130W double mutant is the most promising candidate, achieving a $\Delta\Delta G^\ddagger$ of -8.8 kJ/mol.

In summary, this thesis demonstrates that non-equilibrium alchemical transformations predict enzyme kinetic properties with a precision comparable to first-principles-based methods but at one-tenth of the computational cost. This breakthrough enables high-throughput analyses, facilitating the discovery of new enzymatic variants in the field of protein engineering.

1. Introducción

1.1. Ingeniería de proteínas

La ingeniería de proteínas se entiende como el campo de estudio para la modificación y/o creación de proteínas con un interés específico y una gran aplicación biotecnológica^{1,2}. Dentro de la ingeniería de proteínas, las enzimas han sido el tipo de proteínas más ampliamente estudiadas dadas sus atractivas características como biocatalizadores^{3,4}. Formalmente las enzimas corresponden a catalizadores que aceleran reacciones químicas llevando a los reactantes por un camino alternativo al mecanismo que realizan en solución⁵. Ubicado en el límite entre la heterocatálisis y la catálisis homogénea, la catálisis enzimática posee varias ventajas como la alta selectividad, condiciones de reacción amigables y reacciones del tipo química verde⁶. Sin embargo, la aplicación de tales biocatalizadores a la industria conlleva una serie de desafíos relacionados a las dificultades que pueden poseer, entre las cuales pueden encontrarse su estabilidad térmica, baja actividad catalítica, enantioselectividad reducida, entre otras.

Tales problemáticas pueden ser abordadas desde dos enfoques, la bús-

queda de variantes enzimáticas expresadas en organismos provenientes de diversos ambientes, tales como bacterias extremófilas⁷, o bien, el diseño racional de enzimas^{8,9}, en el cual estas biomoléculas son transformadas a nivel molecular, dando paso así a la ingeniería de proteínas.

El campo de la ingeniería de proteínas actualmente involucra un trabajo multidisciplinar, donde diversas áreas del conocimiento se unen con el fin de generar nuevas biomoléculas. En el marco del diseño de enzimas, existen principalmente dos grandes enfoques que suelen aplicarse en conjunto para obtener mejores resultados: el diseño experimental¹⁰ y el diseño computacional⁸, donde cabe destacar que ambos enfoques han sido galardonados con el premio Nobel de química en los años 2018 y 2024, para Arnold¹¹ y Baker¹² respectivamente.

1.2. Metodologías experimentales en la ingeniería de proteínas

Comenzando con las metodologías experimentales, resulta imposible hablar de ingeniería de proteínas sin mencionar a la evolución dirigida¹³. Este método experimental simula la evolución Darwiniana, generando múltiples variantes enzimáticas que se expresan en microorganismos, se aíslan, purifican y evalúan para comparar sus características con la enzima original. Al

igual que en los procesos evolutivos naturales, las variantes que poseen las características deseadas son seleccionadas, mientras que el resto de variantes fallidas son descartadas o bien sujetas a una nueva tanda de mutaciones aleatorias. La idea original de evolución dirigida consiste en aplicar mutaciones aleatorias a la enzima de interés, a través de distintas herramientas de la biología molecular, tales como la reacción en cadena de la polimerasa sujeta a error^{14,15} (*error-prone PCR*), recombinación de proteínas por secuencia homóloga¹⁶, recombinación de partidores aleatorio¹⁷, entre otras. Esta aleatoriedad impuesta permite obtener una biblioteca de variantes muy diversas. Sin embargo, la producción y posterior análisis de tales variantes requiere gran cantidad de trabajo, sumado a altos costos de ejecución¹⁸. Para sobrellevar la problemática anteriormente expuesta, los diseños semi-rationales han ganado importancia^{19,20}, dado que permiten reducir el número de candidatos considerablemente. Su formulación original consiste en utilizar herramientas de la bioinformática para hacer alineamientos múltiples de secuencia desde donde se reconocen regiones conservadas y no conservadas^{21,22}. Con estas regiones identificadas, se suelen aplicar mutaciones sitio dirigidas a aquellos residuos potencialmente responsables de las características a mejorar, reduciendo el abanico de posibilidades. Hoy en día, el diseño semi-razional también incluye información estructural de la enzima en sí^{23–26}, permitiendo identificar *hotspots* de manera espacial, los cuales suelen aportar información sobre los dominios catalíticos, estructural, regulador, entre otros²⁷. En térmi-

nos concretos se puede decir que el enfoque semi-racional actual, suele ser una combinación de técnicas experimentales y metodologías computacionales apuntadas a optimizar el proceso de ingeniería proteica. Un ejemplo de metodología computacional es el *Docking molecular*^{28,29}, que permite predecir la interacción entre una enzima y su sustrato o inhibidor, determinando sitios de unión, afinidades y posibles conformaciones del complejo³⁰.

1.3. Diseño asistido por computadora

Aunque el diseño semi-racional de proteínas permite reducir el espacio de secuencias a analizar, aún está sujeto a múltiples rondas de mutaciones aleatorias que comprenden etapas de ensayo y error³¹. Este número puede incrementar exponencialmente cuando se quiere explorar un sitio en específico. A modo de ejemplo, si se considera un sitio catalítico compuesto por 8 residuos clave, el número de combinaciones disponibles supera las 10^{10} secuencias distintas. Claramente el número de combinaciones con las que se debe tratar en el ejemplo anterior vuelven impráctico el proceso.

En este contexto, nace un nuevo concepto llamado "diseño asistido por computadora"(CAD, del inglés *Computer-Aided design*), refiriéndonos a la versión que implica estudios netamente *in silico*³². Este enfoque se basa en la utilización de métodos computacionales para poder predecir las propiedades de las proteínas a través de fundamentos estadísticos^{23,33} y/o fisicoquí-

nicos^{34,35}. La constante evolución de las arquitecturas computacionales y la creciente disponibilidad de tales recursos, han hecho de tales herramientas computacionales una de las aristas principales en la ingeniería de proteínas. A modo de ejemplo, desde la primera proteína diseñada completamente desde un enfoque computacional (Top7) en 2003¹², hasta la fecha, decenas de nuevas proteínas diseñadas racionalmente han sido publicadas^{36–43}. Producto de esto, en 2024 David Baker, compartió el premio Nobel de química por el diseño computacional de proteínas, en compañía de Demis Hassabis y John Jumper por la creación de Alpha Fold^{23,44}: una herramienta de inteligencia artificial que predice el plegamiento de las proteínas a partir de su secuencia. Estos hechos demuestran el creciente interés por el desarrollo de herramientas computacionales cada vez más precisas y rápidas para poder estudiar y diseñar proteínas. Esta necesidad se ve justificada por su amplia utilidad en áreas como la medicina⁴⁵, industria alimentaria⁴⁶ y biorremediación⁴⁷.

Actualmente, la cantidad de métodos computacionales utilizados en ingeniería de proteínas es considerable y continúa en expansión⁴⁸. Una forma de poder clasificar los diferentes métodos es basado en los principios que los rigen, separándose en estadísticos, basados en principios fisicoquímicos o híbridos.

1.3.1. métodos computacionales basados en principios estadísticos

Los enfoques estadísticos, utilizan bases de datos experimentales para poder extraer información relevante de los sistemas a estudiar. El análisis filogenético mediante reconstrucción ancestral por ejemplo, se basa en la utilización de datos experimentales generalmente de secuencia, para identificar variantes relevantes de una proteína⁴⁹. Comparando secuencias aminoacídicas con homólogos evolutivos para detectar residuos altamente conservados, se asocian tales secuencias con la estabilidad estructural o la actividad catalítica. De la misma forma, los sitios menos conservados se consideran más tolerantes a mutaciones, siendo los principales objetivos a mutar y evaluar⁵⁰. A pesar de que esta metodología permite diseñar variantes estables mediante mutagénesis sitio-dirigida, no garantizan mejoras en la eficiencia catalítica ni en la estabilidad de la proteína y restringen las mutaciones a aquellas observadas en la naturaleza. A pesar de estas limitaciones, su bajo costo computacional junto a su facilidad de automatización, los convierten en herramientas fundamentales. Algunos ejemplos de métodos estadísticos se encuentran el análisis múltiple de secuencias²², los modelos ocultos de Markov⁵¹ y el análisis de componentes principales⁵².

1.3.2. métodos computacionales híbridos

El segundo tipo de método computacional recibe el nombre de híbrido, esto por poseer características tanto de los métodos estadísticos como también de los basados en principios fisicoquímicos. Estas metodologías ya no solo consideran la secuencia aminoacídica de la proteína de interés si no que también toman en cuenta la estructura, las interacciones y por ende la energía del sistema⁵³. Debido a lo anteriormente mencionado, criterios como el número de puentes de hidrógeno, porcentaje de estructura secundaria o energía libre, son útiles al momento de seleccionar mutaciones con potencial utilidad. Un ejemplo de estas metodologías es Rosetta³⁴, un paquete con múltiples funciones de la que destaca la computación de energías libres relativas de estabilidad. Basado en campos de fuerza estadísticos, utilizando parámetros empíricos obtenidos de diversos estudios⁵⁴, Rosetta permite predecir el cambio en la estabilidad térmica de una proteína sujeta a alguna mutación. Otras metodologías híbridas altamente utilizadas corresponden a las técnicas PBSA (del inglés Poisson-Boltzman surface area) y GBSA (del inglés Generalized Born surface area), los cuales a pesar de poder extraerse de simulaciones basadas en principios fisicoquímicos, usualmente los análisis son realizados utilizando modelos de regresión para disminuir el costo computacional⁵⁵. Los ejemplos mencionados poseen un costo computacional relativamente bajo, sin embargo, los resultados son imprecisos y dependen del sistema particular.

Un importante hecho respecto a estas metodologías es la casi nula capacidad de predecir propiedades cinéticas de las enzimas. Esto radica en el hecho que muchos de los parámetros que utilizan estas herramientas derivan de información experimental y/o computacional de bajo costo y a menos que se tengan datos tanto de reactantes y estado de transición, no es posible obtener información acerca de parámetros cinéticos de la enzima a estudiar, obteniendo conclusiones acerca de la estabilidad, afinidad, solubilidad u otras propiedades termodinámicas.

Lo anterior toma especial relevancia en el marco de este trabajo en el cual justamente son las propiedades cinéticas de las enzimas lo que se busca predecir. Programas tales como FUNCLIB hacen uso de métodos híbridos para seleccionar candidatas que sean estables o posean un correcto plegamiento, para luego aplicar un segundo filtro utilizando métodos basados en principios fisicoquímicos para obtener enzimas catalíticamente superiores. Sin embargo FUNCLIB⁵³ hace uso de la teoría *Empirical Valence Bond*⁵⁶ (EVB) la cual necesita una parametrización específica para cada reacción, lo que requiere de gran experiencia y conocimiento por parte del usuario.

1.3.3. métodos computacionales basados en principios fisicoquímicos

A modo de suplir las falencias descritas con anterioridad, nace el tercer grupo de métodos computacionales, el cual se basa en principios fisicoquími-

cos. Utilizando ecuaciones derivadas de modelos de interacciones moleculares, estas metodologías permiten obtener observables fisicoquímicas, tanto de enzimas silvestres como también de variantes enzimáticas⁵³. Entre las herramientas más utilizadas en este enfoque se encuentran las simulaciones moleculares, que incluyen a las simulaciones de dinámica molecular⁵⁷ y simulaciones de tipo Monte Carlo⁵⁸. Estas simulaciones a su vez pueden dividirse de acuerdo a la precisión con la cual tratan sus interacciones moleculares, donde se encuentran de mayor a menor precisión: la mecánica cuántica⁵⁹, mecánica molecular⁶⁰ y grano grueso⁶¹ respectivamente. En las metodologías más avanzadas se encuentran las simulaciones multiestados, donde más de un modelo de interacción es incluido en el sistema⁶².

Si bien todos los métodos anteriormente descritos permiten en un principio obtener propiedades termodinámicas de las enzimas, sólo un reducido número de técnicas permiten la obtención de propiedades cinéticas. Aquellos métodos con tales atributos deberían en principio ser capaces de capturar la información asociada a la transformación desde reactantes hasta el estado de transición incluyendo los cambios conformacionales y energéticos asociados⁶³.

1.3.4. métodos computacionales basados en mecánica cuántica

Un ejemplo de metodologías que cumplen con los propósitos de la sección anterior corresponde a los métodos basados en mecánica cuántica. Estos métodos permiten la obtención de perfiles de reacción catalizados por enzimas silvestres y sus variantes⁶⁴. A partir de la teoría del estado de transición, es posible determinar el perfil energético que describe el paso de reactantes a productos a lo largo de una coordenada de reacción, generalmente asociada a la formación y ruptura de enlaces⁶⁵. Para ello, una estructura inicial es requerida, correspondiente a la enzima con su sustrato formando el complejo de Michaelis-Menten (CMM). Además, es necesario conocer al menos la estructura del producto de la reacción, para poder interpolar los cambios estructurales del reactante hasta los productos. Entre las metodologías más empleadas para obtener este perfil energético se encuentran las simulaciones de multiestado mecánica cuántica/mecánica molecular (QM/MM)⁶², en las cuales, una parte de la enzima se modela mediante mecánica clásica, incluyendo moléculas de agua e iones para simular su entorno, mientras que el sitio activo y los sustratos son tratados mediante cálculos de mecánica cuántica. Esto permite seguir la evolución temporal de la simulación y obtener un colectivo de configuraciones. Haciendo esto para el estado de reactante, producto y un número finito de estados intermedios es posible obtener un perfil

de energía libre a lo largo de una coordenada de reacción, donde la barrera energética más baja entre dos estados estacionarios corresponde a la energía libre de activación, la cual está relacionada directamente a la constante de velocidad de la enzima dada la ecuación de Eyring⁶⁵.

Realizando el procedimiento múltiples veces para cada una de las mutantes, la comparación directa de los perfiles de reacción permiten inferir si una mutación hace más rápida o lenta a la enzima⁶⁶. La aplicación de las técnicas anteriormente señaladas implica que existe una coordenada de reacción que conecta reactantes con productos y que esta coordenada de reacción es conocida⁶⁷. Esta suposición comprende uno de los mayores desafíos de las metodologías QM/MM donde al igual que en EVB se requiere un conocimiento avanzado en el modelamiento de reacciones enzimáticas. Sin embargo, una vez que la coordenada de reacción es definida, la exploración del espacio de fase a lo largo de esta coordenada puede realizarse utilizando técnicas de muestreo sesgado (*biased sampling*), como metadinámica⁶⁸, *umbrella sampling*⁶⁹ y *steered molecular dynamics*⁷⁰, que vienen a guiar el camino modelado por el potencial QM/MM.

La principal desventaja de los métodos basados en primeros principios, los cuales proporcionan los resultados más precisos, es que representan las herramientas con el mayor costo computacional. Lo anterior limita a estas técnicas a la evaluación de un reducido número de variantes enzimáticas, aún en sus versiones más optimizadas, volviéndose inviable su aplicación

a una exploración más exhaustiva, limitando así, su utilidad como reemplazo al trabajo experimental⁷¹.

1.3.5. Transformaciones alquímicas de no-equilibrio como potencial metodología para la evaluación de variantes enzimáticas

Como alternativa a la anterior problemática, en esta tesis se propone el uso de transformaciones alquímicas basadas en simulaciones de dinámica molecular clásica⁷², como herramienta para predecir los cambios en la energía libre de activación producto de mutaciones. En principio no es necesario conocer el perfil de energía libre de la reacción, si no la diferencia de energía libre entre el estado de transición y el complejo de Michaelis-Menten (para reacciones de un solo paso elemental). Las transformaciones hacen uso de las propiedades de función de estado de la energía libre y obtienen la diferencia entre las energías libres de activación de la enzima silvestre y una variante, como la diferencia de mutar en la simulación a la enzima en el estado de transición y el complejo de Michaelis-menten (CMM) respectivamente (Figura 1.1). Debido a que mutar una enzima a lo largo de una simulación implica transformar elementos químicos, esta técnica recibe el nombre de transformación alquímica. La razón de llevar el proceso por este camino alternativo es netamente computacional. Los campos de fuerza que modelan las interac-

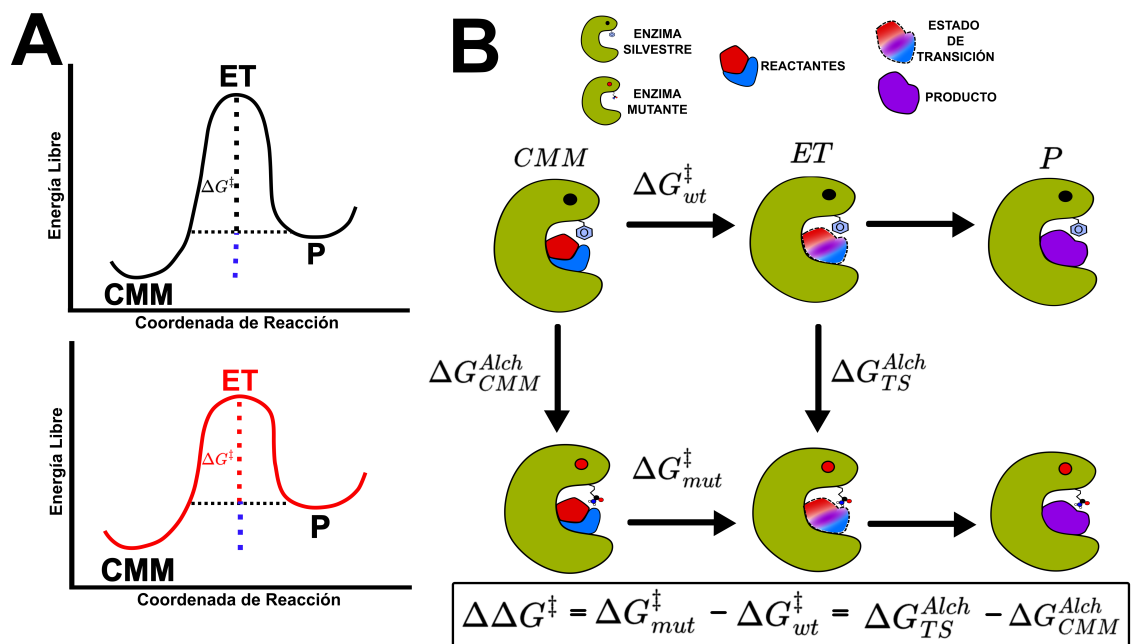


Figura 1.1: A) Perfiles de reacción para una enzima silvestre y una variante catalíticamente más activa. B) Ciclo termodinámico para la estimación de la energía libre de activación de una reacción catalizada por una enzima silvestre (primera fila) o mutante (segunda fila). El ciclo se completa por medio de transformaciones alquímicas en el complejo de Michaelis-Menten (CMM) y en el estado de transición (ET).

ciones en las simulaciones clásicas no permiten representar la formación ni ruptura de enlaces, por lo que simular la reacción química mediante simulaciones clásicas es inviable. Sin embargo las transformaciones de residuos aminoacídicos en simulaciones MD están bien establecidas⁷³⁻⁷⁵. Adicionalmente la transformación de un residuo en el estado de transición y reactante corresponde a una perturbación menor a la transformación desde el CMM al ET. En este contexto, las transformaciones alquímicas de no-equilibrio, aclamadas por ser los métodos alquímicos más rápidos⁷⁶ tienen cabida, permitiendo así el análisis de múltiples variantes a un bajo costo computacional. El principal desafío en esta metodología radica en la generación de un campo de fuerza para el estado de transición. En esta tesis se desarrollan y aplican métodos basados en mecánica cuántica con el fin de generar parámetros de campos de fuerza tanto para el estado de transición como para el complejo de Michaelis-Menten.

Esta metodología asume que el paso limitante de la reacción es conocido, desde donde se extrae el estado de transición para su posterior parametrización. Además, se asume que la estructura del estado de transición no cambia con la mutación de un aminoácido. Una vez definidos los campos de fuerza, simulaciones de dinámica molecular clásica son realizados para cada una de las variantes enzimáticas, seguidas por rápidas transformaciones alquímicas de no-equilibrio que permiten obtener el cambio en la energía libre de activación. Para validar esta nueva metodología se utilizaron dos enzimas. La pri-

mera enzima corresponde a la Crotonil-CoA Carboxilasa/Reductasa (CCR), responsable de fijar CO_2 en diversos organismos, presentando la constante catalítica más alta en comparación a otras enzimas fijadoras de CO_2 ⁷⁷. Las transformaciones alquímicas son aplicadas a CCR considerando cuatro mutantes que poseen valores de energía libre experimentales reportados⁷⁸. Siendo CCR un tetrámero que posee mutaciones en residuos de distinta naturaleza, tales como cargados, hidrofílicos, polares y aromáticos. La predicción de los cambios en la energía libre producto de tales mutaciones implica un desafío para la técnica a implementar. Como segundo método de validación se utiliza a la enzima Dihydrofolato Reductasa (DHFR), la que al igual que CCR posee energías libres de activación reportadas experimentalmente⁶⁶. DHFR es escogida como segundo sistema de validación debido a que es una enzima monomérica con un sitio activo definido, lo que nos permite realizar una evaluación directa del campo de fuerza, eliminando todo efecto producido por una insuficiente exploración del espacio conformacional. En adición, las mutaciones a evaluar en DHFR comprenden cambios en el volumen del aminoácido, lo que permite evaluar el modelo de interacciones débiles tales como la dispersión de London y la Repulsión de Pauli. Dado que en ocasiones estas interacciones pueden quedar enmascaradas por fuertes interacciones no covalentes, como las electrostáticas, el sistema DHFR permite realizar un análisis sistemático.

Luego de ser validada la metodología en CCR y DHFR, se aplica en un

nuevo sistema, la *Old Yellow Enzyme*(OYE). Esta enzima ha sido ampliamente estudiada debido a su promiscuidad por una amplia variedad de compuestos α,β -insaturados, produciendo en algunos casos una preferencia enantiomérica^{79,80}. Particularmente en este trabajo se busca encontrar una variante de la mutación Y187F la cual fué diseñada racionalmente por colaboradores del *Max Planck Institute for terrestrial microbiology*, con el fin de que pudiese catalizar una reacción aldólica con formaldehído, reacción que naturalmente OYE no realizaba. Análisis estadísticos o semiempíricos son inviables en este caso debido a la inexistencia de datos respecto a esta nueva reacción, haciendo de las transformaciones alquímicas la herramienta idónea para encontrar nuevas variantes. Con el objetivo de conocer el perfil de energía libre de la reacción aldólica de OYE, se utiliza el método de la cuerda adaptativa⁸¹, permitiendo identificar el primer paso como el limitante de la reacción y caracterizando al estado de transición(ET). Con esta información, campos de fuerza para el CMM y ET pueden ser derivados, lo que permite aplicar transformaciones alquímicas para encontrar variantes enzimáticas con propiedades cinéticas mejoradas.

En los siguientes capítulos, se presentan los resultados obtenidos a partir de la implementación y validación de las transformaciones alquímicas de no equilibrio en el estudio de sistemas enzimáticos. Primero, se describe la metodología utilizada y los criterios de validación empleados para evaluar su precisión. Luego, se analizan los resultados en términos de la predicción

de cambios en la energía libre de activación debido a mutaciones, comparándolos con datos experimentales cuando están disponibles. Finalmente, se discute el potencial de esta metodología como una herramienta eficiente para la ingeniería de proteínas, destacando su aplicabilidad en el diseño racional de enzimas con mayor eficiencia y estabilidad térmica.

1.4. Hipótesis y Objetivos

Hipótesis

Cálculos de energía libre basados en transformaciones alquímicas de no equilibrio junto con una mejor descripción de las interacciones no covalentes en el sitio activo, predicen la diferencia de energía libre de activación producto de mutaciones en las enzimas CCR y OYE.

Objetivo general

Establecer un flujo de trabajo computacional para predecir diferencias en la energía libre de activación sobre variantes enzimáticas en los sistemas CCR y OYE, basado en simulaciones alquímicas de no-equilibrio y una mejor descripción de las interacciones no covalentes en el sitio activo.

Objetivos Específicos

1. Validar las transformaciones alquímicas basadas en simulaciones de no equilibrio, como metodología para obtener la variación de energía libre de activación, utilizando las variantes N81L, F170A, E171A y H365N de la enzima Crotonil-CoA Carboxilasa/Reductasa (CCR) de *K. setae*, las cuales poseen energías libres de activación calculadas con métodos QM/MM y validadas experimentalmente.

2. Derivar parámetros de campos de fuerza para las interacciones no covalentes de las cadenas laterales de los aminoácidos, sustrato y coenzima, a partir de la densidad electrónica en el ambiente proteico, con el fin de mejorar la descripción de las interacciones en el sitio activo y evaluar su efecto en cálculos de la variación de energía libre de activación sobre el sistema CCR.
3. Calcular el perfil de energía libre de reacción del sistema OYE Y187F mediante simulaciones QM/MM, con el fin de conocer el estado de transición de la reacción.
4. Con la metodología establecida en los objetivos anteriores, identificar variantes enzimáticas de la mutante Y187F del sistema old yellow enzyme (OYE), que posean una actividad aumentada, mediante transformaciones alquímicas basadas en simulaciones de no equilibrio.

2. Metodología

Esta sección contempla la descripción de las metodologías usadas y desarrolladas durante la investigación. Un desarrollo general de los fundamentos de la mecánica cuántica y molecular son descritos, para posteriormente ser utilizados en la definición de los métodos híbridos de mecánica cuántica/mecánica molecular (QM/MM). Se introduce el concepto de campo de fuerza para describir las interacciones en el marco de la mecánica clásica y como sus denominados parámetros son derivados. Posteriormente se describen métodos para la obtención de la energía libre mediante transformaciones alquímicas, en conjunto a una breve descripción de los estimadores. Se hace especial énfasis en las transformaciones alquímicas de no-equilibrio debido a que corresponde a la idea principal del trabajo. Finalmente, se presentan dos métodos para estimar la energía libre de activación: el método de la cuerda adaptativa y la propuesta en esta tesis basada en transformaciones alquímicas de no equilibrio.

2.1. Mecánica cuántica

En el marco de la química cuántica, sistemas moleculares pueden ser descritos mediante la ecuación de Schrödinger independiente del tiempo, la cual nos entrega soluciones estacionarias de la forma:

$$\hat{H}\psi_i = E_i\psi_i \quad (2.1)$$

donde $\hat{H}$ corresponde al Hamiltoniano, mientras que ψ_i junto a E_i son la i -ésima función de onda y nivel de energía del sistema respectivamente. La función de onda y por ende el Hamiltoniano en esta ecuación dependen explícitamente de la posición de los núcleos y de los electrones, por lo que sólo tiene solución para sistemas modelos.

Una forma de sobrellevar este problema es invocando la aproximación de Born-Oppenheimer⁸². En este enfoque se asume que la masa de los electrones es despreciable en comparación a la masa de los núcleos, por lo tanto, el movimiento de ambas partículas ocurre en escalas de tiempo diferentes. Lo anterior nos permite desacoplar el movimiento de los electrones y núcleos, puesto que un pequeño cambio en la posición de los núcleos provocaría un reordenamiento electrónico instantáneo. Como última aproximación, si se consideran los núcleos como cargas puntuales de comportamiento clásico, la ecuación de Schrodinger debe resolverse sólo para los electrones, siendo

las coordenadas y cargas nucleares atómicas parámetros a definir. El enfoque anteriormente descrito construye la base de la mayoría de los métodos de química cuántica y todos los cálculos realizados en esta investigación asumen estas aproximaciones.

2.1.1. Método usados en química cuántica

Existen distintas metodologías para poder resolver el problema electrónico que se pueden agrupar en dos grupos: Cálculos *ab initio* donde se encuentran los métodos Hartree-Fock(HF) y pos-HF y en los cuales la ecuación de schrödinger electrónica es resuelta mediante el teorema variacional bajo ciertas aproximaciones, siendo el método HF el más simple⁸³. En la teoría de HF la función de onda se define como un determinante de Slater construido a partir de spin-orbitales moleculares χ para cada uno de los electrones:

$$|\psi\rangle_0 = |\chi_1, \chi_2, \dots, \chi_n\rangle \quad (2.2)$$

permitiendo que la función de onda sea antisimétrica al intercambio de dos electrones, lo que modela su naturaleza fermiónica. El Hamiltoniano a su vez, es modelado por el operador de Fock que corresponde a una suma de operadores monoelectrónicos y donde la interacción que experimenta un electrón por el resto es aproximada como un campo medio a través del potencial de

Hartree-Fock v^{HF} :

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{1A}} + v^{HF}(i). \quad (2.3)$$

Dependiendo de las aproximaciones aplicadas en el Hamiltoniano y/o la función de onda, estas dan origen a los diferentes métodos pos-HF.

El segundo grupo corresponde a los cálculos basados en la Teoría del funcional de la densidad (TFD) o *DFT* por sus siglas en inglés⁸⁴. Esta teoría se basa en el principio de que el problema electrónico puede resolverse por medio de un funcional (una función que toma otra función como argumento y devuelve un valor numérico) de la densidad electrónica. De la ecuación 2.2 se puede extraer que la energía total del sistema es un funcional del Hamiltoniano, que a su vez es función del número de electrones N , las posiciones de los núcleos R_A y las cargas nucleares Z_A . Debido a que estas 3 variables pueden ser obtenidas de la densidad electrónica, se dice que la energía es un funcional de la densidad $\rho(\mathbf{r})$. Propuesto en un principio por Thomas y Fermi en 1927⁸⁵, pero sin demostrar la existencia de este funcional, no fue hasta 1964 donde se introduce formalmente por Hohenberg y Kohn, estableciendo dos teoremas claves para la DFT⁸⁶:

- 1. El potencial externo $v_{ext}(\mathbf{r})$ es considerado una constante y la densidad electrónica $\rho(\mathbf{r})$ determina de manera única el potencial $v(r)$ (salvo el potencial externo), por lo tanto, determinando completamente el es-

tado fundamental del sistema.

- 2. Existe un funcional de la densidad $F[\rho]$ tal que la energía del estado fundamental se obtiene minimizando este funcional: $E_0 = \min_{\rho \rightarrow N} F[\rho]$ bajo la restricción de normalización de la densidad v -representable: $\int \rho(\mathbf{r}) d\mathbf{r} = N$.

Debido a que la condición de v -representabilidad de la densidad electrónica es limitante para la mayoría de los sistemas de estudio, Mel Levy en 1979⁸⁷ estableció una reformulación de los teoremas de Hohenbeg y Kohn donde la densidad electrónica sólo necesita ser N -representable, lo que implica que esta densidad de probabilidad debe entregar el número de electrones al integrarse, ser mayor o igual a cero en todo el espacio y obtenida de una función de onda antisimétrica:

$$\int \rho(\mathbf{r}) d\mathbf{r} = N; \rho(\mathbf{r}) \geq 0 \quad (2.4)$$

evitando la estricta condición de que la densidad electrónica se deba definir considerando un potencial externo v definido. A pesar de que esta formulación prueba la existencia de un funcional que puede minimizar la energía electrónica utilizando una densidad N -representable, no explica cómo se puede obtener este funcional o cómo se debe describir la densidad electrónica. Sin embargo, en 1965 Kohn y Sham desarrollaron una ecuación análoga a la de Schrödinger para un gas no interactuante de electrones que entrega la mis-

ma densidad electrónica que un sistema molecular conformado por núcleos y electrones. Esta aproximación introduce el concepto de orbitales de Kohn y Sham(KS)⁸⁸, los cuales se utilizan para construir la densidad electrónica:

$$\rho(\mathbf{r}) = \sum_i |\phi_{KS,i}(\mathbf{r})|^2 \quad (2.5)$$

Mediante este concepto es posible obtener el estado fundamental de sistemas moleculares mediante una ecuación similar a la de HF, tanto en formulación como en costo computacional, pero tomando en cuenta efectos de correlación electrónica:

$$\left[-\frac{1}{2}\nabla_i^2 + v_{ext}(\mathbf{r}) + v_{ee}(\mathbf{r}) + v_x(\mathbf{r}) + v_c(\mathbf{r}) \right] \phi_{KS}^* \phi_{KS} = E \phi_{KS}^* \phi_{KS} \quad (2.6)$$

donde ∇^2 , v_{ext} , v_{ee} , v_x y v_c corresponden al operador de energía cinética y a potenciales de interacción núcleo-electrón, repulsión electrón-electrón, intercambio y correlación respectivamente. La Ecuación 2.6 es la forma general de la ecuación de Kohn y Sham asumiendo un gas homogéneo de electrones y que permite obtener la energía del estado fundamental mediante un cálculo de campo autoconsistente.

La forma de los funcionales en la ecuación 2.6 varía dependiendo del método DFT, e incluso muchos de los métodos ocupados en la actualidad no solo necesitan la densidad electrónica del sistema, sino que también utilizan el gradiente o incluso el laplaciano de la densidad de spin (asumiendo un

cálculo de una molécula con electrones desapareados y/o con spin-orbitales no restringidos a poseer el mismo valor)⁸⁹. Lo anterior repercute en la vasta cantidad de funcionales para realizar cálculos DFT, donde se clasifican en: aproximación de densidad local (LDA), aproximación de gradiente generalizado (GGA), meta-GGA, híbridos-GGA, meta-híbridos-GGA, aproximación de fase aleatoria (RPA), entre otros. A pesar de que existen ciertos parámetros para clasificar a estos métodos en cuanto a su calidad, debido a que la mayoría posee parámetros que son ajustados para reproducir bases de datos, hace de esto una tarea difícil, volviéndolo no-sistemático⁹⁰. La definición de los tipos de métodos en TFD van más allá del objetivo de este trabajo, por lo que solo se limita a mencionarlos.

Considerándose un tercer tipo de método, están los enfoques semi-empíricos. Estas metodologías hacen uso de aproximaciones y parámetros que son ajustados para reproducir propiedades moleculares. En un principio los métodos semiempíricos hacían alusión a enfoques basados en orbitales moleculares (OM), donde principalmente se realizaban aproximaciones para resolver las integrales interelectrónicas de sistemas moleculares. Entre sus ejemplos más conocidos se tienen a los métodos MINDO, AM1 o PM3⁹¹. Otro subgrupo de semi-empíricos corresponden a los derivados de la teoría del funcional de la densidad, dentro de los cuales se destaca el método *density-functional tight-binding (DFTB)*⁹², que se deriva de una expansión de segundo orden de la energía total de DFT con respecto a fluctuaciones de la densidad de carga,

en comparación a una densidad de referencia, que usualmente corresponde a una suma de densidades atómicas neutras $\rho_0 = \sum_{\alpha} \rho_0^{\alpha}$:

$$E = \langle \psi_i \hat{H} \psi_i \rangle - \frac{1}{2} \int \int' \frac{\rho_0 \rho_0'}{|\vec{r} - \vec{r}'|} + E_{XC}[\rho_0] - \int V_{XC}[\rho_0] + E_{ii} \\ + \frac{1}{2} \int \int' \left(\frac{1}{|\vec{r} - \vec{r}'|} + \frac{\delta^2 E_{XC}}{\delta \rho_0 \delta \rho_0'} \bigg|_{\rho_0} \right) \delta \rho \delta \rho' \quad (2.7)$$

En la ecuación 2.7, la fluctuación de la densidad de carga $\delta \rho = \rho - \rho_0$ se modela mediante una expansión de Taylor a primer orden, dando lugar a una fluctuación de cargas atómicas $\Delta q_{\alpha} = q_{\alpha} - q_{\alpha}^0$, que son obtenidas mediante el análisis de cargas de Mulliken⁹³. Tanto la doble integral sobre $\vec{r}$, como la segunda derivada del funcional de intercambio y correlación E_{XC} , son aproximados por una función γ que depende de la distancia entre los átomos y en los denominados parámetros de Hubbard U , que corresponden a la segunda derivada de la energía atómica con respecto a la carga total de dicho átomo. Lo anterior lleva a que la contribución a la energía a segundo orden se describa mediante un potencial interatómico de la forma: $E^{2do} = \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta}(U_{\alpha}, U_{\beta}, R_{\alpha\beta}) \Delta q_{\alpha} \Delta q_{\beta}$.

El primer termino de la ecuación 2.7 denominado la contribución a la energía proveniente de un hamiltoniano de orbitales atómicos, es un término de primer orden que depende en su totalidad de la densidad de referencia ρ_0 . Los orbitales de Kohn y Sham usados son obtenidos mediante una combinación lineal de orbitales atómicos(OM-CLOA), donde el set de funciones

bases a usar corresponde a una versión mínima y optimizada según los criterios de Eschrig y Bergert⁹⁴.

Una de las mayores ventajas de DFTB, es que los elementos de la matriz hamiltoniana y de solapamiento son calculadas para distancias interatómicas relevantes y tabuladas, disminuyendo el costo computacional (en conjunto con el uso de un set de bases mínimas) en 2 a 3 órdenes de magnitud con respecto a HF o DFT según previos estudios⁹². Los cuatro términos restantes de la ecuación 2.7 son agrupados y aproximados por un potencial interatómico V_{rep} similar en espíritu a un campo de fuerza clásico. Con todos los componentes definidos, la energía total bajo la aproximación DFTB queda de la forma:

$$E_{total} = \sum_{\mu\nu} c_{\mu} c_{\nu} \hat{H}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta}(U_{\alpha}, U_{\beta}, R_{\alpha\beta}) \Delta q_{\alpha} \Delta q_{\beta} + \sum_{\alpha\beta} V^{rep}(R_{\alpha\beta}). \quad (2.8)$$

En conclusión DFTB emplea parámetros obtenidos de cálculos DFT previos simplificando la descripción electrónica y permitiendo estudiar sistemas de más de 1000 átomos con gran precisión. Recientemente, una versión mejorada que incluye términos de tercer orden⁹⁵ (DFTB3) es popularmente utilizada para modelar sistemas biológicos debido a su gran precisión y bajo costo computacional.

2.2. Mecánica molecular

Como se discutió en la sección anterior, la descripción de sistemas moleculares mediante química cuántica ofrece una alternativa rigurosa para tratar moléculas o sistemas de hasta unos cientos de átomos. Sin embargo, sistemas biológicos conteniendo biomacromoléculas en adición a moléculas de solvente y iones, pueden llegar hasta el millón de partículas con 3 a 4 órdenes de magnitud mayor al tamaño límite de aquellas metodologías. Adicionalmente, los fenómenos de interés, pueden ser mediados por interacciones del tipo no covalentes, que son bien descritas por los campos de fuerza. A modo de ejemplo, están los cambios conformacionales de proteínas⁹⁶ o la unión de una droga a un receptor biológico⁹⁷, donde en la mayoría de los casos son gobernados por interacciones del tipo no-covalente. Tales interacciones han sido ampliamente descritas, ignorando los grados de libertad electrónicos y basados en potenciales interatómicos clásicos.

Estos potenciales comprenden la contribución no enlazante de los denominados campos de fuerza basados en mecánica clásica. En la siguiente sección se analizará la contribución no enlazante, seguido por la descripción de las interacciones covalentes. Finalmente, se explicarán los procesos de parametrización y construcción de los campos de fuerza.

2.2.1. Interacciones no-covalentes

En los campos de fuerza no polarizables, las interacciones no covalentes suelen dividirse en interacciones electrostáticas y de Van der Waals. Estas últimas se modelan a través de un potencial de Lennard-Jones, que incluye un término atractivo, debido a las fuerzas de dispersión y un término repulsivo, que engloba múltiples efectos tales como parte de la repulsión electrostática, la repulsión de Pauli y la contribución por intercambio. Matemáticamente, la energía total de las interacciones no enlazantes se expresa como una suma de potenciales interatómicos:

$$\begin{aligned} E_{nocovalente} &= E_{elec} + E_{vdW} \\ &= \sum_{B>A}^N \frac{q_A q_B}{r_{AB}} + \sum_{B>A}^N \frac{4\epsilon_{AB} \sigma_{AB}^{12}}{r_{AB}^{12}} - \sum_{B>A}^{N_{atoms}} \frac{4\epsilon_{AB} \sigma_{AB}^6}{r_{AB}^6} \end{aligned} \quad (2.9)$$

donde $r_{AB} = |\mathbf{R}_A - \mathbf{R}_B|$ representa la distancia entre los átomos A y B , q es la carga de cada átomo, ϵ es la profundidad del pozo del potencial de Lennard-Jones, y σ es la distancia a la cual el potencial se vuelve 0. Si bien se han propuesto modelos más avanzados, que incluyen efectos como inducción, penetración de carga y polarización, su uso en la práctica es limitado debido al alto costo computacional, la complejidad en su implementación y las dificultades en la parametrización. Por ello, los campos de fuerza clásicos

de primera generación han permanecido sin modificaciones importantes por décadas desde su formulación original.

2.2.2. Interacciones covalentes

El modelamiento de sistemas moleculares utilizando campos de fuerza debe poseer, además de las interacciones no-covalentes, una forma para describir las fuerzas que mantienen unida a una molécula. Tales interacciones se encuentran en el régimen donde las densidades electrónicas de los átomos se solapan, por lo tanto, se definen como fuerzas de corto alcance. En esta escala, los potenciales interatómicos descritos anteriormente fallan, por lo que se describen una serie de nuevas funciones que definen la parte enlazante del campo de fuerza:

$$\begin{aligned}
 E_{covalente} = & E_{enlace} + E_{ngulo} + E_{dihedros} + E_{Impropios} \\
 & \sum^{N_{enlaces}} k_r (r_{ij} - r_{ij}^{eq})^2 + \sum^{N_{angulos}} k_\theta (\theta_{ijk} - \theta_{ijk}^{eq})^2 \quad (2.10) \\
 & + \sum^{N_{dihedros}} k_\phi (1 + \cos[n\phi - \phi^{eq}]) + \sum^{N_{iimproprios}} k_\omega (\omega_{ijkl} - \omega_{ijkl}^{eq})^2
 \end{aligned}$$

El primer y segundo término comprende la energía asociada a la distancia de enlace y ángulo respectivamente, donde el superíndice *eq* indica los valores de equilibrio. Las variables *k* corresponden a constantes de fuerza basados en el modelo armónico. El tercer y cuarto término comprenden la contribución a la energía producto de la variación en los ángulos dihedros y dihedros

impropios. Un ángulo dihedro se define como el ángulo formado por 4 átomos enlazados. Los dihedros convencionales se describen mediante funciones de cosenos donde n es el factor de periodicidad. Por otro lado, los dihedros impropios se basan en un modelo armónico al igual que los ángulos de enlace. La contribución de todos estos componentes en adición a las interacciones no covalentes intramoleculares entre átomos que están a una distancia de 4 o más átomos vecinos conectados (interacciones del tipo 1-4), mantienen la geometría y en teoría deberían ser suficientes para poder reproducir los modos vibracionales normales de la molécula. La ecuación 2.10 comprende la forma general de los campos de fuerza enlazante para los modelos de primera generación. Algunos campos de fuerza biomoleculares incluyen contribuciones extras como es el caso de CHARMM, el cual considera una función adicional para modelar los ángulos de Urey-Bradley⁹⁸. Cabe destacar que las funciones descritas anteriormente corresponden a aproximaciones que impiden la ruptura y formación de enlaces, siendo esta la principal razón de la incapacidad de los campos de fuerza clásicos de primera generación para modelar reacciones químicas. Para abordar esta problemática se han propuesto múltiples enfoques en los cuales se destaca el remplazo de los potenciales armónicos por potenciales de Morse, sin embargo un análisis detallado de estas metodologías queda fuera del objetivo de este trabajo.

2.2.3. Parametrización de un campo de fuerza

En las secciones anteriores se describieron la forma en la cual los campos de fuerza modelan las interacciones covalentes y no-covalentes matemáticamente. Para poder generar tales campos de fuerza, se deben obtener las constantes que poseen cada uno de los términos en las ecuaciones 2.9 y 2.10. Estas constantes reciben el nombre de parámetros del campo de fuerza, los cuales deben ser asignados previo al inicio de una simulación.

Los parámetros de las interacciones no covalentes comprenden las cargas atómicas q y los términos ϵ y σ del potencial de Lennard Jones. Como se aprecia en la ecuación 2.9 estos términos dependen del átomo y si bien podrían ser asignados por elemento en un enfoque minimalista, se suele optar por asignar más de un set de parámetros por elemento para considerar los distintos ambientes químicos caracterizados por los átomos vecinos enlazados en una molécula. Para clarificar la frase anterior, considere el carbono carbonílico y el carbono terminal de un 6-butanal: Intuitivamente, se puede asumir que la carga atómica del carbono carbonílico debe ser positiva dada la sustracción electrónica por parte del oxígeno, mientras que la carga del carbono terminal debe ser ligeramente negativa. De la misma forma, las polarizabilidades y volúmenes de los átomos en una molécula pueden variar y por ende sus parámetros de Lennard-Jones también lo harán.

Las interacciones enlazantes por otro lado, también poseen parámetros,

sin embargo la diferencia radica en que dependen de 2 o más átomos. Las distancias y ángulos de equilibrio en conjunto con las constantes de fuerza comprenden el conjunto de parámetros enlazantes y al igual que en los términos no-enlazantes existe más de uno por elemento. De manera similar al ejemplo dado en el párrafo anterior, se analiza la constante de fuerza y la distancia de equilibrio de un enlace carbono-carbono en un etano y un etino. Existe evidencia espectroscópica que demuestra que el enlace triple en el etino es más corto que el enlace simple en el etano. De la misma forma se puede demostrar las diferencias en la constante de fuerza si se aplica la aproximación armónica a la vibración del enlace.

Anteriormente, se presentó a los campos de fuerza como una alternativa económica para simular sistemas de gran tamaño y complejidad química. Sin embargo, la información experimental disponible para determinar sus parámetros suele ser suficiente solo en casos modelo, como gases diatómicos o hidrocarburos de bajo peso molecular⁹⁹. Para abordar esta problemática, han surgido diversos métodos de parametrización, que pueden agruparse en dos categorías: métodos por ajuste y métodos basados en primeros principios.

Los métodos por ajuste utilizan datos experimentales de referencia, generalmente propiedades termodinámicas, para actualizar los parámetros del campo de fuerza mediante algoritmos de optimización. Esto permite reproducir propiedades deseadas con gran precisión, mientras se evita el sobreajuste a través de técnicas de regularización. Un ejemplo de este enfoque es

la parametrización de los términos de Van der Waals en campos de fuerza generalizados como CGENFF¹⁰⁰ o GAFF¹⁰¹, que emplean entalpías de vaporización y densidades de líquidos orgánicos como referencias. No obstante, una de las principales limitaciones de este método es su baja transferabilidad, ya que su precisión fuera del conjunto de referencia puede ser cuestionable.

En contraste, la parametrización basada en primeros principios se fundamenta en propiedades microscópicas, donde los parámetros se derivan directamente de modelos físicos sin recurrir a ajustes por optimización. Por ejemplo, los mismos parámetros de Van der Waals pueden obtenerse a partir de cálculos de polarizabilidad atómica, utilizando la versión dependiente del tiempo de la teoría del funcional de la densidad¹⁰². Dado que estos parámetros se derivan de principios físicos fundamentales, suelen ser más transferibles. Sin embargo, su exactitud es limitada, ya que los modelos físicos proporcionan soluciones exactas sólo para sistemas pequeños, como átomos o moléculas sencillas. Además, muchas de las propiedades atómicas necesarias para la parametrización, como cargas y polarizabilidades, carecen de una única definición cuando se aplican a moléculas más grandes.

Debido a que ambos enfoques tienen fortalezas y debilidades, la elección del método depende del problema a abordar. De hecho, muchos de los campos de fuerza más utilizados, como AMBER¹⁰³, CHARMM¹⁰⁴, OPLS¹⁰⁵ y GROMOS¹⁰⁶, combinan ambos enfoques para lograr un equilibrio entre precisión y transferibilidad.

Como se menciona en la parametrización por ajuste, los parámetros del campo de fuerza son validados mediante la evaluación de su desempeño al reproducir propiedades termodinámicas. Para lograr esto, se realizan simulaciones de dinámica molecular con los parámetros previamente establecidos y se evalúa la precisión de los resultados. Recientemente se ha comenzado a utilizar cálculos de mecánica cuántica como referencia para evaluar los parámetros del campo de fuerza¹⁰⁷. En las interacciones enlazantes por ejemplo, se busca que los modos de vibración normales de la molécula, obtenidos mediante mecánica molecular, puedan reproducir los modos normales cuánticos. De igual forma los parámetros no enlazantes pueden ser comparados con distintos componentes de un análisis de descomposición de la energía(EDA)¹⁰⁸.

Bajo la premisa del párrafo anterior, y siendo la teoría de interacciones moleculares más rigurosa, se tiene a la teoría de perturbaciones adaptada a la simetría¹⁰⁹ (SAPT) como uno de los métodos EDA más ampliamente utilizado. Este método sienta sus bases en la teoría perturbativa de Rayghley-Schrödinger, donde la interacción entre dos moléculas es obtenida a partir de la perturbación de sus operadores Hamiltoniano de tipo Hartree-Fock(F), denominados los operadores de fluctuación de Moller-pleset(W) en adición a un operador de Coulomb intermolecular(V):

$$E_{SAPT} = F_A + F_B + aW_A + bW_B + cV_{AB} \quad (2.11)$$

donde los subíndices A y B determinan en qué molécula actúan los operadores. el valor de los coeficientes a, b y c determinan el nivel SAPT utilizado, siendo el esquema más simple SAPT0 cuando sólo el término c aparece. Aún en su versión más aproximada, SAPT nos entrega distintas contribuciones a la energía de interacción no covalente las cuáles por simplicidad se agruparán en los siguientes términos:

$$E_{SAPT0} = E_{elec} + E_{disp} + E_{pauli} + E_{pol}. \quad (2.12)$$

Los cuatro términos corresponden a la contribución electrostática, de dispersión, de repulsión de Pauli y finalmente la energía de polarización. Evidentemente los primeros 3 términos constituyen valores que pueden ser contrastados con los términos electrostáticos y de vdw de un campo de fuerza, dando así una validación microscópica de la calidad de los parámetros derivados. Es importante destacar que a pesar de la rigurosidad que presenta SAPT, su alto costo computacional y la imposibilidad de considerar efectos de solvatación, dificultan su aplicación a sistemas de gran tamaño como biomoléculas.

Alternativamente a este tipo de EDA, enfoques basados en la teoría del funcional de la densidad pueden ser utilizados, dado su relativo bajo costo computacional. Recientemente se ha formulado la segunda generación de los métodos *Absolute Localized Molecular Orbital* EDA¹¹⁰ (ALMO-EDA2). Este método utiliza la energía *frozen* clásica, que corresponde a la diferencia

entre la función de onda simétrica construida por los orbitales moleculares de los monómeros aislados(formulación de Heitler-London) y los fragmentos aislados. La energía *frozen* agrupa las interacciones electrostáticas clásicas, de Pauli y de dispersión:

$$E_{frz} = E_{elec}^{cls} + E_{pauli}^{cls} + E_{disp} \quad (2.13)$$

donde la interacción electrostática corresponde a la interacción culómbica entre las distribuciones de carga de dos monómeros aislados:

$$E_{elec}^{cls} = \sum_{A < B} \int_{r_1} \int_{r_2} \rho_A^{tot}(r_1) \frac{1}{r_{12}} \rho_B^{tot}(r_2) dr_1 dr_2 \quad (2.14)$$

con ρ_A y ρ_B representando las densidades electrónicas totales de los monómeros A y B .

Las interacciones clásicas de Pauli y de dispersión se obtienen a partir de la porción restante de la energía *frozen*. Esto proporciona un marco de validación, dado que la mayoría de las implementaciones de campos de fuerza no consideran efectos de polarización en la derivación de cargas atómicas.

Para las interacciones de dispersión, el método ALMO-EDA proporciona un valor no ortogonal mediante la siguiente expresión:

$$E_{disp} = \left(E_{xc}[P_{frz}] - \sum_A E_{xc}[P_A] \right) - \left(E_{xc}^{DF}[P_{frz}] - \sum_A E_{xc}^{DF}[P_A] \right) \quad (2.15)$$

donde P_{frz} corresponde a la densidad congelada antisimetrizada del dímero, mientras que P_A es la densidad no antisimetrizada del monómero A. E_{xc}^{DF} representa un funcional auxiliar sin contribución de dispersión, que corresponde al operador de HF para funcionales híbridos y a revPBE para funcionales semilocales.

Usando la ecuación anterior, la interacción de dispersión puede calcularse explícitamente, mientras que la interacción clásica de Pauli se obtiene de manera implícita restando E_{disp} y E_{elec}^{cls} de E_{frz} :

$$E_{Pauli}^{cls} = E_{frz} - (E_{elec}^{cls} + E_{disp}). \quad (2.16)$$

Ambos métodos permiten obtener contribuciones energéticas para las interacciones no covalentes, siendo un referente al momento de evaluar la calidad de los parámetros derivados desde un punto microscópico, lo que da al campo de fuerza una mayor posibilidad de ser transferible a sistemas más allá del conjunto de ajuste o derivación.

2.3. Metodos de simulación multiescala mecánica cuántica/mecánica molecular (QM/MM)

En las secciones anteriores se presentaron distintas metodologías para modelar interacciones atómicas y moleculares basadas en mecánica cuántica.

ca y clásica. Cada uno de estos enfoques tiene sus propias ventajas y limitaciones, donde la elección del método a utilizar dependerá del sistema que se desea estudiar. Por ejemplo, para reacciones químicas, aunque existen algunos casos donde enfoques clásicos han sido empleados para modelar perfiles de reacción, la mayoría de estos procesos se estudian desde la perspectiva de la mecánica cuántica, o de manera más general, considerando explícitamente los grados de libertad electrónicos. Esto se debe a que la formación y ruptura de enlaces involucra la redistribución de la densidad electrónica, un fenómeno que no puede ser descrito con precisión mediante modelos puramente clásicos. Esto limita el tamaño del sistema que se puede simular a unos cientos o miles de átomos en el mejor de los casos, dificultando el tratamiento explícito de sistemas biolomoleculares como enzimas, ADN u otras especies en fase condensada. Una alternativa a tal problemática es utilizar potenciales de acoplamiento Mecánica Cuántica/Mecánica Molecular (QM/MM). Propuesto por primera vez por Karplus, Levitt y Warshel^{111,112}, esta metodología propone dividir el sistema en dos zonas: A) La zona QM donde los núcleos y electrones son tratados explícitamente permitiendo la ruptura/formación de enlaces químicos y por ende un tratamiento adecuado para reacciones. Esta zona involucra una reducida cantidad de átomos los cuáles por lo general son entendidos a ser participantes de la reacción a estudiar. El resto del sistema comprende la zona MM donde las interacciones son modeladas por campos de fuerza, reduciendo el costo computacional. La zona MM por lo general

tiende a ser mucho mayor que la zona QM, ya que suele incluir moléculas de solvente y iones. Si bien existen distintos enfoques QM/MM se limitará a solo describir el método aditivo, el cual describe un Hamiltoniano efectivo que opera sobre una función de onda dependiendo paramétricamente tanto de los núcleos de la zona QM como de la zona MM:

$$H_{eff}\psi(r_e, R_{MC}, R'_{MM}) = E_{eff}\psi(r_e, R_{MC}, R'_{MM}) \quad (2.17)$$

A partir de la ecuación 2.17 se desprende que el Hamiltoniano consta de tres contribuciones: la energía de la región QM, la energía de la región MM y la energía de acoplamiento entre ambas zonas, expresando la energía total de la siguiente manera:

$$E_{eff} = \langle \psi | H_{QM} + H_{QM/MM} | \psi \rangle + E_{MM}. \quad (2.18)$$

En la última década se han propuesto múltiples formas para definir el Hamiltoniano de acoplamiento QM/MM, lo que lo ha convertido en un área de investigación muy activa. Entre los modelos más empleados destaca el *electrostatic embedding*. En este enfoque, a los átomos de la región QM se les asignan parámetros de Lennard-Jones, mientras que la interacción electrostática incluye las cargas MM en el hamiltoniano electrónico. Esto permite que la densidad electrónica se vea polarizada por las cargas presentes en la región MM, lo que permite una representación precisa de las interacciones

entre átomos de zonas diferentes.

Este método ha demostrado ser exitoso en estudios de enzimas, moléculas de ADN y organocatalizadores¹¹³, entre otros, ya que posibilita la consideración explícita de los grados de libertad electrónicos y la influencia del entorno, aspectos cruciales para comprender los mecanismos de reacciones químicas.

2.4. Dinámica molecular

En la sección anterior se introdujo el concepto de campo de fuerza y cómo este permitía modelar las interacciones intra e intermoleculares con un costo computacional bajo, permitiendo simular sistemas en fases condensadas. Este modelo de interacción es utilizado en la dinámica molecular clásica para obtener una evolución temporal del sistema a una temperatura y presión dada. Entiendase la dinámica molecular como una técnica de simulación computacional que describe el movimiento de todos los átomos que conforman el sistema. Esta evolución temporal es generada al resolverse las ecuaciones de movimiento de Newton:

$$-\frac{d}{dr_i}V(\vec{r}_i, \vec{r}_j, \dots, \vec{r}_N) = m_i a_i^2 = m_i \frac{d^2}{dt^2} \vec{r}_i \quad (2.19)$$

donde la derivada del potencial sobre la posición del átomo i : $\vec{r}_i$, corresponde al negativo de la fuerza aplicado a dicho átomo de masa m_i de acuerdo a la segunda ley de Newton. La ecuación 2.19, posee una forma en la cual el término campo de fuerza resulta de manera natural para describir un potencial de interacción molecular. Si se aplica la ecuación a todos los átomos del sistema, se obtiene un conjunto de posiciones y momentos que generan un espacio vectorial $6N$ dimensional, siendo N el número de átomos del sistema. Este dominio recibe el nombre de espacio de fase y comprende la base de la formulación Lagrangiana y Hamiltoniana de la mecánica clásica. Las simulaciones de dinámica molecular generalmente poseen una conformación inicial que dictamina la posición de cada uno de los átomos. Los momentos por su parte son obtenidos de las velocidades iniciales que se asignan de manera aleatoria generando una distribución de Maxwell-Boltzmann a una temperatura definida. Conocidas las variables iniciales, las ecuaciones de movimiento se integran de forma numérica en función del tiempo $t_{inicial} + dt$ donde el tamaño de paso dt suele ser menor o igual a 2fs cuando se tratan sistemas con hidrógenos o bien gases para poder capturar correctamente sus modos vibracionales y mantener el sistema en un estado conservativo.

La resolución del anterior protocolo da lugar a un sistema donde el número de partículas N , el Volumen del sistema V y la energía E permanece constante. En el campo de la termodinámica estadística este sistema generado recibe el nombre de colectivo microcanónico. En la práctica, los procesos

termodinámicos de interés suelen ocurrir en condiciones de temperatura en lugar de energía constante, dando origen al colectivo canónico (NVT). Adicionalmente, tomando en consideración que los procesos biológicos suelen ocurrir a presión constante también es posible generar sistemas bajo tales condiciones definiendo el colectivo isotérmico-isobárico(NPT).

Para poder mantener la temperatura promedio constante, las simulaciones de dinámica molecular utilizan termostatos, el cual puede entenderse como un baño térmico que intercambia calor con el sistema. Dado que la temperatura T está relacionada directamente con las velocidades de las moléculas, los primeros termostatos corregían las desviaciones de temperatura aplicando penalizaciones a las velocidades atómicas. Este sencillo enfoque es la base del termostato de Berendsen¹¹⁴ el cual a pesar de no ser actualmente utilizado, dado que no logra generar el colectivo correcto, sirvió como inspiración para termostatos más rigurosos como el reescalado de velocidades¹¹⁵. De manera similar al termostato, el barostato de Berendsen mantiene la presión del sistema redimensionando el tamaño de la caja de simulación en conjunto con una re-proyección de las posiciones atómicas, manteniendo una presión promedio constante. Barostatos más elaborados como el Parinello-Rahman son de preferencia¹¹⁶. Sin embargo, una descripción detallada de ellos no será llevada a cabo en esta tesis.

La implementación del termostato y barostato en las simulaciones de dinámica molecular vuelven al sistema no determinista permitiendo aplicar el

formalismo de la termodinámica estadística y poder obtener propiedades macroscópicas tan valiosas como la energía libre.

2.5. Cálculos de energía libre mediante transformaciones alquímicas

Las transformaciones alquímicas corresponden a un tipo de simulación molecular utilizado para predecir valores de energía libre dado un proceso⁷². Su nombre radica del hecho que se simulan intermediarios no físicos, los cuales suelen ser una superposición de los estados de equilibrio para dicho proceso. Esto quiere decir que el sistema en el estado A o parte de él, es transformado hasta el sistema B, mediante cambios en el potencial que gobiernan sus interacciones. Como ejemplo, se puede señalar el caso en que se tiene un péptido que posee un residuo de tirosina y se quiere ver el efecto en la energía libre de estabilidad al mutar este residuo por un residuo de fenilalanina. En este caso, se puede simular el péptido utilizando mecánica molecular y cambiando los parámetros del campo de fuerza para los átomos involucrados en la mutación. Lo anterior generará múltiples estados intermediarios entre la tirosina (*T*) y fenilalanina (*F*). Asumiendo una interpolación lineal entre

ambos estados, se puede escribir el Hamiltoniano del sistema como:

$$H(\vec{r}, \lambda) = (1 - \lambda)H_T(\vec{r}) + \lambda H_F(\vec{r}) \quad (2.20)$$

donde se introduce la variable λ que controla el peso de los estados A y B. De la ecuación 2.20, se puede extraer que λ , oscila en el rango de 0 a 1, donde los límites corresponden a los estados que describen puramente a tirosina y fenilalanina respectivamente¹¹⁷.

Debido a que la energía libre es una función de estado, es importante señalar que no existe un único camino para la transformación de una molécula o sistema en otro. Si bien todas dependen del parámetro λ que va desde 0 para el estado inicial hasta 1 para el estado final siendo los intermedios especies “quiméricas” que poseen parámetros de interacción de ambos estados. El modo en el que cambian las interacciones puede variar dependiendo de la metodología a utilizar.

Independientemente de la metodología, siempre es necesario realizar una simulación de equilibración o dinámica molecular clásica. El objetivo de esto, es poder explorar el espacio de fase con el fin de encontrar conformaciones representativas no correlacionadas, las cuales luego serán sometidas a transformaciones alquímicas para obtener la energía libre, por lo que la precisión de aquella observable es implícitamente dependiente de una exploración del espacio de fase completa. Dicho esto y considerando que se tiene un set de

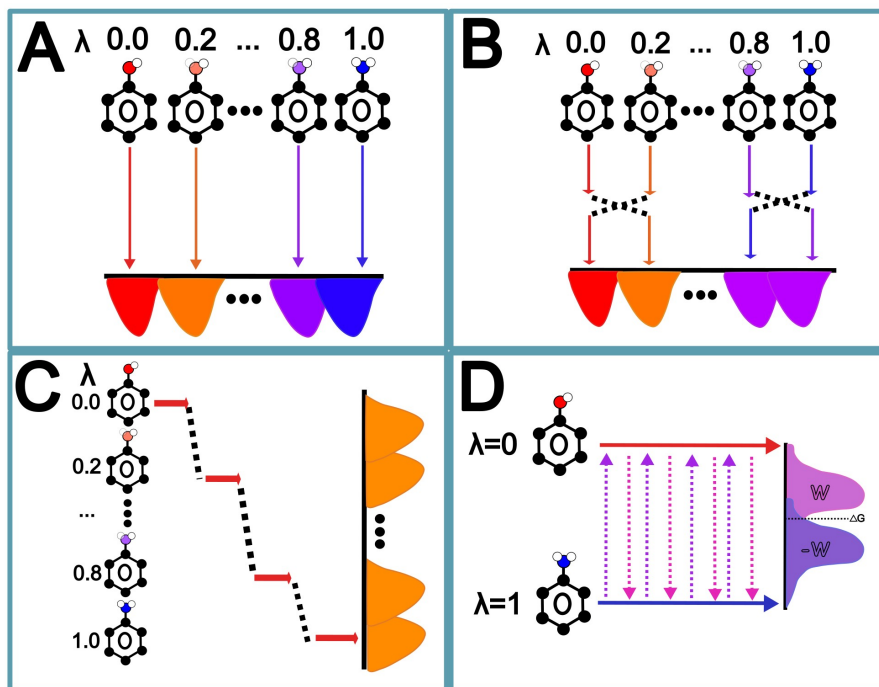


Figura 2.1: Metodologías basadas en transformaciones alquímicas. Los distintos métodos corresponden a: A) Transformaciones alquímicas de equilibrio, B) Intercambio de réplicas aplicadas al Hamiltoniano, C) Muestreo de esquema de réplica única y D) Transformaciones alquímicas de no-equilibrio. Cada uno de los métodos da lugar al cambio de energía libre asociado a transformar un fenol en una anilina.

conformaciones representativas del sistema, los esquemas para realizar transformaciones alquímicas se pueden clasificar en 4 tipos (Figura 2.1). El primer método representa el esquema más simple, en el cual una o más conformaciones son simuladas en cada uno de los estados definidos, desde el estado inicial hasta el final, pasando por los intermediarios¹¹⁸.

El segundo método corresponde a un tipo de intercambio de réplicas conocido como *Hamiltonian Replica Exchange* (H-RE). En este enfoque, se realizan múltiples simulaciones con distintos valores de λ , permitiendo inter-

cambios entre sistemas vecinos, siempre que se cumpla el criterio de aceptación de Metropolis-Hastings. Aquel enfoque permite el intercambio de hamiltonianos, generando un conjunto de réplicas que mantienen una distribución de Boltzmann y mejora la exploración del espacio de fase¹¹⁹.

El tercer método introduce un esquema relativamente nuevo, en el cual un único sistema recorre rápidamente todos los estados mediante simulaciones cortas, intercambiando su estado hasta completar todas las configuraciones posibles. Por esta razón, se le conoce como Esquema de muestreo de réplica única¹²⁰.

Finalmente, el cuarto método corresponde a una estrategia de transformaciones de no-equilibrio. En este enfoque, solo los estados inicial y final son explorados mediante simulaciones de equilibrio, mientras que los estados intermedios se generan a partir de simulaciones rápidas durante la transformación entre los estados 0 a 1, y viceversa. Este proceso permite modelar la evolución del sistema en función de sus interacciones moleculares. Tal evolución ocurre de manera irreversible, llevando al sistema fuera del equilibrio, de donde surge el nombre de la técnica. Dado que este último esquema es la base de la presente tesis, su formalismo se explica en detalle en una sección posterior⁷⁵.

2.5.1. Estimadores de energía libre en transformaciones químicas

Una vez explorado el espacio de fase de la transformación química, es necesario aplicar una metodología para extraer de la simulación la energía libre asociada al proceso. Asumiendo que la transformación química implica un cambio en el volumen despreciable, se puede expresar el cambio de energía libre de Gibbs como:

$$\Delta G_{0 \rightarrow 1} \approx \Delta A_{0 \rightarrow 1} = -k_B T \left(\ln \frac{P(\Theta_1)}{P(\Theta_0)} + \ln(B) \right) \quad (2.21)$$

donde k_B corresponde a la constante de Boltzmann y B es una constante que contiene el número de Avogadro N_A , la concentración estandar y el Volumen del sistema. P por su parte, es la probabilidad de encontrar al sistema en un estado Θ específico que depende de las coordenadas $\vec{q}$:

$$\int_{\Gamma_0} P_0(\Theta(\vec{q})) = \frac{\int_{\Gamma_0} e^{-\beta U(\vec{q})} d\vec{q}}{\int_{\Gamma} e^{-\beta U(\vec{q})} d\vec{q}}. \quad (2.22)$$

donde $\vec{q}$ representa las coordenadas del sistema y β el inverso del término $k_B T$, mientras que Γ representa el espacio de fase y U la función de energía potencial. De la ecuación 2.22 se hace evidente que la estimación de la energía libre se verá afectado por dos factores: la cantidad de configuraciones

correspondientes a los estados inicial y final, junto a la definición de nuestra función de energía potencial, que como se explicó en capítulos anteriores corresponde a un campo de fuerza clásico.

La probabilidad de encontrar al sistema en el estado A se le denomina la función de partición configuracional $Z(\Theta)$, la cual se utiliza para obtener la energía libre desde la formulación del colectivo canónico. Esta formulación se conoce como el enfoque de histogramas de probabilidad, sin embargo debido a su estricto requerimiento de una buena exploración del espacio de fase, se suele optar por métodos donde la energía libre se obtiene por medio de la derivada de la función de partición, :

$$\Delta G = \frac{\partial}{\partial \beta} \ln Z(\beta). \quad (2.23)$$

La forma más simple de este segundo enfoque, corresponde al método de *Free Energy Perturbation*¹²¹ (FEP). La energía libre clásica entre dos estados caracterizados cada uno por un Hamiltoniano, H_0 y H_1 pueden ser expresado a partir de la relación entre sus funciones de partición $Z(\Theta)$, como:

$$\Delta G_{0 \rightarrow 1} = -\beta^{-1} \ln \frac{Z(\Theta_1)}{Z(\Theta_0)} \quad (2.24)$$

En este contexto, Zwanzig postuló una relación que permite calcular la energía libre directamente, mediante el muestro de configuraciones en el estado inicial Θ_0 , para luego evaluar su energía potencial en el Hamiltoniano

del estado final H_1 :

$$\Delta G_{0 \rightarrow 1} = G_1 - G_0 = -\frac{1}{\beta} \ln \langle e^{-\beta(U_1(\vec{q}) - U_0(\vec{q}))} \rangle_0 \quad (2.25)$$

donde $\langle \rangle_0$ indica que se obtiene el promedio sobre el colectivo del estado inicial. Esta relación comprende la base del método FEP, que si bien es sencillo de implementar, ya que sólo necesita recolectar la información de la energía para el conjunto de configuraciones del estado inicial, posee desventajas bastante marcadas frente a otros métodos. La principal problemática es la poca precisión de sus cálculos cuando los espacios de fase entre los estados no están bien solapados. Esto quiere decir que configuraciones de un estado que son muy improbables de poseer el Hamiltoniano del otro estado contribuirán poco al promedio dado el carácter exponencial de la expresión. Lo anterior repercute en la incertidumbre de los resultados, creciendo exponencialmente en relación al bajo solapamiento de los espacios y por lo tanto, convergiendo lentamente. En resumen, FEP corresponde a una metodología en la cual sus resultados son gobernados por configuraciones raras de baja probabilidad.

Otra metodología ampliamente utilizada corresponde a la integración termodinámica, que utiliza la derivada de la energía libre G , respecto al parámetro λ :

$$\frac{dG}{d\lambda} = \frac{d}{d\lambda} \left[-\ln \int_{\Gamma} \frac{e^{-U(\vec{q}, \lambda)}}{Z(\lambda)} \right] d\vec{q} = \left\langle \frac{dU(\vec{q}, \lambda)}{d\lambda} \right\rangle_{\lambda} \quad (2.26)$$

En este método se simula el sistema a valores discretos de λ para determinar

el cambio de la energía interna frente al cambio de λ , para posteriormente integrar numéricamente $dG/d\lambda$ realizando la siguiente integración:

$$\Delta G = \int_0^1 \left\langle \frac{dU(\vec{q}, \lambda)}{d\lambda} \right\rangle d\lambda. \quad (2.27)$$

Una de las ventajas respecto a FEP, recae en el hecho que obtener una correcta distribución del promedio $\left\langle \frac{dU(\vec{q}, \lambda)}{d\lambda} \right\rangle$, requiere de menos configuraciones no correlacionadas debido a su carácter no exponencial, exhibiendo menores variaciones y una distribución más parecida a una gaussiana. El principal problema que limita esta metodología, radica en el sesgo o la pérdida de información al realizar la cuadratura cuando se integra. Este problema numérico debe ser minimizado lo suficiente para estar por debajo del ruido estadístico. Muchos esquemas de integración han sido propuestos, pero el más ampliamente utilizado corresponde a la regla del trapecioide. A modo general, todos los esquemas de integración pueden escribirse como:

$$\Delta F \approx \sum_{k=1}^K w_k \left\langle \frac{dU(\vec{q}, \lambda)}{d\lambda} \right\rangle_k \quad (2.28)$$

donde w_k dependerá del esquema elegido. Si bien esquemas que presentan mejores resultados que la regla del trapecioide han sido postulados, la mayoría requiere una elección específica de los estados intermediarios definidos por λ para evitar el sesgo, lo cual lo vuelve contraproducente en sistemas con intermediarios que poseen amplios y variables niveles de incertidumbre. En

la práctica utilizando la regla de los trapezoides y agregando 2 a 3 intermedios es suficiente para alcanzar la precisión de métodos de integración mas complejos¹²².

Un tercer método ampliamente utilizado es el *Bennett Acceptance Ratio*¹²³ (BAR), el cual se basa en el principio de que, si se conoce la diferencia de energía potencial al pasar del estado inicial al estado final, también se dispone de la diferencia obtenida en el proceso inverso. Al aprovechar esta relación, es posible construir un estimador más preciso y confiable. Este enfoque permite calcular la energía libre reducida entre dos estados a partir de la siguiente ecuación:

$$\sum_{i=1}^{n_i} \frac{1}{1 + \exp \left[\ln \left(\frac{n_i}{n_j} \right) + u_{ij}(\vec{q}) - f_{ij} \right]} = \sum_{i=1}^{n_j} \frac{1}{1 + \exp \left[\ln \left(\frac{n_i}{n_j} \right) - u_{ij}(\vec{q}) + f_{ij} \right]} \quad (2.29)$$

donde n corresponde al número de configuraciones de cada estado. Esta metodología es superior a FEP y converge a esta cuando el número de configuraciones corresponde a 1. Una de las principales ventajas de BAR es que requiere de menos solapamiento entre el espacio de fase de los dos estados, sin embargo, todavía es necesario. Dado que el método BAR solo considera dos estados (inicial y final), la estimación puede verse afectada si no hay un solapamiento adecuado entre ellos, especialmente cuando la transformación implica una perturbación muy grande. Para abordar esto, es necesario introducir una serie de estados intermedios que conecten gradualmente el estado

inicial con el estado final:

$$f_{1 \rightarrow N} = f_{1 \rightarrow 2} + f_{2 \rightarrow 3} + \dots + f_{N-1 \rightarrow N} \quad (2.30)$$

Una de las desventajas de BAR es que la incertidumbre entre $f_{i-1 \rightarrow i}$ y $f_{i \rightarrow i+1}$, están correlacionadas, ya que ambas son obtenidos a partir de la energía u_i . Como resultado, no pueden considerarse variables independientes, lo que impide propagar la incertidumbre de manera convencional, como se haría al calcular la varianza de la energía libre mediante cuadratura en la integración termodinámica. Algunos métodos como *bootstrapping*, son utilizados para obtener la incertidumbre de la energía libre utilizando la metodología BAR¹²³ y es el método utilizado en este trabajo.

En la siguiente sección se revisarán los detalles de las transformaciones químicas irreversibles, mostrando sus fundamentos, estimadores, metodologías, herramientas y las principales ventajas y desventajas con las metodologías clásicas.

2.5.2. Transformaciones de No-equilibrio

En las transformaciones químicas reversibles, se asume que el sistema cambia en pequeñas porciones $\lim_{\lambda \rightarrow 0}$ en las cuales el sistema siempre se encuentra en equilibrio cumpliéndose la igualdad en los procesos que impliquen

un trabajo en el sistema reversible¹²⁴:

$$\langle W(\tau) \rangle \geq \Delta G \quad (2.31)$$

donde τ corresponde al tiempo que toma el proceso a evaluar. En ese sentido la igualdad se cumple para $\tau \rightarrow \infty$, que viene siendo una transformación en un tiempo infinito. En todos los otros casos tendrían lugar procesos irreversibles, donde el valor promedio del trabajo será mayor al valor de energía libre debido a disipación de calor. Sin embargo en 1997 Jarzynski derivó una igualdad que relaciona el promedio de un trabajo irreversible con la diferencia de energía libre en equilibrio para un proceso dado :

$$\left\langle e^{-\beta W(\tau)} \right\rangle = e^{-\beta \Delta G}. \quad (2.32)$$

El trabajo dependerá del camino que conecte el estado inicial y final del proceso. Para una transformación química el estado inicial sería $=0$ y el final $=1$. para transformaciones de no equilibrio el valor de λ varía dinámicamente con el tiempo, permitiendo llegar del esta 0 al 1 en una única simulación. El promedio del lado izquierdo de la igualdad es un promedio de todas las transformaciones de no-equilibrio, que a su vez dependen inicialmente del colectivo de configuraciones obtenido a partir de simulaciones de equilibrio (como se explicó anteriormente, siempre se debe realizar una equilibración de los sistemas mediante simulaciones de dinámica molecular, en este caso

del estado inicial).

El trabajo puede ser obtenido a partir de:

$$W(\tau) = \int_{\lambda=0}^1 \frac{dH(\vec{q}, \lambda)}{d\lambda} d\lambda \quad (2.33)$$

donde el trabajo viene dado por el cambio en el Hamiltoniano en función de λ . Notar que en este caso, para diferenciar del método de integración termodinámica, el trabajo se obtiene de la integración numérica sobre los valores instantáneos del cambio en la energía respecto a λ y no de los promedios. Cabe señalar que en las situaciones límites cuando los cambios son extremadamente rápidos o lentos la igualdad de Jarzynski se reduce a la ecuación de FEP 2.25, que puede entenderse como un cambio instantáneo de la energía potencial y un cambio reversible respectivamente. Ahora de la igualdad de Jarzynski se obtiene la energía libre del proceso como:

$$\widehat{\Delta G} = -k_B T \ln \left[\frac{1}{N} \sum_i^N e^{-\beta W_i} \right] \quad (2.34)$$

donde la suma de los N trabajos irreversibles promediados nos entregan el valor de $\widehat{\Delta G}$. Desafortunadamente, este estimador al ser exponencial presenta los mismos problemas que el método FEP, por lo que el promedio dependerá principalmente de los valores de la cola de la distribución de trabajo, convergiendo lentamente a un valor. Mas aún, debido a que se toma un finito número de configuraciones, el estimador tiende a ser sesgado requiriendo un

gran número de simulaciones de equilibrio para poder explorar el espacio de fase¹²⁵.

Algunos de los problemas anteriormente señalados pueden resolverse aplicando el teorema de fluctuación de Crooks¹²⁶. La diferencia con la igualdad de Jarzynski radica en que no sólo considera la transformación desde un estado 0 a un estado 1, si no que también considera el proceso inverso, obteniéndose una relación entre las distribuciones de trabajo en ambos sentidos y la diferencia de energía libre de la siguiente forma:

$$\frac{P_{\lambda 0 \rightarrow 1}(W)}{P_{\lambda 1 \rightarrow 0}(-W)} = e^{\beta(W - \Delta G)} \quad (2.35)$$

donde $P_{\lambda 0 \rightarrow 1}$ y $P_{\lambda 1 \rightarrow 0}$ son las distribuciones de probabilidad normalizadas de los trabajos en ambos sentidos. De acuerdo a esta igualdad, teniendo un solapamiento entre los trabajos de ambas direcciones la energía libre puede ser obtenida a través de siguiente ecuación:

$$\widehat{\Delta G} = W + k_B T \ln \frac{P_{\lambda 0 \rightarrow 1}(W)}{P_{\lambda 1 \rightarrow 0}(-W)}. \quad (2.36)$$

Aunque el teorema de Crooks mejora la estimación de la energía libre en comparación con la formulación de Jarzynski, aún presenta ciertas limitaciones. Una de ellas es la contribución a la energía libre de estados poco frecuentes en los que la disipación del trabajo es pequeña. Otra limitación es la dificultad para lograr un solapamiento adecuado entre las distribucio-

nes de probabilidad directa e inversa en algunas transformaciones de mayor perturbación¹²⁷. Una posibilidad para tratar tales problemas sería aproximar la distribución de trabajo a una función analítica. Goette y Grubmüller proponen una aproximación Gaussiana llamado enfoque *Crooks Gaussian Intersection*¹²⁸ (CGI). Esta aproximación llevaría a un preciso cálculo de la energía libre siempre y cuando la suposición del comportamiento Gaussiano sea válido, lo cual puede comprobarse utilizando algún test estadístico como el Kolmogorov–Smirnov. Una de las desventajas de este método es que al igual que BAR no entrega un error, siendo en este caso un error analítico, pero puede estimarse de la misma forma utilizando *bootstrapping*.

Por último al igual que en la formulación reversible, para las simulaciones de no-equilibrio la energía libre puede obtenerse a través del estimador BAR. Sustituyendo la diferencia de energía potencial por el trabajo y partiendo del teorema de Crooks, BAR fue adaptado a las simulaciones de no-equilibrio llegando a la siguiente expresión:

$$\sum_{i=1}^{N_{\lambda 0 \rightarrow 1}} \frac{1}{1 + \frac{N_{\lambda 1 \rightarrow 0}}{N_{\lambda 0 \rightarrow 1}} e^{\beta(W_i - \widehat{\Delta G})}} = \sum_{j=1}^{N_{\lambda 1 \rightarrow 0}} \frac{1}{1 + \frac{N_{\lambda 1 \rightarrow 0}}{N_{\lambda 0 \rightarrow 1}} e^{-\beta(W_j - \widehat{\Delta G})}} \quad (2.37)$$

Este estimador no presenta sesgo, su varianza puede determinarse analíticamente y se han propuesto criterios de convergencia para evaluar las simulaciones. Sin embargo, una de sus limitaciones es que asume que los valores de trabajo son estadísticamente independientes, lo cual no siempre es cierto,

especialmente cuando las configuraciones iniciales se obtienen de simulaciones de equilibrio con alta frecuencia de muestreo. Para mitigar este problema, es común extraer configuraciones equidistantes en el tiempo a partir de simulaciones de equilibrio. Por ejemplo, en una simulación de 300 ns, se pueden descartar los primeros 100 ns y seleccionar 100 conformaciones separadas cada 2 ns.

2.6. Cálculos de energía libre de activación $\Delta G^\ddagger$

En la sección anterior se explicó en detalle la teoría para obtener la energía libre de un proceso a través de transformaciones alquímicas. Estos enfoques han tenido éxito en procesos tales como unión de fármacos a dianas biomoleculares, estabilidad térmica de proteínas o energías de solvatación de moléculas orgánicas. Notar que todos los procesos anteriormente mencionados corresponden a transformaciones donde los estados inicial y final se encuentran en equilibrio, por lo tanto, son propiedades termodinámicas. Dado que este trabajo se enfoca principalmente en las propiedades catalíticas de las enzimas, se debe introducir la teoría que conecta la cinética química con la termodinámica. Esta sección busca introducir al lector a la definición de la energía libre de activación $\Delta G^\ddagger$ a través de la teoría del estado de transición.

La velocidad de reacción se entiende por la medida de avance de una reacción en función del tiempo. Para una reacción unimolecular de primer

orden, la velocidad de la reacción viene dado por:

$$[A]_t = [A]_0 k e^{-kt} \quad (2.38)$$

donde la concentración del reactante A a lo largo del tiempo t va desapareciendo, siendo k la constante de la reacción de primer orden. Para la mayoría de reacciones enzimáticas sujetas a ser de carácter "Michaeliano", se puede asumir que son de primer orden y la constante de reacción se define como k_{cat} . Una de las teorías más utilizadas que permite hacer una conexión entre esta propiedad cinética y la mecánica estadística es la teoría del estado de transición⁶⁵ (TST). En esta teoría se asumen 4 suposiciones:

- La aproximación de Born-Oppenheimer es válida a lo largo de toda la reacción
- Los núcleos a lo largo de la reacción exhiben un comportamiento clásico
- Existe un equilibrio térmico entre el estado de reactante y el de transición
- Una vez cruzado el estado de transición este continuará hacia el camino de los productos sin retornar a los reactantes nuevamente

Si estas 4 suposiciones son válidas la constante catalítica puede obtenerse a través de:

$$k_{cat} = \frac{k_B T}{h} e^{-\frac{\Delta G^\ddagger}{k_B T}} \quad (2.39)$$

La ecuación 2.39 nos permite relacionar una propiedad cinética como lo es la constante de velocidad de una reacción química con una propiedad termodinámica llamada la energía libre de activación $\Delta G^\ddagger$. Excluyendo las reacciones catalíticas fotoquímicas o de transferencia electrónica, las 3 primeras condiciones pueden ser asumidas a cumplirse en la mayoría de los casos⁶⁷. La cuarta condición, sin embargo, suele fallar debido a difusiones del sistema hacia coordenadas ortogonales a la coordenada de reacción que permiten volver al estado de reactante. Sin embargo, la ecuación 2.39 puede ser adaptada para considerar este efecto, obteniéndose la denominada ecuación de Eyring:

$$k_{cat} = \frac{\kappa k_B T}{h} e^{-\left(\frac{1}{k_B T} \Delta G^\ddagger\right)} \quad (2.40)$$

donde el término κ corresponde al coeficiente de transmisión que tiene el valor de 1 cuando una vez el cruzado el ET no vuelve a reactante y $0 < \kappa < 1$ en el otro caso. Como se verá en las secciones siguientes, este trabajo está enfocado en calcular las diferencias en la energía libre de activación $\Delta \Delta G$ por lo que asumiendo que κ no varía entre enzimas, la condición cuatro no es necesaria para aplicar la TST.

2.6.1. Método de la cuerda adaptativa

Anteriormente se definió la TST que asume la existencia de un camino de reacción conectando reactantes y productos a través de una coordenada

específica. Considerando el rango del coeficiente de transmisión, se puede deducir que la TST siempre entregará un valor igual o mayor a la constante de velocidad real. Esto define a la teoría del estado de transición como un límite superior al valor real de k_{cat} , permitiendo aplicar el principio variacional a la minimización de la energía. Esto se conoce como el camino de mínima energía libre o *Minimum Free Energy Path*(MFEP) por sus siglas en inglés y comprende la base de muchos métodos para encontrar el perfil de reacción. Para reacciones modelos en vacío como una disociación de un gas bimolecular, el camino de reacción puede ser modelado por una simple coordenada de reacción: la distancia entre los núcleos. Sin embargo, si se aplica el mismo principio a una reacción en disolución o con un gran número de grados de libertad como es el caso de reacciones enzimáticas, el sistema cae en la bien conocida "maldición de la dimensionalidad". En cualquiera de las dos situaciones, sin embargo, es necesario poder modelar la ruptura y formación de enlace, requiriéndose obligatoriamente el uso de enfoques mecanocuánticos, los cuales pueden ser del tipo multiestado QM/MM para sistemas de gran tamaño. Desafortunadamente aún utilizando tal tipo de enfoque el problema comprende un amplio número de variables colectivas a definir (coordenadas geométricas intra e intermoleculares), para poder capturar los grados de libertad asociados con la reacción química. Ejemplificando con un sistema que requiere de 3 variables colectivas (CVs) para describir una reacción química, métodos ampliamente utilizados como *Umbrella Sampling*(US) o metadiná-

mica fallan debido a que su costo computacional crece exponencialmente con cada una de estas variables.

Una alternativa a esta problemática es el método de la cuerda adaptativa⁸¹ (ASM) el cual comienza su formulación utilizando la definición formal de MFEP para una reacción que involucra D variables colectivas($\theta_1, \theta_2, \dots, \theta_D$):

$$\frac{d\mathbf{z}(\alpha)}{d\alpha} \parallel \mathbf{M}(\mathbf{z}(\alpha)) \nabla A(\mathbf{z})(\alpha) \quad (2.41)$$

donde ∇A es el gradiente de la superficie de energía libre D-dimensional en el punto $\mathbf{z}(\alpha)$ y $\mathbf{M}$ es el tensor métrico de distancia que puede entenderse como un objeto matemático que permite encontrar la distancia entre dos puntos en el hiperespacio construido por las CVs y que suelen tener distinta naturaleza (ej: ángulos, distancias euclídeas, dihedros, hibridaciones, etc). La ecuación 2.41 define que para obtener el camino de mínima energía libre el sistema debe evolucionar de tal manera que el gradiente de la energía libre perpendicular($\parallel$) al camino de reacción debe ser minimizado hasta que desaparezca. En este punto el camino que conecta reactantes y productos obtenido es el trayecto más corto y siempre corresponde a un mínimo local en la dirección ortogonal a este. En la práctica este camino se construye a través de un vector $\mathbf{z}(\alpha)$ que se denomina *la cuerda* y que se discretiza en un número finito de n nodos:

$$\mathbf{z}(\alpha) = (\mathbf{z}_1(\alpha), \mathbf{z}_2(\alpha), \dots, \mathbf{z}_n(\alpha)) \quad (2.42)$$

donde α es la longitud del arco formado por la cuerda. De esta forma cada uno de los nodos puede ir minimizando la energía en cada instante t sujeto a la siguiente ecuación:

$$\gamma z'_i(t) = \mathbf{M}(x_i(t)) \mathbf{K}(\theta(x_i) - z_i) \quad (2.43)$$

donde γ corresponde a un factor de amortiguamiento que da cuenta de la velocidad con la que se mueve la cuerda y $\mathbf{M}$ corresponde al tensor métrico anteriormente mencionado donde cada elemento es definido como:

$$M_{ij} = \sum_{k=1}^{3N} \frac{\partial \theta_i(\mathbf{x})}{\partial x_k} \frac{1}{m_k} \frac{\partial \theta_j(\mathbf{x})}{\partial x_k} \quad (2.44)$$

con m_k , siendo la masa asociada con la k -ésima coordenada cartesiana. La ecuación 2.43 da cuenta de un proceso de evolución temporal donde las posiciones de los nodos que conforman el camino de la cuerda son actualizados bajo el gradiente que genera el tensor métrico $\mathbf{M}$. Este movimiento se ve sujeto a la condición de explorar zonas cercanas a la vecindad de la posición del nodo en la iteración anterior z_i . Cualquier desviación de esta posición de referencia genera una fuerza determinada por las constantes contenidas en la matriz $\mathbf{K}$. Un evidente problema con el esquema anteriormente presentado es que inevitablemente la evolución de la cuerda culminará en una acumulación de los nodos en los mínimos más cercanos. Para evitar este colapso, la cuerda es interpolada después de cada actualización de los nodos utilizando la lon-

gitud del arco que forma y posicionando cada nodo de manera equidistante:

$$z'_i = \zeta \left(\frac{i-1}{n-1} L \right) \quad (2.45)$$

donde ζ es una interpolación de la cuerda, n es el número total de nodos que conforman la cuerda y L la longitud del arco. Con la ecuación 2.45 la cuerda puede converger al camino de mínima energía libre que transforma reactantes en productos. En este trabajo se emplea la cuerda adaptativa, la cual genera constantes de fuerza asociadas a los elementos de la matriz $\mathbf{K}$. Estas constantes son distintas para la coordenada de reacción y para la dirección ortogonal a esta, y se actualizan a lo largo de la evolución de la cuerda. Una vez que el camino ha convergido, tanto la posición de los nodos como las constantes de fuerza pueden utilizarse como ventanas para aplicar la metodología de *Umbrella Sampling*. Esto permite obtener con precisión la energía libre a lo largo del camino en cada nodo.

2.7. Cálculos de energía libre de activación relativa mediante transformaciones alquímicas

Anteriormente se detalló como obtener el perfil de energía libre a lo largo de la coordenada de reacción la cual puede involucrar múltiples variables colectivas. Lo anterior nos entrega una forma explícita de calcular la energía

libre de activación $\Delta G^\ddagger$ restando la energía libre del colectivo de estados de transición menos la energía libre del colectivo que conforman los reactantes. La principal desventaja radica en el alto costo computacional asociado. En este trabajo se presenta una alternativa al método de la cuerda para situaciones en las que se requiera calcular la diferencia en la energía libre de activación $\Delta\Delta G^\ddagger$, producto de una mutación o un cambio en el catalizador, en lugar de la energía de activación en sí. Partiendo de la suposición que para reacciones de múltiples pasos el estado de transición y por ende el paso limitante de la reacción no se ve afectado por la mutación, la diferencia en la energía libre de activación puede calcularse mediante transformaciones alquímicas. La figura 1.1 muestra que esta energía libre relativa puede calcularse indirectamente aprovechando que esta propiedad es una función de estado y no depende del camino por el cual es obtenido. De esta forma si se calcula la energía libre asociado a mutar una enzima en el estado de transición y se resta la energía libre del mismo proceso en el estado reactante, se puede obtener la variación de la energía libre de activación debida a la mutación sin la necesidad de describir la ruptura y formación de enlaces. Tales transformaciones comprenden procesos alquímicos que pueden modelarse en su totalidad mediante simulaciones de dinámica molecular clásica con los métodos presentados anteriormente para estimar el cambio de la energía libre, reduciendo el costo computacional considerablemente. Sin embargo, tal enfoque implica ser capaz de poder modelar los estados de transición mediante campos de fuerza.

2.7.1. Derivación de un campo de fuerza para estados de transición

Como se describió en secciones anteriores un campo de fuerza se compone de funciones matemáticas que describen las interacciones no-enlazantes y enlazantes para un determinado estado. Para poder modelar un estado de transición (o cualquier punto estacionario) en principio se necesitan 3 componentes:

- La geometría del estado de transición
- La densidad electrónica de las especies que conforman el ET
- Las fuerzas asociadas a los desplazamientos de los átomos bajo una aproximación armónica y ponderados por la masa, lo que permite construir la llamada matriz hessiana H

Los primeros dos componentes nos permiten obtener los parámetros no enlazantes mediante la teoría de átomos en moléculas¹²⁹. Esta teoría permite asignar densidades de probabilidad en el espacio en cada punto $\mathbf{r}$ a cada uno de los átomos i de acuerdo con la siguiente expresión:

$$\rho_{mol}(\mathbf{r}) = \sum_{i=1}^{N_{atoms}} \rho_i(\mathbf{r}) \quad (2.46)$$

Existen diversas formas de definir la densidad atómica, basadas, por ejemplo, en la existencia o ausencia de solapamiento entre las densidades electrónicas. Cada uno de estos enfoques se conoce como un método de partición. Este trabajo se centrará en la familia de métodos Hirshfeld¹³⁰, los cuales definen la densidad atómica como:

$$\rho_i(\mathbf{r}) = \omega_i(\mathbf{r})\rho_{mol}(\mathbf{r}), \quad \sum_{i=1}^{N_{atoms}} \omega_i(\mathbf{r}) = 1 \quad (2.47)$$

donde $\omega_i(\mathbf{r})$ representa el peso asignado a la densidad electrónica molecular correspondiente al átomo i en el punto $\mathbf{r}$ del espacio, cumpliendo la condición de que la suma de todos los pesos en un mismo punto sea igual a 1. Luego, para un átomo A , con carga nuclear Z_A y densidad electrónica atómica $\rho_A(\mathbf{r})$, la población electrónica atómica (N_A) y la carga atómica (q_A) se calculan mediante:

$$q_A = Z_A - N_A = Z_A - \int \rho_A(\mathbf{r}) d\mathbf{r} \quad (2.48)$$

El volumen atómico V_A a su vez se define como el tercer momento radial de la densidad electrónica atómica:

$$V_A = \int |\mathbf{r} - \mathbf{R}_A|^3 \rho_A(\mathbf{r}) d\mathbf{r} \quad (2.49)$$

La formalización de Tkatchenko-Scheffler¹⁰² se utiliza para calcular el coeficiente de dispersión $C_{6,A}$ de un átomo A en una molécula basado en los

volúmenes calculados. Este coeficiente se obtiene al reescalar el coeficiente $C_{6,A,free}$ del átomo aislado correspondiente, utilizando el cuadrado de la razón de los volúmenes atómicos entre el átomo en la molécula y su estado aislado ($V_{A,free}$):

$$C_{6,A} = \left(\frac{\langle V_A \rangle}{\langle V_{A,free} \rangle} \right)^2 C_{6,A,free} \quad (2.50)$$

Los radios de Van der Waals (R_{vdW}) de los átomos en la molécula pueden ser aproximados mediante la relación de Fedorov-Tkatchenko¹³¹:

$$R_{vdW} = 2,54\alpha^{1/7} \quad (2.51)$$

$$\alpha_A = \alpha_{A,free} \left(\frac{\langle V_A \rangle}{\langle V_{A,free} \rangle} \right)$$

donde α_A y $\alpha_{A,free}$ corresponden a la polarizabilidad estática atómica del átomo A en la molécula y en el vacío, respectivamente. A partir de estos parámetros, es posible determinar los parámetros de la interacción de Lennard-Jones (ϵ y σ) entre dos átomos idénticos. La energía en el mínimo del potencial se define como $V_{\min} = -\epsilon$, y los parámetros se calculan a partir de C_6 y R_{vdW} definidos en las ecuaciones (2.50) y (2.51):

$$\sigma_A = \frac{r_{\min}}{2^{\frac{1}{6}}} = \frac{2R_{vdw,A}}{2^{\frac{1}{6}}} \quad (2.52)$$

$$\epsilon_A = \frac{C_{6,A}}{4\sigma^6} = \frac{C_{6,A}}{2^7 R_{vdw,A}^6}$$

Las interacciones enlazantes por su parte requieren un proceso de dos pa-

sos que implica tener caracterizado el estado de transición en la superficie de energía potencial, el cual es definido como un punto de silla con un modo normal imaginario el cual se proyecta en dirección a la coordenada de reacción. Al ser un máximo en esta dirección, corresponde a un estado de alta energía que sin fuerzas externas cae a los mínimos cercanos. Sin embargo, Rosales y colaboradores¹³² postulan que este punto de silla puede ser transformado en un mínimo artificial a través de una modificación de la hessiana:

$$Hv = \lambda v \quad (2.53)$$

donde H es la Hessiana, v son los vectores propios y λ son los valores propios. De la ecuación 2.53 se obtienen los valores propios a través de la resolución del polinomio característico. Luego, se realiza la modificación de los valores propios negativos, transformándolos a 1.0 unidad atómica:

$$\lambda'_i = \begin{cases} 1,0 & \text{si } \lambda_i < 0, \\ \lambda_i & \text{si } \lambda_i \geq 0, \end{cases} \quad (2.54)$$

donde finalmente, la Hessiana transformada H' se obtiene como:

$$H' = v\Lambda'v^T \quad (2.55)$$

donde Λ' es la matriz diagonal con los valores propios modificados y v son los

vectores propios de H . Este algoritmo sustenta la base del método Q2MM, donde los parámetros no enlazados son obtenidos por ajuste de los modos normales vibracionales clásicos para reproducir los modos normales cuánticos. Dada la poca información de referencia y que la función objetivo en la optimización de los parámetros no tiene solución única (múltiples combinaciones de parámetros enlazantes pueden entregar un mismo valor propio), hace de esta parametrización un método cuestionable.

Sin embargo, teniendo la matriz hessiana calculada, se puede utilizar el método de proyecciones de fuerzas postulado por Seminario¹³³, donde los modos normales son proyectados a lo largo de enlaces y ángulos obteniéndose las constantes de fuerza. Este enfoque comienza dividiendo la hessiana en matrices con dimensiones 3×3 . Estas matrices contienen las fuerzas asociadas al movimiento de dos átomos. De lo anterior resulta evidente que teniendo un átomo A y B es posible proyectar los vectores propios de la matriz 3×3 sobre el vector que forma el enlace entre dichos átomos para obtener la fuerza:

$$K_r = \sum_{i=1}^3 k_i^{AB} |\mu^{AB} \cdot v_i^{AB}| \quad (2.56)$$

donde k_i es el valor propio (constante de fuerza) asociado al vector propio v_i y μ^{AB} corresponde al vector que forma el enlace. Dado que la matriz hessiana se obtiene de una geometría refinada, las distancias de equilibrio son obtenidas directamente de las coordenadas del sistema. Para los parámetros angulares

se construye el ángulo por medio de dos vectores ortonormales a los enlaces que construyen el ángulo. La proyección de las fuerzas sobre dichos vectores permite construir la constante de fuerza angular como:

$$\frac{1}{k_{\theta}} = \frac{1 + \frac{\sum_{i=1}^N |\mu^{PA1} \mu^{PAi}|^2 - 1}{N-1}}{R_{AB}^2 K_{PA}} + \frac{1 + \frac{\sum_{i=1}^M |\mu^{PC1} \mu^{PCi}|^2 - 1}{M-1}}{R_{CB}^2 K_{PC}} \quad (2.57)$$

siendo μ^{PA1} y μ^{PA} los vectores perpendiculares a los enlaces que construyen el ángulo en cuestion y N(M) comprenden el número de ángulos que tienen como átomo central a B y ejercen fuerzas sobre el enlace AB(BC). Por último, K se define como:

$$K_{ab} = \sum_{i=1}^3 k_i^{ab} \left| \mu^{ab} \cdot \nu^{ab} \right| \quad (2.58)$$

de la misma forma como se define la constante de fuerza de enlace en la ecuación 2.56. De esta forma se puede obtener los parámetros mínimos para generar la contribución enlazante del estado de transición. Los diedros son obtenidos directamente de campos de fuerza generalizado, debido a que procesos de optimización por ajuste se encuentran fuera del alcance de este trabajo. Finalmente agrupando los términos enlazantes y no enlazantes, se logra construir un campo de fuerza para el estado de transición basado en su totalidad en primeros principios. El complejo de Michaelis-Menten a su vez, puede ser obtenido de campos de fuerza generalizados y refinar sus parámetros no enlazantes de la misma forma en que se parametrizó el ET. Con los campos

de fuerza derivados, se tienen todos los componentes para poder aplicar las transformaciones químicas a cálculos de energías de activación relativas. En el siguiente capítulo se exponen y discuten los resultados obtenidos por esta metodología aplicada a sistemas enzimáticos.

3. Resultados y Discusión

La catálisis enzimática es uno de los procesos bioquímicos más específicos y eficientes de la naturaleza, utilizado en medicina, industria e investigación¹³⁴. Sin embargo, la creciente demanda de catalizadores más eficientes para la producción de compuestos de interés y la reducción del impacto ambiental ha impulsado el desarrollo de nuevas estrategias. La evolución dirigida, una técnica para mejorar enzimas, requiere múltiples intentos y recursos significativos debido a la naturaleza aleatoria de las mutaciones, lo que la hace costosa y poco viable para aplicaciones industriales frecuentes. Para optimizar este proceso, han surgido enfoques computacionales y estadísticos que buscan predecir mutaciones con características deseadas, reduciendo la experimentación en laboratorio. Dentro de los métodos computacionales, cabe destacar el constante crecimiento en el uso de enfoques basados en inteligencia artificial¹³⁵. Debido a que la catálisis enzimática involucra la ruptura y formación de enlaces químicos, la mayoría de los enfoques computacionales son basados en la química cuántica, lo que conlleva un alto costo computacional. En este trabajo, se propone una alternativa más eficiente basada en transformaciones alquímicas de no-equilibrio para predecir los cambios en la barrera de activación de enzimas inducidas por mutaciones.

Esta sección presenta y analiza los resultados más relevantes de la tesis doctoral. En primer lugar, se describe la implementación de la metodología, junto a un análisis de rendimiento, que permitió optimizar el conjunto de parámetros para equilibrar costo computacional y precisión, seguido de su validación en la enzima Crotonil-CoA Carboxilasa/Reductasa(CCR) de *Kittasatospora Setae*. Posteriormente, se aborda la derivación y evaluación de parámetros no enlazantes para cadenas laterales de aminoácidos, a modo de mejorar la descripción de las interacciones en el sitio activo, para luego poder aplicar estos nuevos parámetros al sistema CCR.

Como segundo caso de validación, se estudia la enzima Dihidrofolato Reductasa (DHFR), para la cual existen datos experimentales y computacionales de energía libre en tres mutantes. Finalmente, la metodología se aplica a la enzima *Old Yellow Enzyme* (OYE), determinando su perfil de energía libre a lo largo de la coordenada de reacción para identificar el paso limitante de una nueva reacción aldólica. Con el estado de transición caracterizado, se emplea el enfoque de no-equilibrio para identificar variantes con mayor eficiencia catalítica que la mutante Y187F, variante diseñada para realizar la catálisis de la nueva reacción aldólica.

3.1. Validación de las transformaciones alquímicas de no-equilibrio en la enzima Crotonil-CoA Carboxilasa/Reductasa CCR

En el contexto de este proyecto, las transformaciones alquímicas de no equilibrio han surgido como una metodología prometedora para evaluar las propiedades cinéticas de las proteínas, proporcionando información detallada sobre la reactividad de variantes enzimáticas. La implementación y validación de estas técnicas resulta fundamental para su aplicación en el diseño racional de proteínas, permitiendo no solo comprender los efectos de mutaciones específicas, sino también optimizar la actividad enzimática con un enfoque cuantitativo. Por lo tanto, el primer objetivo de este trabajo comprende la implementación y validación de las transformaciones alquímicas de no-equilibrio como una herramienta para determinar el cambio en la energía libre de activación producto de mutaciones en sistemas enzimáticos. Para poder obtener esta propiedad se define el ciclo termodinámico que permite determinar la energía libre relativa indirectamente a través de los procesos alquímicos. La figura 1.1 muestra que la energía libre de activación se asocia al proceso de transformar el complejo de Michaelis-Menten (CMM) en el estado de transición (ET).

Para dos sistemas enzimáticos distintos, se deduce que la diferencia en

tales barreras de activación puede obtenerse indirectamente a través de transformaciones alquímicas en el CMM y ET:

$$\Delta\Delta G^\ddagger = \Delta G^\ddagger_{mutante} - \Delta G^\ddagger_{silvestre} = \Delta G_{ET}^{Alch} - \Delta G_{CMM}^{Alch} \quad (3.1)$$

donde ΔG^{Alch} indica el cambio de energía libre producto de mutar un residuo desde la enzima silvestre a la mutante. Este ciclo termodinámico será utilizado en todo este trabajo, por lo que la metodología consiste en realizar simulaciones de los estados de transición y CMM en conjunto a transformaciones alquímicas que consideren mutar aminoácidos en cada uno de estos dos estados.

Todas las mutaciones son realizadas con el paquete PMX⁷³, el cual genera estructuras y topologías híbridas para una descripción basada en mecánica molecular de todo el sistema. Las energías libres asociadas a las mutaciones se obtienen, como es discutido en la metodología, mediante un parámetro λ , el cual interpola las interacciones entre una variante y otra de acuerdo a su implementación en el paquete de dinámica molecular GROMACS¹³⁶.

Definido el ciclo termodinámico, el protocolo para poder obtener las variaciones en la energía libre de activación inicia bajo la suposición de que se cumplen dos condiciones:

- El complejo de Michaelis-Menten y el estado de transición son conocidos

- Las conformaciones para ambos estados se encuentran disponibles

Si estas condiciones son cumplidas, se procede con la parametrización del sistema molecular en el CMM y ET. Los sustratos de reacciones catalizadas por enzimas a menudo involucran moléculas bien comportadas, en el sentido de que pueden ser parametrizadas en su totalidad mediante campos de fuerza generalizados. En este trabajo se utiliza el campo de fuerza CGENFF⁹⁸ para obtener los parámetros iniciales.

Dado que los parámetros no enlazantes en CGENFF se calculan en vacío, no incorporan el efecto del entorno enzimático. Para tener en cuenta esta influencia, los parámetros no enlazantes son modificados y se derivan utilizando una estructura representativa del CMM. A partir de esta conformación, se obtiene la densidad electrónica del sustrato mediante cálculos de punto aislado con la metodología multiestado QM/MM. La densidad electrónica es posteriormente particionada para finalmente obtener los parámetros. Para describir el sustrato en la región QM se utilizó la teoría del funcional de la densidad con el nivel de teoría B3LYP/def2-tzvpd^{137,138}. Como se mencionó en el capítulo anterior, esta densidad se ve polarizada por la zona MM, que en este caso particular incluye la enzima, moléculas de solvente e iones. Luego de esto, la densidad electrónica es dividida y asignada a cada átomo mediante un método de partición basado en la teoría de átomos-en-moléculas¹³⁰ (AIM). Con la densidad electrónica particionada se calculan las nuevas cargas y parámetros de Lennard-Jones de acuerdo a lo establecido en la metodología detallada en

el capítulo 2.7.1, para luego reemplazar los antiguos parámetros del campo de fuerza inicial.

La parametrización del estado de transición por su parte, comprende uno de los mayores desafíos de este trabajo. De forma similar al CMM, utilizando la estructura del estado de transición, se realiza un cálculo QM/MM para obtener la densidad polarizada, la cual posteriormente es particionada, para derivar los parámetros no-enlazantes. Sin embargo, debido a que las estructuras de los estados de transición a menudo violan las reglas químicas con las que se construyen los campos de fuerza generalizados (valencia, geometrías, etc.), no se pueden aplicar estos métodos para obtener parámetros enlazantes iniciales. Como aproximación más simple, se ha postulado el uso de potenciales armónicos que restringen distancias y ángulos para mantener la geometría del ET, a esta aproximación se le llamará la parametrización armónica. En este enfoque se utilizan los parámetros enlazantes de los reactantes y se guía hacia la geometría del ET a través de potenciales externos adicionales, los cuales son escogidos mediante inspección de los grados de libertad más relevantes ligados a la conformación¹³⁹.

Como alternativa a este enfoque, en este trabajo se propone una solución más rigurosa y que representa correctamente los modos normales ortogonales del ET. En comparación a la parametrización por restricciones, este nuevo método no es arbitrario en la elección de cuáles grados de libertad deben ser restringidos para mantener el ET. Por el contrario, todos los parámetros de

enlaces y ángulos involucrados en la formación del ET son derivados desde un enfoque mecanocuántico, confiriendo una generalización al método que no requiere intervención por parte del usuario.

Iniciando de la estructura del estado de transición en la superficie de energía libre, se realiza una búsqueda del ET sobre la superficie de energía potencial (SEP), el cuál se caracteriza por tener una sola frecuencia imaginaria representada como un valor propio negativo en la hessiana. Utilizando el algoritmo de Berny¹⁴⁰ implementado en Gaussian16¹⁴¹ se obtiene la estructura del ET, desde el cual se puede extraer su geometría y la matriz hessiana. Con los resultados del paso anterior y las ecuaciones 2.53, 2.54 y 2.55, se puede transformar la matriz hessiana en una nueva matriz que convierte al punto de silla que define al ET sobre la SEP en un mínimo. Esta transformación permite definir nuevos valores y vectores propios, de los cuales pueden derivarse las fuerzas que confinan al estado de transición en una vecindad cercana, evitando así, que caiga en algún otro mínimo. Esta nueva matriz hessiana que contiene la información respecto al movimiento correlacionado de los átomos es utilizada para obtener los parámetros enlazantes mediante el método de seminario modificado¹³³ (ecuaciones 2.56, 2.57 y 2.58).

Los procedimientos detallados en el párrafo anterior, entregan dos alternativas para poder obtener un campo de fuerza para el estado de transición que puede ser utilizado en las transformaciones alquímicas. Las interacciones entre los átomos que componen la enzima, iones y moléculas de agua se

realizan mediante campos de fuerzas establecidos, siendo el campo de fuerza CHARMM36m¹⁴², el elegido debido a su capacidad para modelar precisamente los cambios conformacionales en proteínas.

Una vez obtenido el campo de fuerza para el sistema completo, la generación de las estructuras híbridas, estructuras que poseen átomos de la enzima silvestre y mutante, y sus respectivas topologías, se procede a preparar y producir simulaciones de equilibrio para cada estado y cada variante enzimática. De las simulaciones de equilibrio se extraen conformaciones que se asumen independientes y permiten construir una distribución de Boltzmann de equilibrio. Estas conformaciones son transformadas de una variante enzimática en otra a través de transformaciones alquímicas de no-equilibrio. El detalle de todas las simulaciones puede ser encontrado en el Anexo C. El cambio de energía potencial en función del parámetro de transformación λ es obtenido e integrado para cada conformación, obteniendo una serie de trabajos, los cuales son utilizados para obtener la energía libre del proceso alquímico mediante algunos de los estimadores descritos en el capítulo 2.5.1. Aplicando este protocolo en el estado de transición y CMM se obtiene el cambio en la energía libre de activación. La figura 3.1 muestra un flujo de trabajo que resume el protocolo establecido.

Una vez definido el protocolo, este es aplicado a la enzima CCR con el fin de validarlo como alternativa para el cálculo de $\Delta\Delta G^\ddagger$. La enzima CCR cataliza la transferencia de hidruro desde la coenzima Nicotinamina Adenina

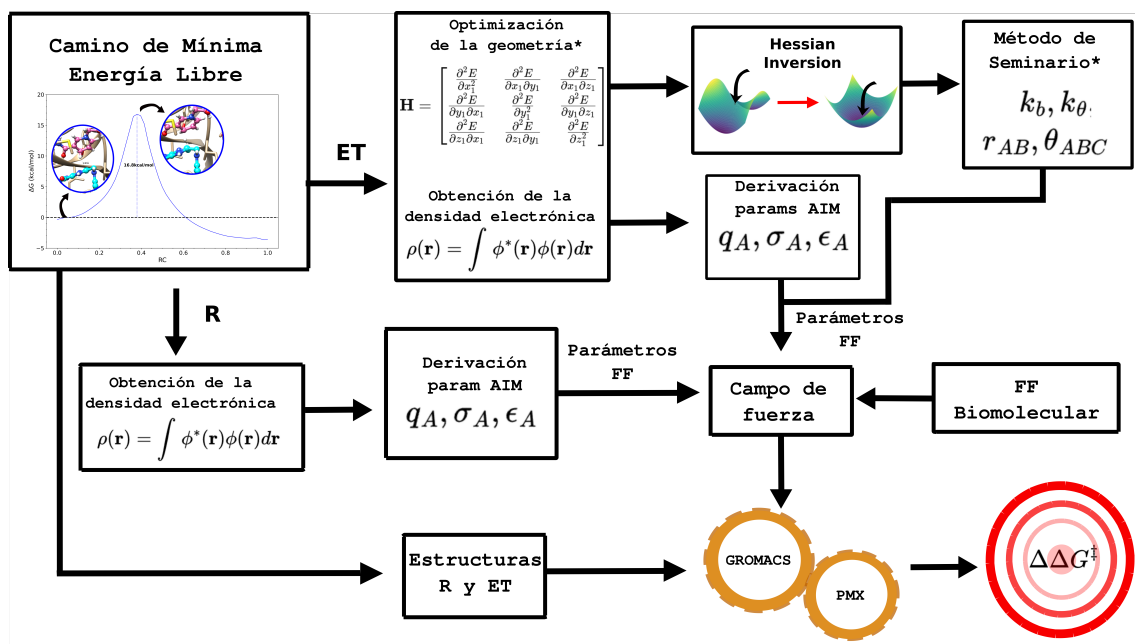


Figura 3.1: Diagrama de flujo para la estimación de $\Delta\Delta G^\ddagger$ mediante transformaciones alquímicas. Utilizando el camino de mínima energía libre, campos de fuerza para el CMM y estado de transición son derivados. En conjunto a estructuras de partida, estos son utilizados para realizar simulaciones de MD en Gromacs. Posteriormente, los datos son procesados con PMX para estimar los valores de energía libre.

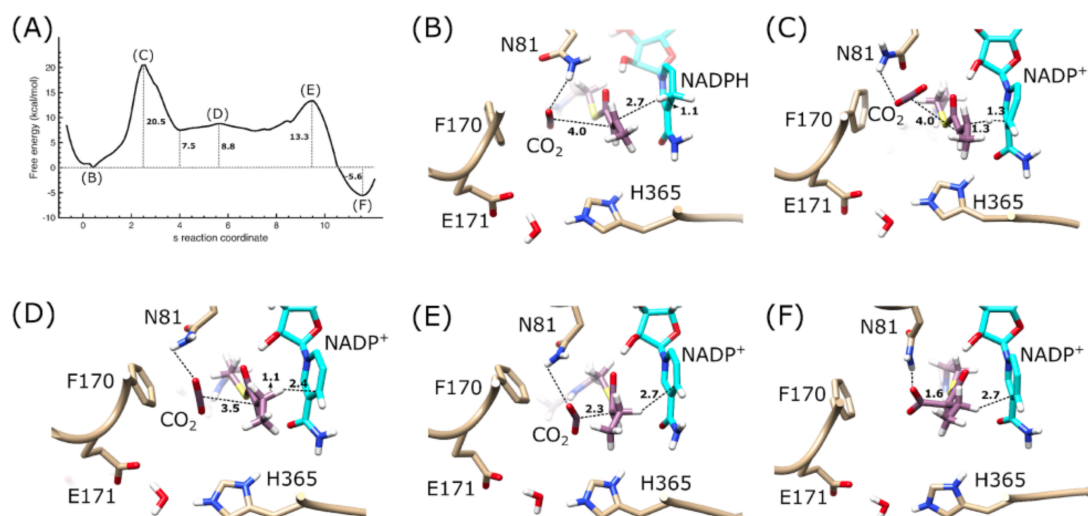


Figura 3.2: La figura A muestra el perfil de reacción de carboxilación para el sistema CCR mostrando una alta primera barrera de activación que corresponde a la formación del enolato. La figura B, C y F, corresponden a las estructuras del complejo Enzima+sustrato, estado de transición y complejo Enzima-producto respectivamente. Las figuras D y E son estructuras intermedias de la reacción. El perfil junto a las estructuras corresponde a resultados del trabajo previo de Recabarren y colaboradores¹⁴³

Dinucleótido Fosfato (NADPH) hacia una molécula de Crotonil-CoA (Figura 3.2). Esta transferencia corresponde al paso limitante de la reacción. En una segunda etapa el enolato formado puede atacar a un hidrógeno electrofílico generando el producto reducido o bien reaccionar con CO_2 para generar el producto carboxilado. La figura 3.2 muestra el perfil de energía libre de la reacción de CCR, donde se observa que la barrera de activación más alta corresponde al primer paso de la reacción. Por otro lado, cercano a la zona donde ocurre la transferencia del hidruro se observan 4 residuos, de los cuales se tiene evidencia experimental y computacional de ser claves en el funciona-

miento de CCR⁷⁸, N81, F170, E171 y H365. El rol de estos aminoácidos fue evaluado al mutar cada uno de los residuos de manera individual por leucina, alanina, alanina y asparragina respectivamente. El estudio anterior permitió obtener las constantes catalíticas. Aplicando la ecuación de Eyring y asumiendo que el factor de transmisión no varía entre mutantes, los valores de $\Delta\Delta G^\ddagger$ se muestran en la tabla 3.1:

Tabla 3.1: Diferencia de energía libre de activación experimental en kJ/mol de las mutantes: N81L, F170A, E171A y H365N. Los valores son obtenidos a partir de las constantes catalíticas reportadas en ref⁷⁸. **N.D**: no determinado

Mutaciones	E171A	H365N	N81L	F170A
$\Delta\Delta G_{exp}^\ddagger$ (kJ/mol)	6.6	7.5	N.D	6.1

El hecho de obtener valores experimentales reportados y que el error asociado a la energía libre de activación relativa sea cercano a cero (dada la relación exponencial entre k_{cat} y $\Delta\Delta G^\ddagger$), vuelve a CCR un sistema idóneo para evaluar y validar el desempeño de la metodología propuesta.

A la fecha de hoy y a conocimiento de los autores este trabajo comprende la primera vez que se postula a las transformaciones alquímicas de no equilibrio como metodología para evaluar la energía libre de activación. Debido a esto, como primer paso, se propuso realizar un análisis de rendimiento para poder definir variables tales como:

- la velocidad de la transformación dado el tamaño de paso $\Delta\lambda$.
- el número de conformaciones de equilibrio a utilizar durante la etapa de

no-equilibrio

Para esto se utilizó la mutante F170A de la enzima CCR debido a que esta mutación implica la mayor perturbación en términos del volumen ocupado de la cadena lateral en el sitio activo. Además, existen estudios que proponen a este residuo como una pieza clave en mantener organizado el sitio catalítico⁷⁸. Todas estas características hacen de F170A un sistema desafiante dado que, como se mencionó en capítulos anteriores, la precisión de las transformaciones alquímicas dependerá del colectivo de conformaciones obtenidas en equilibrio y de la magnitud de la perturbación involucrada, en conjunto a la calidad del campo de fuerza.

Considerando un total de 4 diferentes tiempos de transformación alquímica y 3 tamaños de conjuntos de configuraciones de equilibrio para las transformaciones, la tabla 3.2 muestra el valor de $\Delta\Delta G^\ddagger$ para cada combinación de parámetros:

Tabla 3.2: Análisis de la robustez en la estimación de $\Delta\Delta G^\ddagger$ [kJ/mol], para el sistema F170A. Se evalúan diferentes velocidades de transformación y número de conformaciones utilizadas

Tiempo (ps)	100 Confs	150 Confs	300 Confs
80	0.7± 1.0	-0.2± 1.5	1.0 ±0.9
100	0.2± 0.5	-0.2 ±1.6	-0.2± 1.5
160	0.1±0.7	-0.1 ±1.4	-0.6± 1.2
200	0.3±1.0	-0.1 ±1.2	-0.4± 1.4

Los valores obtenidos corresponden a 3 simulaciones independientes de

F170A de 40ns cada una, donde se descartaron los primeros 15ns y se seleccionaron las conformaciones a transformar de los 25ns restantes. El protocolo anterior aplica tanto para el ET como para CMM en la enzima silvestre y la mutante A170, lo que implica un total de 4 simulaciones de equilibrio, donde se modificaron las cargas atómicas del sustrato en ambos estados y se utilizó la aproximación armónica para el estado de transición. Los valores reportados corresponden al promedio de las 3 réplicas con su respectiva desviación estándar. Los detalles de las simulaciones de equilibrio en conjunto a los parámetros para las transformaciones de no equilibrio, se encuentran en el anexo C. Como se observa en la tabla 3.2, los valores de $\Delta\Delta G^\ddagger$ no varían con el número de conformaciones a transformar o el largo de la simulación de transformación. Los valores oscilan alrededor de 0 kJ muy por debajo del valor experimental, pero dado que esta parte del estudio implica un análisis de la robustez del método, se limita a discutir sólo tal aspecto. En función de los resultados obtenidos y considerando el objetivo de este trabajo, para la validación del método se utilizará el conjunto de parámetros que mejor optimizan el protocolo, siendo el número de conformaciones y el largo de la transformación química 100 estructuras y 80 ps respectivamente. Cabe destacar que valores inferiores a los utilizados en este análisis, no logran convergencia en los resultados según reporta Aldeghi y colaboradores para transformaciones químicas de no equilibrio⁷⁵.

Una vez aplicado el análisis de robustez, se procedió a analizar el desem-

peño del protocolo en las 4 mutantes de CCR. Por motivos computacionales y dada la robustez de los resultados, en lugar de hacer 3 réplicas de 40 ns de las cuales se extraen un total de 75 ns, se optó por realizar una única réplica de 100 ns de los cuales se descartan los primeros 25 ns, tiempo asociado a la equilibración de los sistemas. Posteriormente se generaron las transformaciones alquímicas de 100 conformaciones extraídas de los 75 ns restantes. La elección anterior se realiza con el fin de poder abarcar un conjunto conformaciones sujetas a una distribución de Boltzmann en equilibrio. La tabla 3.3 muestra los valores de $\Delta\Delta G^\ddagger$ utilizando la parametrización armónica (rest) y la parametrización explícita de la contribución enlazante por medio el método de Seminario(sem).

Tabla 3.3: Estimación del cambio en la energía libre de activación para las diferentes mutantes de CCR. La descripción de las interacciones enlazantes en el estado de transición son modeladas usando restricciones mediante potenciales harmónicos externos [rest] o derivando nuevos constantes de fuerza mediante el método de Seminario [sem]

$\Delta\Delta G_{calc}^\ddagger$	E171A	H365N	N81L	F170A
Exp	6.6	7.5	N.D	6.1
q[rest]	$-3.8 \pm 1,0$	$5.0 \pm 1,0$	$6.0 \pm 1,0$	$-0.4 \pm 1,0$
q[sem]	$-0.3 \pm 1,0$	$0.3 \pm 1,0$	$-1.1 \pm 1,0$	$-1.9 \pm 1,0$

El primer sistema a analizar es la mutante E171A, el cual comprende la transformación de un glutamato cargado negativamente en un residuo de alanina. Este residuo se encuentra cercano al sitio activo e interacciona directamente con una histidina de la cual existe evidencia de encontrarse en

su estado doble protonado dado su valor de pK_a ¹⁴³. La interacción anterior es mediada principalmente por un puente salino, fuerza intermolecular que se pierde cuando el residuo de glutamato se sustituye por una alanina. Esta hipótesis repercute directamente en el cálculo de energía libre ya que según el estado de protonación de H365 se obtendrán distintos valores de energía libre. El valor que más se aproxima al resultado experimental corresponde al estado protonado más probable, concordando con hallazgos anteriormente encontrados en el grupo de investigación, lo que sustenta la hipótesis de la doble protonación.

El cálculo de $\Delta\Delta G^\ddagger$ mediante transformaciones alquímicas predice una estabilización del ET proponiendo a la mutante E171A como una mutación beneficiosa, en desacuerdo con la evidencia experimental. Esta discrepancia se vé reducida al aplicar la parametrización enlazante de Seminario, prediciendo que la mutante no genera un cambio significativo en la barrera de activación (-0.3 kJ/mol).

La segunda mutación corresponde a H365N, donde la histidina doblemente protonada mencionada en el párrafo anterior es sustituida por una asparagina. Dado que en ambas variantes (H365 y N365) el residuo de glutamato permanece desprotonado, la transformación implica un cambio neto de la carga total del sistema. Existe evidencia de artefactos generados por el tratamiento de largo alcance de las interacciones electrostáticas cuando se realizan transformaciones alquímicas de sistemas que no son neutros¹⁴⁴. Para sobre-

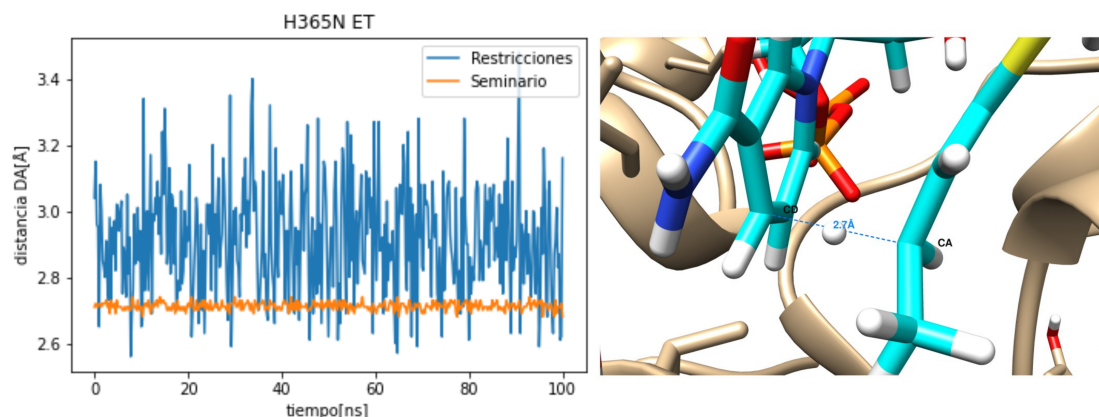


Figura 3.3: Distancia entre el carbono donador y aceptor de hidruro a lo largo de la simulación de equilibrio del estado de transición de la mutante H365N . La curva naranja representa el comportamiento del sistema que posee una parametrización enlazante mediante el método de Seminario. La curva azul, muestra las fluctuaciones en la distancia, cuando se utilizan potenciales armónicos externos para mantener la estructura del ET. La figura de la derecha muestra una conformación de equilibrio para el ET extraído del camino de mínima energía libre, indicando la distancia promedio de 2.7Å.

llevar este problema, el sistema es simulado en conjunto a un tripéptido con la cadena lateral central siendo el aminoácido opuesto al estado del sistema. Cuando se simula a la enzima silvestre, este tripéptido tendrá una asparagina y al momento de transformar la enzima silvestre en mutante, este residuo de asparagina se transformará a su vez en una histidina doblemente protonada, asegurando la neutralidad de la caja de simulación en todo momento.

El valor obtenido para esta mutante es de 5.0 y 0.32 kJ/mol, cuando se aplica la parametrización armónica y de Seminario respectivamente. Si bien pareciera que la parametrización de seminario empeora la predicción, la Figura 3.3 muestra que la distancia entre el carbono donador de NADPH y el

carbono aceptor del crotonil-CoA cuando se aplica el primer tipo de parametrización oscila muy por encima de la distancia requerida para considerarse un ET ($2.7\text{\AA}$), en comparación a la parametrización de Seminario. Lo anterior entrega evidencia de que independiente de obtener un valor cercano al experimental, no existe un correcto muestreo del colectivo de configuraciones que definen al estado de transición, lo que invalida la TST. Cabe destacar que los potenciales armónicos de restricción tienen constantes de fuerza asociadas de 3000 kJ/nm^2 , por lo que la imposibilidad de mantener tales distancias se debe principalmente a la corta distancia que caracteriza al ET y la pobre descripción del modelo. A modo de comparación, el método de Seminario exhibe distancia cercana a los $2.7\text{ \AA}$, en acuerdo con evidencia anteriormente reportada¹⁴³ y posee una baja oscilación alrededor de este valor.

Considerando los resultados expuestos en el párrafo anterior, para las siguientes dos mutantes, se limitará a analizar solo los resultados referentes a la parametrización mediante el método de Seminario. El tercer sistema comprende la sustitución de la asparragina en la posición 81 por un residuo de leucina. Desafortunadamente, esta mutante no posee un valor experimental reportado, debido a que se genera un aducto entre la coenzima y el crotonil-CoA, provocando el bloqueo del proceso catalítico. Esta evidencia experimental implica que en el sistema mutante, el reactante cruza hacia el intermediario una única vez, para luego permanecer en él, pero no da cuenta del tamaño de la barrera, por lo que no existe un valor de referencia para compa-

rar.

Sin embargo, en el marco de la tesis doctoral de Aharon Gomez en el laboratorio de investigación del guía de esta tesis, se comprobó que el residuo de asparagina estaría directamente involucrado en la generación de los sitios de unión de CO_2 , impidiendo que el aducto pueda avanzar hacia los productos.

El cuarto y último sistema corresponde a F170A. Como se explicó anteriormente esta mutante altera el sitio catalítico permitiendo la entrada de moléculas de solvente cuando el residuo 170 es sustituido por alanina. La mutante F170A corresponde al sistema con menor precisión y un valor de $\Delta\Delta G^\ddagger$ de -1.9 kJ/mol, corroborando el gran desafío que implica realizar una modificación altamente disruptiva, dado el tamaño del residuo a mutar. Lo anterior deja expuesta la interrogante de si el modelo de campo de fuerza es suficiente para una correcta predicción.

Si bien todos los valores de $\Delta\Delta G^\ddagger$ calculados no incluyen una reparametrización de los parámetros de LJ, debido a que no existe un análisis sistemático (no así las cargas), de cuál esquema de partición es el más adecuado para derivar tales parámetros, se decidió incluir el análisis de estos parámetros en el siguiente capítulo. Se puede concluir, que dados los modelos de campo de fuerza y esquemas de simulación discutidos hasta ahora, las transformaciones químicas predicen valores de $\Delta\Delta G^\ddagger$ con errores por sobre la precisión química (4.2kJ/mol). Se espera que una mejora de las interacciones no cova-

lentes mediante la reparametrización de las cargas atómicas y coeficientes de LJ pueden reducir la brecha entre los valores calculados y los de referencia, en concordancia con lo propuesto en la hipótesis de este trabajo.

3.2. Derivación de parámetros no enlazantes para aminoácidos mediante la teoría de átomos-en-moléculas (AIM)

El segundo objetivo de esta tesis comprende la derivación de parámetros no covalentes de las cadenas laterales de aminoácidos, el sustrato y la coenzima a partir de la densidad electrónica en el ambiente proteico. Lo anterior con el fin de mejorar la descripción de las interacciones en el sitio activo y evaluar su efecto en cálculos de la variación de energía libre de activación sobre el sistema CCR. Debido a que para obtener tales parámetros es necesario en un primer paso particionar la densidad electrónica y dado que existe más de un método para hacer esto, en este capítulo se expone la evaluación sistemática de una serie de esquemas de partición pertenecientes a la familia Hirshfeld. Los parámetros no enlazantes son utilizados para calcular las energías de interacción electrostáticas, dispersión y repulsión, donde los valores de referencia corresponden a cálculos de interacción basados en el método *Second Generation of Absolute Localized Molecular Orbital Energy Decomposition Analysis*¹⁴⁵ (ALMO-EDA2).

Los sistemas a evaluar comprenden dímeros de cadenas laterales de aminoácidos como también dipéptidos, como un modelo minimalista de interacciones en una proteína. Este conjunto de prueba corresponde a la base de

datos recientemente reportada por Carter-Fenk et al.¹⁴⁶, desarrollada para el estudio de interacciones $\pi - \pi$. Esta base de datos incluye pares de aminoácidos con carga positiva y negativa, hidrofílicos y aromáticos, representados en la Figura 3.4. Para seleccionar la configuración de los dímeros se utilizaron descriptores geométricos, que permiten definir a una interacción del tipo π utilizando distancias y ángulos¹⁴⁷. Este conjunto de datos fué seleccionado con el fin de evaluar los nuevos parámetros no enlazantes en configuraciones desafiantes, donde este tipo interacciones comprenden grupos funcionales problemáticos en terminos de parametrización. En adición, la distancia a la que ocurren las interacciones $\pi - \pi$ ocurren en el regimen donde las densidades electrónicas tienen un solapamiento considerable, haciendo que modelos de interacción como el Coulómbico o el de Lennard-Jones presenten dificultades al reproducir interacciones. Esto permite realizar una evaluación lo suficientemente estricta que permita identificar el conjunto de parámetros no covalentes más óptimo.

Los cálculos de química cuántica para obtener la densidad electrónica se realizaron utilizando el software ORCA 5¹⁴⁸ en el nivel de teoría $U\omega B97X-V/def2-TZVPD$ en fase gaseosa^{138,149}. Los archivos Molden generados son procesados con IODATA¹⁵⁰ para extraer números atómicos, coordenadas y los coeficientes de los orbitales moleculares. Los parámetros no enlazantes se calcularon utilizando herramientas de código abierto, incluyendo HORTON¹⁵¹, ChemTools^{152,153}, Grid¹⁵⁴, GBasis^{155,156} y una versión modificada

del paquete DensPart. Todas las densidades moleculares y promoleculares se calcularon utilizando una grilla de Becke–Lebedev implementada en HORTON 2, con una resolución *insane* y $\text{exp:}5\text{e-}4\text{:}2\text{e}1\text{:}175\text{:}230$ para la grilla molecular y atómica respectivamente¹⁵⁷.

Las energías de interacción de referencia se tomaron de Ref. 146, donde los cálculos son realizados utilizando Q-Chem 5.2¹⁵⁸ al mismo nivel de teoría que los cálculos en ORCA realizados en este trabajo. Las energías de interacción tipo campo de fuerza se calcularon con los parámetros no enlazantes derivados, de acuerdo con la ecuación 2.9, diferenciando entre las interacciones electrostáticas, atractivas de dispersión y repulsivas descritas por el potencial de Lennard-Jones. El código necesario para obtener dichas interacciones fue escrito en el marco de esta tesis.

3.2.1. Interacciones electrostáticas

El primer análisis consiste en evaluar la precisión de las cargas atómicas obtenidas mediante métodos AIM para reproducir la energía de interacción electrostática clásica del análisis ALMO-EDA2^{110,159} en cadenas laterales de aminoácidos. Este análisis emplea la densidad electrónica *frozen* (descrita en el capítulo 2.2.3) de monómeros, para calcular la interacción electrostática entre dímeros, excluyendo efectos de polarización y omitiendo efectos cuánticos como la exclusión de Pauli.

Cuatro esquemas de partición son evaluados correspondientes a: *Iterati-*

ARG	ASN	ASP	GLN	GLU	HIS	PHE	TRP	TYR	
✓	✓	✓	✓	✓	✗	✓	✓	✗	R
	✓	✓	✓	✓	✓	✓	✓	✗	N
		✓	✓	✓	✗	✓	✓	✗	D
			✓	✓	✓	✓	✓	✓	Q
				✗	✗	✓	✓	✓	E
					✓	✓	✓	✗	H
						✓	✓	✓	F
							✓	✓	W
								✓	Y

Figura 3.4: Pares de aminoácidos que forman 610 interacciones entre cadenas laterales en la base de datos de interacciones $\pi - \pi$ ¹⁴⁶. Estas interacciones se utilizan para evaluar la efectividad de los parámetros de fuerza no enlazante derivados de métodos AIM. Los símbolos *check* y *cross* indican la presencia o ausencia de interacciones entre pares respectivamente.

ve Hirshfeld¹⁶⁰ (HI), *Minimal Basis Iterative Stockholder*¹⁶¹ (MBIS) y los esquemas recientemente implementados en este trabajo *Additive Variational Hirshfeld* (AVH) y *Minimal Basis Additive Variational Hirshfeld* (AVH-M)¹⁶². En adición, y a modo de comparación con los parámetros establecidos, las energías de interacción son calculadas usando las cargas de los campos de fuerza AMBER99¹⁶³ y CHARMM36m¹⁴².

La distribución de los errores absolutos y los errores absolutos medios (MAE) en la Figura 3.5 muestran que los métodos AIM HI y MBIS, reproducen la energía electrostática clásica de ALMO-EDA con una precisión comparable al campo de fuerza AMBER99 y con mejoras respecto a CHARMM36, tanto en términos de MAE como de la mediana. Esta tendencia es especialmente notable en aminoácidos cargados (Figura 3.5A). Por otro lado, las interacciones entre cadenas laterales de carácter hidrofílico y aromático (Figura 3.5B) presentan errores absolutos y MAE menores debido a que poseen interacciones electrostáticas débiles.

Al analizar las energías de interacción electrostática calculadas con cargas atómicas de métodos AIM y campos de fuerza convencionales se observa que estas presentan desviaciones sistemáticas respecto a los valores de referencia (Anexo A.1-A.4). Estos errores se ven incrementados a medida que las cadenas laterales se acercan. El incremento en el error es atribuido al efecto de penetración de carga, ya que la aproximación de cargas puntuales no considera la superposición de densidades electrónicas en los dímeros. Esta

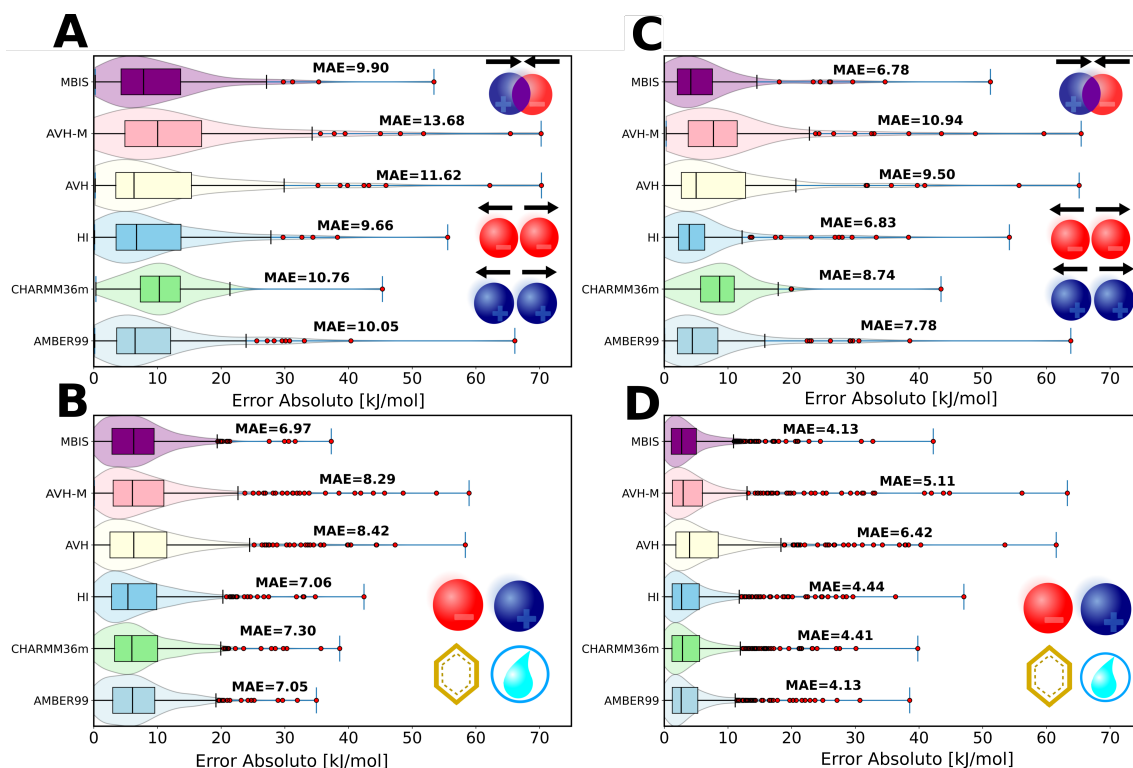


Figura 3.5: Errores absolutos en la energía de interacción electrostática clásica. Las combinaciones de cadenas laterales se agrupan según su naturaleza química: A) Dímeros de residuos cargados, B) Otras combinaciones entre residuos aromáticos, hidrofílicos y cargados. C) y D) muestran la misma distribución de errores absolutos que A) y B), pero considerando efectos de penetración de carga. El error absoluto medio (MAE) para cada método AIM o campo de fuerza se expresa en kJ/mol. La caja se extiende desde el primer cuartil (Q1) hasta el tercer cuartil (Q3), con una línea en la mediana. Los bigotes abarcan hasta el punto de datos más lejano dentro de 1.5 veces el rango intercuartílico (IQR). Los puntos atípicos corresponden a valores fuera del rango de los bigotes.

suposición toma particular sentido al considerar que este efecto se vuelve relevante a distancias cortas, donde la expansión multipolar atómica pierde validez. Es justamente en este rango, donde se observan las desviaciones y por lo tanto se puede aislar selectivamente el problema. Para demostrar esta hipótesis se utilizó una función de energía electrostática clásica que considerara el efecto de la penetración de carga, pero manteniendo el tratamiento clásico de cargas puntuales:

$$E_{elec}^{CP} = \sum_{A,B>A} \frac{Z_A Z_B}{r_{AB}} + \sum_{A,B>A} \frac{Z_A N_B}{r_{AB}} f_{damp}(r_{AB}) + \sum_{A,B>A} \frac{N_A Z_B}{r_{AB}} f_{damp}(r_{AB}) + \sum_{A,B>A} \frac{N_A N_B}{r_{AB}} f_{damp}^{overlap}(r_{AB}). \quad (3.2)$$

Donde f_{damp} son funciones de amortiguamiento que reescalan la interacción en función de la distancia entre los átomos. Z y N por su parte, corresponden a la carga nuclear y población electrónica asignada a cada átomo. En términos simples esta nueva función reescala cualquier interacción que incluya electrones, permaneciendo inalterada la repulsión núcleo-núcleo. Las Figuras 3.5C y D muestran que incluir la corrección por penetración de carga mejora significativamente la descripción de las interacciones electrostáticas entre cadenas laterales cargadas, polares y aromáticas. En dímeros cargados, la corrección impacta principalmente en los valores absolutos más altos, donde la atracción y repulsión son más fuertes (Anexo A.1 y A.2). Para interacciones más débiles, como las polares y aromáticas, la corrección también mejora la

precisión (Anexo A.3 y A.4). Tales resultados coinciden con estudios previos que identifican la penetración de carga como un factor clave en la descripción electrostática de interacciones $\pi - \pi$ en pares de bases de ADN y otros sistemas aromáticos.^{164,165}

La Figura 3.6 por su parte compara la energía electrostática calculada utilizando diferentes cargas atómicas que consideran la corrección por penetración de carga, agrupando las interacciones según su naturaleza química con el fin de tener un análisis más detallado. En esta figura se puede observar que las cargas de AMBER99 presentan un error similar a los métodos MBIS y HI. Por el contrario CHARMM36m muestra desviaciones marcadas, especialmente en las interacciones de residuos con cargas netas similares, donde sobreestima la repulsión. Esta sobreestimación puede deberse al protocolo de parametrización en CHARMM, donde las cargas atómicas son optimizadas para reproducir interacciones entre grupos polares y moléculas de agua⁹⁸. Este proceso se aleja de la metodología basada en una molécula en fase gaseosa (y, por lo tanto, de las propiedades derivadas de tales cálculos) al incorporar una respuesta de polarización implícita debido a la presencia de agua u otros solutos en el entorno químico. Entre los diferentes métodos AIM, MBIS y HI tienen un rendimiento similar, mientras que los métodos variacionales AVH y AVH-M muestran mayores desviaciones en los dímeros cargados que se repelen con cargas iguales.

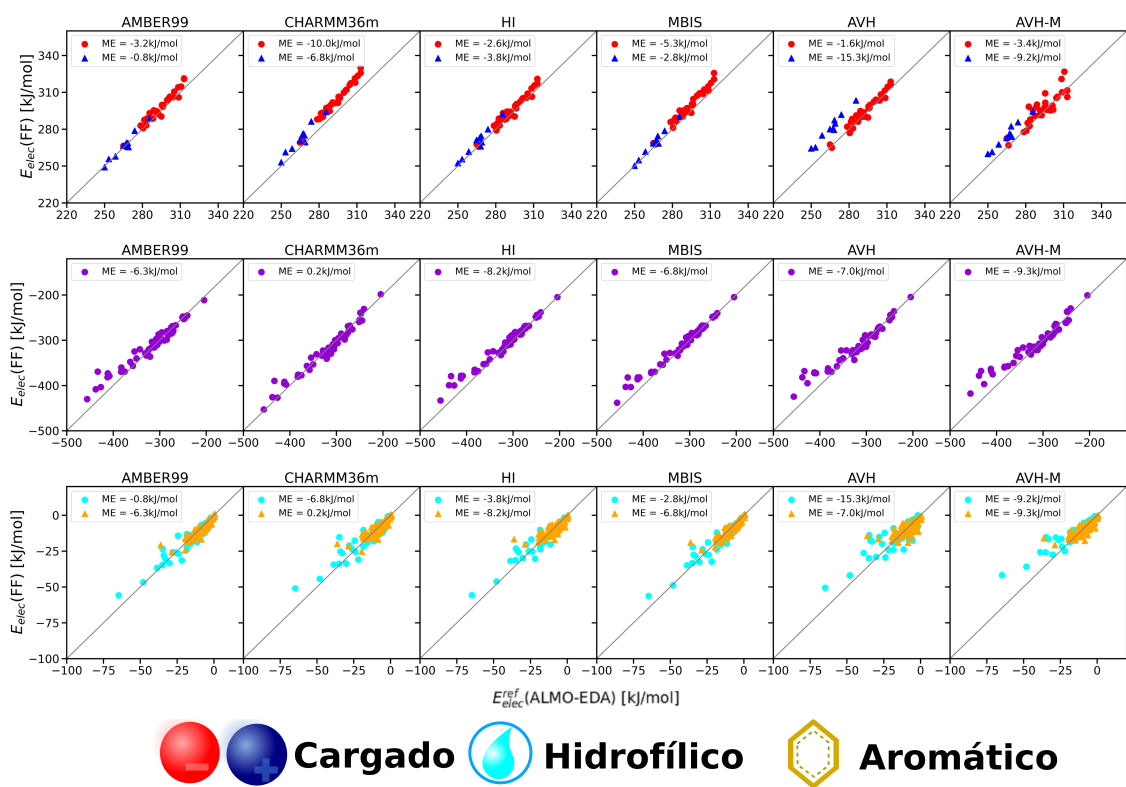


Figura 3.6: Gráfico de paridad de la energía de interacción electrostática clásica, comparando los valores de referencia ALMO-EDA2 con la interacción electrostática que incluye correcciones por penetración de carga. Las interacciones son separadas según los tipos: cargadas(+/-)-cargadas(+/-) (primera fila), cargadas(+)-cargadas(-) (segunda fila), aromáticas-aromáticas e hidrofílicas-hidrofílicas (tercera fila).

3.2.2. Interacciones de dispersión

El segundo tipo de interacción a evaluar corresponde a las interacciones de dispersión. Usando el modelo del oscilador de Drude, Pauli propuso que la interacción de dispersión entre átomos se representa por una combinación de términos que decaen con $1/r^n$ ($n = 6, 8, 10$). Cada término está relacionado con la polarizabilidad atómica debido a fluctuaciones instantáneas en la densidad electrónica, generando dipolos, cuadrupolos y multipolos de rango superior. Mientras que los campos de fuerza solo utilizan el primer término, el coeficiente C_6 , enfoques más rigurosos incluyen términos de orden superior, como la corrección D3 de dispersión en los cálculos de teoría de funcionales de densidad¹⁶⁶.

Usando la regla de combinación Lorentz-Berthelot para combinar los parámetros σ y ϵ entre los átomos A y B de la ecuación 2.9, la Figura 3.7 muestra que MBIS, uno de los sistemas AIM más efectivos para reproducir las energías de interacción electrostática de primeros principios, también es el que mejor reproduce las energías de interacción de dispersión, aunque con una sistemática subestimación reflejada en un error medio de -2.3 ± 0.9 kJ/mol. Tanto AMBER99 como CHARMM36 tienden a sobreestimar las energías de interacción de dispersión para dímeros fuertemente unidos (distancias cortas), alcanzando errores máximos de 11 y 15 kJ/mol, respectivamente. Es importante destacar que esta sobreestimación de la energía de dis-

persión también se observa en todas las demás metodologías AIM, a excepción de MBIS.

Un análisis de los parámetros derivados muestra que los coeficientes C_6 y los radios de Van der Waals R_{vdw} de los hidrógenos en MBIS son considerablemente más pequeños en comparación a los demás esquemas de partición (Anexo A.5 y A.5). Se hipotetiza que esto se debe a la forma en la que se construye la pro-densidad en MBIS utilizando funciones de Slater en comparación a las funciones gaussianas utilizadas en los otros métodos AIM, lo que aumentaría la flexibilidad y reproduciría el decaimiento de la densidad electrónica de mejor forma en estos átomos.

3.2.3. Interacciones de repulsión

Las interacciones de repulsión son un efecto cuántico asociado con la antisimetrización de la función de onda de los monómeros interactuantes. Este efecto aumenta a medida que se superpone la distribución electrónica de las dos moléculas. Los campos de fuerza representan la repulsión atómica utilizando radios efectivos de van der Waals, determinados según el tipo de átomo y su entorno químico. La suma de los radios de dos átomos determina la distancia mínima del potencial de van der Waals. Si las distancias son más cortas, la repulsión predomina, mientras que si son más largas, las interacciones atractivas se vuelven dominantes. A pesar de que un aumento exponencial de la repulsión a distancias menores al radio de van der Waals refleja más fiel-

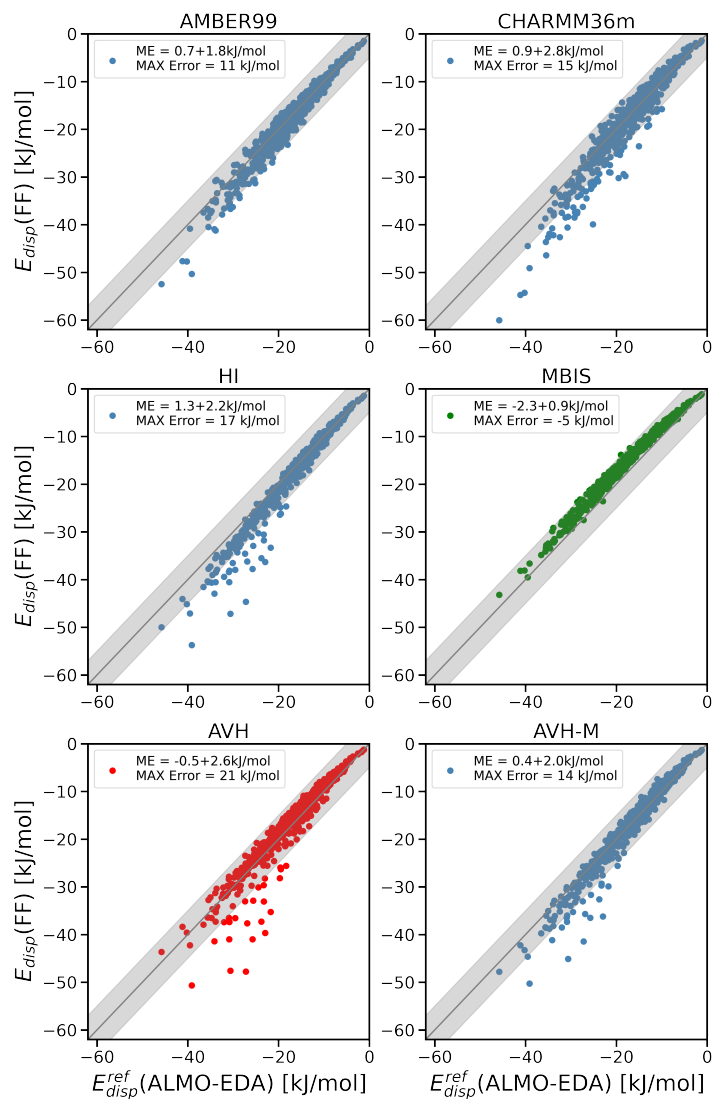


Figura 3.7: Gráfico de paridad de las energías de interacción de dispersión, comparando los valores ALMO-EDA2 con la energía de dispersión clásica calculada utilizando parámetros de LJ. Tales parámetros son combinados con la regla de combinación de Lorentz-Berthelot. Los subgráficos en verde (MBIS) y rojo (AVH) representan los mejores y peores esquemas, respectivamente, según el error medio, la desviación estándar y el error máximo.

mente los resultados ab-initio, los campos de fuerza para proteínas han mantenido predominantemente el término $\frac{1}{r_{AB}^{12}}$ del potencial de Lennard-Jones, principalmente por razones computacionales. En un principio los campos de fuerza parametrizaban los radios de van der Waals por métodos de ajuste según propiedades termofísicas de los fluidos. Sin embargo, Tkatchenko sugirió recientemente una técnica para derivarlos desde primeros principios utilizando enfoques de átomos en moléculas¹³¹. Recientemente, en el laboratorio del profesor guía de esta tesis se demostró que los radios de van der Waals derivados de primeros principios reproducen las densidades y entalpías de evaporación de líquidos orgánicos neutros⁹⁹.

Se utilizan enfoques AIM para evaluar los radios de van der Waals en campos de fuerza para proteínas, asegurando la consistencia con el potencial de Lennard-Jones. Al calcular la parte repulsiva del potencial de Lennard-Jones, se compara esta energía de repulsión con la energía de repulsión de Pauli obtenida mediante ALMO-EDA. Dado que esta contribución es la única que no se calcula explícitamente en ALMO-EDA2 (ecuación 2.16), se compara con las contribuciones derivadas de la teoría de perturbaciones adaptadas a la simetría (SAPT) a orden cero y segundo orden, para una serie de dímeros que se presentan en la tabla 3.4.

Cabe destacar que la energía de Pauli de ALMO-EDA2 está más cercana al valor obtenido por SAPT2+ que SAPT0, demostrando que los valores de referencia son confiables.

Tabla 3.4: Comparación de las contribuciones de energía (electrostática, dispersión, intercambio y energía total de interacción) en kJ/mol obtenidas a partir de los cálculos de ALMO-EDA 2 y sSAPT0 para seis pares de dipeptidos. Los valores entre paréntesis corresponden a la contribución de intercambio de cálculos SAPT2+.

	TRP-ARG	GLN-ARG	ARG-GLU	ASP-GLN	GLN-TRP	GLU-TRP
Elec(ALMO)	-25.8	-57.8	-200.6	19.9	1.2	-17.1
Elec(sSAPT0)	-25.8	-59.3	-199.3	20.8	1.8	-17.4
Δ Elec	0.0	1.5	-1.3	-0.9	-0.6	0.3
Disp(ALMO)	-48.8	-49.6	-6.9	-12.2	-11.4	-27.4
Disp(sSAPT0)	-48.8	-48.8	-5.4	-9.3	-9.9	-25.9
Δ Disp	0.0	-0.8	-1.5	-2.9	-1.5	-1.5
Pauli(ALMO)	76.3	74.8	3.5	6.7	4.7	28.7
Pauli(sSAPT0)	73.3(80.0)*	66.0(78.4)*	2.9(3.5)*	5.8(6.7)*	4.1(5.0)*	28.2(31.0)*
Δ Exch	3.0(-3.7)*	8.8(-3.6)*	0.6(0.0)*	0.9(0.0)*	0.6(-0.3)*	0.5(-2.3)*
Total(ALMO)	-23.8	-71.0	-220.0	9.2	-7.4	-24.3
Total(sSAPT0)	-19.1	-64.1	-214.1	13.0	-5.3	-22.0
Δ Total	-4.7	6.9	-5.9	-3.8	-2.1	-2.3

La Figura 3.8 muestra que tanto los campos de fuerza CHARMM36m como AMBER99 subestiman la energía de repulsión de Pauli en las configuraciones de dímeros de cadenas laterales. Las desviaciones más significativas se encuentran en los dímeros que contienen monómeros cargados con signos opuestos o una mezcla de monómeros hidrofílicos y cargados a corta distancia, donde la energía de repulsión de Pauli de referencia supera los 50 kJ/mol. Curiosamente, para los mismos dímeros, las energías de interacción electrostática atractiva fueron subestimadas debido a la penetración de carga. Por lo tanto, en ciertos escenarios, el efecto combinado de las interacciones repulsi-

vas y electrostáticas atractivas puede coincidir con los energías de interacción ab-initio de referencia debido a la cancelación de errores¹⁶⁷.

Para asegurar la compatibilidad con los campos de fuerza para proteínas, se postula MBIS como el esquema de partición para derivar parámetros de LJ, ya que muestra tendencias similares a las de los campos de fuerza convencionales. Sin embargo, HI podría servir como un método alternativo para derivar parámetros que incrementen el componente repulsivo en el potencial de Lennard-Jones. No obstante, debido a la sobreestimación a distancias cortas se requieren más pruebas y análisis conformacionales.

3.2.4. Interacciones dipéptido-dipéptido

Una vez evaluado los nuevos parámetros en dímeros de cadenas laterales, se extiende el análisis considerando las interacciones dipéptido-dipéptido como un modelo minimalista de una proteína que contiene tanto la cadena lateral como el esqueleto carbonado. Como se muestra en la Figura 3.9 para el dímero de arginina-glutamato, la energía total de interacción (sin incluir la polarización) es sobreestimada, modelando una fuerza sobreatractiva por los dos campos de fuerza convencionales en comparación con los valores de referencia. El mayor aporte al error proviene de una interacción electrostática excesivamente favorable y una subestimación de la repulsión de Pauli. Las energías de interacción calculadas utilizando parámetros derivados de AIM mejoran la descripción de la electrostática, con una ligera subestimación en

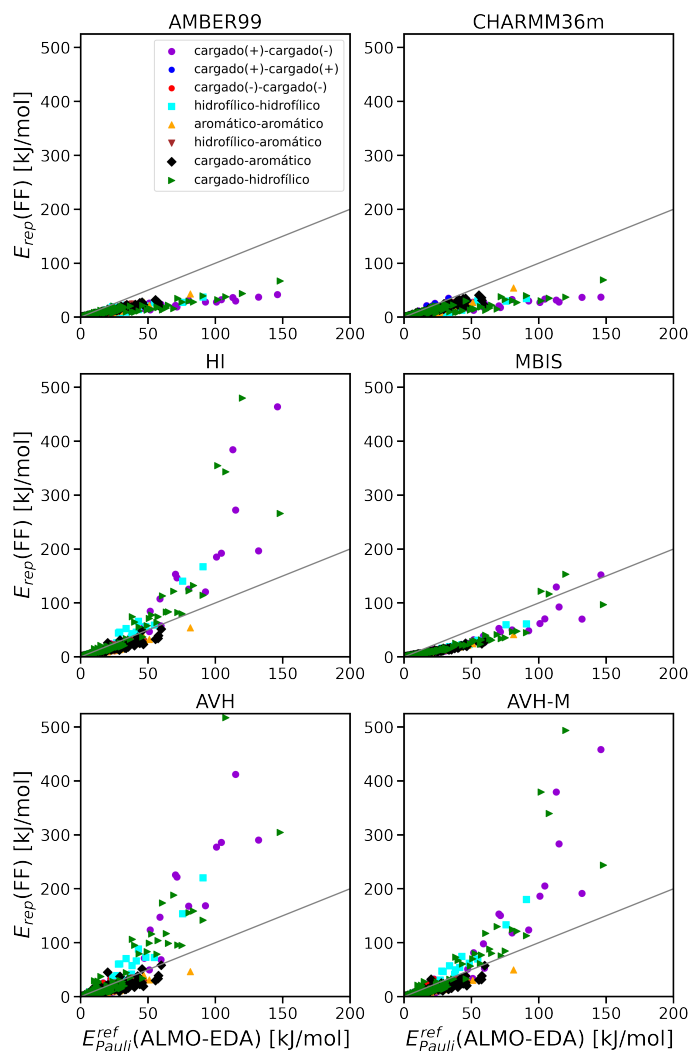


Figura 3.8: Gráfico de paridad de la energía de interacción de repulsión por intercambio, comparando los valores de ALMO-EDA con el término repulsivo del potencial de Lennard-Jones calculado utilizando parámetros de campos de fuerza y diversos métodos AIM. La leyenda muestra la relación entre el color del punto y la naturaleza de los residuos a interacciones.

ambas fuerzas, la interacción electrostática y la repulsión de Pauli. Curiosamente, el error total de interacción es inferior a 1 kJ/mol para HI y MBIS, debido principalmente a la compensación de errores.

En el dímero de triptófano-asparagina, tanto los campos de fuerza convencionales como las cargas AIM describen una interacción electrostática más favorable que el valor de referencia, donde la desviación puede ser resultado de la representación de la densidad electrónica mediante cargas atómicas puntuales en este tipo de sistemas aromáticos, donde la inclusión de multipolos puede ser crucial. En especial, los átomos de nitrógeno y oxígeno con pares de electrones libres poseen un momento dipolar atómico marcado. El error en la interacción de dispersión es pequeño para todos los métodos, excepto para MBIS. Las interacciones de repulsión son subestimadas por todos los métodos, aunque más severamente por los campos de fuerza convencionales y MBIS, tal como se observa en el análisis de las cadenas laterales. Al igual que para los dipéptidos cargados, MBIS presenta el menor error en la energía total de interacción, principalmente por la compensación de errores.

Un quinto esquema llamado c-HI comprende la versión variacional de HI donde las cargas atómicas del esqueleto carbonados son restringidas a reproducir las cargas de AMBER99 para incrementar la compatibilidad de los nuevos parámetros. El desempeño de c-HI es similar a HI y MBIS, sin embargo dependiendo el tamaño del sistema el costo computacional del proceso de derivación puede volverse relevante. En el futuro se espera poder optimizar

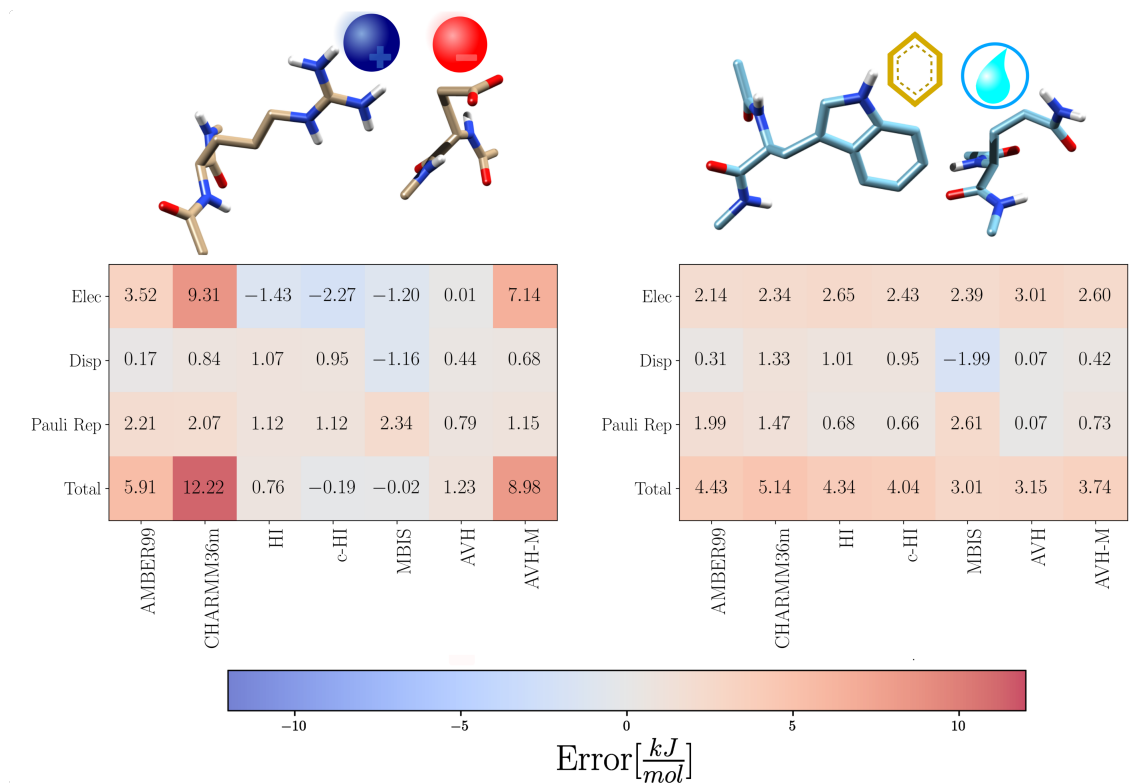


Figura 3.9: Errores en la energía de interacción electrostática, de dispersión y de repulsión de Pauli, en conjunto al error en la energía total de interacción para dos dímeros de dipeptidos. Las energías de referencia corresponde a valores obtenidos mediante ALMO-EDA2.

el código mediante paralelización y soporte con unidades de procesamiento gráfico para poder acelerar el proceso de derivación en métodos que incluyen condiciones.

A modo de conclusión, se tiene que MBIS califica indudablemente como el esquema de partición de elección para derivar cada uno de los parámetros no covalentes utilizados en los campos de fuerza. Los resultados obtenidos en este capítulo permiten decidir el esquema a utilizar en el siguiente capítulo

en el que una reparametrización de la coenzima, el reactante y los residuos a mutar son llevadas a cabo secuencialmente con el fin de evaluar su impacto en el cambio de energía libre de activación.

3.3. Evaluación del efecto de nuevos parámetros no-covalentes sobre la estimación de $\Delta\Delta G^\ddagger$

Hasta el momento los resultados de las transformaciones químicas han tenido un desempeño cuestionable con errores por sobre los 4.2 kJ/mol y prediciendo el efecto contrario dado el signo del $\Delta\Delta G^\ddagger$ experimental. En la sección anterior se obtuvieron resultados que demostraban el potencial del método de partición MBIS para derivar parámetros no covalentes que reproducen energías de interacción calculadas mediante primeros principios. Con la evidencia recopilada, en este capítulo se presentan los resultados en la predicción de $\Delta\Delta G^\ddagger$ cuando se reparametrizan las interacciones no covalentes del campo de fuerza.

En un primer intento, las interacciones no covalentes del ligando y la coenzima fueron modeladas mediante parámetros de Lennard-Jonnes derivados del campo de fuerza CGENFF y cargas atómicas derivadas del método MBIS. Se denomina a este enfoque, el modelo electrostático, ya que se asume que sólo este tipo de interacción se ve modificado en el estado de transición

en comparación al complejo de Michaelis-Menten y comprende los resultados obtenidos en la sección 1 de este capítulo. En una segunda iteración se incluye una modificación de los parámetros de LJ tanto en el estado de transición como en los reactantes a través del esquema de partición MBIS, dando origen al modelo q/LJ.

La tercera fila de la Tabla 3.5 (q/LJ[sem]) muestra una importante mejora al añadir los nuevos parámetros de LJ en comparación al modelo puramente electrostático. Lo anterior se ve reflejado en sus errores menores a 2 kJ/mol para cada una de las mutantes. Una posible explicación a esta mejora radica en la co-dependencia de los parámetros no covalentes en los campos de fuerza convencionales. Dentro del campo de fuerza general de CHARMM, CGENFF, los parámetros de LJ son obtenidos mediante procesos de ajustes para reproducir propiedades termodinámicas. Por lo tanto existen al menos dos razones atribuibles a este efecto: 1) el campo de fuerza nunca fue diseñado para modelar interacciones en configuraciones fuera del equilibrio como el ET y 2) su validez radica en la utilización de las mismas cargas atómicas con las cuales fueron realizados los procesos de ajuste.

Realizando un análisis por sistema, se vé que la mutante E171 posee una mejora substancial de -0.3 a 5.2 kJ/mol, acercándose al valor experimental reportado y prediciendo una disminución de la velocidad de reacción en acuerdo con la evidencia. H365N por su parte posee el valor más cercano al valor de referencia con una diferencia de solo 0.3 kJ/mol entre el valor medio y

$\Delta\Delta G^\ddagger$ [kJ/mol]	E171A	H365N	N81L	F170A
Exp	6.6	7.5	N.D	6.1
q[sem]	-0.3 ± 1.0	0.3 ± 1.0	1.1 ± 1.0	-1.9 ± 1.0
q/LJ[sem]	5.2 ± 1.0	7.2 ± 1.0	-8.3 ± 1.0	4.6 ± 1.0
q/LJ[sem] q[aa]	11.6 ± 1.0	14.6 ± 1.0	-5.2 ± 1.0	-1.4 ± 1.0
q/LJ[sem] q/L[aa]	7.4 ± 1.0	8.8 ± 1.0	-3.0 ± 1.0	2.6 ± 1.0

Tabla 3.5: Cambios en la energía libre de activación para diferentes mutantes de CCR en kJ/mol. Los términos q y LJ representan reparametrización de cargas y parámetros de Lennard-Jones respectivamente. La descripción de las interacciones enlazantes en el estado de transición es obtenida derivando nuevos parámetros enlazantes [sem]. La nomenclatura [aa] indica una reparametrización del aminoácido mutado.

el experimental. N81L parece verse estabilizado aún más con este nuevo set de parámetros alcanzando un valor de -8.3 kJ/mol. Finalmente la mutante F170A exhibe una mejora similar a las otras mutantes con una diferencia de sólo 1.5 kJ/mol y prediciendo el comportamiento correcto.

En un intento por obtener valores aún más cercanos a los valores de referencia, se propuso reparametrizar los aminoácidos que son transformados en los procesos alquímicos. Los resultados corresponden a las filas 4 y 5 de la tabla 3.5. Al igual que en la parametrización del ligando y el sitio activo, la inclusión de sólo nuevas cargas atómicas en los aminoácidos no genera una mejora. Por el contrario, las mutaciones que implican transformaciones de residuos cargados exhiben sobreestimaciones en los valores de $\Delta\Delta G^\ddagger$, mientras

que la mutante F170A presenta una subestimación. Los resultados anteriormente discutidos confirman en cierto modo la hipótesis de que un cambio sólo en la contribución electrostática manteniendo inalterada los parámetros de LJ no lleva a una mejora en las predicciones.

Como análisis final, se tiene la última fila de la tabla 3.5, donde se reparametrizan los parámetros de LJ en conjunto a todos los cambios de las filas anteriores. Los resultados son comparables a los obtenidos cuando se reparametrizan q y LJ en las especies reaccionantes (fila 3), exhibiendo una ligera mejora en E171A, pero disminución de la precisión en N365N y F170A.

A modo de conclusión, se puede afirmar que las transformaciones químicas de no-equilibrio permiten predecir de manera correcta los cambios en las energías libres de activación producto de mutaciones en el sistema CCR, siempre y cuando se realiza una derivación de parámetros no covalentes utilizando el método MBIS. Para reproducir este comportamiento el reactante puede ser descrito simplemente con parámetros enlazantes obtenidos desde CGENFF, mientras que el estado de transición debe ser descrito en su totalidad con parámetros derivados desde primeros principios. Los valores de $\Delta\Delta G^\ddagger$ calculados exhiben diferencias menores a 4.184 kJ/mol alcanzando precisión química. Para poder evaluar de manera sistemática la importancia de la parametrización de los residuos a mutar en el siguiente capítulo se validará la metodología con un segundo sistema, la Dihidrofolato Reductasa (DHFR).

3.4. Validación de la metodología en un segundo sistema enzimático: Dihidrofolato Reductasa (DHFR)

La enzima dihidrofolato reductasa cataliza la transferencia de hidruro desde la coenzima NADPH a una molécula de dihidrofolato, produciendo tetrahydrofolato. Esta enzima ha sido ampliamente estudiada durante más de 40 años debido a su papel en diversas rutas metabólicas ligadas a la embriogénesis y su importancia como diana de fármacos anticancerígenos y antibióticos. Debido a ello, se dispone de una gran cantidad de datos con alta precisión sobre su estructura, propiedades enzimáticas^{168,169} y filogenia¹⁷⁰. Sumado a esto, existen numerosos estudios computacionales donde han analizado su mecanismos de reacción y actividad enzimática.^{171–174}

También se ha estudiado el impacto de 3 mutantes en el residuo 14 de isoleucina (I14) sobre la energía libre de activación. Las mutantes corresponden a valina (V), alanina (A) y glicina (G), disminuyendo sistemáticamente el volumen del residuo en cada mutación (ver Figura 3.10). Los resultados experimentales muestran que una disminución en el volumen del residuo dificulta el acercamiento de los reactantes dificultando la posibilidad de producir la transferencia del hidruro. Esto repercute en un aumento de la energía libre de activación en proporción a la pérdida del volumen en el residuo

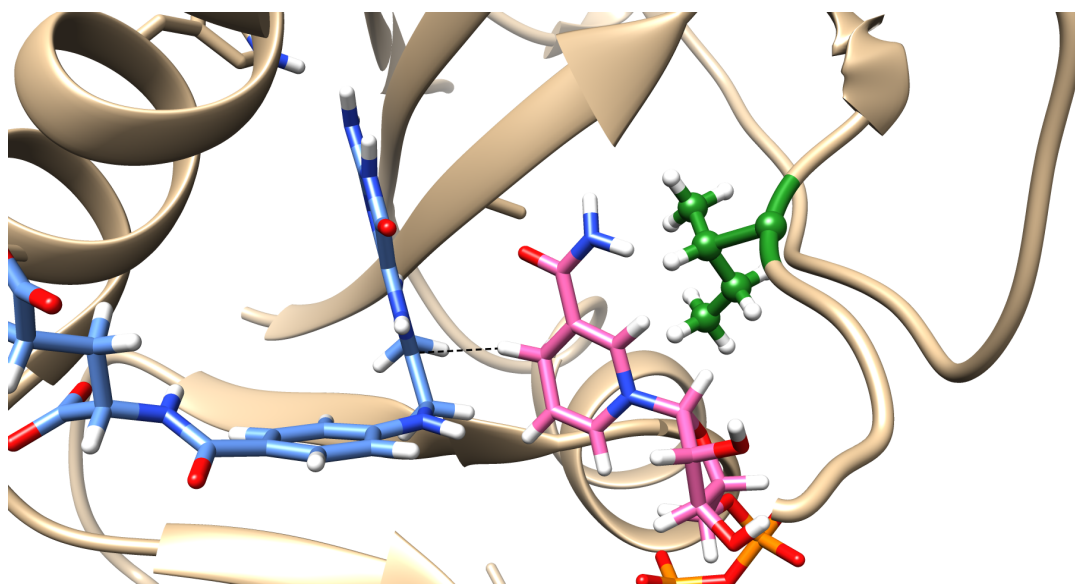


Figura 3.10: Sitio activo de Dihidrofolato Reductasa. El residuo 14 de Isoleucina es coloreado en verde, mientras que el sustrato dihidrofolato y la coenzima NADPH son coloreados en azul y rosa respectivamente.

14, con valores calculados mediante simulaciones QM/MM utilizando métodos semi-empíricos, como también valores experimentales⁶⁶ extraídos de las constantes catalíticas k_{cat} utilizando la ecuación de Eyring (ec. 2.6). Esta serie de características convierte a DHFR en un modelo ideal para realizar una segunda validación de la metodología propuesta en esta tesis. En los siguientes párrafos se detallan los resultados obtenidos en este segundo sistema enzimático. Aunque no es estrictamente necesario determinar un camino de reacción para la enzima DHFR, dado que su paso limitante es conocido de la literatura y además la reacción ocurre en un solo paso, es necesario obtener geometrías representativas de este estado para aplicar las transformaciones alquímicas. El camino de mínima energía libre fue obtenido mediante el mé-

todo de la cuerda adaptativa utilizando el método semi-empírico DFTB3⁹⁵ para modelar la zona QM y el campo de fuerza CHARMM36m¹⁴² es utilizado para modelar las interacciones en la zona MM. Todas las simulaciones fueron realizadas utilizando el paquete de simulación AMBER18¹⁷⁵, y las correcciones a las energías DFTB3 fueron calculadas utilizando el programa Gaussian16 con el nivel de teoría M06-2X¹⁷⁶ y la base def2-TZVP.

La estructura inicial corresponde al estado de reactante obtenido del trabajo de Luk y colaboradores¹⁷⁷. donde se utilizaron los mismos átomos para definir la zona QM. Este estado es minimizado a nivel MM mediante el algoritmo *steepest gradient descent*¹⁷⁸ para disminuir los impedimentos estéricos en las zonas QM y MM. Luego, se realizó una equilibración QM/MM bajo el colectivo NVT a una temperatura de 300 K durante 200 ps con un tamaño de paso de 1 fs.

A partir del reactivo relajado, se buscó la configuración de producto mediante el método de *scan*¹⁷⁹, utilizando la distancia antisimétrica (DH-HA) como coordenada de reacción siendo D el carbono donador en NADPH, H el hidruro a transferir y A el carbono aceptor en dihidrofolato respectivamente. El sistema se minimizó a lo largo de esta coordenada desde los valores -1.4Å(reactivo) hasta 1.4Å, aplicando un potencial de restricción de posición a los átomos a más de 10 Å del hidruro (H4B). Esta restricción de posición se utiliza principalmente para reducir los grados de libertad de los átomos que no están participando directamente de la reacción y así reducir el cos-

to computacional. Los demás átomos fueron minimizados hasta alcanzar el producto en la coordenada 1.4 Å. Luego, el producto fue equilibrado en las mismas condiciones que el reactivo durante 200 ps.

Finalmente, con ambos estados equilibrados, estos son utilizados para determinar el camino de mínima energía libre mediante el método de la cuerda adaptativa. En este paso se utilizaron un total de 40 nodos para describir el camino de mínima energía libre (MFEP por sus siglas en inglés).

Un desafío que se presentó en este paso fue la dificultad de DFTB3 para modelar correctamente la energía potencial cercana a la región del estado de transición. Particularmente para reacciones que involucran transferencias de hidruros se ha encontrado que DFTB3 genera mínimos artificiales cercanos a la geometría del ET¹⁸⁰. Para sobrellevar esto, se han propuesto correcciones realizando cálculos de alto nivel como DFT, los cuales son utilizados para corregir los valores de energía mas no la geometría⁹⁵. En este trabajo el perfil de energía libre fué corregido sumando la diferencia energética entre valores de energía DFTB3 y DFT M06-2X/def2-TZVP para distintos puntos a lo largo del camino de reacción. Estos puntos son obtenidos realizando un *scan* que utiliza los nodos y las variables colectivas de la cuerda, para obtener la coordenada de reacción. La diferencia de energía en cada uno de los nodos es sumada al MFEP lo que elimina el artefacto mencionado anteriormente (Anexo B.1). En adición a este tipo de corrección, en colaboración con investigadores de la Universidad de Talca y Valencia, se probó un enfoque de

corrección más general¹⁸⁰. Este nuevo método consiste en la corrección de los potenciales de repulsión C-C y C-H de la energía DFTB3 (ver ecuación 2.8).

Este método de corrección se aplicó en DHFR para obtener el MFEP obteniéndose el perfil de reacción de la figura 3.4B. La energía libre de activación de 16.8 kcal/mol está en perfecto acuerdo con el valor reportado por Doron y colaboradores⁶⁶. La geometría del estado de transición se caracteriza por una muy corta distancia de 2.7Å entre el carbono donador y el aceptor. Cabe destacar que éste método permite una alternativa generalizada de corrección a sistemas que transfieran hidruros, sin la necesidad de corregir las energías DFTB3 con cálculos DFT que pueden ser computacionalmente costosos.

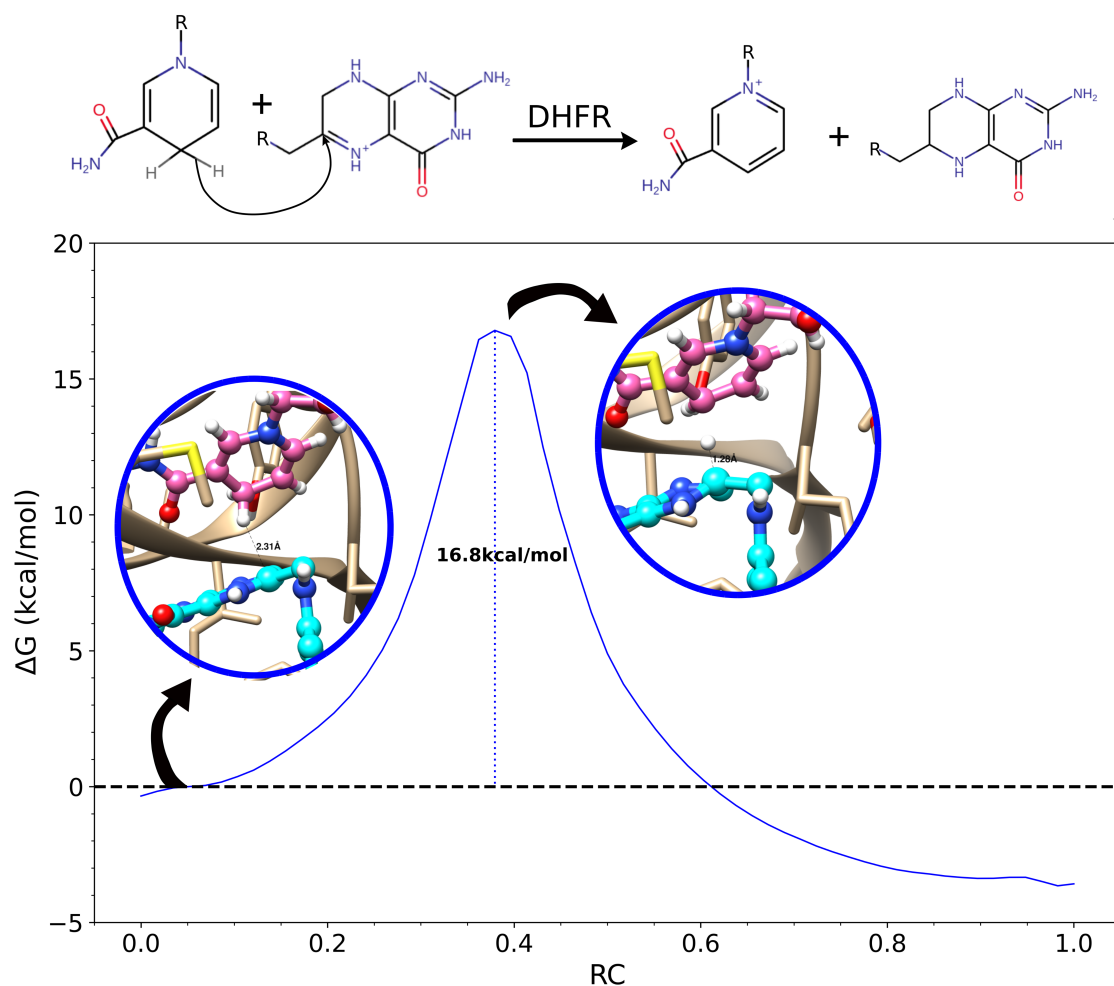


Figura 3.11: Perfil de energía libre de reacción del sistema DHFR a lo largo del camino de mínima energía libre obtenido por el método de la cuerda adaptativa. El mecanismo de la reacción es mostrado en la zona superior del gráfico. Los círculos muestran la organización del sitio activo en el complejo de Michaelis-Menten y el estado de transición. El valor de la energía libre de activación es obtenido de la diferencia entre el valor de energía libre en el reactante y el punto más alto de la curva.

Conocidas las geometrías del estado de transición y del reactante, parámetros enlazantes y no enlazantes fueron derivados bajo las mismas condi-

ciones que las utilizadas en la validación en CCR. Para comparar el método, dos conjuntos de datos son obtenidos de estudios previos. El primer conjunto corresponde a cambios en la energía libre de activación obtenidos de análisis cinéticos experimentales⁶⁶ y el segundo corresponden a energías libres calculadas mediante cálculos QM/MM utilizando el método semiempírico AM1-SRP⁶⁶. A diferencia de CCR y en vista de los resultados de aquel sistema, se utilizaron sólo dos modelos de campo de fuerza para el cálculo de $\Delta\Delta G^\ddagger$ mediante transformaciones alquímicas: A) el modelo q-LJ[sem] que reparametriza todos los componentes no covalentes de las especies en reacción en adición a la reparametrización de la parte enlazante en el estado de transición; B) a este mismo modelo se le agrega la modificación de los parámetros no covalentes de los campos de fuerza que describen los residuos a mutar (q/LJ[sem]q/LJ[aa]).

La tabla 3.6 muestra que el primer modelo de campo de fuerza reproduce correctamente la tendencia de las mutaciones (columna 4). Si bien los errores están por sobre los 4 kJ/mol en comparación a los datos experimentales, las transformaciones alquímicas son capaces de capturar el efecto del tamaño del residuo y su repercusión en la energía de activación. Dado que este efecto está principalmente mediado por interacciones repulsivas entre especies hidrofóbicas, es esperable que todas las predicciones presenten un error sistemático. Lo anterior está en acuerdo con la subestimación de la energía en los campos de fuerza al analizar las interacciones repulsivas de las cade-

nas laterales (capítulo 3.2). Además, aunque el método MBIS aún subestima esta contribución energética, lo hace en menor medida. Este efecto se manifiesta claramente al evaluar el impacto de la modificación de los parámetros no enlazantes de los residuos mediante el método $q/LJ[sem]q/LJ[aa]$. En este caso, se observa una ligera mejora en I14G e I14A, manteniendo una subestimación en comparación a los valores experimentales pero disminuyendo los errores. En particular I14A exhibe un error muy pequeño de tan sólo 1.6 kJ/mol mientras que I14G e I14V presentan errores de 4.8 y 4.4 kJ/mol respectivamente.

Este hallazgo es de suma importancia ya que sustenta la hipótesis sobre la importancia de un campo de fuerza que logre reproducir propiedades microscópicas como energías de interacción en lugar de propiedades macroscópicas que dependen de un colectivo de configuraciones.

En vista de los resultados obtenidos en los capítulos 3.3 y 3.4, se puede validar que las transformaciones alquímicas de no-equilibrio presentan una alternativa de menor costo computacional, económico y precisa para la predicción de cambios en la energía libre de activación producto de mutantes. En el siguiente capítulo se presentarán los resultados de la aplicación del método a otro caso de estudio: la *Old Yellow Enzyme* (OYE). Aquí se utilizará por primera vez el método en la búsqueda de mutantes catalíticamente más veloces que una enzima de referencia, generando así una alternativa para el campo de la ingeniería de proteínas.

	$\Delta\Delta G_{exp}^{\ddagger}$ [kJ/mol]	$\Delta\Delta G_{calc}^{\ddagger}$ [kJ/mol]	$\Delta\Delta G_{calc}^{\ddagger}$ [kJ/mol]	$\Delta\Delta G_{calc}^{\ddagger}$ [kJ/mol]
Mutante	Ref	Ref QM/MM	q/LJ [sem]	q/LJ [sem] q/LJ[aa]
I14G	$16.7 \pm 0,0$	$18.0 \pm n.r.$	$10.5 \pm 1,5$	$11.9 \pm 2,3$
I14A	$8.8 \pm 0,0$	$12.3 \pm n.r.$	$4.6 \pm 1,5$	$7.2 \pm 1,3$
I14V	$4.6 \pm 0,0$	$7.1 \pm n.r.$	$0.5 \pm 1,0$	$0.2 \pm 0,4$

Tabla 3.6: Cambio en la energía libre de activación para diferentes mutantes de DHFR. Los términos q y LJ representan reparametrización de cargas atómicas y parámetros de Lennard-Jonnes respectivamente. La descripción de las interacciones enlazantes en el estado de transición y reactante son modeladas derivando nuevos parámetros enlazantes [sem]. La nomenclatura [aa] indica una reparametrización del aminoácido mutado. Los errores de valores de referencia QM/MM no fueron reportados

3.5. Obtención del camino de reacción para la enzima Old Yellow Enzyme (OYE)

El tercer objetivo de esta tesis comprende la caracterización del mecanismo de reacción de la enzima *Old Yellow Enzyme*, particularmente de la mutante Y187F. En el marco de una colaboración con el Prof. Tobias Erb del MPI for terrestrial microbiology en Marburg Alemania, nuestro laboratorio estudió la enzima Old Yellow Enzyme de *E. Coli*. El objetivo del proyecto consiste en estudiar la mutante artificialmente creada Y187F, la cual permite a la enzima catalizar una reacción aldólica con formaldehído que no realiza la

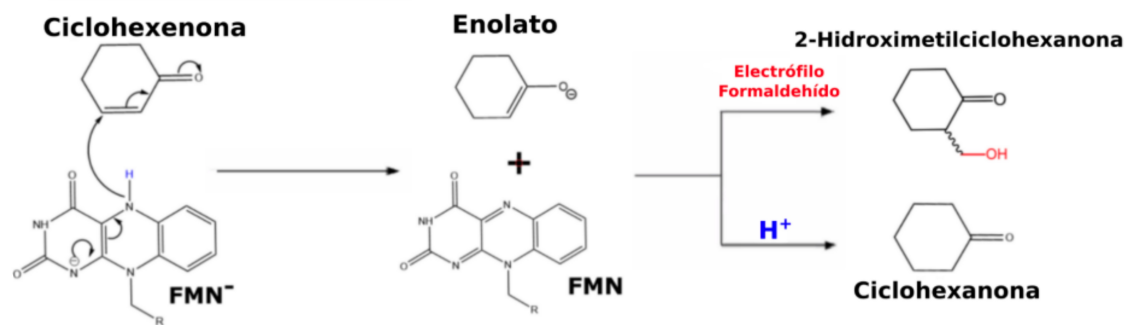


Figura 3.12: Mecanismo de reacción de la variante Y187F de *Old Yellow Enzyme*. La porción aromática de FMN reacciona con el sustrato ciclohexen-2-ona, para dar paso a la formación de enolato. Este intermediario da paso al producto reducido o aldólico dependiendo de si realiza un ataque nucleofílico sobre formaldehído o un protón.

enzima en su estado natural. La figura 3.12 muestra el mecanismo de la reacción en el cual la primera etapa consiste en una transferencia de un hidruro desde la coenzima Flavin Mononucleótido (FMN) hacia la 2-ciclohexenona formando un enolato. Este enolato a su vez reacciona con un protón generando el producto reducido ciclohexenona, siendo OYE naturalmente una reductasa. Estudios estructurales y de carácter computacional han postulado que el protón reaccionante en el segundo paso proviene de una molécula de H_2O o bien del residuo 187 de tirosina⁸⁰. Dada la evidencia encontrada, mutaron esta tirosina por un residuo de fenilalanina removiendo el carácter ácido del residuo, imposibilitando en cierta medida la reducción. Al agregar formaldehído, esta nueva mutante era capaz de catalizar una reacción aldólica entre el intermediario enolato y una molécula de formaldehído (Figura 3.12).

OYE a su vez presenta una gran promiscuidad por una gran cantidad de

sustratos de diversa naturaleza química, reaccionando con más de 20 moléculas α,β -insaturadas, donde se destacan moléculas utilizadas como precursores de fármacos⁷⁹. Esto despierta el interés por generar nuevas variantes de OYE Y187F que tuvieran capacidades catalíticas mejoradas, lo cual supone un atractivo tanto para la investigación como también para industrias. Es por esto que se consideró a OYE como una aplicación directa del flujo de trabajo y metodología propuestos en esta tesis, proponiendo mutaciones que permitan obtener sistemas catalíticos más eficientes.

Como se mencionó en los capítulos de validación, para poder aplicar las transformaciones alquímicas se necesita generar el campo de fuerza del sistema y conocer el paso limitante de la reacción. Dado que esta nueva variante de OYE supone una reacción sin precedentes, se utiliza el método de la cuerda adaptativa para poder encontrar el camino de mínima energía libre que conecta reactantes y producto. Las mismas condiciones de simulación utilizadas en DHFR son aplicadas en OYE, donde el intermediario enolato es relajado y utilizado para encontrar tanto el reactante como el producto para definir los extremos de la cuerda.

De acuerdo a la figura 3.12, la reacción de OYE involucra 3 pasos elementales: la transferencia de hidruro, el ataque nucleofílico del enolato al formaldehído y posteriormente la protonación del grupo alcoxi generado. Este mecanismo de reacción se determinó por simulaciones QM/MM mediante el método de la cuerda adaptativa⁸¹, utilizando el Hamiltoniano semiempírico

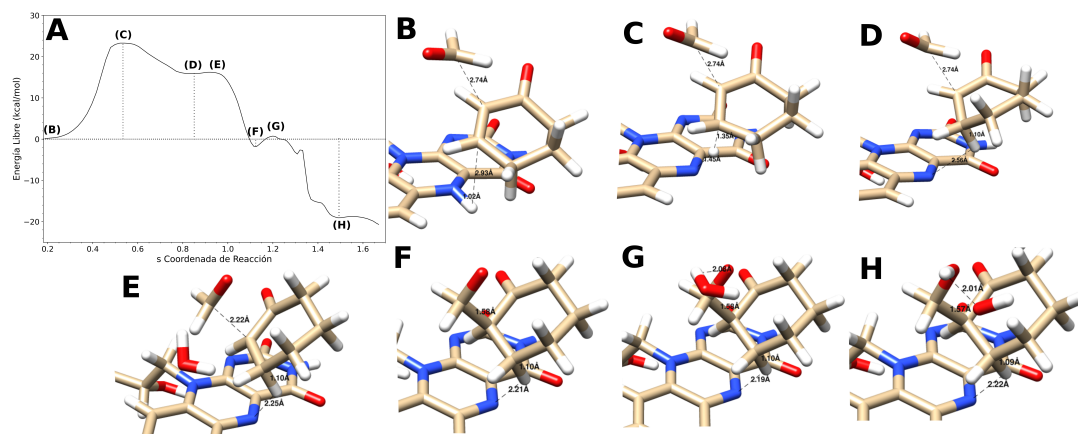


Figura 3.13: A)Perfil de energía libre para la mutante de OYE Y187F. B-H)representación de las estructuras a lo largo de la coordenada de reacción

DFTB3 para la zona QM. Debido a que la reacción involucra más de un paso elemental, para obtener el perfil de energía libre de la reacción catalizada por OYE Y187F, se calcularon 2 cuerdas. La primera corresponde a la transferencia de hidruro por lo que su estado inicial y final son la ciclohexenona y el ciclohexenolato, correspondientes a figura 3.13B y D respectivamente. La segunda cuerda parte del enolato mencionado anteriormente para luego atacar al formaldehído, generando el 2-metoxiciclohexanona (fig 3.13F), el cual es finalmente protonado para formar el 2-hidroximetilciclohexanona (fig 3.13H).

Iniciando desde el complejo enzima sustrato de la estructura cristalina (Figura 3.13B) se realizó un scan para obtener la estructura del enolato (Fig.3.13D),

utilizando la coordenada anti-simétrica como variable colectiva correspondiente a la distancia entre el nitrógeno donador del mononucleótido de Flavina (FMN) y el hidruro, menos la distancia hidruro-Carbono aceptor de la ciclohexenona. De forma similar se obtuvo el estado 3.13F realizando un scan en el cual se iba reduciendo la distancia entre el carbono del formaldehído y el carbono α del enolato. Cada uno de estos estados fueron minimizados y equilibrados a nivel DFTB3/CHARMM36m durante 300 ps. El estado 3.13H surgió espontáneamente durante la equilibración del estado 3.13F tomando un protón de una molécula de agua vecina. Los estados equilibrados sirvieron como puntos finales e iniciales para poder encontrar el camino de mínima energía libre que conectaba cada uno de los estados, utilizando un total de 20 y 24 nodos (estados) para la cuerda 1 y 2 respectivamente.

Para la primera cuerda se utilizaron como variables colectivas (VCs) la distancia donador-hidruro, hidruro-aceptor y donador-aceptor, siendo el donador el FMN y el aceptor la ciclohexen-2-ona (Figura 3.14A). Para la segunda cuerda 5 VCs fueron utilizadas: 1) los dos carbonos que están formando un nuevo enlace, 2) el protón transferido y el átomo de oxígeno del formaldehído, 3) el protón transferido y el átomo de oxígeno del agua y 4) el átomo de oxígeno donador y el átomo de oxígeno aceptor. Como quinta VC, se eligió el ángulo entre el nuevo enlace carbono-carbono y el enlace carbono-oxígeno del formaldehído, ya que este ángulo es conocido en el campo de la química orgánica por ser importante para las adiciones a un grupo carbonilo (trayecto-

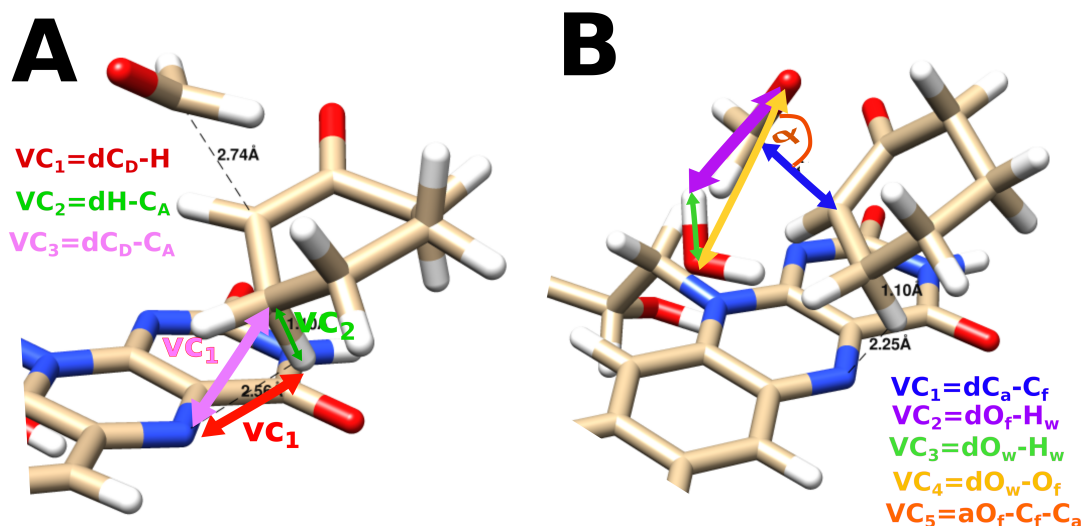


Figura 3.14: A) Variables colectivas utilizadas en la primera cuerda correspondiente a la transferencia de hidruro. B) VCs de la segunda cuerda, involucrando la reacción aldólica con la posterior protonación

ria de Bürgi – Dünitz)(Figura 3.14B). Los valores promedios de cada una de las VC fueron obtenidas de las trayectorias de equilibrio y las cuerdas fueron simuladas hasta que el perfil de energía libre variara en menos de 1 kJ/mol. Una vez cumplido ese criterio se procedió a obtener el perfil de energía libre con los valores de las VC convergidas mediante la técnica de *Umbrella Sampling*⁶⁹, la cual utilizó cada uno de los nodos de la cuerda como una ventana y posteriormente el perfil de fuerza media fue obtenido mediante el estimador *Weighted Histogram Analysis Method*¹⁸¹ (WHAM).

Para las simulaciones QM/MM se utilizó el método DFTB3/CHARMM36m donde la zona de mecánica cuántica (QM) corresponde a la porción aromá-

tica del FMN, la ciclohexen-2-ona, el formaldehído y una molécula de agua. Es conocido de la literatura que el funcional semi-empírico DFTB3 utilizado para tratar la zona QM tiende a sub-estimar las barreras por lo que una corrección a la energía libre fue requerido. Para esto se calcularon estructuras minimizadas de cada nodo a nivel DFT utilizando el funcional M06-2X y la base def2-TZVP. La diferencia en energía entre ambos métodos se le sumó al perfil y se presenta en la figura 3.15, donde se aprecia claramente la sub-estimación de la barrera en la transferencia de hidruro junto a la sobre estabilización del intermediario enolato por DFTB3. La figura 3.13 muestra el perfil de energía libre a nivel M602X-D3/def2-TZVP//DFTB3/CHARMM36m, para el cual la transferencia de hidruro posee una barrera de activación de 22.3 kcal/mol siendo sustancialmente mayor en comparación a la barrera de activación 2 y 3 (3.13E y 3.13G respectivamente). Lo anterior implica que para poder evaluar el efecto de mutaciones sobre la velocidad de reacción en OYE se debe considerar el primer paso de la reacción, generando el campo de fuerza para el estado C y modificando los parámetros del estado B, correspondientes al estado de transición y reactante respectivamente.

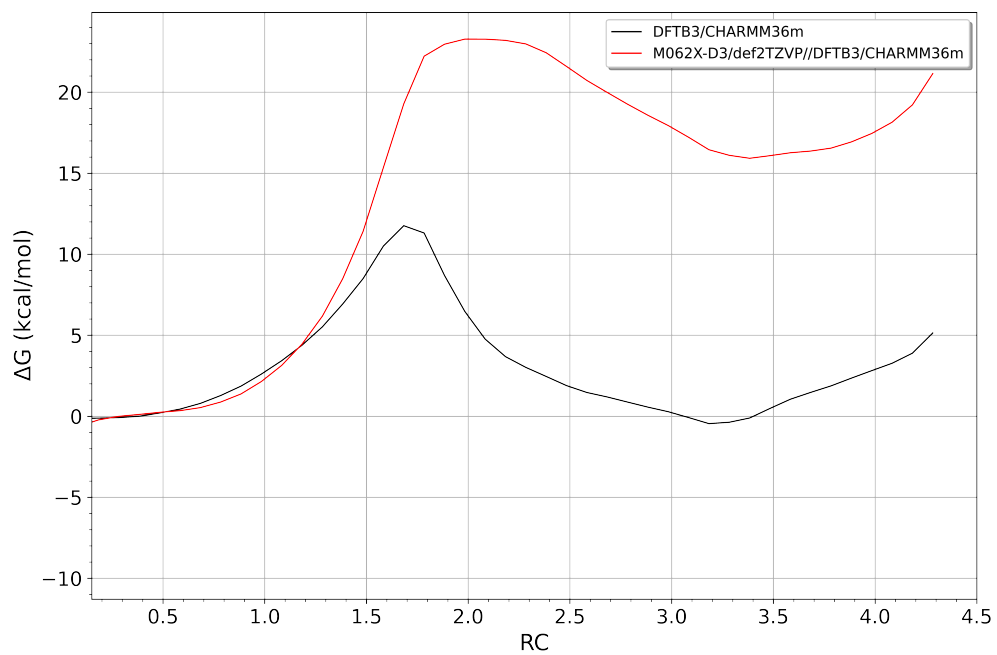


Figura 3.15: Perfil de energía libre de reacción corregido, para la transferencia de hidruro en la mutante Y187F de la enzima OYE. El nivel de teoría corresponde a M062X-D3/def2-TZVP//DFTB3/CHARMM36m

3.6. Aplicación de las transformaciones alquímicas para la búsqueda de nuevas variantes de OYE Y187F

El último objetivo de esta tesis, comprende la búsqueda de nuevas variantes de la enzima OYE Y187F mediante transformaciones alquímicas de no-equilibrio. Y187F corresponde a una mutante diseñada artificialmente para catalizar la reacción de ciclohexen-2-ona con formaldehído, reacción que no ocurría en la variante silvestre. En el capítulo anterior se estableció el mecanismo de la reacción aldólica en OYE Y187F. En base al mecanismo se derivó un campo de fuerza para el primer estado de transición el cual comprende el paso limitante de la reacción. De igual forma que en los capítulos anteriores, los parámetros no enlazantes del reactante fueron modificados para considerar los efectos del ambiente enzimático sobre los ligandos. Las estructuras de partida tanto para el ET como para el reactante son obtenidos de la cuerda adaptativa calculada en el capítulo anterior.

Definidos el campo de fuerza y las estructuras para ambos estados, se procede a la búsqueda de nuevas variantes que puedan tener una actividad catalítica superior a Y187F. Debido a que la catálisis de la reacción aldólica no corresponde a un proceso natural en la familia de enzimas OYE, se opta por un diseño racional de nuevas variantes. La figura 3.16 muestra los resi-

duos a mutar para evaluar su efecto en la energía libre de activación mediante transformaciones alquímicas de no-equilibrio.

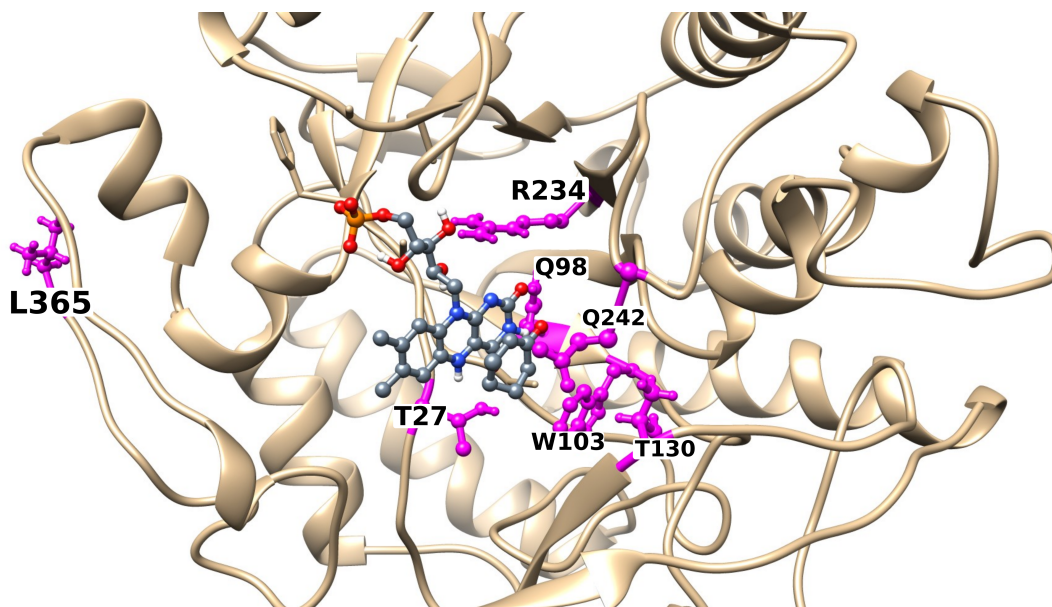


Figura 3.16: Residuos de OYE Y187F a mutar para la búsqueda de variantes catalíticamente más activas. El sustrato junto a la coenzima FMN son coloreados en gris, mientras que todos los residuos a mutar están en color magenta.

La selección de estos residuos fue basada en dos principios. Por un lado, se decidió mutar residuos cercanos o en el mismo sitio activo, ya que estos debiesen tener un rol en los procesos catalíticos dada su ubicación espacial y las interacciones que poseen con los ligandos (W103, Q98, R234 Y Q242). El segundo criterio fue basado en los fundamentos que sustentan un cambio en la energía libre de activación. Para poder disminuir la barrera de activación, la mutación debe ser capaz de alterar la estabilidad del estado de transición y/o el complejo de Michaelis-Menten (CMM) (T27, T130, R234).

A modo de ejemplo, si una mutación logra estabilizar el estado de transición mientras deja inalterado el CMM, el efecto neto será una disminución en la energía libre de activación. De la misma forma, si se desestabiliza el complejo de Michaelis-Menten, volviéndolo más reactivo el efecto sería similar. Un tercer grupo correspondiente a sólo un sitio de mutación (L365) fue seleccionado. Este residuo fué postulado por un método de inteligencia artificial desarrollado por una compañía privada. Nuestros colaboradores experimentales, realizaron ensayos enzimáticos de tales variantes(Anexo E.1).

Las mutaciones elegidas en los sitios marcados en la Figura 3.16, corresponden a modificaciones que desestabilizan el estado de reactante. Para lograr esto, se priorizaron mutantes que implicaran la pérdida de puentes de hidrógeno, puentes salinos, alteración del volumen del residuo, entre otros.

Una forma de corroborar si las mutaciones alteran la estabilidad del CMM, es a través de un ciclo termodinámico como el de la figura 3.17. En este ciclo termodinámico se define la estabilidad del complejo, como la diferencia en energía libre entre este estado y el de la enzima completamente desplegada o denaturada (estado D). Al igual que en el ciclo termodinámico anterior, la diferencia en estabilidad entre la variante Y187F y otra mutante, puede ser calculada indirectamente al realizar una transformación alquímica en el CMM y en el estado desplegado. Asumiendo que el estado desplegado puede modelarse como un tripéptido en solución, los valores de la columna de la izquierda en la figura 3.17, corresponden a datos pre-calculados que pueden

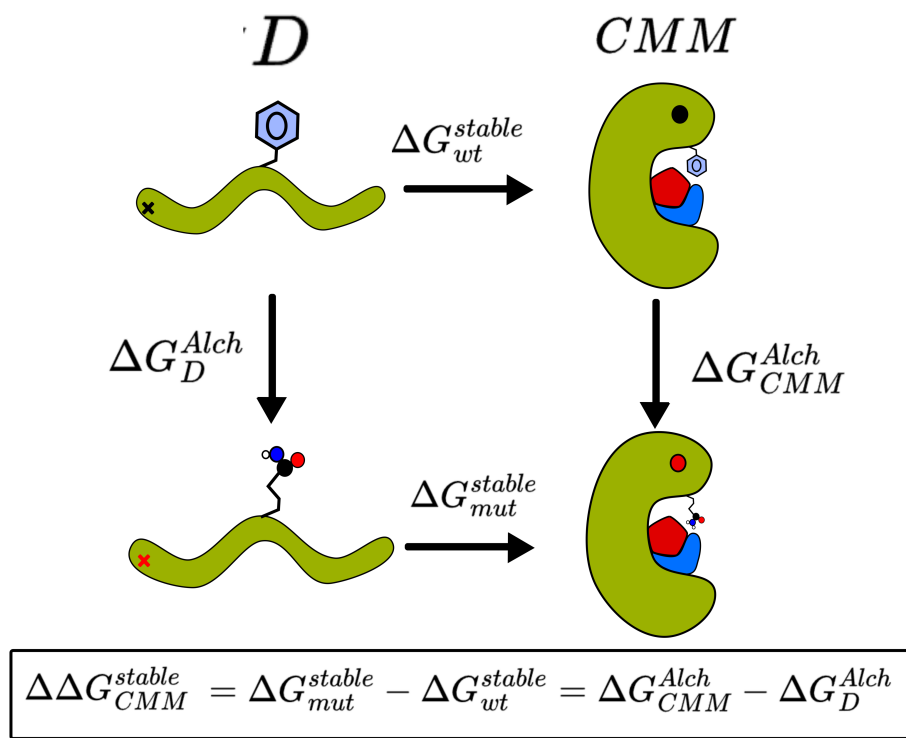


Figura 3.17: Ciclo termodinámico para la estimación de la estabilidad del complejo de Michaelis-Menten. La primera columna corresponde a la transformación alquímica del estado denaturado D, mientras la segunda columna, da cuenta del mismo proceso para el estado de reactante. ΔG^{stable} , se define como la energía libre de estabilidad, de pasar desde un estado desplegado a un complejo de Michaelis-Menten.

obtenerse de la base de datos de PMX⁷³. El uso de tripéptidos como modelo de enzimas denaturadas ha sido ampliamente utilizado en transformaciones alquímicas⁷³, y supone un ahorro en el costo computacional considerable. De esta forma, sin la necesidad de realizar transformaciones alquímicas adicionales, el uso de este nuevo ciclo permite entender en términos energéticos el impacto de las mutaciones a evaluar.

Una vez definidas las mutantes a estudiar, se realizaron las simulaciones

de equilibrio para obtener las conformaciones representativas del colectivo de Boltzmann. Cada estado fue simulado por un total de 300 ns bajo el colectivo NVT, siendo previamente equilibrados a una temperatura de 300 K y 1 bar. Un análisis conformacional reveló que sobre los 250 ns la mayoría de las mutaciones alcanzaban el equilibrio (Anexo D.1 y D.2). De los 50 ns restantes se extrajeron 100 conformaciones equidistantes cada 0.5 ns para realizar las transformaciones alquímicas de 100 ps. La estimación de la energía libre del proceso alquímico es realizado mediante el método BAR¹²³.

Los valores reportados en la tabla 3.7, corresponden a los valores promediados de 3 réplicas de transformaciones alquímicas independientes. El error es estimado como el valor cuadrático medio de las desviaciones estándar para el ET y el complejo de Michaelis-Menten. Un total de 10 diferentes mutantes fueron analizadas. En total se encontraron 2 variantes catalíticamente más activas que Y187F.

Con un valor de $\Delta\Delta G^\ddagger$ igual a -7.5 kJ/mol, Y187F/W103F es la primera mutante predicha a ser más activa. Esta mutación implica la modificación del residuo triptófano 103 por una fenilalanina, otorgándole un carácter completamente apolar. En adición a esto, el valor de $\Delta\Delta G_{CMM}^{stable}$ para esta variante es de 11.7 kJ/mol, indicando una desestabilización del reactante, lo que disminuye la barrera de activación.

Interesantemente, la segunda variante catalíticamente más activa que OYE Y187F, genera un efecto opuesto al de la mutante previa. Y187F/T130W po-

Mutante	$\Delta\Delta G^\ddagger$ [kJ/mol]	$\Delta\Delta G_{CMM}^{stable}$ [kJ/mol]	$\Delta\Delta G_{TS}^{stable}$ [kJ/mol]
Y187F/T27A	9.0±0.8	-12.4±0.2	-3.4±0.2
Y187F/R234K	6.3±2.7	-1.3±2.3	4.8±2.3
Y187F/Q101N/T27N	4.3±1.9	-	-
Y187F/L365A	3.8±0.1	4.4±0.4	8.2±0.4
Y187F/R234Q	3.3±4.4	-	-
Y187F/L365Q	2.7±0.5	-32.2±0.7	-29.5±0.7
Y187F/Q242L	-0.2±0.6	39.0±0.8	38.8±0.8
Y187F/T130H	-0.9±0.4	-13.4±0.6	-14.3±0.6
Y187F/W103F	-7.5±0.8	11.7±0.9	4.2±0.9
Y187F/T130W	-8.8±2.5	-25.4±1.3	-34.2±1.3

Tabla 3.7: Cambios en la energía libre de activación y de estabilidad para diferentes mutantes de la variante de OYE Y187F. - corresponden a energías no disponibles en la base de datos, debido a cambios en la carga neta del residuo, o mutaciones triple.

see un valor de $\Delta\Delta G_{CMM}^{stable}$ de -25.4 kJ/mol, prediciendo un complejo reactante más estable. Sin embargo, la barrera de activación decrece en 8.8 kJ/mol en esta variante, lo que significa que el estado de transición debe tener una estabilización aún mayor. Lo anterior es corroborado al observar el valor de $\Delta\Delta G_{ET}^{stable}$, siendo -34.2 kJ/mol.

A pesar de encontrarse otras variantes con un $\Delta\Delta G^\ddagger$ negativo, los valores son muy cercanos a cero, por lo que tipifican como efecto nulo dentro del error del método. Cabe destacar que las variantes Y187F/L365Q y Y187F/L365A, propuestas por un método de inteligencia artificial, no exhiben una mejora en términos catalíticos. Esto se sustenta con los resultados experimentales facilitados por nuestros colaboradores (Anexo E.1). Lo ante-

rior, pone en evidencia los desafíos que presentan los métodos basados en inteligencia artificial, para la predicción de propiedades cinéticas en sistemas enzimáticos.

En términos computacionales, se tiene que las transformaciones alquímicas de no-equilibrio, permiten obtener valores de $\Delta\Delta G^\ddagger$ en 2.5 horas, partiendo de los estados equilibrados y utilizando 8 procesadores (CPU) y 1 unidad de procesamiento gráfico (GPU). El método de la cuerda adaptativa, tarda 72 horas utilizando por lo menos 20 CPUs, cuando se inicia la minimización del camino de mínima energía libre desde un camino previamente convergido en otra variante. Lo anterior implica que las transformaciones alquímicas son al menos un orden de magnitud (35 veces) más rápidas que la cuerda. Más aún el tiempo utilizado en la cuerda, no considera las correcciones al perfil de reacción, las que son muy dependientes de cada sistema y pueden duplicar el tiempo necesario para obtener un valor de energía libre.

Los hallazgos de este capítulo permiten proponer, de manera rápida y precisa, mutantes con mayor actividad catalítica que las enzimas de referencia. El diseño racional de estas variantes optimiza la búsqueda al enfocarse en modificaciones con efectos específicos en los estados de reactante y/o transición. Sin embargo, los resultados indican que los efectos de una mutación no son siempre aditivos, por lo que una misma perturbación puede mejorar o perjudicar la enzima según su impacto global en la cinética enzimática. Más aún, utilizando el estado desplegado como referencia, se puede explicar el

efecto catalítico de la mutación en términos de la estabilidad del complejo de Michaelis-Menten y el estado de transición. Lo anterior, vuelve a las transformaciones de no-equilibrio, una técnica no solo predictiva, si no también explicativa, logrando entender el impacto de mutaciones en el mecanismo enzimático.

4. Conclusiones

El objetivo principal de esta tesis consistió en desarrollar un método computacional capaz de predecir cambios en la energía libre de activación sujeta a mutaciones para sistemas enzimáticos. El desarrollo de esta herramienta permite acelerar el proceso de búsqueda de variantes enzimáticas con propiedades catalíticas mejoradas en comparación a las metodologías basadas en primeros principios. Esta herramienta facilitará un diseño racional de enzimas a bajo costo y con alta precisión, siendo potencialmente útil en el campo de la biotecnología y la ingeniería de proteínas.

La validación de esta metodología en Crotonil-coA Carboxilasa/Reductasa, muestra que el uso de restricciones de distancia y ángulos para mantener al sistema en el estado de transición no son efectivas, dado lo compacto del estado e interacciones repulsivas demasiado fuertes lo que vuelve al sistema inestable. Este hallazgo se ve respaldado por la alta fluctuación de la distancia entre los carbonos donador y aceptor, en conjunto a simulaciones que fallaron debido a cambios bruscos en la geometría. El método de seminario para modelar las interacciones enlazantes del estado de transición evitó las grandes fluctuaciones en las distancias y ángulos asociados al estado de transición y permitió tener un sistema estable en todas las simulaciones. Dada la evi-

dencia anteriormente detallada, en esta tesis se elige al método de seminario como la metodología para modelar las interacciones enlazantes en los estados de transición.

En cuanto a las interacciones no enlazantes, la evaluación de las transformaciones alquímicas en CCR, mostró que sólo mejorar las interacciones electrostáticas usando diferentes cargas atómicas en el estado de transición y reactante, no era suficiente para poder predecir los cambios en la energía libre de activación con suficiente precisión. Una considerable mejora es lograda cuando también son modificados los parámetros de Lennard-Jones de ambos estados. Este hallazgo refuerza la idea de co-dependencia entre cargas y parámetros de LJ en campos de fuerza convencionales.

En vista de lo anterior, se realizó un análisis sistemático de los parámetros no enlazantes de distintos aminoácidos, centrándose en interacciones del tipo $\pi - \pi$. Los resultados indican que las cargas atómicas derivadas mediante métodos AIM no presentan una precisión significativamente superior a las obtenidas con campos de fuerza convencionales. Sin embargo, todos los esquemas de carga sobrestiman la repulsión electrostática a distancias cortas, problema que pudo mitigarse mediante potenciales alternativos que incorporen efectos de penetración de carga.

En cuanto a las interacciones de dispersión, los parámetros MBIS demostraron mayor precisión, con un error inferior a 4,2 kJ/mol, superando a otros esquemas AIM y a los campos de fuerza convencionales. Por otro

lado, la repulsión de Pauli resultó ser la contribución más difícil de predecir: CHARMM36m y AMBER99sb la subestiman, mientras que los métodos AIM tienden a sobrestimar, salvo MBIS, que mostró el mejor desempeño. No obstante, a distancias cortas, este esquema aún presenta errores superiores a 4,2 kJ/mol.

Este estudio no solo proporciona una guía para seleccionar esquemas de partición en la derivación de parámetros, sino que también resalta las limitaciones y desafíos de las formas funcionales en los campos de fuerza. Se sugiere una reformulación que permita reproducir con mayor fidelidad las contribuciones energéticas obtenidas a partir de primeros principios, mejorando así la transferibilidad de los campos de fuerza.

Una vez identificado MBIS como el esquema óptimo para derivar parámetros no covalentes en aminoácidos, se evaluó su desempeño en CCR modificando los parámetros de los residuos a transformar. La modificación exclusiva de cargas no mejoró la precisión, pero al incluir ajustes en los parámetros de LJ, se observaron mejoras en algunos sistemas, donde el error máximo es de 3.5 kJ/mol en F170A, logrando la precisión química (4.2kJ/mol).

Para evaluar la metodología en un sistema enzimático independiente se selecciono la Dihidrofolato Reductasa (DHFR), cuyo sitio activo bien estructurado permite aislar efectos de interacción. La derivación de parámetros de primeros principios mejoró la predicción en dos de tres mutantes, obteniéndose errores cercanos a 4,2 kJ/mol considerado como el límite de la precisión

química. Sumado a esto, se encontró que todos los valores calculados fueron sistemáticamente menores a los experimentales, concluyendo que las transformaciones químicas no capturan completamente la repulsión en estados de transición, en línea con la subestimación de MBIS a distancias cortas observada en el estudio de los aminoácidos.

Estos hallazgos conectan observaciones microscópicas y macroscópicas, demostrando el potencial de esta nueva metodología y abriendo oportunidades para corregir sus limitaciones en futuros estudios.

Validada la metodología propuesta, se estudió la *Old Yellow Enzyme* (OYE), específicamente su variante Y187F, que acelera la reacción aldólica entre 2-ciclohexenona y formaldehído, que originalmente la enzima silvestre es incapaz de catalizar. Un estudio QM/MM mediante el método de la cuerda permitió caracterizar su perfil de energía libre, identificando la transferencia de hidruro como el paso limitante y obteniendo las conformaciones clave de reactantes y estado de transición. Este análisis proporciona la primera caracterización mecanística de la nueva reacción aldólica en variante Y187F de OYE, sentando las bases para un diseño racional de enzimas aldolasas más eficientes.

Con los estados de transición y reactantes caracterizados, se aplicó la metodología propuesta para evaluar mutantes de OYE basadas en Y187F, buscando reducir la barrera de activación. Además, se analizó la estabilidad de estas mutantes, aprovechando datos disponibles en la plataforma pmx. Los

resultados permiten inferir cómo las mutaciones modulan la energía libre de activación, diferenciando entre efectos desestabilizantes en el complejo de Michaelis-Menten o de estabilización del estado de transición. Las transformaciones alquímicas predicen 2 nuevas variantes con energía libre de activación menor a la mutante simple, Y187F/W103F e Y187F/T130W. Con un $\Delta\Delta G^\ddagger$ de -8.8 kJ/mol la segunda variante se postula a ser la mutante más rápida hasta la fecha.

Como conclusión final se puede establecer que las transformaciones alquímicas de no equilibrio han permitido predecir diferencias en la energía libre de activación en variantes enzimáticas de CCR, DHFR y OYE, validando la herramienta desarrollada en esta tesis para poder predecir con alta precisión el impacto de mutaciones tanto en su estabilidad como también en su desempeño catalítico. Los resultados obtenidos confirman la hipótesis planteada, demostrando que los cálculos de energía libre basados en transformaciones alquímicas de no equilibrio, combinados con una descripción más detallada de las interacciones no covalentes en el sitio activo, son capaces de predecir con alta precisión los cambios en la energía libre de activación producto de mutaciones enzimáticas.

Este trabajo no solo valida el uso de metodologías computacionales basadas en principios fisicoquímicos en la predicción de estabilidad y reactividad enzimática, sino que también sienta las bases para futuras aplicaciones en ingeniería de proteínas, donde la predicción eficiente de variantes enzimáticas

con propiedades optimizadas es clave para el desarrollo de biocatalizadores más eficientes y estables.

Adicionalmente el costo computacional de las transformaciones químicas es al menos un orden de magnitud menor a las simulaciones QM/MM, permitiendo analizar cientos de mutantes, lo que reduciría la labor experimental en el diseño de nuevos biocatalizadores.

4.1. Proyecciones

Los hallazgos y evidencias presentados en esta tesis abren nuevas preguntas y oportunidades para continuar avanzando en este campo. Como primera proyección, y en línea con la metodología desarrollada, surge la necesidad de aplicarla a sistemas enzimáticos que involucren reacciones químicas distintas. Si bien a lo largo de la tesis se aplican las transformaciones alquímicas a 3 enzimas diferentes, todas ellas implican la transferencia de un hidruro como paso limitante de la reacción. Es por esto, que en un futuro se espera ampliar la diversidad química de los sistemas a estudiar, permitiendo realizar análisis cada vez más extensos de la metodología desarrollada.

Como segunda proyección se tiene la mejora en la descripción de las interacciones con el fin de obtener cambios en la energía libre de activación cada vez más precisos. Para lograr tales desafíos se espera poder mejorar tanto la forma funcional de los campos de fuerza así como también sus parámetros. En cuanto a la forma funcional, en el marco de esta tesis se propuso la inclusión de efectos de penetración de carga con el fin de mejorar la descripción de las interacciones electrostáticas. En adición a esta corrección, la inclusión de la contribución de multipolos atómicos en conjunto a las cargas atómicas, mejoraría la descripción electrostática. Para las interacciones de Van der Waals, por otro lado, Tkatchenko y colaboradores han propuesto un potencial

universal que reproduce correctamente la interacción entre dímeros de gases nobles, así como también dímeros de moléculas orgánicas pequeñas¹⁸². La versatilidad de los métodos AIM permiten a su vez derivar cada uno de los nuevos parámetros que deben ser incluidos en los nuevos potenciales mencionados. La derivación de nuevos parámetros en conjunto a las formulaciones de nuevos potenciales de interacción clásicos, abren la oportunidad de generar nuevas proyecciones en el área del desarrollo de campos de fuerza, los cuáles permitirían una estimación cada vez más precisa de la energía libre de activación utilizando transformaciones alquímicas de no-equilibrio.

La última proyección que nace de esta tesis, implica la estimación de propiedades cinéticas más allá de los cambios en la energía libre de activación $\Delta\Delta G^\ddagger$. Las reacciones que generan productos secundarios presentan bifurcaciones en su perfil de reacción. En estos casos, cada producto se forma a partir de estados de transición de etapas que no determinan la velocidad de la reacción, pero sí definen el producto cinético mayoritario. Tomando como ejemplo a la enzima OYE, generar un ciclo termodinámico que implique la transformación del estado de transición de la reacción aldólica en la reacción de reducción, permitiría obtener una medida de la proporción en la cual estas dos reacciones son catalizadas. Aplicando tal ciclo a múltiples variantes, en principio se podría predecir como las mutaciones promueven la generación de un producto por sobre otro, entregando información valiosa para el diseño de nuevos biocatalizadores. Finalmente, para sustratos con centros quirales

o pro-quirales, un tercer ciclo termodinámico que incluya la transformación de los estados de transición que conllevan a distintos productos enantioméricos, permitirían obtener la enantioselectividad que posee cada una de las mutantes.

Estas proyecciones no solo amplían el alcance de la metodología desarrollada, sino que también abren nuevas fronteras en la comprensión y el diseño racional de biocatalizadores. Al mejorar la precisión en la estimación de energías libres de activación, diversificar los sistemas enzimáticos estudiados y aplicar en la predicción de selectividades cinéticas y enantioméricas, este trabajo sienta las bases para una nueva generación de modelos computacionales eficientes y versátiles, aportando así al futuro de la ingeniería de proteínas.

Bibliografía

- [1] Liu, Q.; Xun, G.; Feng, Y. *Biotechnology Advances* **2019**, 37, 530–537.
- [2] Sinha, R.; Shukla, P. *Current Protein & Peptide Science* **2019**, 20, 398–407.
- [3] Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. *Nature* **2012**, 485, 185–194.
- [4] Bommarius, A. S.; Paye, M. F. *Chemical Society Reviews* **2013**, 42, 6534–6565.
- [5] Schramm, V. L. *Annual Review of Biochemistry* **1998**, 67, 693–720.
- [6] Schramm, V. L. *Chemical Reviews* **2018**, 118, 11194–11258.
- [7] Ashaolu, T. J.; Malik, T.; Soni, R.; Prieto, M. A.; Jafari, S. M. *Food Science & Nutrition* **2025**, 13, e4540.
- [8] Zhou, L.; Tao, C.; Shen, X.; Sun, X.; Wang, J.; Yuan, Q. *Biotechnology Advances* **2024**, 73, 108376.
- [9] Song, Z.; Zhang, Q.; Wu, W.; Pu, Z.; Yu, H. *Frontiers in Bioengineering and Biotechnology* **2023**, 11, 1129149.
- [10] Vidal, L. S.; Isalan, M.; Heap, J. T.; Ledesma-Amaro, R. *RSC Chemical Biology* **2023**, 4, 271–291.
- [11] Arnold, F. H. *Angewandte Chemie* **2019**, 58, 14420–14426.
- [12] Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, 302, 1364–1368.
- [13] Wang, Y.; Xue, P.; Cao, M.; Yu, T.; Lane, S. T.; Zhao, H. *Chemical Reviews* **2021**, 121, 12384–12444.
- [14] Zacco, M.; Williams, D. M.; Brown, D. M.; Gherardi, E. *Journal of Molecular Biology* **1996**, 255, 589–603.

- [15] Fujii, R.; Kitaoka, M.; Hayashi, K. *Nature Protocols* **2006**, *1*, 2493–2497.
- [16] Stemmer, W. P. C. *Nature* **1994**, *370*, 389–391.
- [17] Shao, Z.; Zhao, H.; Giver, L.; Arnold, F. H. *Nucleic Acids Research* **1998**, *26*, 681–683.
- [18] Ndochinwa, G. O.; Wang, Q.-Y.; Okoro, N. O.; Amadi, O. C.; Nwagu, T. N.; Nnamchi, C. I.; Moneke, A. N.; Odiba, A. S. *Open Life Sciences* **2024**, *19*.
- [19] Korendovych, I. V. *Protein Engineering*; Springer New York, 2018; Vol. 1685; pp 15–23.
- [20] Lutz, S. *Current Opinion in Biotechnology* **2010**, *21*, 734–743.
- [21] Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Molecular Systems Biology* **2011**, *7*, 539.
- [22] Hogeweg, P.; Hesper, B. *Journal of Molecular Evolution* **1984**, *20*, 175–186.
- [23] Jumper, J. et al. *Nature* **2021**, 1–11.
- [24] Wüthrich, K. *Journal of Biological Chemistry* **1990**, *265*, 22059–22062.
- [25] Papageorgiou, A. C.; Mattsson, J. In *Protein Downstream Processing: Design, Development and Application of High and Low-Resolution Methods*; Labrou, N. E., Ed.; Humana Press, 2014; pp 397–421.
- [26] Nakane, T. et al. *Nature* **2020**, *587*, 152–156.
- [27] Horikoshi, N. et al. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2022790118.
- [28] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *Journal of Molecular Biology* **1982**, *161*, 269–288.
- [29] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Structure, Function, and Bioinformatics* **2002**, *47*, 409–443.

- [30] Peña-Varas, C.; Kanstrup, C.; Vergara-Jaque, A.; González-Avendaño, M.; Crocoll, C.; Mirza, O.; Dreyer, I.; Nour-Eldin, H.; Ramírez, D. *International Journal of Molecular Sciences* **2022**, *23*, 1595.
- [31] Baker, D. *Protein Science: A Publication of the Protein Society* **2010**, *19*, 1817–1819.
- [32] Albanese, K. I.; Barbe, S.; Tagami, S.; Woolfson, D. N.; Schiex, T. *Nature Reviews Methods Primers* **2025**, *5*, 1–28.
- [33] Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y. *Scientific Reports* **2018**, *8*, 6349.
- [34] Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. *Methods in Enzymology*; Elsevier, 2004; Vol. 383; pp 66–93.
- [35] Mignon, D.; Druart, K.; Michael, E.; Opuu, V.; Polydorides, S.; Villa, F.; Gaillard, T.; Panel, N.; Archontis, G.; Simonson, T. *The Journal of Physical Chemistry A* **2020**, *124*, 10637–10648.
- [36] Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387–1391.
- [37] Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D. *Nature* **2008**, *453*, 190–195.
- [38] Correia, B. E. et al. *Structure* **2010**, *18*, 1116–1126.
- [39] Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. *Science* **2016**, *353*, 389–394.
- [40] Gonen, S.; DiMaio, F.; Gonen, T.; Baker, D. *Science* **2015**, *348*, 1365–1368.

- [41] Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; Baker, D. *Nature* **2013**, *501*, 212–216.
- [42] Zhang, Y.; Bartz, R.; Grigoryan, G.; Bryant, M.; Aaronson, J.; Beck, S.; Innocent, N.; Klein, L.; Procopio, W.; Tucker, T.; Jadhav, V.; Tellers, D. M.; DeGrado, W. F. *ACS Chemical Biology* **2015**, *10*, 1082–1093.
- [43] Korendovych, I. V.; Senes, A.; Kim, Y. H.; Lear, J. D.; Fry, H. C.; Therien, M. J.; Blasie, J. K.; Walker, F. A.; DeGrado, W. F. *Journal of the American Chemical Society* **2010**, *132*, 15516–15518.
- [44] Abramson, J. et al. *Nature* **2024**, *630*, 493–500.
- [45] Correia, B. E. et al. *Nature* **2014**, *507*, 201–206.
- [46] Kapoor, S.; Rafiq, A.; Sharma, S. *Critical Reviews in Food Science and Nutrition* **2017**, *57*, 2321–2329.
- [47] Mousavi, S. M.; Hashemi, S. A.; Iman Moezzi, S. M.; Ravan, N.; Gholami, A.; Lai, C. W.; Chiang, W.-H.; Omidifar, N.; Yousefi, K.; Behbudi, G. *Biochemistry Research International* **2021**, *2021*, 5599204.
- [48] Gayathiri, E.; Prakash, P.; Kumaravel, P.; Jayaprakash, J.; Ragu-nathan, M. G.; Sankar, S.; Pandiaraj, S.; Thirumalaivasan, N.; Thiru-vengadam, M.; Govindasamy, R. *Progress in Biophysics and Molecular Biology* **2023**, *185*, 17–32.
- [49] Czech, L.; Stamatakis, A.; Dunthorn, M.; Barbera, P. *Frontiers in Bio-informatics* **2022**, *2*.
- [50] Goldenzweig, A. et al. *Molecular Cell* **2016**, *63*, 337–346.
- [51] Bystroff, C.; Krogh, A. In *Protein Structure Prediction*; Zaki, M. J., Bystroff, C., Eds.; Humana Press, 2008; pp 173–198.
- [52] Ma, S.; Dai, Y. *Briefings in Bioinformatics* **2011**, *12*, 714–722.
- [53] Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Ashani, Y.; Goldsmith, M.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D. L.; Prilusky, J.;

- Amengual-Rigo, P.; Guallar, V.; Tawfik, D. S.; Fleishman, S. J. *Molecular Cell* **2018**, 72, 178–186.e5.
- [54] Leaver-Fay, A. et al. *Methods in Enzymology* **2011**, 487, 545–574.
- [55] Genheden, S.; Ryde, U. *Expert Opinion on Drug Discovery* **2015**, 10, 449–461.
- [56] Warshel, A.; Weiss, R. M. *Journal of the American Chemical Society* **1980**, 102, 6218–6226.
- [57] Karplus, M.; McCammon, J. A. *Nature Structural Biology* **2002**, 9, 646–652.
- [58] Fichtorn, K. A.; Weinberg, W. H. *The Journal of Chemical Physics* **1991**, 95, 1090–1096.
- [59] Neese, F. *Faraday Discussions* **2025**, 254, 295–314.
- [60] Dauber-Osguthorpe, P.; Hagler, A. T. *Journal of Computer-Aided Molecular Design* **2019**, 33, 133–203.
- [61] Liwo, A.; Czaplewski, C.; Sieradzan, A. K.; Lipska, A. G.; Samsonov, S. A.; Murarka, R. K. *Biomolecules* **2021**, 11, 1347.
- [62] Ho, J.; Yu, H.; Shao, Y.; Taylor, M.; Chen, J. *The Journal of Physical Chemistry A* **2025**, 129, 1517–1528.
- [63] Nam, K.; Shao, Y.; Major, D. T.; Wolf-Watz, M. *ACS Omega* **2024**, acsomega.3c09084.
- [64] Gao, J.; Ma, S.; Major, D. T.; Nam, K.; Pu, J.; Truhlar, D. G. *Chemical Reviews* **2006**, 106, 3188–3209.
- [65] Eyring, H. *The Journal of Chemical Physics* **1935**, 3, 107–115.
- [66] Doron, D.; Stojković, V.; Gakhar, L.; Vardi-Kilshtain, A.; Kohen, A.; Major, D. T. *The Journal of Physical Chemistry B* **2015**, 119, 906–916.
- [67] Zinovjev, K.; Tuñón, I. *WIREs Computational Molecular Science* **2018**, 8.

- [68] Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Computational Molecular Science* **2011**, *1*, 826–843.
- [69] Kästner, J. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932–942.
- [70] Patel, J. S.; Berteotti, A.; Ronsisvalle, S.; Rocchia, W.; Cavalli, A. *Journal of Chemical Information and Modeling* **2014**, *54*, 470–480.
- [71] Raghavan, B.; Paulikat, M.; Ahmad, K.; Callea, L.; Rizzi, A.; Ippoliti, E.; Mandelli, D.; Bonati, L.; De Vivo, M.; Carloni, P. *Journal of Chemical Information and Modeling* **2023**, *63*, 3647–3658.
- [72] Mey, A. S. J. S.; Allen, B. K.; McDonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378–18378.
- [73] Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. *Journal of Computational Chemistry* **2015**, *36*, 348–354.
- [74] La Serra, M. A.; Vidossich, P.; Acquistapace, I.; Ganesan, A. K.; De Vivo, M. *Journal of Chemical Information and Modeling* **2022**, *62*, 3023–3033.
- [75] Aldeghi, M.; Gapsys, V.; de Groot, B. L. *ACS Central Science* **2018**, *4*, 1708–1718.
- [76] Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; Vlijmen, H. v.; Tresadern, G.; Groot, B. L. d. *Chemical Science* **2020**, *11*, 1140–1152.
- [77] Erb, T. J.; Berg, I. A.; Brecht, V.; Müller, M.; Fuchs, G.; Alber, B. E. *Proceedings of the National Academy of Sciences* **2007**, *104*, 10631–10636.
- [78] Stoffel, G. M. M.; Saez, D. A.; DeMirci, H.; Vögeli, B.; Rao, Y.; Zarzycki, J.; Yoshikuni, Y.; Wakatsuki, S.; Vöhringer-Martinez, E.; Erb, T. J. *Proceedings of the National Academy of Sciences* **2019**, *116*, 13964–13969.

- [79] Shi, Q.; Wang, H.; Liu, J.; Li, S.; Guo, J.; Li, H.; Jia, X.; Huo, H.; Zheng, Z.; You, S.; Qin, B. *Applied Microbiology and Biotechnology* **2020**, *104*, 8155–8170.
- [80] Lonsdale, R.; Reetz, M. T. *Journal of the American Chemical Society* **2015**, *137*, 14733–14742, Publisher: American Chemical Society.
- [81] Zinovjev, K.; Tuñón, I. *The Journal of Physical Chemistry A* **2017**, *121*, 9764–9772.
- [82] Born, M.; Oppenheimer, R. *Annalen der Physik* **1927**, *389*, 457–484.
- [83] Shavitt, I. In *Advanced Theories and Computational Approaches to the Electronic Structure of Molecules*; Dykstra, C. E., Ed.; Springer Netherlands, 1984; pp 185–196.
- [84] Kohn, W.; Becke, A. D.; Parr, R. G. *The Journal of Physical Chemistry* **1996**, *100*, 12974–12980.
- [85] Thomas, L. H. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, *23*, 542–548.
- [86] Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864–B871.
- [87] Levy, M. *Proceedings of the National Academy of Sciences* **1979**, *76*, 6062–6065.
- [88] Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133–A1138.
- [89] Borioni, J. L.; Puiatti, M.; Vera, D. M. A.; Pierini, A. B. *Physical Chemistry Chemical Physics* **2017**, *19*, 9189–9198.
- [90] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *The Journal of Physical Chemistry A* **2007**, *111*, 10439–10452.
- [91] Thiel, W. *WIREs Computational Molecular Science* **2014**, *4*, 145–157.
- [92] Spiegelman, F.; Tarrat, N.; Cuny, J.; Dontot, L.; Posenitskiy, E.; Martí, C.; Simon, A.; Rapacioli, M. *Advances in Physics: X* **2020**, *5*, 1710252.
- [93] Mulliken, R. S. *The Journal of Chemical Physics* **1955**, *23*, 1833–1840.

- [94] Eschrig, H.; Bergert, I. *physica status solidi (b)* **1978**, *90*, 621–628.
- [95] Gaus, M.; Cui, Q.; Elstner, M. *Journal of Chemical Theory and Computation* **2011**, *7*, 931–948.
- [96] Gomez, A.; Erb, T. J.; Grubmüller, H.; Vöhringer-Martinez, E. *Journal of Chemical Information and Modeling* **2023**, *63*, 7807–7815.
- [97] Vuignier, K.; Schappler, J.; Veuthey, J.-L.; Carrupt, P.-A.; Martel, S. *Analytical and Bioanalytical Chemistry* **2010**, *398*, 53–66.
- [98] Vanommeslaeghe, K.; MacKerell, A. D. *Biochimica et biophysica acta* **2015**, *1850*, 861–871.
- [99] Pulido, J.; Macaya, L.; Vöhringer-Martinez, E. *Journal of Chemical & Engineering Data* **2024**, *69*, 2917–2926.
- [100] Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671–690.
- [101] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.
- [102] Tkatchenko, A.; Scheffler, M. *Physical Review Letters* **2009**, *102*, 073005.
- [103] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.
- [104] Wang, J.; Cieplak, P.; Kollman, P. A. *Journal of Computational Chemistry* **2000**, *21*, 1049–1074.
- [105] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *Journal of the American Chemical Society* **1996**, *118*, 11225–11236.
- [106] Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *Journal of Computational Chemistry* **2004**, *25*, 1656–1676.

- [107] 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.
- [108] Zhao, L.; Von Hopffgarten, M.; Andrada, D. M.; Frenking, G. *WIREs Computational Molecular Science* **2018**, *8*, e1345.
- [109] Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chemical Reviews* **1994**, *94*, 1887–1930.
- [110] Horn, P. R.; Mao, Y.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2016**, *18*, 23067–23079.
- [111] Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227–249.
- [112] Warshel, A.; Karplus, M. *Journal of the American Chemical Society* **1972**, *94*, 5612–5625.
- [113] Amirani, M. C.; Tang, T. *Physical Chemistry Chemical Physics* **2015**, *17*, 7564–7575.
- [114] 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.
- [115] Bussi, G.; Donadio, D.; Parrinello, M. *The Journal of Chemical Physics* **2007**, *126*, 014101.
- [116] Parrinello, M.; Rahman, A. *Journal of Applied Physics* **1981**, *52*, 7182–7190.
- [117] Shirts, M. R.; Mobley, D. L.; Chodera, J. D. In *Annual Reports in Computational Chemistry*; Spellmeyer, D. C., Wheeler, R., Eds.; Elsevier, 2007; Vol. 3; pp 41–59.
- [118] Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M. *Journal of Chemical Information and Modeling* **2019**, *59*, 3128–3135.
- [119] Jiang, W.; Roux, B. *Journal of Chemical Theory and Computation* **2010**, *6*, 2559–2565.

- [120] Tan, Z. *Journal of Computational and Graphical Statistics* **2017**, 26, 54–65.
- [121] Zwanzig, R. W. *The Journal of Chemical Physics* **1954**, 22, 1420–1426.
- [122] Kollman, P. *Chemical Reviews* **1993**, 93, 2395–2417.
- [123] Bennett, C. H. *Journal of Computational Physics* **1976**, 22, 245–268.
- [124] Jarzynski, C. *Physical Review E* **1997**, 56, 5018–5035.
- [125] Jarzynski, C. *Physical Review E* **1997**, 56, 5018–5035.
- [126] Crooks, G. E. *Physical Review E* **1999**, 60, 2721–2726.
- [127] Crooks, G. E. *Physical Review E* **1999**, 60, 2721–2726.
- [128] Goette, M.; Grubmüller, H. *Journal of Computational Chemistry* **2009**, 30, 447–456.
- [129] Bader, R. F. W. *Accounts of Chemical Research* **1985**, 18, 9–15.
- [130] 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.
- [131] Fedorov, D. V.; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. *Physical Review Letters* **2018**, 121, 183401.
- [132] Rosales, A. R.; Quinn, T. R.; Wahlers, J.; Tomberg, A.; Zhang, X.; Helquist, P.; Wiest, O.; Norrby, P.-O. *Chemical communications (Cambridge, England)* **2018**, 54, 8294–8311.
- [133] Allen, A. E. A.; Payne, M. C.; Cole, D. J. *Journal of Chemical Theory and Computation* **2018**, 14, 274–281.
- [134] Newton, M. S.; Arcus, V. L.; Gerth, M. L.; Patrick, W. M. *Current Opinion in Structural Biology* **2018**, 48, 110–116.
- [135] Notin, P.; Rollins, N.; Gal, Y.; Sander, C.; Marks, D. *Nature Biotechnology* **2024**, 42, 216–228.

- [136] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, 1-2, 19–25.
- [137] Becke, A. D. *The Journal of Chemical Physics* **1993**, 98, 5648–5652.
- [138] Weigend, F.; Ahlrichs, R. *Physical Chemistry Chemical Physics* **2005**, 7.
- [139] Bunzel, H. A.; Anderson, J. L. R.; Hilvert, D.; Arcus, V. L.; van der Kamp, M. W.; Mulholland, A. J. *Nature Chemistry* **2021**, 13, 1017–1022.
- [140] Schlegel, H. B. *Journal of Computational Chemistry* **1982**, 3, 214–218.
- [141] Frisch, M. J.; others Gaussian 16 Revision C.01. Gaussian Inc.: Wallingford, CT, 2016.
- [142] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. *Nature Methods* **2017**, 14, 71–73.
- [143] Recabarren, R.; Tinzl, M.; Saez, D. A.; Gomez, A.; Erb, T. J.; Vöhringer-Martinez, E. *ACS Catalysis* **2023**, 13, 6230–6241.
- [144] Petrov, D.; Perthold, J. W.; Oostenbrink, C.; de Groot, B. L.; Gapsys, V. *Journal of Chemical Theory and Computation* **2024**, 20, 914–925.
- [145] Horn, P. R.; Mao, Y.; Head-Gordon, M. *The Journal of Chemical Physics* **2016**, 144, 114107.
- [146] Carter-Fenk, K.; Liu, M.; Pujal, L.; Loipersberger, M.; Tsanai, M.; Vernon, R. M.; Forman-Kay, J. D.; Head-Gordon, M.; Heidar-Zadeh, F.; Head-Gordon, T. *Journal of the American Chemical Society* **2023**, 145, 24836–24851.
- [147] Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. *eLife* **2018**, 7, e31486.
- [148] Neese, F. *WIREs Computational Molecular Science* **2022**, 12.
- [149] Mardirossian, N.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2014**, 16, 9904–9924.

- [150] Verstraelen, T. et al. *Journal of Computational Chemistry* **2021**, *42*, 458–464.
- [151] Chan, M.; Verstraelen, T.; Tehrani, A.; Richer, M.; Yang, X. D.; Kim, T. D.; Vöhringer-Martinez, E.; Heidar-Zadeh, F.; Ayers, P. W. *The Journal of Chemical Physics* **2024**, *160*.
- [152] Heidar-Zadeh, F.; Richer, M.; Fias, S.; Miranda-Quintana, R. A.; Chan, M.; Franco-Pérez, M.; González-Espinoza, C. E.; Kim, T. D.; Lanssens, C.; Patel, A. H. G.; Yang, X. D.; Vöhringer-Martinez, E.; Cárdenas, C.; Verstraelen, T.; Ayers, P. W. *Chemical Physics Letters* **2016**, *660*, 307–312.
- [153] Pujal, L.; Tehrani, A.; Heidar-Zadeh, F. *Conceptual Density Functional Theory*; John Wiley & Sons, Ltd, 2022; Chapter 32, pp 649–661.
- [154] 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.
- [155] Kim, T. D. et al. *The Journal of Chemical Physics* **2024**, *161*, 042503.
- [156] Tehrani, A.; Richer, M.; Heidar-Zadeh, F. *The Journal of Chemical Physics* **2024**, *161*, 072501.
- [157] Becke, A. D. *The Journal of Chemical Physics* **1988**, *88*, 2547–2553.
- [158] Shao, Y. et al. *Molecular Physics* **2015**, *113*, 184–215.
- [159] Mao, Y.; Loipersberger, M.; Horn, P. R.; Das, A.; Demerdash, O.; Levine, D. S.; Prasad Veccham, S.; Head-Gordon, T.; Head-Gordon, M. *Annual Review of Physical Chemistry* **2021**, *72*, 641–666.
- [160] Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. *The Journal of Chemical Physics* **2007**, *126*, 144111.
- [161] 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.

- [162] 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.
- [163] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.
- [164] Parker, T. M.; Sherrill, C. D. *Journal of Chemical Theory and Computation* **2015**, *11*, 4197–4204.
- [165] Rackers, J. A.; Wang, Q.; Liu, C.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. *Physical chemistry chemical physics : PCCP* **2016**, *19*, 276–291.
- [166] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *The Journal of Chemical Physics* **2010**, *132*, 154104.
- [167] Zgarbová, M.; Otyepka, M.; Šponer, J.; Hobza, P.; Jurečka, P. *Physical Chemistry Chemical Physics* **2010**, *12*, 10476–10493.
- [168] Fierke, C. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 4085–4092.
- [169] Cody, V.; Pace, J.; Makin, J.; Piraino, J.; Queener, S. F.; Rosowsky, A. *Biochemistry* **2009**, *48*, 1702–1711.
- [170] Strader, M. B.; Smiley, R. D.; Stinnett, L. G.; VerBerkmoes, N. C.; Howell, E. E. *Biochemistry* **2001**, *40*, 11344–11352.
- [171] Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *ACS Catalysis* **2019**, *9*, 5902–5911.
- [172] Doron, D.; Kohen, A.; Major, D. T. *Journal of Chemical Theory and Computation* **2012**, *8*, 2484–2496.
- [173] RuizPernía, J. J.; Williams, I. H. *Chemistry - A European Journal* **2012**, *18*, 9405–9414.
- [174] Ruiz-Pernia, J. J.; Luk, L. Y. P.; García-Meseguer, R.; Martí, S.; Loveridge, E. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *Journal of the American Chemical Society* **2013**, *135*, 18689–18696.

- [175] Case, D. A. et al. Amber18. University of California, San Francisco, 2018.
- [176] Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2008**, *120*, 215–241.
- [177] Luk, L. Y. P.; Javier Ruiz-Pernía, J.; Dawson, W. M.; Roca, M.; Loveridge, E. J.; Glowacki, D. R.; Harvey, J. N.; Mulholland, A. J.; Tuñón, I.; Moliner, V.; Allemann, R. K. *Proceedings of the National Academy of Sciences* **2013**, *110*, 16344–16349.
- [178] Lalush, D. S.; Wernick, M. N. *Emission Tomography*; Elsevier, 2004; pp 443–472.
- [179] Menon, A.; Pascazio, L.; Nurkowski, D.; Farazi, F.; Mosbach, S.; Akroyd, J.; Kraft, M. *ACS Omega* **2023**, *8*, 2462–2475.
- [180] Velázquez-Libera, J. L.; Recabarren, R.; Saez, D. A.; Castillo, C.; Ruiz-Pernía, J. J.; Tuñón, I.; Vöhringer-Martinez, E. Adapted DFTB3 repulsive potentials reach DFT accuracy for hydride transfer reactions in enzymes. 2025; [chemrxivPreprint](#).
- [181] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1995**, *16*, 1339–1350.
- [182] Khabibrakhmanov, A.; Fedorov, D. V.; Tkatchenko, A. *Journal of Chemical Theory and Computation* **2023**, *19*, 7895–7907.

A. Análisis de los valores de energía y parámetros no enlazantes para las interacciones entre residuos aminoacídicos

En este anexo se añaden resultados complementarios a la sección 3.2. Las figuras A.1, A.2, A.3 y A.4 muestran la estimación de la energía electrostática en presencia y ausencia de correcciones por penetración de carga, utilizando todos los modelos de cargas atómicas estudiadas en esta tesis. Por otro lado las figuras A.5 y A.6 dan a conocer los valores de los parámetros de lennard-Jonnes C_6 y radio de VdW, para los distintos métodos AIM y campos de fuerza convencionales.

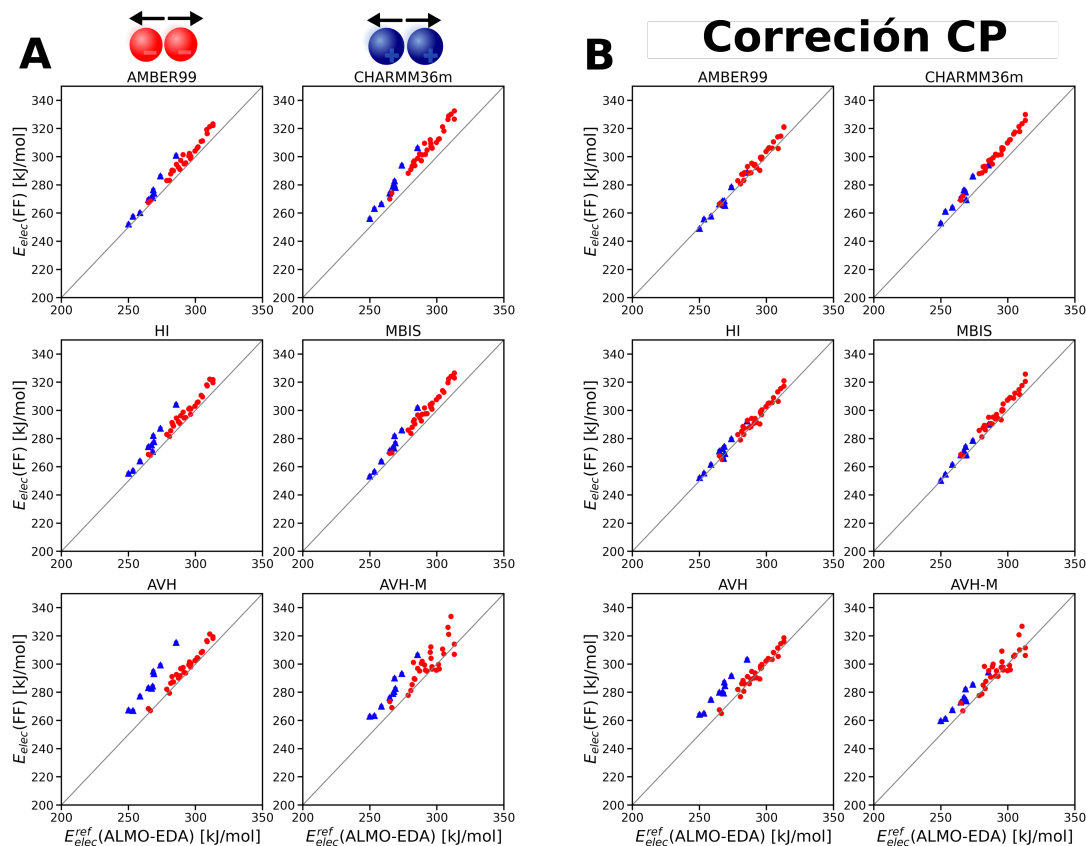


Figura A.1: A) Gráfico de paridad de la energía de interacción electrostática clásica calculada con cargas atómicas de dos campos de fuerza tradicionales (AMBER99, CHARMM36m) o mediante diferentes métodos AIM (HI, MBIS, AVH y AVH-M), en comparación con los valores de referencia obtenidos a partir de cálculos de estructura electrónica ALMO-EDA para dímeros de cadenas laterales cargadas de igual signo. B) Los mismos gráficos de paridad, incluyendo el efecto de la penetración de carga.

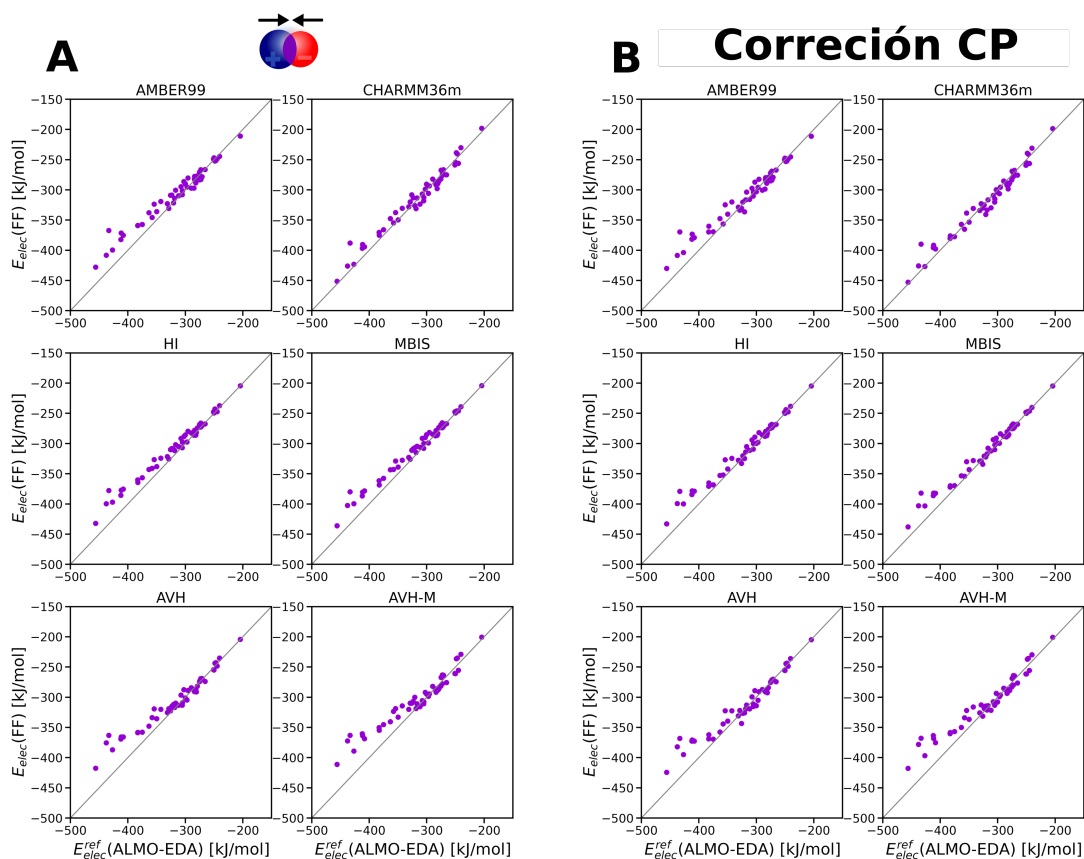


Figura A.2: A) Gráfico de paridad de la energía de interacción electrostática clásica calculada con cargas atómicas de dos campos de fuerza tradicionales (AMBER99, CHARMM36m) o mediante diferentes métodos AIM (HI, MBIS, AVH y AVH-M), en comparación con los valores de referencia obtenidos a partir de cálculos de estructura electrónica ALMO-EDA para dímeros de cadenas laterales cargadas de signo opuesto. B) Los mismos gráficos de paridad, incluyendo el efecto de la penetración de carga.

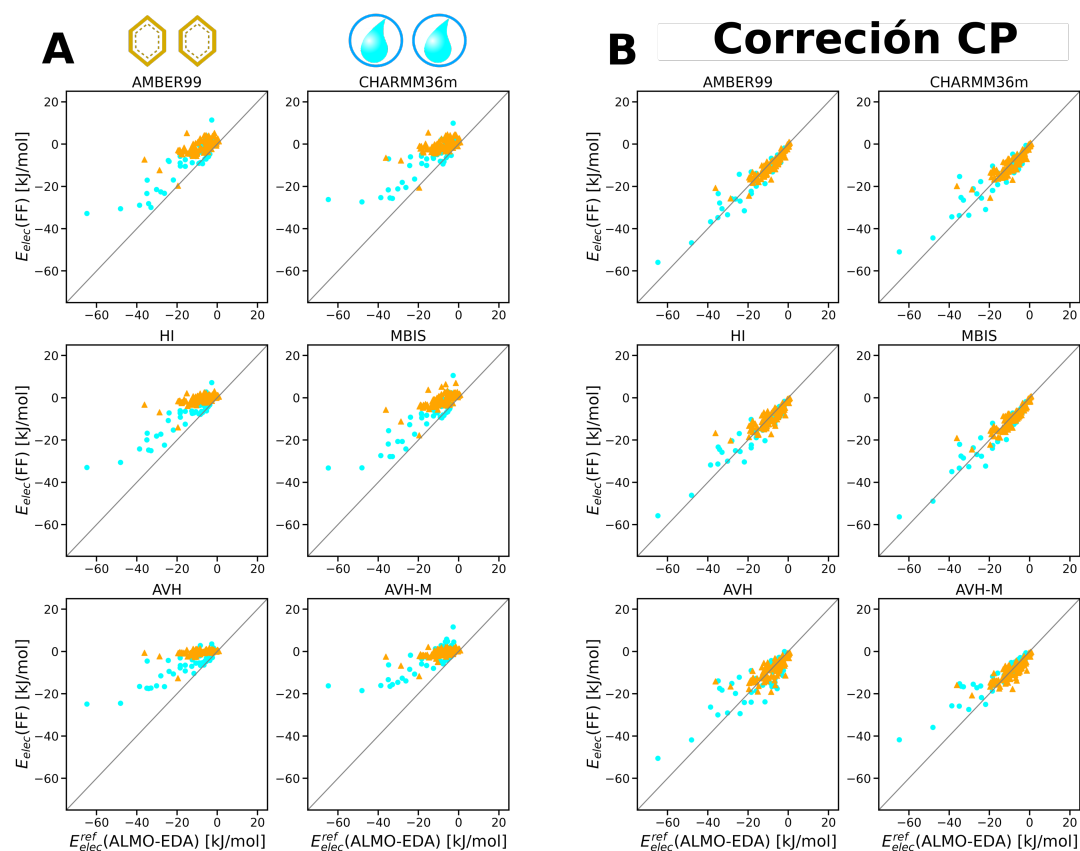


Figura A.3: A) Gráfico de paridad de la energía de interacción electrostática clásica calculada con cargas atómicas de dos campos de fuerza tradicionales (AMBER99, CHARMM36m) o mediante diferentes métodos AIM (HI, MBIS, AVH y AVH-M), en comparación con los valores de referencia obtenidos a partir de cálculos de estructura electrónica ALMO-EDA para dímeros de cadenas laterales aromático-aromático e hidrofílico-hidrofílico. B) Los mismos gráficos de paridad, incluyendo el efecto de la penetración de carga.

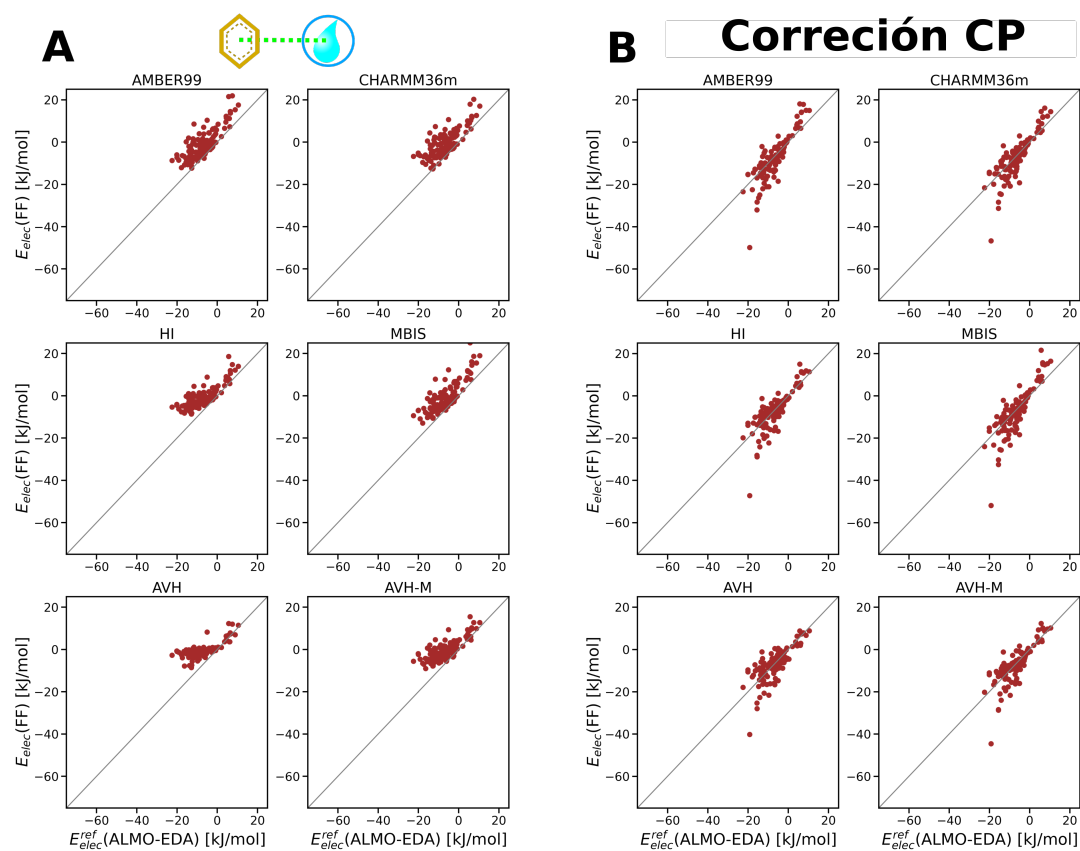


Figura A.4: A) Gráfico de paridad de la energía de interacción electrostática clásica calculada con cargas atómicas de dos campos de fuerza tradicionales (AMBER99, CHARMM36m) o mediante diferentes métodos AIM (HI, MBIS, AVH y AVH-M), en comparación con los valores de referencia obtenidos a partir de cálculos de estructura electrónica ALMO-EDA para dímeros de cadenas laterales aromático-hidrofílico. B) Los mismos gráficos de paridad, incluyendo el efecto de la penetración de carga.

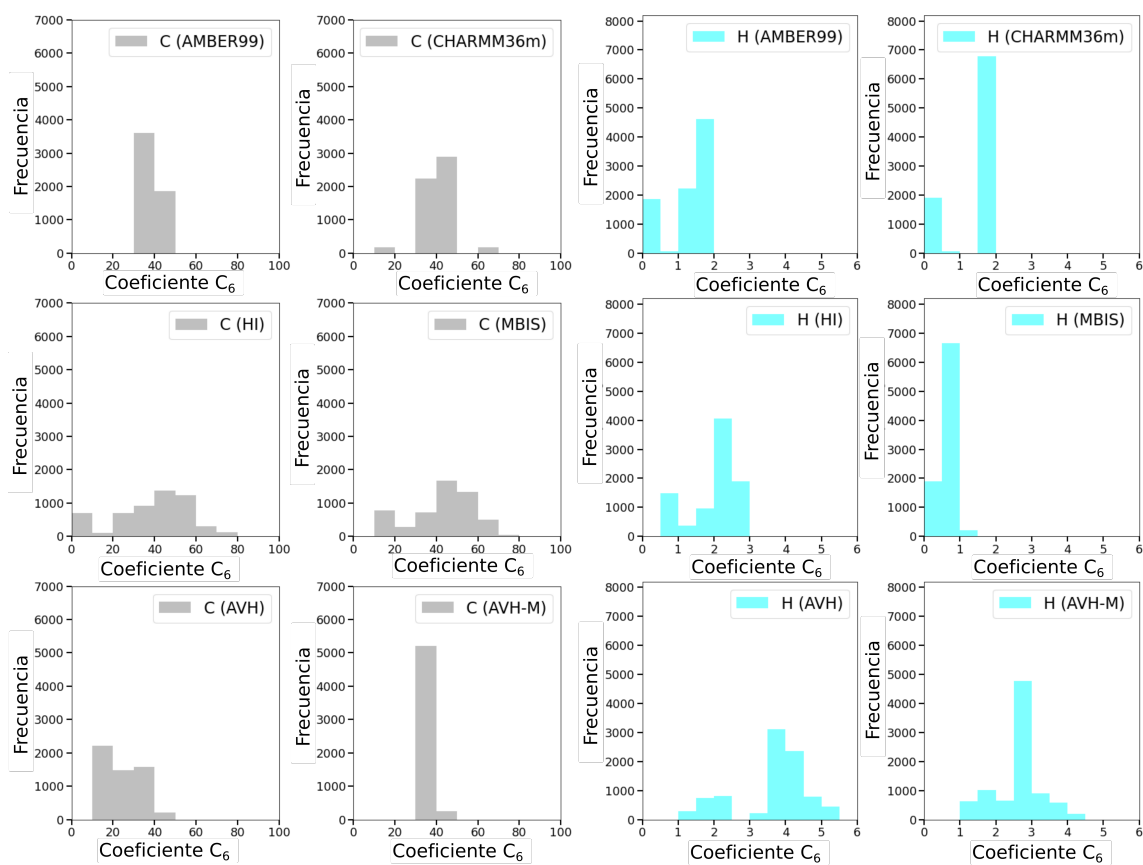


Figura A.5: Distribución de los coeficientes C_6 para los átomos de carbono (C, gris) e hidrógeno (H, cian) según los campos de fuerza AMBER99 y CHARMM36, así como diferentes métodos AIM (HI, MBIS, AVH y AVH-M).

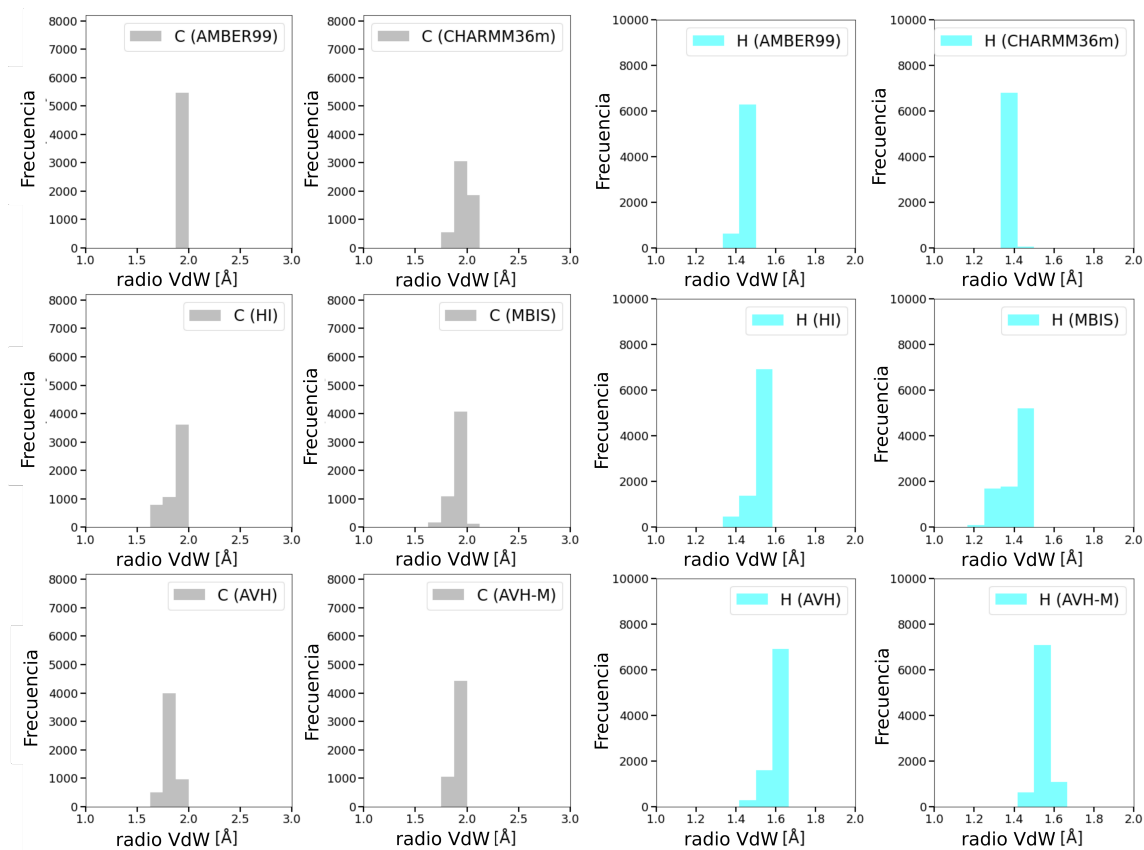


Figura A.6: Distribución de los coeficientes radios de VdW para los átomos de carbono (C, gris) e hidrógeno (H, cian) según los campos de fuerza AMBER99 y CHARMM36, así como diferentes métodos AIM (HI, MBIS, AVH y AVH-M).

B. Corrección al perfil de energía libre usando el método DFTB3

En este anexo se añaden resultados complementarios a la sección 3.4. La figura B.1 muestra los perfiles de energía libre de la enzima DHFR obtenidos con el método de la cuerda y el método semiempírico DFTB3 original (curva negra) y luego de realizarle correcciones sumando la diferencia entre este nivel de teoría y M06-2X/Def2tzvp (curva roja).

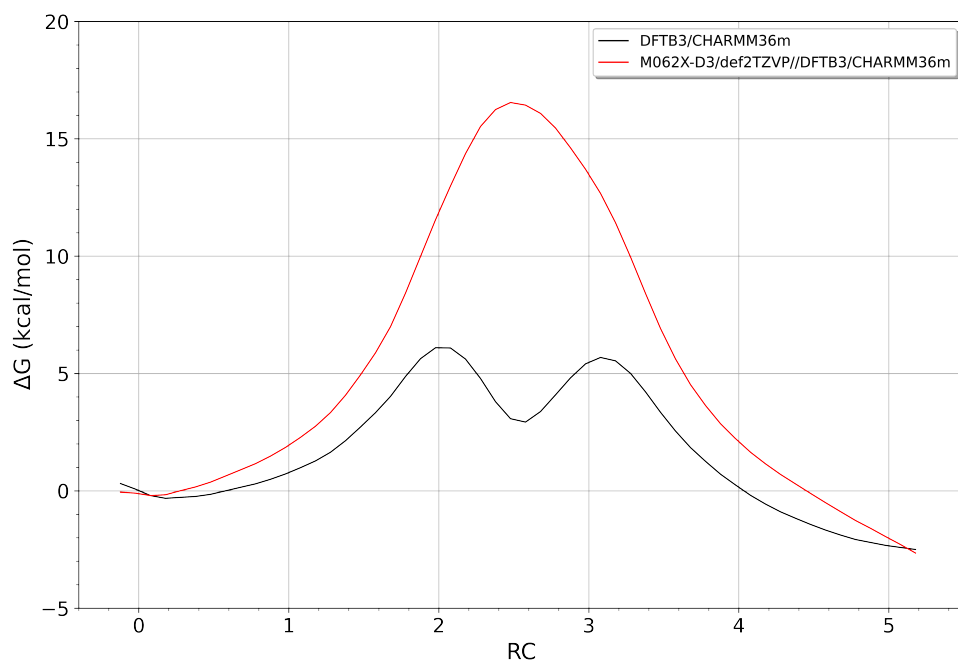


Figura B.1: Corrección al perfil de energía libre de DHFR. La curva negra corresponde al perfil original obtenido mediante el Hamiltoniano DFTB3, exhibiendo un mínimo artificial cercano al ET. La curva roja muestra el perfil corregido por el método DFT M06-2X/def2tzvp.

C. Parámetros de Simulación

Todos los parámetros de simulación referentes a equilibraciones, producciones y transformaciones alquímicas MM han sido almacenados como archivos en formato *mdp* de Gromacs. Los parámetros de las Simulaciones QM/MM en AMBER a su vez, son almacenados en su propio formato. Las estructuras iniciales de cada sistema así como los parámetros de simulación anteriormente mencionados pueden ser descargados del siguiente enlace:

<https://doi.org/10.5281/zenodo.15116078>

D. Análisis Conformacional del sistema OYE

En este anexo se añaden resultados complementarios a la sección 3.6. Las figuras D.1 y D.2 corresponden a cálculos de la desviación cuadrática media (RMSD) para las simulaciones de todas las mutantes en el estado de transición ET y en el complejo de Michaelis-Menten.

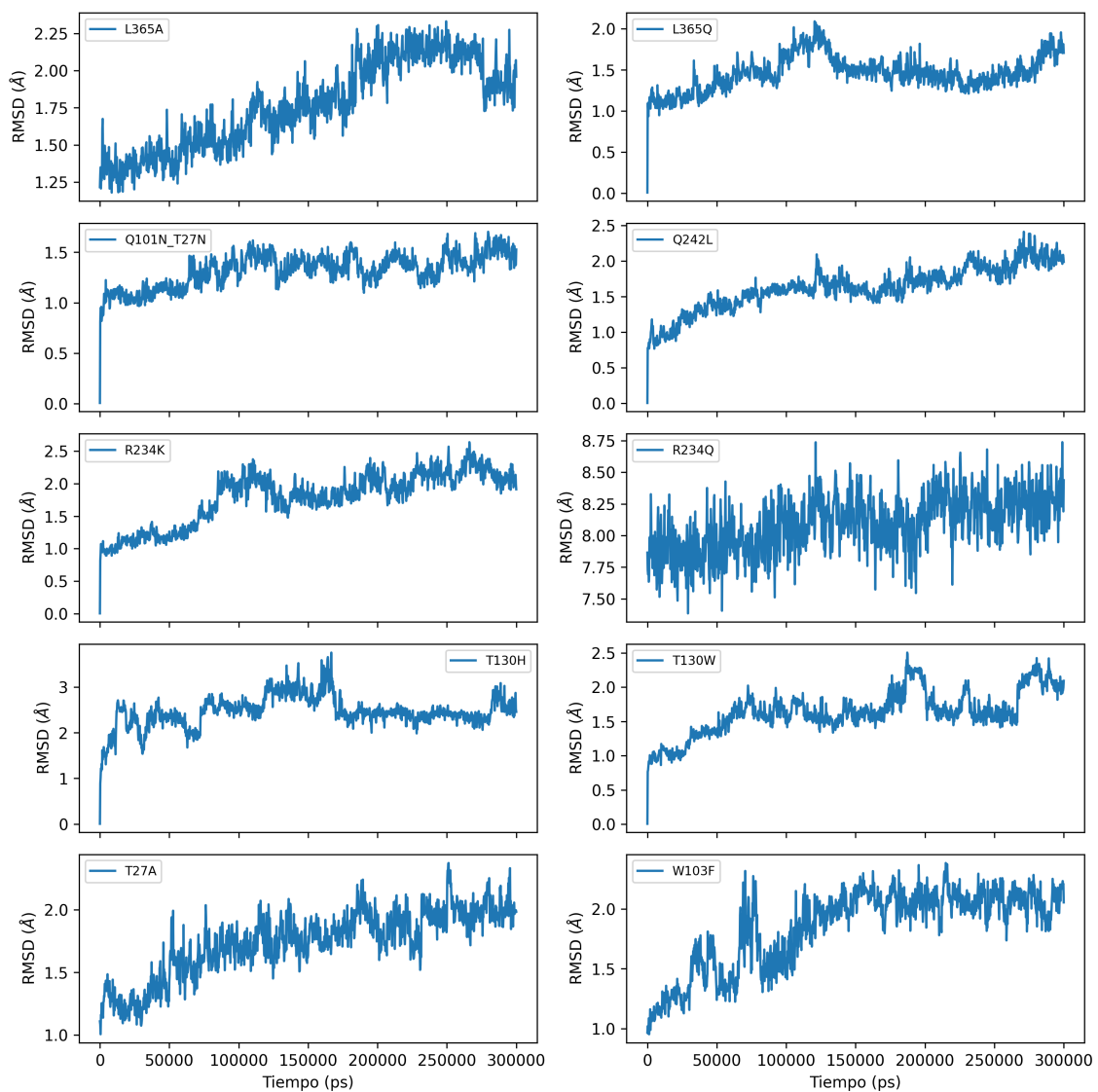


Figura D.1: Evolución del RMSD de la proteína OYE en el complejo de Michaelis-Menten, para diferentes mutantes a lo largo del tiempo. Se muestran los valores de RMSD (en Å) en función del tiempo (en ps) para cada mutante analizado

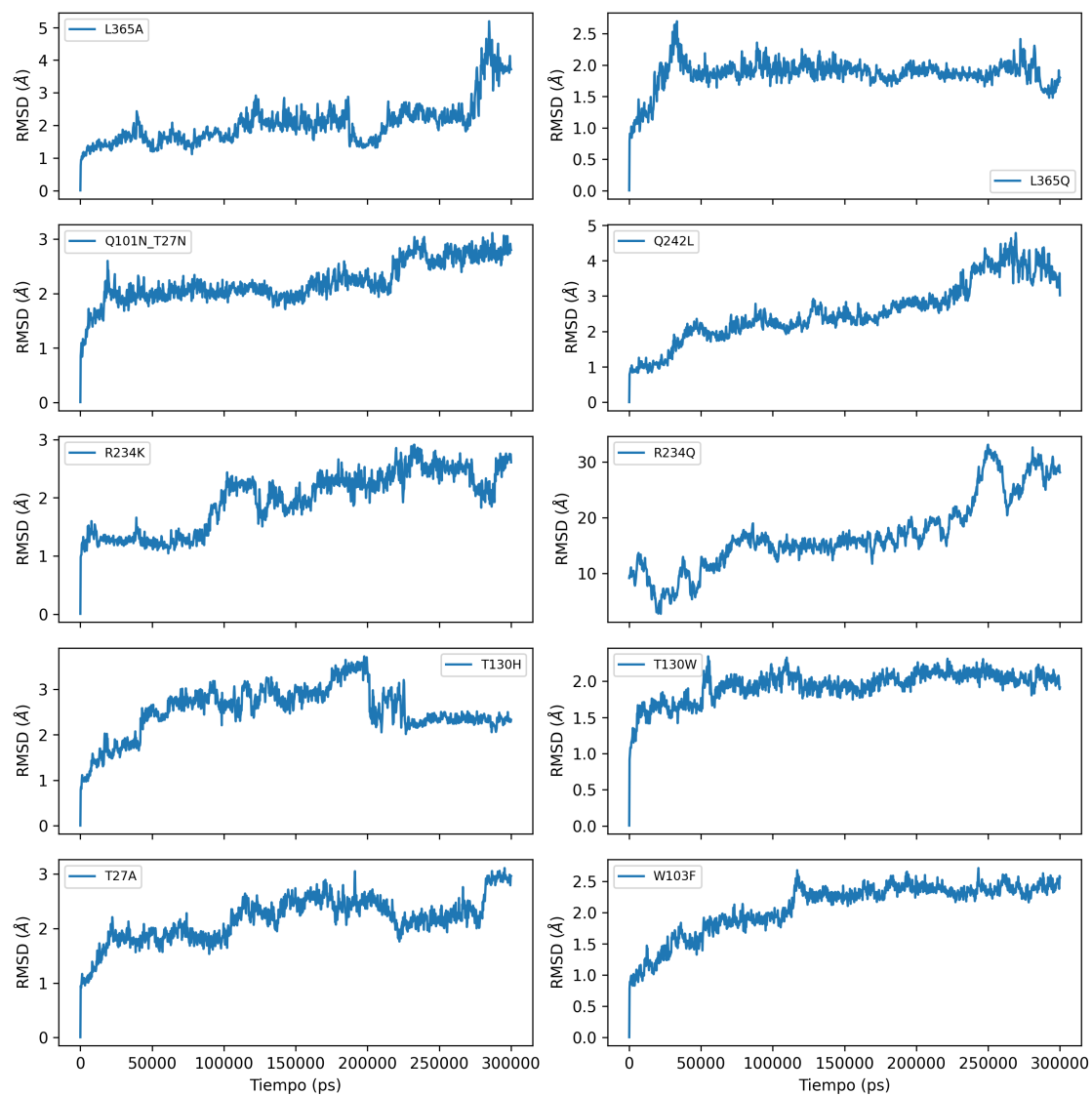


Figura D.2: Evolución del RMSD de la proteína OYE en el estado de transición, para diferentes mutantes a lo largo del tiempo. Se muestran los valores de RMSD (en Å) en función del tiempo (en ps) para cada mutante analizado

E. Resultados Experimentales

Los resultados experimentales presentados en esta sección corresponden al trabajo del estudiante de maestría Tobias Dörer, el cual realizó este trabajo en el marco de su *Lab rotation* en el *Max Planck for Terrestrial Microbiology*

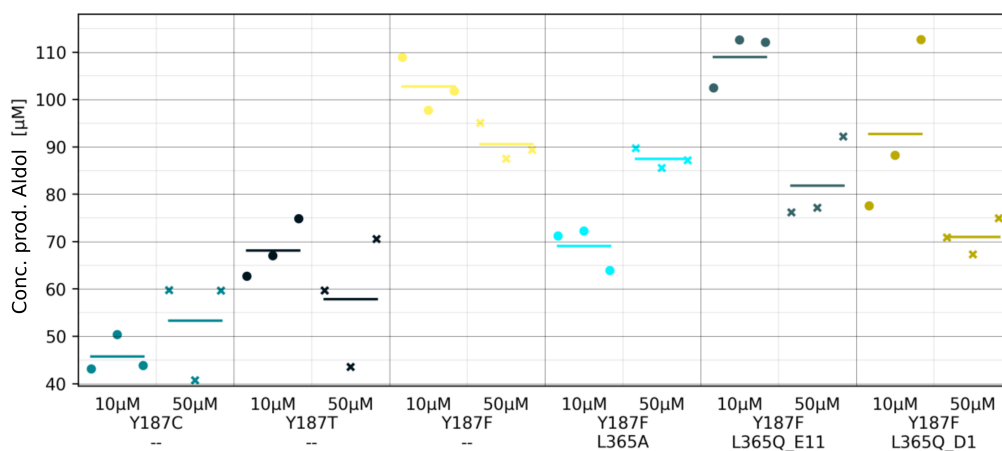


Figura E.1: Actividad aldólica de distintas variantes de OYE. La concentración de producto aldólico [μM] se mide en función de las distintas variantes de OYE Y187F y a dos concentraciones de enzima total distintas 10 y 50 μM . Para la mutante Y187F/L365Q el ensayo se realiza dos veces, donde los códigos E11 y D1 corresponden a la posición en una placa de 96 pocillos, como se suele utilizar en este tipo de ensayos.

F. Presentaciones en congresos y publicaciones

F.1. Congresos

1. *QCMM-2022, 2022*, Universidad de Concepción, Chile, póster titulado *Conformational Changes of Glucose-6-phosphate dehydrogenase induced by NADP⁺ binding and Class I mutations*
2. *19th Toronto Chemical Biophysics Symposium, 2023*, Universidad de Toronto, Canadá, presentación oral titulada *Non-bonded parameters for biological force field derived from Atom-in-molecules approaches*
3. *12^o Workshop en Química Computacional y Espectroscopía Molecular, 2023*, Universidad Andrés Bello, Chile presentación oral titulada *Non-bonded parameters for biological force field derived from Atom-in-molecules approaches*
4. *II Conesul Symposium on Biomolecular Simulation, 2024*, Universidad de Buenos Aires, Argentina, poster titulado *Non-bonded parameters for biological force field derived from Atom-in-molecules approaches*
5. *III Reunión anual de la sociedad chilena de bioinformática, 2024*, Universidad de Concepción, Chile, poster titulado *Non-bonded parameters for biological force field derived from Atom-in-molecules approaches*
6. *LatinXChem24(online), 2024*, poster titulado *Non-bonded parameters for biological force field derived from Atom-in-molecules approaches*

F.2. Publicaciones

1. **Nonbonded Force Field Parameters from Minimal Basis Iterative Stockholder Partitioning of the Molecular Electron Density Improve CB7 Host–Guest Affinity Predictions** Duván González, Luis Macaya, Carlos Castillo-Orellana, Toon Verstraelen, Stefan Vogt-Geisse, and Esteban Vöhringer-Martinez. *Journal of Chemical Information and Modeling* 2022 67 (17), 4162-4174 DOI:[10.1021/acs.jcim.2c00316](https://doi.org/10.1021/acs.jcim.2c00316)
2. **Variational Hirshfeld Partitioning: Mathematical Framework and Additive Variational Hirshfeld Scheme** Farnaz Heidar-Zadeh, Carlos Castillo-Orellana, Maximilian van Zyl, Leila Pujal, Toon Verstraelen, Patrick Bultinck, Esteban Vöhringer-Martinez, and Paul W. Ayers. *Journal of Chemical Theory and Computation* 2024 20 (22), 9939-9953 DOI:[10.1021/acs.jctc.4c01077](https://doi.org/10.1021/acs.jctc.4c01077)
3. **Nonbonded force field parameters derived from Atom-In-Molecules methods reproduce interactions in proteins from first–principles calculations** Carlos Castillo-Orellana, Farnaz Heidar-Zadeh and Esteban Vöhringer-Martinez. *Journal of Chemical Theory and Computation* 2025 21 (4), 2043-2054 DOI:[10.1021/acs.jctc.4c01406](https://doi.org/10.1021/acs.jctc.4c01406)
4. **Non-Covalent Interactions and ChargeTransfer in the CO₂ Activation by Low Valent Group 14 Complexes** Carlos Castillo-Orellana, Esteban Vöhringer-Martinez, and Nery Villegas-Escobar. *Journal of Molecular Modelling* 2024 30 (11), 365 DOI:[10.1007/s00894-024-06150-5](https://doi.org/10.1007/s00894-024-06150-5)
5. **Adapted DFTB3 repulsive potentials reach DFT accuracy for hydride transfer reactions in enzymes** José Luís Velázquez-Libera, Rodrigo Recabarren, David Adrian Saez, Carlos Castillo, J. Javier Ruiz-Pernía, Iñaki Tuñón, Esteban Vöhringer-Martinez. *Submitted to Journal of Computational Chemistry. Preprint* DOI: [10.26434/chemrxiv-2024-kkwwg-v4](https://doi.org/10.26434/chemrxiv-2024-kkwwg-v4)
6. **Permanent Electrostatic Moments Through the Lens of Atoms: How Accurate Are Lower-Rank Approximations of Molecular Moments**

from Variational Hirshfeld Atoms? Maximilian Van zyl, Carlos Castillo-Orellana, Leila Pujal, Farnaz Heidar-Zadeh, Submitted to Physical Chemistry Chemical Physics.

7. **Development of the Biocatalytic Reductive Aldol Reaction** Gabriele M. Stoffel, Matthias Tinzl, David Adrian Saez, Patrick D. Gerlinger, Rodrigo Recabarren, Carlos Castillo, Timothy Bradley, Hendrik Westedt, Pascal Pfister, Aharon Gomez, Nicole Paczia, Marc-Olivier Ebert, Esteban Vöhringer-Martinez, Tobias J. Erb, *Submitted to Nature Catalysis*.

F.3. Pasantías y Reconocimientos

1. Pasantía de investigación en el departamento de química de la Universidad de Queen en Kingston, Canadá, bajo la tutela de la Dra. Farnaz Heidar-Zadeh desde noviembre de 2022 hasta Julio de 2023
2. Beneficiario de la beca canadiense Emerging Leaders in the Americas Program (ELAP) para financiar la estadía en Canadá bajo la tutela de la Dra. Heidar-Zadeh