

UNIVERSIDAD CATÓLICA SANTA MARÍA

FACULTAD DE CIENCIAS FARMACEÚTICAS, BIOQUÍMICAS Y
BIOTECNOLÓGICAS

PROGRAMA PROFESIONAL DE INGENIERÍA BIOTECNOLÓGICA



“Identificación del tipo de interacciones involucradas en la
región poliglutámica de la Ataxina-1 en presencia y ausencia de
histidina mediante herramientas bioinformáticas”

Tesis presentada por el Bachiller:

GIULIANA ELIZABETH ROMERO MARISCAL

Para optar el Título Profesional de

INGENIERO BIOTECNÓLOGO

Asesor: Jaime Barreda del Carpio

AREQUIPA – PERÚ

2012

"Todo lo que una persona puede imaginar, otros pueden hacerlo realidad."

Julio Verne

Agradecimientos:

A mis padres por su apoyo incondicional, gracias a ellos soy la persona que soy hoy en día. Mamá y Papá muchas gracias.

A mis amigos por hacer mi vida tan maravillosa, su compañía en todo momento me ha impulsado hasta aquí.

Al Dr. Badhin Gómez por toda su paciencia, pero sobre todo por sus sabios consejos.

A todos mis compañeros del laboratorio de Bioinformática, los únicos que entienden mi esfuerzo.

Por último a la vida que me ha hecho conocer tantas personas maravillosas y vivir tantos momentos preciosos.

Índice

Índice.....	i
Índice de Figuras.....	iv
Glosario.....	vi
Resumen.....	vii
Abstract.....	viii
Capítulo 1.....	1
Introducción	1
Objetivos	3
1.2.1. Objetivo General.....	3
1.2.2. Objetivos Específicos.....	3
Hipótesis	4
Capítulo 2.....	5
Marco Teórico	5
2.1 Ataxia Espinocerebelosa Tipo 1 (SCA1).....	5
2.2 Ataxina - 1.....	8
2.3 Poliglutamínica.....	10
2.4 Bioinformática.....	11
2.4.1 Dinámica Molecular.....	13
2.4.2 Elementos de la Dinámica Molecular.....	17

2.4.3 Termodinámica Estadística	26
2.4.3.1 Ensamble canónico	30
2.4.4 Potencial Electrostático Molecular.....	33
2.4.4.1 Ecuación de Poisson-Boltzmann	35
Capítulo 3.....	38
Metodología y Detalles Computacionales	38
3.1. Equipos y Software	38
3.2. Elección de las estructuras usadas en las simulaciones.....	42
3.3. Construcción de la Ataxina-1 mutada	44
3.4. Optimización de la geometría.....	44
3.5. Dinámica Molecular de estructuras terciarias	47
3.6. Análisis del Potencial electrostático a través de APBs	49
Capítulo 4.....	51
Resultados y Discusión	51
4.1. Obtención de Estructuras.....	51
4.2. Estabilidad de estructuras terciarias	52
4.2.1. Análisis de la estabilidad de la proteína Ataxina-1	52
4.2.2. Análisis de la estabilidad de la proteína Ataxina-1 mutada	59
4.3. Potencial electrostático.....	66
Conclusiones	77
Sugerencias	79
Referencias	80
Apéndices	89
A. Archivos utilizados	90
B. Ecuaciones Utilizadas	92

C. Secuencias FASTA	93
D. Comandos utilizados en GROMACS	94
E. Diagrama de trabajo	95



Índice de Figuras

2.1. Elementos estructurales de la proteína Ataxina 1	8
2.2. Ataxina-1 Normal (a) y Ataxina-1 Mutada.....	9
2.3. Superficie de energía potencial	19
2.4. Superficie de energía potencial (Puntos Críticos).....	20
2.5. El mínimo global.....	21
2.6. Búsqueda del mínimo local	22
2.7. Método de Steepest Descent	23
2.8. Búsqueda de mínimo local.....	24
2.9. Búsqueda lineal en una dimensión	25
2.10. Representación del Ensamble Canónico	31
2.11. Esquema de la malla utilizada en el programa AMBER para resolver la ecuación de Poisson- Boltzmann	36
3.1. Modelado de Ataxina-1. Realizado con el programa UGENE	43
3.2. Secuencia FASTA de Ataxina-1 mutada	44
3.2. Estructuras tridimensionales (proteína normal y mutada)	46
3.4. Estructuras tridimensionales dinamisadas (proteína normal y mutada)	48
4.1. Estructuras tridimensionales minimizada (proteína normal y mutada).....	51
4.2. Estructuras tridimensionales plegadas (proteína normal y mutada)	52
4.3. Estabilidad Energética de la Ataxina-1	53
4.4. Puentes de Hidrógeno de la Ataxina-1 (Xmgrace)	54
4.5. Puentes de Hidrógeno de la Ataxina-1(Chimera)	55
4.6. RMSD de la Ataxina-1	56
4.7. RMSF de la Ataxina-1	57
4.8. Gráfica de Ramachandran Ataxina-1	58
4.9. Estabilidad Energética de la Ataxina-1 mutada	60
4.10. Puentes de Hidrógeno de la Ataxina-1 mutada (Xmgrace).....	61
4.11. Puentes de Hidrógeno de la Ataxina-1 mutada (Chimera).....	62
4.12. RMSD de la Ataxina-1 mutada	63
4.13. RMSF de la Ataxina-1 mutada.....	63

4.14. Gráfica de Ramachandran Ataxina-1 mutada	65
4.15. Vista isométrica de Ataxina-1 luego de la simulación de DM.....	68
4.16. Potencial Electrostático de Ataxina-1	69
4.17. Vista isométrica de Ataxina-1 mutada luego de la simulación de DM	70
4.18. Potencial Electrostático de Ataxina-1 mutada	71
4.19. Regulación alostérica en enzimas	72
4.20. Potencial Electrostático de Ataxina-1 y su Cadena Poliglutaminica	73
4.21. Potencial Electrostático de Ataxina-1 mutada y su Cadena Poliglutamínica	75



Glosario

1. SCA1. Ataxia Espinocerebelosa tipo 1
2. ATXN-1. Ataxina 1
3. HD. Huntington
4. SBMA. Atrofia muscular espinobulbar o enfermedad de Kennedy
5. DRPLA. Aatrofia dentorrubropalidoluisiana
6. NCBI. National Center for Biotechnology Information.
7. PDB. Protein Data Bank.
8. PB. Poisson-Boltzmann
9. SEP Superficie de energía potencial
10. MM. Mecánica Molecular.
11. DM. Dinámica Molecular.
12. PE Potencial Electrostático
13. RMSD. Root Mean Square Deviation.
14. RMSF. Root Mean Square Fluctuation.
15. GAPDH. Gliceraldehído 3-fosfato deshidrogenasa
16. LANP. Proteína acidica nuclear rica en leucinas

Resumen

En el presente trabajo de investigación se buscó ampliar los conocimientos que se posee sobre la enfermedad Ataxia Espinocerebelosa Tipo1 (SCA1) para lo cual se realizó el modelamiento molecular de dos proteínas, la Ataxina-1 normal y la Ataxina-1 mutada, proteína causante de la enfermedad, esto se debe a la expansión de glutaminas y a la ausencia de histidina en la cadena poliglutamínica.

Una vez realizado el modelamiento se procedió a realizar simulaciones de Dinámica Molecular para ambas proteínas, luego se determino las energías totales de las mismas para poder determinar su estabilidad, después de la Dinámica Molecular se observaron cambios en las estructuras y variaciones de estabilidad, las cuales fueron corroboradas por en análisis de Ramachandran, posteriormente se obtuvieron los potenciales electrostáticos de ambas proteínas.

El potencial electrostático determinado junto con la información bibliográfica permitió descubrir las interacciones que desarrolla la Proteína Ataxina-1, con la proteína Gliceraldehído 3-fosfato deshidrogenasa (GAPDH) y la proteína acidia nuclear rica en leucinas (LANP), siendo la interacción de estas 3 proteínas la que desencadena la enfermedad.

Abstract

In the present investigation sought to expand knowledge about the disease has Spinocerebellar Ataxia Type 1 (SCA1) for which molecular modeling was performed for two proteins, ataxin-1 normal and mutant ataxin-1, causing protein the disease, this expansion is due to the absence of glutamine and histidine polyglutamine chain.

Once the modeling proceeded to Molecular Dynamics simulations for both proteins, is then determined the total energies thereof in order to determine their stability after Molecular Dynamics changes were observed variations in the structure and stability, which in corroborated by Ramachandran analysis subsequently electrostatic potentials were obtained from both proteins.

The electrostatic potential loads allowed determining residues that affect the formation and properties of the links which let us to discover that develops interactions ataxin-1, the protein glyceraldehyde 3-phosphate dehydrogenase (GAPDH) protein and acidia leucine-rich nuclear (LANP), with the interaction of these 3 proteins that triggers the disease.

Capítulo 1

Introducción

En los últimos años, se ha observado que la incidencia de enfermedades degenerativas va en aumento, siendo las neurodegenerativas como el Alzheimer, Huntington y el Parkinson las más comunes (1). La principal característica de estas enfermedades es el rápido deterioro en la calidad de vida de los pacientes.

Recientes estudios han permitido determinar que una gran cantidad de estos trastornos se deben a mutaciones genéticas, siendo la repetición del codón CAG (glutamina) una de las mutaciones más frecuentes (2), estos trastornos, que presentan expansión de glutaminas son conocidos como enfermedades poliglutaminicas (3).

Existen una gran variedad de enfermedades poliglutaminicas, siendo las ataxias hereditarias las de mayor incidencia; estas se caracterizan por presentar manifestaciones clínicas relacionadas a la disfunción del cerebelo (ataxia cerebelosa). Las ataxias hereditarias se han venido estudiando durante las últimas décadas por medio de técnicas de biología molecular y genéticas, lo que ha permitido dar a conocer un mayor número de genes implicados con el desarrollo de estas enfermedades y poder así clasificarlas como enfermedades autosómicas recesivas, autosómica dominante o ligada al cromosoma X (4).

Entre las ataxias autosómicas dominantes, se encuentran las ataxias espinocerebelosas (SCA); las cuales se subdividen según los genes afectados. Se han identificado al menos 21 de los distintos subtipos genéticos de SCAs, siendo una de estas la Ataxia Espinocerebelosa Tipo 1 (SCA1) (5).

La Ataxia Espinocerebelosa Tipo 1, es desarrollada debido a la mutación del gen ATXN-1, localizado en el locus 6p23, este gen es el encargado de codificar la proteína

Ataxina-1, la función de esta proteína es aún desconocida, aunque se cree que podría estar involucrada en la síntesis de proteínas. El gen ATXN-1 mutado presenta una expansión del codón CAG (glutamina) en un rango de 41-81 repeticiones en cambio los alelos normales contienen 6-39 repeticiones de CAG, interrumpidos con secuencias CAT (histidina) (6). El incremento de glutaminas en el gen ATXN-1 genera la producción de una versión anormalmente larga de la proteína Ataxina-1, que se pliega inadecuadamente en su estructura terciaria. Esta proteína anormal se aglomera con otras proteínas formando agregados dentro de los núcleos celulares. Estos agregados causan un mal funcionamiento de la proteína Ataxina-1, lo cual conlleva a la muerte celular (7) y al desarrollo de la enfermedad.

Es por ello que es necesario comprender las interacciones energéticas que genera la presencia y ausencia de la histidina en la región poliglutamínica así como el efecto que produce la expansión de glutaminas en la proteína Ataxina-1 normal y mutada, para así poder dilucidar la función que cumplen estas alteraciones en el plegamiento de la proteína y sus consecuentes interacciones energéticas.

Objetivos

1.1. Objetivo General

Identificar las interacciones energéticas en la región poliglutamínica de la proteína Ataxina-1 normal y mutada en presencia y ausencia del aminoácido histidina, respectivamente.

1.2. Objetivos Específicos

1. Obtener las secuencias de la proteína Ataxina-1 normal en formato .pdb y modelar la estructura de la proteína.
2. Mutar la Ataxina-1 normal, eliminando la histidina y agregando glutaminas en la cadena poliglutamínica. Luego obtener la secuencia en formato .pdb y modelar su estructura
3. Minimizar las estructuras de las proteínas, para obtener su optimización geométrica.
4. Realizar la Dinámica Molecular de nuestras estructuras, utilizando el método de solvente implícito y determinar la estabilidad de las mismas mediante los análisis adecuados.
5. Obtener el potencial electrostático de la proteína Ataxina-1 normal y mutada.
6. Se evaluó los potenciales electrostáticos y la información bibliográfica para determinar las interacciones que desarrolla con la proteína Gliceraldehído 3-fosfato deshidrogenasa (GAPDH) y la proteína acidia nuclear rica en leucinas (LANP).

Hipótesis

Dado que la aplicación de la Bioinformática nos brinda datos de la estabilidad energética, es probable que nos permita determinar mediante el Potencial electrostático la importancia estructural y energética de la histidina presente en la proteína Ataxina-1 normal y la ausencia de la misma en la proteína Ataxina-1 mutada.



Capítulo 2

Marco Teórico

En esta investigación se utilizan métodos de mecánica cuántica para poder realizar un análisis a nivel molecular de la proteína Ataxina-1, es por ello que considero necesario hacer una breve descripción tanto de los métodos aplicados como de la proteína utilizada en nuestra investigación.

Por lo tanto en el presente capítulo, daremos un breve vistazo a la enfermedad neurodegenerativa Ataxia Espinocerebelosa Tipo 1 (SCA1), así como a la proteína causante de la enfermedad, la Ataxina-1, la cual posee dos formas, la normal y la mutada, diferenciadas por el tamaño de la cadena poliglutámica y por la presencia y ausencia de histidina.

También daremos una introducción a los diferentes conceptos teóricos a ser utilizados para la sistematización de la información extraída desde cálculos mecánico cuánticos.

2.1. Ataxia Espinocerebelosa Tipo 1 (SCA 1)

La palabra ataxia, deriva de la palabra griega *taxis* que significa orden, se suele utilizar para referirse a un grupo de enfermedades que cursan con incoordinación motora como manifestación fundamental. (34)

Existen diversos tipos de ataxias, las cuales son clasificadas de acuerdo con la localización cromosómica y el patrón genético: la ataxia autosómica dominante, en la cual la persona afectada hereda un gen normal de un progenitor y un gen defectuoso del otro; y la ataxia autosómica recesiva, en la cual ambos progenitores le pasan una copia del gen defectuoso. Entre las ataxias hereditarias más comunes están la ataxia de Friedreich y la enfermedad de Machado-Joseph. También pueden ocurrir ataxias esporádicas en familias sin historia previa (34).

La ataxia también puede ser adquirida. Las afecciones que pueden causar la ataxia adquirida incluyen los accidentes cerebrovasculares, la esclerosis múltiple, los tumores, el alcoholismo, la neuropatía periférica, los trastornos metabólicos y las deficiencias vitamínicas (34).

La ataxia Espinocerebelosa tipo 1 o SCA1 (siglas en inglés), corresponde al grupo de ataxias autosómicas dominantes; la enfermedad aparece normalmente entre los cuarenta a los setenta y cinco años, pero se han registrado edades de inicio de la enfermedad desde los 18 meses a los 30 años. Afectando por igual a hombres y mujeres (7).

El primer síntoma en aparecer es generalmente la ataxia (carencia de la coordinación de movimientos musculares), que se manifiesta como dificultad, inestabilidad y falta de coordinación al caminar. La falta de coordinación, en un principio, solo afecta a la marcha, pero más tarde puede afectar también a los brazos, manos y tronco (7).

Pueden presentar otros síntomas tales como: deformidades en los pies, escoliosis (encurvamiento de la columna vertebral hacia un lado), debilidad y atrofia (disminución de volumen y peso de un órgano) de los músculos, pérdida de reflejos en rodillas, muñecas y tobillos, pérdida de sensibilidad en las extremidades, nistagmus (movimientos rítmicos, rápidos e involuntarios en los ojos) y disartria (dificultad y lentitud en el habla) (7).

La ataxia, a menudo, va acompañada de afecciones cardíacas: cardiomiopatía hipertrófica, miocarditis (inflamación del músculo cardíaco), taquicardia (latido anormalmente rápido del corazón) pudiendo llegar al fallo cardíaco. Asimismo, desarrollan con frecuencia diabetes y en algunas ocasiones hipotiroidismo (7).

Los síntomas neurológicos en SCA1 se producen por una pérdida progresiva de neuronas en el cerebelo, como las células de *Purkinje*, también degeneran neuronas en el tronco cerebral y se observa una desmielinización frecuente en las columnas posteriores y los tractos espinocerebelosos. Todos estos problemas reducen tanto la calidad como la expectativa de vida de los pacientes (8).

Estudios recientes, muestran que los procesos de neurotransmisión sináptica, regulación transcripcional, homeostasis de calcio, excitotoxicidad y señalización del glutamato, y

degradación proteica son algunos de los mecanismos moleculares y celulares responsables de la patogénesis molecular de la SCA1 (9).

La SCA1 es causada por la expansión del aminoácido glutamina en la proteína Ataxina-1 como consecuencia de la expansión del triplete CAG en la zona codificante del gen SCA1 que se encuentra localizado en el brazo corto del cromosoma 6. Como se menciono antes las células de *Purkinje* son las primeras en ser afectadas, produciéndose una pérdida progresiva éstas, manifestándose en forma de atrofia cerebelosa progresiva (9).

Los síntomas se presentan en individuos que posean entre 39 a 81 repeticiones de glutaminas en el gen ATXN-1 y el desarrollo de los mismos varía de acuerdo al incremento en la longitud de la cadena; siendo predominantes las cadenas largas en pacientes jóvenes. El análisis de este gen ha revelado la presencia de repeticiones del codón CAT (histidina) en un 98 por ciento de los genes normales, los cuales se encuentran repetidos entre 1 a 3 veces, en contraste con los genes mutados, que se caracterizan por la repetición de una serie de glutaminas (CAG) y ausencia de histidina (10). Es debido a esto que se cree que la presencia de histidina es de vital importancia en el mantenimiento de la estabilidad del gen ATXN-1 y la posterior traducción de la proteína Ataxina-1 (11).

Los mecanismos celulares y moleculares que causan neurodegeneración en SCA1 son todavía poco claros. Los estudios genéticos en moscas y ratones han demostrado que la enfermedad se produce por mecanismos tóxicos de ganancia de función debido a que los ratones que no presentan Ataxina-1 no desarrollan ataxia ni signos de neurodegeneración. Los ratones que presentan Ataxia-1 con la expansión de glutaminas han demostrado que la enfermedad no es producida por muerte celular per se, sino por disfunción neuronal y alteraciones morfológicas que ocurren en unos estadios iniciales anteriores a la aparición de ataxia y pérdida neuronal (11).

Aunque no hay todavía fármacos ni tratamientos efectivos para SCA1 u otras ataxias pero se han realizado algunas aproximaciones de intervenciones terapéuticas entre ellas la más prometedora es la que se está realizando con litio. El resto de intervenciones realizadas si bien no alivian todos los síntomas neurológicos, si algunos de ellos, como

es el caso de la Levodopa y los Agonistas Dopaminérgicos que se están empleando con éxito para tratar la distonia y otros síntomas extrapiramidales en SCA1 y otras ataxias (11).

2.2. Ataxina-1

En 1993 se logró clonar el gen de la SCA1 lo cual permitió descubrir que este gen codificaba una proteína, la cual fue designada como Ataxina-1 (12). Esta proteína tiene un tamaño de 90 KDa y se localiza en el núcleo neuronal, en el citoplasma de las células de los tejidos periféricos y en las células de Purkinje.

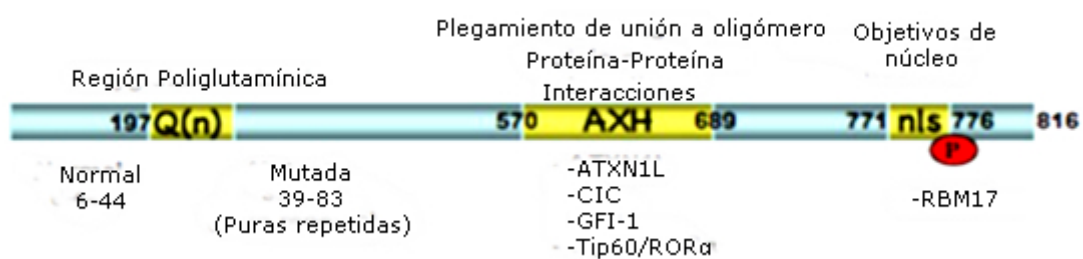


Figura 2.1: Elementos estructurales de la proteína Ataxina 1. Los elementos funcionales están indicados en amarillo. La secuencia muestra el tamaño completo de la proteína con 816 aminoácidos como corresponde a la proteína normal con 30 residuos glutámicos repetidos (197 -226 residuos). El dominio AXH de 570 – 689 residuos, forma un oligonucleótido de unión importante para la dimerización de la proteína así como para su interacción con otras. La región NLS o secuencia de localización nuclear abarca desde los aminoácidos 771 hasta 776. Para finalizar la región C-terminal del NLS es Ser 776, un sitio de fosforilación. (17).

La proteína Ataxina-1 presenta dos dominios diferenciados. En la región amino terminal se encuentra la secuencia de glutaminas mientras que en la región opuesta, carboxilo terminal, se identifica el dominio AXH que es un dominio que participa en la interacción con proteínas. Muy cerca de este dominio también se identifica un sitio de fosforilación por la proteína Kinasa B (13).

La Ataxina-1 regula programas genéticos implicados en las funciones motoras. Y a nivel molecular, juega un papel importante en la regulación de la transcripción de genes que participan en programas de control motor en el cerebelo y en otras regiones del cerebro humano. Entre éstos programas se incluyen los regulados por las rutas de Wnt y

los receptores de hormonas y señalización intracelular. Los datos científicos indican que la Ataxina-1 ejerce de co-regulador de la transcripción de genes que son diana de las hormonas retinoideas y tiroideas en el cerebro. Estas hormonas tienen un papel muy importante en los mecanismos de desarrollo y diferenciación de ciertos tipos de neuronas y partes del cerebro, incluido el cerebelo, y creemos que defectos en estos mecanismos de regulación neuronal podrían inducir disfunciones tempranas en las neuronas donde la Ataxina-1 actúa: principalmente en el cerebelo, en el tronco cerebral, en el mesencéfalo y los ganglios basales, Las cuales son regiones nerviosas imprescindibles para el control de las funciones motoras en vertebrados superiores (17).

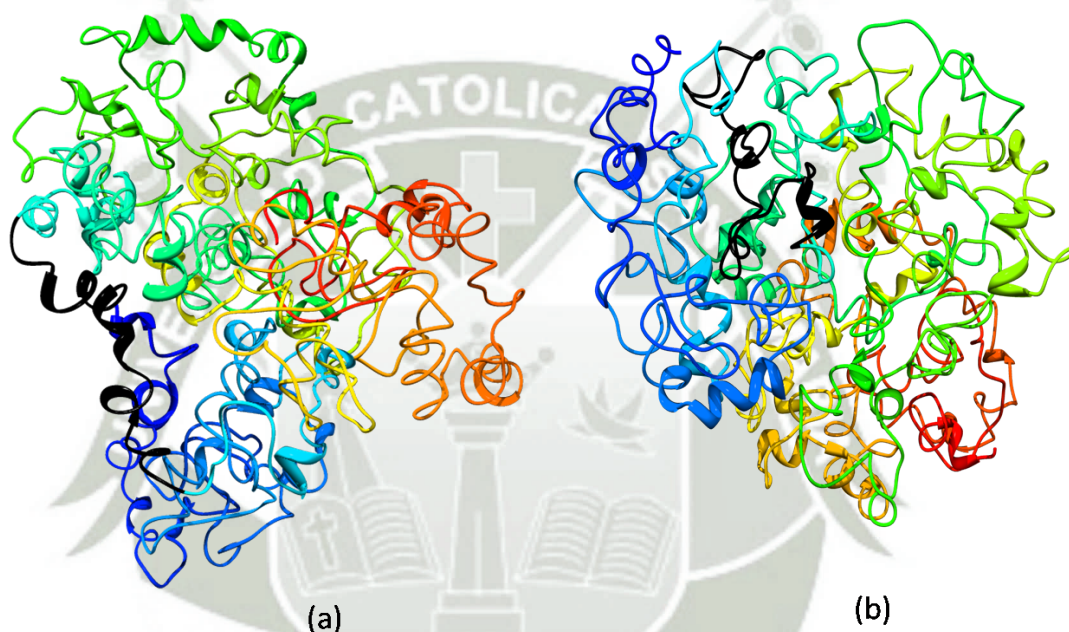


Figura 2.2: Ataxina-1 Normal (a) y Ataxina-1 Mutada, se ha señalado la cadena de poliglutaminas con color negro en ambas, se aprecia la diferencia de longitud. Imágenes obtenidas con Chimera 1.5.3.

La proteína Ataxina-1 normal contiene de 6-39 repeticiones de CAG (glutamina), siendo interrumpidos por una a tres CAT (histidina) a partir de la 21ava repetición. Mientras que la proteína mutada presentan secuencias puras de CAG con repeticiones entre 39 a 82 unidades (14).

La proteína mutada parece asumir una conformación anormal que favorece su acumulación en núcleos y citoplasmas, llamados cuerpos de inclusión (15). A medida que la enfermedad progresa, la Ataxina-1 mutada, cuya estructura tridimensional se modifica por el efecto de contener expansiones de glutaminas, no se elimina

adecuadamente. Se agrega en el núcleo celular y recluta, a través del dominio de poliglutaminas, una variedad de proteínas, incluyendo factores de transcripción, chaperonas (proteínas involucradas en el plegado de proteínas), así como otros componentes de la maquinaria degradadora del proteasoma (16). Dicha agregación conduce a la disminución de la función correguladora de la Ataxina-1 en la síntesis de los genes diana de hormonas retinoideas, tiroideas y del receptor D2 de la dopamina. Siendo la causa principal de la degeneración neuronal progresiva y de los síntomas extrapiramidales (16).

2.3. Poliglutamínica

La cadena poliglutamínica es un fragmento de la proteína, que consiste en la repetición del aminoácido glutamina (Q). En algunas enfermedades neurodegenerativas hereditarias se encuentra una mutación que produce una cantidad excesiva y anormal de glutaminas, produciendo una proteína propensa a agregarse. Estas proteínas agregadas causan la muerte de las células, sobre todo, neuronas, por varios mecanismos. Algunas enfermedades neurodegenerativas hereditarias causadas por esta expansión de glutaminas son la enfermedad de Huntington (HD), la atrofia muscular espinobulbar (SBMA o enfermedad de Kennedy), varias ataxias espinocerebelares (SCA) y la atrofia dentatorrubropalidoluisiana (DRPLA) (18).

Las similitudes identificadas entre las enfermedades poliglutamínicas permiten establecer una hipótesis general acerca de un mecanismo neuropatológico común, según el cual, la toxicidad neuronal puede ser una consecuencia de la expansión poliglutamínica *per se*, con la selectividad regional de degeneración conferida por las secuencias de aminoácidos adyacentes al dominio de poliglutaminas. De hecho, ya ha sido comprobado que la poliglutamina expandida es capaz, por sí misma, de inducir la muerte celular (62).

El número normal de glutaminas repetidas en personas sanas es de 6 a 44 aproximadamente, mientras que en personas afectadas es de 36 a 121, aproximadamente, según la enfermedad. La causa del exceso de glutamina es una expansión de la repetición del triplete CAG, que codifica este aminoácido. Se pensaba que esto ocurría durante la replicación en un proceso conocido como “*replicación*

slippag” pero de acuerdo a diversos estudios se ha determinado que esta expansión ocurre durante la reparación del ADN, debido a que la cantidad de glutaminas tiende a aumentar de una generación a la siguiente, este fenómeno es conocido como “*mutaciones dinámicas*”. Esta es la principal causa de que los síntomas vayan apareciendo antes en las generaciones futuras, relacionándose una mayor cantidad de repeticiones con una mayor gravedad de los síntomas (19).

Las propiedades únicas de la poliglutamina expandida podrían conducir a interacciones proteína-proteína alteradas, incluyendo la agregación. Una hipótesis alternativa es que la poliglutamina expandida constituye un sustrato para las transglutaminasas. De la interacción entre ambas resultaría la unión covalente de las poliglutaminas a otras proteínas o a sí mismas (63).

Uno de los mecanismos que causan la muerte celular es que la poliglutamina expandida interfiere con el sistema de degradación dependiente de ubiquitina, siendo las proteínas mutadas agregadas más resistentes a dicha degradación. En el caso de la enfermedad de la SCA1 se sabe que mientras que la Ataxina-1 normal expandida esta igualmente poliubiquitinada, en neuronas de enfermos con SCA1, la forma expandida es tres veces más resistente a la degradación proteasómica (20). También ha sido demostrado en ratones transgénicos y en cerebros de pacientes SCA1, que la Ataxina-1 expandida causa la represión secuencial de 6 genes neuronales muy abundantes en las células de Purkinje y relacionados con la traducción de señales y la homeostasis del calcio (21).

2.4. Bioinformática

La Bioinformática es una disciplina científica que utiliza técnicas informáticas para organizar, analizar y distribuir información biológica con la finalidad de responder preguntas complejas en biología y microbiología molecular. Es un área multidisciplinaria, la cual puede ser ampliamente definida como una interfase entre la Biología y la Computación. Inicialmente el concepto de Informática se refería sólo a la creación y mantenimiento de bases de datos donde se almacena información biológica, tales como secuencias de nucleótidos y aminoácidos. Pero hoy día, la aplicación de la Bioinformática implica solucionar o investigar problemas sobre escalas de tal magnitud

que sobrepasan las capacidades humanas usando para esto herramientas de sistemas y computación (22).

Las herramientas de software para Bioinformática van desde simples herramientas de línea de comandos hasta complejos programas gráficos y servicios web autónomos situados en compañías de Bioinformática o instituciones públicas. Los principales campos de la Bioinformática incluyen alineamientos de secuencias, predicción de genes, montaje de genomas, alineamientos estructurales de proteínas, predicciones de estructura de proteínas, predicciones de expresión génica, interacciones proteína-proteína, y modelado de la evolución (22).

La Bioinformática se ha empleado en diversas investigaciones biomédicas, sobre todo en enfermedades genéticas es así que se han llevado a cabo investigaciones sobre enfermedades poliglutaminicas y neurodegenerativas (23) como lo es la Ataxia Espino Cerebelosa Tipo 1.

Para el desarrollo de esta investigación se utilizaron varios software, entre los que se encuentra GROMACS, el cual es un motor para realizar simulaciones de dinámica molecular y minimización de energías. Estas son dos de las muchas técnicas pertenecientes al ámbito de la química computacional y modelización molecular.

La modelización molecular pretende describir los sistemas químicos complejos en términos de un modelo atómico realista, con el objetivo de entender y predecir las propiedades macroscópicas basada en un conocimiento detallado a escala atómica (24).

Las propiedades macroscópicas son siempre un promedio sobre un conjunto estadístico representativo (en equilibrio o en desequilibrio) de sistemas moleculares. Para el modelado molecular, esto tiene dos consecuencias importantes (24):

- El conocimiento sobre una estructura aislada, incluso si es una estructura con una energía global mínima, no es suficiente. Es necesario generar un conjunto representativo a una temperatura determinada, para poder calcular sus propiedades macroscópicas (24).

- Aunque las simulaciones moleculares proporcionan en principio detalles atómicos de estructuras y sus movimientos, tales detalles no suelen ser relevantes para analizar las propiedades macroscópicas. Esto abre el camino a la simplificación en la descripción de las interacciones y obviar detalles irrelevantes. La ciencia de la mecánica estática proporciona el marco teórico para hacer estas simplificaciones. Existe una jerarquía de métodos que van desde considerar los grupos de átomos como una unidad, describiendo el movimiento en un número reducido de colectivos coordenadas (24).

Para la generación de un conjunto representativo en equilibrio, hay dos métodos disponibles:

- a) Simulaciones de Monte Carlo
- b) Simulaciones de Dinámica Molecular (DM).

Para la generación de conjuntos en desequilibrio y para el análisis de eventos dinámicos, sólo el segundo método es apropiado.

Mientras que las simulaciones de Monte Carlo son más simples que la DM, porque no requieren el cálculo de fuerzas y porque no ofrecen estadísticas significativamente mejores; es por ello que la Dinámica Molecular es la técnica más universal.

Si la configuración de partida está muy lejos del equilibrio, las fuerzas pueden ser excesivamente grandes, y la simulación DM puede fallar. En estos casos, una fuerte minimización de energía es necesaria. Otra razón para llevar a cabo una minimización de energía, es la eliminación de toda la energía cinética del sistema: si varias instantáneas de simulaciones dinámicas han de compararse, la minimización de energías reduce el ruido térmico y las energías potenciales en las estructuras, de modo que se puedan comparar mejor (24).

2.4.1. Dinámica Molecular (DM)

La dinámica molecular es una técnica de simulación por ordenador en la que átomos y moléculas interactúan durante un período de tiempo por aproximaciones de la física

conocida, dando una visión del movimiento de las partículas. Podríamos describir estas simulaciones por ordenador como una especie de "microscopio virtual" con alta resolución temporal y espacial. Considerando que es posible tomar instantáneas de las estructuras, que no podrían tomarse en experimentos convencionales y permitiendo el acceso a todas las escalas de tiempo y movimiento con resolución atómica.

Por lo tanto, la dinámica molecular permite escudriñar en el movimiento de los átomos individuales de un modo que no es posible en experimentos de laboratorio. Para realizar estas simulaciones, es necesaria la representación en el ordenador de todos los átomos y moléculas que intervienen en el proceso. Estas simulaciones conllevan un gran consumo de recursos, tanto de memoria como de capacidad de procesamiento. Es por tanto habitual, la utilización de ordenadores de gran rendimiento y en muchas ocasiones con grandes capacidades para la computación paralela o distribuida.

La dinámica molecular hace uso de las componentes del sistema en un sencillo esquema formado por: la energía potencial, la ecuación dinámica de cada una de las partículas, donde se obtiene la aceleración y la velocidad para obtener finalmente las coordenadas y por consiguiente las trayectorias.

Integrando las ecuaciones de movimiento de Newton mediante de métodos numéricos; se espera generar trayectorias exactas sobre intervalos largos de tiempo; siendo lo más importante el comportamiento estadístico de la trayectoria para asegurar que las propiedades termodinámicas y dinámicas del sistema estén siendo obtenidas con una predicción adecuada (26).

Cabe señalar que la dinámica molecular no es una simulación de la cinética química, pues su naturaleza no es estadística y macroscópica sino casuística y molecular. Todas las formas de dinámica molecular se basan en representar el comportamiento de los sistemas moleculares de acuerdo con la segunda ley de Newton, donde la fuerza aplicada sobre un sistema es proporcional a su aceleración o variación de su velocidad con el tiempo (25).

Así, los requerimientos para poder realizar una simulación de dinámica molecular son:

- Disponer de una hipersuperficie o función de energía potencial, que exprese las energías entre las partículas individuales de forma suficientemente confiable (25).
- Resolver la ecuación de movimiento de forma tal que si se tienen las posiciones y velocidades de cada partícula en un tiempo t , se pueda obtener cuales serán tales posiciones y velocidades en un tiempo $(t + \Delta t)$, donde el intervalo Δt es finito (25).

El resultado de una simulación de dinámica molecular son las posiciones X y velocidades V de cada átomo de la molécula, para cada instante en el tiempo. A esto se le llama trayectoria (25).

Resumiendo, se puede indicar que las simulaciones de DM consisten en la integración de las ecuaciones de movimiento de Newton para determinar cómo varía la posición de las partículas en función del tiempo.

Una vez conocidas las diferentes trayectorias se pueden calcular diferentes propiedades macroscópicas. Algunas de estas propiedades son:

- Energía (cinética, potencial y total)
- Temperatura
- Presión
- Capacidad calorífica
- Distribución radial
- Constante dieléctrica

Una simulación general de DM normalmente consta de tres etapas (25):

- 1) Primera etapa: iniciación: en esta etapa cada partícula se coloca en un lugar determinado del espacio y se le aplica una velocidad al azar, dependiendo de la temperatura que tenga el sistema. Se deja evolucionar el sistema con el tiempo, logrando que las velocidades de las partículas sean cero, en el momento lineal

del sistema. El propósito de esto es relajar el sistema eliminando el exceso de energía.

- 2) Segunda etapa: esta segunda etapa es la continuación de la primera pero sin renormalizar las velocidades, con el fin de estar seguro que se ha alcanzado el verdadero equilibrio. Durante este proceso se obtiene y conserva la información sobre la historia de las partículas, fundamentalmente sus *trayectorias*, es decir sus posiciones y velocidades en función del tiempo, ya que a partir de ellas se podrán obtener diferentes propiedades del sistema.
- 3) Tercera etapa: Obtención de resultados: en esta etapa se realiza un análisis de las trayectorias para poder determinar algunas de las propiedades dinámicas y estáticas del sistema.



2.4.2 Elementos de la Dinámica molecular

A) Solvatación Implícita (35)

Surge como una alternativa para reducir el costo computacional que genera la solvatación explícita.

Consiste en reemplazar las moléculas de agua por un sistema “virtual” de aguas. Para ello se modela un medio infinito continuo con las propiedades dieléctricas e hidrofóbicas del agua.

Este modelo está basado en la teoría clásica de *Poisson-Boltzmann* (PB). En ella, el soluto se detalla a nivel atómico, mientras que las moléculas de solvente y posibles electrolitos, se tratan como un continuo sin estructura, caracterizado por una constante dieléctrica del solvente (ϵ_s). En el interior de la cavidad del soluto, la constante dieléctrica toma valores característicos de proteínas ($\epsilon_{int}=2-8$) o 1.

Aparte del coste computacional más reducido, este modelo presenta una serie de ventajas frente a la representación explícita del agua como evitar el equilibrado del sistema (temperatura y presión); el soluto puede explorar más rápidamente el espacio de fases debido a la ausencia de viscosidad asociada a los modelos explícitos; se modela la solvatación en un volumen infinito, evitándose artefactos del sistema periódico y se facilita la estimación de energías de estructuras solvatadas.

Sin embargo, por otra parte se pierde también la posibilidad de analizar interacciones estructurales soluto-solvente, como la formación de puentes de hidrógeno.

B) Campos de fuerza (64)

El término campo de fuerza se utiliza para describir al conjunto de ecuaciones potenciales y parámetros que tratan de representar los diversos tipos de interacciones existentes en un sistema. Los campos de fuerza para simulación de proteínas están compuestos por dos componentes, una que modela las interacciones de enlace, funciones que permiten mantener la geometría local de las proteínas, y las interacciones

de no enlace, las cuales modelan la interacción entre átomos que se encuentran separados por más de dos átomos. Los términos que representan las interacciones de enlace describen el estiramiento de los enlaces, el doblamiento de los ángulos y la rotación de los diedros. Mientras que los términos que representan las interacciones de no enlace describen la electrostática, dispersión y exclusiones de *Pauli*. Por lo tanto los términos de Energía para un campo de fuerza se dan de la siguiente manera:

$$E_{\text{bonded}} = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_x [1 + \cos(nX - \sigma)] \quad (2.1)$$

$$E_{\text{nonbonded}} = \sum_{\text{nonbondedpairs } ij} \left(\epsilon_{ij} \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 + \frac{q_i q_j}{r_{ij}} \right) \quad (2.2)$$

La primera de las interacciones enlazantes modela el comportamiento vibracional de los enlaces. En la segunda ecuación se describe las interacciones de no enlace, esta ecuación se compone de dos partes la primera es conocida como la ecuación de *Lennard-Jones* (LJ) y la segunda parte que es la *ley de Coulomb*.

Podemos también describir al campo de fuerza como un campo escalar que implica una superficie de energía potencial que no es posible visualizar debido al gran número de grados de libertad involucrados. Eso se debe a que un campo de fuerza describe una serie de parámetros específicos para cada tipo de átomo.

Los parámetros típicos son la masa atómica, el radio de *Van der Waals* carga parcial para átomos individuales, valores de equilibrio para el largo de los enlaces y los ángulos de enlace.

Los campos de fuerza se basan en numerosas aproximaciones obtenidos de diversos datos experimentales, es por ello que son llamados empíricos. Se han desarrollado diferentes campos de fuerza, pero los más comunes y utilizados son AMBER (Assisted Model Building and Energy Refinement), CHARMM (Chemistry at Harvard Molecular Mechanics) y OPLS (Optimized Potential for Liquid Simulations).

C) Energía Potencial

Es la capacidad que tiene un sistema para realizar un trabajo en función exclusivamente de su posición o configuración. Puede pensarse como la *energía almacenada* en el sistema. Un ejemplo sencillo sería un ladrillo sostenido en alto, este tiene energía potencial debido a su posición en relación al suelo, efectuaría un trabajo si se suelta, ya que caerá al piso debido a la fuerza de gravedad, pudiendo efectuar algún trabajo sobre otro objeto que se interponga en su caída (65).

En dinámica molecular se entiende que la energía potencial, es la energía que la molécula tiene según su posición en un campo de fuerza. La energía potencial que podrá tener una molécula se deberá a las fuerzas de atracción y repulsión con otras moléculas de su entorno (65).

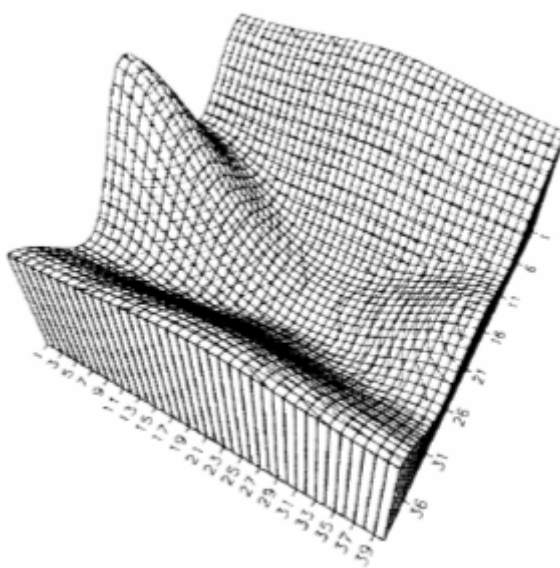


Figura 2.3: Ejemplo de una superficie de energía potencial tridimensional, en la que se muestra un mínimo (66).

Sabemos que la posición relativa de los átomos de una molécula, o de un grupo de moléculas, determina la energía del sistema. Si fuera posible representar gráficamente el cambio en la energía potencial en función de cada uno de los cambios en las geometrías posibles para el sistema molecular, se obtendría como resultado la superficie de energía potencial total correspondiente. Puede hacerse entonces una analogía entre la *superficie de energía potencial* (SEP) y una malla extendida horizontalmente, la cual presenta

deformaciones semejantes a montes y valles. La posición más baja de los valles correspondería a geometrías que son mínimos en la SEP, mientras que los puntos más altos de los montes corresponderían a geometrías que son máximos estados de transición en la misma superficie (66).

La SEP depende de los grados de libertad de un sistema que han sido llevados a un espacio tridimensional: Una molécula o un sistema de varias moléculas con un número N de átomos, tiene en total $3N$ grados de libertad. Cada grado de libertad corresponde a los movimientos posibles de cada uno de los átomos en relación a un sistema de ejes coordenados tridimensional, ya que cualquier movimiento se puede definir en función de un cambio en las coordenadas x , y y z de cada átomo. De los $3N$ grados de libertad, 6 de ellos corresponden a traslaciones y rotaciones del sistema como un todo, es decir a movimientos de todos los átomos en la misma dirección y al mismo tiempo (27).

Estos 6 grados de libertad (5 para sistemas lineales) no cambian la energía de la molécula, pues la posición relativa de los átomos no se modifica. Por lo tanto, el número total de grados de libertad que determinan la SEP de cualquier sistema molecular se reduce a $3N - 6$ ($3N - 5$ para sistemas lineales) (27).

Considerando ahora al sistema formado por dos átomos, ya sea enlazados covalentemente o a través de algún otro tipo de interacción, se tendría un total de $3(2) - 5 = 1$ grado de libertad. Este modo normal correspondería a la variación de la distancia interatómica entre los dos átomos. Luego entonces, para este sistema diatómico, la SEP estaría comprendida únicamente por dos dimensiones. Una dimensión sería la distancia diatómica y la otra la variación de la energía en función del cambio en la distancia. Esto se vuelve mucho más complejo a medida que aumenta el número de átomos que componen el sistema (27).

Como se mencionó anteriormente las superficies de energía potencial permiten determinar los mínimos presentes, los que se encuentran en función de cada una de las dimensiones de la SEP y los máximos o estados de transición, que determinan la velocidad a la cual ocurre la interconversión entre dos mínimos (27).

A medida que el sistema molecular se vuelve más flexible, el número de mínimos tiende

a aumentar, dando como resultado superficies de energías potenciales multidimensionales y con una topología extremadamente compleja. El conocimiento del mayor número de mínimos posible sobre la SEP nos ayuda a comprender mejor las propiedades de un sistema molecular.

D) Puntos estacionarios en una superficie de energía potencial (67)

Se reconocen dos tipos de puntos críticos de la superficie de potencial como importantes para la modelación de reacciones químicas: los mínimos locales o puntos críticos de orden cero, que corresponden a todas las moléculas e intermediarios estables de reacción (reactivos, productos, intermediarios), y los puntos críticos de orden uno o puntos de silla de orden uno, que representan estados de transición.

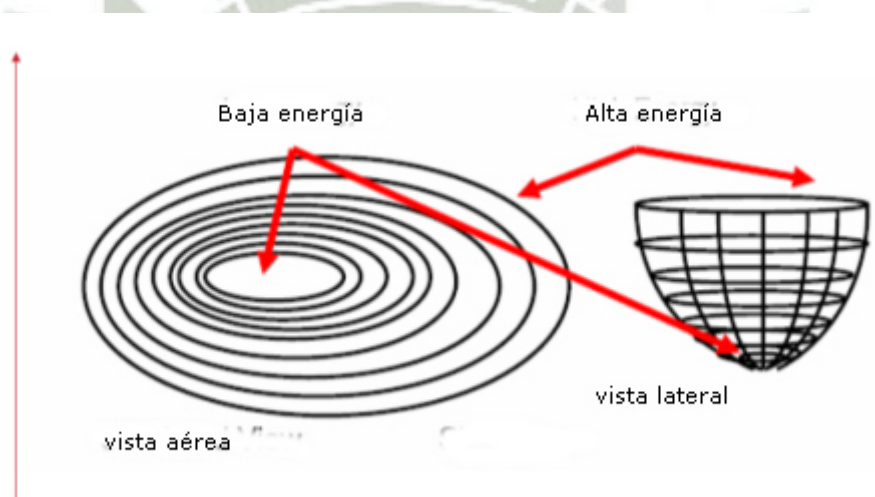


Figura 2.4: Vista superior y lateral de los puntos críticos que se encuentran en la Superficie de Energía Potencial, el punto de menor y mayor energía (67).

El cálculo de puntos críticos de una superficie de energía potencial se hace a partir de las primeras derivadas mediante un algoritmo de minimización local o de minimización restringida (para los estados de transición). En ambos casos se dice que se está llevando a cabo un cálculo de optimización.

Para asegurarse de que los puntos estacionarios determinados son efectivamente mínimos o puntos de silla es necesario calcular las segundas derivadas de la energía (lo que se conoce como un cómputo de la matriz Hessiana) y analizar los signos de esas derivadas en cada una de las direcciones independientes del problema. A esto último se le conoce como un análisis vibracional o un cálculo de las vibraciones moleculares.

E) Minimización

Los modelos tridimensionales obtenidos a partir del formato FASTA de cualquier proteína o estructura adoptan una infinidad de geometrías posibles que no es posible definir las más que en términos probabilísticos. Aunque las conformaciones más probables son aquellas con un menor contenido energético en relación a las otras (67).

También se debe tener en cuenta que la energía del agregado es función tanto de su geometría como de otras variables, por ejemplo, el ambiente químico en que se encuentran. El conocimiento de las geometrías permite identificar aquella que tenga la mayor probabilidad de presentarse, es decir, la geometría de menor energía: El mínimo global (67).

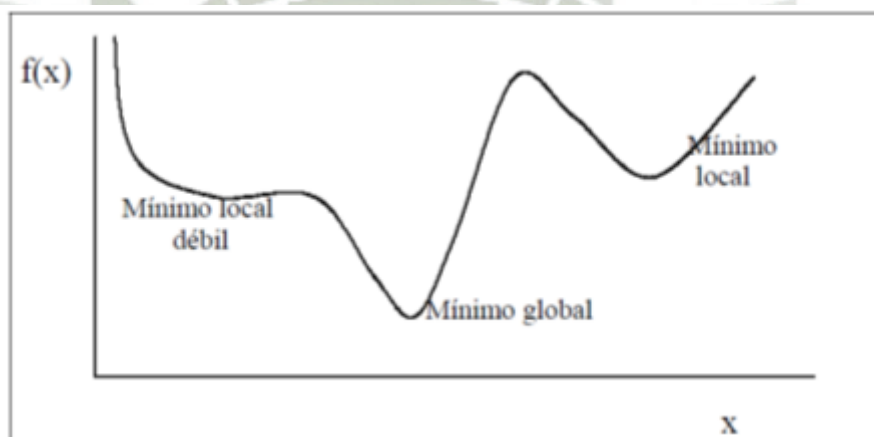


Figura 2.5: El mínimo global, es el menor valor de la función en cualquier otro valor de x , el mínimo local, sólo se cumple en un intervalo que el valor de la función en ese punto es menor que la función en otro punto (67)

El conocimiento del mínimo global, y de otras geometrías de menor energía, permite la elucidación de propiedades del sistema. La tarea de caracterizar el mínimo global por medio de métodos computacionales no es simple, pues requiere de la obtención inicial de los mínimos locales. Si bien existe una mayor probabilidad de identificar al mínimo global con la obtención de un mayor número de mínimos locales, la obtención de éstos depende, finalmente, de la eficiencia del algoritmo computacional utilizado para determinar los diferentes mínimos.(67)

Los métodos de minimización locales son generalmente métodos descendentes, el procedimiento básico consiste en caminar sobre la superficie potencial en la dirección en que la energía decrece de manera que el sistema es llevado a un mínimo de energía

local próximo. La dirección descendente se establece a partir del gradiente, es decir se desciende por el camino de mayor gradiente (67).

Par el proceso de minimización se utilizan métodos del gradiente o derivativos, este tipo de métodos extrae información de la primera y a veces de la segunda derivada para acelerar el proceso de minimización(67).

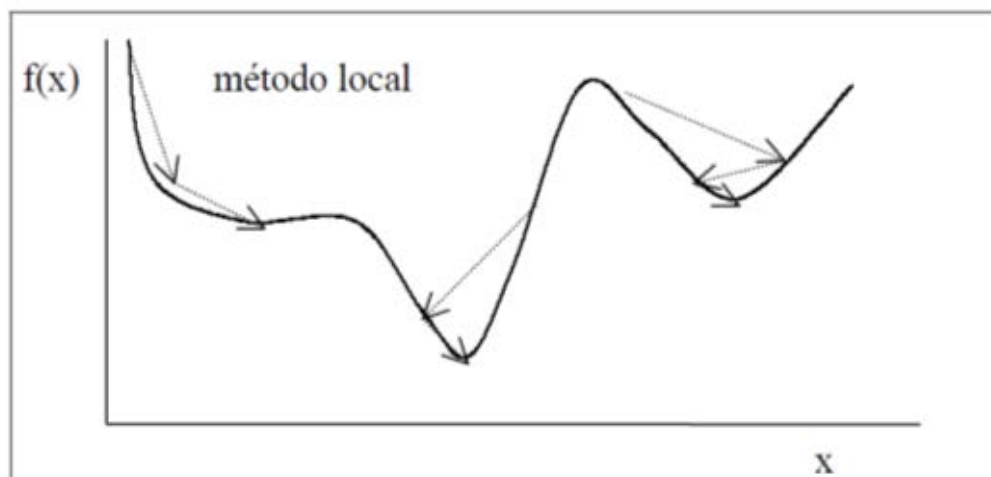


Figura 2.6: Búsqueda del mínimo local, a partir de un valor inicial X_0 se va pasando por diferentes valores hasta alcanzar el valor mínimo de energía (X_m) (67).

El problema de Minimización puede ser formulado de la siguiente manera: Dada una función f que dependa de un conjunto de variables independientes x_1, x_2, \dots, x_i , minimizar la función equivale a encontrar los valores de las variables independientes (x^*) para que f tenga un valor mínimo, se representaría así (68):

$$V(x^*) = \min (V(x)) \quad (2.6)$$

Donde V es la energía potencial de la molécula y “ x ” son las coordenadas atómicas. 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. Donde el punto mínimo de la primera derivada es cero y la segunda derivada son todos los valores positivos (68).

La idea de los procesos de minimización energética es conseguir que el gradiente negativo de la Energía potencial se dirija siempre hacia un descenso en V y de forma

perpendicular a la superficie de energía (68).

Los Métodos de Minimización de Energía más usados para dinámica molecular son aquellos que utilizan información de la forma de la superficie de Energía potencial, en base a las primeras derivadas y segundas derivadas del campo de fuerza. El algoritmo más simple que usa la primera derivada es el método del descenso rápido (*Steepest Descent*) (68).

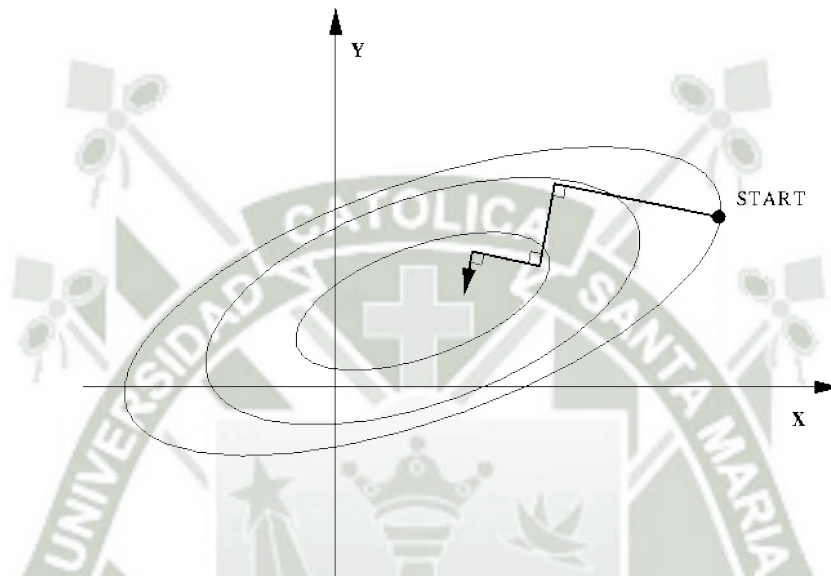


Figura 2.7: El método de *Steepest Descent* se aproxima al mínimo en forma de zig-zag, donde cada nuevo lugar de búsqueda es ortogonal al anterior (68).

F) Método del Descenso por Etapas o *Steepest Descent* (SD) (28)

Este método es el más simple de los métodos por gradiente. La selección de la dirección es hacia donde la función f decrezca con mayor rapidez, la cual es la opuesta al gradiente. La búsqueda empieza en un punto arbitrario y se desliza a través del gradiente, hasta que se esté lo suficientemente cerca de la solución. Una forma gráfica de cómo actúa este método se puede observar en la Figura 2.7.

Un parámetro a definir para este método es de qué tamaño deben ser los pasos (λ_k), con los cuales se quiere mover el punto hacia donde f toma un valor mínimo. Esto se da donde la derivación direccional es cero, lo que indica que $\nabla f(x_{k+1})$ y $g(x_k)$, donde éste último es el gradiente en un punto, deben ser ortogonales. El siguiente paso es dado en la dirección negativa de gradiente en el nuevo punto, describiendo así una trayectoria en

zig-zag como se muestra en la figura 2.8. La iteración continúa hasta que se llegue a un valor aceptable de error.

El tamaño de los pasos se reajusta dependiendo del valor de la energía resultante. Si la energía disminuye, el tamaño del paso aumenta; y si la energía aumenta se asume que la magnitud de descenso ha sido demasiado grande y se reduce el tamaño del paso (generalmente en un 20%).

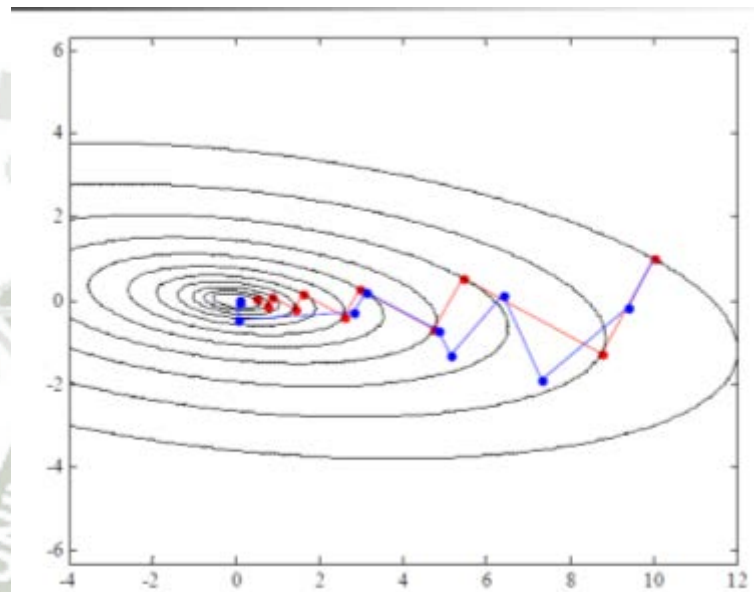


Figura 2.8: Búsqueda de mínimo local, el color rojo representa una búsqueda exacta, el color azul una búsqueda con 50% de error (69).

En forma matemática el proceso de *Steepest Descent* se representa:

Para determinar la dirección del paso o movimiento, la ecuación es la siguiente:

$$s_k = -g_k / |g_k| \quad (2.7)$$

Donde s_k es el vector multidimensional que representa la dirección del movimiento, la cual debe ser opuesta al gradiente (g_k).

Una vez elegida la dirección del paso se debe asignar el tamaño del paso, con la siguiente ecuación:

$$x_{k+1} = x_k + \lambda_k s_k \quad (2.8)$$

Donde λ_k es el tamaño del paso, obviamente el tamaño del paso va a tener que ver con la característica topológica de la superficie cerca del mínimo.

La elección del tamaño del paso es arbitraria. Inicialmente tiene un valor por defecto. Si la primera iteración lleva a una disminución de la energía, el tamaño del paso se incrementa por un factor para la siguiente iteración. Si en cambio, la iteración produce una energía mayor, se reduce el tamaño pues se interpreta que se paso por encima del mínimo en esa dirección.

Teniendo definida la dirección por la cual moverse, es necesario decidir cuán lejos se puede mover a lo largo de la gradiente. Si nosotros imaginamos un corte transversal en la superficie, la superficie pasara a través un mínimo y después incrementara.

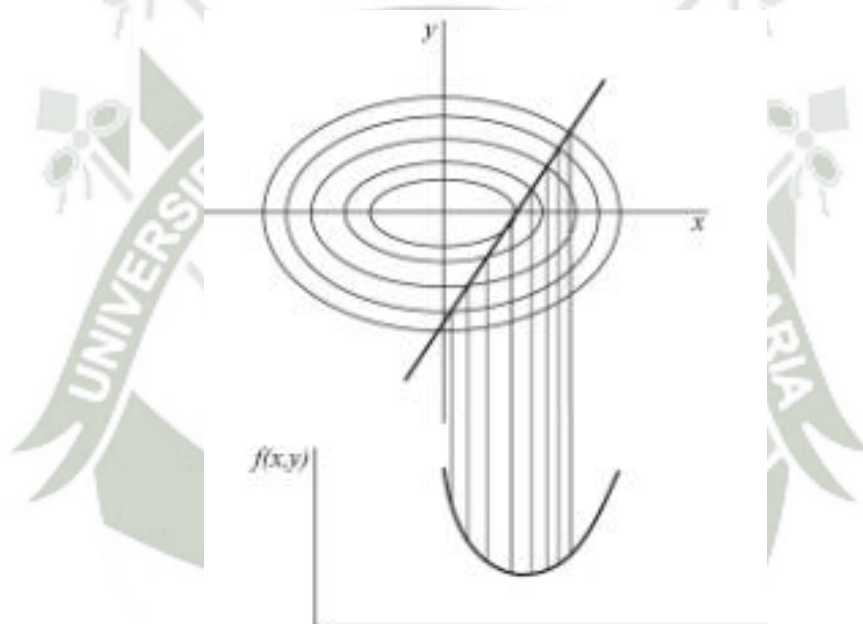


Figura 2.9: Búsqueda lineal en una dimensión (69)

Lo que se hace es encontrar tres puntos a lo largo de la línea de manera que el punto central tenga menos energía que los dos más externos. Si se encuentran, por lo menos debe haber un mínimo entre los dos puntos externos. Se sigue un procedimiento iterativo de manera de disminuir la distancia entre los puntos de manera de restringir gradualmente el mínimo a una región pequeña.

Sin embargo, al calcular la nueva dirección siempre se obtiene una perpendicular a la anterior, no importando la forma del potencial local. Este comportamiento hace que las soluciones oscilen alrededor del mínimo si este es muy amplio. Para evitar este

comportamiento de oscilación cerca del mínimo, se puede utilizar un algoritmo de gradiente conjugado. En este método, el primer paso es igual al del método *steepest descent*, sin embargo la nueva dirección no se hace perpendicular a la anterior si no que se sopesa con la dirección anterior.

G) Método del Gradiente Conjugado (GC) (69)

Este método se desarrollo para poder resolver algunos problemas que el método de *Steepest Descent* no podía, como la cantidad de interacciones necesarias para alcanzar el punto mínimo, Para agilizar este problema se desarrollo el método de *Conjugate Gradient* que utiliza otro algoritmo para la búsqueda del mínimo.

En SD, tanto los gradientes como las direcciones entre dos pasos sucesivos son ortogonales, en cambio en GC, sólo los gradientes son ortogonales, las direcciones son conjugadas.

El primer paso de GC es como en SD , pero subsecuentes búsquedas son resueltas a lo largo de una línea la cual es una mezcla del actual gradiente negativo y las previas direcciones de búsqueda.

$$\mathbf{v}_k = -\mathbf{g}_k + g_k \mathbf{v}_{k-1} \quad (2.9)$$

Donde g_k es la gradiente de la interacción k y \mathbf{v}_{k-1} es la dirección del paso $k-1$.

Existen diversas formas de escoger los valores β . Algunos de los nombre asociados con este método son: *Fletcher-Reeves* y *Polak-Ribiere*. La variante de *Polak-Ribiere* tiene importancia sólo si la función no es cuadrática, ya que funciona mejor en dichos casos.

2.4.3 Termodinámica Estadística (70)

La *mecánica estadística* proporciona el nexo de unión entre la descripción *mecánica* (cuántica o clásica) y *termodinámica* de un sistema macroscópico. Su objetivo es deducir las propiedades macroscópicas de un sistema (entropía, capacidad calorífica, tensión superficial, viscosidad) a partir de las propiedades microscópicas (geometría molecular, interacciones intermoleculares y masas moleculares).

La mecánica estadística se asienta sobre tres pilares: el punto de partida (la mecánica), el punto de llegada (la termodinámica) y el camino entre ambos (la estadística). La mecánica estadística se divide en dos partes:

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

La termodinámica estadística conecta dos descripciones de un sistema macroscópico, es decir, un gran número de partículas o moléculas (N). El estado de un sistema (cada una de las formas en que puede presentarse) se puede especificar de dos maneras distintas, distinguiendo entonces entre:

- Estado Macroscópico o *Macroestado*: El estado del sistema se define por los valores de variables macroscópicas denominadas funciones de estado que vienen relacionadas por una ecuación de estado. Para una sustancia pura en equilibrio el estado de un sistema queda definido por tres variables. Por ejemplo la Presión (P), la Temperatura (T) y el número de moles (n).
Otras variables macroscópicas (el Volumen, V , por ejemplo) pueden obtenerse mediante la ecuación de estado correspondiente. Como la ecuación de volumen para un gas ideal o la de *Van der Waals* si es un gas real.
El macroestado en este caso queda completamente especificado usando tres variables o funciones de estado, ya que cualquier otra magnitud puede obtenerse a partir de ellas.
- Estado microscópico o *Microestado*: La forma de especificar el estado microscópico de un sistema depende de si utilizamos la mecánica clásica o la cuántica en su descripción.
En mecánica clásica el estado microscópico de un sistema de N partículas queda especificado cuando se conocen las coordenadas y velocidades de todas ellas en un instante t dado:

$$x_1, y_1, z_1, x_2, \dots, x_N, y_N, z_N \equiv \vec{r}_1, \vec{r}_2, \dots, \vec{r}_N$$

Se necesitan por tanto $6N$ variables, a partir de las cuales podemos calcular las propiedades del sistema, tales como la energía total, suma de la cinética (K) y potencial (V_p).

$$E = K + V_p = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 + V_p(\vec{r}_1, \dots, \vec{r}_N) \quad (2.10)$$

La energía del sistema dependerá del número de partículas y el volumen disponible $E = E(N, V)$. En mecánica cuántica el estado queda definido por la función de onda del sistema, si hay N partículas necesitaremos $4N$ números cuánticos para especificar la función de onda (por ejemplo para un electrón necesitamos cuatro números cuánticos: n, l, m y s). Conocida la función de onda podemos calcular la energía del sistema siendo la energía, como en el caso anterior, función de N y V .

Por supuesto las descripciones micro y macroscópicas no son independientes entre sí. Así que ¿Cuál es la relación entre ambas?, para explicarlo supongamos que tenemos un sistema en un recipiente cerrado de paredes rígidas y conductoras en contacto con un baño termostático. En este macroestado se puede especificar fácilmente mediante los valores del número de partículas, volumen y temperatura (por ejemplo 1 mol de gas a 298 K y ocupando un volumen de 20 L). Si observamos el sistema en distintos momentos su estado macroscópico no cambiará (está en equilibrio). Sin embargo, las moléculas que lo componen están en continuo movimiento cambiando rápidamente sus coordenadas y velocidades. Es decir, tenemos un único macroestado pero muchos microestados diferentes compatible con él. Si medimos las variables macroscópicas veremos que corresponden a un promedio temporal de los microestados producidos durante la medida.

Como se ha mencionado un sistema posee determinados estados microscópicos y las variables macroscópicas se pueden calcular si conocemos la probabilidad de ocupación de los microestados. Para conocer las propiedades macroscópicas ya no necesitamos

seguir la evolución temporal sino conocer las probabilidades de cada microestado (Imaginemos el lanzamiento de un dado, para conocer el valor medio de varios lanzamientos no hace falta realizarlos y anotar el resultado, sino simplemente conocer la probabilidad de que salga cada uno de los valores). Esta relación se conoce como *hipótesis ergódica*.

Sin embargo, no todas las propiedades macroscópicas del sistema pueden calcularse de esta forma ya que algunas propiedades no tienen su equivalente microscópico, es decir, no están definidos para un determinado microestado. Podremos calcular todas aquellas propiedades que dependan de la energía o de la velocidad/posición de las partículas, tales como la presión o la energía interna, es decir las propiedades llamadas mecánicas. No podremos calcular por ejemplo la entropía, ya que no existe un valor de la entropía para cada microestado. La entropía depende de cómo el sistema se reparte entre los microestados y no del valor que toma una propiedad en cada microestado.

La relación entre estados microscópicos y macroscópicos viene dada por la famosa fórmula de Ludwig Boltzmann de la entropía:

$$S(E, N, V) = k_B \log(\Omega) \quad (2.11)$$

donde Ω es el número de el numero de complejiones o probabilidades termodinámicas y k_B es la constante de Boltzmann, que relaciona la temperatura absoluta y la energía.

En el término de la izquierda tenemos la termodinámica mediante la entropía definida en función de sus variables naturales, lo que da una información termodinámica completa del sistema. A la derecha tenemos las configuraciones microscópicas que definen la entropía mediante esta fórmula.

Ésta es la llamada ecuación puente en el Colectivo Micro Canónico. Existen otros colectivos, como el Colectivo Canónico o el Colectividad macrocanónica, conocidos como ensambles estadísticos, que representa todas las posibles configuraciones del sistema y la probabilidad de realizar cada configuración.

Cada ensamble está asociado con una función matemática que puede ser usada para extraer valores de las propiedades termodinámicas del sistema (29). De acuerdo a la relación del sistema con respecto al universo, uno de los tres tipos generales de ensambles pueden ser utilizados.

- Ensamble microcanónico: describe a sistema aislado completamente, teniendo constante la Energía, por lo que se entiende que no se intercambia Energía o masa con el resto del universo.
- 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 gran canónico: usado para sistemas abiertos en donde hay intercambio de masa y Energía con el exterior.

2.4.3.1 Ensamble canónico (71)

Tratare de explicar matemáticamente como se obtiene el ensamble canónico o NVT , llamado así porque mantiene constante el número de partículas N , el volumen V y la temperatura T .

Para entenderlo pondré un ejemplo sencillo, una bañera caliente, es decir un sistema físico en contacto con un baño térmico. Esto quiere decir que está en contacto con una gran masa a una temperatura dada, y por el principio cero de la termodinámica tendremos por tanto el sistema en equilibrio termodinámico con el baño. En estas condiciones, la energía no está totalmente determinada, sino que es una variable aleatoria que puede tomar una serie de valores. De esta forma, sólo podemos hablar de probabilidad de que el sistema adopte una energía determinada en función de esta temperatura.

Ahora supongamos que la bañera de área transversal A e imaginemos que hacemos k divisiones en la bañera de áreas ak . Luego, supongamos que tenemos N bolas y que

hacemos el experimento de lanzarlas a la bañera. Probablemente, en el primer lanzamiento caerán n_1 bolas en la celda a_1 , n_2 en la celda a_2 y así sucesivamente. Si repetimos el experimento el resultado será diferente, pero si el experimento lo repetimos un número muy grande de veces, encontraremos que siempre habrá una distribución que se repite más veces que las otras. Esta distribución, es la *distribución más probable*.

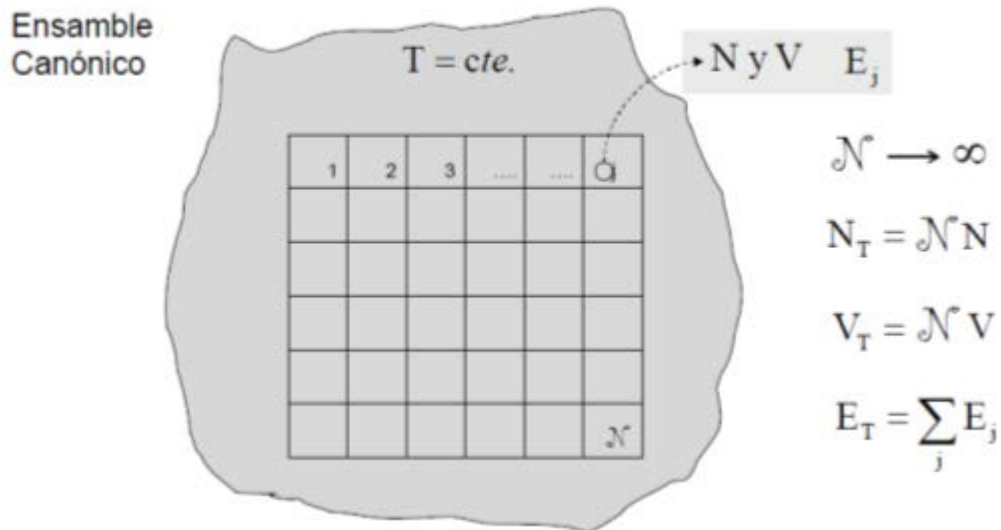


Figura 2.10: Representación del Ensamble Canónico, posee temperatura, número de partículas y volumen constante, permitiendo intercambio de energía, entre el sistema.

Una vez aclarado el principio de probabilidad, supongamos que en vez de tener una bañera tenemos una distribución de energía y que las celdas son particiones de energías ϵ_k . Y que las bolitas que arrojamus son ahora partículas o moléculas.

Como dijimos el sistema tiene N moléculas, donde la molécula n_1 tiene una energía ϵ_1 y la molécula n_2 tiene una energía ϵ_2 y así sucesivamente. Así que las N moléculas se distribuyen con la energía de la siguiente manera:

$$n_1 \epsilon_1 + n_2 \epsilon_2 + \dots + n_k \epsilon_k = \sum n_i \epsilon_i = E \quad (2.12)$$

Ahora para combinar los dos principios que hemos postulado utilizaremos la distribución de *Boltzmann*. La siguiente ecuación demuestra que la probabilidad de que un sistema a temperatura T esté en una configuración de energía E es proporcional al factor de Boltzmann:

$$P_E = \frac{e^{-E/k_B T}}{Z} \quad (2.13)$$

Donde:

P_E es la probabilidad buscada

E es la energía cuya probabilidad estamos buscando

k_B es la constante de Boltzmann ($k_B = 1.381 \cdot 10^{-23} \text{ J/K} = 8.617 \cdot 10^{-5} \text{ eV/K}$)

T es la temperatura

Z es la función de partición canónica

A partir de esta fórmula y la función de partición podemos determinar la energía libre de Helmholtz A , la cual permite ver qué procesos son espontáneos en condiciones de temperatura y volumen constantes. Esta energía no toma en cuenta el proceso que el sistema ha sufrido, sino del estado final e inicial del sistema. La energía libre de Helmholtz se define:

$$A = U - TS \quad (2.14)$$

Donde U es la energía interna, T es la temperatura absoluta del sistema y S es la entropía final.

La energía interna U , podría ser considerada como la energía necesaria para crear el sistema, en ausencia de cambios en la temperatura o el volumen. Pero si el sistema se crea en un ambiente de temperatura T , entonces parte de la energía se puede obtener mediante la transferencia espontánea de calor desde el medio ambiente al sistema. La cantidad de transferencia de energía espontánea es TS , donde S es la entropía final del sistema. En ese caso ya no hay que poner más energía. Note que si se crea un estado final mas desordenado (mayor entropía), se requiere menos trabajo para crearlo. Así pues, la energía libre de Helmholtz es una medida de la cantidad de energía que tenemos que poner para crear un sistema, una vez que ya se ha efectuado la transferencia espontánea de energía del medio ambiente (30).

Una vez que hemos obtenido el valor de la energía de Helmholtz, podemos determinar los otras variables termodinámicas que tenga nuestro sistema, como son la entropía, la energía interna o la capacidad calorífica.

En conclusión el ensamble canónico nos permitirá calcular cantidades termodinámicas en un sistema con temperatura constante, como es el caso de nuestro modelo molecular de Ataxina-1.

2.4.4 Potencial Electrostático Molecular (72)

El Potencial Electrostático Molecular (MEP, de sus siglas en inglés) se ha convertido en los últimos tiempos en una herramienta formidable para el estudio de la reactividad y las interacciones entre moléculas. El potencial electrostático $V(r)$ se genera en el espacio alrededor de una molécula por causa de sus núcleos y sus electrones. El potencial $V(r)$ es una propiedad física real y se puede determinar experimentalmente por métodos de difracción. Se supone que en el acercamiento de una molécula a otra se establece un reconocimiento desde sitios específicos en donde el potencial juega un papel importante. Por ejemplo, cuando un fármaco se acerca a un sitio receptor es de esperar que se establezca una primera atracción a partir de reconocerse ambos desde potenciales de diferente signo. Se entiende también la importancia de esta propiedad en reacciones electrofílicas y nucleofílicas ya que pueden describirse regiones del potencial en las moléculas con tales características.

El potencial electrostático está relacionado con la densidad de carga total $D(r)$ a través de la *ecuación de Poisson*

$$\nabla^2 V(r) = -4\pi D(r) \quad (2.15)$$

y por tal razón ha tenido desde los métodos computacionales un desarrollo sostenido en el área de la Teoría del Funcional de la Densidad.

Cualquier distribución de cargas en el espacio genera un potencial eléctrico $V(r)$. Si tomamos una colección de cargas puntuales Q_i localizadas a posiciones r_i el potencial en un punto dado a una distancia r será

$$V(r) = \sum \frac{Q_i}{|r_i - r|} \quad (2.16)$$

Si la distribución de carga es continua se reemplaza la sumatoria por una integral

$$V(r) = \int \frac{D(r')dr'}{|r'-r|} \quad (2.17)$$

el signo de $D(r)$ puede variar como una función de r .

En el caso de los sistemas moleculares el potencial eléctrico se denomina potencial electrostático y se trata como una distribución estática de carga alrededor de un esqueleto de núcleos y tiene la expresión

$$V(r) = \sum \frac{Z_A}{|R_A-r|} - \int \frac{\rho(r')dr'}{|r'-r|} \quad (2.18)$$

Z_A es la carga sobre el núcleo A localizada a un radio R_A , y $\rho(r')$ es la función densidad electrónica de la molécula. Las ecuaciones están expresadas en unidades atómicas.

Como la densidad electrónica es siempre positiva, a diferencia de $D(r)$ que puede serlo o no, el segundo término de la ecuación anterior tiene antepuesto un signo negativo. El signo del potencial en una zona dependerá del efecto que domine entre los núcleos y los electrones en esa región.

Como el MEP toma diferentes valores en cada punto del espacio, éste se puede representar gráficamente. Una forma es elegir un plano que atraviese la molécula y representar en ese plano las curvas isopotenciales que unen puntos del mismo valor.

Otra forma es representar superficies de alguna propiedad como la densidad electrónica envolviendo una representación tridimensional de la molécula y sobre la misma graficar los valores del MEP asignando colores según el signo del mismo.

Convencionalmente se colorea de rojo los potenciales más negativos y azul los más positivos. Un valor de potencial de especial interés es el mínimo alrededor de la molécula ya que puede indicar, como señalamos anteriormente, la dirección de un posible ataque electrofílico.

Incluso el valor del potencial mínimo puede ser usado para el estudio de relaciones de actividad estructural.

2.4.4.1 Ecuación de Poisson-Boltzmann (72)

La ecuación de Poisson resuelve el potencial electrostático ($\phi(r)$) generado por una distribución de cargas moleculares ($\rho(r)$) dentro de un medio con una determinada constante dieléctrica ($\epsilon(r)$). Si además se considera la presencia de iones, la distribución de los mismos se incluye en la ecuación de Poisson mediante una distribución de Boltzmann, resultando en la ecuación de Poisson-Boltzmann (PB).

$$\nabla - \epsilon(r)\nabla\phi(r) - k'\phi(r) = -4r\rho(r) \quad (2.19)$$

La constante dieléctrica es dependiente de la distancia ($\epsilon(r)$): equivale a la del agua en zonas alejadas del soluto y desciende rápidamente con la distancia en las zonas límite soluto-solvente. El parámetro (k'), inverso de la longitud de Debye-Hückel, adopta valores de 0.1\AA^{-1} en condiciones fisiológicas. Una vez se calcula el potencial ($\phi(r)$), la contribución electrostática a la energía de solvatación viene dada por:

$$\Delta G_{\text{elec}} = \frac{1}{2} \sum_i q_i (\phi(r_i)_{\text{solv}} - \phi(r_i)_{\text{vacío}}) \quad (2.20)$$

Donde q_i es la carga parcial del átomo en la posición r_i que genera la densidad molecular y $(\phi(r)_{\text{vacío}})$ es el potencial electrostático calculado para la misma distribución de cargas pero en ausencia de límites dieléctricos (en vacío, en el que se utiliza una dieléctrica de 1 tanto en la cavidad del soluto como fuera de ella).

La ecuación de Poisson-Boltzmann no es de fácil resolución para los sistemas de interés, por lo que se tienen que aplicar métodos numéricos. Entre ellos, el método de diferencias finitas en una malla (*finite-difference method*, FDPB) es el más usual. Este método se ha implementado en diversos programas como DELPHI203, MEAD204 y UHBD205 y en la versión de AMBER, que es la utilizada en el presente trabajo.

En el método FDPB, se superpone una malla de puntos sobre el soluto y el solvente, asignándose a cada punto de la *grid* los valores de potencial electrostático, densidad de carga, constante dieléctrica y fuerza iónica (Figura 2.11). Cada una de las cargas parciales se distribuye sobre los ocho puntos de la malla que la rodean mediante una ecuación trilineal.

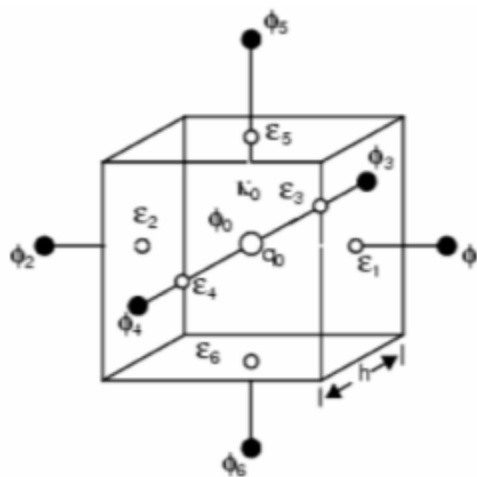


Figura 2.11: Esquema de la malla utilizada en el programa AMBER para resolver la ecuación de Poisson-Boltzmann mediante el método de diferencias finitas

El potencial en cada punto de la malla (ϕ_0) se obtiene según la ecuación (2.21)

$$\phi_0 = \frac{\sum_{i=1}^{\sigma} \epsilon_i \phi_i + 4\pi \frac{q_0}{h}}{\sum_{i=1}^{\sigma} \epsilon_i + Nk_0^2 h^2} \quad (2.21)$$

Donde el sumatorio i se realiza sobre los seis puntos de la *grid* que rodean al punto con carga q_0 , de manera que el potencial en cada punto afecta y es afectado por sus vecinos. Esto se traduce en una resolución del sistema de manera iterativa, hasta que alcanza convergencia. El valor h corresponde a la arista del cubo, k' se calcula a partir de la fuerza iónica y N adopta el valor de 0 cuando la fuerza iónica es nula, 1 para la ecuación lineal o es equivalente a la expansión en serie para la ecuación no lineal (35).

En aplicaciones en dinámica molecular, la ecuación de Poisson-Boltzmann tiene que ser resuelta cada vez que la conformación de la proteína cambia. Así, la ecuación de Poisson-Boltzmann se ha aplicado tradicionalmente en el cálculo de propiedades electrostáticas de configuraciones “estáticas”: cálculo del potencial electrostático, potenciales redox de solvatación, desplazamientos conformacionales inducidos por el

solvente, flexibilidad de proteínas...) y en la determinación del pKa de grupos en proteínas.

En este trabajo, se aplica dicha ecuación para resolver el término de energía de interacción electrostática correspondiente al Potencial electrostático.



Capítulo 3

Metodología y Detalles Computacionales

3.1. Equipos y Software

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

- Computador HP G60-235DX. INTEL PENTIUM DUAL CORE T4200 2.0GHZ, 1MB L2 CACHE, 3072MB DDR2 SDRAM, 320GB
- Software:
 - Gromacs 4.5 (Hess,2008) (73). Disponible en www.gromacs.org

Gromacs es un paquete de programas, que permite realizar Dinámica Molecular, es decir, simular las ecuaciones de Newton del movimiento para sistemas con cientos de millones de partículas, como lo es el caso de las moléculas.

Está diseñado principalmente para moléculas bioquímicas como proteínas, lípidos y ácidos nucleicos, y debido a que ejecuta cálculos rápidamente, sobre todo las complicadas interacciones que poseen los sistemas biológicos, aunque últimamente también es utilizado, en investigaciones en materia no biológica, por ejemplo, polímeros.

Gromacs funciona mediante algoritmos matemáticos que permiten realizar una aplicación moderna de la Dinámica Molecular, utiliza topologías y

archivos de parámetros escritos en formato de texto plano. Los comandos son colocados en una interfaz sencilla con opciones para archivos de entrada y de salida.

- Molden 5.0 (Schaftenaar, 2000) (74). Disponible en ww.cmbi.ru.nl/molden

Molden es un programa de procesamiento de estructura molecular y electrónica general.

Molden permite mostrar la densidad molecular, orbitales moleculares y la densidad de electrones. Puede también animar caminos de reacción y vibraciones moleculares. Así como calcular y mostrar las cargas electrostáticas y potenciales atómicos.

Molden también dispone de un campo de fuerza independiente, *ambfor*, programa que puede optimizar geometrías también dispone de un potente editor de matriz Z que proporcionan un control total sobre la geometría y le permite construir moléculas desde cero, incluidos los polipéptidos.

- UCSF Chimera 1.5.3 (Pettersen, 2004) (75). Disponible en www.cgl.ucsf.edu/chimera

UCSF Chimera es un programa para la visualización y análisis de estructuras moleculares y datos relacionados, incluyendo mapas de densidad, conjuntos supramoleculares, secuencia de alineaciones, trayectorias, y conjuntos conformacionales. Chimera puede generar imágenes de alta calidad y animaciones. Chimera es desarrollado por el recurso de Biocomputación, Visualización e Informática, financiado por los Institutos Nacionales de Salud.

- Unipro UGENE 1.10 (Russian Academy of Sciences,1992) (76), Disponible en *ugene.unipro.ru/*

UGENE es una multiplataforma de código libre utilizado como software de bioinformática. Integra docenas de conocidas herramientas biológicas y algoritmos, proporcionando tanto gráfica de usuario e interfaces de línea de comandos.

Con el fin de proporcionar a software UGENE el máximo rendimiento posible se utiliza CPUs multinúcleo y GPU para optimizar algunos de sus rutinas computacionales.

- APBS (Baker, 2001) (77). Disponible en *www.poissonboltzmann.org/apbs*

APBS es un paquete de software para modelado de solvatación biomolecular a través 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. La electrostática juega un papel importante en varias áreas de la simulación biomolecular, incluyendo:

- simulación de procesos de difusión para determinar la cinética de unión ligando-proteína y proteína-proteína,
- dinámica de solventes implícitos en biomoléculas,
- cálculos de solvatación y de energía vinculante para determinar constantes de equilibrio de unión en interacciones ligando-proteína y proteína-proteína,
- ayuda en el diseño racional de fármacos y estudios biomoleculares.

APBS fue diseñado para evaluar eficientemente las propiedades electrostáticas de tales simulaciones para una amplia gama de escalas de longitud que permite la investigación de moléculas con decenas de millones de átomos. También ofrece modelos implícitos de disolvente no polar de

solvatación que representan con precisión tanto repulsivas y atractivas interacciones soluto-solvente.

- Aplicación online MolProbity (Chen, 2010) (78). Disponible en molprobity.biochem.duke.edu/

MolProbity es una web que ofrece una validación de calidad para las estructuras tridimensionales (3D) de proteínas, ácidos nucleicos y complejos. Esta web proporciona información detallada de los diversos análisis moleculares, y puede calcular y mostrar los enlaces de Hidrogeno y de van der Waals entre los componentes. Un paso esencial en el proceso es la adición y la optimización completa de todos los átomos de hidrógeno, tanto polares como no polares. Los resultados se presentan en múltiples formas: como resultados numéricos en general, como listas, como archivos PDB y archivos gráficos, y sobre todo como informativos, gráficos (imágenes 3D).

- PDB2PQR(University College Dublin, 2002-2011) (79) Disponible en www.poissonboltzmann.org/pdb2pqr

PDB2PQR es un paquete de software Python que automatiza muchas de las tareas en la preparación de las estructuras para que puedan realizar los cálculos de potencial electrostático, proporcionando una utilidad independiente de la plataforma para la conversión de archivos en formato PDB al formato de PQR. Esta acción es necesaria para poder obtener los potenciales electrostáticos con el software APBS:

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

3.2. Elección de las estructuras usadas en las simulaciones

Para esta investigación utilicé la base de datos del NCBI (www.ncbi.nlm.nih.gov/) y

secuencia FASTA de la proteína Ataxina-1 normal.

3.3. Construcción de la Ataxina-1 mutada

La proteína Ataxina-1 obtenida de la base de datos NCBI, es la proteína normal, la que no ha sufrido ninguna mutación. Para construir la proteína Ataxina-1 mutada, usamos como base el formato FASTA de la normal, donde ubicamos la región de poliglutaminas que se encuentra interrumpida por histidinas.

En la secuencia FASTA agregamos 22 aminoácidos glutaminas (Q) en la región de la Cadena Poliglutamínica, al mismo tiempo que se remplazaron las histidinas presentes en esta zona por glutaminas. Es así que obtuvimos una cadena de 50 glutaminas, como se aprecia en la Figura 3.2.

```
MKSNQERSNECLPPKKREIPATSRSSSEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGRRHGPAGTSVELG
LQQGIGLHKALSTGLDYSPPSAPRSVPVATTLPAAYATPQPGTPVSPVQY AHLPHPTFQFIGSSQYSPTYA
SFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEAYSTLLANMGSLSQTPGHKAE QQQQQQQQQQQQQQQ
QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHLSRAPGLITPGSPPPAQQNQYVHISSSPQNTG
RTASPPAIPVHLHPHQTMIPHTLTLGPPSQVVMQYADSGSHFVPREATKKAESSRLQQA IQAKEVLNGEM
EKSRRYGAPSSADLGLGKAGGKSVPHPYESRHVVVHPSPSDYSSRDPSGVRASVMVLPNSNTPAADLEVQ
QATHREASPSTLNDKSLHLGKPGHRSYALSPHTVIQTHSASEPLPVGLPATAFYAGTQPPVIGYLSGQ
QQAITYAGSLPQHLVIPGTQPLLPVGVSTMEASGAAPAVTSSPQFAAVPHTFVTTALPKSENFNPEAL
VTQAAYPAMVQAQIHLPVVQSVASPAAPPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDL
KIDSSSTVERIEDSHSPGVAVIQFAVGEHRAQVSVEVLVEYPPFFVFGQGWSSCCPERTSQLFDLPCSKLSV
GDVCISLTLKLNKNGSVKKGQVPDPASVLLKHSKADGLAGSRHRYAEQENGINQGSQMLSENGELKFPE
KMGLPAAPFLTKIEPSKPAATRKRWSAPESRKLEKSEDEPPLTLPKPSLIPQEVKICIEGRSNVVGK
```

Figura 3.2: **Secuencia FASTA de Ataxina-1 mutada.** Proteína modificada al agregar y sustituir aminoácidos. En amarillo se resalta la cadena Poliglutamínica de 50 aminoácidos. Modificación realizada con el programa MOLDEN 5.0

Luego se procedió a preparar la proteína para su uso en GROMACS como se realizó anteriormente con la proteína normal.

3.4. Optimización de la geometría

Necesitamos hallar la configuración espacial de los núcleos de mínima energía de la molécula. El procedimiento consiste en calcular la función de onda y la energía a partir de la geometría inicial, es decir la distribución espacial de los átomos en la molécula, y luego procede a buscar una geometría de menor energía. Esto se repite

hasta que la geometría de menor energía es encontrada. Es decir se calcula la fuerza sobre cada átomo evaluando el gradiente (primera derivada) de la energía con respecto a las posiciones atómicas. Sofisticados métodos son usados sobre cada paso para seleccionar la nueva geometría. Al final, la fuerza sobre cada átomo de la geometría de mínima energía es cero.

Es importante destacar que este procedimiento no encuentra el mínimo global sino un mínimo local. El procedimiento también se interrumpe cuando encuentra un punto estacionario, es decir el punto donde las fuerzas sobre los átomos son cero, o en punto *saddle* (estructuras de transición). También se puede usar la segunda derivada de la energía con respecto a las coordenadas atómicas, como una mejora adicional a la optimización geométrica.

Se cuenta con las estructuras de 2 moléculas, la Ataxina-1 normal y la Ataxina-1 mutada, las que debemos preparar para la simulación. En primer lugar necesitamos escoger el campo de fuerza a utilizar, para nuestro caso será AMBER, y crear el archivo de topología.

Utilizando el software GROMACS y el comando *pdb2gmx* obtenemos el archivo de topología y el archivo *.gro* que contiene las coordenadas (posiciones y moléculas) por átomos de la proteína.

Debemos limitar el tamaño de nuestro sistema, usando el archivo *.gro* y el comando *editconf*, colocamos el sistema en una “caja” (unidades en nm), la cual debe tener el tamaño adecuado que permita a la molécula moverse libremente. La “caja” tiene una configuración geométrica cúbica.

Luego de colocar las moléculas dentro de las “cajas”, se procedió a emplear el comando *grompp* y el archivo *.mdp*, que contiene los parámetros que se necesitan para la Optimización de la geometría perteneciente a las estructuras a estudiar.

Para empezar el cálculo de Optimización de estructuras se utiliza el comando *mdrun*, que es la principal programa químico computacional con el que cuenta GROMACS y lleva acabo la simulación de Dinámica molecular y minimización de energía,

mediante una *matriz Hessiana*. Este comando generara los archivos necesarios para poder realizar la simulación de Dinámica Molecular.

Realizamos dos procesos de Optimización o Minimización utilizando el método de *Steepest Descent*. El archivo *minima.mdp* empleado se encuentran en el Apéndice.

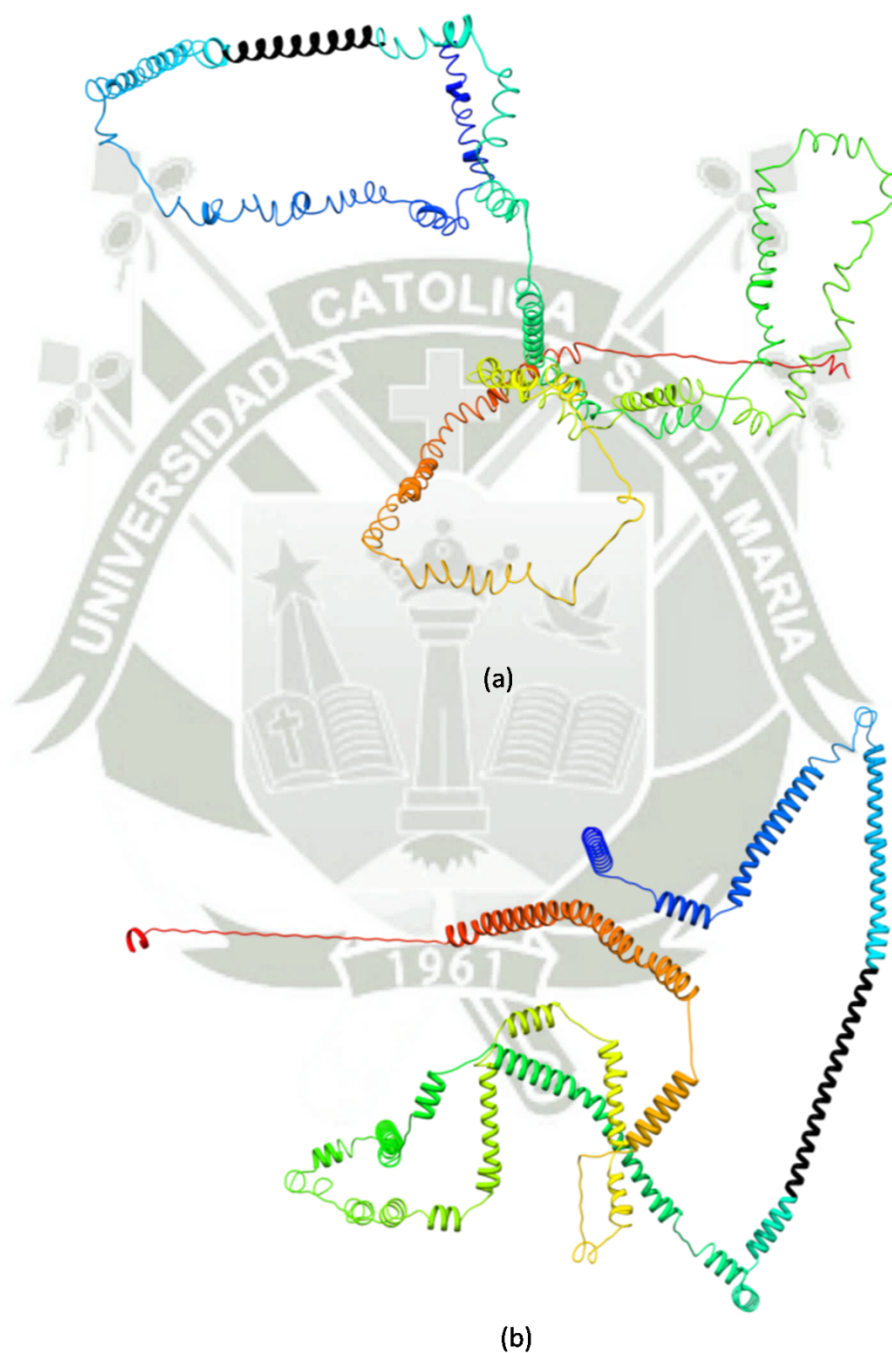


Figura 3.3: **Estructuras tridimensionales.** (a) Proteína Ataxina-1 normal, (b) Proteína Ataxina-1 mutada. En negro se resalta la Cadena Poliglutámica. Imagen obtenida con *Chimera 1.5.3*.

En la Figura 3.3, podemos observar las dos estructuras tridimensionales obtenidas luego de realizar la optimización geométrica con la finalidad de saber si las estructuras mantienen su conformación en el transcurso de la simulación.

Cabe resaltar que se utilizó esta Optimización de Energía después de cada simulación de Dinámica Molecular.

3.5. Dinámica Molecular de estructuras terciarias

Terminada la minimización de las moléculas buscamos que las mismas alcancen una conformación terciaria, para lo cual realizamos una simulación de Dinámica Molecular, empleando el ensamble canónico.

El procedimiento es muy parecido al de la Optimización de la geometría. Mediante el comando *pdb2gmx*, se escoge el campo de fuerza AMBER, nuevamente limitamos el sistema colocándolo en una “caja” mediante el comando *editconf*. Una vez que el tamaño del sistema este definido, el comando *grompp*, nos permite utilizar el archivo *.mdp* perteneciente a la simulación de Dinámica Molecular con un ensamble canónico y solvente implícito. Para empezar el cálculo de la simulación de Dinámica Molecular, se utiliza el comando *mdrun*.

Para el desarrollo de la Dinámica Molecular, se procedió primero a una simulación para lograr la relajación de 30ns (30000 ps), la cual fue desarrollada en 30 millones de pasos, dentro de un ensamble canónico (NVT), manteniendo una temperatura de 309.65 ° K. Terminada esta primera simulación se continuó con la Dinámica Molecular durante un tiempo de 30ns (30000 ps) adicionales, para obtener la estructura en su forma estable, manteniendo las condiciones iniciales, así como el tipo de ensamble (NVT) y la temperatura de 309.65 ° K. Terminada la se obtuvo las estructuras promedio de ambas proteínas, la cual se aprecia en la Figura 3.4.

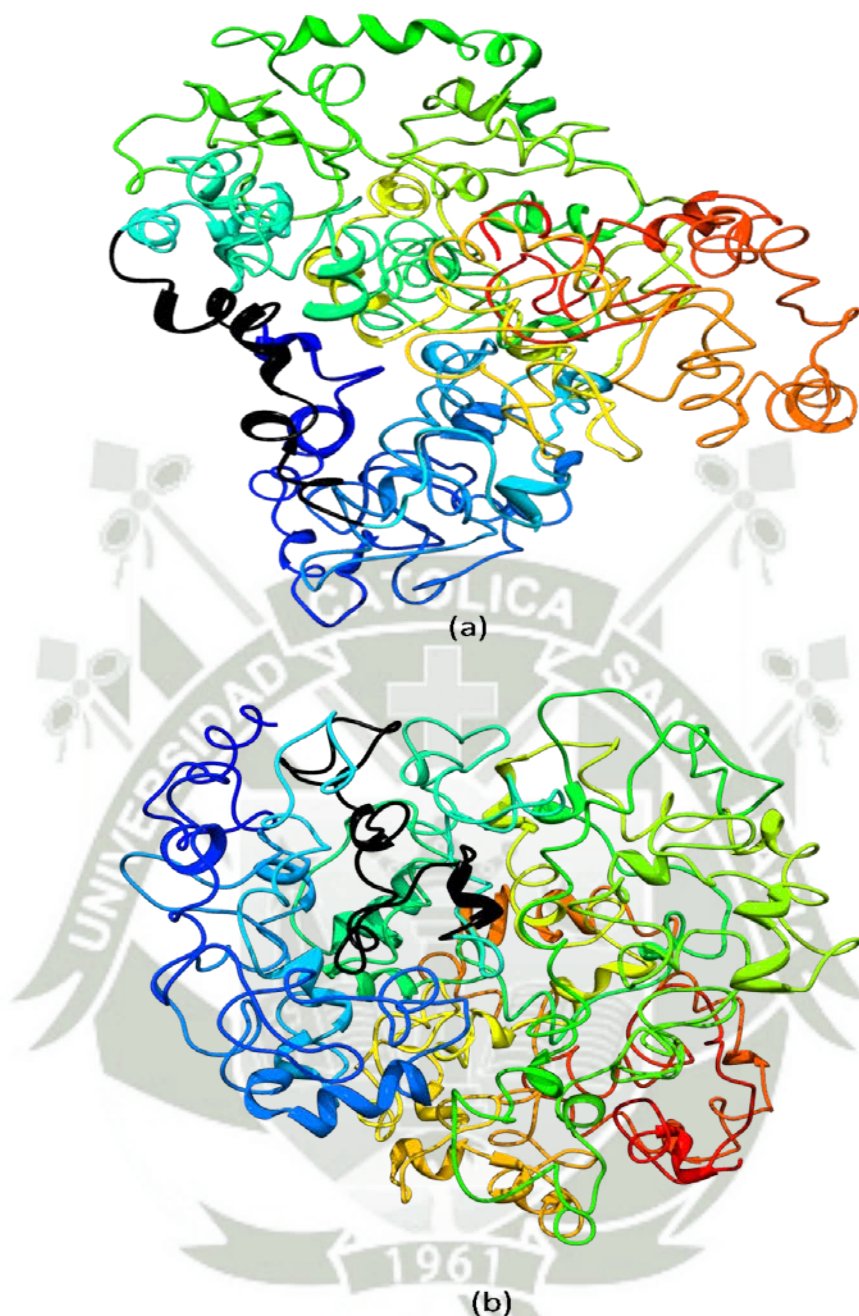


Figura 3.4: **Estructuras tridimensionales dinamizadas.** (a) Proteína Ataxina-1 normal, (b) Proteína Ataxina-1 mutada. En negro se resalta la Cadena Poliglutámica. Imagen obtenida con *Chimera 1.5.3*.

Al final de la simulación de Dinámica Molecular, se obtuvo los archivos que nos permiten determinar la estabilidad estructural de las moléculas. Para comprobar la estabilidad obtenemos el gráfico de *Ramachandran* a partir de la estructura promedio de las moléculas, para tal propósito se utilizó el servidor Molprobit (77) en el cual se analizó la geometría de los átomos para proceder con el cálculo necesario para el Gráfico de *Ramachandran*.

Para llevar a cabo el análisis de la molécula debemos obtener una serie de datos como son la Energía Potencial, Cinética, Total, Temperatura y Presión, para los cuales utilizamos el comando *g_energy* el cual utiliza como input el archivo con extensión *.edr* obtenido al final de la simulación de Dinámica Molecular. Para obtener los gráficos de Radio de Giro, Número de Enlaces de Hidrógeno, Desviación Media Cuadrática de la Distancia entre átomos y para la Desviación Media Cuadrática de la Fluctuación por residuos se utilizan diversos comandos específicos y los archivos *.tpr* y *.trr* obtenidos al final de la simulación.

3.6. Análisis del Potencial electrostático a través de APBs

El potencial electrostático es una poderosa herramienta que ha proporcionado conceptos sobre la asociación intermolecular y propiedades moleculares de pequeñas moléculas, las acciones de las moléculas de drogas y sus análogos, la función biológica de la hemoglobina, y la catálisis enzimática.

Sabemos que las proteínas contienen grupos cargados y polares, y que las interacciones electrostáticas controlan aspectos importantes de su estructura y función. Estas últimas son de largo alcance, sin dirección (escalar) y poseen interacciones coulombianas no específicas, que son importantes para la estabilidad, plegado y función de las proteínas. Estas fuerzas ejercen su influencia a mayores distancias, en contraste con otras fuerzas electrostáticas de corto alcance como los puentes de sal (interacción entre dos residuos cargados en constantes dieléctricas muy bajas).

Es claro que los efectos electrostáticos desempeñan un importante papel en la catálisis de enzimas, la transferencia de electrones, protones de transporte, los canales iónicos, la unión del ligando, ensamblaje macromolecular, la transducción de señales, entre otras cosas.

En un principio los estudios de los potenciales electrostáticos no poseían un modo satisfactorio para representar el potencial electrostático en una forma útil. Este problema se resuelve mediante la visualización efectiva la superficie del potencial electrostático molecular con gráficos de colores interactivos. (37)

En nuestro caso utilizando el software APBS (Adaptive Poisson-Boltzmann Solver), software desarrollado para la evaluación de propiedades electrostáticas en sistemas biomoleculares a nivel nanoescalar, y las estructuras promedio de las moléculas, las cuales se obtuvieron utilizando el comando `g_covar` el cual utiliza los archivos `.tpr` y `.trr`.

Una vez obtenidas las estructuras promedio, mediante el servidor PDB2PQR, el cual es un programa especializado para la configuración, ejecución y análisis de los cálculos electrostáticos utilizando la ecuación de Poisson-Boltzmann; con dicho servidor obtenemos los archivos `.in` y `.pqr`, a este último le modificamos la temperatura a 309.65 °K, que es la que hemos venido empleando.

Aplicando el programa APBS, obtenemos un archivo `.dx`, el cual nos permite visualizar las gráficas de potencial electrostático con el software UCSF Chimera.



Capítulo 4

Resultados y Discusión

4.1. Estructuras

Mediante la utilización del programa Molden se obtuvo la estructura .pdb de la proteínas Ataxina-1 normal y mutada.

Posteriormente se realizó la optimización geométrica de ambas proteínas, lo que dio como resultado la localización de un mínimo de energía, que nos permitió determinar que las estructuras están listas para la Dinámica Molecular.

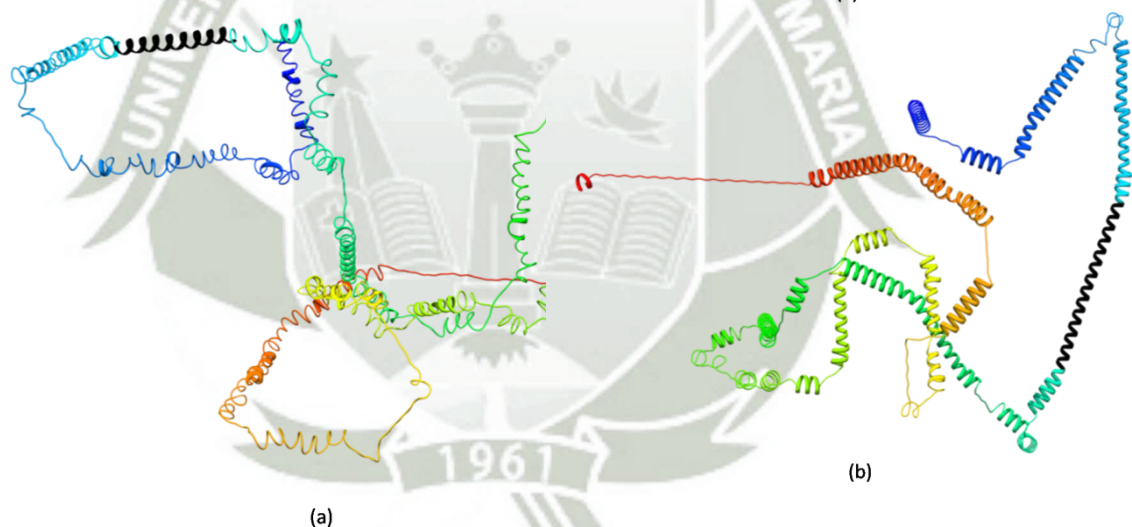


Figura 4.1: **Estructuras tridimensionales minimizadas.** (a) Proteína Ataxina-1 normal, (b) Proteína Ataxina-1 mutada. En negro se resalta la Cadena Poliglutámica. Imagen obtenida con *Chimera 1.5.3*.

En la Figura 4.1 se aprecia claramente la diferencia en el tamaño de la Cadena Poliglutámica, siendo la de la molécula Ataxina-1 mutada mucho más grande que la de la Ataxina-1 normal, se ha coloreado en negro la cadena, para su mejor observación.

4.2. Estabilidad de estructuras terciarias

Una vez obtenidas las estructuras optimizadas se procedió a realizar la Dinámica Molecular, con la finalidad de obtener la estructura en su fase estable. En la Figura 4.2 apreciamos los resultados de la Dinámica Molecular.

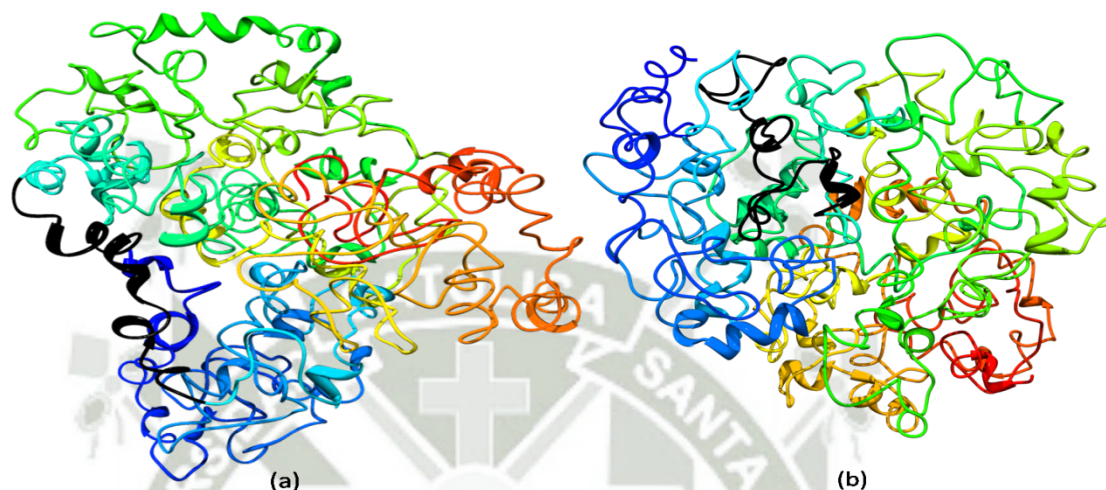


Figura 4.2: Estructuras tridimensionales plegadas. (a) Proteína Ataxina-1 normal, (b) Proteína Ataxina-1 mutada. En negro se resalta la Cadena Poliglutámica. Imagen obtenida con *Chimera 1.5.3*.

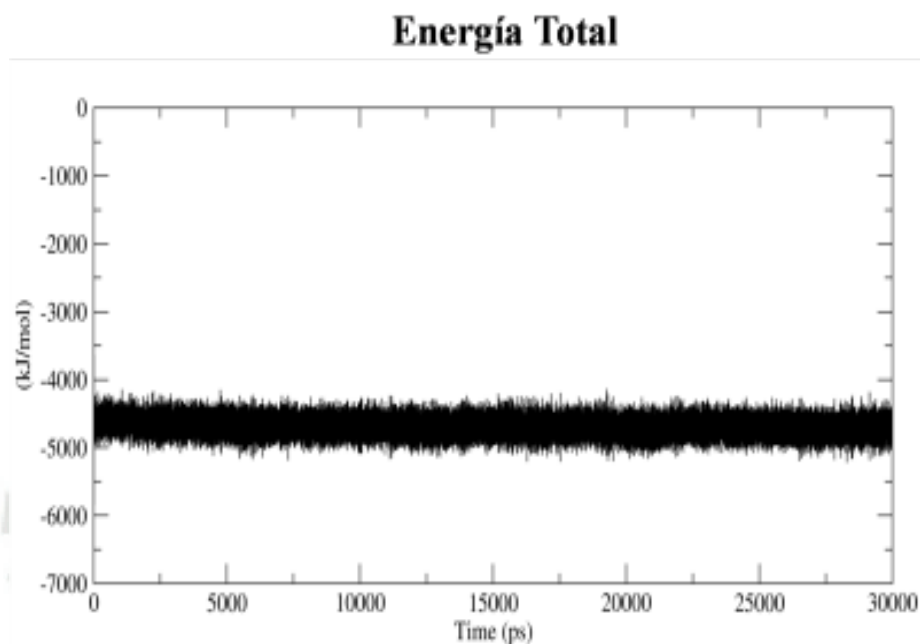
Para verificar la fase de estabilidad, se graficó los datos de **Energía**, tanto como **Energía Potencial**, **Energía Cinética** y la **Energía Total**, el **Número de enlaces de Hidrógeno** que nos indica si las estructuras dependientes de los enlaces de hidrógeno se mantienen constantes durante la simulación; Distancia Media Cuadrática (**RMSD**) y fluctuación media cuadrática (**RMSF**) por residuo.

4.2.1 Análisis de la estabilidad de la proteína Ataxina-1

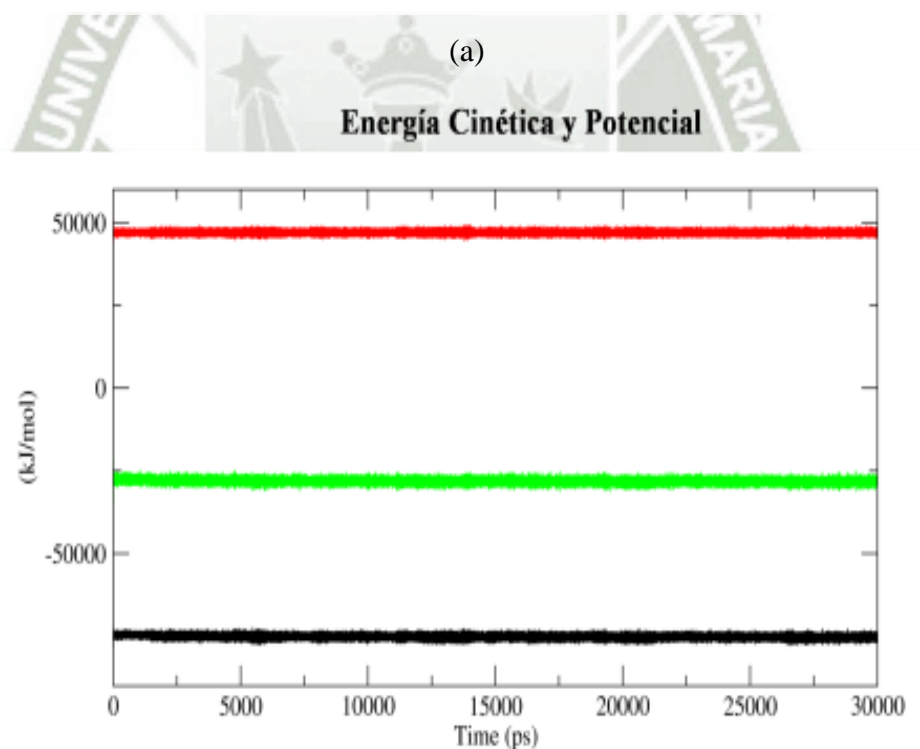
Para poder verificar la estabilidad que ha alcanzado nuestra molécula a través de la Dinámica Molecular, se deben obtener una serie de datos mediante los comandos que nos ofrece el software GROMACS.

Entre estos datos figura la energía cinética, resultante de las energías de traslación, rotación y vibración que poseen las moléculas o átomos que constituyen el sistema, otro dato es la energía potencial intermolecular e intramolecular, proveniente de los

enlaces de la proteína y por último la energía total producto de la diferencia de las dos energías antes mencionadas.



(a)



(b)

Figura 4.3: Estabilidad Energética de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns. (a) Energía Total, (b) Energía Cinética(rojo) y Potencial (Negro). Las gráficas se obtuvieron con el comando *g_energy* de GROMACS. Las gráficas se realizaron con el programa xmgrace

El primero de los datos obtenidos es la energía total que posee nuestro sistema, en esta caso la molécula Ataxina-1 normal. Para ello se utilizó el programa *g_energy*, con el cual generamos el gráfico de Energía total, donde se observa una completa estabilidad a lo largo de los 30ns (30000 ps), siendo su energía Total de -4688.58 KJ/mol la cual se muestra en la Figura 4.3-a; se procedió a desdoblar la energía total en su respectiva energía cinética, la cual es de 7835.1 KJ/mol, y la energía potencial, que es de -12523.7 KJ/mol; con las cuales confirmamos la estabilidad energética, como se aprecia en la Figura 4.3-b.

Con el dato de la estabilidad energética, se procedió al análisis de los puentes de hidrógeno de la molécula, para lo cual se usó el programa *g_hbond*, generando así una gráfica en la cual se observa la transitoriedad y permanencia de los puentes de hidrógeno en la estructura durante los 30ns (30000 ps) de la Dinámica Molecular, se observa que los puentes de hidrógeno se encuentran en una fluctuación constante, lo cual, indica la estabilidad de la estructura terciaria, ya que los puentes de hidrogeno juegan un rol importante en la determinación de las estructuras tridimensionales adoptadas por las proteínas. En la Figura 4.6, se aprecia que a lo largo de los 30ns la fluctuación es de 100 enlaces.

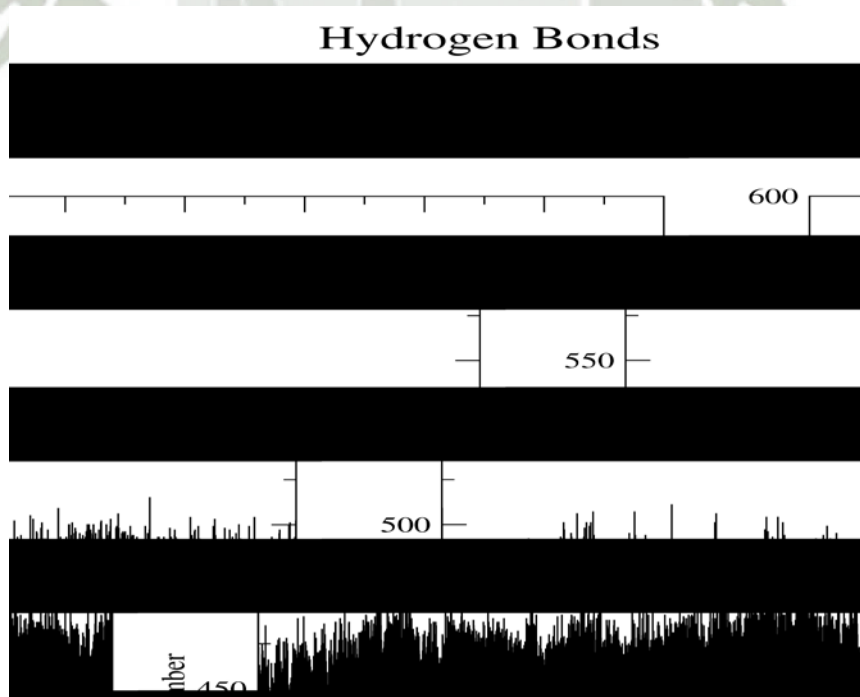


Figura 4.4: Puentes de Hidrógeno de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns. La gráfica se obtuvo con el comando *g_hbond* de GROMACS. Las gráficas se realizaron con el programa *xmgrace*.

Los enlace puente de hidrógeno que encontramos en la molécula Ataxina-1, son producto de la interacción entre dos átomos electronegativos, los donantes y aceptores, a través de un átomo de hidrógeno intermedio que forma un puente entre ellos (36). Con programa *g_hbond* obtuvimos los siguientes resultados, donde encontramos que la proteína Ataxina-1 posee 1137 donadores de hidrógenos y 4580 aceptores de hidrogeno.

En la Figura 4.5 observamos que los enlace puente de hidrogeno, están colocados de una manera predeterminada que permiten proporcionar una forma adecuada a la molécula.



Figura 4.5: Puentes de Hidrógeno de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns. Señalado en negro, observamos la cadena poliglutáminica y en celeste se muestran los enlaces puente de hidrogeno. La gráfica se realizó con el programa Chimera 1.5.3.

Luego utilizando los comandos *g_rms* y *g_rmsf*, para el cálculo de la desviación media cuadrática con respecto a la distancia, los datos generados se utilizaron para la generación de la gráfica de la RMSD, con la cual obtenemos la distancia promedio entre una estructura de referencia y una conformacional, es decir la estructura que está sufriendo las alteraciones de rotación y transformación. En la grafica 4.6 observamos que la estabilidad de la molécula se encuentra desde los 20ns, el cual es un dato crucial ya que es el tiempo tomado para la generación de la estructura promedio.

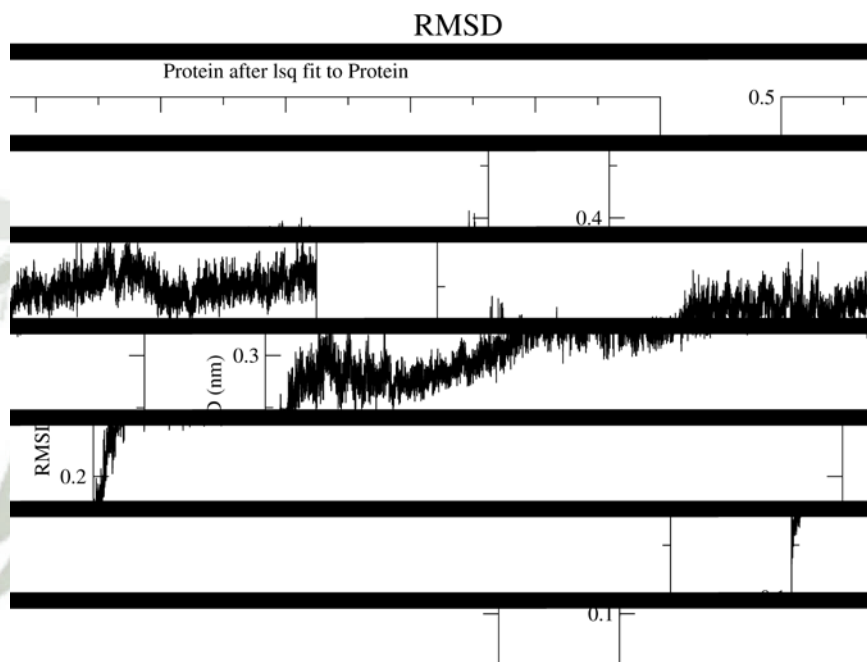


Figura 4.6: RMSD de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns. La gráfica se obtuvo con el comando *g_rms* de GROMACS. Las gráficas se realizaron con el programa *xmgrace*

Se procedió a calcular la fluctuación de la desviación media cuadrática por residuos (RMSF), para obtener información de las estructura que han cambiado su posición, es así que se determinó la zona de la estructura de mayor fluctuación, observándose, que una gran fluctuación entre los aminoácidos 670 al 740 y del 160 al 165, se puede ver que el movimiento de los aminoácidos es de 0,6nm en su pico máximo, esto indica la vibración de la molécula, como se muestra en la Figura 4.7.

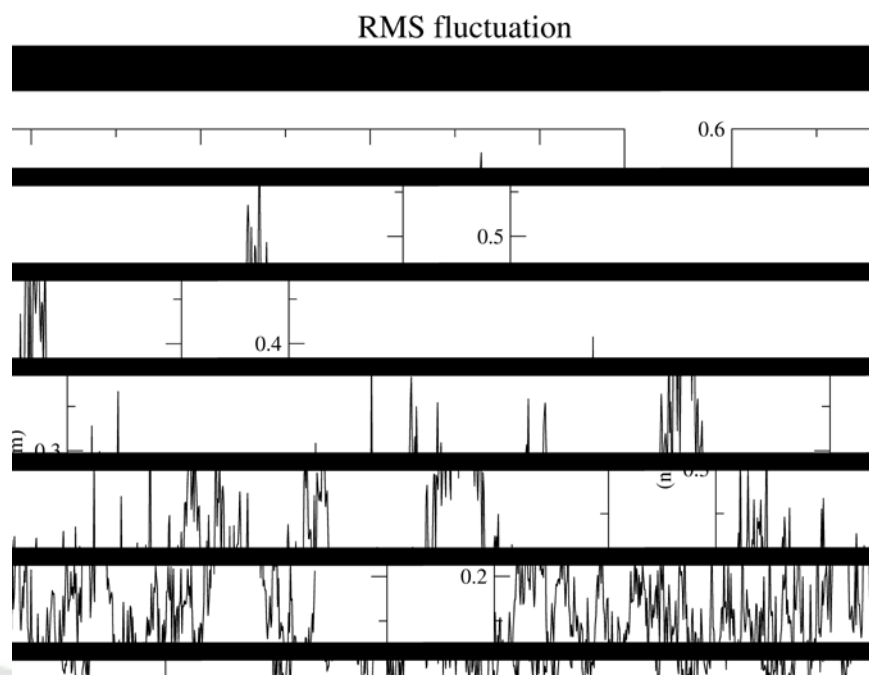


Figura 4.7: RMSF de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns. La gráfica se obtuvo con el comando *g_rmsf* de GROMACS. Las gráficas se realizaron con el programa *xmgrace*

Por último para verificar la congruencia con potenciales resultados experimentales, se realizó el Diagrama de Ramachandran para la Ataxina-1. Debemos resaltar 3 sitios a tomar en cuenta dentro de la molécula, el sitio de las Glicinas, ya que son potenciales regiones para generar metilaciones por sustitución nucleofílica, el sitio son las prolinas, ya que es en estas regiones donde se pueden dar rupturas de anillos o formación de los mismos, por último, se tienen los sitios correspondientes a las regiones de las pre-prolinas, que al igual que en caso anterior, esto pueden formar grupos prolínicos sustituidos, por efecto de la ciclación de los residuos involucrados.

Para nuestra molécula Ataxina-1, obtuvimos un 86.2% de residuos que se hallan en las regiones favorecidas, esto corresponde a 701 de un total de 813, y considerando las regiones permitidas, obtuvimos un 96.8% de nuestros residuos, lo que corresponde a 787 aminoácidos de un total de 813, esto puede ser apreciado en la Figura 4.8.

MolProbity Ramachandran analysis

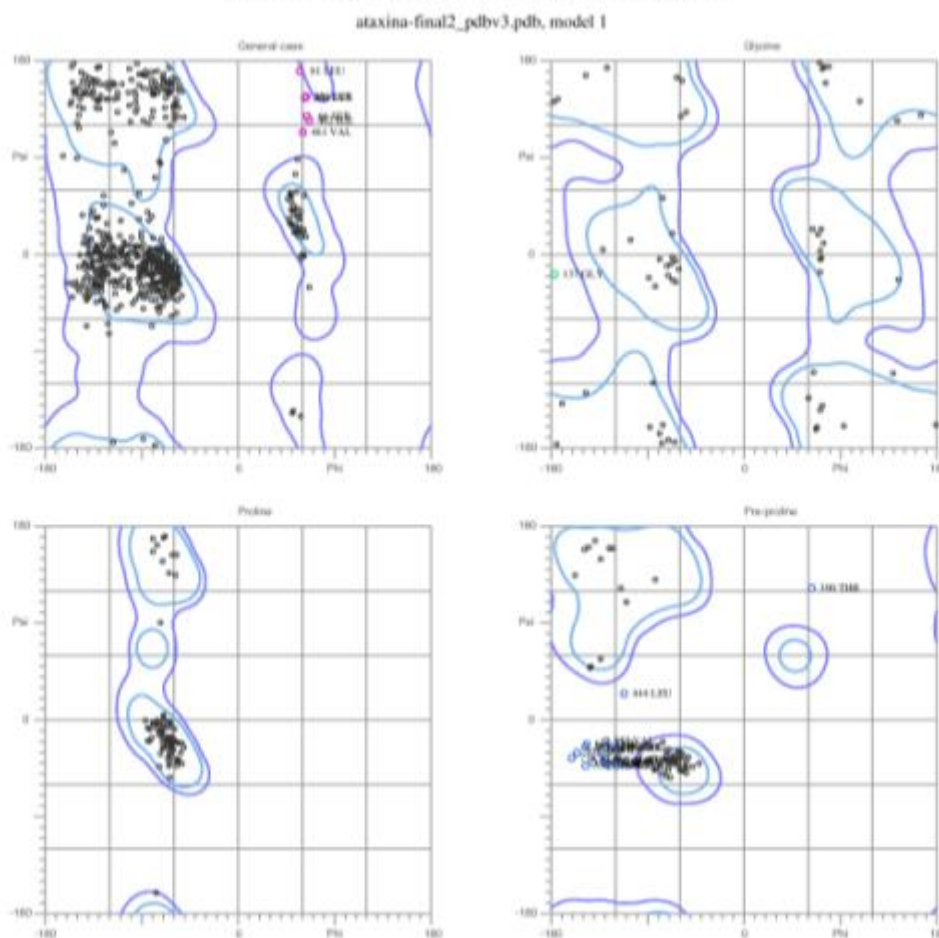


Figura 4.8: **Gráfica de Ramachandran Ataxina-1.** Primer cuadrante corresponde a los casos generales, en donde se ve en la parte superior izquierda las conformaciones β , en la parte inferior izquierda del primer cuadrante las hélices α con orientación a la derecha y en el primer cuadrante parte superior derecha las hélices α con orientación hacia la izquierda. Gráfica obtenida gracias al servidor Online MolProbity (Chen, 2010).

Los residuos que se encuentran fuera de lugar son los siguientes: Serina 66, 116, 262 y 381; Leucina 81, 444 y 570; Treonina 108, 190 y 390; Glicina 137, Isoleucina 143, 148, 462 y 484; Lisina 308; Valina 352 y 461; Histidina 354; Glutamina 488, Tirosina 544 y 648; Alanina 566 y Ácido Aspártico 702.

Cabe resaltar que, la Glicina 137 no se halla en la región favorecida ni permitida, cuando hacemos el análisis de las glicinas, lo cual podría indicar una posible metilación.

Para el caso del análisis de las pre-prolinas, podemos observar que la hay bastantes fuera de las zonas permitidas, siendo las más resaltantes la Treonina 390 y la Leucina 444.

Cuando realizamos el análisis de las prolinas, podemos observar que ninguna se encuentra fuera de las regiones favorecidas o permitidas, lo que indica claramente que nuestra estructura, es estable a mutaciones por efecto de rompimiento de anillos prolínicos, pero inestable para sustituciones de tipo nucleofílica y de generación de anillos prolínicos sustituidos.

Los análisis realizados han permitido determinar que la Proteína Ataxina-1 ha alcanzado la estabilidad energética, siendo este el objetivo que se buscaba alcanzar al momento de someter la molécula al proceso de dinámica molecular

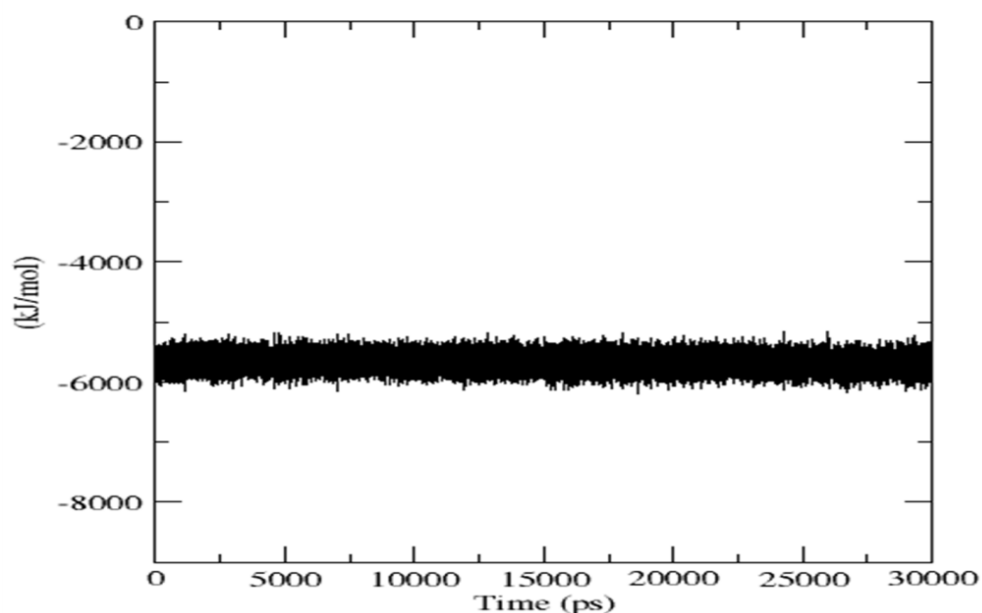
4.2.2 Análisis de la estabilidad de la Ataxina -1 mutada

Al igual que se realizó en la molécula Ataxina-1 normal, en la molécula mutada también obtuvimos datos que permitieran determinar la estabilidad que alcanzó la molécula luego de la Dinámica Molecular. Para ello utilizando nuevamente el programa Gromacs, procedimos a calcular primero la Energía que posee el sistema.

Para verificar la fase de estabilidad energética de la Proteína Ataxina-1 se utilizó el programa *g_energy*, con el cual generamos el gráfico de Energía total, donde se observa una completa estabilidad a lo largo de los 30ns (30000 ps), siendo su energía Total de -5688.69 Kj/mol la cual se muestra en la Figura 4.11. Esta energía comparada con la molécula Ataxina-1 normal, es superior en 988.11 Kj/mol, lo cual indica una mayor estabilidad en la molécula mutada con respecto a la normal.

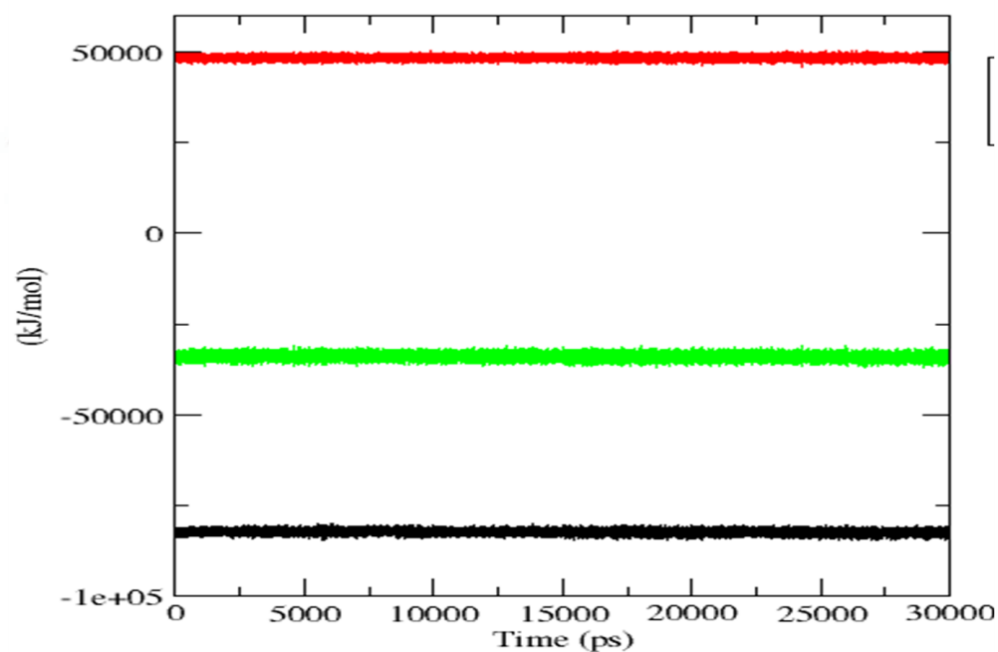
Al proceder en el desdoblamiento de la energía total en sus respectivas energías, observamos que estas también se han incrementado, siendo el valor de 8075.85 Kj/mol para la energía cinética y de -13744.5 Kj/mol para la energía potencial, podemos ver estos resultados en la Figura 4.9-b.

Energía Total



(a)

Energía Cinética y Potencial



(b)

Figura 4.9: **Estabilidad Energética de la Ataxina-1 mutada a 309.65 °K en un ensamble canónico (NVT), por 30 ns.** (a) Energía Total, (b) Energía Cinética (rojo) y Potencial (Negro). Las gráficas se obtuvieron con el comando `g_energy` de GROMACS. Las gráficas se realizaron con el programa `xmgrace`

Luego se procedió al análisis de los enlaces puente de hidrógeno, para lo cual se usó el programa *g_hbond*, generando así una gráfica en la cual se observa la transitoriedad y permanecía de los puentes de hidrógeno en la estructura durante los 30ns (3000ps) de la Dinámica Molecular. En la Figura 4.10, se observa que los puentes de hidrógeno se encuentran en una fluctuación constante, variando en 100 el número total de enlaces a lo largo de la dinámica. El número de variaciones es igual al de la Ataxina-1 normal, pero observamos que el número de puentes de hidrogeno se ha incrementado en 25 enlaces, ya que en la molécula normal el número de enlaces iban de 400 a 500, mientras que en la Ataxina-1 mutada van de 425 a 525; sabiendo que los enlaces puentes de hidrogeno proporcionan la mayor parte de las interacciones de dirección en que están basados el plegamiento de proteínas, la estructura de la proteína y el reconocimiento molecular, podemos afirmar que el incremento en el número de enlaces indicaría una mayor estabilidad molecular.

Hydrogen Bonds

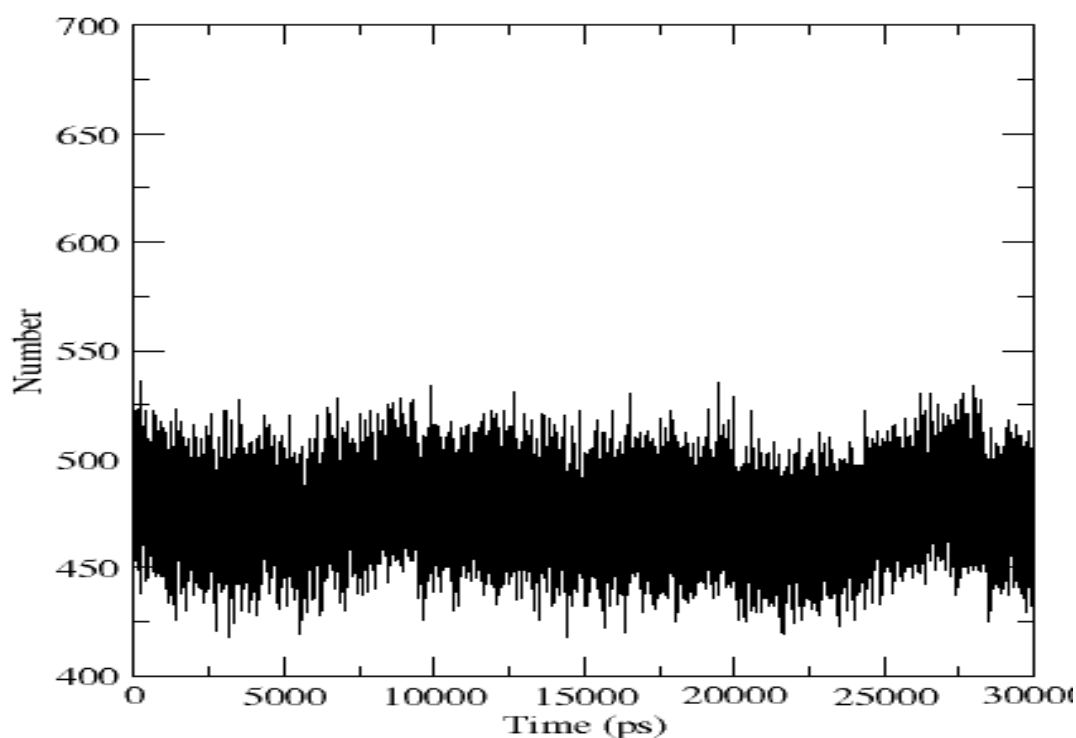


Figura 4.10: **Puentes de Hidrógeno de la Ataxina-1mutada a 309.65 °K en un ensamble canónico (NVT), por 30 ns.** La gráfica se obtuvo con el comando *g_hbond* de GROMACS. Las gráficas se realizaron con el programa xmgrace

Se conoce que la estabilidad estructural de las moléculas son el resultado de diversos enlaces intermoleculares, uno de ellos es el enlace puente de hidrogeno que se

produce gracias a la unión entre un átomo electronegativo y un átomo de hidrógeno unido covalentemente a otro átomo electronegativo. En la Figura 4.11 podemos reconocer los enlaces puente de hidrógeno que posee la molécula. En la proteína Ataxina-1 mutada encontramos 1181 átomos donadores de hidrógeno y 4756 átomos aceptadores de hidrógeno.

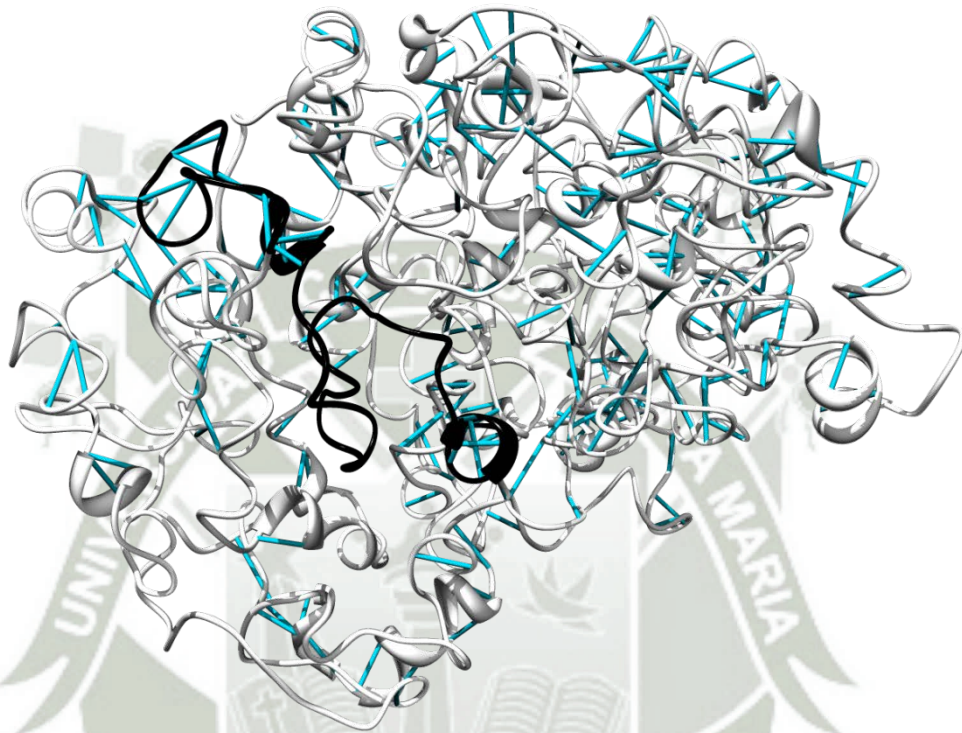


Figura 4.11: **Puentes de Hidrógeno de la Ataxina-1mutada a 309.65 °K en un ensamble canónico (NVT), por 30 ns.** Señalado en negro, observamos la cadena poliglutamínica y en celeste se muestran los enlaces puente de hidrógeno. La gráfica se realizó con el programa Chimera 1.5.3

Por último utilizando los comandos *g_rms* y *g_rmsf*, para el cálculo de la desviación media cuadrática con respecto a la distancia, los datos generados se utilizaron para la generación de la gráfica de la RMSD, con la cual obtenemos la distancia promedio entre una estructura de referencia y una conformacional, es decir la estructura que está sufriendo las alteraciones de rotación y transformación. En la gráfica 4.12 observamos que la estabilidad de la molécula se encuentra desde los 15ns, conociendo este dato podemos generar nuestra estructura promedio.

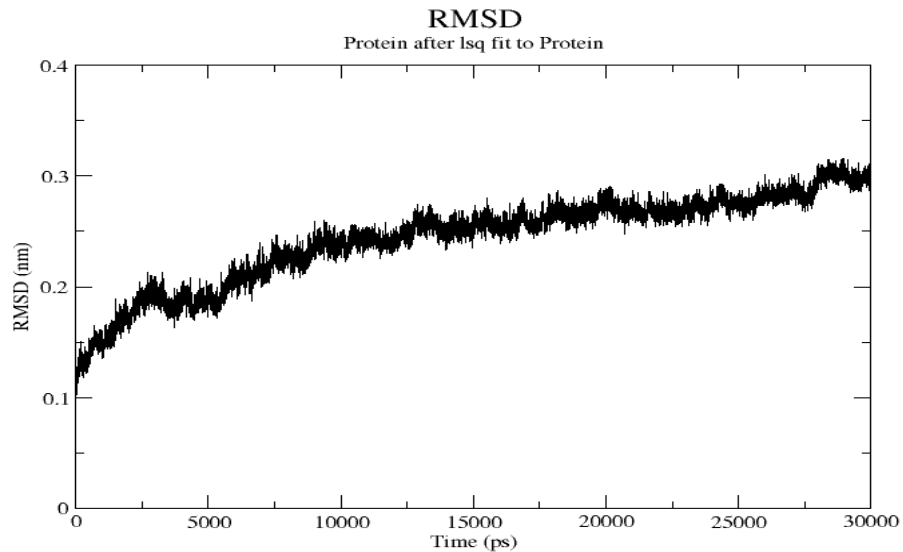


Figura 4.12: **RMSD de la Ataxina-1 a 309.65 °K en un ensamble canónico (NVT), por 30 ns.** La gráfica se obtuvo con el comando *g_rms* de GROMACS. Las gráficas se realizaron con el programa *xmgrace*

Se procedió a calcular la fluctuación de la desviación media cuadrática por residuos (RMSF), para obtener información de las estructura que han cambiado su posición, es así que se determinó la zona de la estructura de mayor fluctuación, observándose, que una gran fluctuación entre los aminoácidos 440 al 460, del 535 al 580 y del 720 al 780, se puede ver que el movimiento de los aminoácidos es de 0,6nm en su pico máximo, esto indica la vibración de la molécula, como se muestra en la Figura 4.13.

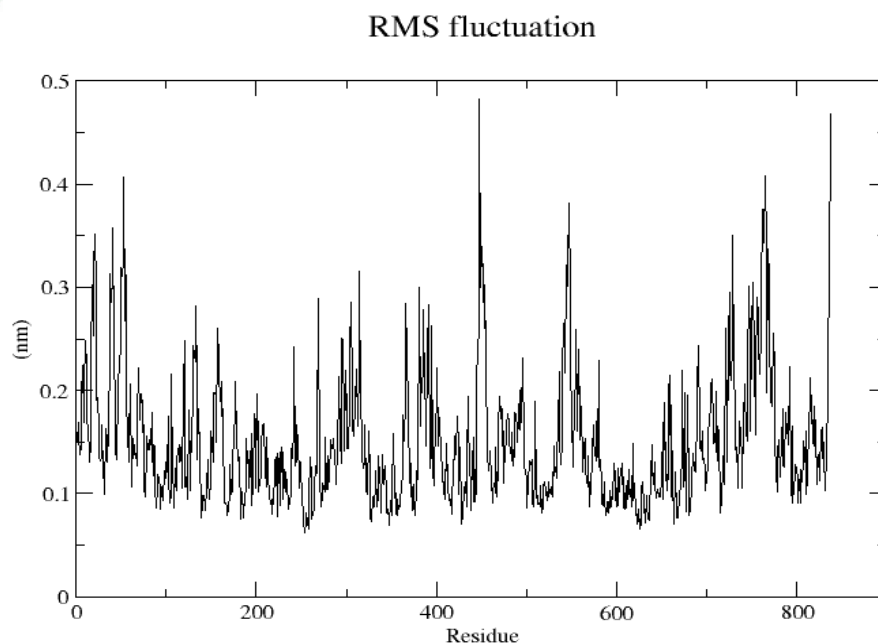


Figura 4.13: **RMSF de la Ataxina-1 mutada a 309.65 °K en un ensamble canónico (NVT), por 30 ns.** La gráfica se obtuvo con el comando *g_rmsf* de GROMACS. Las gráficas se realizaron con el programa *xmgrace*

Para verificar la congruencia con potenciales resultados experimentales, se realizó el Diagrama de Ramachandran para la Ataxina-1 mutada. Este diagrama nos muestra zonas donde los átomos de la molécula están más cercanos que sus radios de Van der Waals. Estas regiones son estéricamente no permitidas para todos los aminoácidos, excepto para Glicina, el cual es el único aminoácido que no presenta cadena lateral. Otra zona corresponde a la región donde no hay impedimentos estéricos, es decir zonas permitidas llamadas conformaciones α -hélices y Hojas- β .

Las regiones no permitidas generalmente involucran impedimentos estéricos entre el carbono metilénico de la cadena lateral y los átomos de la cadena principal. La Glicina no tiene cadena lateral y por lo tanto puede adoptar ángulos phi y psi en los cuatro cuadrantes del gráfico de Ramachandran. Por esto ellas aparecen en regiones de vueltas de la proteína donde cualquier otro residuo estaría estéricamente impedido.

Para nuestra molécula Ataxina-1 mutada, obtuvimos un 85.0% de residuos que se hallan en las regiones favorecidas, esto corresponde a 710 de un total de 835, y considerando las regiones permitidas, obtuvimos un 97.2% de nuestros residuos, lo que corresponde a 812 aminoácidos de un total de 835, esto puede ser apreciado en la Figura 4.14.

Los residuos que se encuentran fuera de lugar son los siguientes: Glicina 54,70, 164 y 764; Glutamina 119 y 207; Serina 133, 260 y 275; Isoleucina 148 y 506; Alanina 252 y 798; Histidina 293 y 376; Valina 323 y 717; Arginina 40; Treonina 412; Leucina 466 y 549; Tirosina 670 y Lisina 826.

Debemos resaltar que, tenemos cuatro Glicinas no se hallan en las regiones favorecidas, lo cual tiende a producir una desestabilización en la α hélice.

Para el caso del análisis de las pre-prolinas, podemos observar que la hay menos prolinas en zonas no permitidas en comparación con la molécula normal, destacando la Serina 275 y la Histidina 376.

Cuando realizamos el análisis de las prolinas, podemos observar que ninguna sale de las regiones favorecidas o permitidas, lo que indica claramente que nuestra estructura, es estable a mutaciones por efecto de rompimiento de anillos prolínicos, pero inestable para sustituciones de tipo nucleofílica y de generación de anillos prolínicos sustituidos.

MolProbity Ramachandran analysis

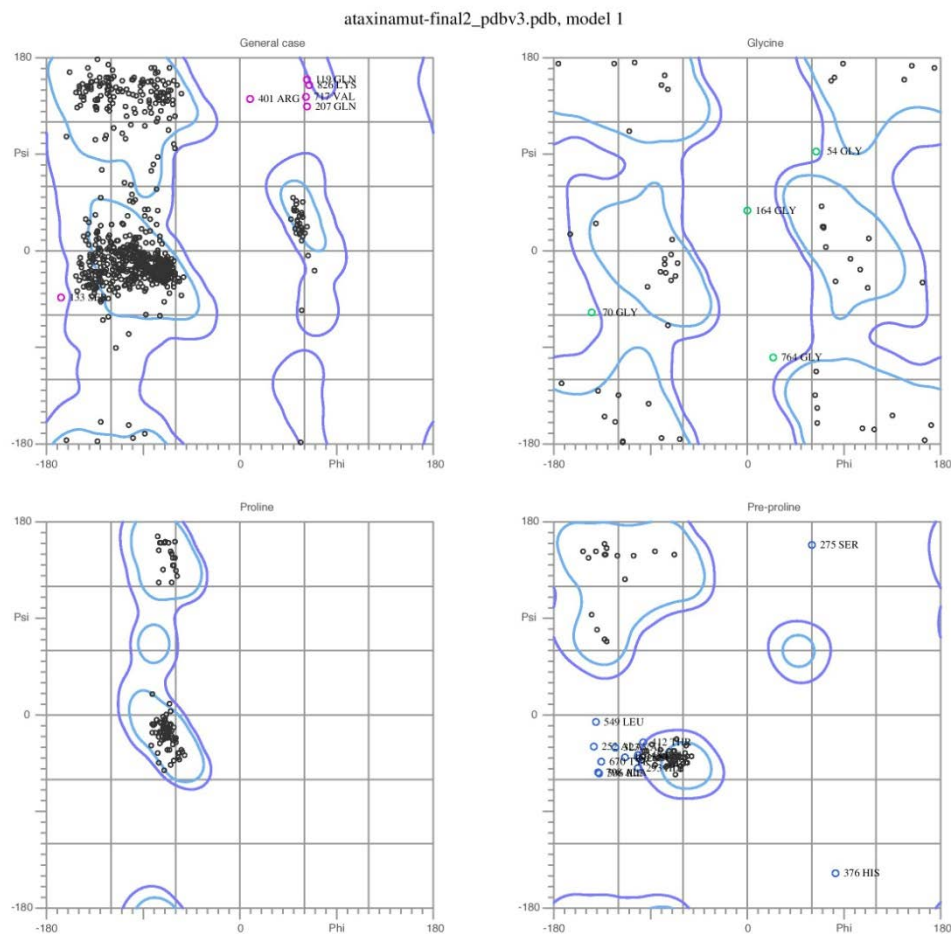


Figura 4.14: **Gráfica de Ramachandran Ataxina-1 mutada.** Primer cuadrante corresponde a los casos generales, en donde se ve en la parte superior izquierda las conformaciones β , en la parte inferior izquierda del primer cuadrante las hélices α con orientación a la derecha y en el primer cuadrante parte superior derecha las hélices α con orientación hacia la izquierda. Gráfica obtenida gracias al servidor Online MolProbity (Chen, 2010).

Los análisis realizados han confirmado la estabilidad energética que ha alcanzado la Proteína Ataxina-1 mutada, siendo este el objetivo que se buscaba obtener al momento de someter la molécula al proceso de dinámica molecular

4.3. Potencial electrostático

Calculamos el Potencial Electrostático utilizando el software APBS, que nos permite obtener las gráficas de las hipersuperficies del **Potencial Electrostático** para las moléculas Ataxina-1 y Ataxina-1 mutada. Esto es de suma importancia cuando se busca zonas de interacción electrostática en una proteína.

El programa APBS se basa en la aplicación de métodos numéricos y la ecuación de Poisson-Boltzmann, esta nos permite calcular el potencial electrostático de las estructuras promedio resultantes de la Dinámica Molecular.

Por medio de las Gráficas de Potencial Electrostático obtenidas, buscamos no solo determinar las zonas de interacción electrostática sino también distinguir la presencia de cavidades o sitios de unión selectiva. Ya que se sabe que todas las proteínas realizan sus funciones por unión selectiva a diferentes moléculas, dependiendo de la función que cumplan; es así que las proteínas estructurales se agregan a otras moléculas de la misma proteína para originar una estructura mayor; mientras que otras proteínas se unen a moléculas distintas, como los anticuerpos a los antígenos específicos, la hemoglobina al oxígeno, las enzimas a sus sustratos, entre otros ejemplos (39).

Esta clase de unión es una de las propiedades que poseen las proteínas, ya que los grupos funcionales de las cadenas laterales R de los aminoácidos definen una superficie activa o centro activo, capaz de interactuar con otras moléculas; el resto de la cadena peptídica solo es necesaria para mantener la forma precisa de la proteína, con el fin de que el centro activo se encuentre en la posición correcta. Es así que la actividad biológica de una proteína se basa en su unión selectiva con otra molécula por geometría complementaria al centro activo de dicha proteína. Esta unión es altamente específica. La especificidad de la unión del centro activo de las proteínas con otras moléculas se basa en el plegamiento particular de cada proteína (estructura terciaria o cuaternaria) que, en último término, dependen de la secuencia de aminoácidos y esta es propia de cada especie (39).

En el caso de la Proteína Ataxina-1 normal, se han realizado estudios para determinar con que proteína interacciona, pero estos estudios aún no han concluido, por ahora se especula que puede interaccionar con la proteína Gliceraldehído-3-fosfato deshidrogenasa (GAPDH) (40), esta enzima, no solo actúa en el proceso de glicolisis, ya que se ha descubierto su acción como un activador de la transcripción, ya que se mueve entre el citosol y el núcleo de la células (41). Se ha demostrado también que la enzima GAPDH inicia la apoptosis, ya que bajo estrés se une a la proteína *Siah1* (E3 ubiquitin ligase seven in absentia homolog-1), que tiene como objetivo degradar proteínas nucleares (42, 59).

Otros estudios revelan que la proteína Ataxina-1 no solo interactúa con la enzima GAPDH, sino también con la proteína acidica nuclear rica en Leucinas, por sus siglas en inglés LANP (leucine-rich acidic nuclear protein). Esta proteína se expresa predominantemente en las células de Purkinje, sitio más que notorio para el desarrollo la enfermedad SCA1. La interacción entre LANP y Ataxina-1 es significativamente más fuerte cuando el número de glutaminas se incrementa, es decir cuando esta mutada. Los estudios de inmunofluorescencia demuestran que tanto LANP y Ataxina-1 se localizan en la matriz nuclear asociada a estructuras subnucleares. (43,44). Las interacciones entre las proteínas Ataxina-1 mutada y LANP, podrían ser la causa de la muerte de las células neuronales de Purkinje, una de las principales características de la SCA1, cuya pérdida causa atrofia cerebelosa progresiva, el primer síntoma de la enfermedad.

Como hemos visto en la Figura 4.4 y como veremos en las Figuras 4.15 y 4.17 la mutación en la proteína Ataxina-1 produce un cambio de conformación y plegamiento comparada con la proteína normal. Diversos estudios han determinado que este cambio induce a la formación de agregados insolubles en el núcleo, participan en el secuestro de factores importantes mediante interacciones proteicas aberrantes impidiendo que realicen sus funciones biológicas normales. Algunos de estos factores secuestrados son los de LANP, que presentan una expresión temporalmente restringida en las células Purkinje del cerebelo, en las células del tronco cerebral y las neuronas de la corteza. Esta selectividad en el sistema nervioso central explicaría la especificidad celular de los procesos neurodegenerativos (45).

En las Figuras 4.15 y 4.17 apreciamos las vistas isométricas de las moléculas de Ataxia-1 normal y mutada respectivamente, donde podemos observar la conformación final que han adquirido ambas moléculas luego de la optimización geométrica. Estas imágenes nos permiten conocer las posiciones en que se encuentran los residuos debajo de la malla generada por el Potencial Electroestático

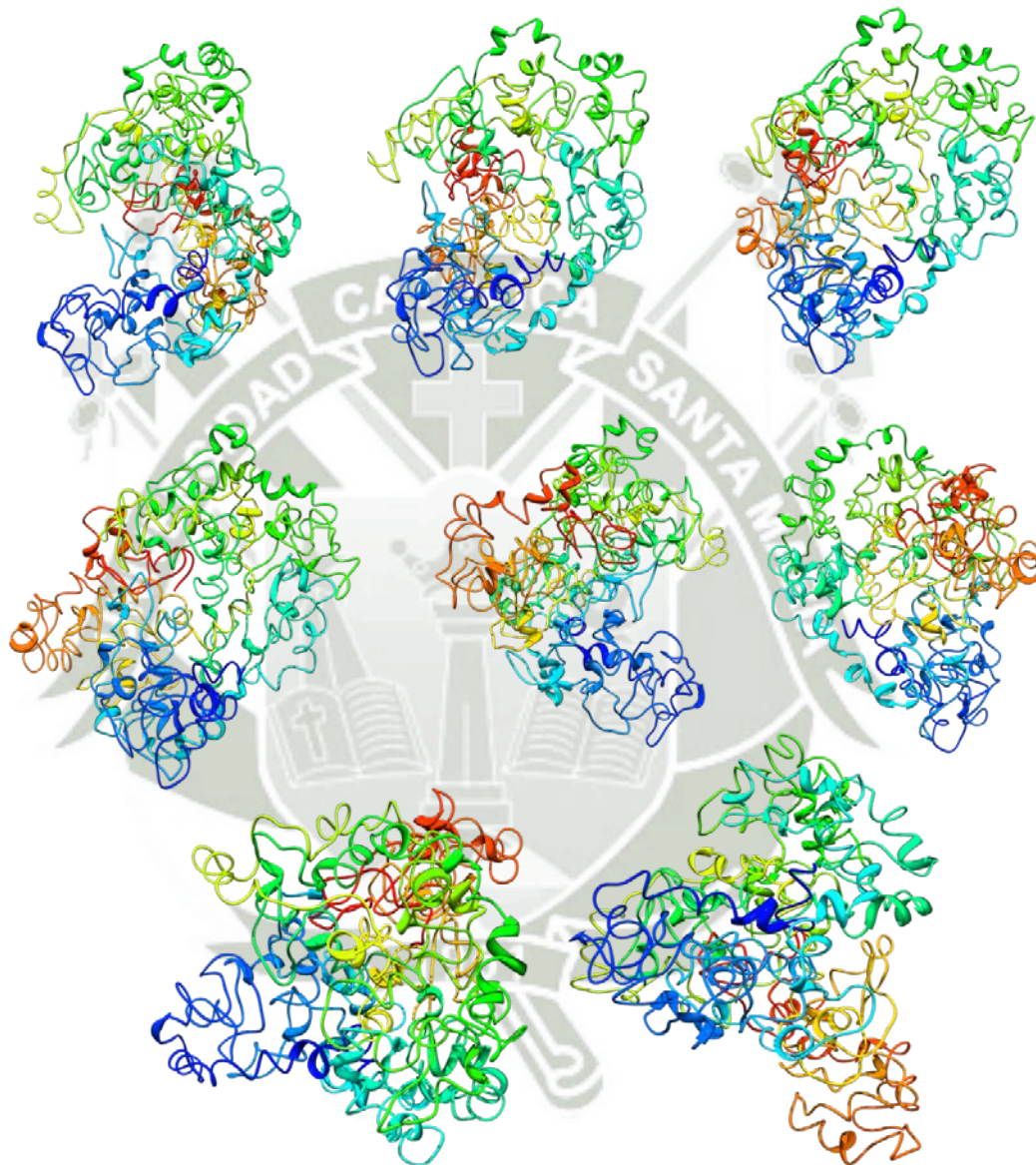


Figura 4.15: **Vista isométrica de Ataxina-1 luego de la simulación de Dinámica Molecular.** En donde vemos la molécula en 6 posiciones diferentes, obtenidas de rotar la molécula 60° . La dos figuras inferiores muestran la vista superior e inferior respectivamente. Figuras obtenidas mediante el programa Chimera 1.5.3

En las Figuras 4.16 y 4.18, se puede observar la superficie del potencial electrostático de las moléculas Ataxina-1 normal y mutada respectivamente; en

donde apreciamos las diferencia entre ambas moléculas sobre el potencial electrostático.

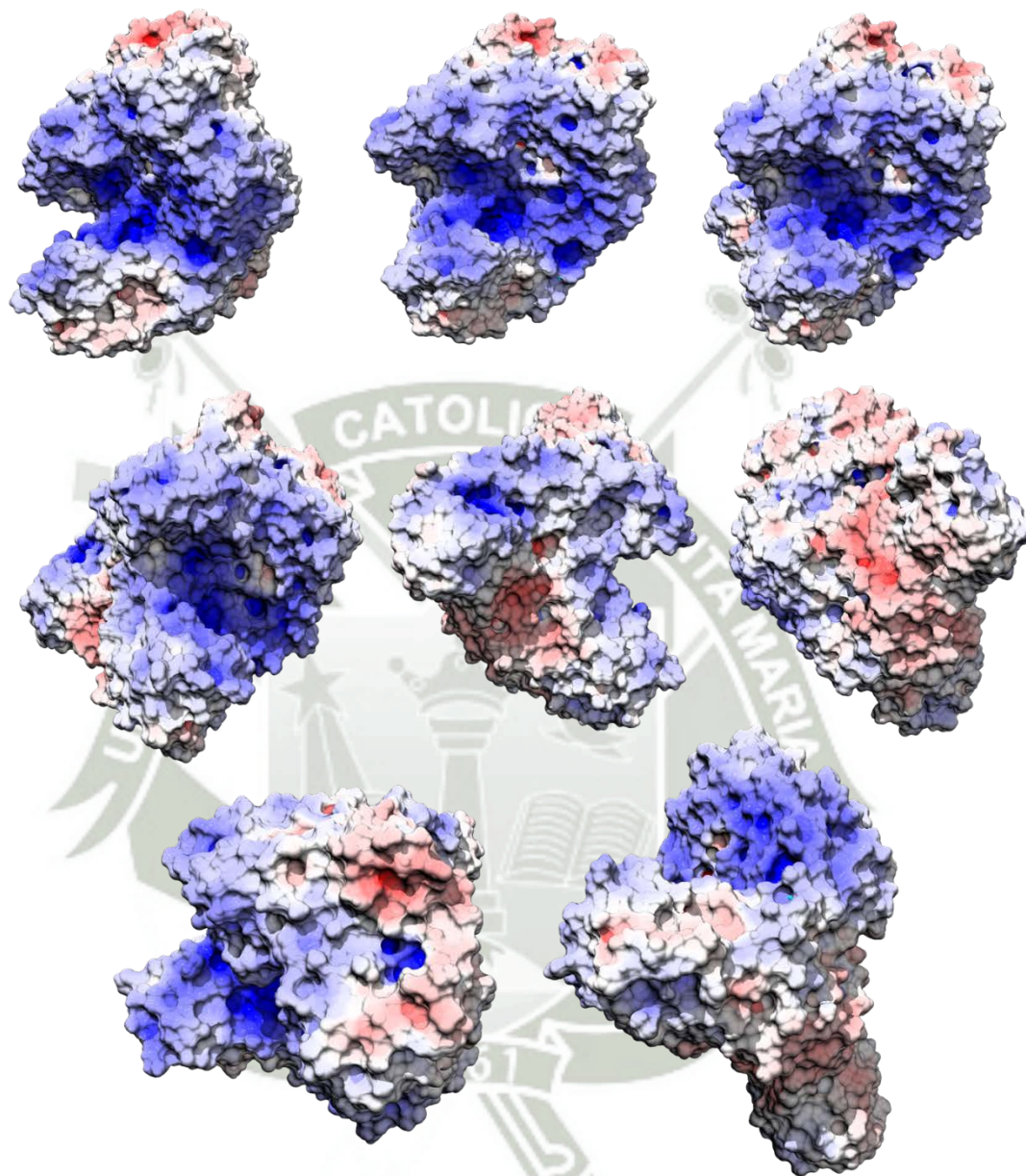


Figura 4.16: **Potencial Electrostático de Ataxina-1.** En esta figura se muestra la superficie de potencial electrostático, el cual se grafico con APBS (Baker, 2001), en donde el potencial electrostático negativo es de color en tonos de rojo y potencial electrostático positivo de color en tonos de azul. Figura obtenida mediante la utilización del Chimera 1.5.3.

En la Figura 4.16, observamos claramente que la Molécula Ataxina-1 posee 2 cavidades, una en la cara anterior con un potencial electrostático positivo y otra en la cara posterior con un potencial electrostático también positivo, al mismo tiempo notamos que en la sexta posición, es decir el lateral derecho, posee un potencial

electrostático negativo. Con estos resultados podemos afirmar que esta molécula atraerá moléculas cargadas negativamente hacia las cavidades, ya que una mayor superficie presenta cargas positivas, y estos sitios son cavidades, lugares que se hallan asociados preferentemente como sitios de unión estructural o sitios activos.

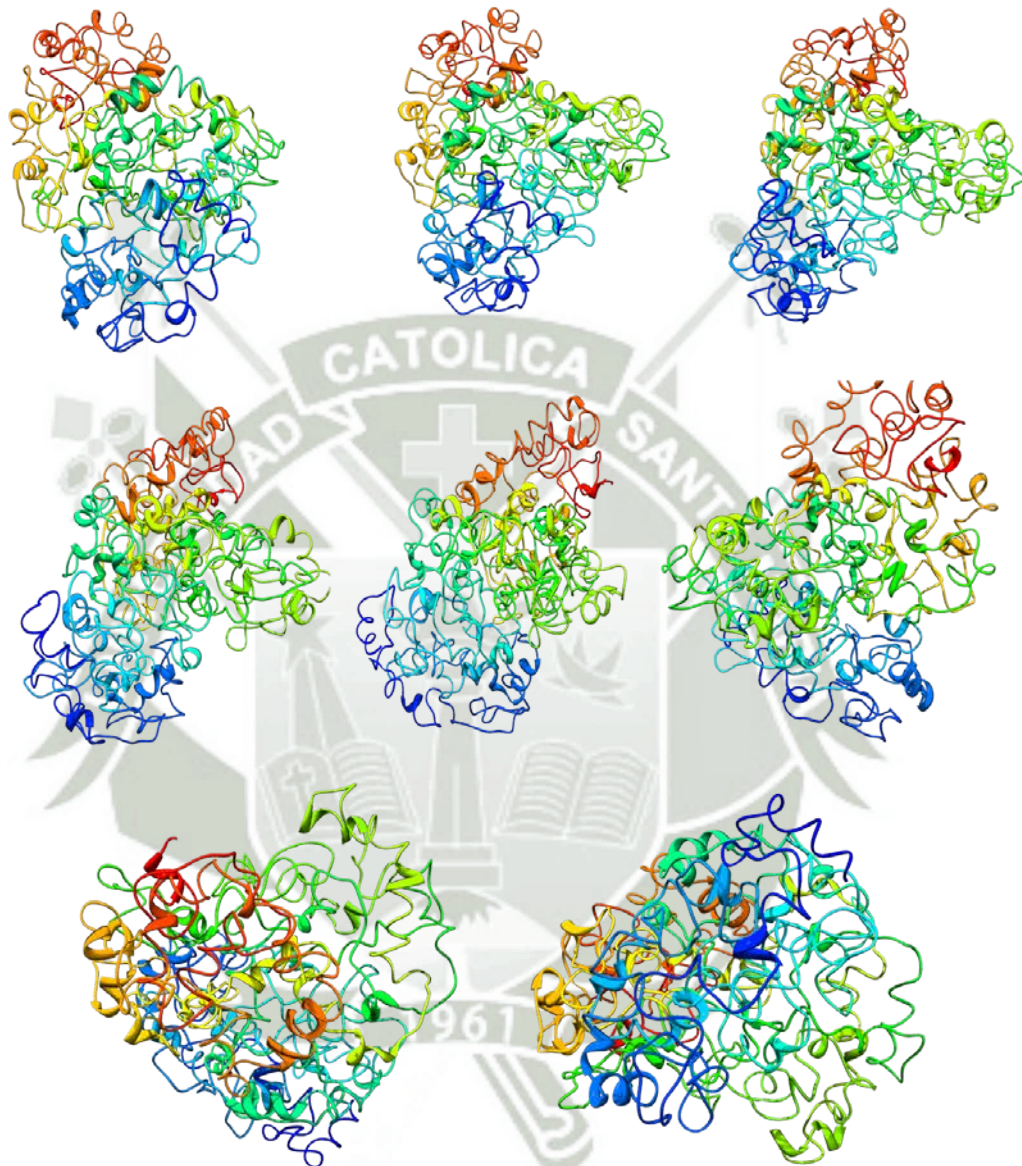


Figura 4.17: **Vista isométrica de Ataxina-1 mutada luego de la simulación de Dinámica Molecular.** En donde vemos la molécula en 6 posiciones diferentes, obtenidas de rotar la molécula 60°. La dos figuras inferiores muestran la vista superior y inferior respectivamente. Figuras obtenidas mediante el programa Chimera 1.5.3

Luego en la Figura 4.18, podemos observar claramente que la Molécula Ataxina-1 mutada no posee ninguna cavidad, como si lo hace su homólogo normal. Notamos además que las caras anterior y posterior presentan un potencial electrostático

positivo y que en forma similar a la estructura normal la sexta posición, es decir el lateral derecho, posee un potencial electrostático negativo. Con estas observaciones podemos afirmar que la molécula Ataxina-1 mutada, que a pesar de estar cargada en los mismos lugares que la molécula normal no presenta cavidades de especificidad estructural, solo posee una superficie plana. Lo cual indica que no funcionan como sitios de unión selectivos, como es el caso de la molécula normal.

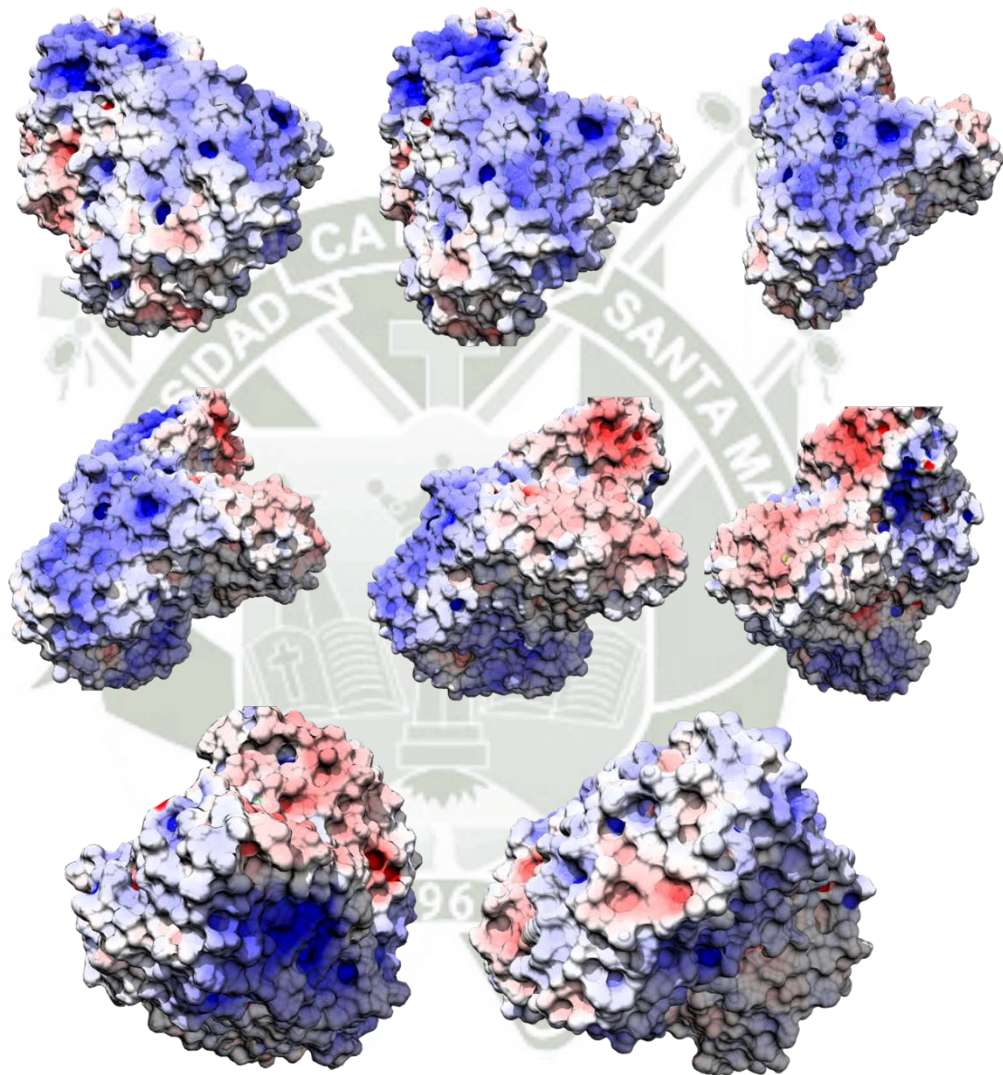


Figura 4.18: **Potencial Electrostático de Ataxina-1 mutada.** En esta figura se muestra la superficie de potencial electrostático, el cual se grafico con APBS (Baker, 2001), en donde el potencial electrostático negativo es de color en tonos de rojo y potencial electrostático positivo de color en tonos de azul. Figura obtenida mediante la utilización del Chimera 1.5.3.

De acuerdo a lo explicado, podemos deducir que la molécula Ataxina-1 interactúa primero con la enzima GAPDH y luego con la proteína LANP bajo el método de interacción no específica de características alostéricas. Una unión alostérica se basa

en el modo de regulación que poseen las proteínas o enzimas, por el cual la fijación de una molécula en una ubicación (sitio catalítico) puede modificar las condiciones de fijación de otra molécula, en una ubicación distante de la enzima (46). En la Figura 4.20 observamos un ejemplo de este proceso.

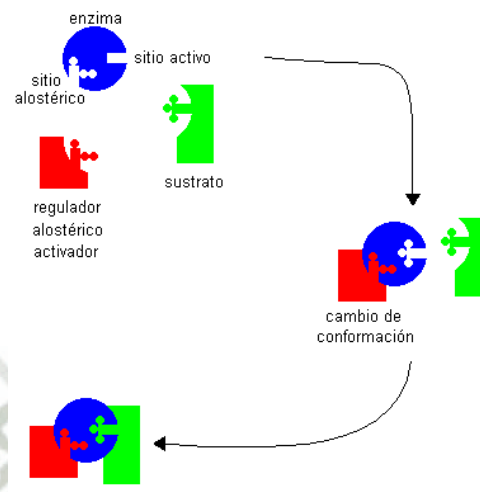


Figura 4.19: **Regulación alostérica en enzimas.** En esta figura se muestra la descripción del Proceso de Regulación alostérica.

En la Figura 4.20 observamos el Potencial electrostático y la ubicación de la cadena poliglutamínica de la Proteína Ataxina-1 normal, la zona de la cadena poliglutamínica se encuentra cargada en su gran mayoría electropositivamente y está ubicada en un costado de la cavidad presente en la molécula. Se determinó que la proteína Ataxina-1 interaccionara con la enzima GAPDH, a través del dominio N-terminal (aminoácidos 1-300) de la proteína Ataxina-1, este dominio se caracteriza por contener la cadena Poliglutamínica; y a través de su dominio de unión nicotinamida adenina dinucleótido (NAD) presente en la enzima GAPDH (51, 52, 53). De lo anterior podemos establecer que la enzima GAPDH se unirá a la proteína Ataxina-1 por su región N-terminal, donde se ubica la cadena poliglutamínica, es decir en un costado de la cavidad o sitio activo de la proteína. La enzima GAPDH, es una enzima deshidrogenasa, la que causará una reacción de reducción al momento de interactuar con la proteína; produciendo que la enzima GAPDH gane hidrogeniones y que la proteína Ataxina-1 los pierda. Esta unión se verá reforzada gracias a la presencia de Histidina, aminoácido que posee un anillo imidazol, altamente protonado, que brindaría los protones a la enzima GAPDH. Esta acción produciría un

cambio en las cargas electrostáticas, sobre todo en el sitio activo de la proteína, que se encuentra bastante cercano al sitio donde interactúan la proteína y la enzima; es así que el sitio activo perdería protones y se convertiría en un sitio neutro.

Estos cambios electrostáticos facilitarían la unión de la proteína LANP al sitio activo, debido a que esta proteína se caracteriza por poseer un dominio N-terminal de repeticiones ricas en leucinas (LRR), dichas repeticiones se producen cinco veces a lo largo de la cadena polipeptídica, comenzando en el residuo 10 y extendiéndose hasta el residuo 128, lo que comprende casi el 50% de toda la proteína (54). Conociendo que la proteína está formada en su mayoría por leucinas, aminoácido que se caracteriza por su cadena alifática y por ser apolar, es decir que presenta una carga neutra, entonces podemos asumir que buscara unirse a un sitio neutro, como lo es el sitio activo de la Ataxina-1.

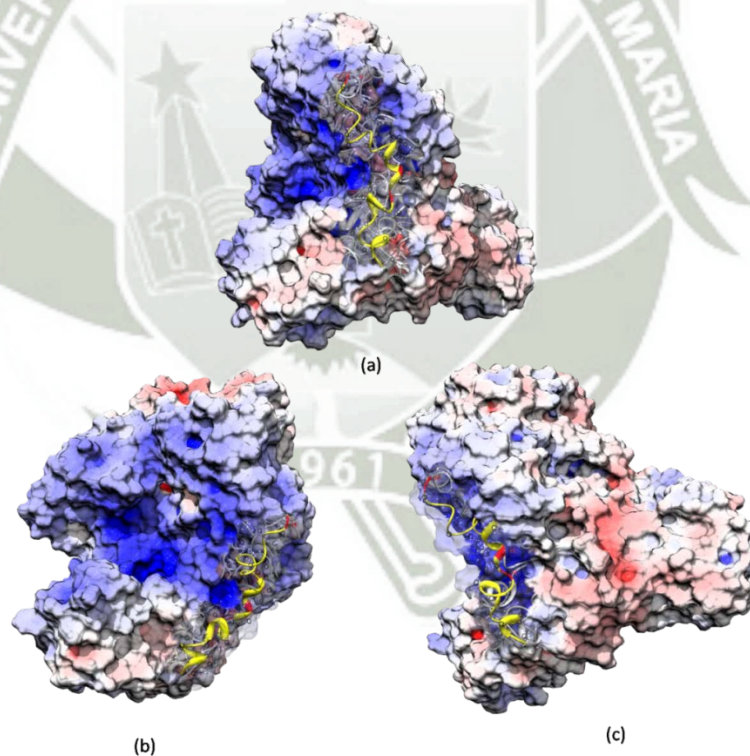


Figura 4.20: **Potencial Electrostático de la proteína Ataxina-1 y señalización de la Cadena Poliglutamínica.** En esta figura se muestra la Cadena Poliglutamínica en amarillo y en rojo la Histidina presente, (a) vista de frente, (b) vista del lado derecho y (c) vista del lado izquierdo. Figura obtenida mediante la utilización del Chimera 1.5.3.

En la Figura 4.21 podemos ver la proteína mutada y la ubicación de su cadena poliglutamínica, que a diferencia de la proteína normal, no se encuentra ubicada en un costado, sino en el medio de la proteína en una zona completamente electropositiva y que esta proteína no presenta ningún potencial sitio activo, como si lo hacia la proteína normal. Notamos también que a diferencia de la Proteína Ataxina-1 normal, cuya cadena presenta interrupciones de residuos histidina, la proteína Ataxina-1 mutada no los presenta, siendo este aminoácido, frecuentemente ubicado en los centros activos de las proteínas (56,57,58), lo cual podría darnos una explicación del porque la molécula Ataxina-1 mutada no presenta ninguno.

En el caso de la enfermedad SCA1, también se ha determinado que las interrupciones de histidina, en la secuencia de repeticiones de glutaminas, juegan un rol de estabilización (47). Además, se cree que la presencia de histidina parece reducir el riesgo de enfermedad, esta afirmación se basa en observaciones realizadas a pacientes con SCA1, los cuales se encuentran desprovistos de inserciones de histidina (48). Y en que algunos individuos que están libres de la enfermedad a pesar de poseer un alelo SCA1 presentan codones de histidina en la cadena poliglutamínica (49).

Con lo establecido, se puede generar la siguiente hipótesis; las interrupciones de histidina presentan un efecto supresor, ya que este aminoácido juega un papel importante en el control de agregados nucleares (50). Siendo los agregados la principal causa de la muerte celular y la consecuente neurodegeneración celular (55), es por ello que al verse interrumpida la agregación nuclear la enfermedad Ataxia Espinocerebelosa tipo 1 no se desarrollaría.

Como se ha descrito la Proteína Ataxina-1 mutada ha cambiado estructuralmente, posee una cadena expandida de glutaminas y no presenta histidinas en la misma; se cree también que la gran mayoría de las proteínas con un dominio poliglutamínico expandido están relacionadas con la regulación de la transcripción (61). Este cambio es el que permite el desarrollo de la enfermedad.

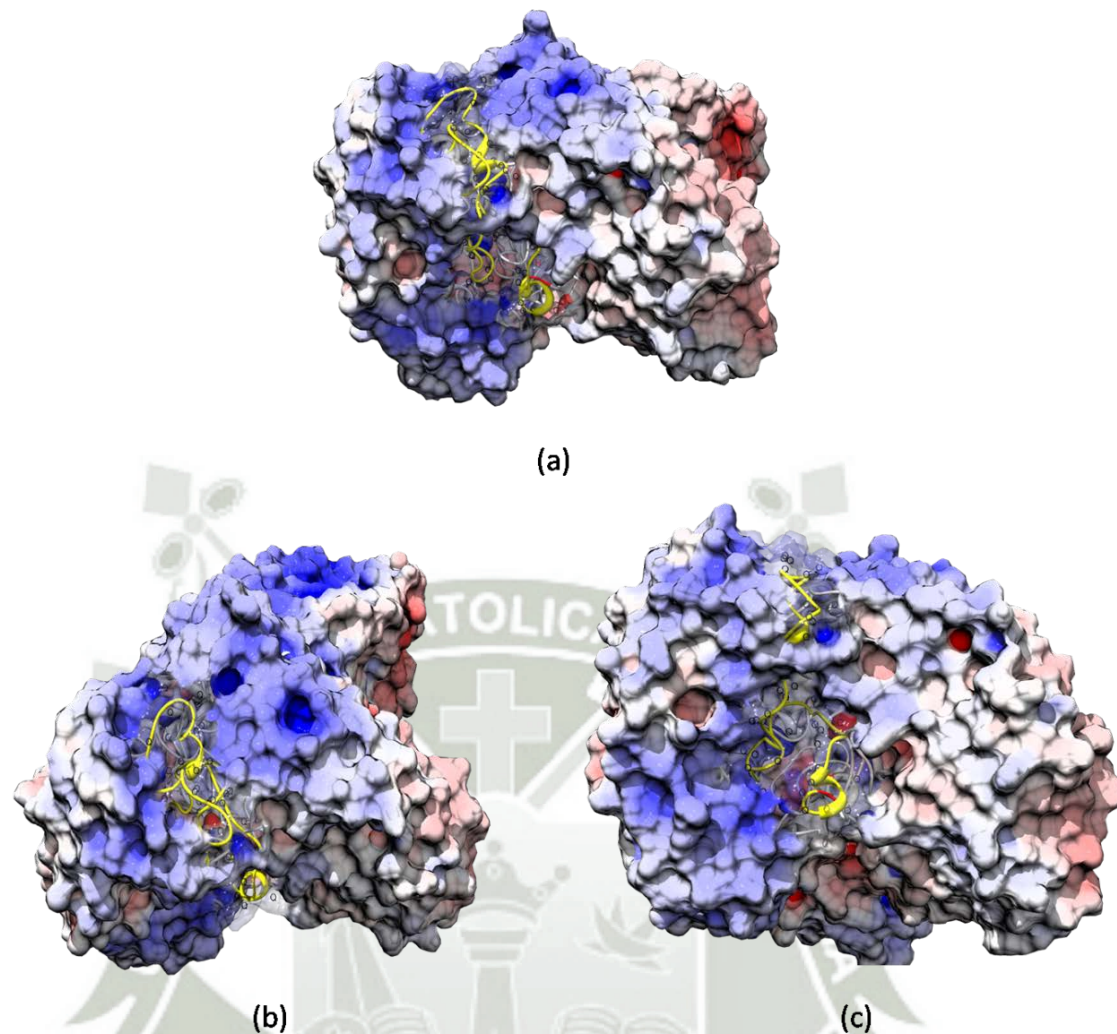


Figura 4.21: **Potencial Electrostático de la proteína Ataxina-1 Mutada y señalización de la Cadena Poliglutamínica.** En esta figura se muestra la Cadena Poliglutamínica en amarillo y en rojo la Histidina presente, (a) vista de frente, (b) vista del lado derecho y (c) vista del lado izquierdo. Figura obtenida mediante la utilización del Chimera 1.5.3.

Por medio de la descripción dada, deducimos que la proteína Ataxina-1 mutada interaccionará con la enzima GAPDH, la cual se unirá en la región donde se ubica la Cadena poliglutamínica, como se aprecia en la Figura 4.21, esta zona se encuentra cargada electropositivamente. Al igual que ocurre en la interacción de la proteína normal con la enzima, en la proteína Ataxina-1 mutada también se producirá un cambio en el potencial electrostático, es así que la molécula perderá protones y se volverá neutra, pero a diferencia de la molécula normal, que posee histidinas que genera una unión entre la molécula y la enzima por medio de puentes de hidrogeno; en el caso de la molécula mutada la unión se generara por la presencia de un número creciente de residuos de glutamina, lo cuales no sólo aumentan el número de enlaces

cruzados por molécula, sino que también vuelven cada residuo más reactivo (60), lo cual producen enlaces más estables e incluso permanentes, a diferencia de los enlaces de la molécula normal. Estos cambios electrostáticos, al igual que en la molécula normal, cambiarían el potencial electrostático de la proteína Ataxina-1 mutada volviéndola más neutra.

Este cambio de potencial favorecerá la unión entre la Ataxina-1 mutada y la proteína LANP ya que la LANP presenta en su conformación una zona rica en leucinas, aminoácido hidrofóbico, apolar y con una carga neutra que buscará interactuar con regiones neutras de la molécula. Entonces en esta zona se formarían uniones por medio del enlace no covalente, es decir fuerzas intermoleculares, que formarían un complejo proteico de gran estabilidad estructural. Por lo mencionado anteriormente deducimos que la proteína LANP actúa como un posible cofactor patogénico, especialmente porque la interacción que se desarrolla en el complejo proteico es mayor a medida que es más larga es la cadena de Poliglutaminas. En forma normal LANP participa en la regulación de la transcripción de proteínas, pero al interactuar con la proteína Ataxina-1 mutada, sus función se ve alterada, lo cual genera agregados nucleares en las células de Purkinje y desencadena la Ataxia Espinocerebelosa tipo 1.

Conclusiones

1. De la NCBI, se obtuvo la secuencia FASTA de la proteína Ataxina-1, se modificó dicha secuencia con el software MOLDEN, agregando glutaminas para obtener la secuencia FASTA de la proteína Ataxina-1 mutada. Se realizó la simulación de Dinámica Molecular con un ensamble canónico y mediante la utilización del software GROMACS y se observó la estabilidad energética de ambas proteínas, la cual se alcanzó a los 30 ns o 30 millones de pasos.
2. Con el software CHIMERA se visualizaron las estructuras terciarias de las proteínas normal y mutada, donde vemos las diferencias estructurales que ha generado la expansión de glutaminas, es así que la proteína Ataxina-1 normal presenta una cadena poliglutamínica con α -hélices ubicado en un costado de la estructura, mientras que la Ataxina-1 mutada presenta una cadena poliglutamínica formada en su mayoría por cuerdas y que se encuentra ubicada en una zona interna de la estructura.
3. Se utilizó el software APBS para obtener los Potenciales Electrostáticos de las proteínas Ataxina-1 y Ataxina-1 mutada, donde visualizamos que la molécula normal posee un sitio activo cargado positivamente y en un costado de este sitio se ubica la cadena poliglutamínica. La molécula mutada no presenta sitio activo alguno y la cadena poliglutámica se encuentra en una zona interna de la molécula cargada en su gran mayoría positivamente.
4. Se determinó por medio de los Potenciales Electrostáticos y la literatura que la proteína Ataxina-1 interactúa con la enzima GAPDH y la proteína LANP, mediante una interacción de características alóstericas, lo cual forma un complejo proteico, en caso de la proteína normal esta interacción es inestable,

mientras que la proteína mutada forma un complejo mucho más estable que lleva al desarrollo la enfermedad Ataxia Espinocerebelosa Tipo1.



Sugerencias

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

- Es importante elegir las secuencias y estructuras de las proteínas de una base de datos confiable. De igual forma es necesario tener en cuenta las referencias Bibliográficas de las secuencias y estructuras.
- Considerando que se ha determinado que la proteína Ataxian-1 interactúa con la enzima GAPDH y la proteína LANP, se sugiere que se realiza el diseño de estas proteínas y luego el docking de las mismas, para poder determinar el lugar exacto de las interacciones, y así poder comprobar nuestras suposiciones.
- Implementar un laboratorio de Bioinformática que este equipado de forma adecuada para realizar cálculos computacionales de alto nivel.

REFERENCIAS

- [1] Rubinsztein, D. C. The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* 2006, 443, 780-786.
- [2] Thompson, L. M. Neurodegeneration: a question of balance. *Nature* 2008, 452, 707-708.
- [3] Mederos, L. E. A.; Zaldivar, Y. G.; Gotay, D. A.; Mesa, J. M. L.; Almarales, D. C. Genes modificadores en enfermedades poliglutaminicas. *Revista Habanera de Ciencias Médicas* 2008, 8 year.
- [4] Palencia, R.; Galicia, G.; Ballesteros, M. A. Ataxias hereditarias. *Boletín de la Sociedad de Pediatría de Asturias, Cantabria, Castilla León* 2004, 44, 120-127.
- [5] Stevanin, G.; Dürr, A.; Brice, A. Clinical and molecular advances in autosomal dominant cerebellar ataxias: from genotype to phenotype and physiopathology. *European Journal of Human Genetics* 2000, 8, 4-18.
- [6] Gustinari, R. S.; Jr., M. C.; Pimentel, M. M.; Duarte, S. F.; Lima, M. A. C. Genetic polymorphism at spinocerebellar ataxia 1 and 2 loci in Brazil. *Genetics and Molecular Research* 2003, 2, 360-365.
- [7] Hong, S.; Kang, S. Molecular pathogenesis of spinocerebellar ataxia type 1 disease. *Molecules and Cells* 2009, 27, 621-627.
- [8] De Michele, G.; Perrone, F.; Filla, A.; Mirante, E.; Giordano, M.; De Placido, S.; Campanella, G. Age of onset, sex, and cardiomyopathy as predictors of disability and survival in Friedreich's disease: a retrospective study on 119 patients. *Neurology* 1996, 47, 1260-1264.

- [9] Maruyama, H.; Izumi, Y.; Morino, H.; Oda, M.; Toji, H.; Nakamura, S.; Kawakami, H. Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia: a study of 1286 Japanese patients. *The American Journal of Medical Genetics* 2002, 578-583.
- [10] Chung, M.; Ban, S.; HT, O. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 1993, 221-226.
- [11] Hellenbroich, Y.; Dalski, A.; Ch, Z. Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. *European Journal of Human Genetics* 2001, 9, 160-164.
- [12] HT, O.; MY, C.; Ban, S.; Kwiatkowski, T. J.; Servadio, A.; Beaudet, A. L.; McCall, A. E.; Duvick, L. A.; Ranum, L. P.; Zoghbi, H. Y. Expansion of an unstable trinucleotide (CAG) repeat in spinocerebellar ataxia type 1. *Nature Genetics* 1993, 221-226.
- [13] MY, C.; Ranum, L. P.; Duvick, L. A.; Servadio, A.; Zoghbi, H. Y.; HT, O. Evidence for a mechanism predisposing to intergenerational CAG instability in spinocerebellar ataxia type 1. *Nature Genetics* 1993, 254-258.
- [14] HT, O.; Zoghbi, H. Y. Pathogenic Mechanisms of a Polyglutamine-mediated Neurodegenerative Disease, Spinocerebellar Ataxia Type 1. *The journal of biological chemistry* 2008, 284, 7425-7429.
- [15] Yoo, S. Y.; Pennesi, M. E.; Weeber, E. J.; Xu, B.; Atkinson, R.; Chen, S.; Armstrong, D. L.; Wu, S.; Zoghbi, H. Y. SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity. *Neuron* 2003, 37, 383-401.
- [16] Matilla, A. Ataxia Espinocerebelosa Autosómica Dominante Tipo 1 (SCA1). *Euro-Ataxia* 2002.

- [17] Huda Y. Zoghbi y Harry T. Orr Pathogenic Mechanisms of a Polyglutamine-mediated Neurodegenerative Disease, Spinocerebellar Ataxia Type 1. *Journal of Biological Chemistry* 2009, 284, 7425 – 7429.
- [18] Ross, C. A y Poirier M. A Protein aggregation and neurodegenerative disease. *Nature Medical*, 2004, 10 (Suppl), S10-S17.
- [19] Perutz, M. F. Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends Biochemical Science*. 1999, 24, 58-63.
- [20] Zhang J, Guenther MG, Carthew RW, Lazar MA. Proteasomal regulation of nuclear receptor corepressor-mediated repression. *Genes Dev* 1998; 12: 1775-1780.
- [21] Lin X, Antalffy B, Kang D, Orr HT, Zoghbi HY. Polyglutamine expansion down-regulates specific neuronal changes in SCA1. *Nat. Neuroscience* 2000; 3(2):157-163.
- [22] Gálvez, E. C. Ph.D. thesis, Universidad Internacional de Andalucía.
- [23] Mooney, *Comprehensive Study Using Bioinformatics Predicts the Molecular Causes of Many Genetic Diseases*. *Biology & Nature* 2010.
- [24] Aliseda Casado Javier Diseño de un simulador de Dinámica Molecular basado en CORBA, 2011.
- [25] Naranjo Feliciano, Dany Análisis in silico de la catepsina B de *Fasciola Hepatica* como diana terapéutica, *REDVET* 2009, Vol 10 N°2.
- [26] Hernando Castañeda M. Wladimir Rodriguez G. y Eliézer Colina M. Clasificación de Trayectorias en Dinámica Molecular Usando Relaciones de Equivalencia Difusa y Análisis de Componentes Principales, 2007.

- [27] Espinoza García Joaquín. Superficies de energía potencial y reactividad química 2da edición Septem ediciones, 2004
- [28] Snyman Jam Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms, 2005
- [29] Frenkel, D., y SMIT, B. Understanding Molecular Simulations, vol. 1. Academic Press, 2002.
- [30] Kerson Huang Statistical Mechanics 2da Edición John Wiley and Sons 1987
- [31] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037-10041 (2001)
- [32] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera a visualization system for exploratory research and analysis *J Comput Chem.* 2004 Oct;25(13):1605-12
- [33] Instituto Nacional de Trastornos Neurológicos y Accidentes (NINDS). 2010 Julio 26.
- [34] Obdulia Rabal Gracia, Jordi Teixidó i Closa Herramientas de cribado virtual aplicadas a inhibidores de tirosina quinasas. Contribución al desarrollo del programa PRALINS para el diseño de quimiotecas combinatorias. (1990)
- [35] Iván Y. Torshin, Irene T. Weber, Robert W. Harrison Geometric criteria of hydrogen bonds in proteins and identification of `bifurcated' hydrogen bonds. Oxford Journals, Vol 15 pg 359-363 (2002).
- [36] Electrostatic Modeling Of Protein Aggregation, Ram Vanam. Indiana University December, 2004

- [37] NCBI [Web en línea]. < > <http://www.ncbi.nlm.nih.gov/protein/189491748>.
[Consulta: 21/10/2011]
- [38] Harvey Lodish, Arnold Berk Biología celular y molecular, 2005
- [39] Koshy, B., Matilla, T., Burrigh, E.N., Merry, D.E., Fischbeck, K.H., Orr, H.T., Zoghbi, H.Y. Spinocerebellar ataxia type-1 and spinobulbar muscular atrophy gene products interact with glyceraldehyde-3-phosphate dehydrogenase. *Human Molecular Genetics*. (1996)
- [40] Zheng L, Roeder RG, Luo Y. S phase activation of the histone H2B promoter by OCA-S, a coactivator complex that contains GAPDH as a key component. *Cell*. 2003 Jul 25;114(2):255-66.
- [41] Hara MR, Thomas B, Cascio MB, Bae BI, Hester LD, Dawson VL, Dawson TM, Sawa A, Snyder SH. Neuroprotection by pharmacologic blockade of the GAPDH death cascade. *Proc Natl Acad Sci U S A*. 2006 Mar 7;103(10):3887-9. Epub 2006 Feb 27.
- [42] Cvetanovic, M., Rooney, R.J., Garcia, J.J., Toporovskaya, N., Zoghbi, H.Y., Opal, P. The role of LANP and ataxin 1 in E4F-mediated transcriptional repression. *EMBO Rep*. (2007).
- [43] Matilla A, Koshy BT, Cummings CJ, Isobe T, Orr HT, Zoghbi HY. The cerebellar leucine-rich acidic nuclear protein interacts with ataxin-1. *Nature*. 1997 Oct 30;389(6654):974-8.
- [44] Matilla Dueñas Antoni. La ataxia espinocerebelosa tipo 1 (SCA1): rutas moleculares, mecanismos patofisiológicos, y estrategias terapéuticas. Jornadas científicas internacionales de euro-ATAXIA.
- [45] J. Monod, J. Wyman, J.P. Changeux. (1965). On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.*, May;12:88-118.

- [46] Klement,I.A., Skinner,P.J., Kaytor,M.D., Yi,H., Hersch,S.M., Clark,H.B, Zoghbi,H.Y. and Orr,H.T. Ataxin-1 nuclear localization and aggregation: Role in polyglutamine-induced disease in SCA1 transgenic mice Cell, Number 95, 41–53.(1998).
- [47] Zoghbi HY, Orr HT. Spinocerebellar ataxia type 1. *Semin Cell Biol.* 1995 Feb;6(1):29-35.
- [48] Quan,F., Janas,J. and Popovich,B.W. A novel CAG repeat configuration in the SCA1 gene: implications for the molecular diagnostics of spinocerebellar ataxia type 1 *Hum. Mol. Genet.*, 4, 2411–2413 (1995).
- [49] Fraser PE, McLachlan DR, Surewicz WK, Mizzen CA, Snow AD, Nguyen JT, Kirschner DA. Conformation and fibrillogenesis of Alzheimer A beta peptides with selected substitution of charged residues. *J Mol Biol.* 1994 Nov 18;244(1):64–73.
- [50] Ercolani, L., Florence, B., Denaro, M. and Alexander, M. (1988) Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. *J. Biol. Chem.*, **263**, 15335–15341.
- [51] Harris, J.I. and Waters, M. (1976) In Boyer, P.D. (ed.), *The Enzymes*. Academic Press, Boca Raton, Vol. 13, pp. 1–49.
- [52] Brodie, A.E. and Reed, D.J. (1987) Reversible oxidation of glyceraldehyde 3phosphate dehydrogenase thiols in human lung carcinoma cells by hydrogenperoxide. *Biochem. Biophys. Res. Comm.*, **148**, 120–125.
- [53] K. Matsuoka, M. Taokat, N. Satozawa, H. Nakayama, T. Ichimurat, N. Takahashi, T. Yamakunit, S.-Y. Songt, T. Isobe A nuclear factor containing the leucine-rich repeats expressed in murine cerebellar neurons. *Biochemistry* Vol. 91, pp. 9670-9674, October 1994
- [54] Davies SW, Turmaine M, Cozens BA, DiFiglia M, Sharp AH, Ross CA, Scherzinger E, Wanker EE, Mangiarini L, Bates GP. Formation of neuronal

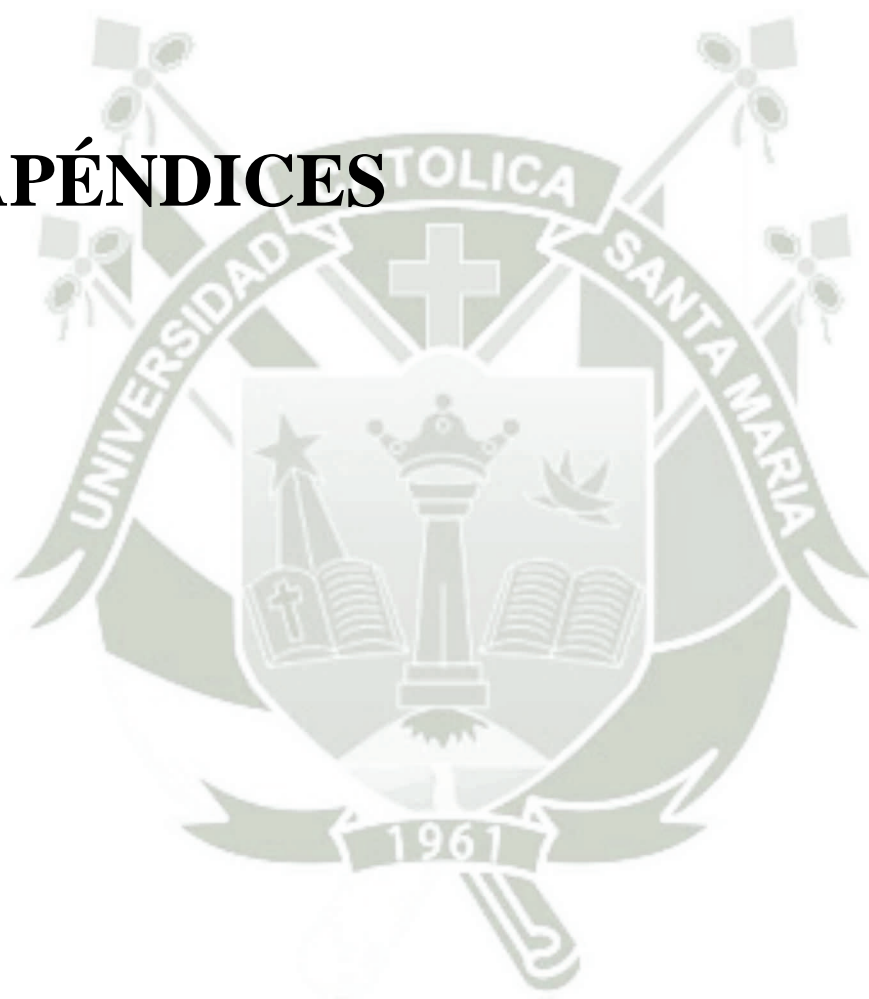
- intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell*. 1997 Aug 8;90(3):537-48.
- [55] Röther D, Poppe L, Viergutz S, Langer B, Rétey J. Characterization of the active site of histidine ammonia-lyase from *Pseudomonas putida*. *Euro Journal Biochemistry*. 2001 Dec; 268(23):6011-9.
- [56] Paul J. Gane, Jim M. Dunwell and Jim Warwickr. Modeling based on the structure of vicilins predicts a histidine cluster in the active site of oxalate oxidase *Journal of Molecular Evolution* Volume 46, Number 4 (1998), 488-493.
- [57] Yogesh C. Awasthi, Aruni Bhatnagar, Shivendra V. Singh. Evidence for the involvement of histidine at the active site of glutathione S-transferase ψ from human liver *Biochemical and Biophysical Research Communications* Volume 143, Issue 3, 30 March 1987, Pages 965–970.
- [58] E. Chepchumba K. Yego, Susanne Mohr. Seven in absentia homolog-1 protein is necessary for high glucose-induced GAPDH nuclear accumulation and cell death in müller cells *J Biol Chem*. 2010 Jan 29;285(5):3181-90. Epub 2009 Nov 23.
- [59] P. Kahlem, C. Terre, H. Green, P. Djian. Peptides containing glutamine repeats as substrates for transglutaminase-catalyzed cross-linking: Relevance to diseases of the nervous system. *Proceedings of the National Academy of Sciences USA* Vol. 93, pp. 14580–14585, December 1996.
- [60] Gerber HP. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 1994; 263: 808-811.
- [61] Bok KS, Rhim H, Yoo YD, Choi E, Ahn K, Kim IH, Kang S. Expanded polyglutamine tract itself induces cell death in cultured cells. *Mol Cells* 1999; 9(4): 398-402.
- [62] Green H. Human genetic disease due to codon reiteration: relationship to an evolutionary mechanism. *Cell* 1993; 74: 955-956.

- [63] M. Selma Arias Pérez. *Mecánica molecular*. Departamento de Química organica- Universidad de Alcalá. 2002, 99-129.
- [64] Feynman, Richard P. (2011). *Work and potential energy. The Feynman Lectures on Physics, Vol. I*. Basic Books. p. 13.
- [65] Joaquín Espinosa García Superficies de energía potencial y reactividad química Segunda Edición. 2004.
- [66] Gloria Moyano, José G. López Superficies de energía potencial Universidad de Antioquia 12 de junio de 2009
- [67] Federico Gago Badenas. Métodos computacionales de modelado molecular y diseño de fármacos. 2005
- [68] McCammon, J. A.; Harvey, S. C. (1987b) "Dynamics of Proteins and Nucleic Acids", Cambridge University Press, Cambridge.
- [69] Huang Kerson Statistical Mechanics (2ed , Wiley, 1987).
- [70] Kittel, Charles; Herbert Kroemer. *Thermal Physics, Second Edition*. San Francisco: W.H. Freeman and Company. (1980)
- [71] Franca Carlos, Estudio teórico de propiedades fisicoquímicas de compuestos de interés farmacológico. 2004.
- [72] Gromacs,[Web en línea]. < > http://www.gromacs.org/About_Gromacs>. [Consulta: 21/10/2012]
- [73] Molden [Web en línea]. < > <http://www.cmbi.ru.nl/molden/>>. [Consulta: 21/10/2012]
- [74] UCSF Chimera [Web en línea]. < > <http://www.cgl.ucsf.edu/chimera/>>. [Consulta: 21/10/2012]

- [75] Konstantin Okonechnikov, Olga Golosova, Mikhail Fursov, UGENE team. Unipro UGENE: a unified bioinformatics toolkit. Oxford Journal, 2012.
- [76] APBS [Web en línea]. < > <http://www.poissonboltzmann.org/apbs>>. [Consulta: 21/10/2012]
- [77] Davis IW, Murray LW, Richardson JS, Richardson DC. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. 2004 Jul 1;32.
- [78] PDB2PQR [Web en línea]. < > <http://www.poissonboltzmann.org/pdb2pqr> >. [Consulta: 21/10/2012]



APÉNDICES



Apéndice 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           = Dinamica Molecular NVT
cpp             = /lib/cpp
include        = -I../top
integrator     = md
dt             = 0.00125
nsteps        = 800000
nstxout       = 2000
nstvout       = 2000
nstlog        = 2000
nstenergy     = 10
nstxtcout    = 10
xtc_grps      =
energygrps    =
nstlist       = 10
ns_type       = grid
fourierspacing = 0.30
pme_order     = 4
ewald_rtol    = 1e-5
pbc           = xyz
tcoupl        = nose-hoover
tc-grps      = protein
tau_t        = 0.5
ref_t         = 291.15
gen_vel       = yes
gen_temp      = 291.15
gen_seed      = 173529
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 = 80
gb_dielectric_offset = 0.09
sa_surface_tension = 2.05016
```

Figura A.2: Archivo *nvt.mpd* con integrador *MD*

Apéndice B

Ecuaciones Utilizadas

Análisis de la Estructura Global de Proteínas en Dinámica Molecular. Manuel de GROMACS (8).

RMSD

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N_{atoms}} d_i^2}{N_{atoms}}}$$

Donde;

$$d_i = \sqrt{\langle (r_f - r_0)^2 \rangle}$$

RMSF

$$RMSF_i = \sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2}$$

Apéndice C

Secuencias FASTA

```
>gi|189491748|ref|NP_001121636.1| ataxin-1 [Homo sapiens]
MKSQNQERSNECLPPKREIPATSRSSSEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGRRHGPAGTSVELG
LQQGIGLHKALSTGLDYSPPSAPRSVPVATTLPAAYATPQPGTPVSPVQYAHLPHTFQFIGSSQYSGTYA
SFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEAYSTLLANMGSLSQTPGHKAEQQQQQQQQQQQQHQ
HQQQQQQQQQQQQQHLSRAPGLITPGSPPPAQQNQYVHISSSPQNTGRTASPPAIPVHLHPHQTMIPHT
LTLGPPSQVVMQYADSGSHFVPREATKKAESSRLQQAIOAKEVLNGEMEKSSRYGAPSSADLGLGKAGGK
SVPHPYESRHVVVHPSVSDYSSRDPSGVRASVMVLPNSNTPAADLEVQQATHREASPSTLNDKSGHLGK
PGHRSYALSPHTVIQTTHSASEPLPVGLPATAFYAGTQPPVIGYLSGQQQAITYAGSLPQHLVIPGTQPL
LIPVGSTDMEASGAAPAIVTSSPQFAAVPHTFVTTALPKSENFNPEALVTQAAYPAMVQAQIHLPPVQSV
ASPAAPPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDLKIDSSTVERIEDSHSPGVAVIQ
FAVGEHRAQVSVEVLVEYPPFFVFGQGWSSCCPERTSOLFDLPCSKLSVGDVCISLTLKNLKNKNGSVKKGQP
VDPASVLLKHSKADGLAGSRHRYAEQENGINQGSQAQMLSENGELKFPEKMGLPAAPFLTKIEPSKPAATR
KRRWSAPESRKLEKSEDEPPLTLPKPSLIPQEVKICIEGRSNVGK
```

Figura C.1 Secuencia en formato FASTA de la proteína ataxin-1

```
>gi|189491748|ref|NP_001121636.1| ataxin-1 mutaded [Homo sapiens]
MKSQNQERSNECLPPKREIPATSRSSSEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGRRHGPAGTSVELG
LQQGIGLHKALSTGLDYSPPSAPRSVPVATTLPAAYATPQPGTPVSPVQYAHLPHTFQFIGSSQYSGTYA
SFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEAYSTLLANMGSLSQTPGHKAEQQQQQQQQQQQQQQ
QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHLSRAPGLITPGSPPPAQQNQYVHISSSPQNTG
RTASPPAIPVHLHPHQTMIPHTLTLGPPSQVVMQYADSGSHFVPREATKKAESSRLQQAIOAKEVLNGEM
EKSSRYGAPSSADLGLGKAGGKSVPHPYESRHVVVHPSVSDYSSRDPSGVRASVMVLPNSNTPAADLEVQ
QATHREASPSTLNDKSGHLGKPGHRSYALSPHTVIQTTHSASEPLPVGLPATAFYAGTQPPVIGYLSGQ
QQAITYAGSLPQHLVIPGTQPLLIPVGSTDMEASGAAPAIVTSSPQFAAVPHTFVTTALPKSENFNPEAL
VTQAAYPAMVQAQIHLPPVQSVASPAAPPTLPPYFMKGSIIQLANGELKKVEDLKTEDFIQSAEISNDL
KIDSSTVERIEDSHSPGVAVIQFAVGEHRAQVSVEVLVEYPPFFVFGQGWSSCCPERTSOLFDLPCSKLSV
GDVCISLTLKNLKNKNGSVKKGQVDPASVLLKHSKADGLAGSRHRYAEQENGINQGSQAQMLSENGELKFPE
```

Figura C.2 Secuencia en formato FASTA de la proteína ataxin-1 mutada

Apéndice D

Comandos utilizados en GROMACS

Minimización

1. De.pdb a .gro y topologías `pdb2gmx -ignh -f .pdb -p .top -o -gro -i .itp`
2. Colocar caja `editconf -nice 0 -box l a a -c -f .gro -o .gro`
3. Expansión de la topología `grompp -f .mdp -c .gro -p .top -o .tpr`
4. Minimización `mdrun -nice 0 -v -s .tpr -o .trr-c .gro -e .edr`

Dinámica

1. Colocar caja `editconf -nice 0 -box l a a -c -f .gro -o .gro`
2. Generación de topologías `pdb2gmx -ignh -f .gro -p .top -o -gro -i .itp`
3. Expansión de topología `grompp -f .mdp -c .gro -p .top -o .tpr`
4. Dinámica Molecular `mdrun -nice 0 -v -s .tpr -o .trr-x .xtc -c .gro -e .edr -cpo .cpt`

Apéndice E

Diagrama de trabajo

