



Universidad Católica de Santa María
Facultad de Ciencias Farmacéuticas, Bioquímicas y Biotecnológicas
Escuela Profesional de Ingeniería Biotecnológica

Informe de Prácticas Pre-Profesionales
Centro de Investigación en Ingeniería Molecular

Quispe Ppacco, David Jonatán
Arequipa, 2023

Índice

1. Introducción

Según la National Human Genome Research Institute, la bioinformática es un vínculo que se encontró entre el campo de la biología y las ciencias informáticas con la capacidad de almacenar, analizar y recopilar datos, este campo está muy relacionado a la genética y genómica que se basa en la manipulación y análisis de datos para obtener como resultado el entendimiento de la fisiología celular cuando ocurre un cambio dentro de su entorno.



Comúnmente este campo nos enlaza con el análisis de secuencias y se aplica a los genes, ADN, ARN o proteínas que resultan especialmente útiles para optimizar los experimentos en muchos ámbitos pueden hacer fricción con la divulgación científica.

El Centro de Investigación en Ingeniería Molecular de la Universidad Católica de Santa María en la dirección del PhD Badhin Gómez Valdes presenta una de las mejores oportunidades para el inicio dentro de la comprensión estructural de la biología computacional a cargo de equipo multidisciplinario ha logrado muchos aportes en el campo de la comprensión de la fisiología celular a través de análisis informáticos.

El objetivo de este informe es presentar la estadía como practicante dentro de el Centro de Investigación en Ingeniería Molecular (CIIM), mostrando el avance de diversas actividades y conocimientos adquiridos hasta el momento de la presentación.

2. Reseña de Actividades

En este apartado mostramos el proceso, con aciertos y desaciertos encontrados en el desarrollo de las actividades realizadas en el CIIM. Cabe resaltar que las actividades realizadas en el centro de investigación postulan logros iniciales con proyección a que en los siguientes informes podamos ver logros realizados completados al igual que habilidades y destrezas desarrolladas:

2.1. Actividad 1: Inducción a la informática

2.1.1. Objetivos

1. Comprender la navegación en sistemas Linux
2. Aplicar comandos usados en softwares sin interfaz gráfica (Gromacs)

2.1.2. Justificación

Para poder iniciar en el mundo de la bioinformática es básico conocer las características del software con el que contamos para poder hacer los cálculos o como coloquialmente se le llama

las corridas de los diferentes métodos como lo es la dinámica molecular que lo explicaremos posteriormente.

Para este proceso es necesario comprender el sistema operativo "Ubuntu" que se usa en el CIIM, el sistema operativo que se usa es una "distro" de Linux que nos facilita la personalización de prioridades en cuanto a procesos ya que si usamos windows este sistema operativo al enfocarse en el confort del usuario genera procesos en segundo plano que no hacen mas que ralentizar los cálculos del análisis, pero allí viene la primera dificultad, el sistema Linux nos da mayor control de procesos al costo de sacrificar el confort.

Linux es un sistema operativo que usa un tipo de lenguaje de programación incluido que es el lenguaje "Bash", es este lenguaje lo que hará la magia posteriormente para el análisis de datos ya que nos ofrece mayor automatización de procesos como lo es la comunicación directa con el programa de dinámica molecular Gromacs, que es un software que se encarga de los cálculos de dinámica molecular.

2.1.3. Metodología

Para aprender este lenguaje se necesito de recurrir a los diferentes fuentes de información que nos puede ofrecer la nube, ya que en ella podemos encontrar los comandos más requeridos o básicos para realizar una buena ejecución en los terminales de Linux lo cual es importante para entender como funcionan los comandos para realizar una dinámica molecular o usar otras herramientas del software Gromacs, del que posteriormente se hablará.

2.1.4. Resultados esperados

Se pudo comprender la importancia de aplicar este lenguaje debido a su amplia versatilidad para interactuar tanto con la cpu como con los diversos softwares usados para la aplicación de el análisis de datos.

Se espera no solo aprender un lenguaje de programación sino también aprender diversos debido a la alta demanda que tienen para optimizar sistemas de análisis de datos genómicos, en la actualidad contamos con múltiples herramientas de análisis de datos como lo es la herramienta Pymol que trabaja principalmente con lenguaje Python3 también tenemos a FastQC que puede combinar Python3 con lenguaje R mientras y también tenemos a Gromacs cuya mayor afinidad se le ha visto con el lenguaje Bash ya que es más intuitivo y fácil de aprender.

La optimización constante de cada uno de los softwares que usamos para el análisis de datos nos impulsará a tener resultados más instantáneos y poder comprender los diversos mecanismos que se encuentran en el organismo a través de softwares informáticos, es por ello que presenta una alta importancia de que una persona que desee iniciar en el campo de la bioinformática tenga conocimiento de como es que las computadoras gestionan los datos ingresados.

2.1.5. Evidencias

Se adjuntan el desarrollo de aprendizaje del lenguaje de programación: ??

```
#!/bin/bash

prot=laki

#Cleaning the water
grep -v 'ANISOU\|HOM\|HET\|CONNECT' $(prot).pdb > $(prot)_clean.pdb
mv $(prot)_clean.pdb > $(prot).pdb

#Structure
gmx pdb2gmx -f $(prot)_clean.pdb -o $(prot)_processed.gro -p $(prot).top -l $(prot).itp -
water spce -ff oplsa
gmx editconf -f $(prot)_processed.gro -o $(prot)_newbox.gro -c -d 1.0 -bt cubic
gmx solvate -cp $(prot)_newbox.gro -cs spc216.gro -o $(prot)_solv.gro -p $(prot).top
gmx grompp -f min.mdp -c $(prot)_solv.gro -p $(prot).top -o ions.tpr -po ions.mdp -maxwarn 1
echo 13 | gmx genion -s ions.tpr -o $(prot)_solv_ions.gro -p $(prot).top -pname NA -nname CL
-neutral -conc 0.154004106
gmx grompp -f min.mdp -c $(prot)_solv_ions.gro -p $(prot).top -o en.tpr -po min.mdp -maxwarn
1
gmx mdrun -nice 0 -v -s -pin on -pinoffset 0 -deffn en -nt 32 |& tee "minmdrun.log"
echo 10 0 | gmx energy -f en.edr -o potential.xvg
gmx grompp -f nvt.mdp -c en.gro -p $(prot).top -o nvt.tpr -po nvt1.mdp -maxwarn 1
gmx mdrun -nice 0 -v -s -pin on -pinoffset 0 -deffn nvt -nt 32 |& tee "nvtmdrun.log"
echo 16 0 | gmx energy -f nvt.edr -o temperature.xvg
gmx grompp -f npt.mdp -c nvt.gro -p $(prot).top -o npt.tpr -po npt1.mdp -maxwarn 1
gmx mdrun -nice 0 -v -s -pin on -pinoffset 0 -deffn npt -nt 32 |& tee "nptmdrun.log"
echo 18 0 | gmx energy -f npt.edr -o pressure.xvg
echo 24 0 | gmx energy -f npt.edr -o density.xvg
gmx grompp -f nvt.mdp -c npt.gro -f nvt.gro -t npt.cpt -p $(prot).top -o md_0_1.tpr -po
nvt1.mdp -maxwarn 1
gmx mdrun -nice 0 -v -s -deffn md_0_1 -c nvt.gro -nt 32 -nb gpu |& tee "nvt_drun.log"
echo 1 0 | gmx trjconv -s md_0_1.tpr -f md_0_1.tpr -o md_0_1n0PBC.xtc -pbc mol -center
echo 4 4 | gmx rms -s md_0_1.tpr -f md_0_1n0PBC.xtc -o rmsd.xvg -tu ns
echo 4 4 | gmx rms -s en.tpr -f md_0_1n0PBC.xtc -o rmsd_xtal.xvg -tu ns
echo 11 | gmx gyrate -s md_0_1.tpr -f md_0_1n0PBC.xtc -o gyrate.xvg

# Create a .gro file with the box parameters
gmx editconf -f pre-$(prot).pdb2gmx.gro -o min-$(prot).gro -resnr 1 -c -d 1.0 -bt cubic

# Add the necessary H2O molecules to the system (solvate)
gmx solvate -cp min-$(prot).gro -cs spc216.gro -o pre-$(prot)_solv.gro -p $(prot).top

# Preprocessing to generate the necessary files for the ion generation
gmx grompp -f min.mdp -c pre-$(prot)_solv.gro -p $(prot).top -o pre-$(prot)_ions.tpr -po
pre-$(prot)_ions.mdp -maxwarn 1

# Add ions to neutralize the system replacing some H2O
echo 13 |
gmx genion -s pre-$(prot)_ions.tpr -o $(prot).gro -p $(prot).top -neutral -conc
0.154004106

}

# Function to execute the energy Minimization
function structureMinimization {

    maxizetefan

    # Preprocessing to generate the necessary files for the min-mdrun after the ions
    generation
}
```

Figura 1: Aplicación de lenguaje Bash

2.1.6. Conclusiones y recomendaciones

Concluimos en que el desarrollo de la ingeniería en bioinformática es un camino que requiere de muchos saberes previos debido a la complejidad que se tiene en el desarrollo, gestión y análisis de la data, dentro de la carrera de Ingeniería Biotecnológica, a la cual pertenezco, se evaluó la importancia de la bioinformática centrado al análisis de secuencias como son los alineamientos, incluyendo la búsqueda de soluciones a la transcriptómica, pero se recomienda ampliar el campo de visión para lo que es la comprensión de los diversos softwares y posiblemente el desarrollo de una nueva optimización de data genómica podría salir a relucir de parte de un ingeniero biotecnólogo.

2.2. Actividad 2: Inducción a la Dinámica Molecular

2.2.1. Objetivos

1. Aplicar los saberes previos de lectura Bash para interpretar los comandos en Gromacs.
2. Comprender la aplicación de los campos de fuerza usados en Gromacs.
3. Analizar los datos resultantes en el ámbito biológico y fisicoquímico.

2.2.2. Justificación

Para una persona que está inspirada en el descubrimiento del funcionamiento de estructuras como lo son las estructuras y maquinarias moleculares de la célula es necesario comprender que existe ciencia detrás del análisis de data de las secuencias y estructuras 3D encontradas en bases de datos como lo es PDB, NCBI, UnitProt, entre otras. Esta ciencia de manipulación de data se basa en el retorno de la estructura a su estado fisiológico.

Pero ¿Por qué transformar una estructura al estado fisiológico? Es algo que también me pregunté, ¿no se supone que la data encontrada ya está lista para usarse?, pues la respuesta corta es "no". La data encontrada en las bases de datos son aproximaciones o en forma mas sencilla de explicar, es una simple foto de un cristal. Dentro de un cristal todo es compacto es decir las moléculas están totalmente compactadas y es así como se encuentran en las estructuras 3D que normalmente vemos y es nuestro deber pasarlas a un estado estable fisiológico que es como comúnmente encontramos a las proteínas en nuestro organismo.

Para ello usamos las herramientas propuestas por un software llamado Gromacs que a través de diversos comandos nos brinda la ayuda para pasar de una molécula compactada a una molécula a la que se le puede captar en una vibración fisiológica ayudándonos a contemplar las características fisicoquímica que presenta estando se encuentra estable.

2.2.3. Metodología

Para superar las dificultades iniciales tuvimos que aprender a instalar el programa Gromacs desde el comienzo, para ello se realizó la aplicación de los diversos comandos anteriormente estudiados, Gromacs es un software que presenta muchas herramientas una de ellas es la fusión de realizar dinámica molecular la cual está dividida en limpieza de estructuras, opciones de capos de fuerza (las más usadas son el campo OPLS no solo empleado en campos proteicos, sino también en ARN), topología, estabilización de la estructura por minimización de energía ??, control de temperatura y presión y finalmente el análisis de resultados ??.

Las formas en las que empezamos a realizar es de manera "steps." paso a paso, como también existe la manera automatizada la cual se tiene que realizar desde un script de bash que gracias a la automatización por uso de sus comandos permite realizar los pasos en uno solo.

2.2.4. Resultados y discusión

Gromacs es un software complejo que nos ayuda a poder encontrar la estabilidad fisiológica de las moléculas y encontrar las características de la dinámica molecular de una proteína y es elegido por muchos por la gran variedad de herramientas que nos ofrece como lo es correr dinámicas moleculares con ligando o mantener encontrar interacciones de dominios transmembranales transmembranales, pero por el momento podemos afirmar que se comprendió el

funcionamiento de la herramienta de dinámica molecular.

Tomamos en cuenta que Gromacs no es la única herramienta que existe pero es la mejor para hacer simulaciones ya que los cálculos que nos brinda son más exactos debido a los campos de fuerza que usa.

El objetivo al que se quiere llegar en un principio es llegar a la minimización y para ello se obtienen gráficos donde se calcula la minimización de energía para eliminar las fuerzas no físicas presentes en la estructura inicial, ya sea por errores en la construcción del sistema o por interacciones atómicas desfavorables.

Dentro del curso de Termodinámica que se encuentra en la curricula de la escuela de Ingeniería Biotecnológica nos demostrar las leyes por las cuales se rige la materia y energía, entre ellas se sabe de la espontaneidad es decir que al minimizar la energía potencial se favorecen estados que tendrán una menor energía libre de Gibbs y serán más estables. Y es lo que se requiere para analizar una proteína en un estado fisiológico estable. ??

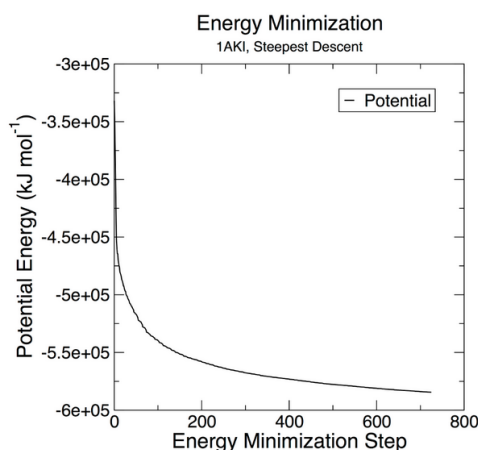
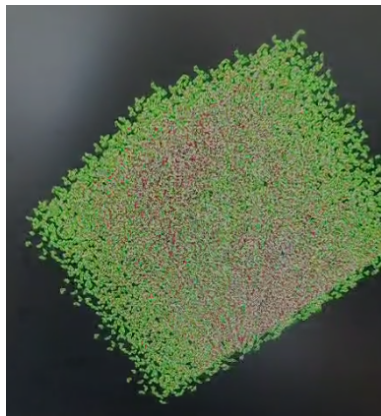


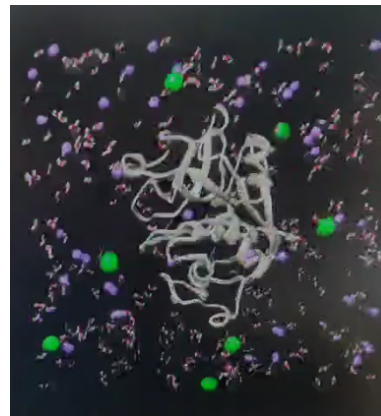
Figura 2: Gráfica de minimización extraída de Gromacs tutorial

Posterior a ello el segundo objetivo es el análisis de la preparación de dinámica gracias a los archivos estándar que nos brinda el mismo Gromacs estos archivos presentan información como el tiempo requerido de dinámica las propiedades del capo de fuerza la administración de la temperatura, presión. dependido de ello tenemos los archivos NVE / NVT / NPT. Donde podemos dilucidar el significado de cada uno de los archivos ya que **N** es número de átomos, **T** Temperatura, **E** Energía, **P** Presión y **V** volumen, cada uno de estos pasos son importantes para iniciar con la dinámica molecular.

Usamos este tipo de archivos para desarrollar la dinámica molecular que nos mostrará el comportamiento de los átomos y la macromolécula generando una película y dándonos a conocer las gráficas.



(a) Caja de agua previa



(b) Film posterior a la dinámica

Figura 3: Desarrollo de la dinámica molecular con Gromacs

Finalmente se obtuvo la gráfica donde podemos ver la diferencia entre las métrica de la estructura cristalina que se tenía al comienzo. A esta métrica se le conoce como RMSD o (Root Mean Square Deviation). ??

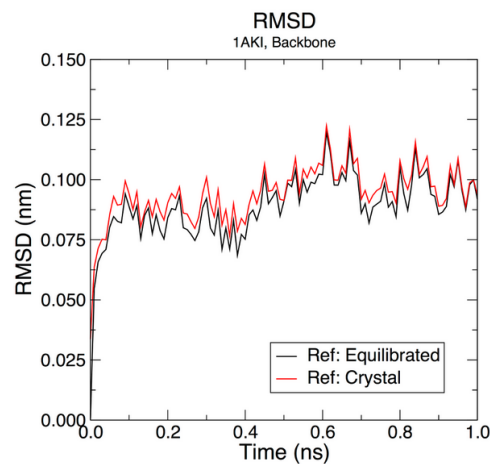


Figura 4: Gráfica RMSD en comparación cristal y dinámica molecular

Ya para culminar nos damos cuenta que es nuestra estructura proteica está completamente lista para usar las diversas herramientas posteriores como lo son la localización en membrana si lo requiere o si se necesita o un docker simple al igual que un Screen masivo dependiendo de los experimentos que se requieren.

2.2.5. Conclusiones y recomendaciones

Podemos concluir que es necesario comprender no solo los comandos de mencionado software como lo es Gromacs, sino también cuales son los resultados y como llegó a ese objetivo.

2.3. Actividad 3: Planteamiento de un proyecto de Investigación

2.3.1. Objetivos

1. Analizar la bibliografía inicial de manera exhaustiva
- 2.

2.3.2. Justificación

2.3.3. Metodología

2.3.4. Resultados y discusión

2.3.5. Evidencias

2.3.6. Conclusiones y recomendaciones

2.4. Actividad 4: Asistencia de investigación para estudiantes de intercambio

2.4.1. Objetivos

2.4.2. Justificación

2.4.3. Metodología

2.4.4. Resultados y discusión

2.4.5. Evidencias

2.4.6. Conclusiones y recomendaciones

3. Discusión

En el campo de la biotecnología podemos encontrar varias ramas en diferentes disciplinas como la Biotecnología ambiental, industrial, entre otras. Una de las disciplinas actuales que es muy requerida por la facilidad que tiene en internarse dentro de experimentos de alta complejidad, la comprensión de sistemas celulares completos, la bioquímica de las proteínas, la

regulación de la expresión de genes, así como la ayuda en la optimización de terapias contra diferentes patologías, es la bioinformática.

Existe muchas discusiones sobre el tema de la derivación de la bioinformática a partir de la biología computacional, o por lo contrario se afirma de una gran diferencia entre ambas disciplinas. Personalmente, mi opinión es que la bioinformática es muy requerida por la gran comprensión que puede darnos a conocer a partir del análisis de datos informáticos y por ello se requiere la comprensión del sistema completo celular y para que eso pueda suceder es necesario el conocimiento de la biología computacional que nos aporta las herramientas para simular las estructuras celulares con la alta complejidad fisicoquímica y matemática que tiene cada estructura proteica y sus diferentes alteraciones o interacciones.

Son estas inquietudes que demuestran que un internamiento a la bioinformática debe empezar por una comprensión completa de las interacciones y comportamientos de diferentes genes del organismo, es decir, la parte de la biología computacional o como otras personas lo llaman "la parte teórica" que abarca la proteómica, transcriptómica, metabolómica, entre otras disciplinas adscritas al grupo de las omicas por ende interesadas en dilucidar los comportamientos moleculares.

Dentro de la Universidad Católica de Santa María existe un centro especializado en la comprensión de diferentes estructuras celulares como lo es el Centro de Investigación en Ingeniería Molecular (CIIM) a cargo del doctor "Badhin Gomez Valdezrenombrado ponente especializado en química computacional que se encargó de investigar las interacciones y comportamientos de diferentes estructuras proteicas con diversos ligandos, proteínas o comprender las patologías desde un punto de vista informático.

Es allí donde se inició la aventura de mis prácticas pre-profesionales con el objetivo expreso de aprender todo lo necesario para comprender la fisiología celular iniciando con un adiestramiento informático, pasando por un proceso de comprensión y análisis bibliográfico y culminando con la comprensión del uso de diferentes herramientas como lo es el alineamiento de secuencias, interacciones proteína-proteína, proteína-ligando para culminar con un entendimiento cuántico completo sobre los comportamientos que pueden tener diferentes proteínas a cambios dentro de su entorno ya sea por un fármaco proteína o alteraciones en su vía de señalización.

El objetivo de este informe es narrar las actividades realizadas y logros encontrados en estas primeras 5 semanas como practicante pre-profesional en el Centro de Investigación en Ingeniería Molecular (CIIM).

4. Evidencias

En esta sección, se adjuntan documentos y evidencias relevantes que respalden las actividades y logros mencionados en la sección anterior. Esto puede incluir informes técnicos, presentaciones, fotografías, resultados de experimentos, tablas, gráficas, entre otros.

5. Conclusiones y Recomendaciones

6. Anexos

Agregue algunos anexos de ser necesario.