

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE MECÁNICA

VALPARAÍSO – CHILE



“COMPARACIÓN DE ENERGÍAS DE
SOLVATACIÓN CALCULADAS CON POISSON-
BOLTZMANN Y DINÁMICA MOLECULAR PARA
MOLECULAS PEQUEÑAS”

RODRIGO ANDRÉS HUENCHO MORA

MEMORIA DE TITULACIÓN PARA OPTAR AL
TÍTULO DE INGENIERO CIVIL MECÁNICO

PROFESOR GUÍA

PH.D. CHRISTOPHER COOPER

PROFESOR CORREFERENTE

PH.D. HARVEY ZAMBRANO

NOVIEMBRE 2023

Agradecimientos

Agradecer a mis amigos por hacer la estadía fuera increíble y estar ahí cuando más lo necesitaba.

A mis ex-compañeros de Civil Industrial que siguieron conmigo una vez que me fui, a los mecánicos que me aceptaron rápidamente en su grupo, a quienes compartí clases y me ayudaron durante esta etapa, a todos ellos que hacen que me sienta contento y orgulloso de ser parte de esta universidad.

Al tremendo grupo de basquet y los eternos partidos en la 5x5. A los de "Huenshow F.C", los bicampeones del baby-industrial. A mis compañeros de Rishilagoza F.C" que me enseñaron que por mas que cueste, siempre hay que dar la cara. Y especial mención a los chilotes de "El Sapete F.C" por tantos años de amistad y fútbol.

A mi familia en Chiloé por su apoyo y sacrificio que me permitió disfrutar de una tranquila vida universitaria llena de oportunidades. A mi segunda familia en Santiago por aguantarme estos últimos años y en parte de la pandemia.

Al profesor Cooper, por su apoyo y tiempo durante todo el trabajo.

Índice

Introducción	1
Capítulo 1: Marco Teórico	
1.1. Modelo de Solvente Implícito	3
1.1.1. Electrostática en el vacío	3
1.1.2. Electrostática en un medio continuo dieléctrico	4
1.1.3. Ecuación de Poisson-Boltzmann	5
1.2. Modelo de Solvente Implícito en Sistemas Moleculares	6
1.2.1. Cálculo de la Energía	9
1.3. Método de Elementos de Borde	10
1.3.1. Ecuación de Laplace	11
1.3.2. Sistema Lineal	15
1.4. Estrategia de aceleración - Treecode	16
1.4.1. Treecode	17
1.5. Estado del Arte	20
Capítulo 2: Metodología	
2.1. Modelo de Superficie Simple	22
2.2. Energía en los Bordes	24
2.3. Python GPU for Boundary elements - PyGBe	24
2.4. Generación de Superficie	26
2.5. Trabajo para Múltiples Moléculas	27
2.5.1. Conversión de Formatos	28
2.5.2. Generación de Parámetros y Configuración	29
2.5.3. Generación de Geometría	30
2.5.4. Generación de Base de datos	30
2.5.5. Comparación Método Implícito y Método Explícito	32
Capítulo 3: Resultados y Análisis de Resultados	
3.1. Comparación entre métodos en base a Número de átomos, Volumen y Superficie	35
3.2. Comparación entre métodos en base a presencia de átomos o enlaces	38
3.2.1. Nitrógeno y Oxígeno	40
3.2.2. Doble Enlace - = ...	41
3.3. Comparación entre métodos en base a un Grupo Funcional	42

3.3.1. Compuestos Aromáticos	44
3.3.2. Compuestos No-Aromáticos	46
3.3.3. Ácidos y Aminas	49
3.3.4. Alcanos	52
3.4. Resultados finales.	53
Capítulo 4: Conclusiones	

Índice de figuras

1.1. Definición de Superficies. Fuente: [8]	7
1.2. Definición de Superficies. Fuente: [8]	8
1.3. Singularidad en superficie. Fuente: [8]	13
1.4. Algoritmo Treecode. Fuente [8]	19
2.1. Discretización de la superficie de la proteína. Fuente:[8]	23
2.2. Diagrama generación de superficie	27
2.3. Documento xxx”.config”.	30
2.4. Diagrama de flujo de programa 1.	31
2.5. Diagrama de flujo de programa 2.	32
2.6. Gráfico de Resultados BEM (Pygbe) vs Dinámica Molecular, echo en el programa energy plots.	33
3.1. Gráfico de Energy Plots. $\epsilon = 1, 35$; Barra Color = N° de Átomos: de 4 a 9; Coeficiente de Correlación = 0,9758; RMSD = 0,4566; MUE = 0,2814.	36
3.2. Gráfico de Energy Plots. $\epsilon = 1, 25$; Barra Color = N° de Átomos: de 10 a 15; Coeficiente de Correlación = 0,9465; RMSD = 0,8660; MUE = 0,5783.	36
3.3. Gráfico de Energy Plots. $\epsilon = 1, 20$; Barra Color = N° de Átomos: de 16 a 18; Coeficiente de Correlación = 0,9424; RMSD = 0,9303; MUE = 0,7086.	37
3.4. Gráfico de Energy Plots. $\epsilon = 1, 20$; Barra Color = N° de Átomos: de 19 a 20; Coeficiente de Correlación = 0,9241; RMSD = 1,0861; MUE = 0,8122.	37
3.5. Gráfico de Energy Plots. $\epsilon = 1, 15$; Moléculas con átomos de Carbono; Coeficiente de Correlación = 0,9468; RMSD = 0,9098; MUE = 0,6254. Barra de color representa el número de átomos.	39
3.6. Gráfico de Energy Plots. $\epsilon = 1, 15$; Moléculas con átomos de Nitrógeno; Coeficiente de Correlación = 0,8824; RMSD = 1,1425; MUE = 0,8164. En rojo destacadas las moléculas con presencia de Triple Enlace.	40
3.7. Gráfico de Energy Plots. $\epsilon = 1, 20$; Moléculas con átomos de Oxígeno; Coeficiente de Correlación = 0,8704; RMSD = 1,1434; MUE = 0,9110. Barra de color representa el número de átomos.	40
3.8. Gráfico de Energy Plots. $\epsilon = 1, 15$; Moléculas con Doble Enlace; Coeficiente de Correlación = 0,9426; RMSD = 0,9330; MUE = 0,7143. Barra de color representa el número de átomos.	41
3.9. Familias de grupos funcionales.	43

- 3.10. Gráfico de Energy Plots. $\epsilon = 1, 25$; Compuestos Aromáticos; Coeficiente de Correlación = 0,9391; RMSD = 0,9320; MUE = 0,7431. Barra de color representa el número de átomos. 44
- 3.11. Gráfico de Energy Plots. Moléculas Aromáticas Cuadro 3.3; Coeficiente de Correlación = 0,9854; RMSD = 0,3759; MUE = 0,2750. Barra de color representa el número de átomos. 46
- 3.12. Gráfico de Energy Plots. Moléculas Alifáticas Cuadro 3.4; Coeficiente de Correlación = 0,9838; RMSD = 0,3671; MUE = 0,2214. Barra de color representa el número de átomos. 47
- 3.13. Gráfico de Energy Plots. $\epsilon = 1, 35$; Alcoholes; Coeficiente de Correlación = 0,8527; RMSD = 1,0688; MUE = 0,8217. Barra de color representa el número de átomos. 48
- 3.14. Gráfico de Energy Plots. $\epsilon = 1, 10$; Ácidos y Derivados; Coeficiente de Correlación = 0,8420; RMSD = 1,2251; MUE = 0,9641. Barra de color representa el número de átomos. 50
- 3.15. Molécula de Amoniaco. 51
- 3.16. Gráfico de Energy Plots. $\epsilon = 1, 10$; Aminas; Coeficiente de Correlación = 0,8372; RMSD = 1,2008; MUE = 0,9842. Barra de color representa el número de átomos. 51
- 3.17. Gráfico de Energy Plots. 22 Moléculas Alcanas estudiadas; Coeficiente de Correlación = 0,6201; RMSD = 0,0420; MUE = 0,0393. Barra de color representa el número de átomos. 53
- 3.18. Gráfico de Energy Plots. Moléculas pertenecientes a grupos funcionales poco solubles en Agua y Alcanos; Coeficiente de Correlación = 0,9888; RMSD = 0,4231; MUE = 0,2701. Barra de color representa el número de átomos. . . . 54
- 3.19. Gráfico de Energy Plots. Moléculas solubles de grupos funcionales presentes en el Cuadro 3.5; Coeficiente de Correlación = 0,8831; RMSD = 1,2327; MUE = 0,7828. Barra de color representa el número de átomos. 55

Índice de cuadros

2.1. Estructura documento ".pqr".	28
2.2. Parámetros PyGBe en documento ".param".	29
3.1. Resultados obtenidos con el cálculo de 18 moléculas con presencia de Triple Enlace.	42
3.2. Compuestos Estudiados.	44
3.3. Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Aromáticos. Grupo Fenol (b) es el grupo Fenol sin considerar 2-nitrophenol; 2,5-dimethylphenol y 2,6-dimethylphenol.	45
3.4. Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Alifáticos.	47
3.5. Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Alifáticos poco predecibles.	48
3.6. Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, Ácidos y Derivados.	50
3.7. Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, Ácidos.	52
3.8. Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, 22 Compuestos Alcanos evaluados con distinta constante dieléctrica.	53

Listado de Símbolos

ϕ	Potencial electrostático promedio
$\epsilon \frac{\partial \phi}{\partial n}$	Desplazamiento eléctrico
ϵ	Permitividad del medio
ϵ_o	Permitividad del vacío
λ	Largo de Debye
κ	Inversa del largo de Debye
Ω_t	Dominio de t , 1 para el soluto y 2 para el solvente.
Γ	Frontera de Ω_1
ρ	Distribución de cargas
q	Carga del átomo
$\mathbf{r}_q$	Posición de la carga q
$\mathbf{F}$	Fuerza
$\mathbf{E}$	Campo eléctrico
c_{0i}	Número promedio de especies i
e^+	Carga de un protón
k_B	Constante de Stefan Boltzmann
T	Temperatura Absoluta
$\mathbf{n}$	Normal unitaria apuntando hacia afuera de Ω_{in}
$\tilde{K}$	Operador de potencial Double-Layer
$\tilde{V}$	Operador de potencial Single-Layer
K	Operador Double-Layer
V	Operador Single-Layer
γ	Operador de traza
G_A	Función de espacio libre de Green del operador diferencial lineal A
W	Trabajo
$\mathcal{D}$	Operador diferencial lineal

Glosario

BEM	Método de elementos de borde
PBE	Ecuación de Poisson Boltzmann
DM	Dinámica Molecular
GMRES	Método generalizado de residual mínimo
RMSD	Root mean square deviation
MUE	Mean unsigned error
SES	Superficie excluída por el solvente
MSMS	Maximal Speed Molecular Surface

Resumen Ejecutivo

La mayoría de las interacciones químico-biológicas que existen en la naturaleza se producen en medios acuosos. Para la predicción en la magnitud de las reacciones se han desarrollado diversos métodos y modelos como los explícitos, con la Dinámica Molecular, o los implícitos con el desarrollo de la ecuación de Poisson Boltzmann.

Al igual que en muchas áreas de la ciencia computacional, constantemente se discute el rendimiento y la eficacia de los diversos modelos que puedan existir, cuanto se puede sacrificar la precisión con tal de obtener una mejor eficiencia. Es por ello que se ha desarrollado una herramienta que permite acelerar la generación y recopilación de resultados, y otra para facilitar el análisis de los datos al permitir comparar por categorías lo obtenido al utilizar diversos métodos. La comparación entre el método implícito BEM y el método explícito DM se ha llevado a cabo utilizando la base de datos de cerca de 500 moléculas recopiladas por David L. Mobley.

Los análisis resultan poco concluyentes al categorizar las moléculas por su número de átomos o la presencia de alguno específico, como el Nitrógeno, así como la presencia del al menos un enlace doble no garantiza una buena correlación entre los métodos utilizados. Por otro lado, al comparar según el grupo funcional de cada molécula, es posible notar grupos que poseen una alta correlación y bajo RMSD al utilizar una determinada constante de permittividad en BEM y lo obtenido en DM. También se puede apreciar que el común en todos los grupos que poseen una alta correlación es su baja solubilidad en agua, la correlación lineal entre los métodos de todas estas moléculas es de 0,9888 contando con una RMSD de 0,4231; en contraste de la correlación de 0,8831 y RMSD de 1,237 en los grupos con alta solubilidad en agua.

Abstract

Most of the biomolecular interaction in the nature occur in an aqueous environment. For prediction in the magnitude of these reactions a variety of methods and models have been developed, explicit methods such as Molecular Dynamics or implicit methods using the Poisson-Boltzman equation.

As many disciplines in the computational science, there is a continuous discussion between performance and effectiveness, how much precision can be sacrificed for a better efficiency. That is why a tool has been developed to accelerate the generation and compilation of results, and another one to ease the data analysis by comparison in different categories between diverse methods. The analysis is performed using the David L. Mobley database of 500 molecules by comparing the implicit method BEM results and the explicit method MD calculated by Mobley.

The analysis are quite inconclusive when the molecules are categorized by number of atoms or the presence of one them, such as Nitrogen, as well as the presence of at least one double bond does not guarantee a correlation between BEM and MD. On the other hand, when comparing molecules according functional groups, it is possible to identify some of them with high correlation and low RMSD when a determined permittivity constant is used in BEM. Also, what the groups with a high correlation between methods have in common is low solubility in water, the correlation among these functional groups is 0,9888 with a RMSD of 0,4231; in contrast to the poor correlation of 0,8831 and RMSD of 1,237 in highly soluble compounds.

Introducción

La electrostática posee un importante rol en diversos aspectos químicos y biológicos. Entender las propiedades electrostáticas es básico para comprender e investigar los procesos moleculares, como lo puede ser la solvatación, unión y adsorción. Dado que la gran mayoría de estas interacciones ocurren en medios acuosos, es que se han desarrollado distintos modelos y métodos con la finalidad de obtener la energía de estos procesos, y así aplicarlas en la investigación de fármacos y otros procesos [2][17]

Dentro de los modelos desarrollados más populares existen los explícitos e implícitos. La principal diferencia entre ambos modelos radica en la forma de considerar las moléculas. En el modelo implícito se identifican 2 regiones, solvente y proteína, en los cuales se aplican ecuaciones electrostáticas como si fueran un dieléctrico. En cambio el modelo explícito trata el solvente en todo su detalle molecular. Ambos modelos poseen ventajas y desventajas, el hecho de que el solvente sea tratado en detalle en un modelo explícito, ofrece la capacidad de reflejar de mejor manera la reacción soluto-solvente que un modelo implícito, a cambio se tiene un mayor tiempo de cálculo computacional. Las limitantes computacionales hacen que los modelos explícitos estén acotados a moléculas pequeñas, ya que el costo del tiempo no vale su precisión. Por otro lado, el modelo implícito que se utiliza resuelve la ley de Gauss (Ecuación de Poisson) en el soluto, representado por puntos de carga o átomos, mientras que la presencia de sal en el solvente transforma la ley de Gauss en una ecuación de Poisson-Boltzmann.

En el presente informe, se estudia el uso de los modelos implícitos que implementan el método de elementos de frontera (BEM) para resolver las ecuaciones de Poisson y Poisson-Boltzmann. Para realizar los cálculos se utilizará PyGBe [9], un programa escrito principalmente en Python y C++. PyGBe trabaja mayormente en la unidad gráfica (GPU) mediante PYCUDA, es posible utilizar únicamente la CPU a falta de compatibilidad con la GPU, a cambio se incrementan los tiempos de cálculo.

Objetivos

Este trabajo tiene como objetivos:

1. Estudiar el modelo matemático que representa las interacciones electrostáticas entre moléculas y una solución iónica.
2. Aplicar el método de elementos de borde al caso de la electrostática en biomoléculas utilizando el código PyGBe.
3. Elaborar un código en el lenguaje de programación Python que permita transformar formatos y generar los parámetros aceptados por PyGBe.
4. Elaborar un código que permita automatizar la ejecución de los programas cuando se estudia una gran cantidad de moléculas.
5. Realizar un análisis de datos para obtener los parámetros que generen los mejores resultados según el tipo de molécula.

En el presente trabajo se realizará una comparación entre los modelos explícitos e implícitos, se contrastarán los resultados obtenidos mediante el método de Dinámica Molecular realizado por David L. Mobley [16] y los resultados obtenidos al utilizar PyGBe. El análisis considerará los grupos funcionales, como los átomos pertenecientes de las moléculas, para así intentar obtener los parámetros que mejor se adapten a cada situación.

Capítulo 1

Marco Teórico

La electrostática tiene un rol importante en la física molecular, la química y sus respectivas aplicaciones. Dada la gran cantidad de átomos que se encuentran enlazados entre sí producto de las fuerzas generadas por sus propias cargas, es que existen distintas formas de modelar la interacción de estas con su entorno. Este tipo de métodos son los explícitos, donde cada átomo se considera de manera independiente, y los implícitos, en donde el soluto y el solvente forman dos regiones distintas. Cada método posee ventajas y desventajas en términos de precisión, recursos computacionales y tiempo de proceso.

1.1. Modelo de Solvente Implícito

Biológicamente, las proteínas se encuentran en un medio acuoso que determinan su funcionalidad, en este caso incluir los efectos del agua es fundamental para el funcionamiento de cualquier modelo [4]. Ya se ha mencionado que el funcionamiento de un modelo de solvente implícito hace uso de la electrostática en un medio continuo dieléctrico (distinto nivel de polarización: ϵ) para representar el medio acuoso, y resuelve para el campo de potencial electrostático medio. A continuación se realizará una breve introducción a la teoría de la electrostática [8][13].

1.1.1. Electrostática en el vacío

Para comprender la electrostática es fundamental conocer los conceptos de carga (q) y como este reacciona en presencia de un campo eléctrico ($\mathbf{E}$, un vector dado por la fuerza sobre unidad de carga en un punto). La perturbación de la carga dentro de un campo electrostático está representado por la fuerza $\mathbf{F} = q\mathbf{E}$. En las ecuaciones de Maxwell y por ley de

Gauss, el campo eléctrico está relacionado con la densidad de distribución de carga (ρ) y la permisividad en el vacío (ϵ_0)

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \quad (1.1)$$

La ecuación (1.1) también puede expresarse de forma integral, aplicando una integral sobre el volumen en cada lado y utilizando el teorema de la divergencia en el lado izquierdo, se obtiene

$$\oint_{\Gamma} \mathbf{E} \cdot \mathbf{n} d\Gamma = \int_{\Omega} \frac{\rho}{\epsilon_0} d\Omega \quad (1.2)$$

donde Γ es la superficie que encierra a Ω , y $\mathbf{n}$ es el vector normal a la superficie Γ . El término izquierdo de la ecuación (1.2) es conocido como flujo eléctrico.

Ya que en electrostática los efectos de los campos magnéticos no son considerados, el campo eléctrico es irrotacional ($\nabla \times \mathbf{E} = 0$), esto indica que existe una función potencial ϕ tal que

$$-\nabla\phi = \mathbf{E}. \quad (1.3)$$

Al utilizar las ecuaciones (1.1) y (1.3), se puede obtener la ecuación electrostática de Poisson

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0} \quad (1.4)$$

donde ϕ es el potencial electrostático, que en caso de ser considerada como una distribución de cargas puntual $\rho(\mathbf{x}) = q\delta(\mathbf{r}')$ y con δ como la función Delta de Dirac, asume como la Ley de Coulomb

$$\phi(\mathbf{r}) = \frac{q}{4\pi\epsilon_0|\mathbf{r} - \mathbf{r}'|} \quad (1.5)$$

1.1.2. Electrostática en un medio continuo dieléctrico

Lo que se expuso en la sección (1.1.1) solo puede tener sentido y utilidad cuando entre las dos cargas hay vacío. Para utilizar esta teoría en arreglos macroscópicos se necesita contar cada molécula con tal de obtener una correcta representación del sistema, lo que conlleva a un

gasto elevado de tiempo y recursos computacionales. Es por ello que existe motivación en el desarrollo de teorías de la electrostática macroscópica donde el campo eléctrico actúa sobre el material dieléctrico. Si se aplica esto en la ecuación (1.4), se obtiene la versión macroscópica de la Ecuación de Poisson para electrostática

$$\nabla^2 \phi = -\nabla \cdot \mathbf{E} = -\frac{\rho}{\epsilon} \quad (1.6)$$

donde ϵ es la constante de permisividad de la región.

Cálculo de energía en un medio dieléctrico

Para el cálculo de energía se debe considerar la utilizada para la polarización del medio. Si se integra el trabajo echo por un cambio infinitesimal en la distribución de la carga, junto con relaciones del desplazamiento eléctrico ($\mathbf{D}$), se puede obtener que la energía promedio está dada por

$$W = \frac{1}{2} \int_{\Omega} \phi(\mathbf{r}) \rho(\mathbf{r}) d\Omega \quad (1.7)$$

solo para un material dieléctrico.

1.1.3. Ecuación de Poisson-Boltzmann

Generalmente los solventes consisten en agua con sal, cuando la sal es disuelta produce iones (Cloro Cl^- y Sodio Na^+) que son capaces de moverse libremente por el fluido, haciendo que la distribución de densidad de carga de la ecuación (1.6) dependa del potencial electrostático ($\rho = \rho(\phi)$). Con la finalidad de simplificar el problema, se asume que los iones son esferas solidas, el solvente no tiene comportamiento molecular y que es un continuo incluso a pequeñas distancias. Para obtener la ecuación de Poisson-Boltzmann se debe considerar que las fuerzas que afectan a los iones son únicamente de naturaleza electrostática y de larga distancia, aunque en general los iones cuentan con interacciones a corta distancia como lo son los efectos de van der Waals y solvatación [8].

Aplicando funciones de distribución radial en la ecuación (1.6), para N tipos de ion, se obtiene que la densidad es

$$\rho(\phi) = \sum_i^{N_s} q_i c_i(\mathbf{r}) = \sum_i^{N_s} q_i c_{0i} e^{-\frac{\phi q_i}{k_B T}} \quad (1.8)$$

donde q_i y c_{0i} son la carga y la densidad de iones de la especie i . Para la solución de agua con sal, solo hay dos tipos de ion, sodio y cloro con cargas e^+ y e^- respectivamente ($e^+ = -e^-$). Iones que al provenir de la misma fuente y compartir igual cantidad de átomos en la estructura, se es correcto afirmar que las concentraciones c_{0Na} y c_{0Cl} sean iguales (c_0). Para este caso, la ecuación (1.6) se vuelve

$$\begin{aligned} \nabla^2 \phi &= -\frac{n_0}{\epsilon} (e^+ e^{-\frac{\phi e^+}{k_B T}} + e^- e^{-\frac{\phi e^-}{k_B T}}) = -\frac{e^+ n_0}{\epsilon} (e^{-\frac{\phi e^+}{k_B T}} - e^{-\frac{\phi e^-}{k_B T}}) \\ \nabla^2 \phi &= \frac{2n_0 e^+}{\epsilon} \sinh\left(\frac{\phi e^+}{k_B T}\right) \end{aligned} \quad (1.9)$$

para las situaciones donde $\phi q \ll k_B T$ la ecuación (1.9) puede ser linealizada y escrita como

$$\nabla^2 \phi = \frac{\phi}{\epsilon k T} \sum_i^{N_s} n_{0i} q_i = \kappa^2 \phi \quad (1.10)$$

donde κ es la inversa del largo de Debye que representa el efecto de apantallamiento, que para el caso que es de relevancia se vuelve

$$\kappa^2 = \frac{2n_0 e^+}{\epsilon k T} \quad (1.11)$$

1.2. Modelo de Solvente Implícito en Sistemas Moleculares

El modelo de solvente implícito usa teoría electrostática continua en medio dieléctrico para representar el solvente alrededor de la proteína. Las bio-moléculas generalmente se encuentran disueltas en medios acuosos, un grupo importante de bio-moléculas son las proteínas, cadenas de aminoácidos que se encuentran compactadas de maneras muy específicas. Dentro de las proteínas encontramos una distribución de carga producto de la presencia de

Hidrógenos y de los enlaces químicos. Debido a su estructura específica y su interés de estudio es que muchas proteínas son almacenadas para libre disposición en la Protein Data Bank (PDB) [12], además de existir modelos que permiten obtener su radio de van der Waals, la carga de cada átomo y su posición.

En el dieléctrico se pueden identificar 2 regiones, el solvente y la proteína. La región del solvente está compuesta generalmente por agua con sal, la cuál posee la constante dieléctrica del agua ($\epsilon \approx 80$) y los iones de sal son considerados en la ecuación de Poisson-Boltzmann linealizada (1.10). Para la proteína la situación cambia y se considera una constante mucho menor ($\epsilon \approx 2 - 4$) por su distribución de carga modelada de forma puntual en las ubicaciones de los átomos.

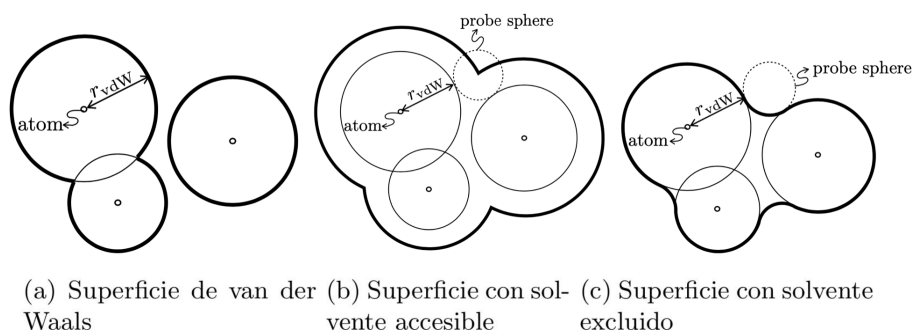


Figura 1.1: Definición de Superficies. Fuente: [8]

Existen diferentes formas de representar la interfaz entre el solvente y la proteínas, de los más recurrentes se encuentran los propuestos en la figura 1.1, donde el común de todos estos modelos es la dependencia del radio de van der Waals. La representación a utilizar en este trabajo es la mostrada en la figura 1.1.c, aquí la proteína se modela como un conjunto de esferas con una superficie que excluye al solvente ("Solvent-Excluded Surface", SES). El SES a diferencia del SAS ("Solvent-Accessible surface", figura 1.1.b), que es generado por el centro de una esfera del tamaño del agua ($\sim 1.4\text{\AA}$) que gira al rededor de la proteína; se genera utilizando la misma esfera, pero definiéndola de forma en que esta sea excluida de la proteína (lo mas cerca que puede estar la molécula de agua de la proteína). En este trabajo se

hace uso de MSMS [21], un software libre que permite obtener la superficie de la proteína de la forma SES.

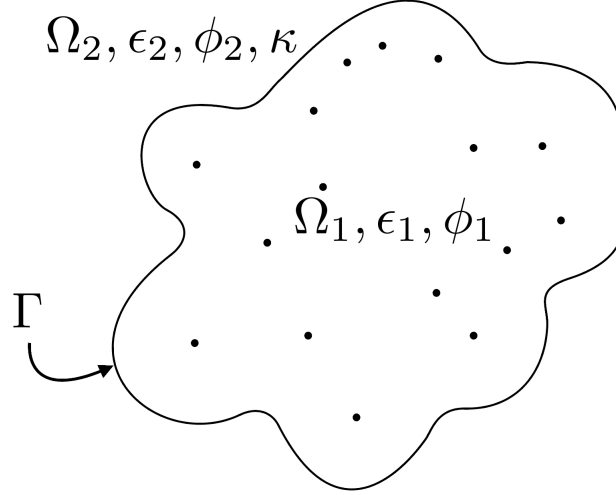


Figura 1.2: Definición de Superficies. Fuente: [8]

Cada región del sistema, proteína (Ω_1) y agua (Ω_2), cuenta con su propia permisividad y potencial electrostático (ϵ_i, ϕ_i), además la presencia de iones en Ω_2 se asocia con un largo de Debye κ^{-1} (sólo presente en la región Ω_2). Por las condiciones que se establecen, la interfase Γ debe cumplir con las condiciones de:

- El potencial electrostático sea continuo.
- No existan cargas explícitas en el SES.

Esto puede ser escrito como:

$$\begin{aligned}
 \nabla^2 \phi_1(\mathbf{r}) &= - \sum_k \frac{q_k}{\epsilon_1} \delta(\mathbf{r}, \mathbf{r}_k) \quad \text{en proteína } (\Omega_1) \\
 \nabla^2 \phi_2(\mathbf{r}) &= \kappa^2 \phi_2(\mathbf{r}) \quad \text{en solvente } (\Omega_2) \\
 \phi_1 &= \phi_2 \quad \text{en interfase } \Gamma \\
 \epsilon_1 \frac{\partial \phi_1}{\partial \mathbf{n}} &= \epsilon_2 \frac{\partial \phi_2}{\partial \mathbf{n}}
 \end{aligned} \tag{1.12}$$

1.2.1. Cálculo de la Energía

Dentro de la solvatación de la proteína se pueden identificar 2 tipos de contribuciones a la energía libre, polar y no polar. El tipo de contribución que ha sido objeto de diversos estudios y de mayor interés es la polar, cuya fuente es la energía electrostática. La energía electrostática tiene orígenes distintos:

- Energía de Coulomb, producto del desplazamiento de los puntos de carga.
- Energía de solvatación, proveniente del solvente fuera de la región de la proteína.

Energía de Coulomb

La energía de Coulomb es la que se libera o absorbe gracias al movimiento de una carga dentro de un campo eléctrico. Este campo eléctrico genera una fuerza de atracción o repulsión dependiendo si la carga es positiva o negativa. Esta fuerza es conocida como Ley de Coulomb, que para el caso de una distribución de cargas se escribe como

$$F_{Coulomb} = \frac{1}{2} \sum_i \sum_j \frac{q_i \cdot q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1.13)$$

Energía de Solvatación

La energía libre de solvatación se obtiene al momento de la disolución de la proteína y el solvente se polariza. Existen situaciones donde se disuelve sólo una proteínas, múltiples biomoléculas y además hay superficies cargadas. El caso trabajado fue la disolución de una proteína, donde la energía de solvatación es la diferencia entre la energía de la proteína en el vacío (Coulomb) y la energía liberada en la disolución. Para un arreglo general, la formulación es

$$F_{solv} = \frac{1}{2} \int_{\Omega} \rho \phi_{reac} = \sum_{k=0}^{N_q} q_k \phi_{reac}(\mathbf{r}_k) \quad (1.14)$$

Esta ecuación indica que la energía es generada por el potencial de reacción ($\phi_{reac} = \phi_{total} - \phi_{Coulomb}$) dentro de la proteína, y en caso de que la distribución de carga (ρ) sea puntual, la integral toma la forma de una sumatoria.

Energía Libre

Finalmente la energía libre está dada por la suma de todas las energías de contribución polar,

$$F_{total} = F_{Coulomb} + F_{solv} \quad (1.15)$$

1.3. Método de Elementos de Borde

Los métodos de elementos de borde (Boundary Element Method, BEM) es un método numérico para la resolución de Ecuaciones en Derivadas Parciales (EDP) lineales que son formuladas como ecuaciones integrales. Para realizar esta formulación se utilizan las condiciones de borde del problema para resolver una ecuación integral solo en la frontera, y luego de tener un resultado, aplicar esta ecuación a los puntos interiores del dominio de forma numérica. Este método es aplicable a problemas que se les conozca una solución fundamental (función de Green) y de preferencia que se trabaje sobre un campo linear.

La utilización de este método conlleva ventajas y desventajas sobre los métodos de elementos y diferencias finitas. Las principales son que representan de mejor forma la superficie molecular, ya que trabajan principalmente en esta, además son mas eficientes para las situaciones que el cociente superficie/volumen es bajo [14]. Por otro lado, éste método genera matrices densamente pobladas, por lo que buscar un adecuado método de aceleración es fundamental. En este capítulo se realizará una introducción básica de este método para su mejor comprensión.

1.3.1. Ecuación de Laplace

Uno de los problemas de importancia a resolver mediante el uso de BEM es la ecuación de Laplace ($\nabla^2 \phi = -\rho/\epsilon$), para ello se toma un ejemplo básico con tal de derivar una solución,

$$\begin{aligned}\nabla^2 \phi &= 0 \quad \text{dentro de } \Omega \\ \phi &= \phi_0 \quad \text{en } \Gamma, \text{ o} \\ \frac{\partial \phi}{\partial \mathbf{n}} &= \frac{\partial \phi_0}{\partial \mathbf{n}} \quad \text{en } \Gamma\end{aligned}\tag{1.16}$$

Para Formular la ecuación integral, se integra la ecuación de Laplace en la región Ω junto a una función peso a definir.

$$\int_{\Omega} \nabla^2 \phi(\mathbf{r}'_{\Omega}) w(\mathbf{r}'_{\Omega}) d\Omega' = 0\tag{1.17}$$

Se sabe que para utilizar este método el problema debe contar con una función de Green, esta función para el problema toma el valor de la función peso (w) y está definida como

$$G_L(\mathbf{r} - \mathbf{r}') = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|}\tag{1.18}$$

por lo que Ec (1.17) se reescribe

$$\int_{\Omega} \nabla^2 \phi(\mathbf{r}'_{\Omega}) G_L(\mathbf{r}'_{\Omega}) d\Omega' = 0\tag{1.19}$$

Para la resolución del problema es recomendable recordar ciertos teoremas e identidades matemáticas para facilitar la comprensión de la derivación,

$$\begin{aligned}
\iiint_{\Omega} \nabla \cdot \mathbf{F} d\Omega &= \iint_{\partial\Omega} \mathbf{F} \cdot \mathbf{n} d\Gamma \\
\nabla \cdot (\mathbf{A} \cdot \mathbf{B}) &= \nabla \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \nabla \mathbf{B} \\
\int_R f(x) \delta(x - x_0) dx &= f(x_0) \quad \text{si } x_0 \text{ está en } R
\end{aligned} \tag{1.20}$$

la última ecuación es anotada, ya que la función de Green cumple

$$\nabla^2 G_L(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r}) \tag{1.21}$$

mediante la manipulación de la ecuación (1.19) y el uso las identidades (1.20)

$$\begin{aligned}
\int_{\Omega} [\nabla \cdot (\nabla \phi(\mathbf{r}'_{\Omega}) G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Omega})) - \nabla \phi(\mathbf{r}'_{\Omega}) \nabla G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Omega})] d\Omega' &= 0 \\
\int_{\Omega} [\nabla \cdot (\nabla \phi(\mathbf{r}'_{\Omega}) G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Omega})) - \nabla \cdot (\phi(\mathbf{r}'_{\Omega}) \nabla G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Omega})) + \phi(\mathbf{r}'_{\Omega}) \nabla^2 G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Omega})] d\Omega' &= 0
\end{aligned} \tag{1.22}$$

$$\tag{1.23}$$

Si el teorema de la divergencia se le aplica a (1.23), se puede ver que los 2 primeros términos pasan de ser integrales de volumen a integrales de superficie

$$\int_{\Gamma} \nabla \phi(\mathbf{r}'_{\Gamma}) \cdot \mathbf{n} G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Gamma}) d\Gamma' - \int_{\Gamma} \phi(\mathbf{r}'_{\Gamma}) \nabla G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Gamma}) \cdot \mathbf{n} d\Gamma' - \int_{\Omega} \phi(\mathbf{r}'_{\Omega}) \delta(\mathbf{r}_{\Omega}) d\Omega' = 0 \tag{1.24}$$

donde $\mathbf{n}$ es el vector normal a la superficie Γ apuntando hacia fuera del dominio Ω , y $\mathbf{r}_{\Gamma}$ es la ubicación en la frontera de Γ . Por otro lado si aplica una de las propiedades de la función delta de Dirac (1.20), la ecuación (1.24) se puede reescribir como

$$\phi(\mathbf{r}_{\Omega}) = \int_{\Gamma} G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Gamma}) \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}'_{\Gamma}) d\Gamma' - \int_{\Gamma} \phi(\mathbf{r}'_{\Gamma}) \frac{\partial}{\partial \mathbf{n}} G_L(\mathbf{r}_{\Omega}, \mathbf{r}'_{\Gamma}) d\Gamma' \tag{1.25}$$

En esta derivación $\mathbf{r}$ corre sobre el dominio de la región Ω mientras que $\mathbf{r}'$ solo recorre la superficie de la región (Γ). Pero que sucede que cuando $\mathbf{r}$ se ubica en la superficie Γ ($\mathbf{r} = \mathbf{r}'$), por como esta definida la función de Green (1.18), en los puntos de nuestra superficie se generan singularidades. Para lidiar con este inconveniente, se integra en Γ' en lugar de Γ . Γ' tiene la particularidad de ser igual a Γ en todos los puntos con la diferencia que en $\mathbf{r}$ posee una semi-esfera de radio ε . Por lo que si se separa la superficie de integración, se obtiene que

$$\Gamma = \Gamma' + \Gamma_\varepsilon, \quad \lim_{\varepsilon \rightarrow 0} \Rightarrow \Gamma = \Gamma' \quad (1.26)$$

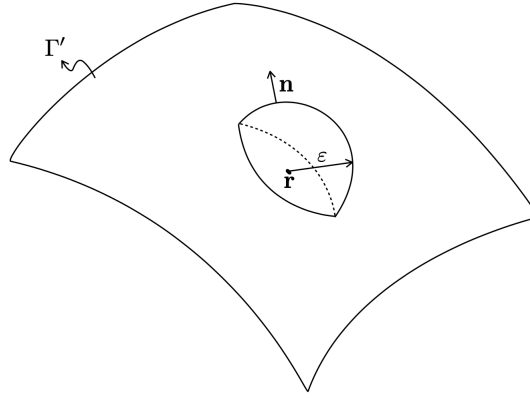


Figura 1.3: Singularidad en superficie. Fuente: [8]

Con esto definido, es recomendable separar las integrales y resolver. Se sabe que $\varepsilon = |\mathbf{r} - \mathbf{r}'|$, parte de la ecuación (1.25) se puede anotar como

$$\begin{aligned} \int_{\Gamma_\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}') G_L(\mathbf{r}, \mathbf{r}') d\Gamma' &= \frac{1}{4\pi} \int_{\Gamma_\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} d\Gamma' \\ &= \frac{1}{4\pi\varepsilon} \int_{\Gamma_\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}') d\Gamma' \end{aligned} \quad (1.27)$$

ahora, si se intenta tomar toda la superficie, es decir $\varepsilon \rightarrow 0$

$$\begin{aligned} \frac{1}{4\pi\varepsilon} \int_{\Gamma_\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}') d\Gamma' &= \lim_{\varepsilon \rightarrow 0} \frac{1}{4\pi\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}') \int_{\Gamma_\varepsilon} d\Gamma' \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{4\pi\varepsilon} \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}) 2\pi\varepsilon^2 = 0 \end{aligned} \quad (1.28)$$

Si esto procedimiento es realizado en el mismo espacio para la integral con el término $\partial G_L / \partial \mathbf{n}$, resulta

$$\begin{aligned} \int_{\Gamma_\varepsilon} \phi(\mathbf{r}') \frac{\partial}{\partial \mathbf{n}} G_L(\mathbf{r}, \mathbf{r}') d\Gamma' &= -\frac{1}{4\pi} \int_{\Gamma_\varepsilon} \phi(\mathbf{r}') \frac{(\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}}{|\mathbf{r} - \mathbf{r}'|^3} d\Gamma' \\ &= -\frac{1}{4\pi} \int_{\Gamma_\varepsilon} \phi(\mathbf{r}') \frac{\varepsilon}{\varepsilon^3} d\Gamma' \end{aligned} \quad (1.29)$$

si $\varepsilon \rightarrow 0$

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} -\frac{1}{4\pi} \int_{\Gamma_\varepsilon} \phi(\mathbf{r}') \frac{\varepsilon}{\varepsilon^3} d\Gamma' &= -\frac{\phi(\mathbf{r})}{4\pi} \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} \int_{\Gamma_\varepsilon} d\Gamma' \\ &= -\frac{\phi(\mathbf{r})}{4\pi} \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^2} 2\pi\varepsilon^2 = -\frac{\phi(\mathbf{r})}{2} \end{aligned} \quad (1.30)$$

Al utilizar los resultados (1.29) y (1.30), la ecuación (1.25) puede ser reformulada con las singularidades resueltas

$$\frac{\phi(\mathbf{r}_\Gamma)}{2} + \int_{\Gamma} \phi(\mathbf{r}'_\Gamma) \frac{\partial}{\partial \mathbf{n}} G_L(\mathbf{r}_\Gamma, \mathbf{r}'_\Gamma) d\Gamma' = \int_{\Gamma} G_L(\mathbf{r}_\Gamma, \mathbf{r}'_\Gamma) \frac{\partial}{\partial \mathbf{n}} \phi(\mathbf{r}'_\Gamma) d\Gamma' \quad (1.31)$$

La Ec (1.31) es conocida como la formulación integral de Poisson, a pesar de que esta derivación fue hecha para (1.18), es totalmente válida para la ecuación de Poisson-Boltzmann lineal (1.10). Para ello la función de Green varía a:

$$G_Y(\mathbf{r}, \mathbf{r}') = \frac{\exp(-\kappa|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|} \quad (1.32)$$

1.3.2. Sistema Lineal

BEM resuelve las ecuaciones de Poisson y PB de forma integral al utilizar las condiciones de borde en la superficie, es por esto que los puntos interiores de la región deben calcularse mediante métodos numéricos. Antes de poder aplicar un Sistema Lineal que permita solucionar estas ecuaciones, es necesario definir los operadores que componen la matriz [8]. Los Operadores que se utilizan se conocen como:

- Single-Layer Potential Operator
- Double-Layer Potential Operator

Single-Layer Potential Operator

$$V_L^{\mathbf{r}_\Gamma}(\psi(\mathbf{r}_\Gamma)) = \int_{\Gamma} \psi(\mathbf{r}'_\Gamma) G_L(\mathbf{r}_\Gamma, \mathbf{r}'_\Gamma) d\Gamma \quad (1.33)$$

Double-Layer Potential Operator

$$K_L^{\mathbf{r}_\Gamma}(\psi(\mathbf{r}_\Gamma)) = \int_{\Gamma} \psi(\mathbf{r}'_\Gamma) \frac{\partial}{\partial \mathbf{n}} G_L(\mathbf{r}_\Gamma, \mathbf{r}'_\Gamma) d\Gamma \quad (1.34)$$

Con los operadores definidos en las ecuaciones 1.33 y 1.34, es posible formular el sistema lineal ($Ax = f$).

$$\left[\frac{\mathbb{I}}{2} + K_L^{\mathbf{r}_\Gamma} \right] (\phi_\Gamma) = V_L^{\mathbf{r}_\Gamma} \left(\frac{\partial}{\partial \mathbf{n}} \phi_\Gamma \right) \quad (1.35)$$

Discretización del Sistema

Al discretizar la superficie se utilizan paneles triangulares, en cada uno de estos paneles se asume que ϕ y $\partial\phi/\partial\mathbf{n}$ son constantes, al discretizar los operadores se obtiene:

$$V_{L,disc}^{\mathbf{r}_i} \left(\frac{\partial}{\partial\mathbf{n}} \phi(\mathbf{r}_\Gamma) \right) = \sum_{j=1}^{N_p} \frac{\partial}{\partial\mathbf{n}} \phi(\mathbf{r}_\Gamma) \int_{\Gamma_j} G_L(\mathbf{r}_i, \mathbf{r}_{\Gamma_j}) d\Gamma_j \quad (1.36)$$

$$K_{L,disc}^{\mathbf{r}_i}(\phi(\mathbf{r}_\Gamma)) = \sum_{j=1}^{N_p} \phi(\mathbf{r}_\Gamma) \int_{\Gamma_j} \frac{\partial}{\partial\mathbf{n}} [G_L(\mathbf{r}_i, \mathbf{r}_{\Gamma_j})] d\Gamma_j \quad (1.37)$$

1.4. Estrategia de aceleración - Treecode

Uno de los grandes problemas de BEM es el tamaño y lo densamente poblada que resulta la matriz del sistema después de haber aplicado la discretización. La dimensión de estas hace que su solución mediante método de eliminación Gaussiana requiera un gran almacenamiento $\mathcal{O}(N^2)$ y tiempo de computo para biomoléculas de medio tamaño. El método iterativo que se utiliza es GMRES (Generalized Minimal Residual Method), donde el escalamiento cae a $\mathcal{O}(N^2)$ por iteración. Este problema se soluciona con la aplicación de aceleradores que hacen que las computaciones escalen a $\mathcal{O}(N \log N)$ y que sus requerimientos de memoria sean $\mathcal{O}(N)$, adaptando así el método a moléculas de mayor tamaño.

Los métodos de multipolo son los aceleradores más populares que se aplican en BEM, estos aproximan interacciones de objetivos lejanos a la fuente con una expansión de multipolo usando expansiones de Taylor. Estos algoritmos son útiles para aproximar integrales con reglas de cuadratura Gaussiana, los cuales convierten un sistema Matriz-Vector en cuerpos de N cálculos. El tipo de algoritmo utilizado en PyGBe es Treecode, el cual es explicado en mas detalle a continuación [3].

1.4.1. Treecode

Treecode es un algoritmo de suma rápida que capaz de trabajar en $\mathcal{O}(N \log N)$ el siguiente patrón $\mathcal{O}(N^2)$:

$$V(x_i) = \sum_{j=i}^N q_j \psi(x_i, y_j) \quad (1.38)$$

donde q_j es el peso, ψ el núcleo, y_j la ubicación de las fuentes y x_i la ubicación de los objetivos.

En este algoritmo se agrupan las fuentes (y) acorde su posición espacial en las celdas de un árbol (es decir, celdas contenidas en celdas mas grandes). Con las celdas formadas se calculan las contribuciones cercanas directamente, y las contribuciones mas alejadas se aproximan con series de Taylor utilizando las fuentes alrededor de la celda con centro en y_c .

$$\psi(x_i, y_j) = \sum_{\|k\|}^P \frac{1}{k!} D_y^k(x_i, y_j) (y_c - y_j)^k \quad (1.39)$$

con P el orden de la expansión, $D_y^k = D_{y_1}^{k_1} D_{y_2}^{k_2} D_{y_3}^{k_3}$ operador diferencial y $k = (k_1, k_2, k_3)$. Se define el operador k como $k! = k_1! k_2! k_3!$ e $y^k = y_1^{k_1} y_2^{k_2} y_3^{k_3}$. Al arreglar los términos de (1.39)

$$\begin{aligned} V(x_i) &= \sum_{\|k\|}^P \frac{1}{k!} D_y^k \psi(x_i, y_c) \sum_j q_j (y_c - y_j)^k \\ a^k(x_i, y_c) &= \frac{1}{k!} D_y^k \psi(x_i, y_c) \\ m_c^k &= \sum_j q_j (y_c - y_j)^k \end{aligned} \quad (1.40)$$

donde a^k y m_c^k son los coeficientes de expansión y los multipolos respectivamente.

Para cada objetivo, se decide si una celda (con fuentes) está suficientemente lejos como para realizar una aproximación de Taylor aceptable (1.40), en caso de que una aproximación no sea recomendable se realiza la misma pregunta a las celdas hijas (celdas que están dentro de celdas de mayor tamaño). Este proceso se aplica de manera hasta llegar al menor nivel de

las celdas (celdas más pequeñas). La toma de esta decisión se basa en el criterio de aceptación de multipolo θ (multipole-acceptance criterion, MAC), el cual debe ser mayor que el cociente entre el tamaño de la celda (r_b) y la distancia objetivo - centro celda (r)

$$\frac{r_b}{r} < \theta \quad (1.41)$$

Los valores más comunes de θ se encuentran entre $\frac{1}{2}$ y $\frac{2}{3}$.

El algoritmo de Treecode se puede dividir en:

1. Generación del árbol
2. Partícula a multipolo (P2M)
3. Multipolo a multipolo (M2M)
4. Multipolo a partícula (M2P)
5. Partícula a partícula (P2P)

Generación del árbol. Las fuentes son agrupadas en celdas dentro de un árbol. Estas celdas a su vez contienen celdas de menor tamaño, esto continúa así hasta llegar a las celdas de menor nivel, asegurando que las celdas mas pequeñas tengan un menor número de partículas que las establecidas por el número crítico (NCRIT).

P2M. Los multipolos son calculados por los niveles mas bajos de las celdas usando la tercera ecuación de (1.40). Las fuentes son expandidas por el correspondiente centro de la celda.

M2M. Multipolos son calculados por las celdas restantes usando caja padre de referencia. Para mejor eficiencia, la expansión de Taylor se modifica a

$$m_p^k = \sum_{\text{hijos}} \sum_{k'_1 \leq k_1, k'_2 \leq k_2, k'_3 \leq k_3} m_c^k \binom{k_1}{k'_1} \binom{k_2}{k'_2} \binom{k_3}{k'_3} (y_{cp} - y_{cc})^{k-k'} \quad (1.42)$$

donde m_p^k es el orden del multipolo k de la celda padre (celda que posee celdas más pequeñas), y_{cp} es la ubicación de la celda padre e y_{cc} es la ubicación de la celda hija.

M2P. Cálculo de la expansión de Taylor para las interacciones de la partícula objetivo y las celdas que cumplan con el criterio θ . Si la partícula está lo suficientemente lejos para no interactuar con ninguna celda interior que cumpla con el criterio θ , ésta puede interactuar con la celda más grande.

P2P. Como su nombre indica, se calcula la interacción directa entre las partículas siempre y cuando el objetivo esté tan cerca de una celda como para no cumplir con θ .

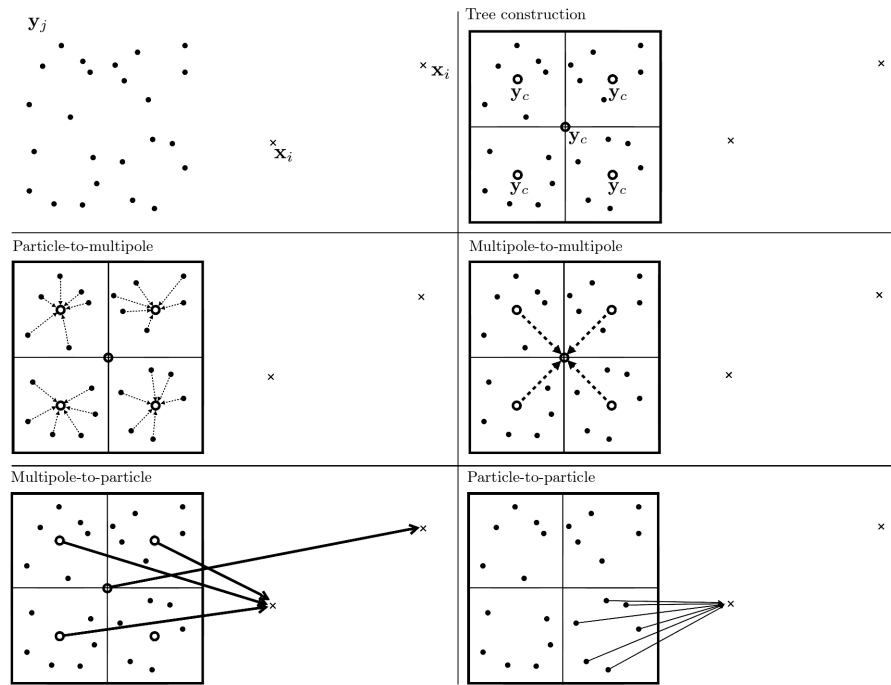


Figura 1.4: Algoritmo Treecode. Fuente [8]

Cada paso es mostrado de manera simplificada en la figura 1.4. Aquí se genera el árbol con 2 niveles, el primero es la caja padre que contiene 4 celdas, el segundo nivel son las cuatro celdas más pequeñas que contienen las fuentes. En cada celda se calculan las interacciones de las fuentes contenidas en ellas, donde se puede decir que su efecto se concentra en el centro

de cada una de estas. Con los multipolos listos, se busca la interacción que hay usando la celda padre (nivel 1) de referencia. Luego se ve el efecto de las celdas con el objetivo que cumpla el criterio θ . Para el caso particular de la figura 1.4, ya que hay una partícula que se encuentra lo suficientemente lejos como para no verse afectado por las celdas mas pequeñas, se utiliza el multipolo de la caja más grande. Por otro lado, la partícula más cercana está lo suficientemente cerca para verse afectado por celdas de forma individual (cumple con θ), pero demasiado cerca de una celda en específico, por lo que su interacción con esta es de manera individual con cada una de sus fuentes.

Finalmente, la forma de aplicar treecode a un problema de BEM, es mediante la resolución por cuadratura Gaussiana de las integrales presentes en la matriz del sistema. La cuadratura Gaussiana transforma los operadores expresados como integral a una suma. Por ejemplo

$$V_{L,ij}^\Gamma = \int_{\Gamma_j} G_L(\mathbf{r}_{\Gamma i}, \mathbf{r}_{\Gamma j}) d\Gamma_j = \sum_{k=1}^{N_k} w_k G_L(\mathbf{r}_{\Gamma i}, \mathbf{r}_{k\Gamma j}) \quad (1.43)$$

el cual se multiplica por un vector $\mathbf{m}$ para luego ser calculado de la forma de (1.39)

$$V_{L,ij}^\Gamma m_j = \sum_{j=1}^{N_p} m_j \sum_{k=1}^{N_k} w_k G_L(\mathbf{r}_{\Gamma i}, \mathbf{r}_{k\Gamma j}) = \sum_{l=1}^{N_p \cdot N_k} M_l G_L(\mathbf{r}_{\Gamma i}, \mathbf{r}_l) \quad (1.44)$$

en el cual los puntos de Gauss N_p ubicados en $\mathbf{r}_l$ asumen como las fuentes y_j , objetivos x_i son los puntos de colocación y el núcleo ψ es la función libre de Green.

1.5. Estado del Arte

La química (bioquímica) computacional es una ciencia que estudia las propiedades e interacción de las moléculas (energía, momentos, frecuencias, etc.) que se encuentra en constante crecimiento, para lograr su implementación se han desarrollado diversos softwares utilizando distintos métodos explícitos e implícitos. Dentro de los modelos explícitos el método más utilizado es el de Dinámica Molecular, donde existen software ampliamente utilizados como lo pueden ser CHARMM, GROMACS, NAMD, entre otros. Por el otro lado, los soft-

wares de modelos implícitos desarrollados se centran principalmente en resolver la ecuación de Poisson-Boltzmann formulada al momento de considerar el conjunto soluto-solvente como un material dieléctrico. Entre los softwares y paquetes más aceptados se encuentran APBS, BEMPP o PyGBe, este último el utilizado en este trabajo.

A pesar de la cantidad de programas y paquetes que se han creado, existen pocos estudios acerca de su funcionamiento sistemático y que tanto difieren los resultados al comparar distintos modelos. David Mobley es uno de los exponentes que más han desarrollado esta clase de estudios, y en su trabajo principal [16] se han recopilado más de 500 moléculas orgánicas pequeñas con sus resultados experimentales, creando una base de datos [18] utilizando lo obtenido mediante un modelo de solvente explícito TIP3P y los datos experimentales. Con esta base, él busca la eficacia de su modelo, además de un comportamiento sistemático de las moléculas al ser separadas por grupos funcionales. Al ser el trabajo más destacado en su tipo, es que algunos investigadores han realizado estudios comparando sus resultados (generalmente obtenidos con modelos implícitos) y los de Mobley. Entre los trabajos que más resaltan, se puede encontrar el modelo AGBNP2 de Gallicchio [10] y el estudio de la eficiencia de la energía de solvatación mediante "R6 Generalized Born Model" de Boris Aguilar [1].

Capítulo 2

Metodología

En la presente sección se aplican los conceptos mencionados y desarrollados en el Marco teórico. El procedimiento de comparación se realizará utilizando moléculas pequeñas con valores experimentales conocidos y calculados utilizando modelos explícitos, las moléculas han sido recopiladas y calculadas por David Mobley [16]

2.1. Modelo de Superficie Simple

Se ha mencionado en la sección 1.3.1 que la derivación de la ecuación de Laplace también era válida para el problema de Poisson-Boltzmann, siempre que sean realizadas las adaptaciones necesarias [6]. Junto a esto, en la sección 1.3.2 se han definidos los operadores necesarios para que los cálculos operen en un espacio común y los resultados posean coherencia. Al aplicar esto, se puede definir un modelo de Superficie Simple

$$\phi_{1,\Omega_1} + K_L^{\Omega_1}(\Omega_{1,\Gamma}) - V_L^{\Omega_1} \left(\frac{\partial}{\partial \mathbf{n}} \phi_{1,\Gamma} \right) = \frac{1}{\epsilon_1} \sum_{k=0}^{N_q} \frac{q_k}{4\pi |\mathbf{r}_{\Omega_1} - \mathbf{r}_k|} \quad (2.1)$$

$$\phi_{2,\Omega_2} - K_Y^{\Omega_2}(\phi_{2,\Gamma}) + V_Y^{\Omega_2} \left(\frac{\partial}{\partial \mathbf{n}} \phi_{2,\Gamma} \right) = 0 \quad (2.2)$$

donde la ecuación (2.1) se define en la proteína (Ω_1) y (2.2) lo hace en el solvente (Ω_2), ϕ_{i,Ω_i} es el potencial evaluado en cualquier punto de la región en cuestión ($\phi_{1,\Omega} = \phi_1(\mathbf{r}_{\Omega_1})$) y $\phi_{i,\Gamma}$ en la interfaz soluto-solvente ($\phi_{1,\Gamma} = \phi_1(\mathbf{r}_\Gamma)$). Ahora si solo se recorre la superficie de la

proteína y se imponen las condiciones de borde (Ec 1.12), se logra

$$\begin{aligned} \frac{\phi_{1,\Gamma}}{2} + K_L^\Gamma(\phi_{1,\Gamma}) - V_L^\Gamma\left(\frac{\partial}{\partial \mathbf{n}}\phi_{1,\Gamma}\right) &= \frac{1}{\epsilon_1} \sum_{k=0}^{N_q} \frac{q_k}{4\pi|\mathbf{r}_\Gamma - \mathbf{r}_k|} \\ \frac{\phi_{1,\Gamma}}{2} - K_Y^\Gamma(\phi_{1,\Gamma}) + \frac{\epsilon_1}{\epsilon_2} V_Y^\Gamma\left(\frac{\partial}{\partial \mathbf{n}}\phi_{1,\Gamma}\right) &= 0 \end{aligned} \quad (2.3)$$

el cual puede ser escrito como un sistema lineal

$$\begin{bmatrix} \frac{1}{2} + K_L^\Gamma & -V_L^\Gamma \\ \frac{1}{2} - K_Y^\Gamma & \frac{\epsilon_1}{\epsilon_2} V_Y^\Gamma \end{bmatrix} \cdot \begin{bmatrix} \phi_{1,\Gamma} \\ \frac{\partial}{\partial \mathbf{n}}\phi_{1,\Gamma} \end{bmatrix} = \begin{bmatrix} \sum_{k=0}^{N_q} \frac{q_k}{4\pi|\mathbf{r}_\Gamma - \mathbf{r}_k|} \\ 0 \end{bmatrix} \quad (2.4)$$

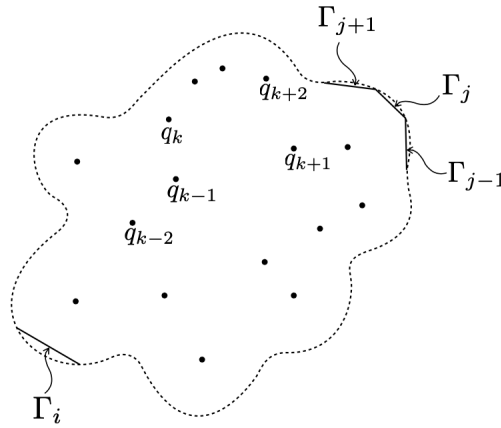


Figura 2.1: Discretización de la superficie de la proteína. Fuente:[8]

Ya que el sistema (2.4) debe ser resuelto de manera numérica, se utilizará la versión discretizada de los operadores K y V (transformando las integrales en sumas mediante cuadratura Gaussiana, sección 1.4.1). Al realizar esta acción, es necesario discretizar la superficie de la proteína en N_p paneles (utilizando el software MSMS [21]), cada panel con centro $\mathbf{r}_i$. Esto transforma cada componente de la matriz en sub-matrices de tamaño $N_p \times N_p$, donde los elementos ij son integrales sobre el panel Γ_j con $\mathbf{r}_i$ ubicado en el panel Γ_i . Cada una de estas integrales puede expresarse como una suma mediante cuadratura Gaussiana. Finalmente la

descretización de los elementos de la matriz resulta en

$$\begin{aligned}\phi_{1,\Gamma_1} &= \phi_1(\mathbf{r}_i) \\ \frac{\partial}{\partial \mathbf{n}} \phi_{1,\Gamma_1} &= \frac{\partial}{\partial \mathbf{n}} \phi_1(\mathbf{r}_i) \\ \sum_{k=0}^{N_q} \frac{q}{4\pi |\mathbf{r}_{\Gamma_1} - \mathbf{r}_k|} &= \sum_{k=0}^{N_q} \frac{q}{4\pi |\mathbf{r}_i - \mathbf{r}_k|}\end{aligned}\quad (2.5)$$

2.2. Energía en los Bordes

La energía de solvatación se definió como $\phi_{reac} = \phi_{total} - \phi_{Coulomb}$, por lo que se puede obtener el potencial de reacción a partir de la ecuación (2.1)

$$\phi_{reac,\mathbf{r}_k} = -K_L^{\mathbf{r}_k}(\phi_{1,\Gamma_1}) + V_L^{\mathbf{r}_k} \left(\frac{\partial}{\partial \mathbf{n}} \phi_{1,\Gamma_1} \right) \quad (2.6)$$

donde el término de la derecha de (2.1) corresponde al potencial de Coulomb.

2.3. Python GPU for Boundary elements - PyGBe

PyGBe [9] es un programa escrito mayormente en Python con algunas secciones en C++ y CUDA. El programa está optimizado para trabajar de mejor manera en la GPU, aunque es posible utilizar solamente una CPU a falta de compatibilidad con PYCUDA. PyGBe resuelve el sistema (2.4) para una sola interfaz. Para hacer correr el código, se necesitan varios archivos de entrada, estos son:

- “.pqr”
- “.param”
- geometry
- “.config”

“.pqr”. El archivo “.pqr” es un documento de texto que ubica en columnas la posición, carga y el radio de Van der Waals de todos los átomos presentes en la proteína. Como ya se

ha mencionado en secciones anteriores, esta se obtiene utilizando los servidores PDB2PQR ingresando un documento ".pdb" disponible en Protein Data Bank.

".param". Es un documento de texto que contiene los parámetros en los cuales el programa trabaja. Los parámetros tienen relación con el código del Treecode y como se resuelven las ecuaciones, es decir, los puntos para realizar la cuadratura de Gauss, el tamaño de las celdas y la distancia de los objetivos a la celda. Alguno de estos parámetros son:

- K: Número de puntos de Gauss por elemento (1,3,4 y 7)
- tolerance: Tolerancia de GMRES.
- P: Orden de expansión de Treecode.
- NCRIT: Máximo número de elementos por celda de la estructura del Treecode.
- θ : Criterio de aceptación de multipolo del Treecode.

geometry. Esta carpeta contiene los archivos ".vert" y ".face" generados con MSMS mediante un documento ".xyzr".

".config". Este documento contiene la configuración de la geometría y la interacción entre ellas. Las interacciones existen entre superficies Padre e Hija con constantes dieléctricas distintas. Una superficie o región padre es la que contiene otra en su interior, y una hija es la que se encuentra dentro de una región padre. La superficie padre es la región del solvente (Agua-Sal) y la superficie hija es la región del proteína. Acá se configuran algunos parámetros y las ecuaciones que se deben resolver para cada región. Alguna de las cosas a definir junto a las superficies padre e hija son:

- LorY: 1 para resolver la ecuación de Laplace (G_L , Ec 1.18) o 2 para la ecuación de Yukawa (G_Y , Ec 1.32)
- E?: 1 para calcular energía en la región, 0 para no calcular energía
- charges?: 0 si no hay presencia de cargas, 1 si hay cargas en la región (archivo ".pqr")

- coulomb?: 1 si se desea calcular la energía de coulomb en la región, 0 para omitir este cálculo
- κ, ϵ : inverso del largo de Debye, constante dieléctrica.

El nombre de cada uno de estos archivos corresponde al código de la proteína, por ejemplo si utilizamos la proteína 4rl6, los archivos serían "4rl6.pdb", "4rl6.pqr", "4rl6.param", "4rl6.vert", etc.

2.4. Generación de Superficie

Como se ha mencionado en secciones anteriores, para definir la superficie de la proteína se utilizará el programa MSMS [21], el cual genera una superficie de solvente excluido. MSMS genera 2 tipos de documentos, uno con terminación ".vert" y otro ".face", estos corresponden a la ubicación de cada vértice de la malla y al centro de cada panel respectivamente. Para lograr esto es necesario entregar la ubicación de los átomos y sus radios mediante un archivo en formato ".xyzr". Este formato de archivos no están disponibles, pero su obtención se puede realizar de manera sencilla a través de un documento ".pdb". Los archivos ".pdb" están en línea y a libre disposición en Protein Data Bank. Este algoritmo simplificado se puede ver en la Figura 2.2.

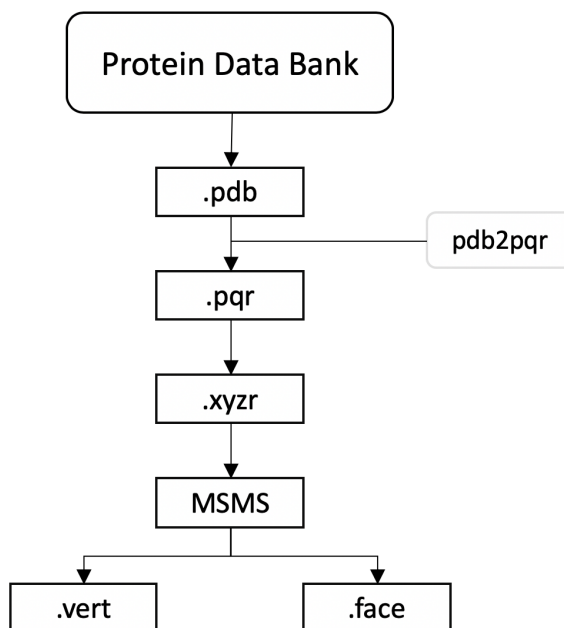


Figura 2.2: Diagrama generación de superficie

Con `pdb2pqr` un paso para separar las ubicaciones, cargas y radios del resto de los datos presentes en `".pdb"`. `MSMS` junto con el archivo `".xyzr"` requieren otros parámetros de entrada, los parámetros son el radio con el cual se genera la superficie (probe radius `-p`) y la densidad (`-d`), con valores $1,5[\text{\AA}]$ para el caso del "probe radius" y la densidad 1,0 para moléculas de más de 1.000 átomos ó 3,0 para moléculas más pequeñas.

2.5. Trabajo para Múltiples Moléculas

Como se ha mencionado, `PyGBe` requiere una extensión de documentos predeterminada y proveniente de documentos `".pqr"` para poder funcionar. Mobley [16] en su recopilación de pequeñas moléculas ha utilizado los formatos `".prmtop"` ("Parameters-Topology ") y `".crd"` ("coordinates"), por lo que para poder realizar los cálculos con `Pygbe` y la posterior comparación es necesario transformar los documentos de un formato `".prmtop"` y `".crd"` a `".pqr"`.

Dada la gran cantidad de moléculas recopiladas, se ha desarrollado mediante Python un programa que sea capaz de analizar todas las moléculas contenidas en un directorio y auto-

matizar las siguientes funciones:

1. Convertir Formatos de ".prmtop" - ".crd" a ".pqr".
2. Generar Documentos con configuración de Superficies y Parámetros.
3. Generar Geometría con MSMS en base al documento ".pqr".

Los documentos convertidos y generados se almacenan automáticamente en un nuevo directorio para ser trabajados. Con otro programa, de manera automática todos los archivos convertidos son ingresados individualmente a PyGBe, cada resultado es anotado en un documento de texto que es guardado para ser comparado con los datos obtenidos por Mobley y su método explícito.

Dado que la finalidad de este trabajo es la obtención de los parámetros del modelo implícito que entreguen los resultados más cercanos a un método explícito, es que se realiza una base de datos con los resultados obtenidos por PyGBe al cambiar la constante dieléctrica.

2.5.1. Conversión de Formatos

Los documentos ".pqr" ("Position, Charge, Radius") tienen una estructura de columnas como las exhibidas en el Cuadro 2.1

Cuadro 2.1: Estructura documento ".pqr".

Atom	N° de at	Tipo at	Algo	Tipo at (N°)	Posición (x y z)	Carga	Radio
ATOM	1	C1	Algo	1	0.611 -0.742 0.525	0.0344	1.908
ATOM	2	C2	Algo	1	1.117 -1.812 -0.412	-0.0268	1.908

Este tipo de documentos puede ser formados a través de los documentos de parámetros y coordenadas (.prmtop" y ".crd"), para ello es necesario el uso de las librerías PARMED AMBER y Numpy.

Para realizar la conversión es necesario tomar desde ".prmtop" los parámetros necesarios para obtener las primeras 5 columnas, esto se realiza con Pyhton y la librería PARMED

AMBER. Luego mediante la librería Numpy, es posible tomar las coordenadas presentes en ".crd" y copiarlas en el nuevo documento ".pqr". Finalmente es necesario volver a utilizar el documento ".prmtop" para obtener la carga y el radio correspondiente a cada átomo.

2.5.2. Generación de Parámetros y Configuración

Los nuevos documentos se generan para cada molécula presente en el directorio principal. Los documentos ".param" que contienen los parámetros de funcionamiento de PyGBe son realizados con los mismos valores y de manera automática para todas las moléculas. Los parámetros establecidos se pueden encontrar en el cuadro 2.2.

Cuadro 2.2: Parámetros PyGBe en documento ".param".

Precision	double
K	4
Nk	9
K_fine	37
thresold	0.5
BSZ	128
restart	100
tolerance	1e-6
max_iter	1000
P	8
eps	1e-12
NCRIT	500
theta	0.5
GPU	0
polar_eps	1e-2

El documento ".config" se debe generar en base a una única interfaz, por lo que solo se tiene una región padre y otra hija. La región hija siempre es la compuesta por la molécula a analizar, debido a que se desea analizar la influencia de la constante dieléctrica (ϵ) en la energía de solvatación es que este parámetro se varía cada vez que se deseen ingresar las moléculas a PyGBe, este valor oscila entre $\epsilon = 1,00$ y $\epsilon = 1.50$. La forma en que el documento de texto con extensión ".config" debe ser escrito se puede apreciar en la figura 2.3

FILE	geometry/xxx		dielectric_interface								
PARAM	LorY	E?	Dielec	kappa	charges?	coulomb?	charge_file	Nparent	parent	Nchild	children
FIELD	1	0	80	1e-12	0	0	NA	0	NA	1	0
FIELD	1	1	eps	1e-12	1	1	xxx.pqr	1	0	0	NA

Figura 2.3: Documento xxx".config".

donde "xxx" representa el nombre de la molécula a tratar.

Ya que Mobley utiliza agua sin sal como solvente para realizar las simulaciones de dinámica molecular, es que PyGBe resolverá la ecuación de Laplace (Ecuación (1.19)), o bien la ecuación de Poisson-Boltzmann con un $\kappa \approx 0$ (Ecuación (1.32)).

2.5.3. Generación de Geometría

Cada documento ".xyzr" pasa por MSMS para generar los documentos ".vert" y ".face", a partir del documento ".pqr" el procedimiento es el mismo que el descrito en la sección 2.4 y en la figura 2.2.

Los parámetros establecidos para generar la superficie son:

- probe radius: 1,5
- d: 3,0

2.5.4. Generación de Base de datos

PyGBe es capaz de procesar una molécula a la vez, es por ello que mediante el uso de Python se automatiza este proceso para todas las moléculas a las cuales se ha realizado una conversión de formatos, generación de superficies, parámetros y configuraciones. Al activar el programa, se ingresa una molécula y la energía resultante es anotada en un único documento de texto, una vez se realiza este proceso, se trabaja con la siguiente molécula presente en el directorio. El documento que recopila la información producida por PyGBe para cada

molécula y se escribe de la forma "eps.txt" (Sí $\epsilon = 1.25 \Rightarrow$ "125.txt"). El proceso de generar los archivos por molécula y la generación de base de datos se pueden observar en los diagramas de flujo de las figuras 2.4 y 2.5 respectivamente.

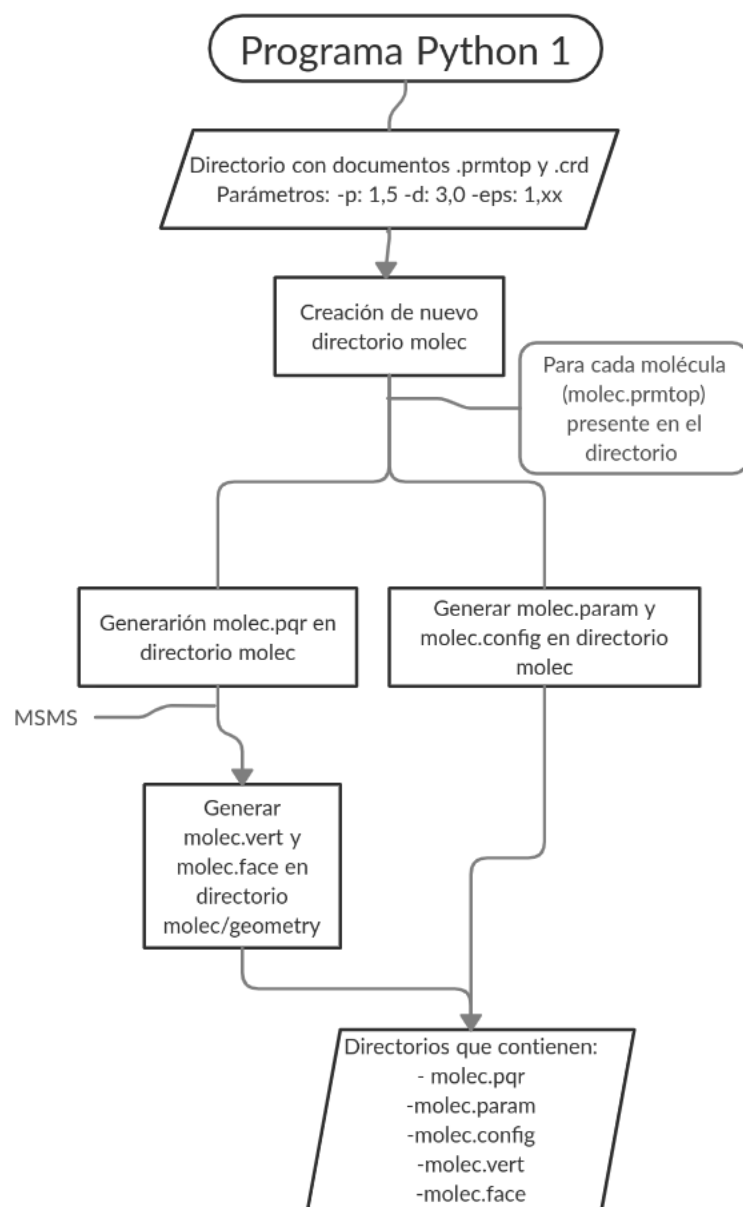


Figura 2.4: Diagrama de flujo de programa 1.

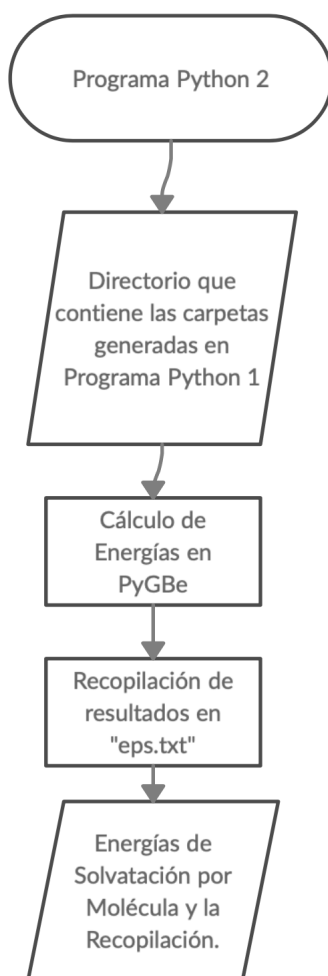


Figura 2.5: Diagrama de flujo de programa 2.

2.5.5. Comparación Método Implícito y Método Explícito

Para realizar la comparación entre ambos métodos, se realiza un programa con el lenguaje de programación Python mediante Jupyter Notebook. El programa utiliza la base de datos creada por Mobley y las creadas mediante PyGBe. En el programa, cada molécula se representa de manera puntual utilizando como referencia los resultados obtenidos por Dinámica Molecular (Eje X) y los obtenidos por el método implícito (Eje Y). Junto con la representación gráfica, se entregan:

- Coeficiente de Correlación: Donde 1,000 representa una correlación total y 0,000 una correlación nula.

- RMSD ("Root Mean Square Deviation of Atomic Position"): A menor valor, menor diferencia entre resultados obtenidos por método implícito y explícito $RMSD(X, Y) = \sqrt{\frac{1}{N_X} \sum_{i=1}^{N_X} ||x_i - y_i||}$.
- MUE ("Mean Unsigned Error"): A menor valor, menor error medio entre métodos. $MUE = \frac{\sum |X-Y|}{N_X}$.

La representación puede ser observada en la Figura 2.6.



Figura 2.6: Gráfico de Resultados BEM (Pygbe) vs Dinámica Molecular, echo en el programa energy plots.

A pesar de que las moléculas trabajadas están compuestas en base a Carbono (C) y todas tienen un número reducido de átomos, debido al gran número de moléculas estudiadas, es necesario categorizar cada una de ellas. El programa ha sido diseñado para clasificar a las moléculas, dependiendo la entrada deseada, en las siguientes categorías:

- Rango de Número de átomos.
- Rango de Volumen o Superficie.

- Presencia de un átomo en específico, Doble o Triple Enlace.
- Nombre o Nomenclatura.
- Grupo Funcional.

Dentro de la química orgánica, la mayoría de moléculas pueden ser clasificar por sus grupos funcionales, estos grupos pueden ser uno o más átomos que le dan reactividad y sus propiedades, he ahí su interés de estudio. Dentro de los grupos funcionales se pueden encontrar los aromáticos, alcoholes, nitros, ácidos carboxílico, etc [23]. La clasificación de cada molécula fue recopilada por David Mobley utilizando el software libre Checkmol. La base de datos se encuentra a libre disposición en el trabajo realizado en conjunto con Guthrie, FreeSolv [18].

Capítulo 3

Resultados y Análisis de Resultados

En este capítulo se muestran los resultados obtenidos al comparar la base de datos creada con PyGBe y la echa por Mobley usando Dinámica Molecular. El análisis se divide en 3 secciones; Número de átomos, Volumen y Superficie; Presencia de un átomo o enlace; y Grupo Funcional. En cada sección se buscará la constante dieléctrica ϵ que entregue los resultados que mejor relacione ambos métodos.

3.1. Comparación entre métodos en base a Número de átomos, Volumen y Superficie

Para iniciar con esta comparación se debe ingresar una de las variables "n", "sup" ó "vol"; dependiendo de si se desea analizar el número de átomos, la superficie o el volumen. A pesar de que el tamaño de la superficie y el volumen dependen de varios factores como los enlaces presentes, posicionamiento y estructura atómica; se puede decir que la dimensión de estos dependen directamente de la cantidad de átomos presentes en la molécula, por lo que las conclusiones de un análisis echas para el número de átomos son replicables tanto para la superficie como para el volumen.

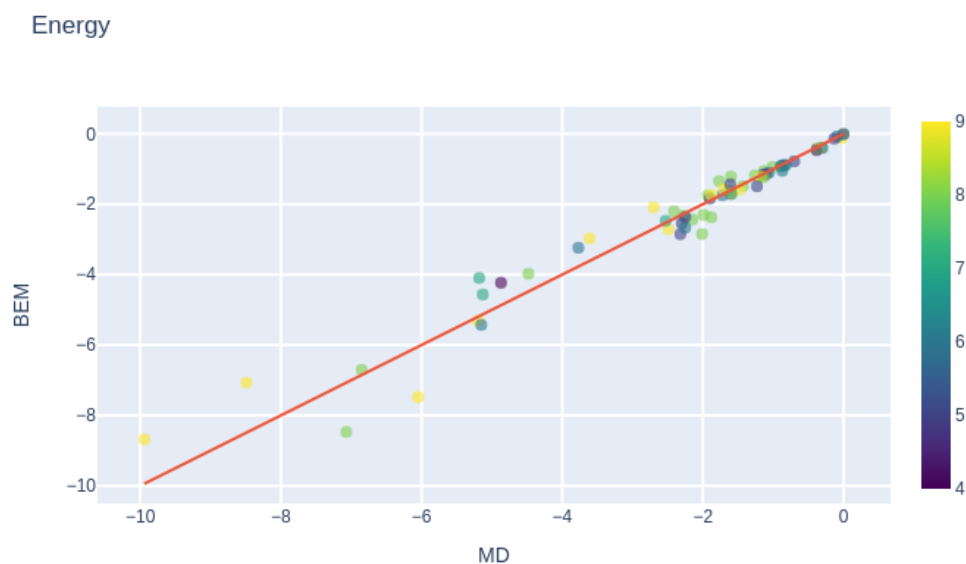


Figura 3.1: Gráfico de Energy Plots. $\epsilon = 1,35$; Barra Color = N° de Átomos: de 4 a 9; Coeficiente de Correlación = 0,9758; RMSD = 0,4566; MUE = 0,2814.

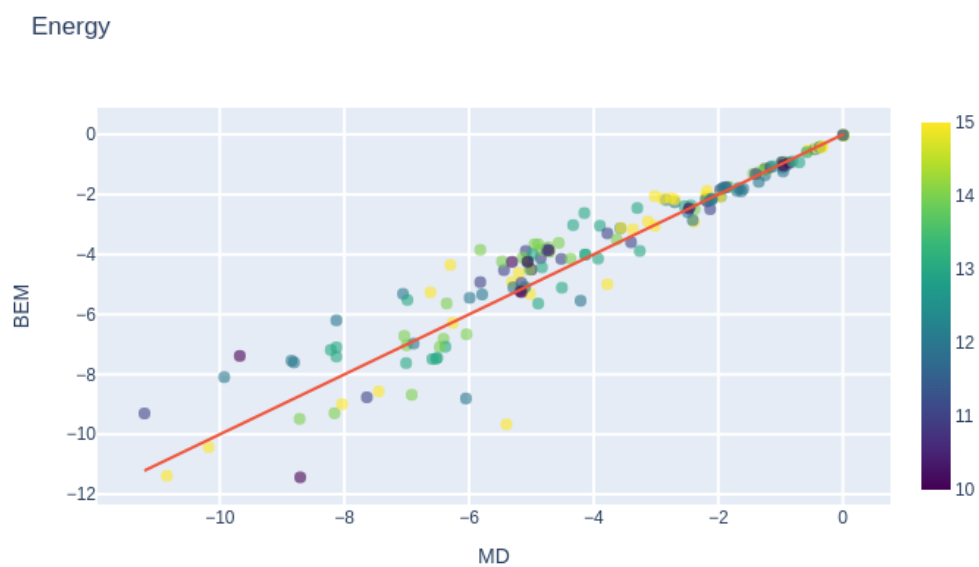


Figura 3.2: Gráfico de Energy Plots. $\epsilon = 1,25$; Barra Color = N° de Átomos: de 10 a 15; Coeficiente de Correlación = 0,9465; RMSD = 0,8660; MUE = 0,5783.

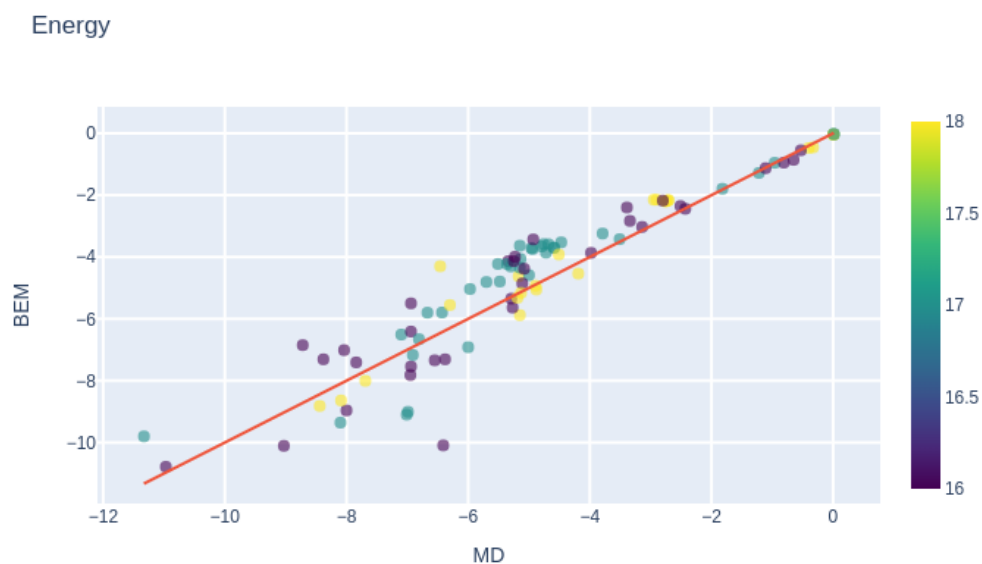


Figura 3.3: Gráfico de Energy Plots. $\epsilon = 1, 20$; Barra Color = N° de Átomos: de 16 a 18; Coeficiente de Correlación = 0,9424; RMSD = 0,9303; MUE = 0,7086.

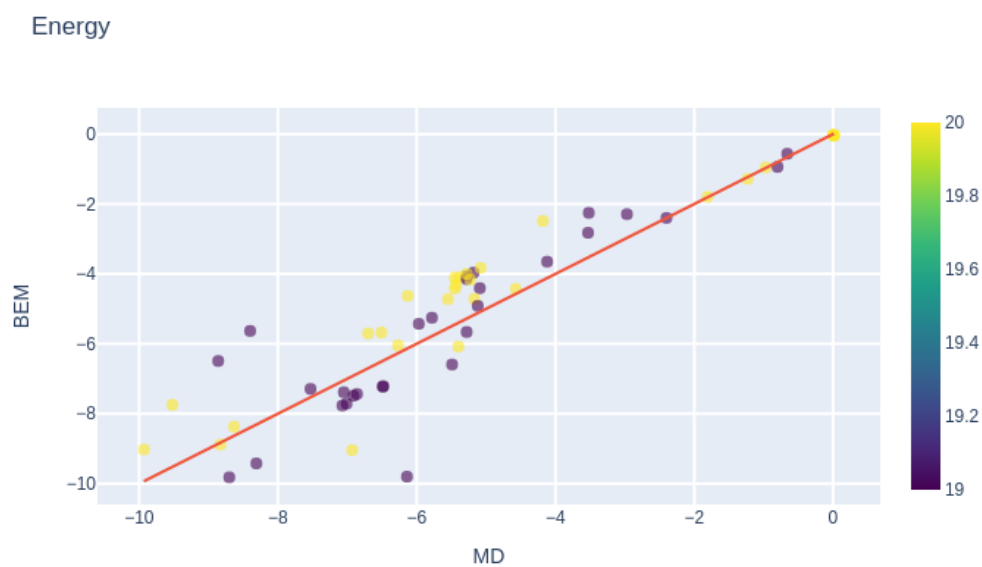


Figura 3.4: Gráfico de Energy Plots. $\epsilon = 1, 20$; Barra Color = N° de Átomos: de 19 a 20; Coeficiente de Correlación = 0,9241; RMSD = 1,0861; MUE = 0,8122.

Los gráficos que se presentan se crearon utilizando la constante ϵ que entregará los mejores coeficientes. Al observar los gráficos de las Figuras 3.1 a la 3.4, se es difícil apreciar una tendencia clara respecto al aumento en el número de átomos con la constante dieléctrica (ϵ). Por un lado, las moléculas más pequeñas (Rango de átomos de 4 a 9) no se ven afectados en gran medida al cambio de la constante dieléctrica, y la predicción de la energía de solvatación es relativamente precisa respecto al método de Dinámica Molecular. Por otra parte, las moléculas a medida que aumentan su tamaño se vuelven más complejas de trabajar y se vuelve difícil encontrar los parámetros que se ajusten a mejores resultados, a pesar de que cada vez se acoten mas los rangos de trabajo.

Dado que las propiedades electroquímicas de las moléculas dependen de distintas variables, y no solo del número de átomos, no se puede encontrar una constante dieléctrica óptima que entregue los mejores resultados según la cantidad de átomos.

3.2. Comparación entre métodos en base a presencia de átomos o enlaces

Por norma general, uno o más átomos unidos a una cadena de carbono forma parte de un grupo funcional, estos son responsables de diversas propiedades químicas en las moléculas. Los grupos funcionales serán tratados de forma más precisa en la sección 3.3. El carbono forma parte de todas las moléculas estudiadas, los resultados obtenidos por PyGBe para todas las moléculas se muestran en la Figura 3.5. Se entrega este gráfico para reconocer los átomos, enlaces o grupos funcionales que puedan ser complejos de calcular y generen una mayor dispersión en los resultados en las siguientes secciones.

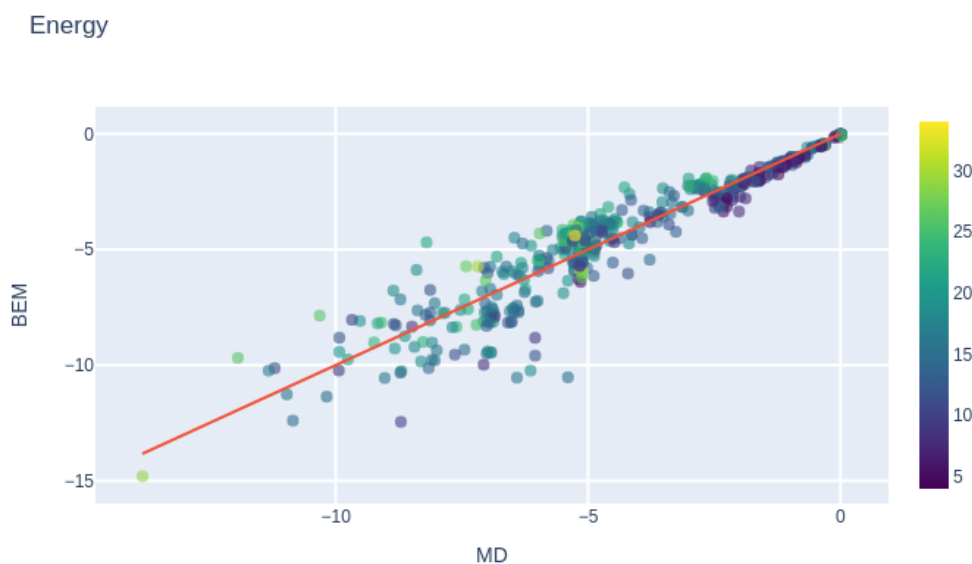


Figura 3.5: Gráfico de Energy Plots. $\epsilon = 1,15$; Moléculas con átomos de Carbono; Coeficiente de Correlación = 0,9468; RMSD = 0,9098; MUE = 0,6254. Barra de color representa el número de átomos.

En esta sección se tratarán los átomos y enlaces más comunes cuya presencia no signifique necesariamente formar parte de un grupo en específico, como lo pueden ser el Oxígeno ("O"), Nitrógeno ("N"), etc. Además, se verá solo la influencia de un átomo por separado, ya que varias moléculas poseen Nitrógeno en conjunto de Oxígeno y enlaces dobles.

3.2.1. Nitrógeno y Oxígeno



Figura 3.6: Gráfico de Energy Plots. $\epsilon = 1,15$; Moléculas con átomos de Nitrógeno; Coeficiente de Correlación = 0,8824; RMSD = 1,1425; MUE = 0,8164. En rojo destacadas las moléculas con presencia de Triple Enlace.



Figura 3.7: Gráfico de Energy Plots. $\epsilon = 1,20$; Moléculas con átomos de Oxígeno; Coeficiente de Correlación = 0,8704; RMSD = 1,1434; MUE = 0,9110. Barra de color representa el número de átomos.

Se puede observar que la presencia de Nitrógeno como Oxígeno generan inestabilidad y dispersión en los resultados, pero como se ha mencionado, las propiedades electroquímicas de una molécula dependen de los átomos que la conforman, de su estructura y en la forma que estos están enlazados. En el gráfico de la Figura 3.6 se puede apreciar una clara correlación en las moléculas resaltadas en rojo (Moléculas con presencia de Triple Enlace, se ajustan mejor a la recta con $\epsilon = 1,25$). Este conjunto de moléculas se caracterizan por tener un átomo de Nitrógeno unido a otro átomo mediante un triple enlace, conformando el grupo funcional de los Carbonitrilos.

Por esto se puede concluir que, para el método de BEM se vuelve complicado predecir una constante dieléctrica que mejor represente a un conjunto si sólo se considera la presencia de un único átomo.

3.2.2. Doble Enlace - =

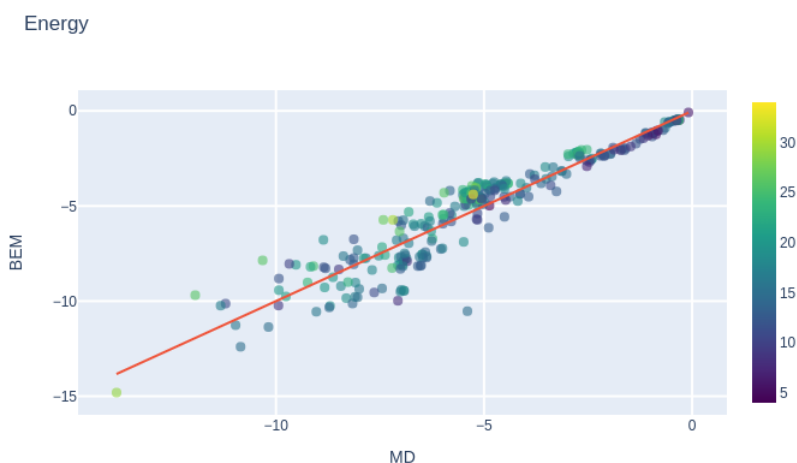


Figura 3.8: Gráfico de Energy Plots. $\epsilon = 1,15$; Moléculas con Doble Enlace; Coeficiente de Correlación = 0,9426; RMSD = 0,9330; MUE = 0,7143. Barra de color representa el número de átomos.

Cuadro 3.1: Resultados obtenidos con el cálculo de 18 moléculas con presencia de Triple Enlace.

ϵ	Coef Corr	RMSD	MUE
1,25	0,9893	0,4371	0,3314

Se puede apreciar tanto en la Figura 3.8 como en el Cuadro 3.1 que la presencia de enlaces más fuertes tienden a disminuir la dispersión en los resultados. Esto puede tener relación con la estabilidad que le dan a la molécula la presencia de estos.

Las moléculas con doble enlace forman cerca del 50 % de las moléculas estudiadas, por lo que es normal ver un coeficiente de correlación, RMSD y MUE similares entre la totalidad de las moléculas estudiadas y las que únicamente poseen doble enlace. Por otro lado las moléculas con triple enlace representan tan sólo 18 de las cerca de 500 moléculas estudiadas, pero los resultados obtenidos muestran una alta correlación, escasa dispersión y bajo error, lo que confirma la idea de que a mayor estabilidad en la estructura molecular, es más sencillo encontrar y utilizar una constante (ϵ) que mejor represente el nivel de polarización de la misma.

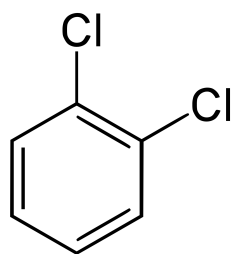
A pesar de que se puedan obtener resultados más o menos precisos dependiendo la situación, basar el análisis en una única variable no sería correcto por las demás variables ignoradas, por ello no es posible encontrar una única constante dieléctrica que mejor represente a cada conjunto estudiado en la sección (a excepción de utilizar el triple enlace).

3.3. Comparación entre métodos en base a un Grupo Funcional

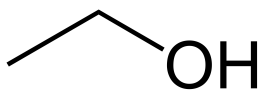
Un grupo funcional es un conjunto de átomos y enlaces distintivos, se encuentran unidos a una molécula mediante un enlace covalente y le dan las propiedades químicas a esta. El grupo funcional es generalmente la parte de la molécula donde ocurren las reacciones, el grupo sufrirá las mismas reacciones independientemente del resto de la molécula a la cual está acoplada. El hecho de que la mayoría de los compuestos orgánicos formen parte de algún

grupo funcional, junto a las propiedades que entrega cada uno de estos, es que permite el estudio y predicción de forma sistemática a sus reacciones químicas y el comportamiento de sus compuestos [15].

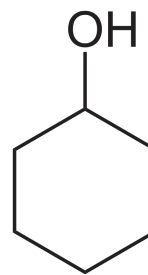
Dentro de las moléculas estudiadas hay una gran variedad de grupos presentes, a excepción de un pequeño grupo de moléculas y los alcanos, todas pertenecen a un único grupo funcional. Por otro lado, los grupos funcionales forman parte de dos familias principales, estas son los Compuestos Aromáticos y los Compuestos No-Aromáticos.



(a) Aromático: Dicloroben-
ceno (Aril).



(b) No-Aromático: Etanol
(Alcohol).



(c) No-Aromático: Ciclo-
benzanol (Alcohol).

Figura 3.9: Familias de grupos funcionales.

Los compuestos aromáticos representan cerca de la mitad de los compuestos orgánicos, son cíclicos y poseen una gran estabilidad. Los compuestos aromáticos son representados en el gráfico de la Figura 3.10, dentro de estos se encuentran distintos grupos como lo pueden ser los Heterocíclicos Aromáticos, Arilos o el grupo Fenol.

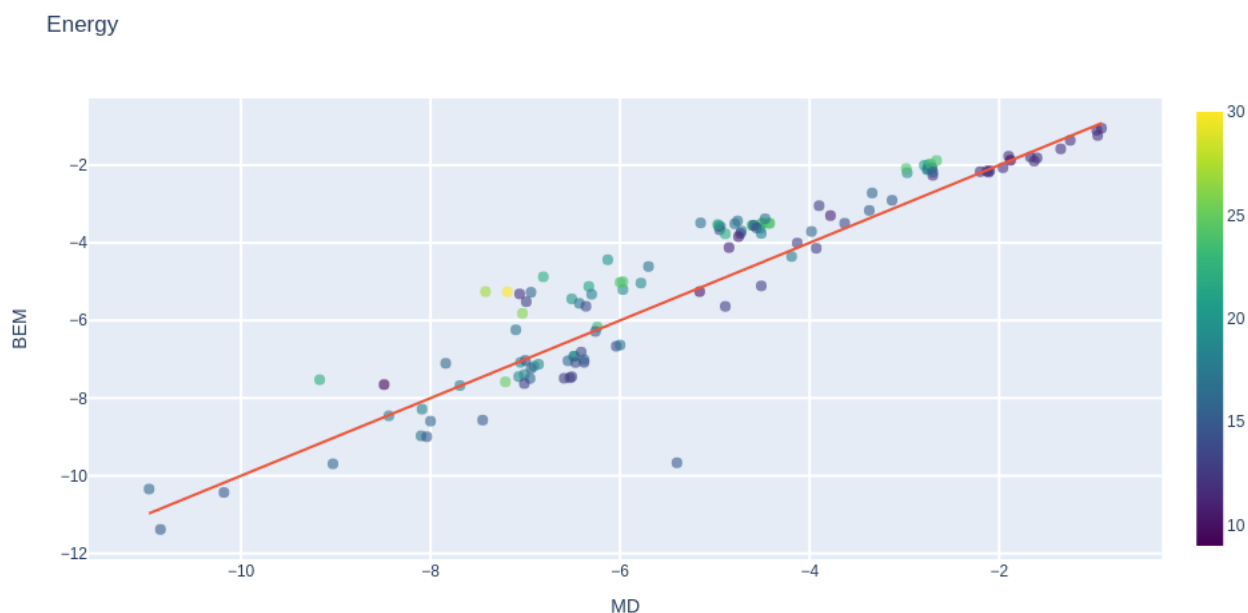


Figura 3.10: Gráfico de Energy Plots. $\epsilon = 1,25$; Compuestos Aromáticos; Coeficiente de Correlación = 0,9391; RMSD = 0,9320; MUE = 0,7431. Barra de color representa el número de átomos.

Cuadro 3.2: Compuestos Estudiados.

Compuestos Aromáticos	Compuestos No-Aromáticos
Arilos	Alcoholes - Aldehidos
Fenoles	Alquenos - Haloalcanos
	Carbonitrilos -Cetonas
	Éter - Nitro

El Cuadro 3.2 muestra los grupos funcionales estudiados por Mobley [16]. Los grupos con mayor presencia se estudian en la presente sección.

3.3.1. Compuestos Aromáticos

Los compuestos aromáticos representan el mayor de los grupos estudiados, dentro de este tipo de compuesto se encuentran los grupos Arilos y Fenoles, con escasa presencia de

moléculas de Aminas y otros pequeños grupos. Para que un compuesto pueda ser considerado como aromático debe cumplir los siguientes criterios [23]:

- La estructura debe ser cíclica y contener cierta cantidad de enlaces π conjugados (enlaces covalentes donde 2 lóbulos de una orbita solapan con otros 2 de una órbita distinta).
- Cada átomo del anillo debe tener un orbital p sin hibridar.
- Los orbitales p atómicos sin hibridar se deben traslapar para formar un anillo continuo de orbitales paralelos.
- La deslocalización de los electrones π en el anillo debe disminuir la energía electrónica.

Por norma general los compuestos aromáticos son bastante estables e insolubles en agua (a menos que sea un grupo funcional con algunos elementos fuertemente polares, como el fenilo), tienen presencia en las hormonas y vitaminas, condimentos y perfumes. Por otro lado, gran parte de estos compuestos pueden ser dañinos para la salud como el benceno y tolueno, incluso hasta cancerígenos como lo puede ser el alfa-benzopireno (presente en el humo del tabaco) y el mismo benceno. Al existir distintos grupos dentro de los compuestos Aromáticos, se pueden encontrar diferentes constantes que mejor representen a cada grupo (Cuadro 3.3).

Cuadro 3.3: Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Aromáticos. Grupo Fenol (b) es el grupo Fenol sin considerar 2-nitrophenol; 2,5-dimethylphenol y 2,6-dimethylphenol.

	N° Molec	ϵ	Coef Corr	RMSD	MUE
Arilos	32	1.35	0,9908	0,3394	0,2437
Fenoles	28	1.35	0,8513	0,7738	0,4069
Fenoles (b)	25	1.35	0,9842	0,2549	0,2230
Heterociclico Aromático	25	1.00	0,9424	0,4988	0,3672

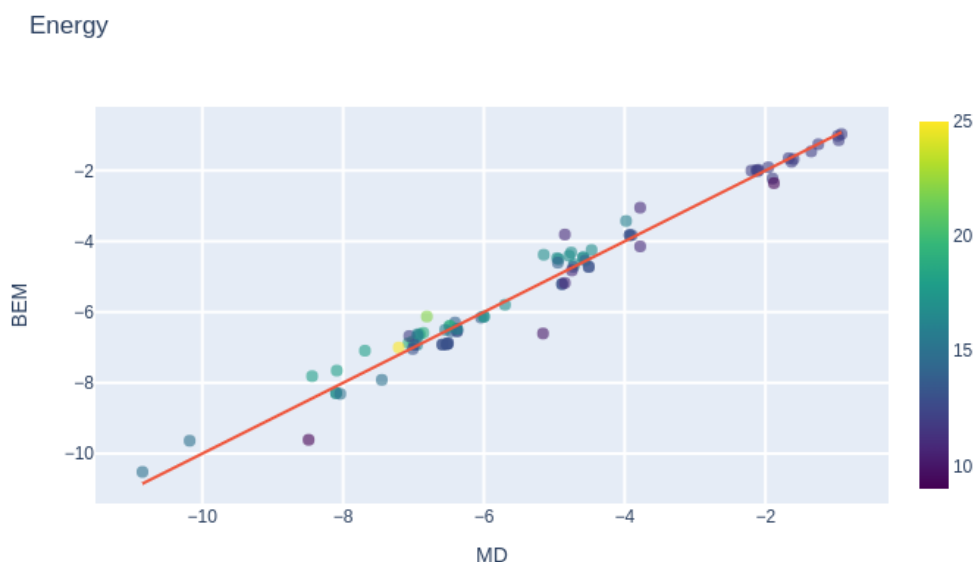


Figura 3.11: Gráfico de Energy Plots. Moléculas Aromáticas Cuadro 3.3; Coeficiente de Correlación = 0,9854; RMSD = 0,3759; MUE = 0,2750. Barra de color representa el número de átomos.

Puesto que los compuestos aromáticos tienen reglas bien definidas en como están conformadas, son grupos estables en lo que respecta a sus enlaces y reacciones, en general es posible identificar constantes que representen a cada grupo de manera correcta, como lo pueden ser los casos de los grupos representados en el Cuadro 3.3. Las excepciones vienen cuando las moléculas forman parte de más de un grupo o grupos más complejos, como lo pueden ser las aminas o ácidos aromáticos.

3.3.2. Compuestos No-Aromáticos

Dentro de los compuestos No-Aromáticos se encuentra una gran variedad de grupos funcionales, estos grupos se caracterizan por no cumplir los criterios establecidos en la sección 3.3.1. Estos compuestos también son conocidos como alifáticos y pueden ser abiertos (cadena) o cerrados (cíclicos). A pesar de que la mayoría de los Hidrocarburos Alifáticos son inflamables, estos tienen distintas propiedades y usos según su grupo funcional. Así como es

posible representar ciertos grupos acorde a una constante dieléctrica, también existen grupos más complejos que se pueden volver más difíciles de trabajar como lo puede ser los alcoholes, cetonas o las aminas alifáticas.

Cuadro 3.4: Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Alifáticos.

	N° Molec	Epsilon	Coef Corr	RMSD	MUE
Alqueno	31	1.25	0,9903	0,1495	0,0993
Haloalcano	63	1.30	0,9877	0,2912	0,2158
Carbonitrilos	7	1.25	0,9963	0,3747	0,3341
Derivado de Halógenos	20	1.40	0,9717	0,6814	0,3892

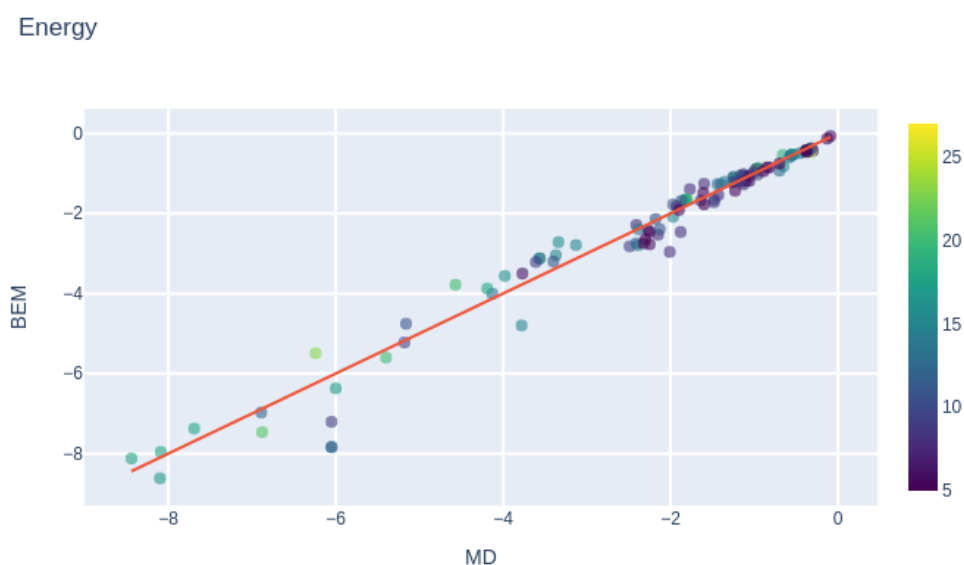


Figura 3.12: Gráfico de Energy Plots. Moléculas Alifáticas Cuadro 3.4; Coeficiente de Correlación = 0,9838; RMSD = 0,3671; MUE = 0,2214. Barra de color representa el número de átomos.

La relación que poseen los grupos que pueden ser bien representados (Cuadro 3.3 y Cuadro 3.4) es que todos, a excepción de los fenoles, son insolubles en agua pero solubles en otros compuestos orgánicos. Dentro del grupo fenólico existen diversos tipos de moléculas y

grados de solubilidad, pero al ser compuestos aromáticos se permite una buena representación en su generalidad.

Cuadro 3.5: Resultados Obtenidos con Pygbe vs Dinámica Molecular para Compuestos Alifáticos poco predecibles.

	N° Molec	Epsilon	Coef Corr	RMSD	MUE
Alcohol	35	1.35	0,8527	1,0688	0,8217
Cetona	19	1.00	0,2454	0,4392	0,3769
Nitro	6	1.20	0,6948	0,1157	0,0749
Ether	30	1.05	0,8780	1,4969	0,8194

Los compuestos presentes en el Cuadro 3.5 son difíciles de tratar en las condiciones trabajadas, poseen una muy baja correlación en sus resultados respecto al método de Dinámica Molecular y por otro lado presentan una gran dispersión en los datos. A diferencia de los compuestos en los que hay una alta correlación de resultados, estos grupos son altamente solubles en agua.

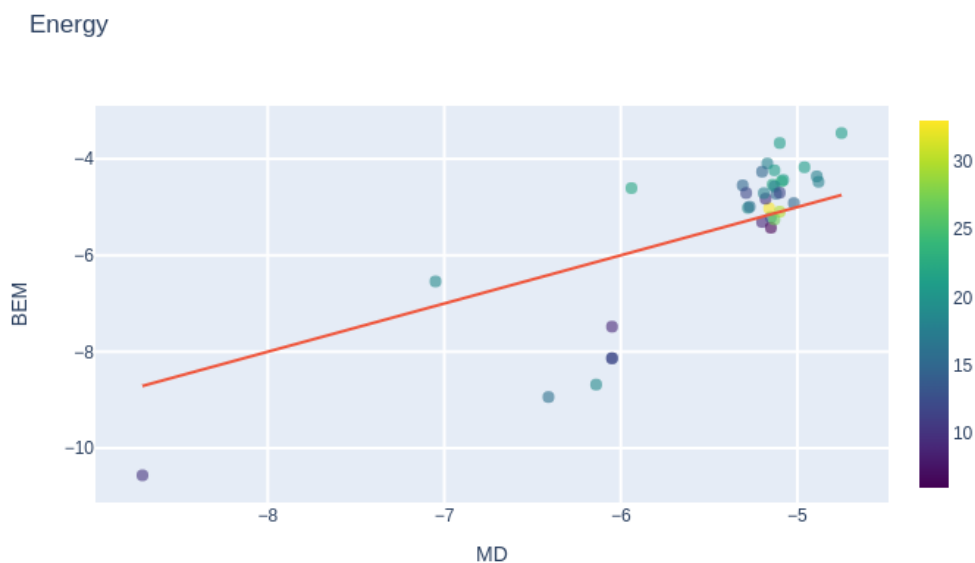


Figura 3.13: Gráfico de Energy Plots. $\epsilon = 1,35$; Alcoholes; Coeficiente de Correlación = 0,8527; RMSD = 1,0688; MUE = 0,8217. Barra de color representa el número de átomos.

Por otro lado, los grupos presentes en el Cuadro 3.5 se comportan de manera similar, al igual que en el gráfico de de la Figura 3.13 tanto en el grupo Nitro como el de las Cetonas, las moléculas se tienden a acumular en su mayoría en una zona. Si bien estos últimos 2 grupos poseen una baja dispersión y error, las energías tienden a ser pequeñas y no poseen una mayor correlación en los resultados.

Finalmente, en el grupo de los Éteres existe una puñado de moléculas no consideradas en el Cuadro 3.5 los cuales si pueden ser representados de manera correcta, estas moléculas pertenecen al grupo de los "alkyl aryl ether"(Ariilos) las cuales son aromáticas.

3.3.3. Ácidos y Aminas

Estos grupos están clasificados como complejos ya que una proporción no menor de moléculas son aromáticas, por otro lado existe una variedad de ácidos y aminas derivadas del grupo principal, que si bien no cambian algunas de las propiedades principales del grupo funcional, si tienen diferencias los unos de los otros.

Ácidos y Derivados

La mayoría de los ácidos estudiados son derivados de ácidos carboxílicos, en su mayoría pertenecen al grupo de ésteres de ácidos carboxílicos. Los ácidos en general son solubles en agua, al igual que en casos anteriores se generan problemas al momento de calcular la energía de solvatación mediante el uso de PyGBe.

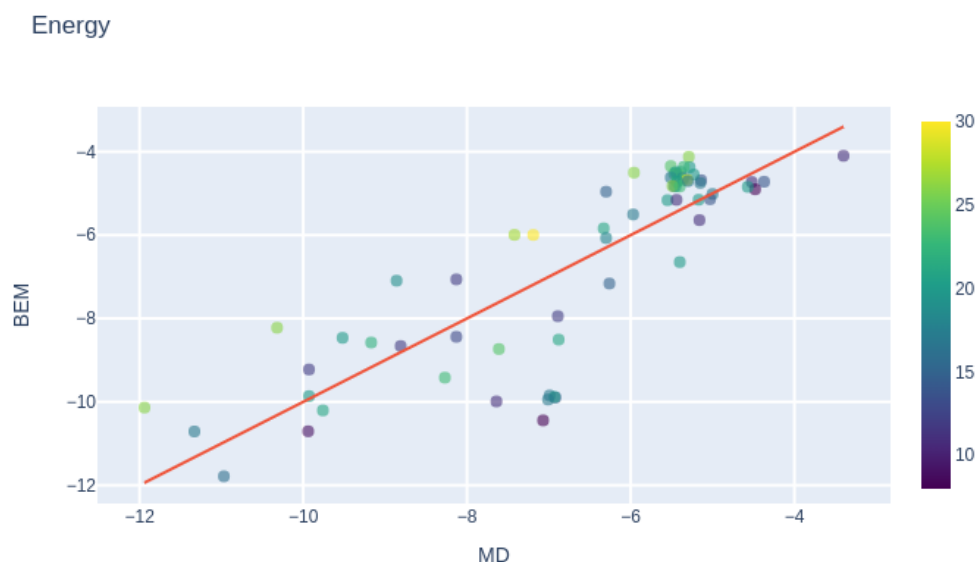


Figura 3.14: Gráfico de Energy Plots. $\epsilon = 1,10$; Ácidos y Derivados; Coeficiente de Correlación = 0,8420; RMSD = 1,2251; MUE = 0,9641. Barra de color representa el número de átomos.

Cuadro 3.6: Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, Ácidos y Derivados.

	N° Molec	Epsilon	Coef Corr	RMSD	MUE
Ácidos Carboxílicos	6	1.50	0,1354	0,3282	0,3068
Ester de Ácido Caboxílico	44	1.05	0,9265	0,7452	0,6275
Amida de Ácido Carboxílico	7	1.10	0,9368	0,9683	0,8803
Ácidos Aromáticos	7	1.10	0,9172	0,8998	0,7316

En el Cuadro 3.6 se puede apreciar distintos derivados de los Ácidos Carboxílicos, todos estos tienen distintas propiedades y formas de producirse. Los compuestos derivados de mayor presencia entre las moléculas estudiadas es el grupo de los Ésteres de Ácido Carboxílico, estos Ésteres son producidos mediante la reacción de un Ácido Carboxílico y un Alcohol; a pesar de ser formado por dos grupos fuertemente solubles en agua, este grupo es hidrofóbico (poco soluble o insoluble en agua), esto se ve reflejado en unos resultados relativamente buenos y mejores que los otros compuestos estudiados en esta sección.

Aminas

Las Aminas son compuestos orgánicos que derivan del amoníaco (Fig 3.15), se forman por la sustitución de uno de sus Hidrógenos por otros radicales (sustituyentes). En caso de que sólo un Hidrógeno sea sustituido, se considera al nuevo compuesto como una Amina Primaria, y una Amina Secundaria o Terciaria si dos o tres Hidrógenos son reemplazados.

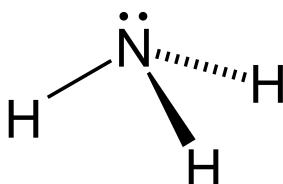


Figura 3.15: Molécula de Amoníaco.

Las Aminas son consideradas simples si los sustituyentes son el grupo Alquilo (Metilo, Etilo, Propilo, etc), Aromáticas si los sustituyentes son Arilos y Mixtas si los grupos son distintos a los Alquilos o Arilo.

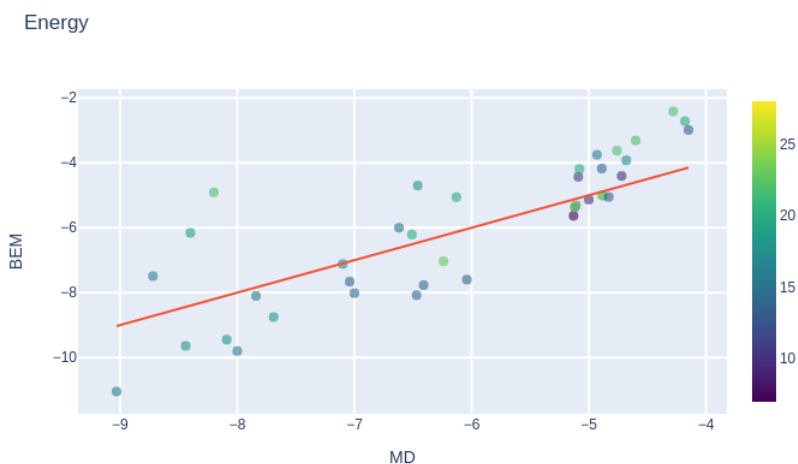


Figura 3.16: Gráfico de Energy Plots. $\epsilon = 1, 10$; Aminas; Coeficiente de Correlación = 0,8372; RMSD = 1,2008; MUE = 0,9842. Barra de color representa el número de átomos.

Las propiedades de las Aminas varían dependiendo de cuantos Hidrógenos sean sustitui-

dos, también la solubilidad de la molécula se reduce a medida que las Aminas poseen más átomos.

Cuadro 3.7: Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, Ácidos.

	N° Molec	Epsilon	Coef Corr	RMSD	MUE
Aminas Alifáticas Primarias	10	1.15	0,9100	0,3659	0,1975
Aminas Alifáticas Secundarias	10	1.00	0,9466	0,7199	0,5558
Aminas Alifáticas Terciarias	6	1.00	0,9368	0,9683	0,8803
Aminas Aromáticas Primarias	11	1.30	0,8706	0,5526	0,4254

Aunque las propiedades principales y su estabilidad no se vean altamente afectadas, el hecho de que las Aminas puedan ser formadas por múltiples grupos dificulta la posibilidad de que se puedan representar en su generalidad, a pesar de estar clasificadas en Aminas primarias, secundarias o terciarias.

3.3.4. Alcanos

Los Alcanos son compuestos que solo poseen Carbono e Hidrógeno enlazados únicamente por enlaces covalentes simples. Son los únicos compuestos que carecen de grupo funcional, son hidrofóbico y casi totalmente insolubles en el agua, por lo que una reacción con esta es muy poco probable. Como el proceso que se ha realizado en el presente trabajo y en el de Mobley es de hidratación (proceso en el que se le agrega agua a un compuesto); ya que los alcanos son prácticamente insolubles en agua, es natural ver las componentes electrostáticas de la energía de solvatación equivalente a 0 (Figura 3.17).

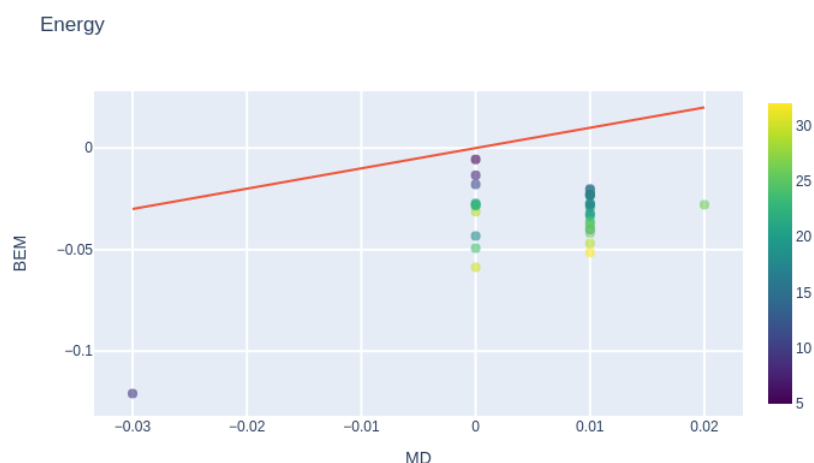


Figura 3.17: Gráfico de Energy Plots. 22 Moléculas Alcanas estudiadas; Coeficiente de Correlación = 0,6201; RMSD = 0,0420; MUE = 0,0393. Barra de color representa el número de átomos.

Por otro lado, los alcanos son un grupo fuertemente no-polar, por lo que la variación de la constante dieléctrica (constante de polarización, ϵ) no influye en gran medida en los cálculos y sus resultados, como se puede apreciar en el Cuadro 3.8.

Cuadro 3.8: Comparación de Resultados Obtenidos con Pygbe vs Dinámica Molecular, 22 Compuestos Alcanos evaluados con distinta constante dieléctrica.

ϵ	Coef Correl	RMSD	MUE
1.00	0,6196	0,0575	0,0531
1.40	0,6201	0,0420	0,0393

3.4. Resultados finales.

Para presentar los resultados finales se separan las moléculas pertenecientes a los grupos funcionales insolubles en un documento y los solubles presentes en el Cuadro 3.5 en otro. La recopilación de las moléculas se muestran en las Figuras 3.18 y 3.19.

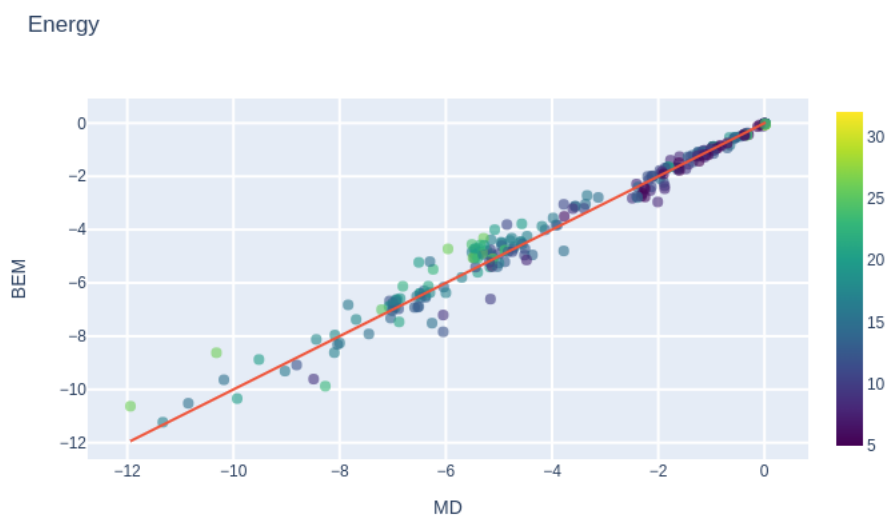


Figura 3.18: Gráfico de Energy Plots. Moléculas pertenecientes a grupos funcionales poco solubles en Agua y Alcanos; Coeficiente de Correlación = 0,9888; RMSD = 0,4231; MUE = 0,2701. Barra de color representa el número de átomos.

Los resultados muestran claramente que los grupos funcionales no-polares o insolubles en agua pueden ser calculados con una escasa dispersión y una alta correlación respecto a los obtenidos en Dinámica Molecular. Es importante destacar que para obtener todos estos resultados se deben utilizar distintas constantes dieléctricas acorde al grupo con el cuál se trabaja. A diferencia de las moléculas no-polares, los grupos fuertemente polares tienen distintos comportamientos, como acumularse en zonas y presentar una baja correlación ó una alta dispersión, estos grupos son complejos de tratar y el cálculo de su energía de solvatación con PyGBe es poco recomendable.

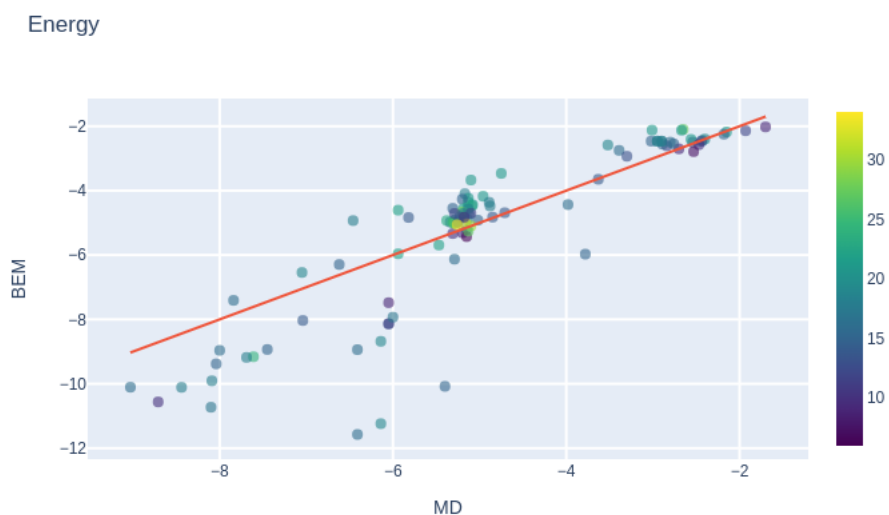


Figura 3.19: Gráfico de Energy Plots. Moléculas solubles de grupos funcionales presentes en el Cuadro 3.5; Coeficiente de Correlación = 0,8831; RMSD = 1,2327; MUE = 0,7828. Barra de color representa el número de átomos.

Al observar los resultados se pueden observar distintas variables que pueden influir en estos, dentro de los importantes a destacar se encuentra la estabilidad de los enlaces y la estructura, pero principalmente la solubilidad. Es difícil explicar por qué los compuestos menos polares (o no-polares) pueden ser mejor representados que los fuertemente polares al utilizar este modelo, la razón de ello puede ser un buen material de estudio.

Capítulo 4

Conclusiones

La computación y sus aplicaciones son áreas que se encuentran en constante crecimiento, y el cálculo de la energía de solvatación mediante métodos computacionales no es una excepción; en la búsqueda de la eficiencia y precisión se han desarrollado distintos estudios y modelos como los pueden ser los presentados por David Mobley, Emilio Galiccio o el utilizado en este trabajo, PyGBe, entre otros.

Todas las tareas se desarrollaron mediante el lenguaje de programación Python, un lenguaje sencillo de implementar y de gran versatilidad, permitiendo que usuarios de poca experiencia puedan acceder al desarrollo y mejoramiento de funciones sin mayores dificultades. Si bien PyGBe fue desarrollado en base a un método implícito para trabajar de manera más rápida con biomoléculas de gran tamaño, utilizar esta herramienta en un gran número de distintas moléculas más pequeñas permite tener una mejor comprensión al momento de implementar el código en compuestos más complejos.

Con el fin conocer las condiciones en las que PyGBe obtiene los resultados de la manera más precisa, se realizaron cálculos a la energía de solvatación a poco menos de 500 moléculas orgánicas de tamaño pequeño; inicialmente las moléculas se clasificaron respecto al número de átomos y la presencia de algún elemento específico, obteniendo resultados con escasa relación y poco concluyentes. En segunda instancia, se clasificaron las moléculas en base a sus respectivos grupos funcionales llegando a conclusiones que pueden ser de interés. Como se ha mencionado en las secciones anteriores, los grupos funcionales es un conjunto de elementos que le dan las propiedades a una molécula orgánica independiente del tamaño y el resto de

su composición. Dentro de los grupos presentes en las moléculas estudiadas, existen algunos que con determinada constante entregan excelentes resultados, siendo el común entre todos estos la poca solubilidad de los compuestos en agua, es decir, ser compuestos no-polares.

Hay que destacar que los buenos resultados se dan al momento de compararlos con su símil en Dinámica Molecular, un modelo que en su generalidad es más preciso. Lamentablemente este modelo tampoco está exento de errores, así como en varios grupos se logra replicar su precisión, también son replicados los errores sistemáticos de sus resultados, como lo es en el caso de los haloalcanos de bromo ("alkyl bromide") y los carbonitrilos.

Ya que el foco principal de este trabajo era encontrar parámetros que mejor representen a las moléculas acorde a una clasificación, se es difícil decir por qué los compuestos poco solubles en agua obtienen buenos resultados, altos niveles de correlación y escasa dispersión con una determinada constante dieléctrica mientras los compuestos con mayor solubilidad no. Sin duda alguna las conclusiones de este trabajo pueden ser un material de estudio interesante y explicar a que se deben los valores de la constante dieléctrica puede ser un tema interesante a desarrollar.

Referencias

- [1] Aguilar, B. and Onufriev, A. (2012). Efficient Computation of the Total Solvation Energy of Small Molecules via the R6 Generalized Born Model. *Journal of Chemical Theory and Computation*. 8. 2404–2411.
- [2] Baker, Nathan. (2004). Poisson-Boltzmann Methods for Biomolecular Electrostatics. *Methods in enzymology*. 383. 94-118. 10.1016/S0076-6879(04)83005-2.
- [3] Barnes, j. and Hut, P. (1986). A hierarchical $O(N\log N)$ force-calculation algorithm. *Nature*, 324:446-449
- [4] Bellissent-Funel, Marie-Claire and Hassanali, Ali and Havenith, Martina and Henchman, Richard and Pohl (2016). Water Determines the Structure and Dynamics of Proteins. *Chemical Reviews*, Vol. 116, (13), 7673-7697.
- [5] Brooks, B. Brucoleri, R. Olafson, B. (2004). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry*. 4. 187 - 217.
- [6] Byung Jun & Yoon A. M. Lenhoff (1990). A boundary element method for molecular electrostatics with electrolyte effects.
- [7] C.A. Brebia & J. Dominguez (1991). *Boundary elements: An Introductory Course*. Boston, Southampton: WIT PRESS.
- [8] Christopher Cooper Villagrán. *Biomolecular Electrostatics with Continuum models: A Boundary Integral Implementation and application to Biosensors*.
- [9] Cooper, C. D., Clementi, N. C., Forsyth, G., & Barba, L. A. (2016). PyGBe: Python, GPUs and Boundary elements for biomolecular electrostatics. *The Journal of Open Source Software*.

-
- [10] Gallicchio, Emilio and Paris, Kristina and Levy, Ronald M. (2009) The AGBNP2 Implicit Solvation Model. *Journal of Chemical Theory and Computation*, Vol. 5, (9) 2544-2564.
- [11] Haider, N. (2003) Checkmol/Matchmol
- [12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242. (rcsb.org).
- [13] Jackson, J. D. (1998). *Classical Electrodynamics*. Wiley, 3rd edition.
- [14] John T. Katsikadelis, "Boundary Elements Theory and Applications", Elsevier, Oxford, 2002
- [15] March, Jerry and Smith, Michael B. (2007). *March's Advanced Organic Chemistry: Reactions, Mechanisms and Structure*. 6th Edition. Wiley-Interscience.
- [16] Mobley, David L. and Bayly, Christopher I. and Cooper, Matthew D. and Shirts, Michael R. and Dill, Ken A. (2009) Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *Journal of Chemical Theory and Computation*, Vol. 5, (2), 350-358.
- [17] Mobley, David L. and Klimovich, Pavel V. (2012). Alchemical free energy calculations for drug discovery. *The Journal of Chemical Physics*, Vol. 137, (23).
- [18] Mobley, David L. and Guthrie, J. (2014). FreeSolv: a database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design*, Vol. 28, (7), 711-720.
- [19] Nelson MT, Humphrey W, Gursoy A. (1996). NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *The International Journal of Supercomputer Applications and High Performance Computing*, 10(4), 251-268.
-

-
- [20] Rizzo, Robert C. and Aynechi, Tiba and Case, David A. and Kuntz, Irwin D. (2006) Estimation of Absolute Free Energies of Hydration Using Continuum Methods: Accuracy of Partial Charge Models and Optimization of Nonpolar Contributions. *Journal of Chemical Theory and Computation*, Vol. 2, (1), 128-139.
- [21] Sanner, M.F., Spohner, J.-C., and Olson, A.J. (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, Vol. 38, (3),305-320.
- [22] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. (2005). GROMACS: fast, flexible, and free. *J Comput Chem*. 26 (16): 1701–18.
- [23] Wade, L.G. (2004). *Química Orgánica*. Pearson Educación.