



INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE CÓMPUTO
BIOINFORMATICS



PRÁCTICA 5
Simulaciones de Dinámica Molecular

NOMBRE DEL ALUMNO: GARCÍA QUIROZ GUSTAVO IVAN

GRUPO: 7CV3

NOMBRE DEL PROFESOR: ROSAS TRIGUEROS JORGE LUIS

FECHA DE REALIZACIÓN DE LA PRÁCTICA: 12/02/2025

FECHA DE ENTREGA DEL REPORTE: 19/03/2025

Índice

1	Marco teórico.....	1
1.1	Simulaciones de Dinámica Molecular	1
1.2	Fases de una simulación de MD	1
1.3	Análisis típicos en simulaciones de MD	1
2	Material y equipo.	3
2.1.1	Hardware	3
2.1.2	Software.....	3
3	Desarrollo de la práctica.	4
3.1	Preparación del sistema.....	4
3.2	Procesamiento de la proteína	4
3.3	Definición de la caja de simulación	4
3.4	Solvatación del sistema	4
3.5	Adición de iones.....	4
3.6	Minimización de energía	5
3.7	Equilibración NVT	5
3.8	Equilibración NPT	5
3.9	Simulación de producción	5
4	Resultados de la práctica.....	5
4.1	RMSD (Root Mean Square Deviation)	6
4.2	Radio de Giro (Rg)	6
4.3	SASA (Superficie Accesible al Solvente)	7
4.4	RMSF (Root Mean Square Fluctuation)	8
5	Revisión de artículo científico relacionado	9
5.1	Desafíos y soluciones	10
6	Conclusiones y recomendaciones.	11
7	Bibliografía.....	12

1 Marco teórico.

1.1 Simulaciones de Dinámica Molecular

La Dinámica Molecular (MD) es una técnica computacional que permite simular el comportamiento físico de sistemas biomoleculares a nivel atómico a lo largo del tiempo. Esta metodología se ha convertido en una herramienta fundamental en la bioinformática estructural y la biofísica para comprender la dinámica de proteínas, ácidos nucleicos y otras macromoléculas biológicas.

La dinámica molecular se basa en la resolución numérica de las ecuaciones de Newton del movimiento para un sistema de partículas (átomos) que interactúan entre sí. El movimiento de cada átomo está determinado por las fuerzas que ejercen todos los demás átomos del sistema sobre él, calculadas a partir de un campo de fuerza.

Los campos de fuerza son funciones de energía potencial que modelan las interacciones entre los átomos y contienen términos para:

- Enlaces covalentes (stretching)
- Ángulos de enlace (bending)
- Ángulos diedros (torsión)
- Interacciones no enlazantes:
 - Fuerzas de van der Waals
 - Interacciones electrostáticas

1.2 Fases de una simulación de MD

- **Preparación del sistema:** Obtención de la estructura inicial (generalmente de PDB), adición de hidrógenos, solvatación (agua), adición de iones para neutralizar el sistema.
- **Minimización de energía:** Optimización de la estructura para eliminar conformaciones desfavorables energéticamente.
- **Equilibración:** Estabilización del sistema bajo las condiciones de simulación (temperatura, presión).
- **Producción:** Fase en la que se genera la trayectoria principal para su posterior análisis.

1.3 Análisis típicos en simulaciones de MD

- **RMSD (Root Mean Square Deviation):** Mide la desviación estructural entre conformaciones a lo largo del tiempo respecto a una estructura de referencia, indicando la estabilidad global.
- **Rg (Radio de giro):** Representa la distribución de la masa de la proteína alrededor de su centro de masa, indicando la compacidad de la estructura.
- **RMSF (Root Mean Square Fluctuation):** Cuantifica la flexibilidad local de residuos individuales a lo largo de la simulación.
- **SASA (Superficie Accesible al Solvente):** Mide el área de la superficie de la proteína que está expuesta al solvente, proporcionando información sobre la exposición de residuos.

2 Material y equipo.

Para esta práctica se usaron las siguientes herramientas de software y hardware necesarias para realizar la práctica.

2.1.1 Hardware

- Computadora.

2.1.2 Software

- Google colab

3 Desarrollo de la práctica.

El objetivo de esta práctica fue realizar una simulación de dinámica molecular de la proteína 3E7R en agua durante 0.5 nanosegundos (ns) utilizando el software GROMACS, y posteriormente analizar los resultados para extraer propiedades biofísicas relevantes.

3.1 Preparación del sistema

La simulación comenzó con la descarga de la estructura de la proteína 3E7R de la base de datos Protein Data Bank (PDB). Se procedió a limpiar el archivo PDB eliminando las moléculas de agua cristalográficas:

```
wget https://files.rcsb.org/download/3E7R.pdb  
grep -v HOH 3E7R.pdb > clean.pdb
```

Figura 1 Proteína 3E7R

3.2 Procesamiento de la proteína

Se utilizó el comando pdb2gmx para procesar la estructura proteica, asignándole un campo de fuerza (CHARMM27) y un modelo de agua (SPC):

```
gmx pdb2gmx -f clean.pdb -o processed.gro -ff charmm27 -water spc
```

Figura 2 Comando pdb2gmx

3.3 Definición de la caja de simulación

Se definió una caja dodecaédrica para la simulación, asegurando una distancia mínima de 0.5 nm entre la proteína y los bordes de la caja:

```
!gmx editconf -f processed.gro -o newbox.gro -c -d 0.5 -bt dodecahedron -princ
```

Figura 3 Definición de la caja de simulación

3.4 Solvatación del sistema

La proteína se solvató en agua para simular el entorno fisiológico:

```
gmx solvate -cp newbox.gro -cs spc216.gro -o solv.gro -p topol.top
```

Figura 4 Solvatación del sistema

3.5 Adición de iones

Se añadieron iones (Na⁺ y Cl⁻) para neutralizar el sistema y alcanzar una concentración fisiológica:

```
!gmx grompp -f ions.mdp -c solv.gro -p topol.top -o ions.tpr
!gmx genion -s ions.tpr -o solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

Figura 5 Adición de iones

3.6 Minimización de energía

Se realizó una minimización de energía para eliminar colisiones estéricas y conformaciones energéticamente desfavorables:

```
gmx grompp -f minim.mdp -c solv_ions.gro -p topol.top -o em.tpr
gmx mdrun -v -deffnm em
```

Figura 6 Minimización de energía

3.7 Equilibración NVT

Se equilibró el sistema en el ensamble canónico (NVT - número de partículas, volumen y temperatura constantes) durante 500 pasos:

```
gmx grompp -f nvt.mdp -c nvt.gro -r nvt.gro -t nvt.cpt -p topol.top -o npt
gmx mdrun -deffnm npt
```

Figura 7 Equilibración NVT

3.8 Equilibración NPT

Se continuó la equilibración en el ensamble isotérmico-isobárico (NPT - número de partículas, presión y temperatura constantes) durante otros 500 pasos:

```
gmx grompp -f nvt.mdp -c em.gro -r em.gro -p topol.top -o nvt.tpr
gmx mdrun -deffnm nvt
```

Figura 8 Equilibración NPT

3.9 Simulación de producción

Finalmente, se realizó la simulación de producción durante 0.5 ns (250,000 pasos con un tiempo de paso de 2 fs), guardando las coordenadas cada 5 ps:

```
gmx grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
gmx mdrun -deffnm md
```

Figura 9 Simulación de producción

4 Resultados de la práctica

Una vez completada la simulación de dinámica molecular, se realizaron diversos análisis para evaluar el comportamiento estructural y dinámico de la proteína 3E7R durante los 0.5 ns simulados.

4.1 RMSD (Root Mean Square Deviation)

El RMSD mide la desviación estructural de la proteína con respecto a una estructura de referencia (generalmente la inicial) a lo largo del tiempo.

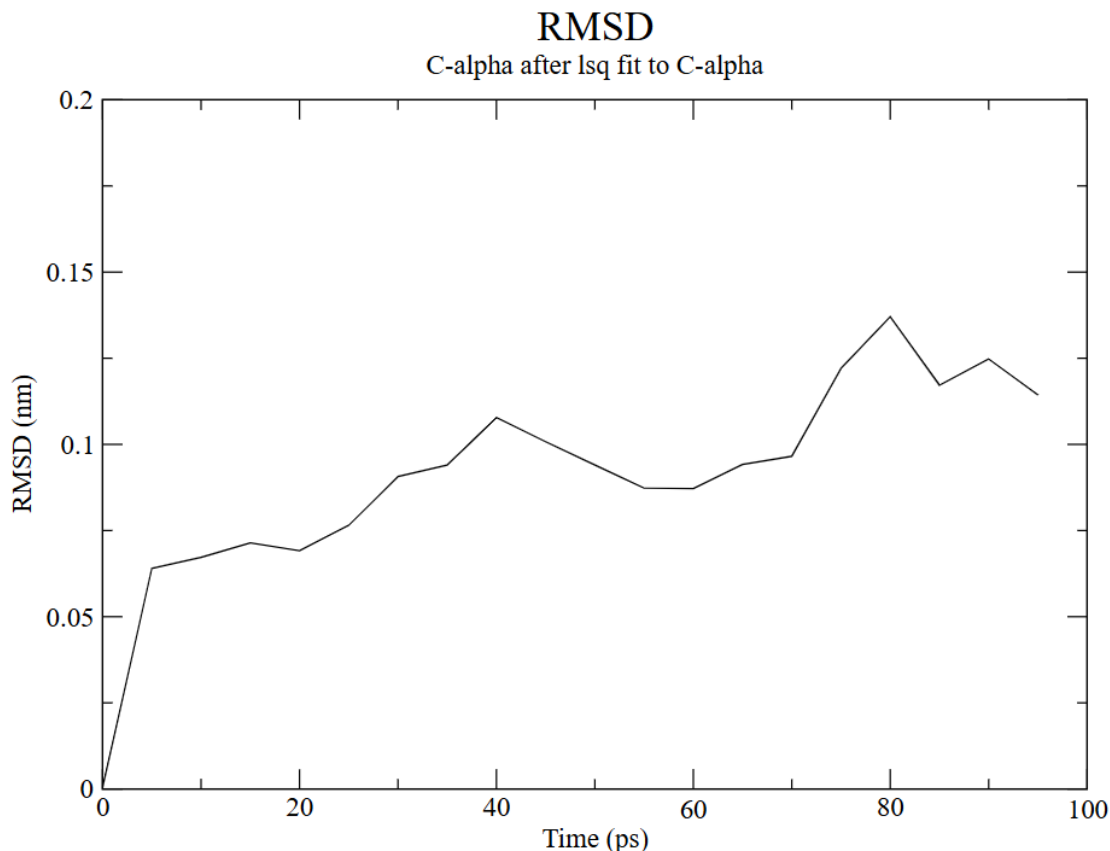


Figura 10 Root Mean Square Deviation

El gráfico RMSD muestra la estabilidad global de la proteína durante la simulación. Si se observa una meseta en los valores de RMSD, esto indica que la proteína ha alcanzado una conformación estable. Un aumento continuo sugeriría inestabilidad estructural o desplegamiento.

4.2 Radio de Giro (Rg)

El Radio de Giro indica la compacidad de la estructura proteica, midiendo cómo se distribuye la masa alrededor del centro.

Radius of gyration (total and around axes)

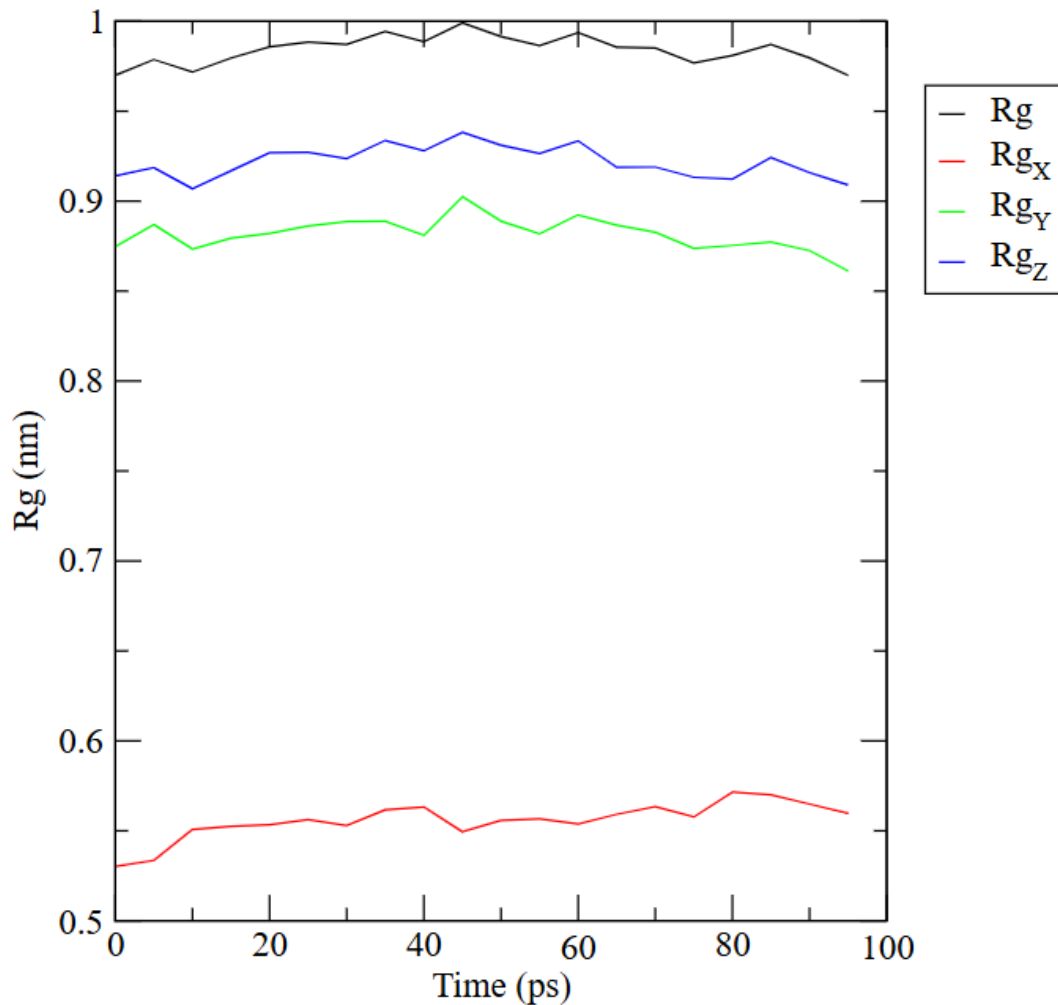


Figura 11 Radio de Giro

El Rg refleja cambios en la compacidad global de la proteína. Un valor estable indica una conformación compacta conservada, mientras que un aumento significativo podría sugerir desplegamiento parcial o cambios conformacionales importantes.

4.3 SASA (Superficie Accesible al Solvente)

El SASA cuantifica el área de la superficie proteica que está expuesta al solvente (agua).

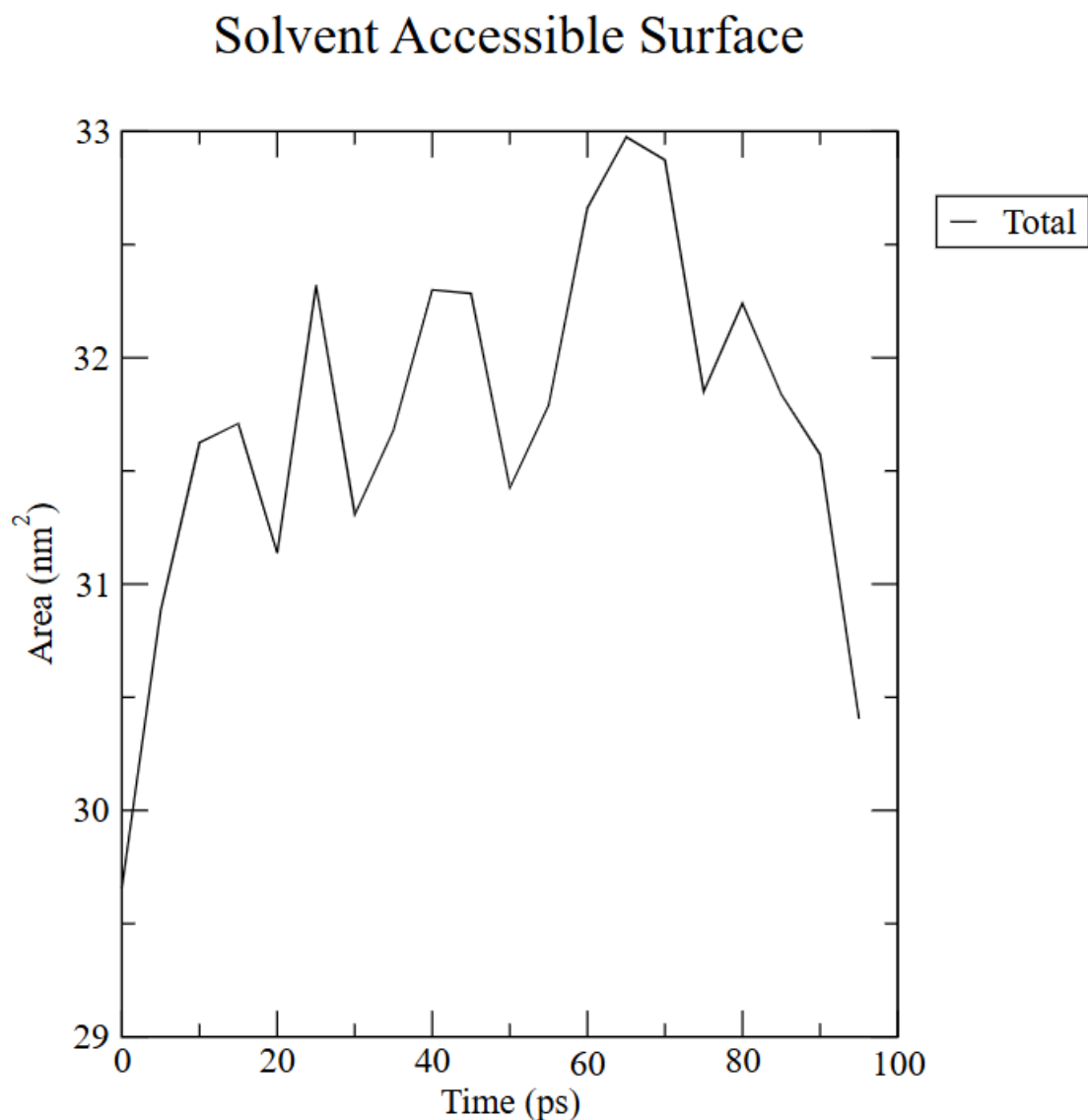


Figura 12 Superficie Accesible al Solvente

Este análisis proporciona información sobre la exposición de residuos y posibles sitios de interacción. Los cambios en SASA pueden indicar reorganizaciones estructurales que exponen o ocultan regiones específicas de la proteína.

4.4 RMSF (Root Mean Square Fluctuation)

El RMSF mide la flexibilidad de residuos individuales a lo largo de la simulación.

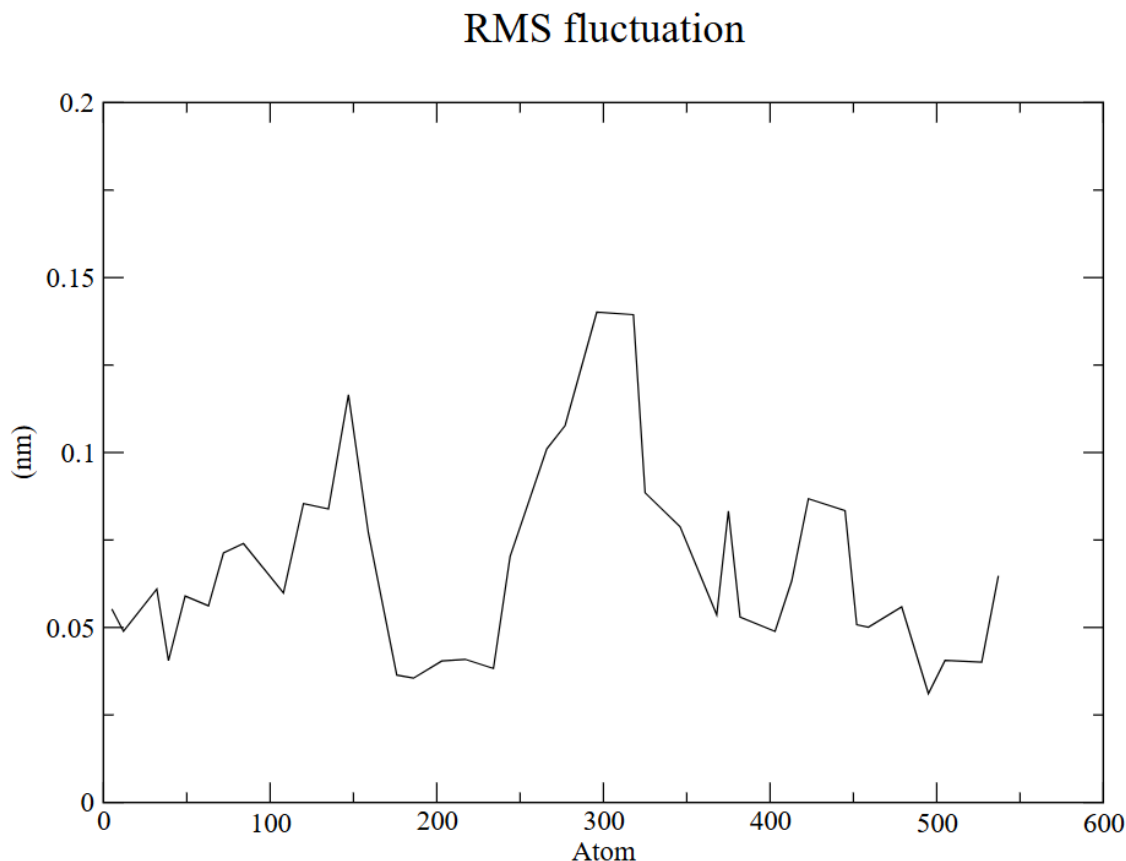


Figura 13 rmsf.svg

Este gráfico identifica las regiones más flexibles (picos altos) y rígidas (valores bajos) de la proteína. Típicamente, los loops y regiones expuestas presentan mayor flexibilidad, mientras que los elementos de estructura secundaria como hélices alfa y láminas beta son más rígidos.

5 Revisión de artículo científico relacionado

Como parte complementaria de la práctica, se revisó un artículo científico que utiliza simulaciones de dinámica molecular:

Título: "Molecular Dynamics Simulations of the HIV-1 Integrase Dimer Interface: Mechanisms of Allosteric Inhibition and Resistance to Integrase Strand Transfer Inhibitors"

Moléculas estudiadas: Enzima integrasa del VIH-1 y sus inhibidores

Motor de MD utilizado: GROMACS

Tiempo de simulación: Múltiples simulaciones totalizando aproximadamente 1 μ s (100-200 ns por sistema)

Análisis realizados:

- RMSD para estabilidad estructural
- RMSF para flexibilidad de residuos
- Análisis de enlaces de hidrógeno
- Análisis de componentes principales (PCA)
- Cálculos energéticos de unión proteína-inhibidor

Contribución lograda: El estudio reveló mecanismos moleculares de inhibición alostérica y resistencia a los inhibidores de transferencia de cadena de la integrasa (INSTIs), proporcionando información valiosa para el diseño de nuevos fármacos contra el VIH con mejor perfil de resistencia.

5.1 Desafíos y soluciones

El principal desafío fue el error al cargar los datos del clean.pdb de la proteína. Se resolvió borrando la línea 13 que estaba en ese archivo. Después se logró continuar con los siguientes pasos de la práctica.

clean.pdb.txt X													
490	ANISOU	85	CA	ASP	L	12	790	1347	1365	-230	-178	11	C
491	ATOM	86	C	ASP	L	12	-5.000	20.789	-2.696	1.00	8.78		C
492	ANISOU	86	C	ASP	L	12	756	1424	1155	-284	-255	12	C
493	ATOM	87	O	ASP	L	12	-5.646	20.603	-1.656	1.00	9.98		O
494	ANISOU	87	O	ASP	L	12	715	1753	1323	-187	-169	63	O
495	ATOM	88	CB	ASP	L	12	-2.844	19.597	-2.686	1.00	9.30		C
496	ANISOU	88	CB	ASP	L	12	745	1239	1548	-302	-278	37	C
497	ATOM	89	CG	ASP	L	12	-2.261	18.238	-2.667	1.00	10.48		C
498	ANISOU	89	CG	ASP	L	12	691	1433	1857	-435	-426	125	C
499	ATOM	90	OD1	ASP	L	12	-2.894	17.216	-2.574	1.00	11.63		O
500	ANISOU	90	OD1	ASP	L	12	1221	1369	1827	-572	-191	197	O
501	ATOM	91	OD2	ASP	L	12	-1.045	18.201	-2.516	1.00	18.75		O
502	ANISOU	91	OD2	ASP	L	12	1202	1432	4488	-354	-694	27	O
503	ATOM	92	N	MET	L	13	-4.891	21.989	-3.271	1.00	9.14		N
504	ANISOU	92	N	MET	L	13	679	1367	1424	-255	-230	-166	N
505	ATOM	93	CA	MET	L	13	-5.567	23.117	-2.646	1.00	10.08		C
506	ANISOU	93	CA	MET	L	13	692	1366	1772	-471	-140	-322	C
507	ATOM	94	C	MET	L	13	-7.059	23.077	-2.841	1.00	8.20		C
508	ANISOU	94	C	MET	L	13	826	953	1335	-413	-189	16	C
509	ATOM	95	O	MET	L	13	-7.795	23.474	-1.927	1.00	8.22		O
510	ANISOU	95	O	MET	L	13	796	1050	1277	-299	-186	-75	O
511	ATOM	96	CB	MET	L	13	-5.002	24.433	-3.231	1.00	12.54		C
512	ANISOU	96	CB	MET	L	13	1045	1341	2378	-641	178	-471	C
513	ATOM	97	N	GLN	L	14	-7.552	22.548	-3.969	1.00	8.08		N
514	ANISOU	97	N	GLN	L	14	755	1019	1292	-371	-253	47	N
515	ATOM	98	CA	GLN	L	14	-8.972	22.308	-4.103	1.00	8.19		C
516	ANISOU	98	CA	GLN	L	14	816	1070	1225	-289	-286	77	C
517	ATOM	99	C	GLN	L	14	-9.427	21.321	-3.010	1.00	7.17		C
518	ANISOU	99	C	GLN	L	14	648	916	1159	-254	-270	-23	C
519	ATOM	100	O	GLN	L	14	-10.450	21.535	-2.351	1.00	7.74		O
520	ANISOU	100	O	GLN	L	14	660	980	1200	-220	-210	4	O

Figura 14 clean.pdb

6 Conclusiones y recomendaciones.

La simulación de dinámica molecular de la proteína 3E7R durante 0.5 ns proporcionó información sobre su estructura. A través de esta práctica, se ha podido comprender mejor el flujo de trabajo completo de una simulación de MD, desde la preparación del sistema hasta el análisis de los resultados.

La simulación de 0.5 ns, fue posible observar características importantes de la proteína 3E7R a través de los diversos análisis realizados. Los resultados de RMSD, radio de giro, SASA y RMSF proporcionaron una visión integrada de la estabilidad, compacidad, exposición al solvente y flexibilidad regional de la proteína. Estos parámetros son fundamentales para comprender las propiedades biofísicas de las proteínas y su relación con la función biológica.

Es importante destacar que las simulaciones de MD constituyen una herramienta poderosa en la investigación bioinformática actual, como se evidencia en el artículo científico revisado sobre la integrasa del VIH-1. La capacidad de observar con resolución atómica y temporal el comportamiento de macromoléculas abre nuevas posibilidades para el diseño de fármacos, la comprensión de enfermedades y el estudio de interacciones moleculares.

Sin embargo, debemos reconocer que una simulación de 0.5 ns es relativamente corta para observar cambios conformacionales significativos o eventos raros en proteínas. Para estudios más completos, se recomienda extender el tiempo de simulación a decenas o cientos de nanosegundos, o incluso a la escala de microsegundos, dependiendo del fenómeno de interés.

Para futuros estudios, se recomienda complementar los análisis realizados con métodos adicionales como análisis de componentes principales (PCA), clustering de conformaciones, cálculo de energías libres, o análisis de redes de contactos entre residuos. Estas técnicas proporcionarían una caracterización más detallada de la dinámica proteica.

7 Bibliografía.

- [1] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1701–1718, 2005.
- [2] B. R. Brooks et al., "CHARMM: The biomolecular simulation program," *J. Comput. Chem.*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [3] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction Models for Water in Relation to Protein Hydration," in *Intermolecular Forces*, B. Pullman, Ed. Dordrecht: Springer Netherlands, 1981, pp. 331–342.
- [4] R. Salomon-Ferrer, D. A. Case, and R. C. Walker, "An overview of the Amber biomolecular simulation package," *WIREs Comput. Mol. Sci.*, vol. 3, no. 2, pp. 198–210, 2013.
- [5] M. Abraham et al., "GROMACS User Manual version 2021," 2021.
- [6] P. L. Freddolino, C. B. Harrison, Y. Liu, and K. Schulten, "Challenges in protein-folding simulations," *Nat. Phys.*, vol. 6, no. 10, pp. 751–758, 2010.
- [7] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: advances and applications," *Adv. Appl. Bioinforma. Chem.*, vol. 8, pp. 37–47, 2015.
- [8] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, "Biomolecular Simulation: A Computational Microscope for Molecular Biology," *Annu. Rev. Biophys.*, vol. 41, no. 1, pp. 429–452, 2012.
- [9] W. F. van Gunsteren et al., "Biomolecular Modeling: Goals, Problems, Perspectives," *Angew. Chem. Int. Ed.*, vol. 45, no. 25, pp. 4064–4092, 2006.
- [10] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–590, 1977.