

Universidad Católica de Santa María

Facultad de Ciencias Farmacéuticas, Bioquímicas y Biotecnológicas

Programa Profesional de Ingeniería Biotecnológica



Evaluación de las propiedades termodinámicas
en el proceso de trimerización de la proteína
CusC perteneciente a la *Acidithiobacillus
ferrooxidans*, comparada con la proteína
CusC de *Escherichia coli* usando herramientas
bioinformáticas.

Tesis presentada por la Bachiller:

Roxana Gabriela Cari Vargas

Para Optar el Título Profesional de:

Ingeniero Biotecnólogo

Asesor: Ing. Jaime Barreda

Coasesor: PhD. Badhin Gómez Váldez

Arequipa – Perú

2012

Agradecimientos

A Dios, por ayudarme a superar cada una de las barreras que se presentaron a lo largo de mi carrera.

A mi madre, Jeanette Vargas, que fue siempre una figura ejemplar, gracias a sus consejos y amor me ayudaron a alcanzar esta meta que no será la última (un beso le mando para el cielo). A mi padre Carmelo Cari, por brindarme todo su apoyo y por depositar toda su confianza en mí y a mis hermanos, Gabriel y Mauricio por estar allí en los momentos difíciles y también en los más felices.

Al Dr. Badhin Gómez Váldez, por haberme asesorado y orientado siempre hacia el camino correcto. Le doy las gracias por haberme brindado toda la ayuda académica necesaria para la culminación de mi tesis.

A todos los profesores que de alguna u otra manera me ayudaron a alcanzar esta meta.

A mis amigos del grupo de bioinformática, mis hermanos académicos, por todo el apoyo que me brindaron y por hacer placentera mi vida en la universidad.

Contenido

Contenido	I
Índice de Figuras	V
Índice de Tablas	IX
Glosario	X
Resumen	XII
Abstract	XIII
1. Introducción	1
Objetivos	4
1.0.1. Objetivo General	4
1.0.2. Objetivos Específicos	4
Hipótesis	5
2. Marco Teórico	6
2.1. Biominería	6

CONTENIDO

II

2.1.1.	Biolixiviación	7
2.1.2.	Acidithiobacillus ferrooxidans	11
2.1.3.	Escherichia coli	12
2.2.	Homeostasis del cobre	13
2.3.	Mecanismos de Resistencia /Tolerancia a metales	16
2.3.1.	Facilitadores de difusión de cationes (CDF)	17
2.3.2.	ATPasas tipo P	17
2.3.3.	Bomba de flujo de salida CBA	18
2.4.	Bioinformática	23
2.5.	Modelado Molecular	25
2.5.1.	Mecánica Molecular	26
2.5.2.	Campos de Fuerza para Modelamiento Molecular	29
2.5.3.	Optimización geométrica	31
2.5.4.	Dinámica Molecular	37
2.5.5.	Métodos de cálculo con disolvente	39
2.5.6.	Mecánica Estadística	41
3.	Materiales y Métodos	45
3.1.	Equipos y Software	45
3.1.1.	Hardware	45
3.1.2.	Bases de Datos	45
3.1.3.	Software	46
3.1.4.	Servidores online	48
3.2.	Búsqueda de las secuencias usadas para las simulaciones	50

CONTENIDO	III
3.3. Obtención de las estructuras secundarias	51
3.4. Minimización de las estructuras	51
3.5. Dinámica molecular de las estructuras	53
3.6. Análisis de la estabilidad	54
3.7. Construcción de los trímeros	57
3.8. Potencial Electrostático	58
3.9. Obtención y Estudio de Parámetros Termodinámicos	59
4. Resultados y Discusión	60
4.1. Proteínas CusC de <i>Escherichia coli</i> y <i>Acidithiobacillus ferrooxidans</i> .	60
4.1.1. Búsqueda de las secuencias usadas para las simulaciones . . .	60
4.1.2. Obtención de las estructuras secundarias	61
4.1.3. Minimización de las estructuras	63
4.1.4. Dinámica Molecular de estructuras	63
4.1.5. Análisis de la estabilidad	64
4.2. Trímeros de las proteínas CusC de <i>E. coli</i> y <i>A. ferrooxidans</i>	73
4.2.1. Construcción de los trímeros	73
4.2.2. Minimización de las estructuras	76
4.2.3. Dinámica Molecular de estructuras	76
4.2.4. Análisis de la estabilidad	76
4.2.5. Potencial electrostático	85
4.2.6. Parámetros termodinámicos	90
Conclusiones	94

CONTENIDO	IV
Sugerencias	95
Bibliografía	96
Anexos	102
A. Archivos utilizados	103
B. Ecuaciones utilizadas	108
C. Datos de la Simulación	109
D. Programa para construir el trímero	110
E. Conceptos adicionales	113



Índice de Figuras

2.1. Interacciones más comunes bacteria-metal. ⁷	7
2.2. Estructura propuesta del complejo de proteico que incluye: (1) un transportador de membrana citoplasmática del tipo RND (blanco), (2) una proteína de fusión de membrana (MFP; negro) y (3) un factor de membrana externo (OMF, sombreado). ¹¹	19
2.3. Modelo funcional del sistema Cus. Las tres partes de este complejo consisten en la bomba de la membrana interna CusA, la proteína periplasmica CusB y la proteína de membrana externa CusC las cuales forman un canal al espacio periplásmico. La entrada del cobre puede ocurrir a partir del periplasma (A), a partir del citoplasma (B) o por la vía de la chaperona CusF (C). ³⁴	21
2.4. La Bioinformática como convergencia multidisciplinar	24
2.5. Interacciones más importantes que se consideran en mecánica molecular. ⁵²	27
2.6. Búsqueda lineal es usada para localizar el mínimo en la función en la dirección del gradiente	35
3.1. Rotación de coordenadas.	47
3.2. Diagrama de Ramachandran caso general.	50
3.3. Imagen del servidor Robetta donde se copia la secuencia FASTA. . . .	51
3.4. Criterio geométrico de los enlaces de hidrógeno.	56

3.5. Imagen del servidor MolProbity donde se sube la estructura promedio para el diagrama de Ramachandran.	56
3.6. Diagrama de flujo del programa que se usó para construir el trímico. .	57
3.7. Imagen del servidor PDB2PQR donde se sube la estructura promedio.	58
4.1. Secuencias de las proteínas CusC de <i>A. ferrooxidans</i> y <i>E. coli</i> en formato FASTA	61
4.2. Estructuras secundarias de las proteínas CusC. Foto obtenida con Chimera 1.6.1	62
4.3. Gráficos de energía pertenecientes a la proteína CusC de <i>Escherichia Coli</i> . Gráfico obtenido con xmgrace	65
4.4. Gráficos de energía pertenecientes a la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráfico obtenido con xmgrace	65
4.5. Gráficos pertenecientes a la proteína CusC de <i>Escherichia coli</i> . Gráficos obtenidos con xmgrace	66
4.6. Gráficos pertenecientes a la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráficos obtenidos con xmgrace.	68
4.7. Gráficos de los puentes de hidrógeno en ambas proteínas CusC. Gráficos obtenidos con xmgrace.	68
4.8. Estructuras promedio de las proteínas CusC. Foto obtenida con Chimera 1.6.1	69
4.9. Ploteo de Ramachandran de la proteína CusC de <i>Escherichia coli</i> . Gráfico obtenido con Molprobity	70
4.10. Ploteo de Ramachandran de la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráfico obtenido con Molprobity	73

4.11. Trímero de la proteína CusC de <i>Escherichia Coli</i> . Gráfico obtenido con Chimera 1.6.0	75
4.12. Trímero de la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráfico obtenido con Chimera 1.6.0	75
4.13. Gráficos de energía pertenecientes al trímero de la proteína CusC de <i>Escherichia Coli</i> . Gráfico obtenido con xmgrace.	77
4.14. Gráficos de energía pertenecientes al trímero de la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráfico obtenido con xmgrace.	77
4.15. Gráficos pertenecientes al trímero de la proteína CusC de <i>Escherichia Coli</i> Gráficos obtenidos con xmgrace.	78
4.16. Gráficos pertenecientes a la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráficos obtenidos con xmgrace.	79
4.17. Gráficos de los puentes de hidrógeno. Gráficos obtenidos con xmgrace.	80
4.18. Estructura promedio del trímero de la proteína CusC de <i>Escherichia coli</i> . Gráficos obtenidos con Chimera 1.6.1.	81
4.19. Estructura promedio del trímero de la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráficos obtenidos con Chimera 1.6.1.	82
4.20. Ploteo de Ramachandran de la proteína CusC de <i>Escherichia coli</i> . Gráfico obtenido con Molprobit	83
4.21. Ploteo de Ramachandran de la proteína CusC de <i>Acidithiobacillus ferrooxidans</i> . Gráfico obtenido con Molprobit	84
4.22. Representación de la superficie del potencial electrostático de la proteína CusC perteneciente a <i>Escherichia coli</i> en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1	86

4.23. Representación de la superficie del potencial electrostático de la proteína CusC perteneciente a <i>Acidithiobacillus ferrooxidans</i> en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1	86
4.24. Representación de la superficie del potencial electrostático del trímero perteneciente a <i>Escherichia coli</i> en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1	88
4.25. Representación de la superficie del potencial electrostático del trímero perteneciente a <i>Acidithiobacillus ferrooxidans</i> en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1	89
A.1. Archivo minima.mpd con integrador steepest descent	104
A.2. Archivo minima.mpd con integrador l-bfgs	105
A.3. Archivo nvt.mpd con integrador MD	107
C.1. Resultados de la temperatura promedio a las que se trabajó las cuatro proteínas	109
C.2. Resultados de la presión promedio a las que se trabajó las cuatro proteínas	109
D.1. Primera parte del programa	111
D.2. Segunda parte del programa	112

Índice de Tablas

2.1. Resistencia al cobre en microorganismos que viven en ambientes bio- mineros.	10
2.2. Proteínas de la familia de factor de membrana externa (OMF).	23
4.1. Resultados de las Energías	90
4.2. Resultados Parámetros Termodinámicos	92
A.1. Relaciones Termodinámicas	106

Glosario

1. ATP. Adenosin-trifosfato.
2. MIC. Minimal Inhibitory Concentration.
3. AF. Acidithiobacillus ferrooxidans.
4. EC. Escherichia coli.
5. TIGR. The Institute for Genomic Research.
6. NCBI. National Center for Biotechnology Information.
7. PDB. Protein Data Bank.
8. ECOR. E. coli Reference Collection.
9. MLEE. Multi-Locus Enzyme Electrophoresis.
10. STM. Segmentos transmembranales.
11. CDF. Cation Diffusion Facilitator.
12. MFP. Membrane Fusion Protein.
13. RND. Resistance-Nodulation-Division.
14. OMF. Outer Membrane Factor.
15. CFF93. Consistent Force Field 93.
16. UFF. Universal Force Field.

17. CHARMM. Chemistry at Harvard Molecular Mechanics.
18. GAFF. General AMBER force field.
19. AMBER. Assisted Model Building with Energy Refinement.
20. OPLS. Optimized Potentials for Liquid Simulations.
21. DFP. Davison-Fletcher-Powel.
22. BFGS. Broyden-Fletcher-Goldfarb-Shanno.
23. MS. Murtaugh-Sargent.
24. BLAST. Basic Local Alignment Search Tool.
25. PSI. Position-Specific Iterated.
26. MSA. Multiple Sequence Alignment.
27. MM. Mecánica Molecular.
28. DM. Dinámica Molecular.
29. BG. Born Generalizado.
30. RMSD. Root Mean Square Deviation.
31. RMSF. Root Mean Square Fluctuation.
32. Rg. Radius of Gyration.

Resumen

En el presente trabajo de investigación se buscó simular el proceso de trimerización de las proteínas CusC pertenecientes a las bacterias *Acidithiobacillus ferrooxidans* y *Escherichia coli*, para lo que se diseñó un programa que permite construir las estructuras de ambos trímeros, nuestras simulaciones fueron realizadas con la finalidad de obtener un mejor entendimiento de las propiedades termodinámicas y electrostáticas.

Para lo cual, se realizaron simulaciones de dinámica molecular tanto para los monómeros como los trímeros de ambas bacterias. Una vez obtenidas las proteínas estables, los cambios estructurales fueron evaluados mediante los diagramas de Ramachandran. Luego se calcularon los valores de las energías totales promedio; a partir de los cuales se obtuvo la energía interna (ΔU), entropía (ΔS), entalpía (ΔH) y energía libre de Gibbs (ΔG) para los dos procesos de trimerización, comprobando la espontaneidad de estos procesos, además los parámetros termodinámicos antes mencionados indican que las reacciones son exotérmicas y tienden al orden. Siendo el trímero de *Acidithiobacillus ferrooxidans* el más estable termodinámicamente y por ende el más factible en la formación de los trímeros.

Abstract

The present investigation sought to simulate the process of trimerization of CusC proteins belonging to the bacteria *Acidithiobacillus ferrooxidans* and *Escherichia coli*, for which a program was designed to build the structures of both trimers, these simulations were in order to gain a better understanding of the thermodynamic and electrostatic properties.

To do this, Molecular Dynamics simulations were performed for monomers as the trimers of the bacteria. Once obtained the stable proteins, the structural changes were evaluated through the Ramachandran diagram. Then the average total energy values were calculated; from whom were obtained the internal energy, entropy, enthalpy, and free energy of Gibbs for the two processes of trimerization, proving the spontaneity of these processes, the thermodynamic parameters mentioned above also indicates that reactions were exothermic and tend to order. Being the trimer of *Acidithiobacillus ferrooxidans* the most stable thermodynamically and therefore more feasible in the formation of trimers.

Capítulo 1

Introducción

Después del oro y la plata, el cobre es el metal industrial más antiguo usado por el hombre. A partir de la segunda mitad del siglo XIX, hasta hoy, el cobre de alta pureza ha aumentado su demanda con el surgimiento de nuevas tecnologías que requieren de; alta capacidad de transmisión de ondas electromagnéticas como el teléfono, la radio, los computadores y los dispositivos móviles,^{1,2} también por sus propiedades de resistencia a la corrosión y conductividad térmica,³ es utilizado en las tuberías de agua, en las casas y edificios, en motores eléctricos, cables de conducción eléctrica, aparatos electrodomésticos, incluso en el cuidado de la salud como bactericida.⁴

El 90 % de las reservas mundiales están concentradas en cinco áreas: la cuenca de las Montañas Rocosas en Estados Unidos, la prolongación de la cadena en Canadá, los Andes de Perú y Chile, Indonesia y Zambia en la cuenca centro/sur de África.⁵ Alrededor del 80 % mundial de cobre se encuentra como minerales sulfurados, que frecuentemente son mezclas compuestas de sulfuros de cobre y hierro, combinados con compuestos de otros elementos. Este mineral no se disuelve fácilmente en soluciones acuosas, por lo que su extracción se da mediante la pirometalurgia.

El otro 20 % se encuentra como minerales oxidados de cobre que se originan en la descomposición y oxidación de los minerales sulfurados. Estos fueron los primeros explotados, y en este caso la extracción se da por hidrometalurgia¹ La producción

mundial de cobre asciende a 12 millones de toneladas al año y las reservas explotables son alrededor de 300 millones de toneladas, lo que se espera que dure solo otros 25 años.⁶

Las tendencias en el consumo global de cobre refinado están progresando en forma alarmante. Anualmente entre los años de 1980 y 2010 el consumo de cobre se incrementó en 10.1 %.⁶ La producción industrial no mantiene el ritmo de consumo de este metal. Por lo que el mundo está comenzando a sentir el impacto de los retos en el abastecimiento, encontrándose la siguiente dificultad actualmente; los yacimientos de alta ley son menos frecuentes, quedando reservas de mineral de baja ley. Este problema ha llevado a la búsqueda de una mejora continua y al surgimiento de nuevas tecnologías de explotación como la biolixiviación; la cual permite tratar minerales oxidados, mixtos y sulfuros secundarios, de manera más económica y con menos efectos ambientales nocivos.²

Por otro lado el cobre también es un elemento esencial requerido para la vida de todos los organismos, por esta razón las células han desarrollado mecanismos para controlar los niveles de cobre libre cuando la concentración de este metal excede los niveles aceptables, estos mecanismos ayudan a los microorganismos a sobrevivir en ambientes adversos. Por ejemplo bacterias neutrófilas pueden crecer en un rango de concentración de cobre entre 1 y 8 mmol/l dependiendo de las especies. Mientras que, acidófilos como *Acidithiobacillus ferrooxidans* o *Sulfolobus metallicus* puede resistir concentraciones de cobre hasta 800 mmol/l y 200 mmol/l, respectivamente.⁷

Por estas razones la lixiviación bacteriana de minerales ha sido reconocida como una contribución para la extracción de metales desde el descubrimiento de *Acidithiobacillus ferrooxidans* en 1947.⁸ A pesar que la biolixiviación se ha estudiado por más de 60 años; todavía no se han reportado mejoramientos reales en la recuperación del cobre, el problema que persiste en este proceso es que todavía no se explota al máximo la capacidad de las bacterias, y es que a pesar que los microorganismos lixiviantes se desarrollan naturalmente en soluciones ácidas con alto contenido de metales, todavía

no se tiene claro los mecanismos moleculares por los cuales atacan y solubilizan los minerales, sobretodo el rol de las proteínas encargadas de los mecanismos de resistencia y transporte de cobre.

Dentro de los avances en la secuencia genómica completa de la *Acidithiobacillus ferrooxidans*, se ha informado de proteínas que son expresadas en su superficie cuando se expone a altas concentraciones de cobre.⁹ Varias de estas proteínas periplásmicas se desconoce su función y alrededor del 26.1 % son proteínas sin homólogos en la base de datos, muchas de las cuales pueden ser características de este microorganismo.¹⁰ Una de las que se ha identificado es la proteína CusC, que es un factor de membrana externa, y aunque se han identificado más proteínas de este tipo en diferentes bacterias gram-negativas, como *Escherichia coli*, ninguna es usada en biominería.¹¹

Estos hallazgos no sólo contribuyen a la comprensión de la fisiología de *A. ferrooxidans*, también serán importantes para la anotación genómica de las nuevas proteínas periplásmicas de este y otras bacterias extremófilas presentes en el ambiente biominero. Es así que una de las vías descritas para resistencia al cobre que ha sido postulada es que pueden bombear el metal a partir del citoplasma y el periplasma al espacio extracelular por sistemas de proteínas transportadoras de resistencia-nodulación-división celular, como el sistema Cus de *Escherichia coli* que es codificado por el operón cusCFBA. Por otro lado *A. ferrooxidans* contiene un operón formado por los genes cusCBA¹² que dan lugar a las proteínas CusA, CusB y CusC. Estas 3 proteínas formarían un complejo que bombea cobre fuera del periplasma manejado por un gradiente de protones, siendo CusC la proteína anclada a la membrana externa probablemente formando un canal que expulsa el cobre al exterior. Un gran avance que se tiene sobre esta proteína en *Escherichia coli* es que forma un trímero.

Objetivos

1.0.1. Objetivo General

Evaluar las propiedades termodinámicas de la proteína CusC de *Acidithiobacillus ferrooxidans* y *Escherichia Coli* en su trímero estructural, relacionada con el transporte de cobre.

1.0.2. Objetivos Específicos

1. Construir la proteína CusC de la *Acidithiobacillus ferrooxidans* y *Escherichia Coli*.
2. Obtener la estabilidad energética y estructural de las estructuras proteicas CusC mediante una simulación de dinámica molecular, para lo cual se empleará el software GROMACS y un ensamble canónico.
3. Modelar la trimerización de las proteínas CusC mediante parámetros de simetría.
4. Analizar las propiedades estructurales involucradas en el Proceso de trimerización.
5. Evaluar las propiedades termodinámicas de energía interna, entropía, entalpía y energía libre de Gibbs de los trímeros a partir de los datos de energía total obtenidos durante las simulaciones de dinámica molecular y su factibilidad en la migración de iones positivos.

Hipótesis

Mediante el uso de técnicas bioinformáticas, es probable que podamos evaluar las propiedades termodinámicas del proceso de trimerización de la proteína CusC perteneciente a la *Acidithiobacillus ferrooxidans* comparada con el trímero CusC de *Escherichia Coli*.



Capítulo 2

Marco Teórico

Para poder entender adecuadamente el desarrollo de esta investigación, es importante hacer una descripción de los métodos aplicados por la bioinformática, para resolver los diversos problemas a nivel molecular, así como de la importancia de las moléculas que son la materia de nuestra investigación. Para ello, en el presente capítulo, daremos una breve pero completa revisión a los conceptos sobre las metaloproteínas CusC de *Acidithiobacillus ferrooxidans* y de *Escherichia coli*, así como a la homeostasis del cobre. También daremos una introducción a los diferentes conceptos teóricos a ser utilizados para el desarrollo de cálculos de mecánica molecular y dinámica molecular.

2.1. Biominería

La biominería es la rama de la biotecnología aplicada a la industria minero-metalúrgica que busca la solución de sus diferentes problemas ¹³ desde los procesos que intervienen en la explotación de los minerales abarcando la concentración de las especies de interés (a través de la bioflotación), la recuperación de los elementos presentes en ellas o facilitando su recuperación (biolixiviación y biooxidación respectivamente) hasta su acción en tareas de remediación ambiental (biorremediación).¹⁴

Estos procesos microbianos están ganando un creciente interés en los últimos años.

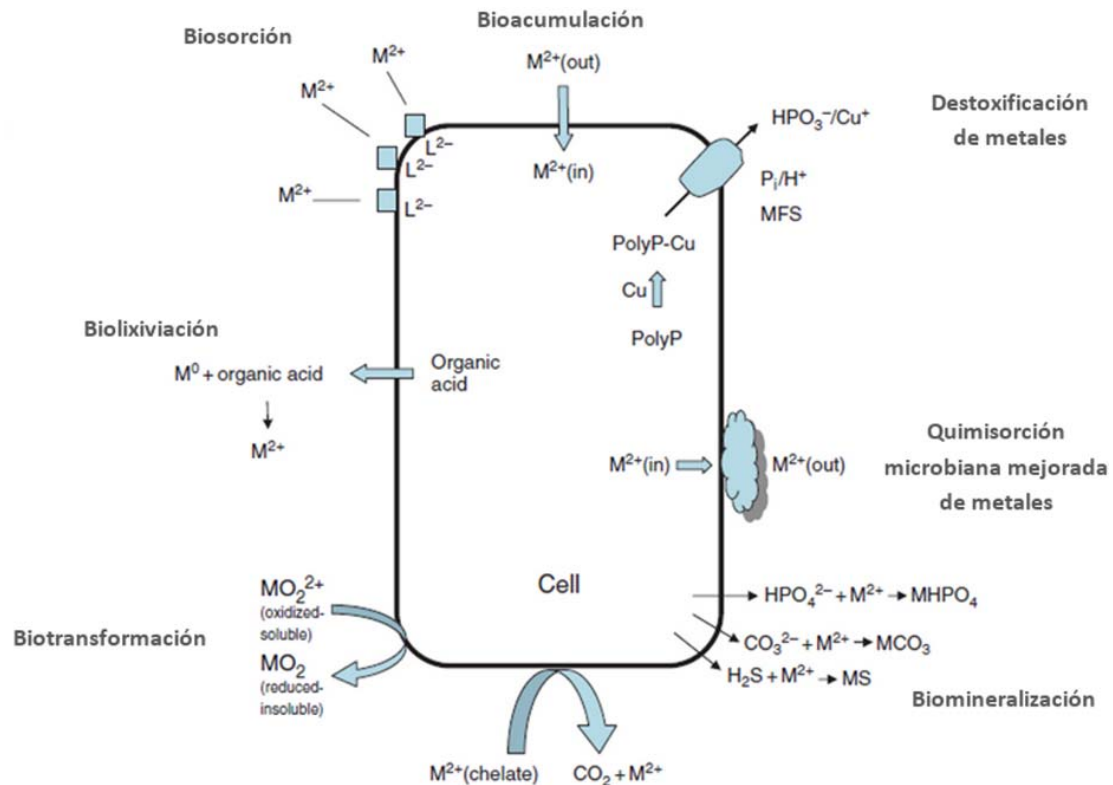


Figura 2.1: Interacciones más comunes bacteria-metal.⁷

Las aplicaciones potenciales actuales incluyen la extracción de oro, cobre y otros metales pesados, desulfuración de carbón y petróleo, recuperación terciaria de petróleo y biosorción de iones metálicos. En la actualidad, la lixiviación bacteriana de cobre y la biooxidación de oro son procesos establecidos a gran escala que se llevan a cabo en pilas o tanques reactores.¹⁵ Principalmente nos enfocaremos en la recuperación del cobre, por ello a continuación se hará una descripción de la biolixiviación.

2.1.1. Biolixiviación

EL término biolixiviación hace referencia a un proceso en el cual se emplean microorganismos para disolver los minerales, liberando el metal de valor presente. Esta tecnología es aplicada cuando los yacimientos tienen una concentración menor a 0.5% de cobre.¹⁴ Como es el caso de los minerales sulfurados, en los cuales se necesita aumentar la cinética de su disolución. Con esta tecnología se han conseguido reali-

zar explotaciones económicamente más rentables, que por los métodos tradicionales físico-químicos de extracción supondrían pérdidas de gran cuantía. Actualmente la biolixiviación representa el 10 % de la producción de cobre en todo el mundo.¹⁶

Las bacterias que intervienen en los procesos de lixiviación son conocidas como quimiolitotóxicas ya que usan los minerales como combustible. Su oxidación genera electrones para obtener ATP y el carbono es obtenido de la fijación del CO₂ del aire.⁷

También se les llama microorganismos sulfuro y ferro-oxidantes, porque oxidan el mineral para producir el ión férrico y ácido sulfúrico, necesarios para las reacciones de biolixiviación. El ión férrico, es un agente fuertemente oxidante, que permite oxidar los minerales de sulfuro de cobre a sulfato de cobre que es soluble. Así realizan un trabajo útil, liberando metales y calor, sin necesitar una aplicación externa de energía al proceso, en este tipo de procesos, no son necesarias altas energías de activación.¹⁷

Por otro lado existen dos mecanismos dominantes, que se consideran estar involucrados en la biolixiviación. Estos son los mecanismos directos e indirectos. En el mecanismo directo, la lixiviación ocurre por el ataque enzimático de las bacterias sobre los componentes del mineral que son susceptibles de oxidar. Los electrones liberados son transportados a través del sistema proteico de la membrana celular y de ahí a su aceptor terminal. Este mecanismo supone la adhesión de los microorganismos en el sólido. El sulfuro de cobre, CuS, es uno de los minerales que pueden ser convertidos en una forma soluble del metal, que en este caso es cobre. Mediante una reacción de oxidación, las bacterias extraen los electrones y disuelven el sulfuro de cobre (CuS), que es sólido, obteniendo una solución de sulfato de cobre (CuSO₄) a partir de la cual se puede recuperar el cobre como metal.

También las mismas especies bacterianas pueden oxidar azufre elemental hasta ácido sulfúrico, según la reacción siguiente:

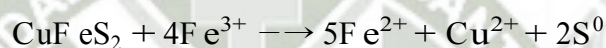


En el mecanismo indirecto, no ocurre un ataque frontal de la bacteria sobre la estructura atómica del mineral, en su lugar los minerales son lixiviados solo por iones férricos, y el rol de las bacterias es oxidar los iones ferrosos. De acuerdo a este mecanismo, el Fe(II) que aparece en las soluciones debido a la disolución de los sulfuros por acción bacteriana directa, puede ser oxidado por las bacterias hierro-oxidantes, según la siguiente reacción:^{18,19}



Finalmente, el Fe(III) producido puede actuar como fuerte oxidante de los sulfuros según la siguiente reacción donde se toma como ejemplo la calcopirita:²⁰

Calcopirita:



Ya que en estas reacciones se produce ácido sulfúrico, los microorganismos viven en condiciones extremas, a pH ácido (usualmente de 1-3) y altas concentraciones de metales.²¹ Ellos poseen mecanismos de resistencia al metal o desintoxicación.⁷ Estas bacterias se aíslan, generalmente, de lugares donde ocurre la oxidación natural de minerales, como minas, aguas de manantiales ricas en azufre y depósitos de carbón.²² En la Tabla 2.1 se tienen un listado de los microorganismos que se han logrado aislar de estos lugares y la concentración de cobre que resisten.

Uno de sus principales exponentes es la bacteria *Acidithiobacillus ferrooxidans*, aislada por primera vez desde las aguas de una mina de carbón, cuyo descubrimiento se dio a conocer en 1947. Así fue como se encontró la primera bacteria identificada capaz de lixiviar el cobre. La *Acidithiobacillus ferrooxidans*, ha sido la bacteria más estudiada para biolixiviación y por consiguiente de la que existe mayor información.²¹

Tabla 2.1: Resistencia al cobre en microorganismos que viven en ambientes biomine-
ros.

23

Microorganismo	Cobre MIC (mM)	T ^o óptima (°C) ^c	Capacidad Oxidativa ^d		Referencia ^e
			Fe ²⁺	S ⁰	
<i>Acidithiobacillus ferrooxidans</i> ^{b*}	800	30-35	+	+	Harvey and Crundwell, 1996
<i>Acidimicrobium ferrooxidans</i> ^b aislado N39-30-03	≥ 786	45-50	+	+	Watkin et al., 2008
<i>Sulfobacillus thermosulfidooxidans</i> aislado N19-45-01 ^b	786	45-50	+	+	Watkin et al., 2008
<i>Leptospirillum ferrooxidans</i> ^{b**}	393	28-30	+	-	Hallmann et al., 1993
<i>Ferroplasma acidarmanus</i> ^{a**}	312	42	+	-	Baker-Austin et al., 2005
<i>Sulfobacillus thermosulfidooxidans</i> DMS 9293 ^{T b}	300	45-50	+	+	Watkin et al., 2008
<i>Sulfolobus metallicus</i> ^a	200	65	+	+	Remonsellez et al., 2006
<i>Sulfobacillus montserratensis</i> ^b	100	37	+	+	Shippers, 2007
<i>Acidithiobacillus caldus</i> DSM8584 ^{T b**}	24	45	-	+	Watkin et al., 2008
<i>Thiobacillus prosperus</i> ^b	16	33-37	+	+	Shippers, 2007
<i>Metallosphaera sedula</i> ^{a*}	16	75	+	+	Huber et al., 1989
<i>Acidimicrobium ferrooxidans</i> DSM10331 ^{T b}	9,4	45-50	+	+	Watkin et al., 2008
<i>Thiomonas cuprina</i> ^b	7,9	30-36	-	+	Shippers, 2007
<i>Acidithiobacillus ferriphilum</i> ^{b**}	ND	28-30	-	+	Waksman and Joffe, 1992
<i>Leptospirillum ferriphilum</i> ^{b**}	ND	30-37	+	-	Coram and Rawlings, 2002
<i>Acidianus infernus</i> ^a	ND	90	+	+	Segeer et al., 1986

^aArchaea, ^bBacteria, en las columnas c y d los datos fueron tomados a partir Shippers A.2007, la columna ^ereferencias donde fueron tomados los valores de MIC, * secuencias y genomas disponibles y publicados, ** genomas para estos microorganismos han sido secuenciados y los datos no han sido publicados, MIC: mínima concentración inhibitoria, ND: no determinado.

2.1.2. *Acidithiobacillus ferrooxidans*

Acidithiobacillus ferrooxidans es considerada la bacteria biominera modelo. Las células de *A. ferrooxidans* tienen forma de cilindro, con diámetros de 0,3 a 0,8 micrómetros (micras) y una longitud de 0,9 a 2 micras. La movilidad de la célula se logra por medio de un solo flagelo polar.¹⁹ Las características de esta bacteria son:

- Es gram-negativa, quimiolitotrofa; obtiene su energía de la oxidación de hierro ferroso, azufre elemental, o componentes de azufre parcialmente oxidados.¹⁰
- Autótrofa porque usa el dióxido de carbono como fuente de carbono, también necesita nitrógeno y fósforo como nutrientes para el crecimiento y la síntesis junto con minerales traza como, Mg, Na, Ca y Co.
- Aeróbica; el oxígeno es esencial como aceptor de electrones.¹⁹
- Es mesófila, puede crecer a temperaturas entre 20°C y 40°C, prospera óptimamente a 30°C. También es acidófila creciendo en un rango de pH entre 1 a 4.0, siendo el óptimo a pH 2.

Esta bacteria *A. ferrooxidans* es abundante en ambientes naturales asociados con minerales como piritita, depósitos de carbón, y drenajes ácidos.¹⁶ Las células de *A. ferrooxidans* crecen en forma de una biopelícula, probablemente como una monocapa, esto se ha observado en la superficie de minerales como piritita. Además se ha demostrado que poseen un sistema quimiosensorial que les permite tener la quimiotaxis, es decir, la capacidad para detectar gradientes de sustratos oxidables siendo extraídos de los minerales tales como iones Fe (II)/Fe (III), tiosulfato, y otros.⁷

A. ferrooxidans fue el primer microorganismo biominero cuyo genoma fue secuenciado por TIGR (The Institute for Genomic Research).¹⁰ Esta información ha sido muy útil para hacer búsquedas en todo el genoma de los genes candidatos de importantes vías metabólicas y varias funciones fisiológicas importantes, además, las predicciones para las funciones de muchos genes se puede hacer.⁷ Se han identificado en la *A.*

ferrooxidans ATCC 23270, 138 vías con 658 enzimas catalizadoras, 752 reacciones enzimáticas con 631 compuestos involucrados. Entre los cuales se consideran parte de: ciclo de Calvin, la vía de las pentosas fosfato, la glucólisis, ciclo de Krebs, la síntesis de los precursores, la construcción de bloques y componentes celulares. Esto se hizo con el PGDB BioCyc (versión 11.5 desde agosto de 2007); PGDB se refiere pathway/genome database y BioCyc que también es una base de datos.²⁴

2.1.3. *Escherichia coli*

E. coli es una de las especies bacterianas más minuciosamente estudiadas, y no solamente por sus capacidades patogénicas, sino también como sustrato y modelo de investigaciones metabólicas, genéticas, poblacionales y de diversa índole.²⁵ Forma parte de la familia Enterobacteriaceae. Se trata de bacterias de rápido crecimiento y amplia distribución en el suelo, el agua, vegetales y gran variedad de animales.

E. coli es un bacilo corto Gram–negativo, catalasa positivo, oxidasa–negativo y anaerobio facultativo. La mayoría de las cepas fermentan la lactosa, aunque algunas son fermentadoras lentas de este azúcar y algunas son anaerogénicas. Las cepas de *E. coli* se pueden diferenciar serológicamente en relación a los antígenos somáticos, flagelares y capsulares. Produce reacción positiva de rojo de metilo, y negativa de Vogues-Proskauer. Es inhibida por KCN e incapaces de crecer en medio con citrato como única fuente de carbono y energía, pero sí en caldo acetato. Son H₂S, ureasa y fenilalanina negativos, pero en general son indol positivos y decarboxilan la lisina.²⁶

La colección de referencia de *E. coli* (ECOR por sus siglas en inglés) de 72 cepas de diversos orígenes naturales se cree que representa la diversidad genética de las especies. Esto ha hecho posible la generación de grandes cantidades de datos comparativos de muchas cepas de las especies. Se utilizó inicialmente electroforesis de enzimas multilocus (MLEE) para producir una topología de las cepas ECOR.²⁵

Como sabemos *E. coli* es uno de los microorganismos los cuales son habitantes normales en el colon de mamíferos de sangre caliente, pero la cepa de *E. coli* K-12 es una

cepa debilitada que no suele colonizar el intestino humano. También se ha demostrado que sobreviven muy poco tiempo en el medio ambiente, tiene un historial de uso comercial seguro, y se sabe que no tiene efectos adversos sobre los microorganismos o plantas. A esto se debe su amplio uso como un organismo modelo en la investigación en la genética y la fisiología microbiana, y su uso en aplicaciones industriales, siendo la cepa de *E. coli* K-12 uno de los microorganismos más estudiados.²⁷

Entre las ventajas que presenta esta cepa son: representa el mejor organismo genéticamente comprendido, es fácilmente modificable por muchos métodos genéticos, y está clasificada como vehículo biológicamente seguros para la propagación de la clonación de genes y vectores de expresión en todas las grandes guías nacionales e internacionales sobre la seguridad biológica para el trabajo con la tecnología del ADN recombinante.²⁸

2.2. Homeostasis del cobre

Los microorganismos para llevar a cabo sus funciones metabólicas requieren de la presencia de ciertos iones inorgánicos esenciales, muchos de ellos pertenecen al grupo de los alcalinos y alcalinoterreos como; calcio, magnesio, sodio, potasio, etc., así como algunos metales de transición como: manganeso, zinc, cobre, etc.. Por otro lado, existen otros elementos que son intrínsecamente tóxicos y carecen de actividad biológica como; plomo, mercurio, cadmio.²⁹ Desde el punto de vista fisiológico, estos metales se dividen en tres grupos:

- Grupo I, el cual incluye metales esenciales caracterizados por su baja toxicidad, como el Fe, Mo.
- Grupo II, el cual incluye metales con moderada importancia como elementos traza pero que podrían ser potencialmente tóxicos al aumentar su concentración como; Zn, Ni, Cu, V, Co, W y Cr.

- Grupo III, el cual incluye metales con ningún rol y que son tóxicos como; Hg, Cd, Sb y Pb.³⁰

Asimismo, los iones metálicos se encuentran de manera natural en la biosfera normalmente en niveles de concentración bajos, de tal manera que no representan un factor de toxicidad para los organismos vivos, más bien actúan como constituyentes de las proteínas cumpliendo una variedad de tareas, como; componentes estructurales de las biomoléculas, moléculas de señalización, y cofactores catalíticos en reacciones reversibles de oxidación-reducción y la hidrólisis, y en reajustes estructurales de las moléculas orgánicas.³¹ Sin embargo, en ciertas áreas, como algunos yacimientos minerales o en zonas con una elevada actividad industrial, los niveles de dichos iones metálicos pueden llegar a ser elevados. Pudiendo desencadenar en los microorganismos mecanismos de toxicidad como; interferencia con el transporte y la función de los iones fisiológicos esenciales.²⁹ En esta situación los mecanismos de homeostasis deben asegurarse que la adquisición de iones metálicos sea suficiente y no excesiva.³⁰

A continuación definiremos a la homeostasis como; el mantenimiento de una concentración biodisponible óptima, mediada por el equilibrio de la captación de metal y el tráfico intracelular con los procesos de flujo de salida/ almacenamiento de modo que las necesidades de la célula con el ion metálico se cumplan, es decir, el metal correcto se inserta en la macromolécula correcta en el momento oportuno. Así como una escasez de un ion metálico particular induce una respuesta de estrés que puede conducir a la reprogramación del metabolismo celular para minimizar las consecuencias del agotamiento de un ion metálico en particular. La homeostasis del metal es gobernada por la formación de complejos específicos de coordinación proteína-metal, que son usados para efectuar el consumo, expulsión, tráfico intracelular dentro de compartimientos y almacenamiento.³¹

Por la naturaleza minera de nuestro país, y la gran cantidad de minas de cobre, nos enfocaremos en él. Este es un micronutriente esencial para los organismos, ya que es requerido por proteínas involucradas en una variedad de procesos biológicos como

el transporte de electrones y transformaciones redox; su deficiencia puede causar la alteración de sus actividades³². El cobre es esencial pero también puede ser tóxico incluso a bajas concentraciones. Debido a esta naturaleza dual, los mecanismos para la homeostasis del cobre se pueden encontrar en todos los reinos. Tanto en procariotas como eucariotes, en éste último desde las levaduras hasta los humanos.³³

La homeostasis del cobre es un requisito para la evolución del metabolismo aeróbico, porque éste es un metal de transición redox-activo. El cobre parece ser el responsable de la generación intracelular de superóxido y otras especies reactivas del oxígeno. Además, en condiciones anaeróbicas el cobre parece desplazarse desde el Cu^{+2} a Cu^{+1} y se convierte en mucho más tóxico, posiblemente debido a Cu^{+1} puede difundir a través de la membrana citoplasmática. Por lo tanto, las concentraciones intracelulares de cobre deben ser regulados dentro de límites muy estrechos,³⁴ ya que cuando la concentración de éste excede los niveles aceptables puede dañar las membranas celulares, también la estructura de ácidos nucleicos así como alterar la especificidad de enzimas y otras funciones celulares en general. Debido a esto, los mecanismos de resistencia son activados para prevenir su acumulación.^{23,29} Sin embargo todos estos mecanismos involucrados en el transporte y homeostasis de cobre en bacterias solo son parcialmente comprendidos.

Los sistemas de absorción de cobre deben ser específicos, a fin de asegurar un suministro adecuado de este en presencia de niveles variables de cationes relacionados, tales como el níquel o zinc. Una vez dentro de las células, se necesitan proteínas adicionales para el secuestro y el tráfico de cobre. El Cu^{+2} no unido intracelularmente puede resultar dañino. Sin embargo, se ha demostrado que los sistemas celulares de flujo de salida remueven los niveles excesivos de cationes y protegen el medio intracelular.³⁴

2.3. Mecanismos de Resistencia /Tolerancia a metales

Las bacterias han desarrollado diversos mecanismos de resistencia para tolerar los efectos nocivos de los metales tóxicos. Entre ellos se encuentran principalmente los que involucran:

- a Componentes celulares que capturan a los iones, neutralizando su toxicidad.
- b Enzimas que modifican el estado redox de los metales o metaloides, convirtiéndolos en formas menos tóxicas.
- c Transportadores de la membrana que expulsan las especies nocivas del citoplasma celular.³⁵ Hay tres principales mecanismos de expulsión que han sido descritos: bomba de flujo de salida CBA, familia de facilitadores de difusión de cationes (CDF), y ATPasas tipo P. Estos sistemas pueden o no estar vinculada al ATP, y son muy específicos para el catión o anión transportados.³⁰

Una alternativa importante a la contribución de la resistencia a metales en acidófilos, que son los microorganismos de interés industrial y en los que está enfocada esta tesis, parece ser el potencial de membrana interior positivo (el cual es opuesto al interior negativo de neutrófilos). Esto parece una adaptación esencial para generar y mantener un gradiente amplio de pH (ΔpH), el cual puede alcanzar hasta 4 unidades de pH entre el exterior y su citoplasma. El ΔpH genera la fuerza motriz de protones la cual es una característica energética esencial para la supervivencia de acidófilos. El gran ΔpH podría también mejorar la resistencia a la acumulación de cationes y también se cree que podría disminuir el requerimiento de energía para mantener el gradiente, dado que los cationes serían sacados en contra de su gradiente de concentración, pero a lo largo de su gradiente electrogénico.

Muy pocos estudios relacionados a los mecanismos de resistencia al cobre en microorganismos biomineros se han llevado a cabo. En lo que respecta a la principal bacteria biominera *A. ferrooxidans*, se ha informado de proteínas que se expresan en

su superficie cuando se expone al cobre. Además, cuando se expone a este metal, *A. ferrooxidans* pierde estructuras extracromosómicas, lo que sugiere que si las bacterias poseen los genes que codifican para proteínas implicadas en la resistencia al cobre, éstas estarían presentes en su genoma.^{12,23,36}

A continuación se hará una breve descripción de los mecanismos de transporte, tocando con más detalle la bomba de flujo de salida CBA que es la que nos interesa.

2.3.1. Facilitadores de difusión de cationes (CDF)

Son proteínas que se encuentran distribuidas en los tres dominios de la vida (Bacteria, Archaea y Eucariota). Generalmente transportan zinc pero también pueden expulsar otros cationes como Cadmio, Cobalto, Níquel e incluso Hierro. Las proteínas CDF son polipéptidos de tamaño mediano (menos de 400 residuos) que poseen seis segmentos transmembranales (STM), funcionan como homodímeros en la membrana interna, y expulsan los sustratos al espacio periplásmico. El transporte de los iones se asocia con un proceso quimiosmótico que involucra el intercambio con protones en el cual participan residuos de histidina, aspartato y glutamato.

2.3.2. ATPasas tipo P

Estas proteínas constituyen una superfamilia de transportadores de metales que son energizadas por la hidrólisis del ATP. Como las proteínas CDF, las ATPasas tipo P (llamadas así porque el γ -fosfato liberado en la hidrólisis del ATP se une en forma covalente a la enzima) están ampliamente distribuidas en los tres dominios de la vida y sus sustratos son iones como H^+ , Na^+ , K^+ , Mg^{2+} , Ca^{2+} , Cu^+ , Ag^+ , Zn^{2+} y Cd^{2+} . Este tipo de ATPasas se localizan en la membrana interna y pueden transportar iones hacia el interior celular, comúnmente iones fisiológicos como Mg^{2+} , o bien funcionar como sistemas de expulsión, removiendo metales tóxicos al espacio periplásmico. Al igual que los transportadores CDF, estas ATPasas deben asociar su función con proteínas de

la membrana externa (por ejemplo, las porinas) para lograr expulsar el sustrato tóxico hasta el exterior celular.³⁵ En bacterias gram-negativas como *Enterococcus hirae* la ATPasa tipo P conformada por las proteínas CopA y CopB encargadas de la expulsión de cobre es la más estudiada, este sistema recientemente ha sido caracterizado en *A. ferrooxidans*.²⁹

2.3.3. Bomba de flujo de salida CBA

La bomba de flujo de salida denominada también sistema CBA es exclusivo de bacterias gram-negativas³⁴. Este enorme complejo proteico funciona como una eficiente bomba de expulsión que transporta los iones tóxicos como iones metálicos, xenobióticos o drogas desde el citoplasma hasta el exterior de la célula bacteriana (a diferencia de los otros sistemas de expulsión que sólo translocan su sustrato al espacio periplásmico). Como ocurre con las proteínas CDF, los transportadores CBA funcionan como antiports quimiosmóticos de cationes y protones.^{30,35}

Estos sistemas quimiosmóticos que intercambian ion/protón están compuestos por 3 polipéptidos:

- Una proteína de membrana interna (subunidad A) de más de 1000 residuos de aminoácidos.
- Una proteína de membrana externa (subunidad C).
- Una proteína de unión (subunidad B) conectando las 2 subunidades en el espacio periplásmico, también llamada proteína de fusión de membrana (MFP)

La subunidad A pertenece a la superfamilia RND, un grupo de proteínas que transportan metales pesados, solutos y factores de nodulación en bacterias y arqueas, así como esteroides en Eucariotas.³⁰ Las proteínas RND son proteínas integrales de membranas y el componente central de este tipo de sistemas, actuando como transportadores secundarios, utilizando la energía del gradiente de concentración de la membrana

citoplasmática, a través del antiport de protones (a la vez que sale el sustrato entra un protón). Estas proteínas confieren especificidad al sustrato.^{37,38} Por ejemplo los miembros de la familia de transportadores RND relacionados con el flujo de salida de metales pesados son muy específicos para sus sustratos y poseen la habilidad de diferenciar entre iones monovalentes y divalentes.³⁸

La subunidad B corresponde al grupo de las proteínas MFP que se agrupan en la familia de proteínas de flujo de salida periplásmico o proteínas adaptadoras periplásmicas. Mientras que la subunidad C son proteínas factor de membrana externo (OMF) que abarcan el espacio periplásmico. Las proteínas OMF a menudo son intercambiables ya que los miembros de la familia OMF podrían funcionalmente sustituir uno por el otro.³⁴ Por el contrario, generalmente los genes que codifican para los componentes RND y MFP están transcripcionalmente acoplados, y ocasionalmente asociado con un gen que codifica para la OMF.³⁷

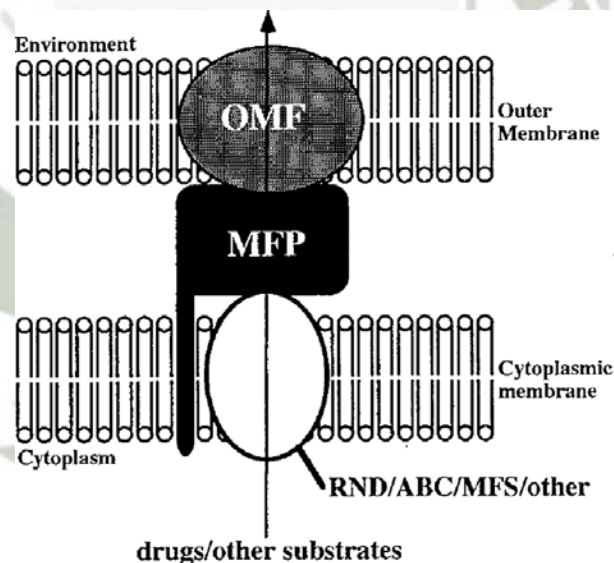


Figura 2.2: Estructura propuesta del complejo de proteico que incluye: (1) un transportador de membrana citoplasmática del tipo RND (blanco), (2) una proteína de fusión de membrana (MFP; negro) y (3) un factor de membrana externo (OMF, sombreado).¹¹

Durante el transporte, el sustrato inicialmente unido a la proteína RND se transfiere directamente al componente OMF para exportarlo al exterior. La proteína MFP sirve

como un enlazador para unir la proteína RND a la proteína OMF, que está anclada en la membrana externa y se proyecta a través del periplasma, formando un "tubo molecular".

Las bacterias gram-negativas típicamente codifican más de un sistema de este tipo para sustratos específicos, los cuales son requeridos para sobrevivir en sus nichos ecológicos.³⁷ Los miembros mejor estudiados de este tipo de transporte han sido los implicados en resistencia a múltiples fármacos (por ejemplo, AcrAB-TolC de *E. coli*). Otros microorganismos donde se han encontrado este tipo de sistema son en *Pseudomonas*, *Ralstonia*, *Synechococcus* y *Salmonella*.³⁴ En cuanto a los sistemas de resistencia a metales se cree que tienen muchas similitudes en la estructura y la función de estos sistemas.³⁹ Por ejemplo en la actualidad, el sistema Czc de *Ralstonia metallidurans* es el transportador de metales mejor caracterizado de este tipo, este microorganismo muestra resistencia al Zinc, Cadmio, y Cobalto.³⁴ También se tiene al sistema Cus de *E. coli* relacionado con la resistencia del cobre, y recientemente se han encontrado tres genes relacionados con este sistema en *A. ferrooxidans*, los cuales son: cusAaf, cusBaf, and cusCaf.^{12,23} A continuación será descrito el sistema Cus.

Sistema Cus

Se denomina Cus por Cu-sensing o en español detección de cobre. Este complejo de transporte, bombea el Cu desde el citoplasma a través de la envoltura celular al espacio extracelular usando un gradiente de protones que cruzan la membrana interna, como fuente de energía.^{33,38,39}

Este sistema está compuesto por las proteínas: CusA que pertenece a las proteínas RND-(Resistencia, Nodulación, División celular) antiport de H⁺/Cu(mecanismo de acoplamiento de transporte de dos compuestos a través de una membrana en direcciones opuestas), CusB que pertenece a la familia de proteínas de fusión de membrana (MFP), miembros de la familia de proteínas de fusión de membrana normalmente están anclados en la membrana interna con un largo dominio de extensión y CusC que es una

proteína factor de membrana externo (OMF).^{11,40}

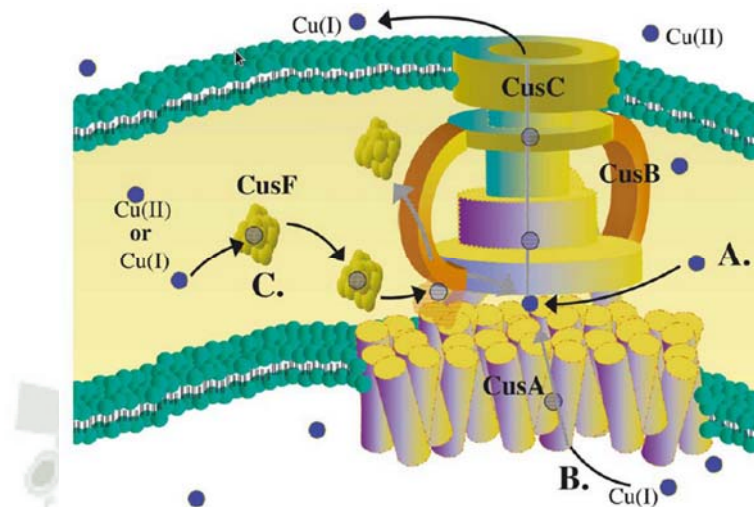


Figura 2.3: Modelo funcional del sistema Cus. Las tres partes de este complejo consisten en la bomba de la membrana interna CusA, la proteína periplásmica CusB y la proteína de membrana externa CusC las cuales forman un canal al espacio periplásmico. La entrada del cobre puede ocurrir a partir del periplasma (A), a partir del citoplasma (B) o por la vía de la chaperona CusF (C).³⁴

El sistema Cus, para el caso de *E. coli*, está compuesto por dos operones, *cusRS* y *cusCFBA*.⁴¹ Las proteínas codificadas CusR y CusS activan la expresión del operón *cusCFBA* en respuesta a concentraciones elevadas de cobre. El sistema *cusCFBA* también puede ser activado por Plata, aunque en concentraciones más altas que las observadas para el cobre.³³ En este sistema aparte de las proteínas CusC, CusB y CusA se codifica también la proteína CusF, la cual es una proteína periplásmica de unión al cobre. Se cree que CusF cumple la función de chaperona transportando el cobre al complejo *cusCBA* y así facilita la desintoxicación de cobre del periplasma.^{34,42}

En lo que respecta a *A. ferrooxidans* una búsqueda bioinformática en su genoma mostró la presencia de al menos 10 genes que están posiblemente relacionados con la homeostasis de Cu. De los cuales tres están relacionados con el sistema de la familia RND involucrada en la expulsión del Cu fuera de la célula (*cusAaf*, *cusBaf*, y *cusCaf*), estos tres genes están contenidos en un operón.^{12,23} *A. ferrooxidans* también posee un

ORF que codifica para una proteína con 25 % de identidad a la CusF de *E. coli* pero con diferente organización genómica ya que se encuentra distancialmente localizada de *afcusCBA*.²³

■ Proteína CusC

Hasta la fecha, todas las proteínas, factor membrana externa (OMF) que se han caracterizado estructuralmente porque son β -barriles de 8 a 22 hebras, que se insertan en la membrana y además son β -cadenas.^{43,44} Formando un trímero a través del cual los solutos se bombean fuera del citoplasma o membrana citoplasmática. La membrana externa de β -barril se ensambla a partir de los tres monómeros con cada uno aportando cuatro cadenas- β . Los miembros de la familia OMF se encuentran en la mayoría de las clases de proteobacterias, en las cianobacterias, espiroquetas, y en las especies de *Deinococcus*. Estas proteínas tienen entre 347 a 541 aminoácidos de longitud.

A pesar de la simplicidad estructural de estas proteínas, que realizan una plétora de funciones biológicas en la membrana externa: OMFs se han identificado como adhesinas, proteínas arquitectónicas, poros pasivos de difusión, receptores de sideróforos, canales de flujo de salida, poros de translocación de proteínas y enzimas, por ejemplo lipasas, proteasas y palmitoil transferasas.⁴⁴

CusC es una proteína factor de membrana externo (OMF) que puede asociarse con una porina en la membrana externa o puede funcionar como un canal a través de esta membrana, como ha sido postulado para otros miembros de la familia OMF.⁴⁵

Tabla 2.2: Proteínas de la familia de factor de membrana externa (OMF).

Proteína OMF	Función	Fuente	#AA ^b
TolC (Eco)	Exportación de drogas	Escherichia coli	495
AprF(Pae)	Exportación de proteasa alcalina	Pseudomona aeruginosa	481
PrtF(Ech)	Exportación de proteasas	Erwinia chrysanthemi	462
CnrC(Aeu)	Exportación de Cobalto/Níquel	Alcaligenes eutrophus	481
11 NccC(Axy)	Exportación de Cobalto/Níquel/ Cadmio	Alcaligenes xylooxidans	438
CyaE(Bpe)	Exportación de ciclolisina	Bordetella pertussis	474
Czc(Aeu)	Exportación de cobalto/Cadmio/Zinc	Alcaligenes eutrophus	345
MtrE(Ngo)	Exportación de multidroga	Neisseria gonorrhoeae	467
OprM(Pac)	Exportación de multidroga	Pseudomona aeruginosa	477
YjeP(Eco)	Resistencia al ácido fusárico	Burkholderia cepacia	433
FusA(Bce)	Exportación de multidroga	Burkholderia cepacia	512

#AA^b, número de aminoácidos en la proteína OMF.

2.4. Bioinformática

La bioinformática es la aplicación a problemas biológicos moleculares de la modelización y simulación de sistemas dependientes del tiempo y la posición; corresponde como tal a una disciplina científica que utiliza tecnología de la información para organizar, analizar y distribuir información de Biomoléculas con la finalidad de responder preguntas complejas. Las herramientas bioinformáticas han nacido de la necesidad de trabajar específicamente con una gran cantidad de secuencias de DNA y Proteínas que se encuentran almacenadas en bases de datos; esta información se puede utilizar para buscar las propiedades de una secuencia, comparándola con secuencias semejantes con propiedades conocidas o a través de reglas para predecir propiedades de secuencias nuevas. Y luego de la comparación, la bioinformática, a pasado a utilizar los modelos físicos y matemáticos para describir fenómenos biológicos de índole molecular.

La bioinformática abarca el uso de técnicas y herramientas utilizadas en tres disciplinas separadas; la Biología Molecular (donde se originan los datos a analizar; genética, proteómica, ciencias biomédicas), la Computación (que proporciona el hardware,

las vías de comunicación de los resultados entre investigadores) y el análisis de datos mediante algoritmos (que entrega los programas y resultados a analizar; para lo que se necesita de matemática, física y estadística).^{46,47}

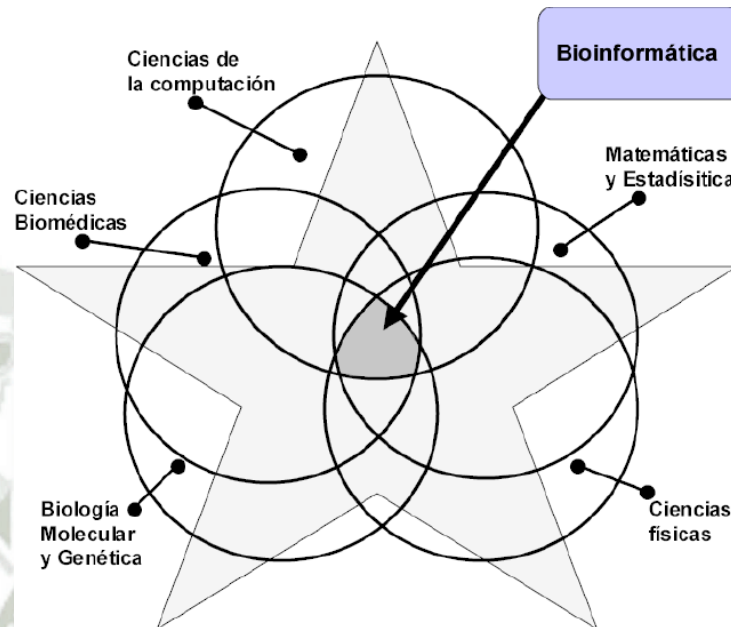


Figura 2.4: La Bioinformática como convergencia multidisciplinaria.

47

Existen varias razones por lo que es adecuada la utilización de software libre en la bioinformática, entre las que tenemos es que los sistemas operativos libres y sus programas asociados son en general más estables y potentes. En misiones de alta exigencia de cómputo esta diferencia se hace crucial (muchos proyectos que requieren procesamiento en paralelo se llevan a cabo con software libre), aplicaciones que no requieren muchos recursos también aumenta el rendimiento considerablemente. Otra gran característica de los sistemas libres es su flexibilidad. En ellos se puede ajustar absolutamente todo el sistema en función de los conocimientos y necesidades del usuario.

Afortunadamente, en bioinformática se han desarrollado formatos de archivo manipulables en cualquier sistema operativo, esto debido a que la gran mayoría del software utilizado en el área se encuentra disponible tanto para plataformas libres como propie-

tarias.

Los dos métodos de simulación computacional más usados actualmente son el de la dinámica molecular, que es de carácter determinista, y el de Montecarlo, que es de carácter probabilístico. Ambos pueden considerarse como métodos para generar diferentes configuraciones de un sistema de partículas. Los sistemas estudiados van desde cientos a miles y últimamente incluso a decenas de millones de átomos. Los aspectos estudiados incluyen propiedades estructurales, termodinámicas, mecánicas y cinéticas. Los ensambles que se usan en la simulación son; microcanónico (NVE), canónico (NVT), isobárico-isotérmico (NPT) y gran canónico (N, μ, E). 48

2.5. Modelado Molecular

En principio, modelado molecular fue el nombre dado a la aplicación de la mecánica molecular para la determinación estructural; hoy en día se incluye en los métodos de la química cuántica, y se caracteriza por el uso de técnica gráfica para representar la estructura y comportamiento de las moléculas.

Los desafíos que se plantean dedicados al modelado molecular consisten en caracterizar y predecir la estructura y estabilidad del sistema, estimar las diferencias energéticas entre distintos estados, y explicar, mecanismos y procesos a nivel atómico y molecular. El propósito último (en cuanto a la aplicación práctica) es tratar de eliminar costosos experimentos en términos económicos y de tiempo.

Para calcular las propiedades de un sistema molecular se requiere de una estructura bien definida. Los cálculos de propiedades requieren a menudo que las estructuras utilizadas tengan la energía más baja posible, aunque a menudo nos interesa estudiar la manera en que una de dichas conformaciones se transforma en otra. Esto es importante porque nos permite entender el porqué de determinadas propiedades macroscópicas.

La mayoría de estudios de modelado molecular implica la elección del modelo para describir las interacciones intra e intermoleculares en el sistema. Los dos más comunes

son las mecánica molecular y la mecánica cuántica. Estos modelos permiten calcular la energía de cualquier arreglo de átomos y moléculas en el sistema, así como determinar cómo esta energía varía al cambiar la posición de las partículas. Otra etapa es el cálculo de la minimización de la energía o una búsqueda conformacional. Finalmente el cálculo debe ser analizado no sólo para calcular propiedades, sino para verificar que se ha realizado apropiadamente.⁴⁹

2.5.1. Mecánica Molecular

La mecánica molecular (MM) es uno de los grandes métodos de cálculo disponibles en la actualidad usados para calcular geometrías moleculares y energías. El método fue desarrollado por Westheimer, Hendrickson, Wilberg, Allinger, Warshel y otros, y es aplicable a estados electrónicos fundamentales. En MM, se ignora el movimiento electrónico y se calcula la energía de un sistema solo como una función de la posición nuclear,^{50,51} considerándose a la molécula como una colección de partículas (normalmente átomos) unidas por fuerzas armónicas o elásticas (un ejemplo gráfico es considerar un modelo de esferas unidas por resortes). Estas fuerzas se pueden describir como funciones de energía potencial, de modo que a cada aspecto estructural (estiramiento del enlace (tensión), ángulos de enlace, diedros, interacciones no enlazantes, flexión fuera del plano, rotación interna (torsión), etc), se le puede asignar un tipo de función.⁵² La suma de estas contribuciones da la energía potencial de la mecánica molecular (llamada también energía estérica). Además al conjunto de estas funciones de energía potencial se le denomina "Campo de Fuerza".

Una forma funcional para un campo de fuerza que se puede utilizar para modelar moléculas individuales o conjuntos de átomos y/o moléculas es:

$$V(r^N) = \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,o})^2 + \sum_{angulos} \frac{k_i}{2} (\theta_i - \theta_{i,o})^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \sum_{i=1}^N \sum_{j=i+1}^N (4\epsilon_{ij} [(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\epsilon_o r_{ij}})$$

(2.1)

$V(r^N)$ denota la energía potencial, la cual es una función de las posiciones (r) de las N partículas (usualmente átomos). Las varias contribuciones son esquemáticamente representadas en la figura de abajo. El primer término de la ecuación (2.1) es la interacción entre pares de átomos enlazados, modelados aquí por el potencial harmónico que da el incremento en energía como la longitud del enlace l_i que se desvía del valor de referencia $l_{i,0}$. El segundo término es la suma de todos los ángulos de valencia en la molécula, de nuevo modelado usando un potencial harmónico (un ángulo de valencia es el ángulo formado entre tres átomos A-B-C en el cual A y C son ambos unidos a B). El tercer término en la ecuación (2.1) es el potencial de torsión que modela como la energía cambia cuando el enlace rota. La cuarta contribución es el término no-enlazante. Esto es calculado entre todos los pares de átomos (i y j) que están en diferentes moléculas o que están en la misma molécula pero separados por al menos tres enlaces. En un simple campo de fuerza el término no-enlazante es usualmente modelado usando el potencial de Coulomb para interacciones electrostáticas y el potencial de Lennard-Jones para las interacciones de van der Waals.

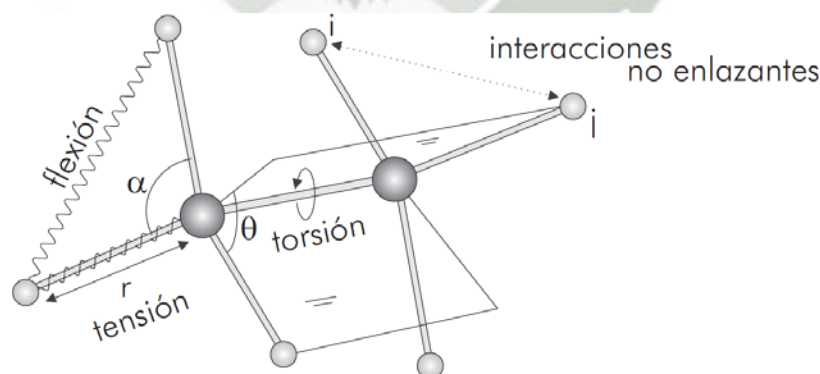


Figura 2.5: Interacciones más importantes que se consideran en mecánica molecular.⁵²

El método MM a veces se denomina método del campo de fuerza. Los campos de fuerza se usan no sólo para cálculos de mecánica molecular de moléculas simples de diferencias de energía, geometrías y frecuencias vibracionales, sino también para simulaciones de dinámica molecular de líquidos y disoluciones, donde se integra la segunda ley de Newton para seguir los movimientos de los átomos con el tiempo, en sistemas que contienen cientos de moléculas.⁵¹

Un atributo clave de un campo de fuerza es la transferibilidad, esto permite a un conjunto de parámetros desarrollarse y probarse en un número relativamente pequeño de casos para ser aplicado en un amplio rango de problemas. Por otra parte, los parámetros desarrollados a partir de datos en moléculas pequeñas pueden ser usados para estudiar moléculas más grandes como polímeros.⁵⁰ Debido a que los cálculos de mecánica molecular son mucho más rápidos que los cálculos mecano- cuánticos, se pueden tratar sistemas de hasta 10 000 átomos, siendo la mayoría de las aplicaciones en compuestos orgánicos, pero están creciendo a compuestos organometálicos y compuestos de coordinación de metales de transición.⁵¹

El punto más delicado en un campo de fuerza es la parametrización. La parametrización consiste en tomar un conjunto de moléculas (de geometría y energía conocidas experimentalmente) e ir modificando las constantes arbitrarias (parámetros) que aparecen en los diferentes potenciales hasta reproducir lo mejor posible la geometría, energía y otros valores de ese conjunto de moléculas tomadas como modelo. Posteriormente, una vez parametrizado el método, se podrán predecir propiedades de otras moléculas. En la parametrización es muy importante la elección adecuada del conjunto de moléculas. Si el conjunto es relativamente pequeño, un valor experimental erróneo puede producir serias distorsiones en los resultados. Esto se puede evitar si se emplean conjuntos muy grandes con lo cual los errores se promedian y disminuye su importancia.

Otra característica importante de los campos de fuerza es que cada uno asigna un número (código) a cada átomo en la molécula según el tipo al que pertenece y en función de su número atómico, carga y ambiente molecular. Cada tipo corresponde

a diferentes características y comportamientos. Por ejemplo, algunos de los tipos de átomos comúnmente usados en un campo de fuerzas en compuestos orgánicos son carbono sp^3 (saturado), carbono sp^2 (doble enlace), carbono sp (triple enlace).

En un cálculo de MM se deben especificar no sólo las coordenadas atómicas iniciales, sino también la conectividad, es decir, determinar a qué átomo o átomos está unido cada átomo de la molécula, de tal manera que se pueda clasificar correctamente por su tipo, con la finalidad de asignarle los parámetros adecuados y con ellos construir en forma apropiada la expresión matemática por resolver. Esto resulta cada vez más sencillo cuando se usa una interfaz gráfica para construir la molécula.⁵³

2.5.2. Campos de Fuerza para Modelamiento Molecular

Actualmente los campos de fuerza pueden ser clasificados en función de diferentes criterios. Atendiendo al origen de los datos utilizados en el proceso de parametrización, se pueden distinguir fundamentalmente dos tipos de campos de fuerza.

1. Campos de Fuerza Experimental: intenta predecir de forma precisa las estructuras moleculares y sus propiedades. Un buen ejemplo son los trabajos de Allinger y colaboradores, que han desarrollado los campos de fuerza; MM1, MM2 (para compuestos orgánicos de tamaño pequeño y medio), MM3 (compuestos orgánicos pequeños, polipéptidos y proteínas) y MM4 (versión mejorada para hidrocarburos),^{51,54} los cuales se han parametrizado a partir de datos experimentales de calores de formación y frecuencias vibracionales. MMX es otro campo de fuerza de este tipo, mezcla de MM2 y MMP1, añadiendo, este último, términos para el cálculo de hidrocarburos aromáticos e introduciendo el concepto de parámetros genéricos, para aquellos casos en los que no se tengan datos para algunos tipos de átomos.
2. Campos de Fuerza Mecanocuánticos: Las constantes de fuerza son obtenidas mediante cálculos ab initio. Un ejemplo es el campo CFF93 (Consistent Force

Field 93).

También podemos clasificar los campos de fuerza según para lo que están diseñados.

1. Generales o Universales: Son campos de fuerza de carácter general. Tal son los casos de los campos de fuerza UFF (Universal Force Field) y DREIDING, o bien de GAFF, son válidos tanto para el estudio de compuestos orgánicos como de compuestos inorgánicos y estructuras formadas con complejos de metales de transición.⁵² Debido a la simplicidad de este tipo de campos de fuerza, no constituyen un método válido para la determinación de las frecuencias vibracionales.⁵⁵
2. Específicos: AMBER, CHARMM y OPLS son los más representativos de esta categoría. La forma funcional de estos campos de fuerza es la más usada para simulación de proteínas debido a que su relación eficiencia-exactitud es suficientemente buena para llevar a cabo simulaciones de sistemas con muchas partículas. Al hacer detallado el modelo, es posible reproducir mayor número de propiedades de las moléculas. También se suelen utilizar en simulaciones de dinámica molecular.
 - AMBER (Assisted Model Building with Energy Refinement): es el más popular de los campos de fuerza diseñado para proteínas y ácidos nucleicos. Los parámetros de este campo fueron ajustados tanto con datos experimental obtenidos a partir de microondas, difracción de neutrón y estudios precisos mecano cuánticos. Estimaciones preliminares de longitud de enlaces y ángulos fueron adoptadas a partir de estudios de difracción de neutrones y microondas, y las cargas atómicas fueron obtenidas por ajuste de cálculos mecano cuánticos ab initio. El potencial fue incluido para la descripción de puentes de Hidrógeno. El campo de fuerza considera efectos del solvente implícito usando distancias dependientes de la constante dieléctrica.⁵⁶ Hay un número de versiones diferentes de Amber algunas de estas son: Amber 2, Amber 3, Amber para sacaridos, Amber 94 y Amber 96.⁵⁷

- OPLS (Optimized Potentials for Liquid Simulations): Es un campo de fuerza desarrollado a principios de 1980, al igual que el AMBER, el campo de fuerza OPLS se diseñó para el estudio de proteínas y ácidos nucleicos. La diferencia fundamental con él radica en la inclusión de parámetros para las interacciones no enlazantes que fueron cuidadosamente desarrollados a partir de simulaciones Monte Carlo en estado líquido de pequeñas moléculas. De esta forma, este campo de fuerza es capaz de describir más adecuadamente las simulaciones donde se incluye explícitamente al disolvente y donde las interacciones no enlazantes son particularmente importantes. Con este proceso de desarrollo, es lógico que la expresión de la energía OPLS sea la misma que la utilizada en el método AMBER con la diferencia de que el término de interacción por puente de hidrógeno ha sido eliminado. Los parámetros y funciones para los términos de tensión, flexión y torsión son los mismos que los desarrollados en el campo de fuerza AMBER.

2.5.3. Optimización geométrica

Cuando empezamos un estudio de modelamiento molecular lo primero que hacemos es generar un modelo de la molécula en la computadora para definir la posición relativa de los átomos en el espacio por un conjunto de coordenadas cartesianas. Una geometría razonable y fiable inicial, determinará la calidad de las siguientes investigaciones. Esta puede ser obtenida a partir de varias fuentes. Los tres métodos básicos para la generación de estructuras moleculares 3D son:

1. Uso de base de datos de cristalografía de rayos-X.
2. Compilación a partir de librería de fragmentos con geometrías estándar.
3. Dibujo simple de 2D de estructuras.

Estos modelos generados en 3D de una molécula no tienen una geometría ideal; por lo tanto, la optimización de la geometría debe ser desarrollada subsecuentemente.⁵⁸

La optimización geométrica por métodos de mecánica molecular tiene como objetivo encontrar un mínimo local de la energía V . Debido a que V posee una expresión analítica, la primera y segunda derivadas de V pueden ser evaluadas analíticamente, lo cual facilita la minimización de la energía. Este proceso predice la geometría de equilibrio del conformero en estudio.

Se debe hacer énfasis en que el valor numérico de la energía de equilibrio del conformero no tiene un significado físico por sí mismo. El nivel cero de V corresponde a la molécula ficticia, en la cual todas las longitudes de enlace y los ángulos tienen su valor de referencia y donde las interacciones torsionales, de Van der Waals y electrostáticas están ausentes. Se han construido muchos programas de mecánica molecular con métodos de búsqueda que localizan muchos conformeros de baja energía.

51

Minimización de la Energía

En modelamiento molecular se busca los puntos mínimos en la superficie de energía. Los arreglos de menor energía de los átomos corresponden a estados estables del sistema, algún movimiento lejos del mínimo da una configuración con una alta energía. Puede haber un gran número de mínimos en la superficie de energías. El mínimo más bajo será conocido como un mínimo de energía global.

La minimización de un modelo es hecho en dos pasos. Primero, la expresión de la energía (una ecuación describiendo la energía del sistema como una función de sus coordenadas) debe ser definida y evaluada para una conformación dada. Luego, la conformación es ajustada para bajar el valor de la expresión de la energía. Un mínimo puede ser encontrado después de un ajuste o puede requerir muchas miles de iteraciones, dependiendo de la naturaleza del algoritmo, la forma de la expresión de la energía, y el tamaño del modelo. Esto quiere decir que la computadora hace pequeños cambios en la posición de todos los átomos y calcula la energía después de cada movimiento. El movimiento es mantenido si la energía es baja, de lo contrario el átomo es

retornado a su posición original. Este proceso es repetido muchas veces hasta que un mínimo de energía se alcanza. Un ciclo completo, donde cada átomo se mueve una vez, es llamado un paso de minimización. Miles de pasos pueden ser necesarios para encontrar una estructura razonable para una molécula.^{50,59}

El punto alto en el camino entre dos mínimos es de especial interés y es conocido como un punto de silla (saddle point). Ambos puntos mínimos y puntos de silla son puntos estacionarios en la superficie de energía, donde la primera derivada de la función energía es cero con respecto a todas las coordenadas.

El problema de minimización puede ser formulado de la siguiente manera: Dada una función f que dependa de una o dos variables independientes $x_1, x_2 \dots x_i$, encontrar los valores de las variables para que f tenga un valor mínimo. En un punto mínimo la primera derivada de la función con respecto a cada una de las variables es cero y la segunda derivada son todos los valores positivos.

$$\frac{\partial f}{\partial x_i} = 0; \frac{\partial^2 f}{\partial^2 x_i} > 0 \quad (2.2)$$

La función de mayor interés para nosotros es la de energía con respecto a mecánica molecular con las variables x_i de las coordenadas cartesianas o internas de átomos. Las minimizaciones de mecánica molecular son casi siempre desarrolladas con coordenadas cartesianas en donde la energía está en función de variables $3N$. Para funciones analíticas, el mínimo de una función puede ser encontrada usando métodos de cálculos estándar. Sin embargo, esto no es posible para sistemas moleculares debido a la complejidad en la que la energía varía con las coordenadas.

La energía potencial mínima se localiza usando métodos numéricos, que gradualmente cambian de coordenadas para producir configuraciones con energía cada vez más baja, hasta que se alcanza la energía mínima.

1. Métodos de Minimización Derivativa

Las derivadas mejoran la información que puede ser muy útil en minimización

de energía, además las derivadas son usadas por los métodos más populares de minimización. La dirección de la primera derivada de la energía (el gradiente) indica donde el mínimo se encuentra, y la magnitud del gradiente indica la inclinación de la pendiente local. La segunda derivada indica la curvatura de la función, información que puede ser usada para predecir donde la función cambiará de dirección.

a) Métodos de la primera derivada

- Steepest Descent: La energía es calculada para la geometría inicial y denuevo cuando uno de los átomos ha sido movido en un pequeño incremento en una de las direcciones del sistema de coordenadas.

Este proceso puede ser repetido para todos los átomos, los cuales son finalmente movidos a la nueva posición cuesta abajo de la superficie de energía. El procedimiento se detendrá si la condición de mínimo pre-determinada es cumplida. El proceso de optimización es lento cerca del mínimo, por eso el método del steepest descent es frecuentemente usado para estructuras lejanas del mínimo. Este es el método más probable para generar estructuras de baja energía de datos cristalográficos refinados pobremente. En muchos casos, esta minimización es usada como una primera aproximación y una corrida introductoria seguida por una subsecuente minimización usando un algoritmo más avanzado como gradiente conjugado.

Este método se mueve en una dirección paralela a la malla de fuerza, que en una analogía geográfica, sería como caminar directamente colina abajo. Al moverse en la dirección contraria del gradiente, la energía disminuye rápidamente, por ello este algoritmo es eficiente para optimizar configuraciones de alta energía. Teniendo definida la dirección por la cual moverse, es necesario decidir cuan lejos se puede mover a lo largo de la gradiente. La ventaja es que algoritmo es muy simple, y solamente requiere el almacenamiento del vector gradiente.

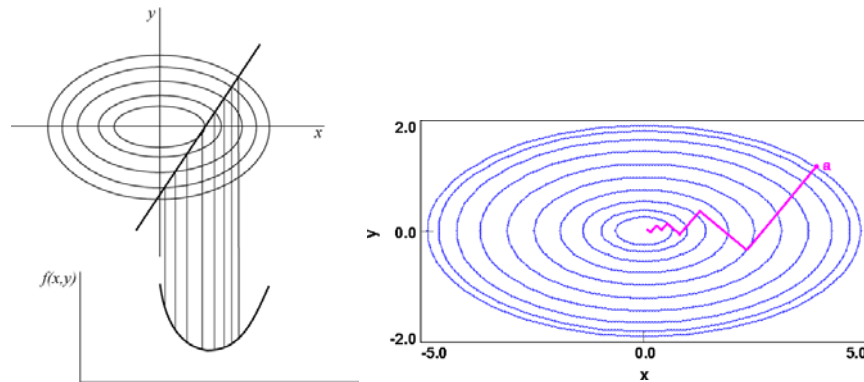


Figura 2.6: Búsqueda lineal es usada para localizar el mínimo en la función en la dirección del gradiente

- **Gradiente Conjugado:** El método del gradiente conjugado acumula la información acerca de la función de una iteración a la siguiente. Con este procedimiento, el reverso del progreso hecho en una iteración temprana puede ser evitado. Para cada paso de minimización, el gradiente es calculado y usado como información adicional para computar el nuevo vector dirección del procedimiento de minimización. Así cada paso sucesivo continuamente refina la dirección hacia el mínimo.

En el gradiente conjugado, los gradientes de cada punto son ortogonales pero las direcciones son conjugadas, a diferencia del steepest descent que ambos; el gradiente y la dirección de pasos sucesivos son ortogonales. El método del gradiente conjugado, esencialmente, combina la información obtenida del vector gradiente con la información acerca del vector gradiente de iteraciones previas. Lo que hace el método es calcular la nueva dirección de búsqueda utilizando una combinación lineal del gradiente en la etapa considerada y el de la etapa anterior.

El esfuerzo computacional y los requerimientos de almacenaje del método del gradiente conjugado son mayores que en el steepest descent pero el gradiente conjugado es el método de elección para siste-

mas grandes. El mayor gasto computacional total para la mayor eficiencia y el largo tiempo por iteración son más que compensados por la mayor eficiencia.

b) Método de la segunda derivada

- Newton-Raphson: el método de la segunda derivada mejora la información de la curvatura de la función para identificar la dirección de búsqueda, lo que permite predecir donde la función pasa a través de un mínimo. La eficiencia del método de Newton-Raphson se incrementa a medida que la convergencia se aproxima. El esfuerzo computacional y los requerimientos de almacenaje para cálculo de sistemas grandes son desventajas de este método. Para estructuras con alta tensión, el proceso de minimización puede llegar a ser inestable, así la aplicación de este algoritmo es limitada por problemas donde la rápida convergencia de una geometría preoptimizada en un mínimo extremadamente preciso se requiere. El método Newton-Raphson es el más adecuado para moléculas pequeñas (usualmente menos de 100 átomos o menos). Ya que este método calcula la matriz Hessiana esto puede ser computacionalmente demandante para sistemas con muchos átomos y también puede requerir una cantidad significativa de almacenaje.^{50,58,59} Este cálculo puede ser potencialmente una operación que consume tiempo representando esto una desventaja significativa para los métodos derivativos de la segunda derivada como Newton-Raphson. Ante esto surge el método Quasi-Newton el cual será explicado a continuación.

2. Quasi-Newton: Este método también conocido como método de la métrica variable, gradualmente construye el inverso de la matriz Hessiana en iteraciones sucesivas. En cada iteración k , las nuevas posiciones x_{k+1} son obtenidas a partir de las posiciones actuales x_k , el gradiente g_k y la aproximación actual de la matriz

Hesiana inversa H_k :

$$x_{k+1} = x_k - H_k g_k \quad (2.3)$$

Esta línea de búsqueda es desarrollada a lo largo del vector $x_{k+1} - x_k$. Habiendo movido a la nueva posición x_{k+1} , H es actualizado a partir de su valor en el paso previo de acuerdo a la fórmula dependiente del método específico que se utiliza. Los métodos de Davidon-Fletcher-Powel (DFP), Broyden-Fletcher-Goldfarb-Shanno (BFGS) y Murtaugh-Sargent (MS) son los más comunes de este tipo.⁵⁰

Método Quasi-Newtoniano l-bfgs: Este método quasi-Newton constituye un procedimiento más sofisticado para la minimización sin restricción con memoria limitada.⁶⁰ El término memoria limitada procede del hecho de que se almacena una aproximación de rango bajo de la inversa de la matriz hessiana, en lugar de la matriz completa. Cada iteración del algoritmo es una búsqueda a lo largo de una recta, cuya dirección se calcula a partir de la matriz inversa aproximada del hessiano. La función objetivo decrece siempre tras cada búsqueda exitosa a lo largo de la recta.⁶¹

En este método realmente no se calcula el hessiano ni su inversa, sino que se va construyendo una aproximación a él teniendo en cuenta los pasos anteriores de la minimización.⁶²

2.5.4. Dinámica Molecular

La dinámica molecular (DM) es un campo multidisciplinario⁶³. Sus leyes y teorías provienen de las matemáticas, física y química. Emplea algoritmos de las ciencias de la computación y teoría de la información. Permite entender a los materiales y las moléculas no como entidades rígidas, sino como numéricas.

Es una técnica que permite simular el movimiento de un sistema de partículas y puede aplicarse por igual a sistemas de distintos tamaños. Lo esencial para la simula-

ción en DM es conocer los potencial de interacción de las fuerzas entre moléculas o partículas y las ecuaciones de movimiento que rigen su dinámica.

Las simulaciones en dinámica molecular se inician con el conocimiento de la energía del sistema como una función de las coordenadas atómicas. La superficie de energía potencial determina la estabilidad relativa de las diferentes estructuras posibles, estables o metaestables.

Las simulaciones de DM calculan el movimiento de moléculas individuales de sólidos, líquidos y gases. La idea clave aquí es el movimiento, el cual describe como las posiciones, velocidades, y orientación cambian con el tiempo. En efecto, la DM constituye una película que sigue a las moléculas.⁶⁴

Una simulación en dinámica molecular requiere:

1. Estructura inicial: La estructura a utilizar podrá determinarse a partir de los datos experimentales o con la definición de una estructura de mínima energía mediante un análisis de búsqueda conformacional.
2. Temperatura/energía: La energía de la estructura depende de la temperatura del sistema, y se divide entre energía potencial y cinética.
3. Tamaño del paso: La trayectoria de las partículas es discreta en el tiempo. Normalmente se elige un paso de tiempo suficientemente pequeño (p.ej. 1 femtosegundo) para evitar errores numéricos.
4. Duración de la simulación: depende de la información requerida.
5. Otro parámetro que podría considerarse es la presión.^{49,50}

La DM se basa en la segunda ley del movimiento de Newton, $F=ma$, donde F es la fuerza ejercida sobre una partícula, m es su masa y a su aceleración. Conociendo la fuerza ejercida sobre cada átomo es posible determinar su aceleración en el sistema. La integración de las ecuaciones del movimiento, producirá entonces una trayectoria

que describe las posiciones, velocidades y aceleración de las partículas en función del tiempo. A partir de esta trayectoria, se pueden determinar los valores promedio de las propiedades del sistema. El método es determinístico: una vez que se conocen las posiciones y velocidades de cada átomo, se puede predecir el estado del sistema para cualquier valor del tiempo.

El método consiste en la resolución del sistema de ecuaciones diferenciales dada por la segunda ley de Newton:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{f}_i \quad \text{para } i = 1, \dots, N \quad (2.4)$$

donde m_i y \vec{r}_i son la masa y posición (coordenadas) de la partícula i , y la fuerza viene expresada en función del gradiente de la energía potencial $V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$:

$$\vec{f}_i = -\nabla_{\vec{r}_i} V \quad (2.5)$$

El resultado de la integración nos dará información sobre cómo varían las posiciones y velocidades de las partículas del sistema en función del tiempo, a lo cual se le denomina trayectoria.⁵²

2.5.5. Métodos de cálculo con disolvente

De manera general, existen dos modelos teóricos para tener en cuenta el efecto del disolvente. Uno, trata el disolvente de forma explícita; y el segundo modelo, que lo que hace es considerado de forma implícita.

1. Disolvente explícito: En la simulación, sus moléculas se sitúan alrededor del soluto.
2. Disolvente implícito: Los modelos con solvente implícito proponen una manera eficiente de representar los efectos electrostáticos de las moléculas solventes, mientras ahorran una gran parte de los cálculos involucrados. Este modelo ofrece

varias ventajas en comparación con la solvatación explícita, incluyendo la eliminación de la necesidad del equilibrio del agua alrededor del soluto, y la ausencia de la viscosidad, lo cual permite a la proteína explorar más rápido el espacio conformacional.

En general, un tratamiento implícito de solvatación considera la influencia media del solvente a través de la estimación directa de la energía libre de solvatación, definida como el trabajo reversible necesario para transferir el soluto en una configuración del vacío para la solución. Este modelo también es conocido como “modelo continuo”, un modelo continuo es aquel en el que el número de grados de libertad de las partículas se describen de forma continua, normalmente por una función de distribución.⁶⁵

En el modelo continuo un concepto básico es la cavidad que consiste en colocar una molécula de soluto (o unas pocas) en una cavidad hueca, inmersa en un medio dieléctrico continuo que imita el disolvente. La física de la interacción electrostática soluto-disolvente es simple. La distribución de carga del soluto dentro de la cavidad polariza el medio continuo dieléctrico, el cual a su vez polariza la distribución de carga del soluto. Este es un proceso autoconsistente que ha de resolverse numéricamente mediante un proceso iterativo.

Son varios los métodos que existen, sin embargo una revisión exhaustiva de los mismos queda fuera del alcance de este tesis por lo que nos limitaremos al modelo generalizado de Born que es el que aplica el programa usado para la DM.⁵²

- Modelo de Born Generalizado (BG): este modelo es un aproximado de las interacciones ion-disolvente, considerando al disolvente como un continuo de alta constante dieléctrica y provisto de cavidades de constante dieléctrica menor, donde los átomos del soluto están dispuestos.

Para obtener el potencial electrostático ϕ del modelo de dos dieléctricos, con una carga posicionada en el interior de la cavidad de la constante dieléctrica menor, se debe resolver la ecuación de Poisson:

$$\nabla[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \quad (2.6)$$

donde ρ es la distribución de carga. La constante dieléctrica vale ϵ_{in} en el interior de la cavidad donde se encuentran las moléculas del soluto y ϵ_{ex} en el exterior de la cavidad. Para el soluto en el gas se tiene $\epsilon_{ex} = 1$, en cuanto que para la solución acuosa $\epsilon_{ex} = \epsilon_w$, la constante dieléctrica del agua.

Para el caso de un único ion de radio a y carga q , los potenciales pueden ser calculados analíticamente y un componente electrostático de la energía libre de solvatación es dado por la conocida fórmula de Born:

$$\Delta G_{Born} = \frac{-q^2}{2a} \left(1 - \frac{1}{\epsilon_w}\right) \quad (2.7)$$

El modelo BG es capaz de reproducir relativamente bien la estructura y la energía de solvatación de diversas clases de moléculas. Esos parámetros son desarrollados comparando la estructura del soluto observada utilizando modelos de solvatación explícita, y su energía libre de solvatación obtenida a través de las soluciones numéricas de la ecuación de Poisson.

52,66

2.5.6. Mecánica Estadística

La mecánica estadística es una disciplina científica que pretende predecir las propiedades macroscópicas (entropía, capacidad calorífica, tensión superficial, viscosidad . . .) de un sistema a partir de las propiedades moleculares (geometría molecular, interacciones intermoleculares, masas moleculares . . .), en otras palabras nos proporciona el nexo de unión entre la descripción mecánica (cuántica o clásica) y termodinámica de un sistema macroscópico.

El objetivo de la mecánica estadística es derivar todas las propiedades de equilibrio de un sistema molecular macroscópico a partir de las leyes de dinámica molecular; no solo las leyes generales de la termodinámica sino también las funciones termodinámi-

cas específicas de un sistema dado. Sin embargo, la mecánica estadística, no describe como el sistema se aproxima al equilibrio. Se limita a indicar que la situación de equilibrio es para un sistema dado.⁶⁷

El estado termodinámico de un sistema se define normalmente por un pequeño conjunto de parámetros, como por ejemplo la temperatura T , la presión P y el número de partículas N , pudiéndose derivar, de éstos, el resto de propiedades termodinámicas. Por otra parte, el estado microscópico de un sistema se define por las posiciones atómicas \tilde{r} y sus momentos \tilde{p} , las cuales se consideran coordenadas en un espacio multidimensional denominado espacio de fase.

El desarrollo de la mecánica estadística depende del cálculo de la probabilidad de una distribución particular de puntos en el espacio y el cálculo del promedio de propiedades a partir del conocimiento de muchas distribuciones probables. Estos cálculos son basados en dos axiomas de mecánica estadística.

- Axioma I: El estado de equilibrio es el estado más probable. El estado más probable es el cual da el máximo número de estados en una distribución particular.
- Axioma II: Las propiedades de equilibrio son las propiedades promedio del sistema asumidas en un tiempo infinito. La termodinámica estadística esta basada en el principio que las propiedades termodinámicas son el promedio de las propiedades moleculares, y ella establece un régimen para el cálculo de estos promedios.⁶⁸

Habitualmente la mecánica estadística se divide en dos partes:

- La mecánica estadística de no equilibrio dedicada al estudio de los fenómenos de transporte (de calor, materia . . .) y de las reacciones químicas.
- La mecánica estadística de equilibrio (o termodinámica estadística) que se ocupa de sistemas en equilibrio termodinámico.

Nos enfocaremos en la termodinámica estadística ya que los sistemas en equilibrio son los que nos interesan para esta tesis. Siendo su postulado fundamental que todos los microestados posibles de un conjunto aislado son igualmente probables. El número de microestados igualmente probables que corresponde a un determinado macroestado k se denomina probabilidad termodinámica W_k del macroestado. El número total de posibles microestados Ω de un conjunto o probabilidad termodinámica del conjunto es igual a la suma de todas las probabilidades termodinámicas de cada macroestado:

$$\Omega = \sum_k W_k \quad (2.8)$$

1. Función de partición: La función de partición o suma sobre estados es una fórmula matemática que expresa una suma de factores de probabilidad de los diferentes microestados. Los términos de esta suma regulan el modo como se distribuyen las partículas del colectivo en los diferentes microestados. La función de partición es la herramienta matemática principal de la termodinámica estadística. A partir de ella se obtiene toda la información termodinámica sobre el sistema.⁶⁹ La función de partición se obtiene a partir de la distribución de la estadística de Maxwell-Boltzmann.

La función de partición se representa por Z ,

$$Z \equiv \sum_j g_j \exp \frac{\epsilon_j}{k_B T} \quad (2.9)$$

La función de partición depende sólo de la temperatura T y de los parámetros que determinan los niveles de energía.

2. Ensamble estadístico: Se denomina ensamble al conjunto de todos los posibles sistemas que tienen diferentes microestados, pero que tienen un mismo estado termodinámico macroscópico. Cada ensamble está asociado con una función matemática que puede ser usada para extraer valores de las propiedades

termodinámicas del sistema.⁷⁰ De acuerdo con la relación del sistema con el universo, uno de los tres tipos generales de ensambles deberían ser usados.

- Ensamble microcanónico: considera que cada sistema aislado, el cual tiene como requisito de que todos los sistemas tienen que tener la misma energía.
- Ensamble canónico: describe un sistema en equilibrio termodinámico con su entorno. Se debería tener como único intercambio de energía, el calor intercambiado con el exterior.
- Ensamble grancanónico: usado para sistemas abiertos donde el volumen y la temperatura son los mismos, lo que quiere decir que la materia es capaz de intercambiarse entre los sistemas; la composición de cada sistema individual fluctuará, pero ahora el potencial químico será el mismo para todos los sistemas.

Ensamble canónico: Considera que todos los sistemas cerrados idénticos pueden estar en contacto térmico con cualquier otro, de forma que pueden intercambiar energía. La energía total de todos los sistemas es E y puesto que están en equilibrio térmico, tienen la misma temperatura T . Algunas configuraciones serán mucho más probables que otras. La cantidad Q , que será función de la temperatura, se llama función de partición. canónica.⁷¹

$$Q = \sum_j \exp^{-\beta \epsilon_j}, \quad \beta = \frac{1}{k_B T} \quad (2.10)$$

Capítulo 3

Materiales y Métodos

3.1. Equipos y Software

Dentro de los requerimientos para poder realizar la parte computacional, necesitaremos los siguientes equipos:

3.1.1. Hardware

Computador Lenovo IdeaPad Z570. INTEL i7-2670M 2.2GHz(3.1GHz c/TB)/4GB-DDR3/HD750GB

3.1.2. Bases de Datos

1. NCBI (National Center for Biotechnology Center): Es una base de datos almacena y actualiza constantemente la información referente a secuencias genómicas, artículos científicos referentes a biomedicina, biotecnología, bioquímica, genética y genómica, una recopilación de enfermedades genéticas humanas, además de otros datos biotecnológicos de relevancia en diversas bases de datos. Todas sus bases de datos están disponibles en línea de manera gratuita, y son accesibles usando el buscador Entrez (www.ncbi.nlm.nih.gov/).

2. PDB (Protein Data Bank): Es una base de datos de la estructura tridimensional de las proteínas y ácidos nucleicos. Estos datos, generalmente son obtenidos mediante cristalografía de rayos X o resonancia magnética nuclear. Están bajo el dominio público y pueden ser usados libremente; (www.pdb.org/pdb/home/home.do).

3.1.3. Software

El Sistema Operativo utilizado en esta investigación es Linux, distribución Ubuntu 11.04.

1. Gromacs 4.5.5.: es un paquete de dinámica molecular diseñado para la simulación de moléculas bioquímicas tales como las proteínas, los lípidos y los ácidos nucleicos. Es un software libre. Disponible en www.gromacs.org.
2. Molden 5.0.: es un programa usado como interfaz gráfica, que permite el procesamiento de las estructuras moleculares y electrónicas, se puede visualizar los resultados de densidades electrónicas y orbitales moleculares.
Disponible en www.cmbi.ru.nl/molden
3. Chimera 1.6.: es un programa altamente extensible para la visualización interactiva y el análisis de estructuras moleculares y datos relacionados, incluyendo mapas de densidad, secuencia de alineaciones, resultados de acoplamiento, trayectorias, y conjuntos de conformación. Además se pueden generar imágenes de alta calidad y animaciones. Disponible en www.cgl.ucsf.edu/chimera
4. APBS 1.3.: es un paquete de software para el modelado biomolecular de solvatación a través de la solución de la ecuación de Poisson-Boltzmann (PBE), uno de los modelos continuos más populares para describir las interacciones electrostáticas entre solutos moleculares en medios salinos y acuosa.
Disponible en www.poissonboltzmann.org/apbs

5. Grace 5.1.22.: es una herramienta para hacer ploteos de dos dimensiones de datos numéricos. Este programa es un software libre
Disponible en <http://plasma-gate.weizmann.ac.il/Grace/>
6. Programa para armar el trímtero. Desarrollo propio, en el programa se trabajó con los siguientes conceptos:

- Rotación de coordenadas: El vector r es un objeto geométrico independiente del sistema de coordenadas. Dejamos a r en dos sistemas diferentes, uno rota en relación al otro. Considerando el caso de dos dimensiones, si las coordenadas X, Y son rotadas en sentido contrario a través del ángulo ϕ , manteniendo r fijo, conseguimos la siguientes relaciones entre los componentes resueltos en el sistema original y esto resuelve el nuevo sistema rotado⁷²:

$$x' = x \cos \phi + y \sin \phi \quad y' = -x \sin \phi + y \cos \phi \quad (3.1)$$

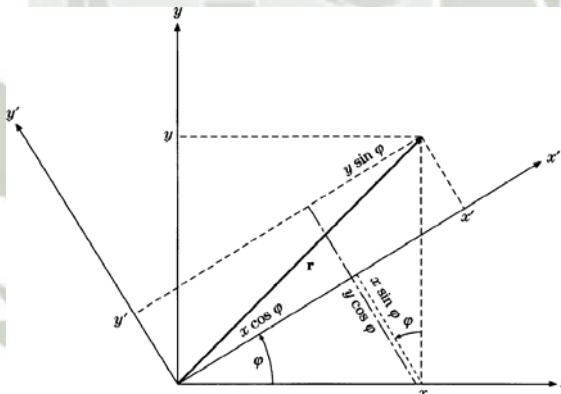


Figura 3.1: Rotación de coordenadas.

- Simetría molecular: es la simetría de la estructura formada por los diferentes núcleos de la molécula en su posición de equilibrio. Otro concepto importante en este punto es operación de simetría, que es aquella transformación de un cuerpo de tal manera que la posición final es indiscernible con respecto a la inicial, las operaciones de simetría típicas son las

rotaciones e inversiones. Existe un elemento de simetría correspondiente a cada operación de simetría, que es el punto, la línea o el plano respecto al que se aplica la operación de simetría.⁷¹

3.1.4. Servidores online

1. Aplicación online ROBETTA Server: Este servicio es ofrecido por el laboratorio Baker y se encuentra disponible en [http : //robetta.bakerlab.org/](http://robetta.bakerlab.org/), proporciona herramientas para la predicción de estructura de proteínas, realizando un análisis de las cadenas de proteínas en los dominios putativos haciendo uso de los siguientes métodos:⁷³

- Protocolo Ginzu: intenta determinar las regiones de una cadena de proteína que se plegara en unidades globulares, llamados "dominios". Se analiza la secuencia de la cadena de proteína con métodos sucesivos de detección para determinar algunos homólogos con estructuras determinadas experimentalmente, como el PDB-BLAST. Después que los homólogos son identificados, la búsqueda de las demás regiones se hace con HMMER contra el Pfam; una base de datos de la familia de proteínas. Por último, la alineación de secuencias múltiples PSI-BLAST se utiliza para asignar las regiones de mayor probabilidad de poseer un dominio contiguo basado en secuencia de las agrupaciones. El último paso consiste en seleccionar los puntos de corte entre los dominios (y, posiblemente, la definición de nuevos dominios basados en los fuertes puntos de corte para los tramos restantes de la secuencia que no lo ha combinado con un homólogo de una estructura o Pfam) utilizando el PSI-BLAST MSA.⁷³
- Cálculos de Ab initio: Constan de tres etapas: aproximación a la estructura molecular, optimización para encontrar la geometría de energía mínima y cálculo frecuencias.^{74,75}

- Modelado por homología: Generalmente, el proceso de obtención de un modelo proteico virtual a través de la ejecución de la estrategia de modelado por homología involucra cuatro etapas principales: búsqueda de proteínas homólogas, alineamiento de las secuencias, construcción y optimización de los modelos, y validación de los mismos.⁷⁶
2. Aplicación online MolProbity: es un servicio que ofrece la validación de la calidad para las estructuras tridimensionales (3D) de proteínas, ácidos nucleicos y complejos, (disponible en molprobity.biochem.duke.edu/), este servidor online realiza un diagrama de Ramachandran, el cual será explicado a continuación:
- Este diagrama es una herramienta útil en el análisis estructural cuantitativo de las conformaciones de proteínas que representa los valores permitidos de los ángulos de rotación ϕ y ψ , estos valores corresponden a los dos ángulos dihedros definidos entre enlaces peptídicos colindantes.⁷⁷
- Ramachandran, un científico hindú, uso modelos de pequeños polipéptidos para variar sistemáticamente estos ángulos con el objeto de encontrar conformaciones estables. Para cada una de las conformaciones, la estructura fue examinada para detectar los contactos cercanos entre los átomos. Los átomos fueron considerados como esferas rígidas con dimensiones correspondientes a su radio de van der Waals, por lo tanto, los ángulos que causaban una colisión de las esferas correspondían a ángulos prohibidos y correspondían a conformaciones no permitidas de la cadena polipeptídica.
- En este diagrama vemos que el eje horizontal son los valores de phi (ϕ), mientras que el eje vertical muestra los valores de psi (ψ). Las zonas blancas corresponden a regiones que son no permitidas para todos los aminoácidos, excepto para Glicina, el cual es el único aminoácido que no presenta cadena lateral. Las regiones rojas corresponden a conformaciones donde no hay impedimentos, es decir estas son zonas favorecidas llamadas conformaciones α -hélices y hojas- β . Las zonas amarillas muestran las regiones permitidas. También tenemos una región

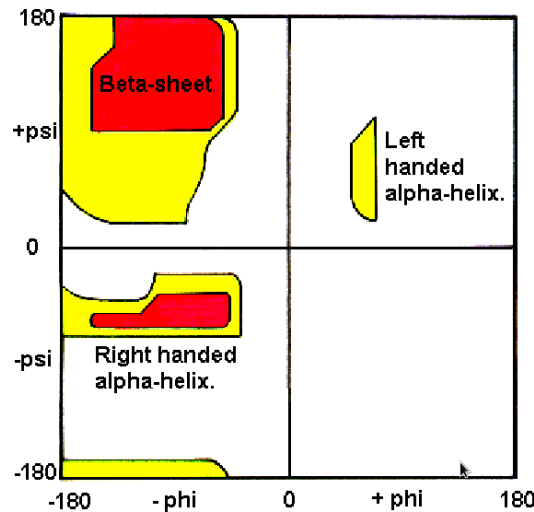


Figura 3.2: Diagrama de Ramachandran caso general.

que corresponde a una hélice con giro hacia la izquierda, los L-aminoácidos no pueden formar una hélice con giro a la izquierda pero ocasionalmente residuos individuales pueden adoptar esta conformación, estos residuos generalmente son glicina pero también pueden ser asparragina y ácido aspártico en los que la cadena lateral forma un puente hidrógeno con la cadena principal y así estabiliza esta conformación desfavorable.⁷⁸

3. Servidor PDB2PQR: es un servicio que permite obtener a partir del archivo .pdb, los archivos necesarios para el cálculo del potencial electrostático, disponible en <http://nbc-222.ucsd.edu/pdb2pqr1,8/>.

3.2. Búsqueda de las secuencias usadas para las simulaciones

Para la siguiente investigación se usó la base de datos del NCBI, por ser la base de datos más completa en la actualidad, referente a secuencias de nucleótidos y proteínas. De estos sitios se obtuvo la secuencia de la proteína CusC de *Escherichia coli* y *Acidithiobacillus ferrooxidans*. Ambas secuencias se obtuvieron en formato FASTA.

3.3. Obtención de las estructuras secundarias

Se buscó en la base de datos PDB, por ser la principal base de datos de estructuras tridimensionales de proteínas, en esta base encontramos la estructura secundaria de la proteína CusC de *Escherichia coli* la cual fue obtenida experimentalmente por cristalografía de rayos X. El formato de este archivo es .pdb.

En el caso de la proteína CusC de *Acidithiobacillus ferrooxidans* se trabajó con el servidor online Robetta, se coloca en este servidor la secuencia de la proteína CusC de *Acidithiobacillus ferrooxidans* en formato FASTA. Las estructuras luego de que el servidor realiza el análisis, se encuentran en formato .pdb. Este formato es un estándar para archivos que contienen coordenadas atómicas.



Figura 3.3: Imagen del servidor Robetta donde se copia la secuencia FASTA.

3.4. Minimización de las estructuras

Para estos pasos se usa el software GROMACS 4.5, el cual se puede descargar de la página <http://www.gromacs.org>, donde también se encuentra su modo de instalación, este programa se trabaja en la terminal.

1. Una vez que tenemos los archivos con las estructuras secundarias de las proteínas en formato .pdb, se procede a limitar el sistema, para esto usamos el comando `editconf`, este nos permite colocar la proteína en una caja y darle un tamaño adecuado (unidades en nm), lo suficientemente amplia para que la proteína pueda moverse.

- `editconf -box 20 20 20 -c -f molecula.pdb -o molecula-caja.pdb`

Luego de poner las proteínas dentro de sus cajas, el siguiente paso es prepararlas para la minimización de energía.

2. Se usa el comando `pdb2gmx`, el cual nos permite escoger el campo de fuerza, crear los archivos con la topología de la proteína (.top y .itp) y convierte el formato .pdb a formato (.gro) que contiene las coordenadas por átomos de las proteínas, el cual es compatible con el software GROMACS.

- `pdb2gmx -ignh -f molecula-caja.pdb -p molecula.top -o molecula.gro -i molecula.itp`

3. El siguiente comando es `grompp`, acá necesitamos el archivo .mdp, el cual contiene los parámetros que se necesitan para la minimización.

- `grompp -f minima.mdp -c molecula.gro -p molecula.top -o molecula.tpr`

4. Para que el programa empiece el cálculo se utiliza el comando `mdrun`, mediante el cual se puede obtener los archivos necesarios para monitorear el cálculo usando un visualizador como Chimera.

- `mdrun -nice 0 -v -s molecula.tpr -o molecula.trr -c molecula-minima.gro`

Las pautas para la construcción del archivo .mdp se encuentran en el manual del GROMACS,⁷⁹ y una muestra de como es este archivo se encuentra en la página web del GROMACS, <http://manual.gromacs.org>, seguidamente describiremos los parámetros que se consideran en estos archivos.

- Preprocesamiento: se definen los directorios que usará el software.
- Control del cálculo: especificamos el tipo de cálculo, en este caso minimización tipo l-bfgs.
- Energía de minimización: número de pasos, tamaño de los pasos.
- Control de salida: los archivos que se obtendrán al finalizar el cálculo.
- Búsqueda de vecinos: se refieren a las parámetros de los límites de la caja de la molécula.

3.5. Dinámica molecular de las estructuras

Continuando con el uso del software GROMACS 4.5, se buscó la estabilidad de las estructuras de las proteínas realizando una simulación de dinámica molecular, utilizando el ensamble canónico.

El procedimiento para la simulación de dinámica molecular, es muy parecido al de minimización.

1. Se procede a limitar el sistema mediante el comando `editconf` para generar una caja. Para la dinámica molecular, se puede utilizar una caja de gran tamaño de esta manera la proteína podrá acomodarse y moverse sin rozar con los límites de la caja.
 - `editconf -box 20 20 20 -c -f molecula-minima.gro -o molecula-minima-caja.gro`
2. Usamos el comando `pdb.gmx`, luego, el comando `grompp`, aquí se usó también un archivo `.mdp` el cual es diferente al usado en la minimización, en este archivo estarán los parámetros para trabajar con un ensamble canónico y solvente implícito, además de la temperatura y el tiempo. Para empezar el cálculo de la simulación de dinámica molecular, se utiliza el comando `mdrun`.

- `pdb2gmx -ignh -f molecula-minima-caja.pdb -p molecula.top -o molecula.gro -i molecula.itp`
- `grompp -f dinamica.mdp -c molecula.gro -p molecula.top -o molecula.tpr`
- `mdrun -nice 0 -v -s molecula.tpr -o molecula.trr -x molecula.xtc -c molecula-final.gro -e molecula.edr`

Para la construcción del archivo `.mdp` se tomaron en cuenta los siguientes parámetros

- Preprocesamiento: se definen los directorios que usará el software.
- Control del cálculo: el cálculo es dinámica molecular.
- Dinámica molecular: número de pasos, tamaño de los pasos.
- Control de salida: los archivos que se obtendrán al finalizar el cálculo.
- Búsqueda de vecinos: se refieren a las parámetros de los límites de la caja de la molécula.
- Solvente implícito: se define los parámetros para trabajar con solvente implícito.

3.6. Análisis de la estabilidad

Al final de la simulación de dinámica molecular, se obtuvo los archivos que permiten la evaluación de la estabilidad estructural de las moléculas. (`.edr`, `.xtc`, `.trr`, `.top`)

1. Obtenemos los datos de energía potencial, cinética y total utilizando el comando `g_energy`, el cual utiliza el archivo con extensión `.edr` obtenido al final de la simulación de dinámica molecular.

- `g_energy -f molecula.edr -o energy.xvg`

2. Gráfico RMSD (Root Mean Square Deviation): este gráfico realiza comparaciones entre las estructuras que se van formando durante la trayectoria de la dinámica molecular.
 - `g_rms -f molecula.tpr -s molecula.xtc -o rmsd.xvg`
3. Gráfico RMSF (Root Mean Square Fluctuation): RMSF es una medida de flexibilidad local de una estructura. A partir de una simulación de dinámica molecular, el RMSF del átomo *i* da cuenta del grado de fluctuación de las coordenadas de dicho átomo respecto a su promedio temporal.
 - `g_rmsf -f molecula.tpr -s molecula.xtc -o rmsd.xvg`
4. Número de Enlaces de Hidrógeno: este gráfico analiza todos los puentes de hidrógeno existentes entre dos grupos de átomos posibles donadores D y aceptores A, de la siguiente manera:
 - Distancia Donador-Aceptor (τ) distribución de todos los puentes de hidrógeno.
 - Angulo Hidrógeno-Donador-Aceptor (α) distribución de todos los puentes de hidrógeno.
 - El número total de puentes de hidrógeno en cada período de tiempo.

Para determinar, si existe un puente de hidrógeno, un criterio geométrico es usado:⁷⁸

$$\tau \leq \tau_{HB} = 0,35\text{nm} \quad \alpha \leq \alpha_{HB} = 30^\circ \quad (3.2)$$

- `g_hbond -f molecula.xtc -s molecula.tpr -o hbond.xvg`

Para visualizar todos estos gráficos mencionados usamos `xmgrace`, colocando el siguiente comando:

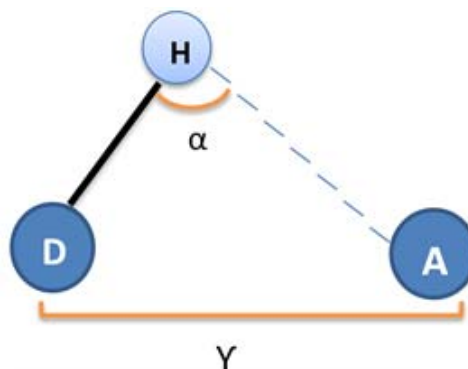


Figura 3.4: Criterio geométrico de los enlaces de hidrógeno.

- xmgrace hbond.xvg

5. Obtener la estructura promedio de las proteínas, para esto se usó con el siguiente comando:

- `g_covar -f molecula.xtc -s molecula.tpr -av average.pdb -o eigenval.xvg -c eigenvec.trr -l covar.log`

6. Diagrama de Ramachandran: luego que obtenemos el archivo de la estructura promedio en formato .pdb procedemos a ingresarlo en el servidor MolProbity y como resultado nos da un archivo en .pdf.

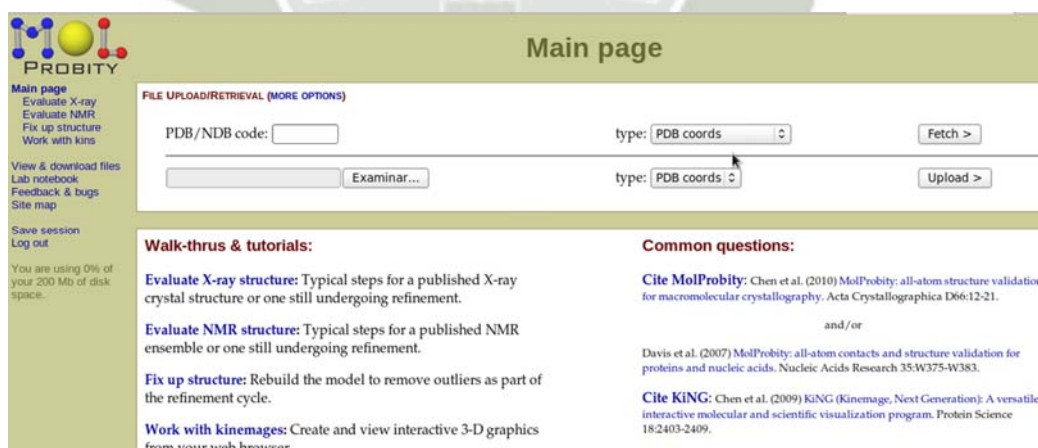


Figura 3.5: Imagen del servidor MolProbity donde se sube la estructura promedio para el diagrama de Ramachandran.

3.7. Construcción de los trómeros

Para esto se diseñó un programa en lenguaje de programación fortran, el cual es un lenguaje de alto nivel, orientado a facilitar las cosas al usuario, y que permite escribir el programa de manera casi idéntica a como se escriben las fórmulas en un papel. En la figura 3.6 se muestra un diagrama del programa.

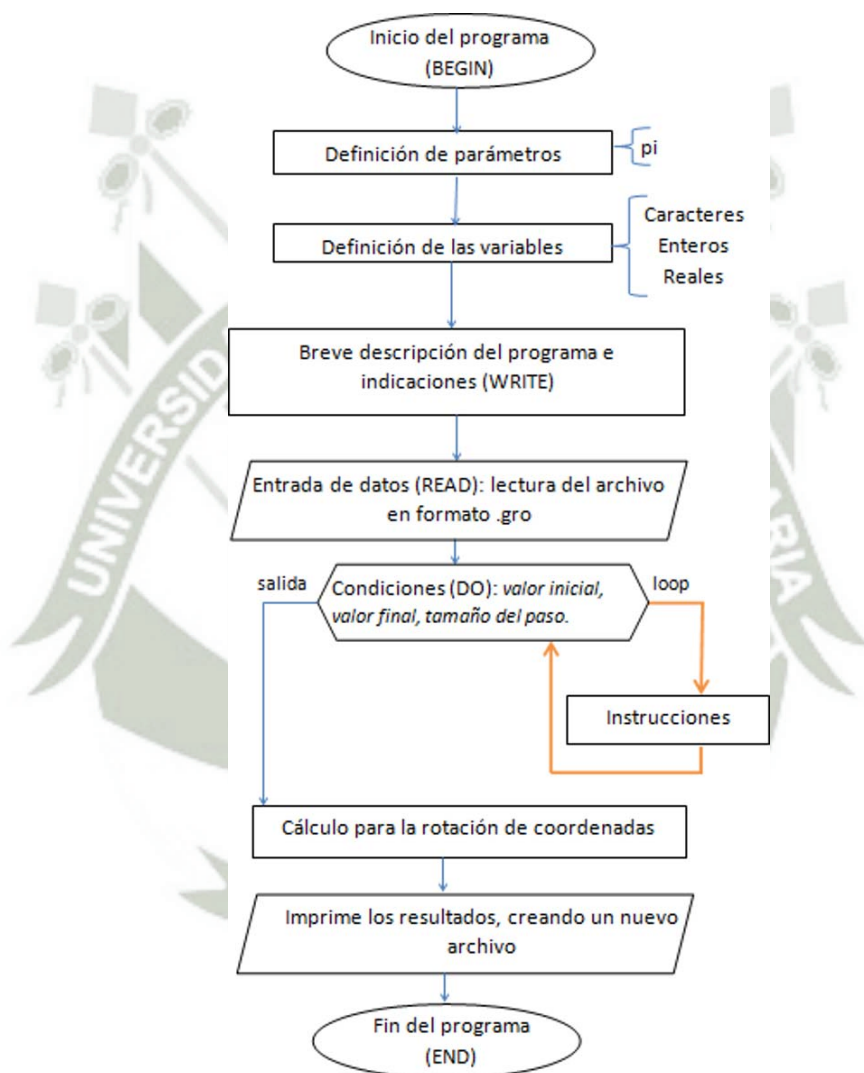


Figura 3.6: Diagrama de flujo del programa que se usó para construir el trómero.

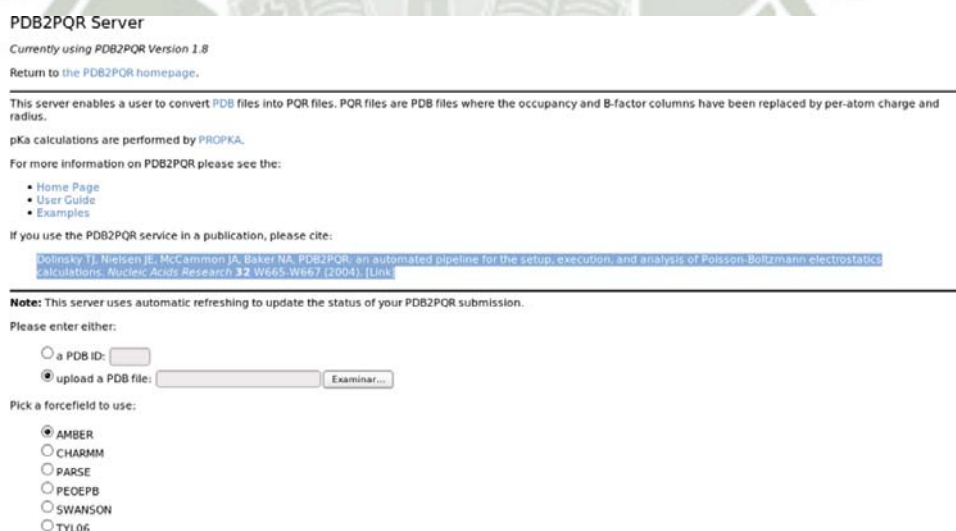
Los archivos de los trómeros se obtienen en formato .gro y luego con el comando editconf los transformamos a formato .pdb, estos archivos tenemos que prepararlos para la minimización y para la dinámica molecular, para esto nombramos cada una de las

cadena del trímero, de este modo el programa GROMACS reconocerá al trímero como tres moléculas unidas y no como una sola, en consecuencia nos generará tres archivos .itp para cada trímero, este formato de archivo es usado para definir componentes individuales de una topología como un archivo separado, este archivo contiene solo la información para una molécula particular, en nuestro caso tenemos 3 de estos archivos por cada trímero. Estos archivos .itp, están incluidos dentro del archivo .top, en el cual está definida la topología del sistema completo

Luego se realizan los mismos pasos expuestos anteriormente para la minimización, dinámica y el análisis de la estabilidad.

3.8. Potencial Electrostático

Antes de realizar el cálculo del campo de potencial electrostático sobre la superficie de las proteínas, fue necesario obtener los archivos .pqr y .in, para ello se utilizó el servidor en línea PDB2PQR, donde se sube el archivo de la estructura promedio en formato .pdb.



PDB2PQR Server
Currently using PDB2PQR Version 1.8
Return to the PDB2PQR homepage.

This server enables a user to convert PDB files into PQR files. PQR files are PDB files where the occupancy and B-factor columns have been replaced by per-atom charge and radius.
pKa calculations are performed by PROPKA.
For more information on PDB2PQR please see the:

- Home Page
- User Guide
- Examples

If you use the PDB2PQR service in a publication, please cite:
[Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research. 32: W665-W667 \(2004\). \[Link\]](#)

Note: This server uses automatic refreshing to update the status of your PDB2PQR submission.

Please enter either:

a PDB ID:

upload a PDB file:

Pick a forcefield to use:

- AMBER
- CHARMM
- PARSE
- PEOEPB
- SWANSON
- TYL06

Figura 3.7: Imagen del servidor PDB2PQR donde se sube la estructura promedio.

Luego en el archivo .in corregimos la temperatura, para el cálculo del potencial electrostático, previamente se instaló el programa APBS, el cual se descargó de la

página www.poissonboltzmann.org/apbs, haciendo uso del siguiente comando se puede dar inicio al cálculo:

- `apbs molecula.in`

Al finalizar el cálculo obtenemos un archivo `pot.dx` el cual contiene los datos de la hipersuperficie generada que puede ser visualizada con el programa Chimera.

3.9. Obtención y Estudio de Parámetros Termodinámicos

- Para los cálculos usamos los datos de energía total por tratarse de un ensamble canónico. Si tomamos en cuenta que la función de partición describe las propiedades de un sistema termodinámicamente en equilibrio⁷¹, el promedio de las Energías Totales de los microestados para cada sistema, se puede considerar como la Energía Total del Sistema. Luego de la simulación de dinámica molecular se obtiene, datos referentes a la energía del sistema. Calculamos:

- La entropía (S):

$$\Delta S = \frac{\Delta H}{T}$$

- La entalpía (H):

$$H = U + P V$$

- La capacidad calorífica (C_V), a volumen constante se define como,

$$C_V = \frac{\Delta U}{\Delta T}_v$$

Capítulo 4

Resultados y Discusión

4.1. Proteínas CusC de *Escherichia coli* y *Acidithiobacillus ferrooxidans*

4.1.1. Búsqueda de las secuencias usadas para las simulaciones

Para iniciar nuestra investigación se procedió a la búsqueda en las bases de datos de las secuencias tanto para la proteína CusC de la *Escherichia coli* y de la *Acidithiobacillus ferrooxidans*, esta búsqueda la hicimos en la base de datos NCBI donde se encontró la secuencia de las proteínas CusC de *E. coli* con el código 3PIK_A, esta proteína tiene 446 aminoácidos. En cuanto a la secuencia de la proteína de *A. ferrooxidans* se encontró en el NCBI con el código Lferr_1619 en la *A. ferrooxidans* ATCC 53993, esta secuencia cuenta con 427 aminoácidos. Cabe destacar que anteriormente esta secuencia fue proporcionada por el Dr. Carlos Jerez de la Universidad de Chile, la secuencia que él mandó tiene el código AFE_1143 en la *A. ferrooxidans* ATCC 23270, siendo reportada en la base de datos TIGR.¹² Ambas secuencias se obtuvieron en formato fasta, se utiliza este formato por su simplicidad lo que hace fácil el manipular y analizar las secuencias usando herramientas de procesado de textos. Las secuencias se pueden observar en la figura 4.1.

Acidithiobacillus ferrooxidans ATCC 53993

>gi|198248248|gb|ACH83841.1| outer membrane efflux protein

```
MYTTIRGCLTPPAWGGICFLAGVLLLLSTTHASAAPLSLQDAEAIALGKNPGLGAIKQKVVELRHQAVAV
AQLPDPHLDLGAANLPLNSFSMNQQQMSLSVGLSQTFFPSFGKLGKGGQQAGVEAQAADTLRGQS
AELVLLRRRAWLQALYAEDAETIRHQEQQLQAESVQAALALYRSAQGSQADVLRAQLARDSLANDISKL
QAERASDLAQIAQILNLPKPPSIEHQWPNLSPITLVQAEARLSGQPLLRAAQAQTRAAQVGVKVKARG
YWPEVTVSVDYGQDFYPGSPNWLSAGVDLSLPIFPGNRQDQDVAAARARALRAQYRYDDQHLAMIQ
QVRANFARYVALKTELEERTDQRLPTARNAFSATLAAYAVGRAELSAVLRQTQKEVLDYALSRLQYRRDLA
LSAAELDFLT TEGGARP
```

(a)

Escherichia coli

>gi|321159981|pdb|3PIK|A Chain A, Outer Membrane Protein Cusc

```
CSLAPDYQRPAMPVPPQQFSLSQNGLVNAADNYQNAGWRTFFVDNQVKTLLISEALVNNRDLRMATLKV
QEARAQYRLTDADRYPQLNGEGSGSWGNLKGNTATTREFSTGLNASFDLDFGRKKNMSEAERQNYL
ATEEAQRAVHILLVSNVAQSYFNQQLAYAQQLIAEETLRNYQQSYAFVEKQLLTGSSNVLALEQARGVIES
TRSDIAKRQGELAQANNALQLLGSYGKLPQAQTVNSDSLQSVKLPAGLSSQILLQRPDIMEAEHALMAA
NANIGAARAAFFPSISLTSGISTASSDLSSLFNASSGMWNFIPKIEIPIFNAGRQANLDIAEIRQQQSVVNY
EQKIQNAFKEVADALALRQSLNDQISAQQRYLASLQITLQRARALYQHGAVSYLEVLD AERSLFATRQTL
DLNARQVNEISLYTALGGGGHHHHHH
```

(b)

Figura 4.1: Secuencias de las proteínas CusC de *A.ferrooxidans* y *E. coli* en formato FASTA

4.1.2. Obtención de las estructuras secundarias

Una vez que se obtuvo las secuencias, se procedió a generar un modelo de las proteínas, en el caso de la proteína CusC de *E. coli* se encontró su estructura en la base de datos PDB, dicha estructura fue obtenida por cristalografía de rayos-X; su código es el mismo que encontramos en el NCBI, según código 3PIK_A. El archivo que descargamos se encuentra en el formato .pdb en el cual están definidas las posiciones de los átomos en el espacio por un conjunto de coordenadas cartesianas e identificadas con nombres en base a campos de fuerza.

Por otro lado la estructura secundaria de la proteína CusC de *A. ferrooxidans* se predijo subiendo la secuencia fasta al servidor Robetta, este programa fue desarrollado por el laboratorio Baker de la Universidad de Washington, el fundamento de este pro-

grama fue descrito en la sección de metodología. Luego de finalizada la predicción, el programa nos dio 5 estructuras similares, para elegir la que usaremos en los cálculos posteriores se verifico las que tuvieran la secuencia completa, después se realizó una primera minimización de todas las estructuras restantes y elegimos la de menor energía. En la figura 4.2 se puede ver ambas estructuras seleccionadas.

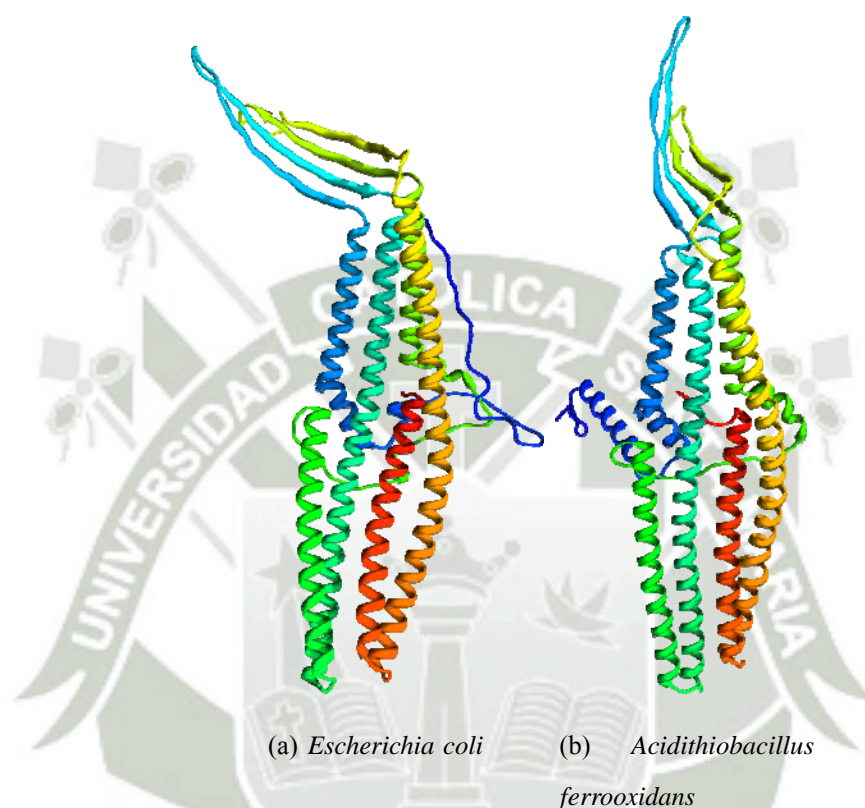


Figura 4.2: Estructuras secundarias de las proteínas CusC. Foto obtenida con Chimera 1.6.1

Como se puede observar ambas estructuras secundarias son muy similares, entre sus características tenemos que ambas poseen 4 láminas β antiparalelas estas conformaciones se hallan unidas transversalmente por enlaces de hidrógeno intercatenarios, estas cadenas para *A. ferrooxidans* se forman en las siguientes posiciones: HIS 77-ASN 88, GLN 94-THR 107, GLU 278-ASP 288 y ASN 295-SER 304; para *E. coli* las láminas β se encuentran en las siguientes posiciones: GLN 84-SER 94, THR 104-ASP 116, SER 290-SER 301 y GLY 313-GLU 322. También se pueden observar cadenas

largas α hélice. Este tipo de proteínas pertenecientes a la familia OMF se sabe por bibliografía que forman un β barril cuando se forma un trímero, esto será visto más adelante.

4.1.3. Minimización de las estructuras

Una vez obtenidas las estructura de nuestras dos proteínas CusC, se procedió a la minimización de sus energías, este paso se realiza para lograr una optimización geométrica; esta etapa del proceso es importante porque la estructura de la proteína de *E. coli* proveniente de datos de cristalografía, no tiene una geometría asociada a un sistema en solución, ya que las fuerzas internas son influenciadas por pequeñas desviaciones de las longitudes “ideales” de los enlaces. Además la minimización se realiza para que las estructuras se relajen y evitar la distorsión de los sistemas, de lo contrario las fuerzas asociadas a los átomos serán demasiado grandes y los algoritmos de integración podrían fallar.

Para la minimización se utilizó primero el método steep descent y luego el método quasi-newton l-bfgs, siendo el número de pasos para ambos métodos de 100000. Esta última minimización se realizó unas 5 veces para asegurarnos que la molécula se ubique en un mínimo local.

4.1.4. Dinámica Molecular de estructuras

Posteriormente se realizó la dinámica molecular, en esta etapa se procedió primero a una simulación de 50000ps (50ns) para lograr una relajación de la estructura, en una primera instancia se trabajó con el campo de fuerza OPLS-AA/L en un ensamble canónico (N V T), manteniendo una temperatura de 303.15 K. Una vez culminada la relajación de la molécula, se procedió con una segunda dinámica molecular durante 70000ps (70ns) con los mismos parámetros anteriores pero esta vez con el campo de fuerza AMBER99SB-ILDN. Se cambió de campo de fuerza ya que AMBER genera

una mejor repuesta cuando se trabaja con solvente implícito además hace una sobrees-
timación de los puentes de hidrógeno con el solvente, de forma que para estructuras
en las cuales los puentes de hidrógeno son importante, es muy útil.

La dinámica molecular se desarrolla con el propósito de que las proteínas lleguen a
la fase de equilibrio, es decir permitir al sistema evolucionar desde su configuración
inicial hasta la búsqueda del equilibrio, la cual se halla relacionado al mínimo global
o estructura estacionaria.

4.1.5. Análisis de la estabilidad

El equilibrio de una proteína se consigue cuando los valores del conjunto de pro-
piedades monitoreadas llegan a ser estable. Las propiedades que monitoreamos de las
moléculas resultantes serán descritas a continuación, en primer lugar se realizó la
gráfica de los datos de energía que incluyen: energía potencial, energía cinética y la
energía total.

Desde el punto de vista termodinámico, la energía total es mandatoria, esto im-
plica que es el parámetro a observar cuando deseamos conformaciones estructurales
estables, es así que sistemas conformacionales con la mayor energía negativa, estarán
asociadas a sistemas más estables.

Para el caso de *E. coli*, como se muestra en la figura 4.3, la energía total (ver figu-
ra 4.3-a) continuó bajando hasta los 14000ps (14ns) y de ahí en adelante permanece
constante, por está razón todos los gráficos posteriores relacionados a la proteína
CusC de *E. coli* (monómero) se tomarán desde los 14000ps (14ns). En la figura 4.3-b,
donde se presenta la energía cinética (línea roja), potencial (línea negra) y total (línea
verde), también podemos apreciar que la relajación se logra a los 14000ps (14ns) para
los dos componentes de la energía total.

Los resultados de la dinámica molecular para la *A. ferrooxidans*, como se muestra en
la figura 4.4-a, la energía total continuó descendiendo hasta los 20000ps (20ns), luego

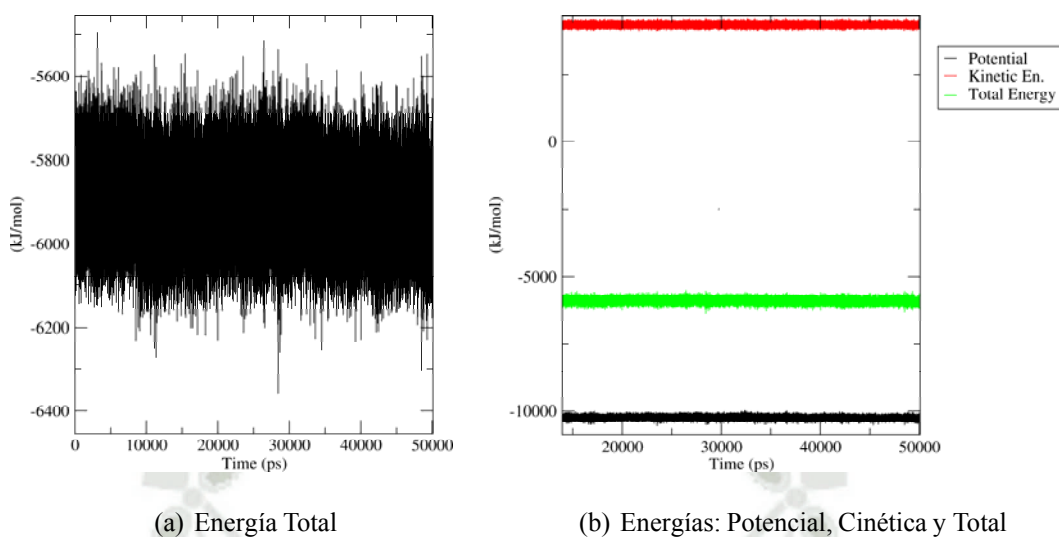


Figura 4.3: Gráficos de energía pertenecientes a la proteína CusC de *Escherichia Coli*.

Gráfico obtenido con xmgrace

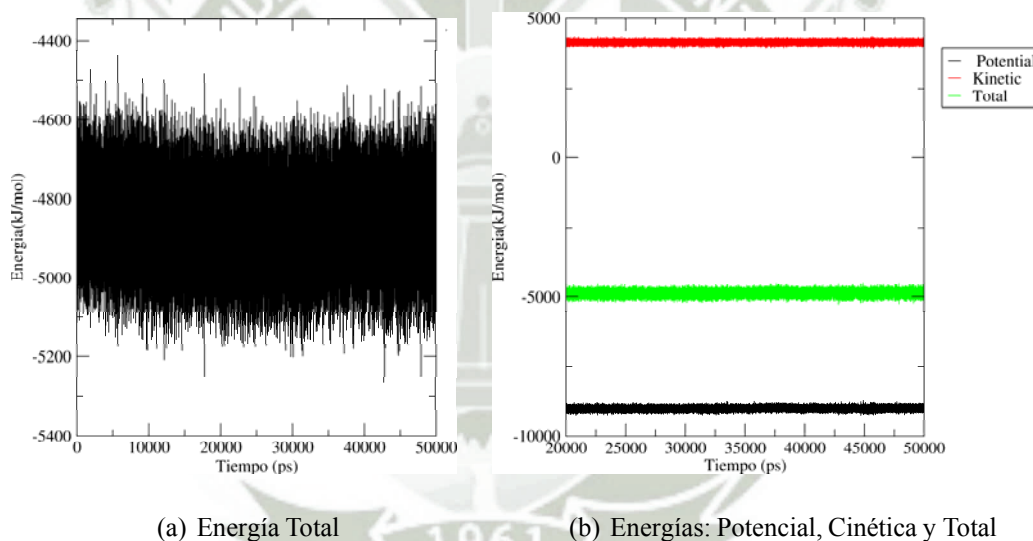


Figura 4.4: Gráficos de energía pertenecientes a la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráfico obtenido con xmgrace

de lo cual la energía total se vuelve constante, los datos a partir de los 20000ps (20ns) son los que se tendrá en cuenta para los posteriores gráficos de la *A. ferrooxidans*, los componentes de la energía total (línea verde) como son: la energía cinética (línea roja) y la energía potencial (línea negra) se pueden apreciar en la figura 4.4-b, en cuyo caso también a partir de los 20000ps (20ns) se puede observar que la estructura logra el

equilibrio.

El gráfico RMSD (root-mean-square deviation) de las posiciones atómicas en una estructura proteica predicha a partir de sus coordenadas nativas es comúnmente reportado para indicar el éxito de la simulación de plegamiento. Además sirve para juzgar la calidad de una estructura proteica, siendo este gráfico el más usado como medida de similaridad geométrica. Para tener una referencia de los valores que son aceptables para el RMSD tenemos que; para proteínas cortas (aprox. 150 aa) la dispersión de RMSD menor que 3Å ($0,3\text{nm}$) son consideradas muy buenas predicciones, valores hasta de $4,5\text{Å}$ ($0,45\text{nm}$) son consideradas aceptables y útiles y predicciones con desviaciones por encima de 5Å ($0,5\text{nm}$) son consideradas solo como informativas.⁸⁰

El gráfico de RMSF (root-mean-square fluctuation) por residuos, nos sirve para indicar las zonas dentro de la estructura donde se presenta mayor movimiento de los aminoácidos durante el tiempo de la dinámica molecular.

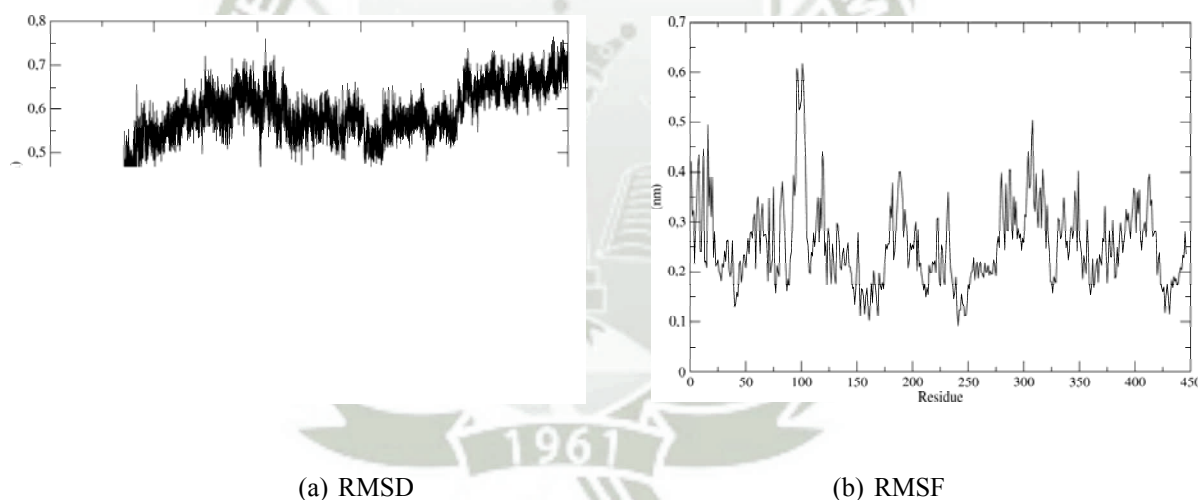


Figura 4.5: Gráficos pertenecientes a la proteína CusC de Escherichia coli. Gráficos obtenidos con xmgrace

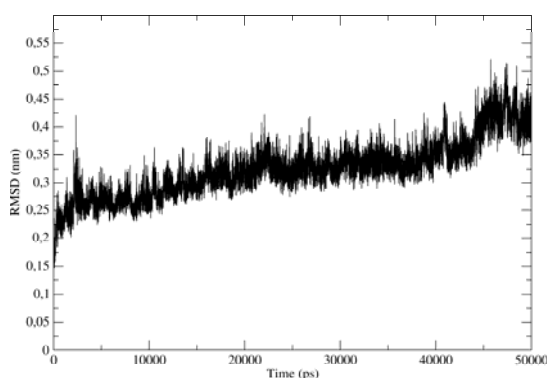
En el gráfico de RMSD de E. coli ilustrado en la figura 4.5-a se puede ver una amplia dispersión hasta los primeros 14000ps (14ns), esto significa que la proteína está en un proceso de plegamiento donde hay bastante movimiento de sus átomos y enlaces después de este tiempo se tiene una desviación de $0,2\text{nm}$ (2Å), lo que indica que se tiene una estructura estable.

En el gráfico RMSF por residuos de *E. coli* que se muestra en la figura 4.5-b podemos ver que la región entre los 75 a 100 aa es la que presenta un mayor movimiento ya que en el gráfico observamos un gran pico de que alcanza el valor de 0,6nm, esta región en la estructura corresponde a una zona que comprende parte de una cadena α hélice (ARG-75 a ARG-81) y la primera hoja β de la estructura (GLN-84 a SER-94) los residuos entre estas estructuras forman cuerdas, también observamos otro pico de 0,5nm que abarca la región de los 275 hasta los 325 aa esto comprende una pequeña porción de una cadena α hélice (ALA-275 a PHE-287) y una zona conformada por dos hojas β (SER-290 a SER-301 y GLY-313 a GLU-322), los residuos entre estas estructuras secundarias forman cuerdas y zonas de giros.

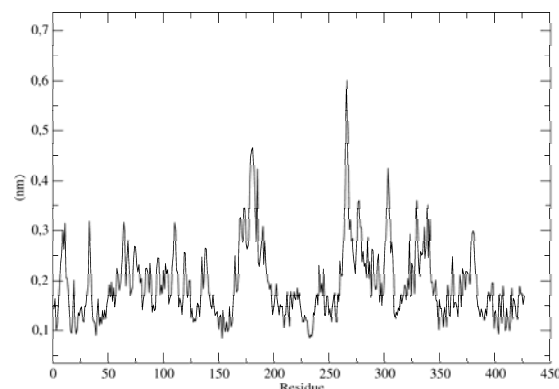
Para el gráfico de RMSD de *A. ferrooxidans* ilustrado en la figura 4.6-a se puede ver que la proteína logrará estabilizarse a partir de los 20000ps (20ns) teniendo valores de desviación de 0,15nm (1,5Å), esto corrobora los resultados de energía donde a partir de los 20000ps (20ns) la proteína ya se encuentra estable.

Por otro lado en el gráfico RMSF por residuos de *A. ferrooxidans* ubicado en la figura 4.5-b podemos observar que hay dos grandes picos, el primero es de 0,46nm abarcando una región entre los 150 a 200 aa que corresponde a una zona de torsión que forma dos cadenas α hélices, la primera va desde ALA-150 a SER-180 luego viene un bucle y la segunda cadena α hélice de GLN-185 a ASN-200. El otro pico que es el más alto, alcanza los 0,6nm este comprende entre los 250 a 300 aa y la zonas que corresponde en la estructura son las siguientes; una cadena α hélice (GLY-250 a VAL-270), a una hoja β (GLU-278 a ASP-288) y parte de la hoja β siguiente (ASN-295 a GLY-300).

Los puentes de hidrógeno desempeñan un papel vital en la estabilización de las estructuras tridimensionales de las moléculas ya que la conformación más estable es la que permite la formación del máximo número de puentes de hidrógeno dentro de la proteína, por eso que monitorear el número de estos enlaces durante la dinámica molecular es importante con este gráfico podemos saber el número de puentes de



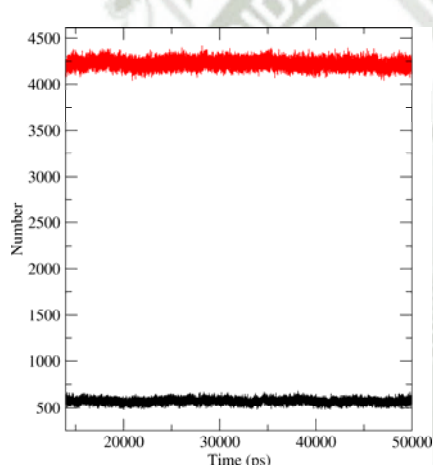
(a) RMSD



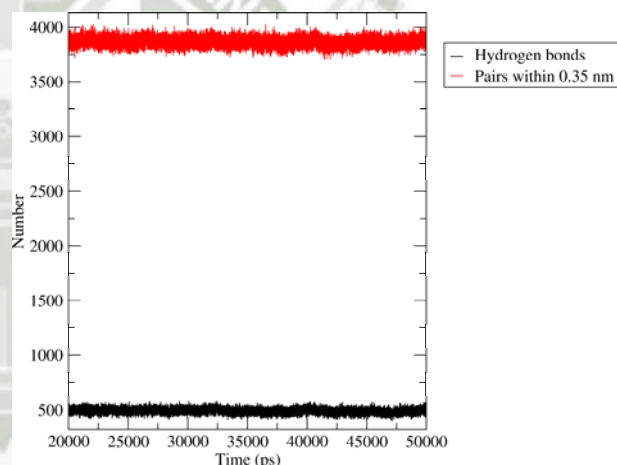
(b) RMSF

Figura 4.6: Gráficos pertenecientes a la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráficos obtenidos con xmgrace.

hidrógeno que se formaron en cada proteína y observar si estos permanecen constantes, disminuyen o aumentan durante la dinámica.



(a) *Escherichia Coli*



(b) *Acidithiobacillus ferrooxidans*

Figura 4.7: Gráficos de los puentes de hidrógeno en ambas proteínas CusC. Gráficos obtenidos con xmgrace.

En ambas gráficas de la figura 4.7 se observa los puentes de hidrógeno transitorios (línea roja) y los puentes de hidrógeno permanentes (línea negra), en el caso de los puentes de hidrógeno transitorios este enlace se establece cuando la separación donora-aceptor es menor que la distancia límite 0,35nm (3,5Å). Podemos ver en la figura 4.7, que tanto en *E. coli* como *A. ferrooxidans* los puentes de hidrógeno permanentes como transitorios se encuentran en una fluctuación constante indicando la estabilidad

de las estructuras terciarias. Para *E. coli* que es la figura 4.7-a, en el cálculo de puentes de hidrógeno se encontró 695 donores y 2616 aceptores, teniendo un promedio de 573.726 puentes de hidrógeno permanentes. En el cálculo de puentes de hidrógeno de *A. ferrooxidans*, mostrado en la figura 4.7-b tenemos como resultado 630 donores y 2412 aceptores, con un promedio de 489.028 puentes de hidrógeno permanentes.

Luego de hacer todos los análisis que nos sirven para evaluar la estabilidad de nuestras proteínas (monómeros), hemos procedido a obtener las estructuras promedio de las proteínas CusC de *E. coli* y *A. ferrooxidans* las cuales se muestran en la Figura 4.8.

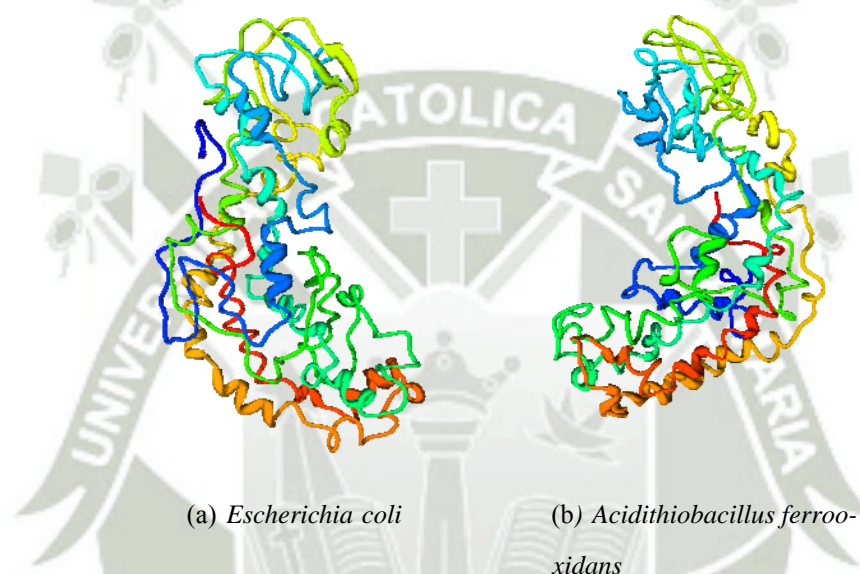


Figura 4.8: Estructuras promedio de las proteínas CusC. Foto obtenida con Chimera 1.6.1

Estas estructuras promedio corresponden a las estructuras terciarias, las cuales describen el plegamiento de la cadena polipeptídica para ensamblar los diferentes elementos de la estructura secundaria en un ordenamiento particular.

Podemos apreciar en la figura 4.8 las proteínas CusC de *E. coli* y *A. ferrooxidans* plegadas, estas proteínas plegadas provienen de un estado de energía alta y finalmente llegan a un estado de energía baja. La estabilidad de nuestras estructuras proteicas se debe generalmente a fuerzas no covalentes como los puentes de hidrógeno, además intervienen fuerzas covalentes como enlaces bisulfuro, enlaces por fuerza de van der Waals; sin embargo las fuerzas no covalentes son más fuertes al ser más numerosas.

Estas fuerzas que actúan para estabilizar las proteínas, puede dejar la falsa impresión de que las proteínas tienen estructuras fijas y rígidas, al contrario las proteínas son moléculas flexibles y fluctuantes, cuyas moviidades estructurales son funcionalmente significativas.

Para corroborar los datos de estabilidad de las estructuras promedio de las proteínas, generamos los Diagramas de Ramachandran, el cual nos brindará información de las regiones permitidas y favorecidas de la estructura, las cuales se hallan relacionadas con la estabilidad estructural y por tanto con la existencia misma de la estructura estudiada.

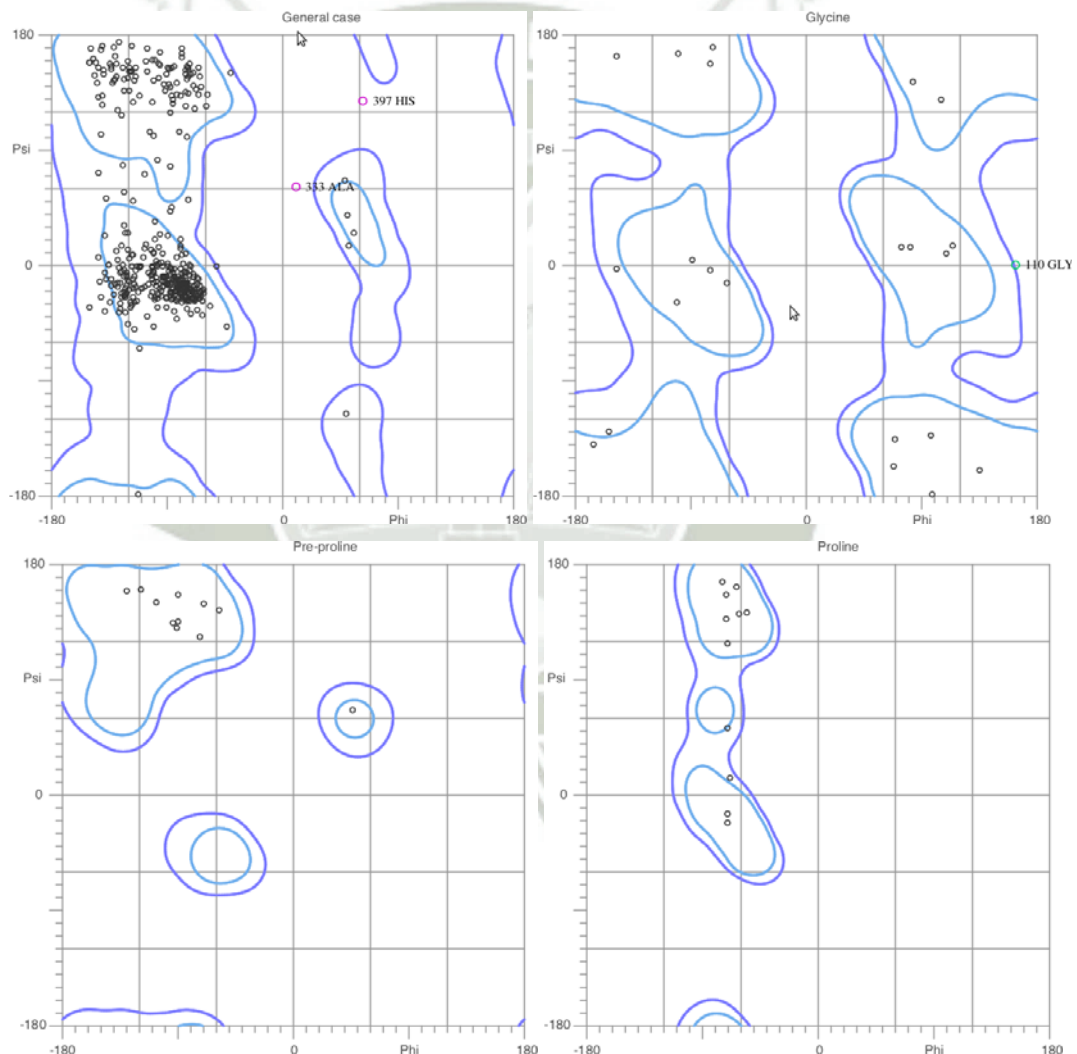


Figura 4.9: Ploteo de Ramachandran de la proteína CusC de *Escherichia coli*. Gráfico obtenido con Molprobit

Para el Diagrama de Ramachandran de *Escherichia coli* ilustrado en la figura 4.9, observamos que está dividido en cuatro cuadrantes, comenzaremos por explicar el primero. Este cuadrante representa el caso general de los aminoácidos indicándonos que de los 446 aa, el 92,8 % de todos los residuos se encuentran en regiones favorecidas, este valor correspondería a 414 aa y 99,3 % de todos los residuos se encuentran en regiones permitidas, esto sería 442 aa; podemos ver que los únicos residuos que no se encuentran en estas zonas son ALA-333 y HIS-397. En el segundo cuadrante podemos observar que corresponde al análisis de las glicinas, en este cuadro podemos ver que una glicina no se encuentra en la zona favorecidas ni en la zona permitida, es la GLY-110, este análisis de las glicinas se hace porque este aminoácido presenta zonas de posible conformación con una simetría central, mucho más amplias que las de los demás aminoácidos, debido a que tiene solo Hidrógeno como la cadena lateral. La glicina es el único residuo sin un átomo de C_{β} , por lo que está mucho menos impedido desde el punto de vista estérico que los otros residuos aminoacídicos. La glicina con frecuencia ocupa posiciones en las que un esqueleto polipeptídico hace un giro pronunciado, que con cualquier otro residuo estaría sujeto a interferencia estérica⁷⁸. Adicionalmente, podemos asociar las glicinas como puntos susceptibles de ataques nucleofílicos, lo cual estaría asociado con reacciones de Sustitución Nucleofílica tanto de primer y segundo orden.

En el tercer cuadro, el análisis de las pre-prolinas nos indica que no tenemos ninguna ellas fuera de las zonas permitidas o favorecidas, cuando se realiza este análisis, tenemos que tener presente que los aminoácidos considerados en este caso, son aquellos de cadena lateral larga que pudieran sufrir procesos de ciclación, generando prolinas sustituidas. Por último, el análisis de las prolinas, donde podemos apreciar que ninguna prolina se halla fuera de las zonas permitidas o favorecidas, este último análisis es porque debido a su cadena lateral cíclica de la prolina, es el residuo con mayores restricciones de conformación.

A continuación, describiremos las posiciones de los aminoácidos que resultaron fuera de las zonas permitidas y favorecidas en el diagrama de Ramachandran. En la

estructura secundaria, la GLY-110 se ubica en una hoja β , la ALA-333 se ubica en una α hélice y la HIS-397 se ubica en una zona de giro entre dos α hélices, luego de la dinámica está estructura se pliega y la estructura anterior cambia por lo que ahora, la GLY-110 se ubica en una zona de curvatura, mientras que ALA-333 y HIS-397 se encuentran cercanas a curvaturas.

En el Diagrama de Ramachandran de *A. ferrooxidans* que se muestra en la figura 4.10, en el caso general tenemos los siguientes resultados: de los 427 aa el 90,6 % de todos los residuos se encuentran en regiones favorecidas, este valor corresponde a un total de 387 aa, mientras que 419 aa del total que representa el 98,3 % de todos los residuos se encuentran en regiones permitidas. En este cuadrante los aminoácidos que no encuentran en las zonas favorecidas ni permitidas son; GLN-185, ALA-186, ASN-221, ARG-330 y ILE-339. Estos resultados nos indican que nuestras estructuras son estables y que existen en la naturaleza. En los análisis de Glicinas y Prolinas, observamos que no hay residuos fuera de las zonas favorecidas ni permitidas. Y en el análisis de las pre-prolinas son tres los aminoácidos que se ubican en zonas no favorecidas ni permitidas, estos son los siguientes; THR-10, ALA-35, PHE-108. Estos aminoácidos pueden sufrir procesos de ciclación, generando prolinas sustituidas.

Describiremos las posiciones de los aminoácidos que se encuentran fuera de las zonas favorecidas y permitidas. En la estructura secundaria, THR-10 se encuentra en al inicio de la proteína (cola), ALA-35 se encuentra en un giro entre dos α hélices, PHE-108 está ubicada en un giro entre un α hélice y una hoja β , GLN-185, ALA-186, ARG-330 y ILE-339 están conformando α hélices y ASN-221 al principio de una α hélice pero sin formar parte de está estructura. Luego que se desarrolló la dinámica molecular y la proteína se plegó la ubicación de los aminoácidos es la siguiente: THR-10, ALA-35 PHE-108 GLN-185, ALA-186, ASN-221 y ILE-339 se encuentran en zonas de giro o torsión y ARG-330 se encuentra en una zona cercana a una α hélice.

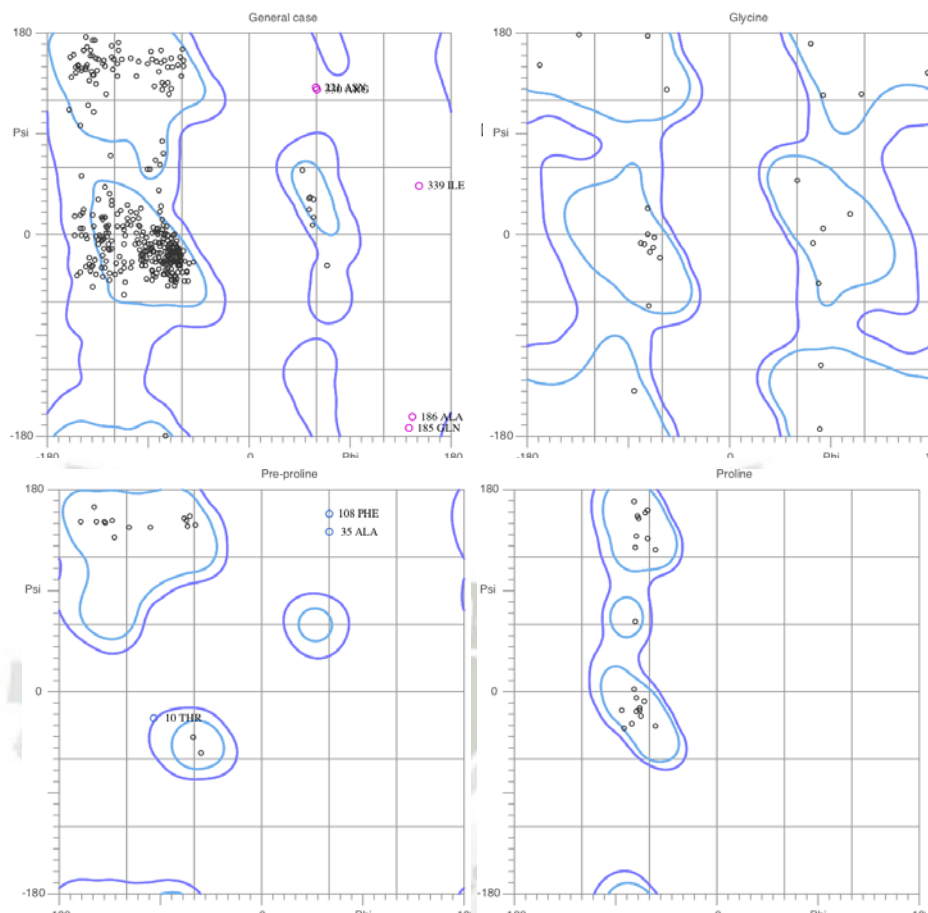


Figura 4.10: Ploteo de Ramachandran de la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráfico obtenido con Molprobit

4.2. Trímeros de las proteínas CusC de *E. coli* y *A. ferrooxidans*

4.2.1. Construcción de los trímeros

Para la formación de los trímeros de las proteínas CusC tanto de *E. coli* y *A. ferrooxidans* se desarrolló un programa usando el lenguaje de programación Fortran; para aplicar este programa partimos de las estructuras secundarias minimizadas y en formato .gro las cuales ubicamos en una caja ajustada al tamaño de cada proteína. Luego de aplicar el programa se obtienen las estructuras mostradas en las figuras 4.11 y 4.12; estos archivos posteriormente fueron preparados para la minimización y para la

dinámica molecular.

En las figuras 4.11 y 4.12 podemos apreciar las estructuras de partida para la minimización y dinámica molecular de cada trímero, en ambas imágenes se muestra tanto una vista frontal como superior de los trímeros, cada monómero está representado por un color diferente: cadena-A azul, cadena-B verde y cadena-C roja. A continuación, describiremos las características de estos trímeros, empezaremos diciendo que tienen una simetría C_{3V} , con un eje principal del tipo $3V$, este nos indica que en giros de 120° se reproduce la estructura, es decir que sus subunidades se relacionan en un solo eje de rotación. Podemos apreciar que tienen forma un túnel esto se debe a las estructuras secundarias α helicoidales-barril, además en la parte superior podemos apreciar un barril β , en el cual cada monómero aporta 4 hojas β , teniendo un total de 12, es por esta estructura que la proteína está anclada a la membrana externa. A la fecha, todas las proteínas factor de membrana externo (OMF) que han sido caracterizadas estructuralmente son β barriles de 8 a 22 hojas β las cuales está insertadas en la membrana⁴³.

El trímero de *Escherichia Coli* está formado por 1338 aa, tiene un diámetro entre $33,82\text{\AA}$ a $37,75\text{\AA}$ y un largo de 134\AA , por cristalografía se sabe que experimentalmente esta proteína tiene un diámetro de aproximadamente 30\AA ⁴⁰. Por otro lado, el trímero de *Acidithiobacillus ferrooxidans* consta de 1281 aa, posee un diámetro entre 36\AA a 44\AA y tiene un largo de $154,02\text{\AA}$ (las medidas se hicieron usando molden 5.0).

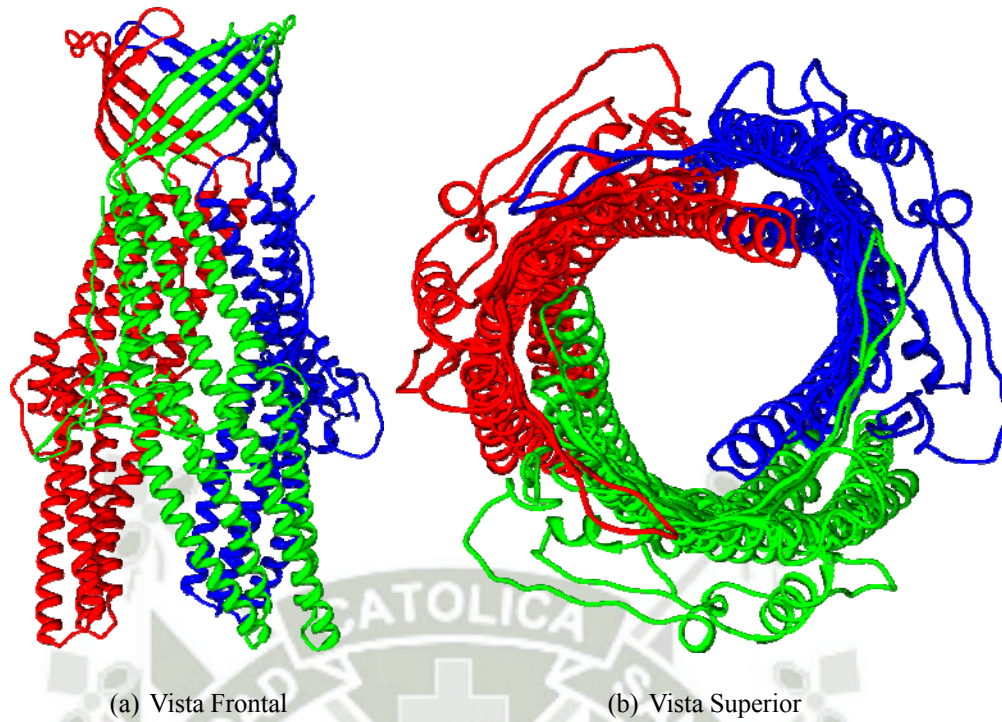


Figura 4.11: Trímero de la proteína CusC de *Escherichia Coli*. Gráfico obtenido con Chimera 1.6.0

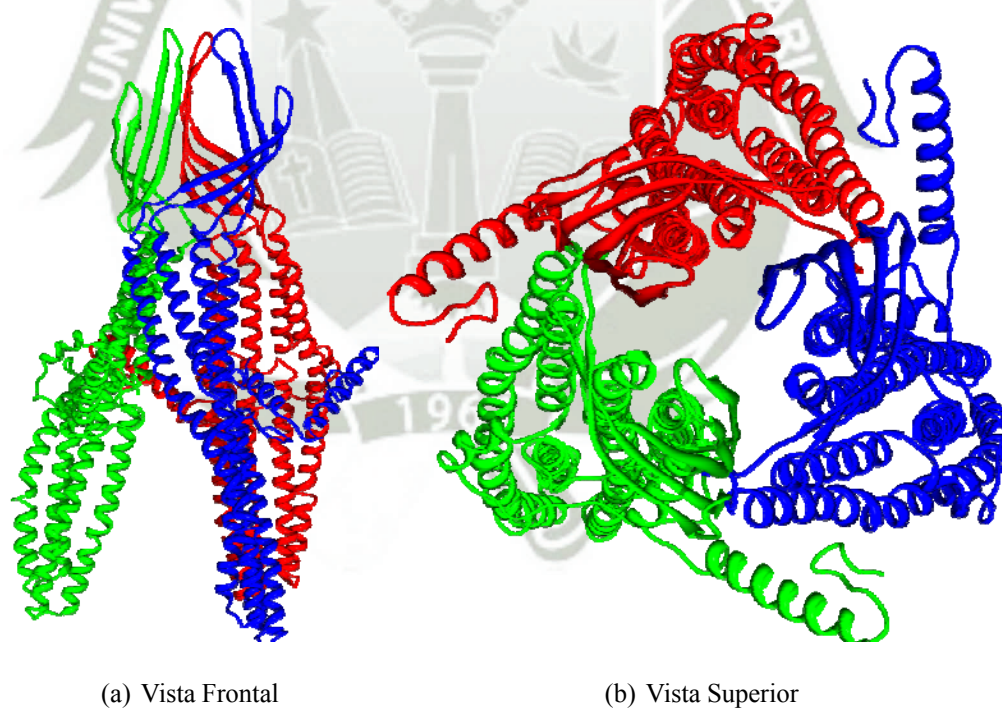


Figura 4.12: Trímero de la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráfico obtenido con Chimera 1.6.0

4.2.2. Minimización de las estructuras

La minimización se realizó con los mismos parámetros anteriores; primero con el método steep descent y luego con el método quasi-newton l-bfgs y el número de pasos fue de 100000 para ambos casos. Se realizaron varias minimizaciones para asegurarnos que las moléculas estén en un mínimo local.

4.2.3. Dinámica Molecular de estructuras

Para la dinámica molecular se trabajó con el campo de fuerza AMBER99SB-ILDN, en un ensamble canónico (N V T), manteniendo una temperatura de 303.15 °K. La dinámica molecular duró 30000ps (30ns) para ambas moléculas. Se escoge el campo de fuerza AMBER ya que como se indicó anteriormente este campo sobreestima los puentes de hidrógeno, lo cual nos interesa porque precisamente son las fuerzas no covalentes, las fuerzas estabilizadoras de nuestros trímeros.

4.2.4. Análisis de la estabilidad

Al igual que en las anteriores dinámicas para evaluar la estabilidad de los trímeros necesitamos de los siguientes gráficos: energía total, gráfico de los componentes de la energía, RMSD, RMSF y puentes de hidrógeno. A continuación comenzaremos por el gráfico de energía.

En el gráfico de energía perteneciente a *E. coli* que se muestra en la figura 4.13, la energía total (ver figura 4.13-a) presenta un descenso de energía hasta los 15000ps (15ns) después de este tiempo observamos que la energía se vuelve constante. En la figura 4.13-b, están las gráficas de la energía cinética (línea roja), potencial (línea negra) y total (línea verde), en esta gráfica se confirma que la relajación se logrará a partir de los 15000ps (15ns), por esta razón en los posteriores gráficos solo se mostraran los datos a partir de este tiempo.

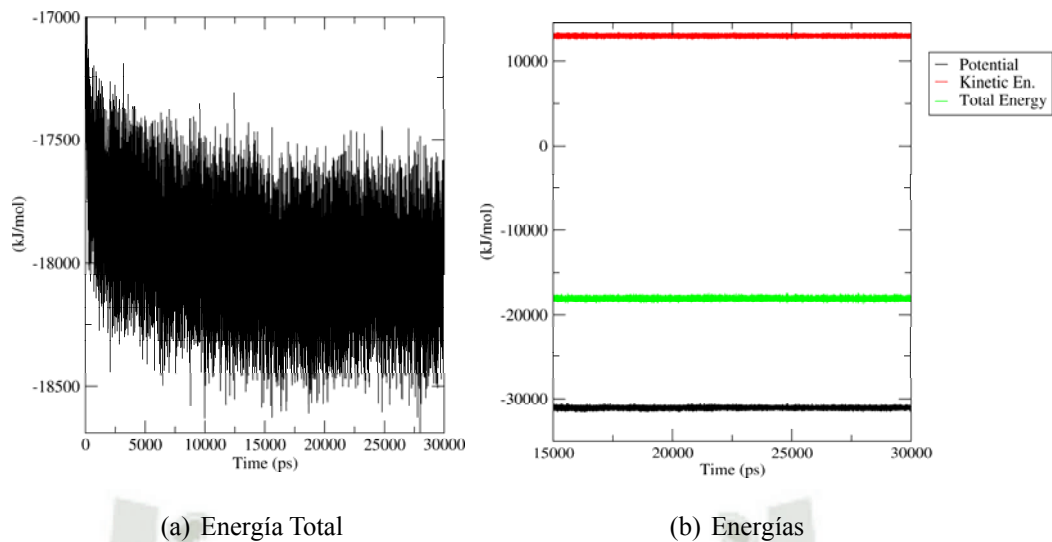


Figura 4.13: Gráficos de energía pertenecientes al trímero de la proteína CusC de *Escherichia Coli*. Gráfico obtenido con xmgrace.

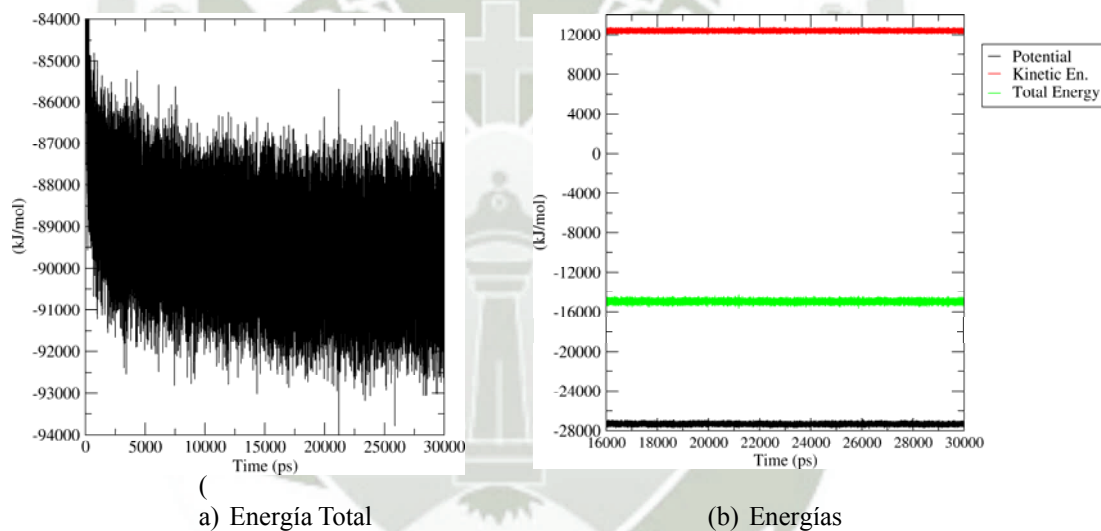
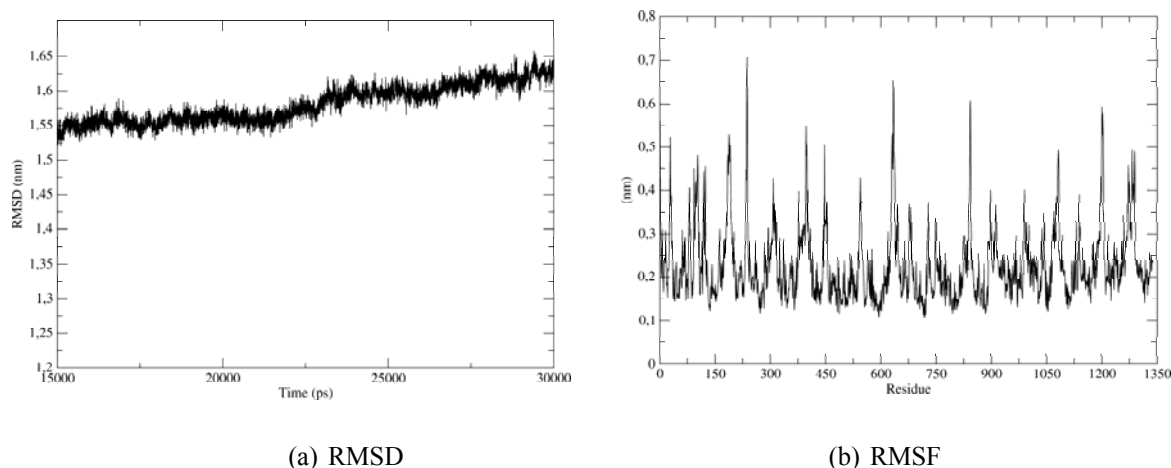


Figura 4.14: Gráficos de energía pertenecientes al trímero de la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráfico obtenido con xmgrace.

En los resultados de energía para *A. ferrooxidans*, que se observan en la figura 4.14-a, la energía total presenta una disminución hasta los 16000ps (16ns), siendo continua en el tiempo restante de la dinámica molecular, en la figura 4.14-b podemos ver la energía cinética (línea roja), la energía potencial (línea negra) y total (línea verde) donde también podemos apreciar que a partir de los 16000ps (16ns) se logró el equilibrio, por lo tanto los siguiente gráficos para esta estructura mostrarán los datos a partir de este tiempo.



(a) RMSD

(b) RMSF

Figura 4.15: Gráficos pertenecientes al trímero de la proteína CusC de *Escherichia Coli*. Gráficos obtenidos con xmgrace.

En el gráfico de RMSD de *E. coli*, que se muestra en la figura 4.15-a se puede ver que desde los 15000ps (15ns) se tiene una desviación constante, la cual tiene un valor de 0,125nm (1,5Å), esto nos indica que nuestra estructura se encuentra estable.

En el gráfico RMSF por residuos de *E. coli* que se muestra en la figura 4.15-b, podemos ver en una primera zona que corresponde a la cadena A de 0 a 446 aa, un pico de 0,7nm que es el más elevado y abarca desde los 225 hasta los 262 aa, está zona corresponde a una cuerda que consiste de giros y α hélices formadas en los siguientes aminoácidos: LEU-225 a LEU-229, SER-243 a SER-245 y SER-256 a GLN-262. En la segunda zona que es la cadena B de 447 a 892 aa tenemos dos picos, uno de 0,66nm que incluye desde los 600 a 675 aa, esto corresponde a dos cadenas α hélices, una desde GLN-600 a GLY-630 y consta de 8 giros, luego viene un bucle y seguida la otra cadena de α hélice, de VAL-634 a GLY-670 (forma 10 giros); el segundo pico en esta cadena tiene un valor de 0,62nm que va desde los 820 a 858 aa, corresponde también a cadenas de α hélices, de TYR-820 a GLN-836 (5 giros) y TYR-842 a THR-858 (5 giros), en el medio de estas dos cadenas encontramos un bucle. Por último tenemos a la cadena C, que va desde los 893 hasta los 1338 aa presenta un pico de 0,6nm que abarca desde los 1163 a 1217 aa, corresponde a dos hojas β antiparalelas, formadas entre los siguientes aminoácidos: SER-1170 a SER-1181 y GLY-1193 a GLU-1202; los aminoácidos no mencionados forman cuerdas.

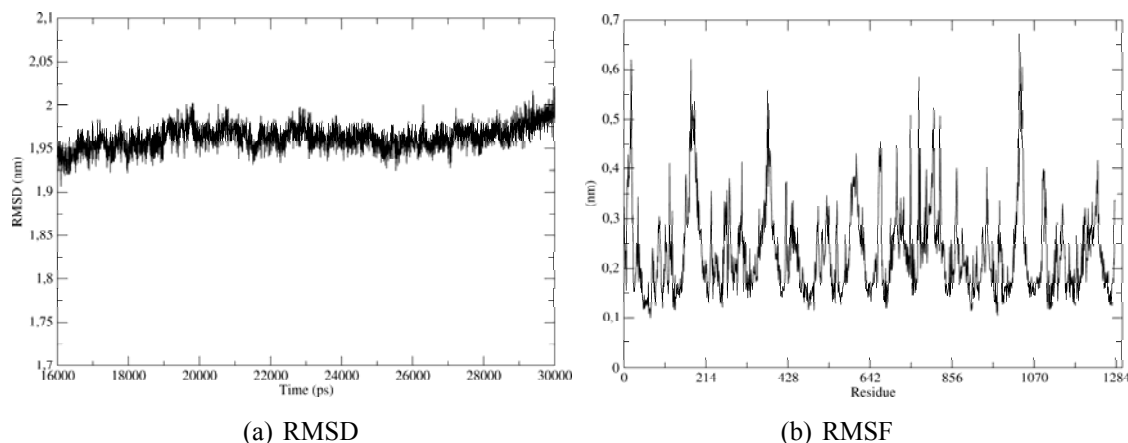


Figura 4.16: Gráficos pertenecientes a la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráficos obtenidos con xmgrace.

En el gráfico de RMSD de *A. ferrooxidans*, ilustrado en la figura 4.16-a se puede ver los datos de RMSD desde los 16000ps (16ns) en este período de tiempo vemos que la desviación es constante, la cual tiene un valor de 0,075nm (0,75Å), estos datos confirman los resultados de energía que este tiempo la estructura ya se encuentra estable.

En el gráfico RMSF por residuos de *A. ferrooxidans* que se muestra en la figura 4.16-b, nos muestra las zonas en la estructura con mayor fluctuación. Las tres zonas que conforman el trímero están distribuidas de la siguiente manera: zona A que va desde 1 a 427 aa, la zona B de 428 a 854 aa y por último la zona C de los 854 a 1281 aa. En la zona A observamos un primer pico de 0,62nm que abarca de 1 a 26 aa, es el comienzo de la proteína que tiene como estructura una cadena α hélice de ALA-13 a LEU-26 (4 giros) y el resto es una cuerda, y el segundo pico de 0,62nm que va desde 135 a 214 aa, acá encontramos 2 cadenas α hélices en la parte inferior de la estructura, una desde GLN-135 a TYR-178 (12 giros) y la otra de GLN-185 a ALA-214 (9 giros), ambas cadenas unidas por un bucle; por último tenemos un pico de 0,56nm que va desde los 331 a los 427 aa, que corresponde a otras dos cadenas α hélices, que van de TYR-331 a VAL-380 (13 giros) y LEU-385 a PHE-418 (9 giros) unidas también por un bucle, la última parte corresponde a una cuerda. En la zona B tenemos un pico de 0,58nm esta región abarca desde los 749 a los 854 aa, acá encontramos dos cadenas α

hélices ubicadas en la parte inferior ubicadas entre los siguientes aminoácidos: ALA-742 a TYR-805 (15 giros) y LEU-812 a PHE-845 (9 giros) y todo esto unido por un bucle; luego de estas estructuras viene una cuerda desde LEU-846 a PRO-854. Por último en la zona C tenemos un pico de 0,68nm, que va desde los 1017 a los 1070 aa, está región está constituida por dos cadenas α hélices, formadas en las siguientes posiciones: GLN-1017 a TYR-1032 (5 giros) y GLN-1038 a ILE-1070 (9 giros), estas dos cadenas están unidas por un bucle.

En las gráficas de la figura 4.17 se tiene los puentes de hidrógeno transitorios (línea roja) y los puentes de hidrógeno permanentes (línea negra). En esta gráfica tanto en *E. coli* como *A. ferrooxidans* los puentes de hidrógeno permanentes como transitorios se encuentran en una fluctuación constante indicando la estabilidad de los trímeros. Para *E. coli* que es la figura 4.7-a, en el cálculo de puentes de hidrógeno se encontró 2058 donores y 7848 aceptores, teniendo un promedio de 1687.987 puentes de hidrógeno permanentes. En el cálculo de puentes de hidrógeno de *A. ferrooxidans*, mostrado en la figura 4.7-b tenemos como resultado 1890 donores y 7236 aceptores, con un promedio de 1524.357 puentes de hidrógeno permanentes.

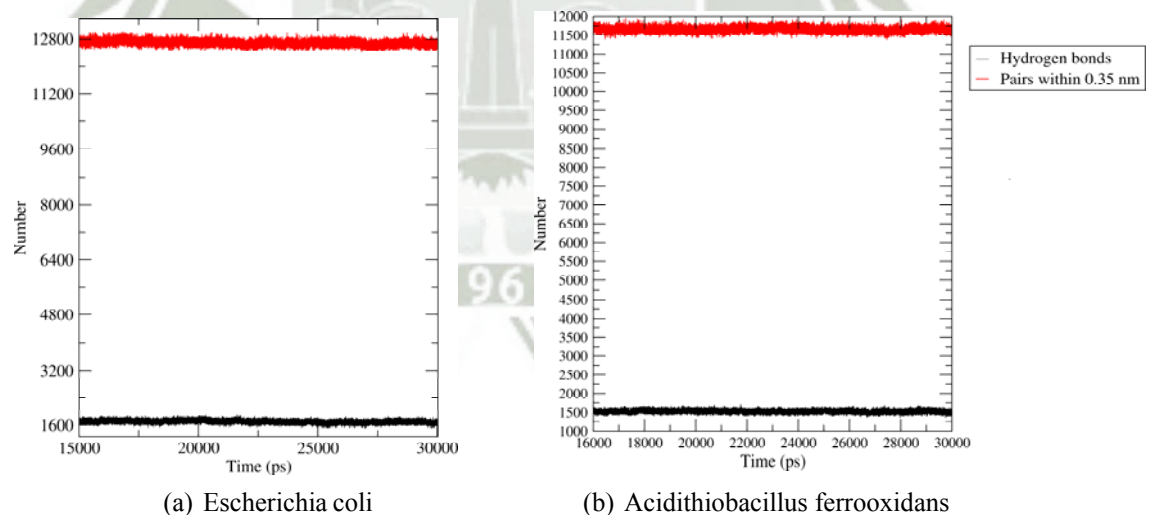


Figura 4.17: Gráficos de los puentes de hidrógeno. Gráficos obtenidos con xmgrace.

Luego del análisis de los gráficos anteriores que nos permiten determinar la estabilidad de las proteínas, hemos procedido a obtener las estructuras promedio de ambos trímeros. Estas estructuras promedios corresponden a las estructuras cuaternarias,

también las podemos denominar estructuras nativas, ya que estas son las estructuras funcionalmente activas, además son las conformaciones más estable termodinámicamente.

La formación de las estructuras cuaternarias se da porque las superficie de las proteínas son irregulares, esto es lo que permite las uniones y asociaciones específicas con otras proteínas, nuestra estructura cuaternaria es el resultado de la unión de tres monómeros iguales de la proteína CusC, también puede ser llamado homotrímero. Si comparamos la estructura de partida del trímero con las estructuras obtenidas (figura 4.18, 4.19), claramente vemos que el diámetro del trímero disminuyó notablemente. Siendo los valores para el trímero de *Escherichia coli* de 6,84Å y para el trímero de *Acidithiobacillus ferrooxidans* de 2,23Å.

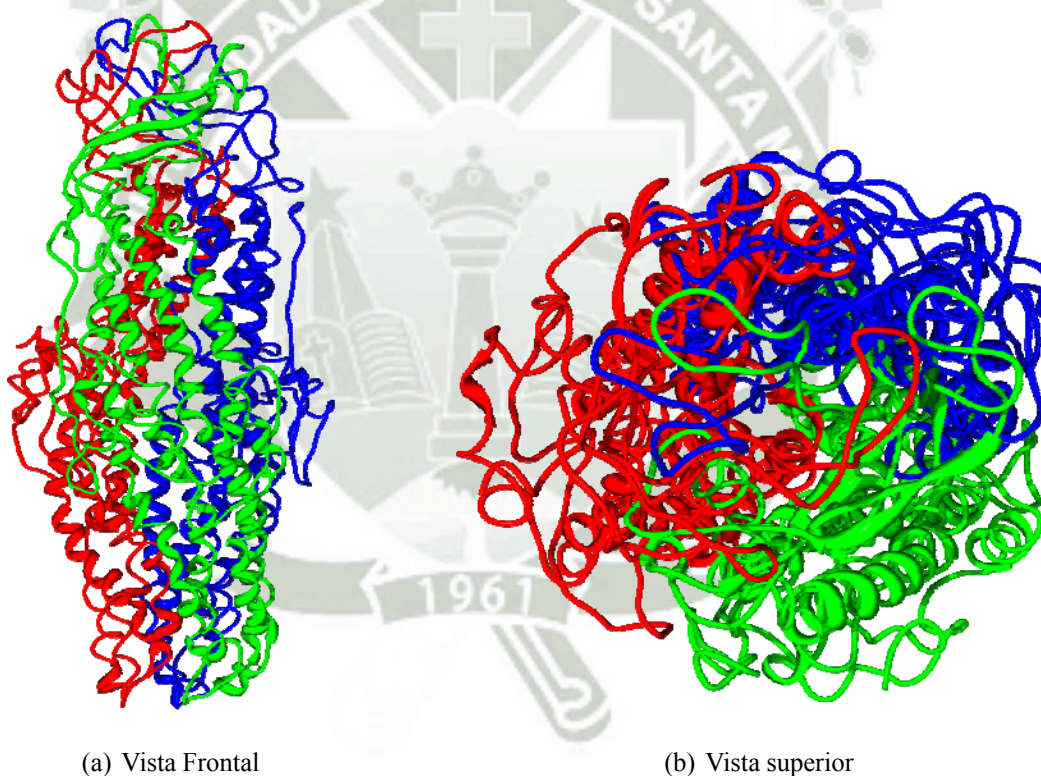


Figura 4.18: Estructura promedio del trímero de la proteína CusC de *Escherichia coli*. Gráficos obtenidos con Chimera 1.6.1.

En el Diagrama de Ramachandran de *E. coli* que se muestra en la figura 4.20, en el caso general tenemos los siguientes resultados: de los 1338 aa el 90,5 % de todos los residuos se encuentran en regiones favorecidas, este valor corresponde a un total

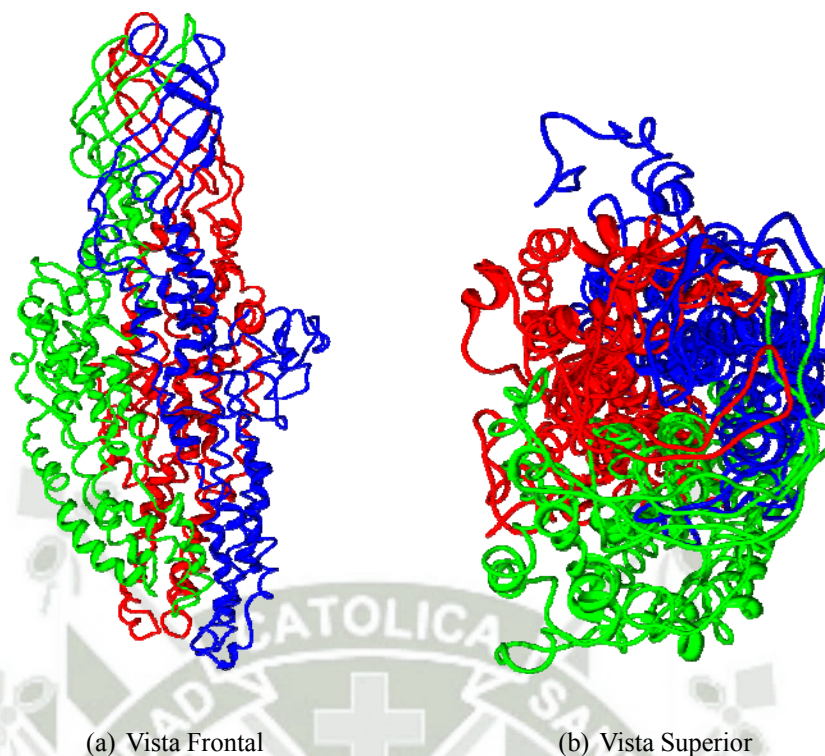


Figura 4.19: Estructura promedio del trimero de la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráficos obtenidos con Chimera 1.6.1.

de 1206 aa, mientras que 1327 aa del total que representa el 99,6 % de todos los residuos se encuentran en regiones permitidas, los residuos que no se encuentran en estas regiones son: THR-109, SER-302, SER-312, ASP-1010. En los análisis de Glicinas tenemos a la GLY-775 fuera de las zonas permitidas y favorecidas. Para las prolinas y pre-prolinas observamos que no hay residuos fuera de las zonas favorecidas ni permitidas.

Describiremos que estructuras conforman los aminoácidos que se encuentran fuera de las zonas permitidas y favorecidas. En la cadena A tenemos: THR-109, el cual forma parte de un hoja β , SER-302 y SER-312, se encuentran en zonas de giro cercanas a hojas β . En la cadena B tenemos a la GLY-775 que se encuentra en una zona de giro que conforma una cuerda. Por último en la cadena C tenemos a ASP-1010 ubicado en una zona de giro entre una α hélice y una hoja β .

En el Diagrama de Ramachandran de *A. ferrooxidans* que se muestra en la figu-

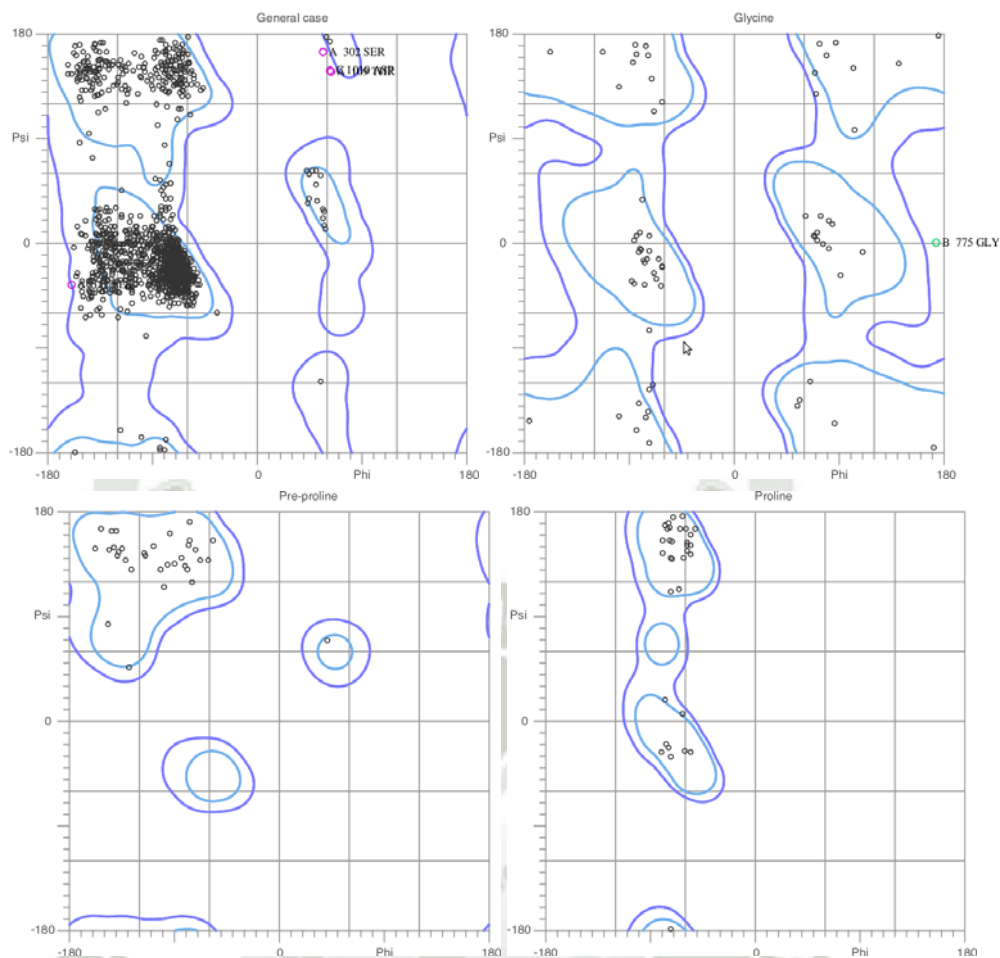


Figura 4.20: Ploteo de Ramachandran de la proteína CusC de *Escherichia coli*. Gráfico obtenido con Molprobit

ra 4.21, en el caso general tenemos los siguientes resultados: de los 1281 aa el 93,3 % de todos los residuos se encuentran en regiones favorecidas, este valor corresponde a un total de 1190 aa, mientras que 1260 aa del total que representa el 98,8 % de todos los residuos se encuentran en regiones permitidas, los residuos que no se encuentran en estas regiones son: LYS-49, LEU-116, GLN-185, ALA-263, ILE-307. En los análisis de Glicinas tenemos tres fuera de las zonas permitidas y favorecidas, estas son: GLY-442, GLY-479, GLY-1128, estas glicinas están predispuestas a sufrir un ataque nucleofílico. Para las prolinas observamos que no hay residuos fuera de las zonas favorecidas ni permitidas. Por último en las pre-prolinas tenemos seis aminoácidos fuera de las zonas permitidas y favorecidas, que corresponden a: TYR-290, SER-293, LEU-

364, THR-437, LEU-500, TYR-1144, LEU-1218, estos últimos residuos son los que pueden sufrir ciclación y formar prolina sustitutorias.

A continuación, describiremos en detalle en que estructuras se localizan los residuos que se encuentran fuera de las zonas permitidas y favorecidas, empezaremos por los que se encuentran en la cadena A; tenemos LYS-49, LEU-116, GLN-185, ALA-263, LEU-364, se encuentran formando α hélices, TYR-290, SER-293, ILE-307, se encuentran en zonas de giro.

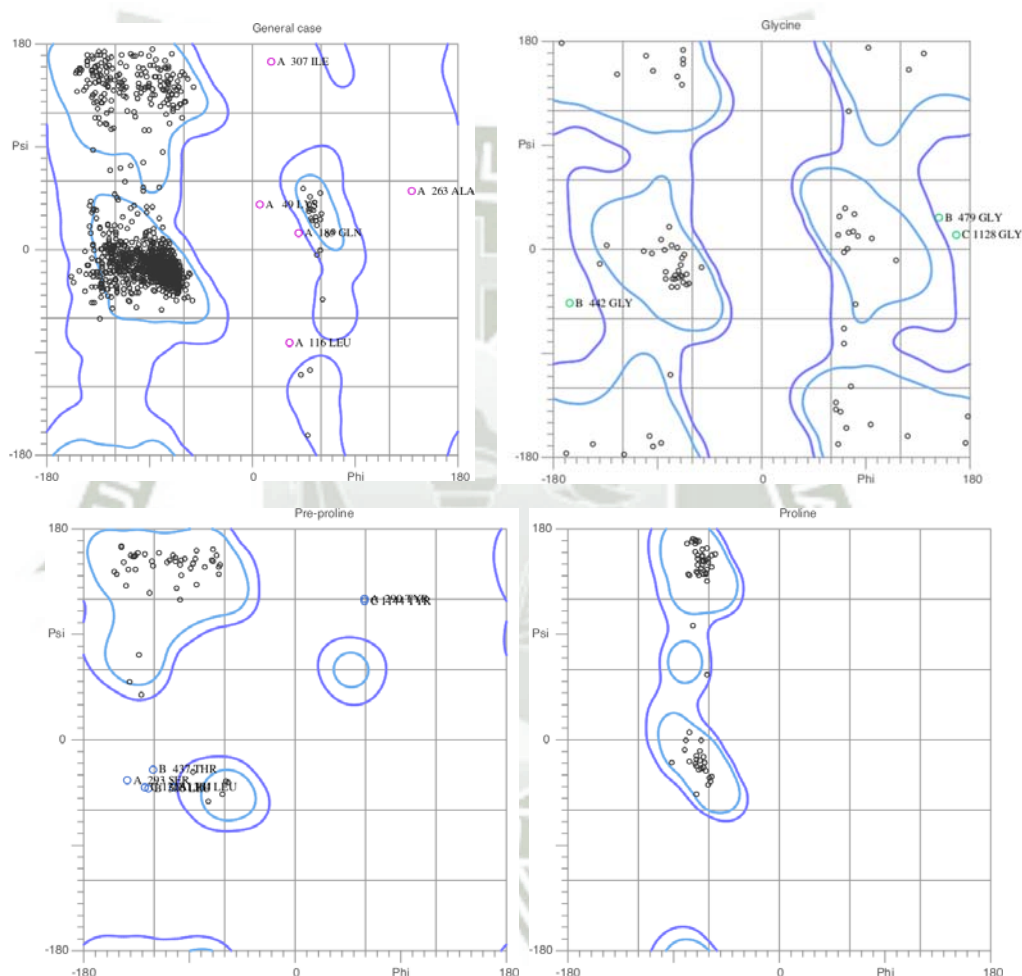


Figura 4.21: Ploteo de Ramachandran de la proteína CusC de *Acidithiobacillus ferrooxidans*. Gráfico obtenido con Molprobit

4.2.5. Potencial electrostático

Después de los ploteos de Ramachandran, necesitamos evaluar el potencial electrostático de las proteínas tanto de los monómeros como los trímeros. Este se calculó usando la aproximación de Poisson Boltzmann implementada en el programa APBS. Partimos de los archivos de las estructura promedios en formato .pdb. Cada uno de estos archivos se transfirió al servidor del programa PDB2PQR (<http://pdb2pqr.sourceforge.net/>); el cual transforma los archivos .pdb de las proteínas a formato .pqr, que contiene información adicional sobre las cargas y los radios de cada uno de los átomos presentes en la molécula proteica y también adiciona hidrógenos a la estructura molecular optimizándolos. Además de este archivo nos proporciona otro archivo más, con extensión .in que contiene las dimensiones de la malla que ajusta a la superficie molecular de la proteína, este archivo considera una temperatura de 298.15K, la cual tenemos que cambiar por la temperatura a la que estamos trabajando que es 303.15K.

El potencial electrostático se ha convertido en los últimos tiempos en una herramienta formidable para el estudio de la reactividad y las interacciones entre moléculas. Este se genera en el espacio entorno a una molécula por causa de sus núcleos y sus electrones. Es importante saber el potencial ya que nos ayuda a identificar las zonas donde puede ocurrir reacciones electrofílicas y nucleofílicas.

Comenzaremos describiendo las figuras 4.22, 4.23 donde observamos las superficies del potencial para los monómeros. Convencionalmente las zonas de color rojo son los potenciales más negativos y las regiones de color azul son las zonas más positivas, en este sentido los lugares neutros o hidrofóbicos se hallaran con potenciales en torno a cero, y estarán visualizadas en color blanco. De acuerdo a esto podemos decir, que para la proteína CusC de *Escherichia coli* que se observa en la figura 4.22, apreciamos que la mayoría de la superficie tiene zonas negativas,⁴⁰ por el contrario, la mayor parte de la superficie de la proteína CusC de *Acidithiobacillus ferrooxidans* es positiva (ver figura 4.23).

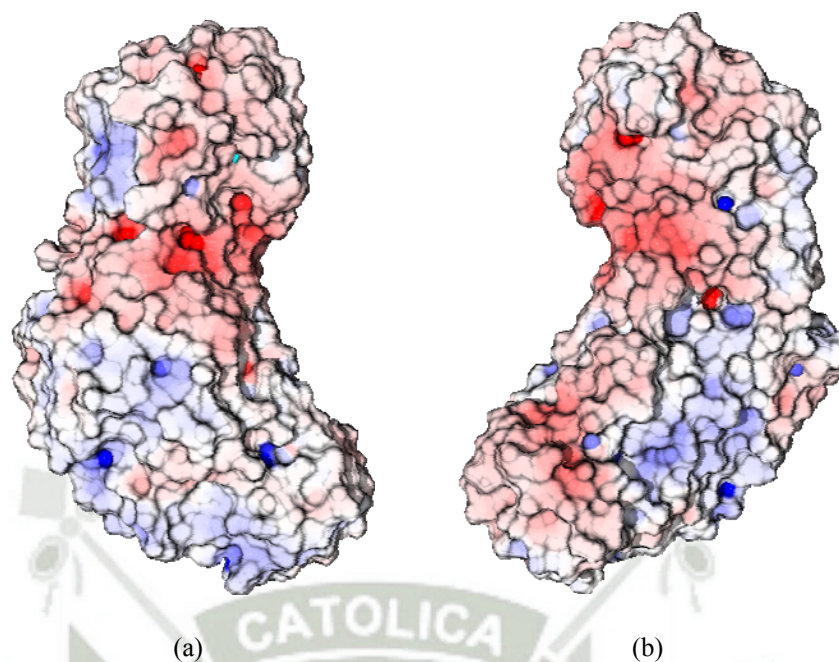


Figura 4.22: Representación de la superficie del potencial electrostático de la proteína CusC perteneciente a *Escherichia coli* en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1

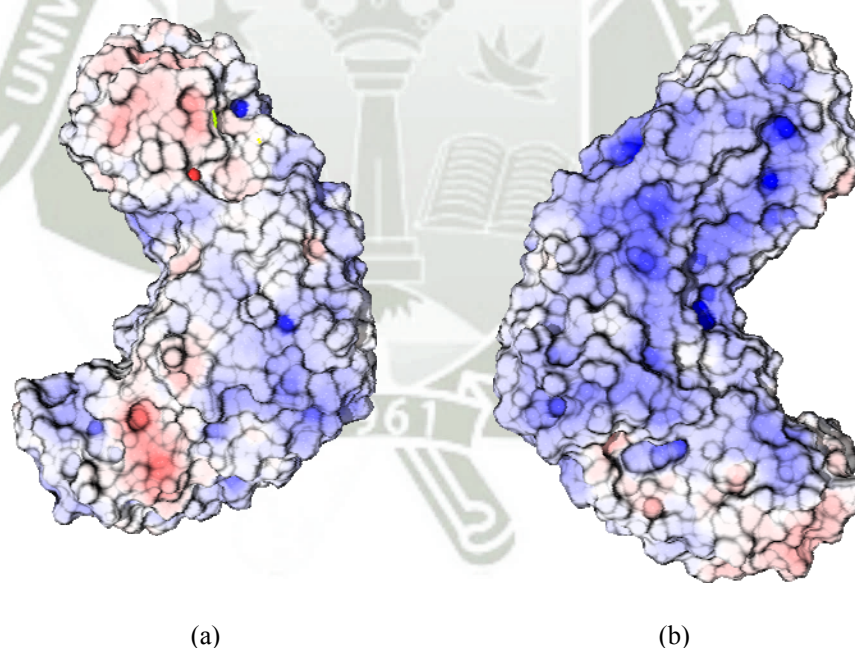


Figura 4.23: Representación de la superficie del potencial electrostático de la proteína CusC perteneciente a *Acidithiobacillus ferrooxidans* en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1

La distribución de los residuos polares y no polares, pone en claro una faceta clave de la arquitectura proteica, en un entorno acuoso, el plegamiento proteico está dirigido por la fuerte tendencia de los residuos hidrofóbicos a ser excluidos del agua. Un sistema es más estable termodinámicamente cuando los grupos hidrofóbicos están agrupados en vez de expuestos al entorno acuoso. Nuestro trímero es una proteína transmembrana y esta clase de proteínas posee una orientación específica en la bicapa lipídica, ambos trímeros (ver figuras 4.24, 4.25) están anclados a esta capa por el barril β , por lo tanto, las cadenas laterales de los aminoácidos expuestos a la parte interna del barril, que además tapizan el canal acuoso, son esencialmente hidrofílicas, mientras que las cadenas expuestas al exterior del barril entran en contacto con el núcleo hidrófobo de la bicapa lipídica y por tanto como observamos en ambos trímeros (ver figuras 4.24, 4.25) esta parte está cargada ligeramente negativa y además posee zonas hidrófobas o apolares.

Siguiendo con el análisis de las superficies de potencial electrostático de ambos trímeros, se observa en la figura 4.24, la superficie del potencial electrostático del trímero de *E coli*, en la región externa tanto como interna de la proteína se puede apreciar que, tiene extensas zonas electronegativas, esto ha generado interesantes preguntas de como los cationes de cobre son transportados sin quedar atrapados dentro del barril. Una posible respuesta que se tiene es debido al diámetro que es demasiado amplio, tanto el valor que se tiene experimentalmente como el calculado después de la dinámica molecular ($37,75\text{Å}^{40}$ y $6,84\text{Å}$ respectivamente), ya que el diámetro de un catión de Cu^{+2} es tan solo de $0,071\text{nm}$.

A diferencia del trímero de *E coli*, el trímero de *A ferrooxidans* que se observa en la figura 4.25 posee tanto en la región exterior como interior extensas zonas electropositivas, que nos habla de la aceleración del catión por efecto del entorno.

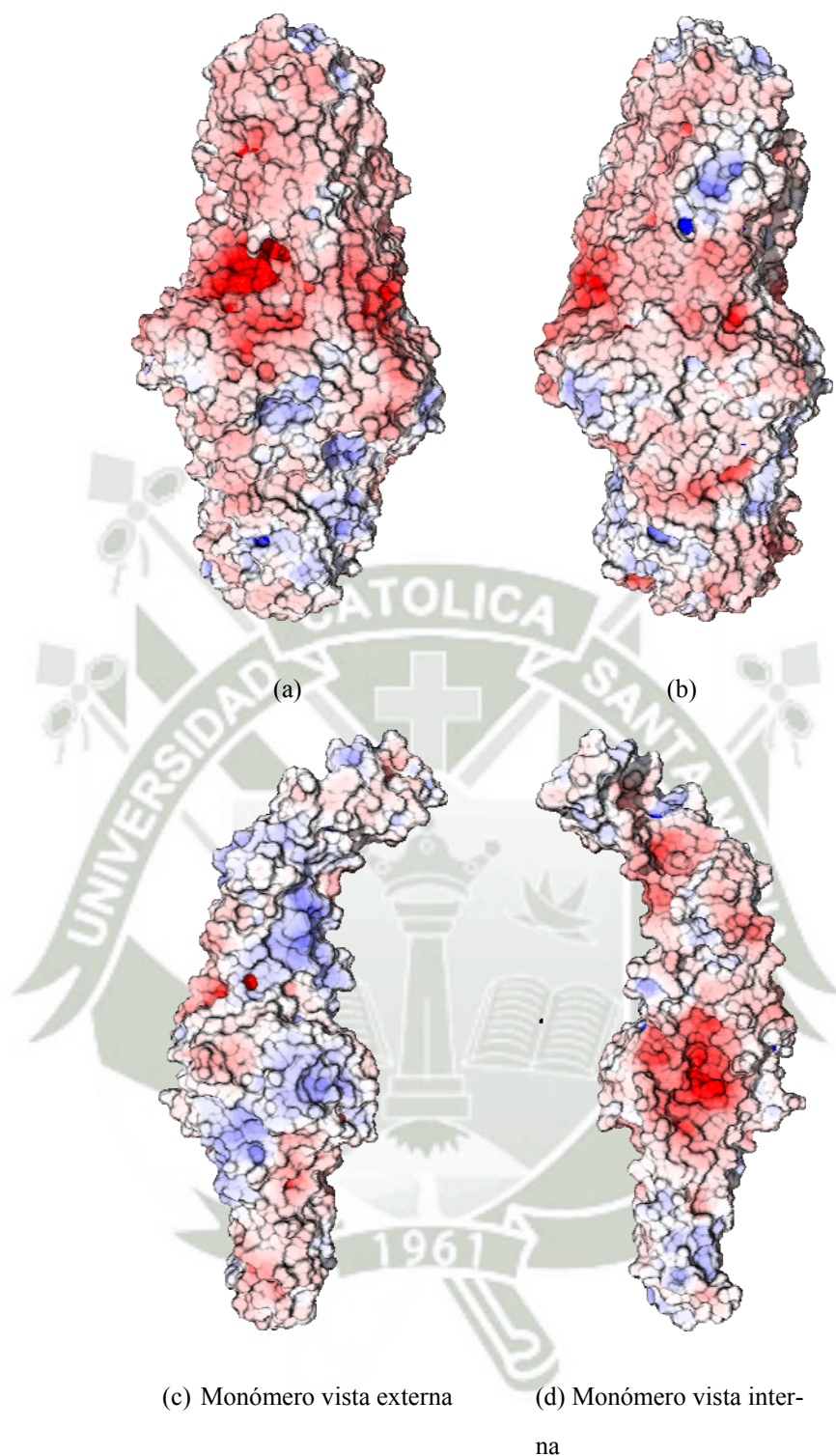


Figura 4.24: Representación de la superficie del potencial electrostático del trímero perteneciente a *Escherichia coli* en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1

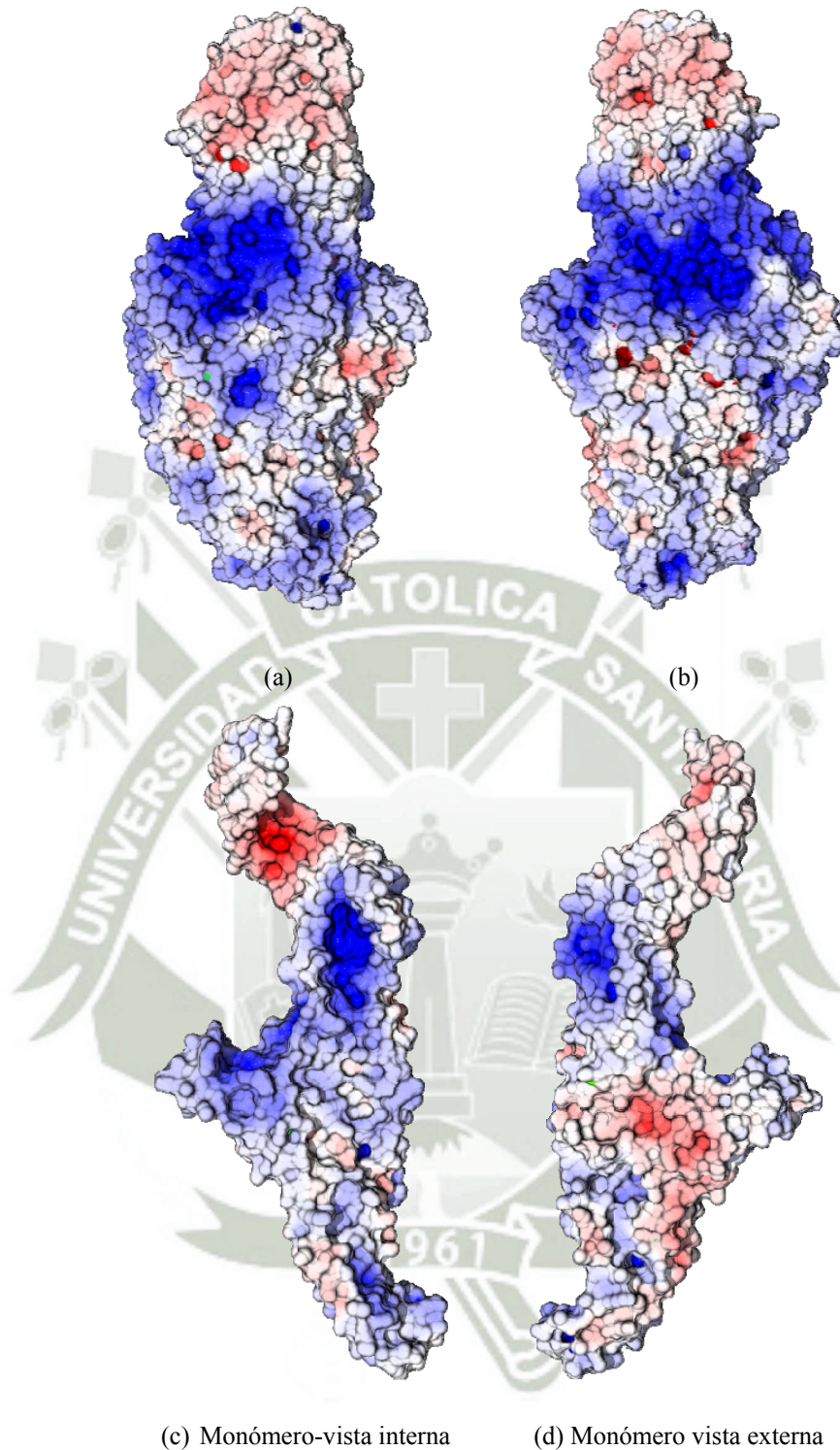


Figura 4.25: Representación de la superficie del potencial electrostático del trímero perteneciente a *Acidithiobacillus ferrooxidans* en donde los colores representan diferentes propiedades de hidrofobicidad. Foto obtenida con Chimera 1.6.1

4.2.6. Parámetros termodinámicos

La conformación nativa de una proteína es la de más baja energía total. El proceso de plegamiento, como cualquier otro proceso biológico, se encuentra bajo control termodinámico o cinético, y es un proceso que está claramente favorecido en condiciones fisiológicas. En otras palabras, el proceso de plegamiento hasta alcanzar la estructura nativa debiera presentar un ΔG negativo, pero en nuestro caso ya que estamos evaluando en el punto de equilibrio bajo el ensamble canónico, el ΔG es igual a cero, este parámetro más adelante será comprobado.

Tabla 4.1: Resultados de las Energías

Proteínas	Energía Total (KJ)	Energía Cinética (KJ)	Energía Potencial (KJ)
Monomero _{af}	-29246.8	24806	-54052.8
Monomero _{ec}	-35433.2	26110.3	-61543.5
Trímero _{af}	-89760.2	74425.4	-164186
Trímero _{ec}	-108236.6	78236.4	-186473

En la tabla 4.1, se muestran la energía total, cinética y potencial de las estructuras estudiadas, estos resultados lo calculamos mediante el programa Gromacs. El valor de energía total nos sirve para calcular los demás parámetros termodinámicos, tenemos que tener en cuenta que trabajamos con un ensamble canónico (NVT), lo primero es saber y tener presente que cuando realizamos una dinámica molecular en este tipo de ensamble (NVT), quiere decir que el estado macroscópico queda fijado al conocer el número de partículas, volumen y temperatura, por lo tanto, podemos decir que la energía total es aproximadamente igual a la energía interna del sistemas como se muestra en la ecuación 4.1.

La energía total es el resultado de la suma de la energía cinética y potencial, esta última es negativa y mucho mayor que la cinética, esto se debe a que la energía potencial es igual al trabajo que se debe hacer frente a esa fuerza (campo de fuerza) para mover la proteína de un punto de referencia, a otra posición (plegamiento). La fuerza que se debe ejercer para este proceso deberá ser igual pero de sentido opuesto, y ello es el origen del signo negativo, además el signo nos indica que la proteína se encuentra en

un pozo de energía, esto es como consecuencia que la fuerza ejercida por el campo de fuerza siempre tiende hacia una energía más baja y actuará para reducir la energía potencial, por esta razón nuestra energía total también es negativa. En la tabla 4.1 podemos observar que para las proteína CusC de *E. coli* tanto el monómero como trímero se alcanzó energías mucho menores que para *A. ferrooxidans*, esto puede deberse a la diferencia de tamaños de ambas proteínas ya que al tener más aminoácidos la proteína CusC de *E. coli* se aplica una mayor fuerza para reducir la energía potencial.

$$\Delta E \approx \Delta U \quad (4.1)$$

A partir de esta aproximación y recordando la relación entre entropía y energía interna que nos proporciona la termodinámica tenemos la ecuación 4.2, de donde podemos hallar la variación en la entropía (ΔS) del sistema, la temperatura de referencia fue 303.15 K.

$$\frac{\Delta S}{\Delta U}_{N,V} = \frac{1}{T} \quad (4.2)$$

Para hallar la entalpía de nuestros sistemas usamos la ecuación 4.3

$$\Delta H = \Delta U + P V \quad (4.3)$$

determinamos que para un sistema en ensamble canónico, $\Delta H = \Delta U$, esto se debe a que la presión es despreciable al ser demasiado pequeña.

Adicionalmente, se obtuvo los valores de la capacidad calorífica a volumen constante (C_v) usando la siguiente ecuación 4.4

$$C_v = \frac{\Delta U}{\Delta T}_v \approx \frac{\Delta U}{T} \quad (4.4)$$

En la tabla 4.2 se muestran los resultados de los cálculos de los parámetros termodinámicos que consideramos importantes en nuestros sistemas.

Desde el punto de vista entrópico, el proceso de plegamiento supone una disminución neta de entropía hasta alcanzar la única estructura nativa y como podemos ver en la tabla 4.2 la variación de la entropía en el proceso de trimerización, es negativa lo

Tabla 4.2: Resultados Parámetros Termodinámicos

Proteínas	Energía Interna ΔU KJ	Entalpía ΔH KJ	Entropía ΔS J/K	Cv KJ/mol K	Energía de Gibbs ΔG KJ
Monomero _{af}	-29246.8	-2019.600	-6662.048	-16.0794	0.0000
Trímero _{af}	-89760.2			-49.3485	
Monomero _{ec}	-35433.2	-1937.000	-6389.576	-19.4805	0.0000
Trímero _{ec}	-108236.6			-59.5068	

que nos indica que estás tienden al orden, siendo para la *A. ferrooxidans* mucho menor por lo que este proceso tiende más al orden en *A. ferrooxidans* que en *E. coli*.

Si bien un cambio de entropía negativo, es termodinámicamente poco favorable, para que la estructura cuaternaria corresponda a un estado termodinámicamente más favorable es necesario que este efecto entrópico sea compensado por una contribución negativa en el cambio de la energía interna del sistema (ver tabla 4.2) de forma que el cambio de la energía libre del sistema sea negativo, esta contribución energética debe provenir del modo de reacción de los átomos dentro de la macromolécula y consecuentemente, de la redistribución de las interacciones presentes. Para nuestro caso ΔG es cero que corresponde a un estado de equilibrio y como se puede ver en la tabla 4.2 este valor está comprobado para los dos procesos.

Para poder tener un descenso global en el proceso es necesario que el incremento de entalpía sea negativo, los resultados de la variación de la entalpía en el proceso de trimetrización, son negativos (ver tabla 4.2), siendo para *A. ferrooxidans* mucho menor, este es otro indicador que esta proteína es más estable que la de *E. coli*; la principal contribución entálpica al proceso de plegamiento la constituyen la formación de interacciones no covalentes, que estabilizan la estructura nativa y las interacciones hidrofóbicas entre las cadenas laterales apolares, estas entalpías negativas también indican que el proceso es exotérmico, es decir que la energía liberada se convierte en calor (energía del movimiento molecular).

En resumen podemos decir que ambos procesos de trimerización están favorecidos entálpicamente, y entre los dos trímeros el que se encuentra más estable es el trímero

de *Acidithiobacillus ferrooxidans*, ya que sus valores son mucho más negativos que para *Escherichia coli* tanto entropía como entalpía.

La capacidad calorífica es importante porque nos habla de un estado de la materia, es decir que debiera haber un C_v constante para cada estado de las proteínas (estructuras terciarias y cuaternarias).



Conclusiones

1. El diseño de la proteína CusC de *Escherichia coli* se obtuvo a partir de la estructura cristalizada, que se encuentra en la página del Protein Data Bank con el código 3PIK. El diseño de la proteína CusC de *Acidithiobacillus ferrooxidans* se obtuvo usando el servidor Robetta que nos dio su estructura secundaria.
2. Las estructuras terciarias de las proteínas CusC de *Escherichia coli* y *Acidithiobacillus ferrooxidans*, se obtuvieron luego de la simulación de dinámica molecular en un ensamble canónico y haciendo uso del software GROMACS donde la estabilidad energética y estructural fue a los 14 ns y 20 ns respectivamente.
3. El modelamiento de los trímeros de ambas proteínas, se realizó aplicando el programa desarrollado en fortran.
4. El estado de relajación energética de las estructuras cuaternarias de los trímeros de *Escherichia coli* y *Acidithiobacillus ferrooxidans* fue a partir de los 15 ns y 16 ns respectivamente luego de la simulación de dinámica molecular en un ensamble canónico y haciendo uso del software GROMACS, logrando las estructuras en equilibrio en el mínimo estructural.
5. Se obtuvieron los datos de energía de trimerización; a partir de los cuales se obtuvo la energía interna (ΔU), entropía (ΔS), entalpía (ΔH). Posteriormente, para los dos procesos de trimerización se obtuvo una ΔG igual a 0, los parámetros termodinámicos antes mencionados indican que los procesos tienden al orden; siendo el trímero de *Acidithiobacillus ferrooxidans* el que presenta una mayor estabilidad termodinámica y estructural.

Sugerencias

De lo estudiado en la presente tesis, surgen algunas sugerencias para posteriores estudios.

- Es de suma importancia en estudios bioinformáticos, trabajar con estructuras optimizadas. Esto evita que se extienda el tiempo de cálculo y memoria de disco.
- Es importante elegir las secuencias y estructuras de las proteínas de una base de datos confiable. De la misma manera se debe tener en cuenta las referencias bibliográficas de las secuencias y estructuras.
- Sería bueno hacer un estudio del proceso de trimerización a diferentes temperaturas, de esta manera lograr ubicar la temperatura que sea óptima para las proteínas y también determinar hasta qué punto la proteína es termoresistente.
- Hacer una dinámica molecular para observar mejor el comportamiento del cobre.

Bibliografía

- [1] Codelco, Biolixiviación, 2010. <http://www.codelcoeduca.cl>.
- [2] Moussa, N. CEPAL 1999, 1–69.
- [3] Biswas, A.; Davenport, W. Extractive Metallurgy of Copper, 4th ed.; Pergamon: USA, 2002.
- [4] BinaryBag, Cobre Bactericida: Plataforma de Conocimiento de la Propiedad del Cobre, 2009.
- [5] Arlandis, J. Boletín Informativo de la Asociación de Ingenieros Industriales de Madrid 2011, 13–15.
- [6] Inc., A. C. Fundamentos del Cobre, 2010. www.aqmcopper.com.
- [7] Jerez, C. A. Encyclopedia of Microbiology 2009, 1, 407–420.
- [8] Barrett, J.; Hughes, M.; Karavaiko, G.; Spencer, P. Metal extraction by bacteria oxidation of minerals.; Ellis Horwood: New York, 1993.
- [9] Das, A.; Monak, J.; Natarajan, K. Minerals Engineering 1997, 10, 743–749.
- [10] Jerez, C. Hidrometallurgy 2008, 94, 162–169.
- [11] Paulsen, I.; Park, J.; Choi, P.; Saier, M. H. FEMS Microbiology Letters 1997, 156, 1–8.
- [12] Navarro, C.; Orellana, L.; Mauriaca, C.; Jerez, C. APPLIED AND ENVIRONMENTAL MICROBIOLOGY 2009, 75, 6102–6109.

- [13] Arrascue, M. L. Biolixiviación a Nivel Industrial, X Encuentro Científico Internacional ECI 2003, 2002.
- [14] Donatti, E. Biominería: una tecnología alternativa, FUNDAMIN, 2010.
- [15] Acevedo, F. EJB Electronic Journal of Biotechnology 2000, 3, 1–11.
- [16] Valdés, J.; Pedroso, I.; Quatrini, R.; Dodson, R.; Tettelin, H.; Blake, R.; Eisen, J.; Holmes, D. BMC Genomics 2008, 9, year.
- [17] de España, I. T. G. Minería Química; ITGE: España, 1992.
- [18] Dresher, P. W. Producing Copper Nature's Way: Bioleaching, 2004.
- [19] Jonglertjunya, W. Ph.D. thesis, Escuela de Ingeniería: La Universidad de Birmingham, 2003.
- [20] Navarro, G.; Osses, M. Diagnóstico integral proceso lixiviación bacteriana en pilas y dump, Biotecnologías antofagasta s.a. technical report, 2010.
- [21] Lagos, C.; Gúzman, X. Biolixiviación: Desarrollo actual y sus expectativas, Comisión chilena de cobre technical report, 2009.
- [22] Rodríguez, Y.; Blázquez, M.; Ballester, A.; González, F.; Muñoz, J. Revista de Metalurgia 2001, 37, 616–627.
- [23] Orell, A.; Navarro, C.; Arancibia, R.; Mobarec, J.; Jerez, C. Biotechnology Advances 2010, 28, 839–848.
- [24] Hold, C.; Andrews, B.; J.A.Asenjo, Biotechnology and Bioengineering 2009, 102, 1448–1459.
- [25] Lecointre, G.; Rachdi, L.; Darlu, P.; Denamur, E. Molecular Biology and Evolution 1998, 15, 1685–1695.
- [26] Sánchez, J.; Serrano, S.; Marfil, R.; Jodral, M. Patógenos emergentes en la línea de sacrificio de porcino: Fundamentos de seguridad alimentaria; España, 2009.

- [27] Escherichia coli K-12 Derivatives Final Risk Assessment, U.S. Environmental Protection Agency, 1997. <http://epa.gov>.
- [28] Kuhnert, P.; Nicolet, J.; Frey, J. Applied and Environmental Microbiology 1995, 61, 4135–4139.
- [29] Vargas, E.; Álvarez, A.; Cervantes, C. Revista Latinoamericana de Microbiología 1998, 40, 53–71.
- [30] Hunter, C.Ñ.; Daldal, F.; Thurnauer, M. C.; Beatty, J. T. Advances in Photosynthesis and Respiration: The Purple Phototrophic Bacteria; Springer, 2009; Vol. 28.
- [31] Ma, Z.; Jacobsen, F.; Giedroc, D. Chemical Reviews 2009, 109, 4644–4681.
- [32] Peña, M.; Lee, J.; Thiele, D. JN the journal of nutrition 1999, 129, 1251–1260.
- [33] Outten, F. W.; Huffman, D. L.; Hale, J. A.; O'Halloran, T. V. The Journal of Biological Chemistry 2001, 276, 30670–30677.
- [34] Rensing, C.; Grass, G. FEMS Microbiology Reviews 2003, 27, 197–213.
- [35] Cervantes, C.; Espino-Saldaña, A.; Acevedo-Aguilar, F.; León-Rodríguez, I.; Rivera-Cano, M.; Avila-Rodríguez, M.; Wróbel-Kaczmarczyk, K.; Wróbel-Zasada, K.; Gutiérrez-Corona, J.; Rodríguez-Zavala, J.; Moreno-Sánchez, R. Revista Latinoamericana de Microbiología 2006, 48, 203–210.
- [36] Luo, Y.; Liu, Y.; Zhang, C.; Luo, H.; Guan, H.; Liao, H.; Qiu, G.; Liu, X. Current Microbiology 2008, 57, 274–280.
- [37] Pontel, L.; Audero, M. P.; Espariz, M.; Checa, S.; Soncini, F. Molecular Microbiology 2007, 66, 814–825.
- [38] Marrero, K.; Fando, R. Revista CENIC Ciencias Biológicas 2009, 40, 186–198.
- [39] Bagai, I.; Rensing, C.; Blackburn,Ñ.; McEvoy, M. Biochemistry 2008, 4, 11408–11414.

- [40] Kulathila, R.; Kulathila, R.; Indic, M.; Van den Berg, B. PLOS ONE 2011, 6, e15610–.
- [41] Franke, S.; Grass, G.; Rensing, C.; Nies, D. Journal of Bacteriology 2003, 185, 3804–3812.
- [42] Boal, A.; Rosenzweig, A. Chemical Reviews 2009, 109, 4760–4779.
- [43] Koronakis, V.; Sharff, A.; Koronakis, E.; Luisi, B.; Hughes, C. Nature 2000, 405, 914–919.
- [44] Mogensen, J.; Otzen, D. Molecular Microbiology 2005, 57, 326–346.
- [45] Storz, G.; Hengge-Aronis, R. Bacterial stress responses; American Society for Microbiology: USA, 2000.
- [46] Martínez, J. Ciencia y Trabajo 2006, 8, 159–163.
- [47] Coltell, O. M.Sc. thesis, Universitat Jaume I: Castellón, España, 2004.
- [48] Rappaport, D. The art of molecules dynamics simulation, 2nd ed.; University of Cambridge Press: United Kingdom, 2004.
- [49] Nicolás, M.; Marín, E.; Castro, F.; Miranda, R. Algunos Aspectos Básicos de la Química Computacional, Primera ed.; Universidad Nacional Autónoma de México: Cuautitlán Izcalli, Estado de México, 2006.
- [50] Leach, A. R. Molecular Modelling, 2nd ed.; Pearson Education., 2001.
- [51] Levine, I. Química Cuántica, Quinta ed.; Pearson Educación.: Madrid, 2001.
- [52] Aguirre, A. M.Sc. thesis, Universidad de Granada: España, 2009.
- [53] Cuevas, G.; Cortés, F. Introducción a la Química Computacional, Primera ed.; Fondo de Cultura Económica: México DF, 2003.
- [54] Lii, J.-H.; Allinger, N J. Am. Chem. Soc. 1989, 111, 8576–8582.

- [55] Bort, J. A.; Beltrán, J. Química Teórica y Computacional; Universitat Jaume I, 2000.
- [56] Handbook of Computational Chemistry; Leszczynski, J., Ed., Tercera ed.; Springer, 2012.
- [57] Hinchliffe, A. Molecular Modelling for Beginners.; Wiley Publishing, Inc., 2008.
- [58] Holtje, H.-D.; Folkers, G. Molecular Modelling: Basic Principles and Applications; VCH: New York, USA, 1997; Vol. 5.
- [59] Cerius, Forcefield-Based Simulations, 1998. <http://www.chem.cmu.edu>.
- [60] Lévy, J.; Flórez, R.; Fernández, J. Las redes neuronales artificiales: Fundamentos teóricos y aplicaciones prácticas., 1st ed.; Netbiblo, España, 2008.
- [61] Dodier, R. Manual de Maxima Ver. 5.26; 2011.
- [62] Sánchez, D. M.Sc. thesis, Universidad de Cádiz, 2000.
- [63] Dodson, G.; Lane, D.; Verma, C. Molecular simulations of protein dynamics: new windows on mechanisms in biology., University of York, York, uk technical report, 2008.
- [64] Haile, J. Molecula Dynamics Simulation: Elementary Methods; John Wiley & Sons, Inc.: Clemson, South Carolina, 1992.
- [65] Namba, A.; da Silva, V.; da Silva, C. Eclética Química 2008, 33, 13–24.
- [66] Taidi, M. Ph.D. thesis, Universidade Estadual de Campinas: Campinas, São Paulo, 2005.
- [67] Huang, K. Statistical Mechanics, 2nd ed.; John Wiley & Sons, Inc.: Canada, 1987.
- [68] Gupta, M. Statistical Thermodynamics, 2nd ed.; K.K Gupta: New Delhi, 2003.

- [69] Moyano, G. Cálculos Básicos de Termodinámica Estadística, Universidad de Antioquia, 2011.
- [70] Frenkel, D.; Smit, B. Understanding Molecular Simulations; Academic Press, 2002.
- [71] Atkins, P.; de Paula, J. Atkin's Physical Chemistry; Oxford, 2002.
- [72] Arfken, G.; Weber, H. Methods for Physicists, Sixth ed.; Elsevier Inc.: USA, 2005.
- [73] Robetta Beta Full-chain Protein Structure Prediction Server, 2004.
- [74] Cálculos Ab initio. <http://www.luventicus.org>.
- [75] Berg, J. M.; Tymoczko, J. L.; Stryer, L.; Macarulla, D. J. Bioquímica, Sexta ed.; Reverté, España, 2008.
- [76] da Silva, V.; Silva, C. Revista Electrónica de Farmacia 2007, 4, 15–26.
- [77] Modern Methods for Theoretical Physical Chemistry of Biopolymers; Eugeni B., James P., L.; Shigenori., Eds., 1st ed.; ELSEVIER B. V.: The Netherlands, 2006.
- [78] Voet, D.; Voet, J. Bioquímica, Tercera ed.; Editorial Médica Panamericana: Buenos Aires, 2006.
- [79] Apol E., Apostolov R., F. G. B. B.; others., GROMACS: Groningen Machine for Chemical Simulations; 4th ed., 2010.
- [80] The New Avenues in Bioinformatics; Seckbach, J.; Rubin, E., Eds.; Kluwer Academic: Netherlands, 2004.
- [81] McQuarrie, D.; Simon, J. Molecular Thermodynamics.; University Science Books, 1999.



Anexo A

Archivos utilizados



```
title           = Minimizacion
cpp             = /lib/cpp
include        = -I../top
integrator      = steep
nsteps         = 20000
nstcomm        = 1
emtol          = 41,8
emstep         = 0.01
nstcgsteep     = 1000
nstxout        = 10
nstvout        = 10
nstfout        = 0
nstlog         = 100
nstenergy      = 1
nstxtcout      = 1
xtc_precision  = 1000
xtc_grps       = Protein
energygrps     = Protein
nstlist        = 5
ns_type        = grid
pbc            = xyz
rlist          = 1.31
coulombtype    = PME-Switch
rcoulomb_switch = 0.1
rcoulomb       = 0.9
epsilon_r      = 78.0
epsilon_rf     = 1
vdwtype        = Shift
rvdw_switch    = 0.1
rvdw           = 1.0
DispCorr       = Ener
constraints     = none
optimize_fft   = yes
fourierspacing = 0.12
```

Figura A.1: Archivo minima.mpd con integrador steepest descent

```

title           = Minimizacion
cpp             = /lib/cpp
include        = -I./top
define         =
integrator      = l-bfgs
nsteps         = 100000
nstcomm        = 1
emtol          = 0.1
emstep         = 0.01
nstcgsteep     = 1000
nstxout        = 10
nstvout        = 10
nstfout        = 0
nstlog         = 100
nstenergy      = 1
nstxtcout      = 1
xtc_precision  = 1000
xtc_grps       =
energygrps     =
nstlist        = 5
ns_type        = grid
pbc            = xyz
optimize_fft   = yes
fourierspacing = 0.3
DispCorr       = Ener
constraints    = none
implicit_solvent = GBSA
coulombtype    = Cut-off
rlist          = 1.0
rcoulomb       = 1.0
rgbradii       = 1.0
vdwtype        = Cut-off
rvdw           = 1.0
gb_algorithm    = HCT
nstgbradii     = 1
gb_epsilon_solvent = 78
gb_dielectric_offset = 0.09
sa_surface_tension = 2.05016
    
```



Figura A.2: Archivo minima.mpd con integrador l-bfgs

Tabla que presenta las Relaciones Termodinámicas de Maxwell.⁸¹

Tabla A.1: Relaciones Termodinámicas

U	$\Delta E = T\Delta S - P\Delta V$	$\frac{\delta E}{\delta S}_V = T$ $\frac{\delta E}{\delta V}_S = -P$
$H = E + PV$	$\Delta H = T\Delta S + V\Delta P$	$\frac{\delta H}{\delta S}_P = T$ $\frac{\delta H}{\delta P}_S = V$
$A = U - TS$	$\Delta A = \Delta U - T\Delta S$	$\frac{\delta A}{\delta T}_P = -S$ $\frac{\delta G}{\delta P}_T = V$



```

title                = Dinamica Molecular NVT
cpp                  = /lib/cpp
include              = -I./top
integrator           = md
dt                   = 0.001
nsteps               = 20000000
nstxout              = 400
nstvout              = 400
nstlog               = 400
nstenergy            = 40
nstxtcout            = 400
xtc_grps             =
energygrps           =
nstlist              = 10
ns_type              = gnd
fourierspacing       = 0.30
pme_order            = 4
ewald_rtol           = 1e-5
pbc                  = xyz
tcoupl               = nose-hoover
tc-grps              = protein
tau_t                = 0.5
ref_t                = 303.15
gen_vel              = yes
gen_temp              = 303.15
gen_seed              = 173
constraints           = none
implicit_solvent     = GBSA
coulombtype          = Cut-off
rlist                = 1.0
rcoulomb              = 1.0
rgbradii              = 1.0
vdwtype              = Cut-off
rvdw                  = 1.0
gb_algorithm          = HCT
nstgbradii           = 1
gb_epsilon_solvent   = 78
gb_dielectric_offset = 0.09
sa_surface_tension    = 2.05016
    
```



Figura A.3: Archivo nvt.mpd con integrador MD

Anexo B

Ecuaciones utilizadas

1. RMSD (Root Mean Square Deviation): El valor de RMSD corresponde a la distancia euclidiana entre dos vectores, normalizada por el número de dimensiones de estos. Una estructura de proteína puede considerarse un vector de $3N$ dimensiones donde N corresponde al número de átomos de la estructura, de este modo, para dos conformaciones a y b de una proteína, su RMSD está dado por:

$$\begin{aligned} RMSD(a, b) &= \sqrt{\frac{1}{3N} \sum_{i=1}^n (b_{ix} - a_{ix})^2 + (b_{iy} - a_{iy})^2 + (b_{iz} - a_{iz})^2} \\ &= \sqrt{\frac{1}{3N}} \cdot D(a, b) \end{aligned} \quad (B.1)$$

2. RMSF (Root Mean Square Fluctuation): RMSF es una medida de flexibilidad local de una estructura. A partir de una simulación de dinámica molecular, el RMSF del átomo i da cuenta del grado de fluctuación de las coordenadas de dicho átomo respecto a su promedio temporal:

$$RMSF(i) = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - [x_i])^2 + (y_i - [y_i])^2 + (z_i - [z_i])^2} \quad (B.2)$$

Donde los términos entre corchetes corresponden a los promedios temporales.

Anexo C

Datos de la Simulación

Proteína	Temperature	Err.Est.	RMSD	Tot-Drift
Monómero(AF)	303.151	0.00058	3.60403	0.00193334 (K)
Monómero(EC)	303.15	0.00069	-nan	0.00114795 (K)
Trímero(AF)	303.15	0.00049	2.09803	0.000206776 (K)
Trímero(EC)	303.15	0.00036	2.17011	0.00172398 (K)

Figura C.1: Resultados de la temperatura promedio a las que se trabajó las cuatro proteínas.

Proteína	Pressure	Err.Est.	RMSD	Tot-Drift
Monómero(AF)	0.00320114	0.0021	6.66389	-0.000330311 (bar)
Monómero(EC)	-0.000728929	0.0018	-nan	-0.000963274 (bar)
Trímero(AF)	0.00804209	0.002	9.25135	-0.00387893 (bar)
Trímero(EC)	-0.0141788	0.005	11.9883	0.0284106 (bar)

Figura C.2: Resultados de la presión promedio a las que se trabajó las cuatro proteínas.

Anexo D

Programa para construir el trímtero



```

Program trimerio
-----
parameter (maxv=50000)
parameter (pi=3.14159265358979323846)
-----
character nombre*80, linea*80, respuesta*1
character res(maxv)*8, tip(maxv)*7
-----
integer num, ato(maxv), grp(maxv), amn, am1, nul
-----
real cox(maxv),coy(maxv),coz(maxv),lax,lay,laz
real clx(maxv),cly(maxv),clz(maxv),c2x(maxv),c2y(maxv),c2z(maxv)
-----
999 continue
write(*,*)'-----'
write(*,*)'Este programa tomara una estructura y producira su'
write(*,*)'trimerio tomando como referencia coordenadas xyz'
write(*,*)'La extension que tiene tu archivo debe ser gro'
write(*,*)'-----'
-----
write(*,*)'Los datos se encuentran en: '
read(*,*) nombre
open(60,FILE=nombre,IOSTAT=ierror,STATUS="OLD",ACTION="READ")
if(ierror.ne.0) then
write(*,*)'Error en el proceso de lectura del archivo'
stop
end if
read(60,FMT="(A)") linea
read(60,100) num
-----
do i=1,num
res(i)=""
tip(i)=""
end do
do i=1,num
ato(i)=0
grp(i)=0
end do
do i=1,num
cox(i)=0.0
coy(i)=0.0
coz(i)=0.0
clx(i)=0.0
cly(i)=0.0
clz(i)=0.0
c2x(i)=0.0
c2y(i)=0.0
c2z(i)=0.0
end do
-----
do i=1,num
read(60,101)grp(i),res(i),tip(i),ato(i),cox(i),coy(i),coz(i)

```

Figura D.1: Primera parte del programa

```

end do
read(60,102)lax,lay,laz
close(60)
C-----
    amn = grp(num)
    am1 = 2*amn
    nul = 2*num
C-----
C-----
C-----
    do i=1,num
    c1x(i) = cox(i)*cos(2*pi/3) + coy(i)*sin(2*pi/3)
    c1y(i) = -cox(i)*sin(2*pi/3) + coy(i)*cos(2*pi/3)
    c1z(i) = cozi(i)
    c2x(i) = cox(i)*cos(4*pi/3) + coy(i)*sin(4*pi/3)
    c2y(i) = -cox(i)*sin(4*pi/3) + coy(i)*cos(4*pi/3)
    c2z(i) = cozi(i)
    end do
C-----
C-----
C-----
C-----
    write(*,*)'Donde quieres que guarde tus resultados'
    read(*,*)nombre
    open(61,FILE=nombre,IOSTAT=ierror,STATUS="REPLACE",ACTION="WRITE")
    write(61,*)'Produciendo un trimero con nuestro programa'
    write(61,100)3*num
    do i=1,num
    write(61,101)grp(i),res(i),tip(i),ato(i),cox(i),coy(i),cozi(i)
    end do
    do i=1,num
    write(61,101)grp(i)+amn,res(i),tip(i),ato(i)+num,c1x(i),c1y(i),c1z(i)
    end do
    do i=1,num
    write(61,101)grp(i)+am1,res(i),tip(i),ato(i)+nul,c2x(i),c2y(i),c2z(i)
    end do
    write(61,102)3*lax,3*lay,laz
    close(61)
C-----
100  format(i10)
101  format(i5,a3,a7,i5,f8.3,f8.3,f8.3)
102  format(f10.5,f10.5,f10.5)
C-----
    write(*,*)'Desea continuar (s / n):'
    read(*,*) respuesta
    if(respuesta .eq. 's') goto 999
C-----
end

```

Figura D.2: Segunda parte del programa

Anexo E

Conceptos adicionales

- Formato FASTA: es un formato para representar las secuencias de nucleótidos o péptidos, en el cual los pares de base o aminoácidos son representados usando códigos de una sola letra. La secuencia en formato FASTA comienza con una línea de descripción, seguida por las líneas de las secuencias de datos. La línea de descripción se distingue de la secuencia por el siguiente símbolo: <<, al inicio. Es recomendable que las líneas del texto tengan una longitud menor a 80 caracteres.
- BLAST: es un programa de búsqueda proteína-proteína.
- PDB-BLAST: es una interfaz que solo permite búsqueda de proteínas. ([http : //www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).
- HMMER: es un programa usado para la búsqueda de secuencias en bases de datos para homólogos de secuencias de proteínas, y también permite realizar alineamientos de proteínas. HMMER es más preciso que BLAST y más capaz de detectar homólogos remotos. (<http://hmmer.janelia.org/>).
- PSI-BLAST: es una herramienta usada para buscar posibles homólogos en organismos muy lejanos entre ellos, filogenéticamente hablando. ([http: //www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).

- PSI-BLAST-MSA: es un método de elección para detectar regiones conservadas en el ADN y proteínas, usualmente asociadas con; estructura como plegamiento y regiones de interacción, reactividad química como sitios catalíticos.

