



Instituto Politécnico Nacional
Escuela Superior de Cómputo



Bioinformatics

Lab Session 4. Molecular Dynamics Simulations

González Bocio Erik Alexander 2020630163

Jorge Luis Rosas Trigueros

Fecha de realización: 7 de marzo de 2022

Fecha de entrega: 28 de marzo de 2022

Marco Teórico:

Cuando alguien (desde personas que estudian la biología hasta alumnos) quiere saber cómo funciona una proteína u otra biomolécula se enfrenta a un gran desafío. Una estructura a nivel atómico es tremendamente útil y, por lo general, genera información sustancial sobre cómo funciona la biomolécula. Sin embargo, los átomos en una biomolécula están en constante movimiento, y tanto la función molecular como las interacciones intermoleculares dependen de la dinámica de las moléculas involucradas. A uno le gustaría no solo una imagen instantánea estática, sino la capacidad de observar estas biomoléculas en acción, perturbarlas a nivel atómico y ver cómo responden. Desafortunadamente, observar los movimientos de los átomos individuales y perturbarlos de la manera deseada es difícil. Una alternativa atractiva es trabajar con una simulación por computadora a nivel atómico de las biomoléculas relevantes.

Las simulaciones de dinámica molecular (MDS) predicen cómo se moverá cada átomo en una proteína u otro sistema molecular con el tiempo, basándose en un modelo general de la física que gobierna las interacciones interatómicas (Karplus y McCammon, 2002). Estas simulaciones pueden capturar una amplia variedad de procesos biomoleculares importantes, incluido el cambio conformacional, la unión al ligando y el plegamiento de proteínas, revelando las posiciones de todos los átomos a una resolución temporal de femtosegundos. Es importante destacar que tales simulaciones también pueden predecir cómo responderán las biomoléculas, a nivel atómico, a perturbaciones como la mutación, la fosforilación, la protonación o la adición o eliminación de un ligando. Las simulaciones de MD se utilizan a menudo en combinación con una amplia variedad de técnicas experimentales de biología estructural, incluida la cristalografía de rayos X, la microscopía crioelectrónica (crio-EM), la resonancia magnética nuclear (RMN), la resonancia paramagnética electrónica (EPR) y la transferencia de energía de resonancia de Förster (FRET)

Las simulaciones md no son nuevas. Las primeras simulaciones MD de gases simples se realizaron a finales de la década de 1950 (Alder y Wainwright, 1957). La primera simulación MD de una proteína se realizó a finales de la década de 1970 (McCammon et al., 1977), y el trabajo preliminar que permitió estas simulaciones fue uno de los logros reconocidos por el Premio Nobel de Química 2013 (Levitt y Lifson, 1969; Lifson y Warshel, 1968). Sin embargo, las simulaciones de MD se han vuelto sustancialmente más populares y visibles en los últimos años, particularmente desde la perspectiva de los biólogos moleculares experimentales. Las simulaciones han comenzado a aparecer con frecuencia en los trabajos experimentales de biología estructural, donde se utilizan tanto para interpretar los resultados experimentales como para guiar el trabajo experimental. Esta tendencia es particularmente notable en neurociencia: se han utilizado simulaciones para estudiar proteínas críticas para la señalización neuronal (Dawe et al., 2016; Delemotte et al., 2011; Dror et al., 2013; Jensen et al., 2012; Shi et al., 2008), para ayudar en el desarrollo de fármacos dirigidos al sistema nervioso (Manglik et al., 2016; McCorvy et al., 2018; Spahn et al., 2017), para revelar mecanismos de agregación de proteínas asociados con trastornos neurodegenerativos (Khandogin y Brooks, 2007; Wu y Shea, 2013), y proporcionar una base para el diseño de herramientas optogenéticas mejoradas (Takemoto et al., 2015; Kato et al., 2018).

La simulación de dinámica molecular ha avanzado desde la simulación de varios cientos de átomos hasta sistemas con relevancia biológica, incluyendo proteínas enteras en solución

con representaciones explícitas de solventes, proteínas incrustadas en membrana o grandes complejos macromoleculares como nucleosomas o ribosomas. La simulación de sistemas que tienen $\sim 50,000$ - $100,000$ átomos es ahora rutinaria, y las simulaciones de aproximadamente $500,000$ átomos son comunes cuando las instalaciones informáticas apropiadas están disponibles. Esta notable mejora es en gran parte una consecuencia del uso de la computación de alto rendimiento (HPC) y la simplicidad del algoritmo md básico. Un modelo inicial del sistema se obtiene a partir de estructuras experimentales o datos de modelado comparativo. El sistema simulado podría representarse con diferentes niveles de detalle. La representación atomística es la que conduce a la mejor reproducción de los sistemas reales. La representación solvente es un tema clave en la definición del sistema. Se han ensayado varios, pero, de nuevo, la más efectiva es la más simple, la representación explícita de moléculas de disolvente, aunque a expensas de aumentar el tamaño de los sistemas simulados. El disolvente explícito es capaz de recuperar la mayoría de los efectos de solvatación del disolvente real, incluidos los de origen entrópico como el efecto hidrófobo. Una vez construido el sistema, las fuerzas que actúan sobre cada átomo se obtienen derivando ecuaciones, los campos de fuerza, donde se deduce la energía potencial de la estructura molecular. Los campos de fuerza son ecuaciones complejas, pero son fáciles de calcular. La simplicidad de la representación del campo de fuerza de las características moleculares: resortes para la longitud y los ángulos de enlace, funciones periódicas para rotaciones de enlaces y potenciales de Lennard-Jones, y la ley de Coulomb para van der Waals e interacciones electrostáticas, respectivamente, asegura que los cálculos de energía y fuerza sean extremadamente rápidos incluso para sistemas grandes. Los campos de fuerza utilizados actualmente en las simulaciones moleculares atomísticas difieren en la forma en que se parametrizan. Los parámetros no son necesariamente intercambiables, y no todos los campos de fuerza permiten representar todos los tipos de moléculas, pero las simulaciones realizadas utilizando campos de fuerza modernos son normalmente equivalentes. Una vez que se obtienen las fuerzas que actúan sobre los átomos individuales, la ley clásica del movimiento de Newton se utiliza para calcular aceleraciones y velocidades y para actualizar las posiciones de los átomos. Como la integración del movimiento se realiza numéricamente, para evitar la inestabilidad, se debe usar un paso de tiempo más corto que los movimientos más rápidos en la molécula. Esto se clasifica normalmente entre 1 y 2 fs para simulaciones atomísticas, y es el principal cuello de botella del procedimiento de simulación. Las simulaciones de microsegundos de duración, que apenas arañan las escalas de tiempo de los procesos biológicos, requieren iteración a lo largo de este ciclo de cálculo 10^9 veces. Este es uno de los puntos fuertes de las estrategias de grano grueso. A medida que se utiliza una representación más simplificada del sistema, es posible que se realicen pasos de tiempo mucho más grandes y, por lo tanto, la duración efectiva de las simulaciones se amplía drásticamente. Por supuesto, esto se puede obtener a expensas de la precisión del conjunto de simulación. Los avances algorítmicos, que incluyen el ajuste fino de los cálculos de energía, la paralelización o el uso de unidades de procesamiento gráfico (GPU), han mejorado en gran medida el rendimiento de las simulaciones MD.

Material y equipo:

- Computadora
- Cuenta de Google
- Red de Internet

Desarrollo de la práctica:

Primero se llevo a cabo la prueba de como es que se llevaba a cabo la simulación, para esto se hizo uso de dos links en Google Colaboratory los cuales nos ayudarían con un ejemplo, la ubiquitina, proteína que ya hemos revisado en prácticas anteriores, realizando una simulación de dinámica molecular de ubiquitina en agua que abarca 100 ps.

```
#"Step","Potential Energy (kJ/mole)","Temperature (K)"
1000,-252871.8589338677,249.131956414742
2000,-240974.5005435872,285.08127746356536
3000,-237629.3577852986,298.9617951696606
4000,-236261.27040506588,295.82488533360856
5000,-236109.8714116046,300.95641631037006
6000,-235967.6785400842,297.14217581978005
7000,-237152.94749906872,303.088995498368
8000,-236796.11532949843,297.9392100064164
9000,-236647.89479971086,302.9656534380835
10000,-236659.60163939634,298.22332859637675
11000,-235454.48836654008,296.62069412211906
12000,-235946.1979689501,300.49362720469
13000,-236197.10495394503,300.11734100823674
14000,-235516.77354254114,299.068113143187
15000,-235911.5501176156,301.4578996192586
16000,-236001.03219743728,304.4800390181645
17000,-236116.29202703058,301.3100938585054
18000,-236197.01587398906,300.66920062642225
19000,-235788.30158982894,304.4475861421453
20000,-236382.5264532216,302.3110290572627
21000,-236333.30928279445,296.73763611499794
22000,-236902.7966367706,298.26728068075255
23000,-236723.44751028734,301.7886049891761
24000,-236746.4580297311,299.79627089841296
25000,-235463.18734006275,298.7809279603054
26000,-236178.96154481533,299.90973191648186
27000,-236758.9430138057,298.2300767589556
28000,-237204.89071314922,299.21757178610994
29000,-236295.69067112167,302.021240817176
30000,-235437.25358800386,300.0792234179134
32000,-235992.5204596096,297.8132779944658
33000,-236482.79882167547,297.5185973169784
34000,-236631.12876844715,303.61002580140945
35000,-235664.289084305,301.0166367917833
36000,-235253.76025857992,300.68413051309574
37000,-236355.7828572484,301.5572755326789
38000,-236584.3384375453,300.17917157614727
39000,-236915.59519134072,299.6809717666284
40000,-236035.46197816145,294.71157141750956
41000,-236310.2187340767,299.08119133347543
42000,-236476.95803603268,296.456399755067
43000,-236761.60929605097,302.8573724161512
44000,-236609.85090541854,302.8621163746616
45000,-236214.76602859292,298.88240332721506
46000,-235884.26870156598,300.6913409013581
47000,-235678.8461521467,295.3270228996617
48000,-235593.45985874176,300.8870632573906
49000,-236117.39805667056,300.3430158948293
50000,-236294.81907640683,297.37423795545703
```

Fig. 1. Cálculo de la energía potencial y temperatura en 50000 pasos

Por lo que tenemos la siguiente simulación:



Fig. 2. MDS de la ubiquitina en 100 ps

Seguido de eso continuamos analizando los resultados, los cuales son los siguientes: Calculando el área de superficie accesible al disolvente de cada átomo o residuo en cada marco de simulación, tenemos:

```
<mdtraj.Trajectory with 500 frames, 18388 atoms, 5795 residues, and unitcells>
sasa data shape (500, 18388)
(500,)
```

Fig. 3.SASA DATA SHAPE

Luego se hace la gráfica de “TOTAL SASA” en función del tiempo, obteniendo:

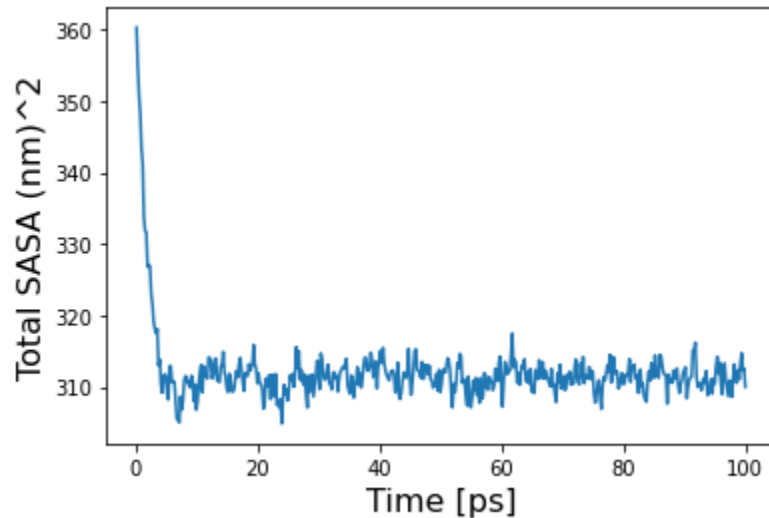


Fig. 4. Gráfica "TOTAL SASA"

En esta grafica podemos ver que hay una caída casi al momento de iniciar, teniendo muchas altas y bajas, pero manteniéndose relativamente abajo, dándonos a entender que no esta tan expuesto al agua.

Seguido se calcula el RSMF y su gráfica:

(76,)

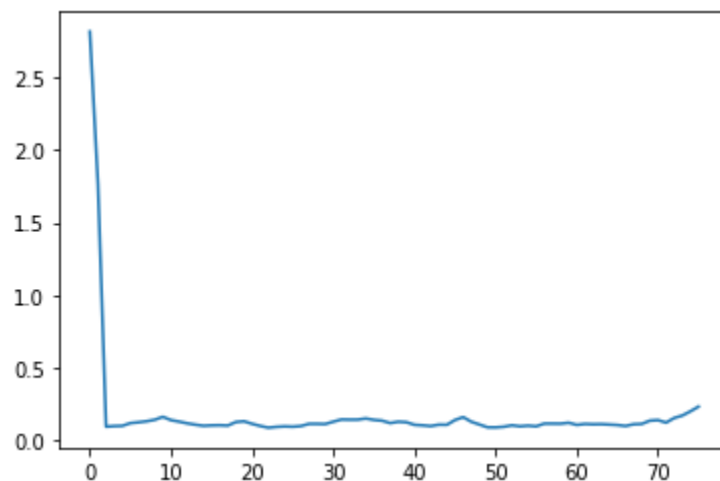


Fig. 5.Cálculo del RMSF y gráfica de este

Al igual que la anterior podemos ver en esta que los valores se mantienen muy abajo, aunque en ciertos momentos esta sube pero se mantiene casi en 0, diciéndonos que no hay mucha desviación con una partícula dada.

Luego el RMSD y su gráfica igualmente:

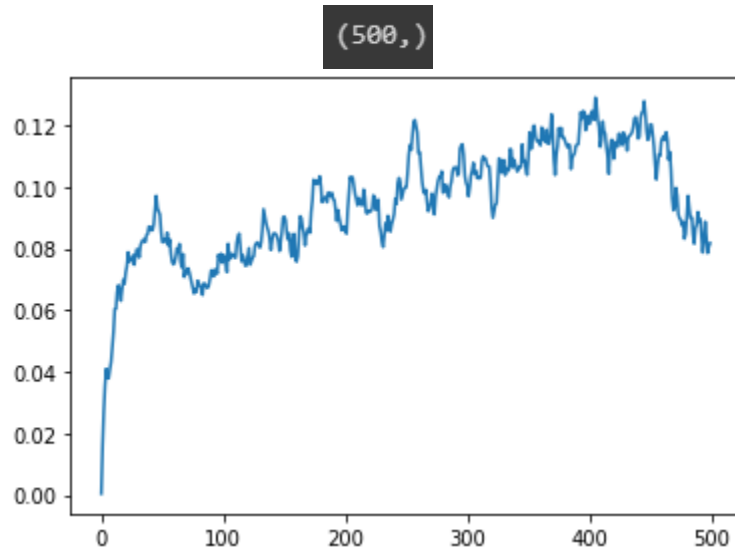


Fig. 6. Cálculo del RMSD y su gráfica

A diferencia de las anteriores podemos ver que aquí la gráfica va hacia arriba, teniendo muchos cambios y yendo poco a poco, lo cual nos dice que la distancia media entre átomos es grande, ya que la gráfica se eleva y se mantiene relativamente arriba.

Y finalmente calculamos el radio de giro para cada fotograma y se lleva a cabo su gráfica:

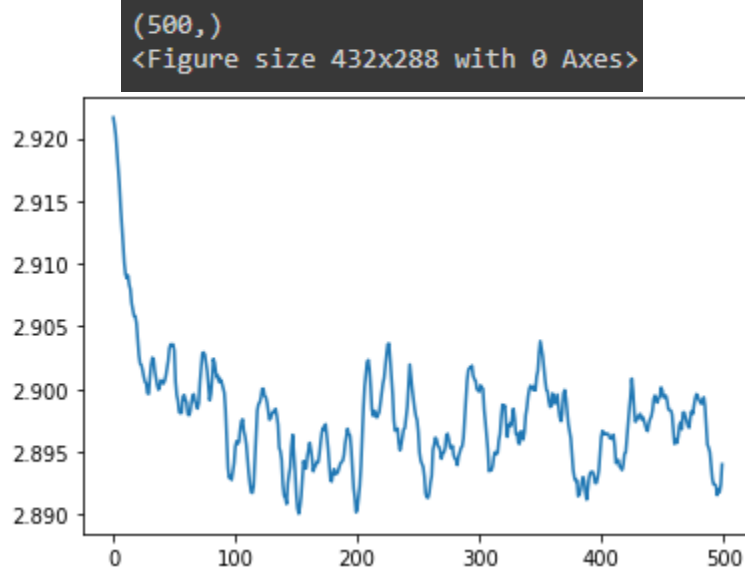


Fig. 7. Radio de Giro para cada fotograma y su gráfica

En esta también hay una gráfica que va hacia abajo, teniendo subidas, pero manteniéndose con valores medios.

Una vez hecho esto lo llevamos a cabo con otra proteína, en este caso se eligió el tripsinógeno bovino(1tgn) que es una proteína presente en los bovinos, así que iniciamos con esta proteína en agua que abarca 100 ps.

#Step	Potential Energy (kJ/mole)	Temperature (K)
1000	-398996.0407031269	249.80132299813653
2000	-379213.7706308671	285.403357197061
3000	-374441.4052322904	295.97444403978886
4000	-372300.293289623	300.3828548377194
5000	-372602.77670313366	297.02478507488377
6000	-372603.16811789194	299.19085359282917
7000	-371389.3701017747	300.71459564168885
8000	-371255.59763674455	299.6157371304681
9000	-372572.08861505496	300.7972952938129
10000	-371843.3147810162	298.50495643046605
11000	-372801.152664486	304.4289749661103
12000	-372728.5943783305	300.07295084886846
13000	-371738.00350843003	299.4360777026074
14000	-372198.40164967254	300.11629296822537
15000	-373619.8922478589	299.28464323504613
16000	-372551.63270334323	300.0799875188904
17000	-372369.33558043186	299.5162739164799
18000	-372010.6009593336	303.69156477162517
19000	-372020.8504328387	298.98364414298345
20000	-372432.9821953877	299.14658252800933
21000	-372665.82243954955	299.1762280621706
22000	-372456.8057031925	298.18473928012855
23000	-372743.3232472204	301.54684735759764
24000	-371193.88988304045	300.0165930983387
25000	-371539.3390740073	301.22447239446694
26000	-372086.03012202913	301.67010350578187
27000	-371202.01289108215	300.5617678280223
28000	-371643.8275460735	298.83480675188605
29000	-372203.78922130505	298.8666365156437
30000	-374572.21595405665	299.6036543059164
31000	-372821.07604572625	301.7422694309348
32000	-372310.8281718253	302.93159401504477
33000	-373014.5669617951	300.5277570665424
34000	-373060.62438668613	300.29432070602024
35000	-373059.5822415946	298.8358675401089
36000	-373531.5693066069	299.91830204665087
37000	-373508.958280592	298.588020743935
38000	-373026.9101054189	300.76527775909636
39000	-373346.3034771847	301.3592453665084
40000	-372918.9009920442	299.2896343395871
41000	-372824.8942290943	302.5062293203155
42000	-372197.99842798634	299.5774836401722
43000	-373847.97458146285	303.160507958421
44000	-372989.3719370983	299.8413425039322
45000	-371877.9841112064	299.89504283798624
46000	-372992.1398108946	300.8079359466464
47000	-372223.58278673247	299.61554521954076
48000	-373102.4748008553	299.81345770497296
49000	-372924.4714408078	299.16997267506173
50000	-373139.11326852767	297.5020535527254

Continuamos llevando a cabo la simulación, la cual queda de la siguiente manera:

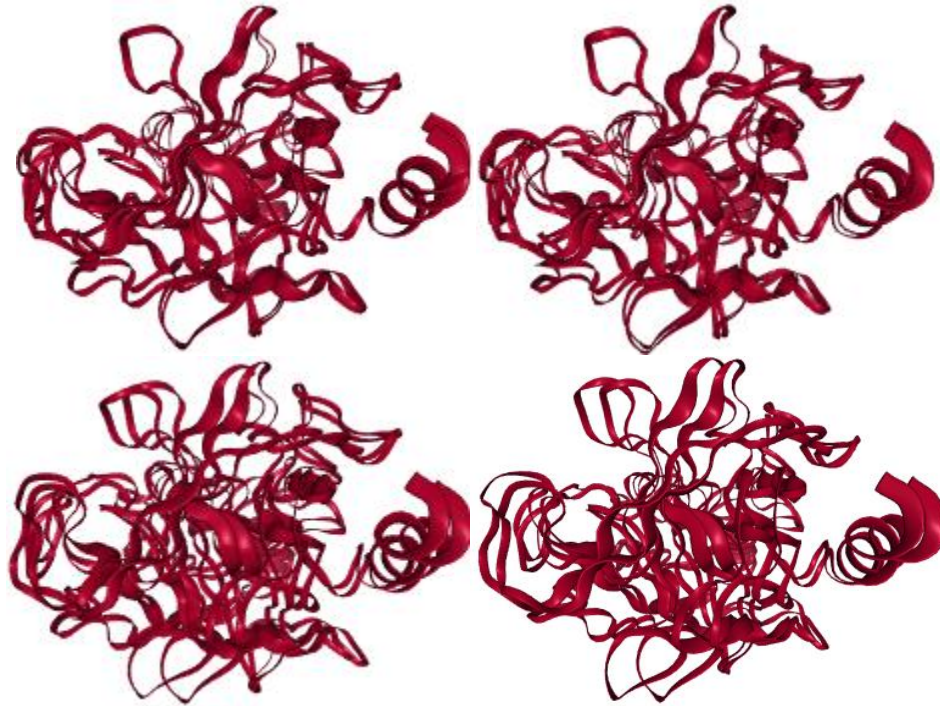


Fig. 8. MDS de 1TGN

Calculando el área de superficie accesible al disolvente de cada átomo o residuo en cada marco de simulación, tenemos:

```
<mdtraj.Trajectory with 500 frames, 30165 atoms, 9214 residues, and unitcells>
sasa data shape (500, 30165)
(500,)
```

Fig. 9. SASA DATA SHAPE

Luego se hace la gráfica de “TOTAL SASA” en función del tiempo, obteniendo:

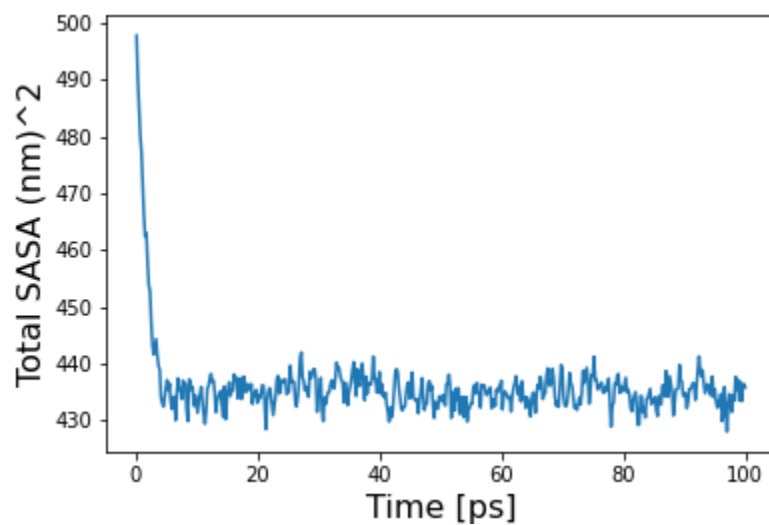


Fig. 10. Gráfica "TOTAL SASA"

Podemos ver en la gráfica que en cierto tiempo el valor SASA baja considerablemente manteniéndose casi así todo el tiempo hasta los 100 ps por lo que no esta tan expuesto al agua.

Seguido se calcula el RSMF y su gráfica:

(222,)

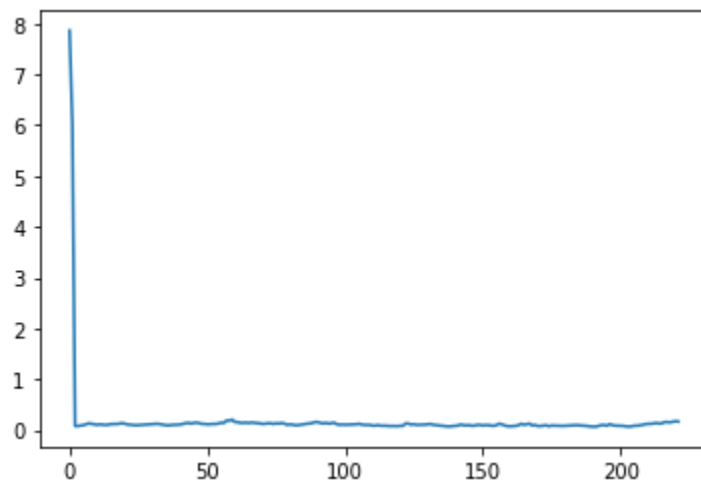


Fig. 11. Cálculo del RMSF y gráfica de este

Con esta grafica podemos ver que no hay mucha desviación de una partícula dada, ya que la gráfica no muestra un incremento y se mantiene prácticamente igual.

Luego el RMSD y su gráfica igualmente:

(500,)

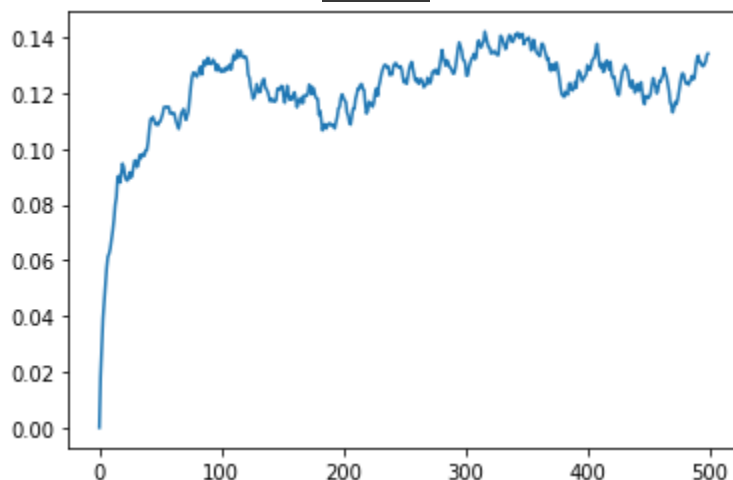


Fig. 12. Cálculo del RMSD y su gráfica

En el RMSD podemos ver que la distancia media entre átomos es grande, ya que la gráfica se eleva y se mantiene relativamente arriba.

Y finalmente calculamos el radio de giro para cada fotograma y se lleva a cabo su gráfica:

(500,)

<Figure size 432x288 with 0 Axes>

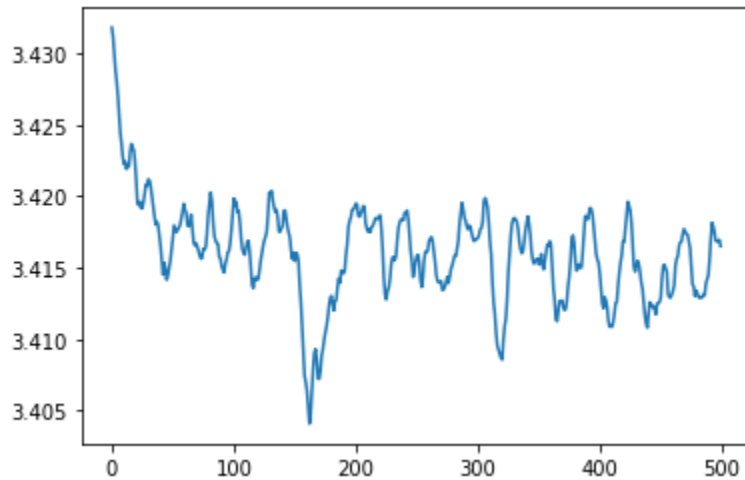


Fig. 13. Radio de Giro para cada fotograma y su gráfica

El radio de giro tiene muchos cambios, como vemos hay muchas subidas y bajadas, pero se mantiene relativamente en medio.

Por último, se nos pidió encontrar un trabajo de investigación que presente un estudio de dinámica molecular de una proteína e informe lo siguiente:

- Motor MD usado
- Tiempo de simulación
- Análisis realizados
- Contribución lograda

Por lo que se encontró la simulación del ADN y sus contraiones, viendo el comportamiento de los contraiones móviles, Na y K, analizándolo alrededor de una doble hélice de ADN-B con la secuencia CCATGCGCTGAC en solución acuosa durante dos trayectorias dinámicas moleculares de 50 ns de largo.

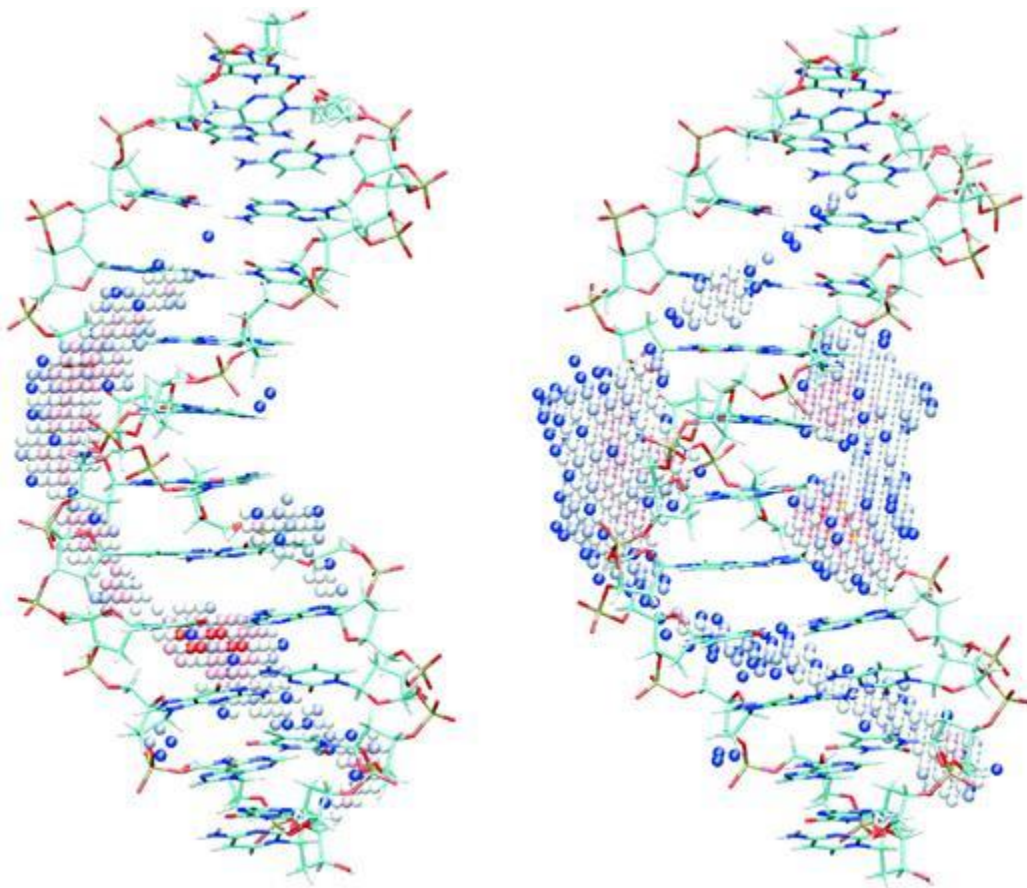


Ilustración 14. Distribución del contraión alrededor de una estructura de ADN promedio para Na (izquierda) y K (derecha). Las posiciones mostradas corresponden a aproximadamente 700 a 3500 visitas por 1 Å+3 simulación de más de 50 ns.

La construcción de modelos y simulaciones se realizaron utilizando el conjunto de programas AMBER y el conjunto de parámetros parm99 un conjunto de programas de simulación biomolecular.

Se realizaron simulaciones sobre una proteína dodecamérica con la secuencia d(CCATGCGCTGAC) que se construyó en una conformación estándar de ADN-B.

El oligómero fue neutralizado con 22 iones Na o K (colocados siguiendo valores de potencial electrostático) y solvatado con más de 5000 moléculas de agua TIP3P en una caja octaédrica truncada, lo que permitió al menos una capa de agua de 10 Å alrededor del soluto. Se eligieron las condiciones mínimas de sal porque la descripción adecuada de los iones en las simulaciones representa partes altamente controvertidas de los campos de fuerza actuales. Las simulaciones de dinámica molecular se llevaron a cabo a temperatura constante (300 K) y presión (1 bar) utilizando el algoritmo de Berendsen y condiciones de contorno periódicas. Se utilizó un paso de tiempo de integración de 2 fs y todas las longitudes de enlace que involucraban hidrógenos se restringieron usando SHAKE. Las interacciones de largo alcance se trataron mediante el enfoque PME con un corte de espacio directo de 9 Å. La lista de pares no enlazados se actualizó heurísticamente, siempre que cualquier átomo se haya movido más de la mitad de una distancia de "piel" de 2 Å, y el centro de movimiento de masa se eliminó cada 10 ps durante la simulación. Inicialmente, las moléculas de agua y los iones se relajaron mediante la minimización de energía y permitieron equilibrarse a 300 K alrededor del oligómero de ADN fijo durante 100 ps a volumen constante. Todo el sistema se calentó de

100 a 300 K durante 10 ps y se equilibró durante 40 ps con restricciones armónicas de 5.0 kcal / mol / Å⁺⁺² en los átomos de soluto a volumen constante. Posteriormente, la simulación se continuó a presión constante. Las restricciones se eliminaron gradualmente durante un período de 250 ps y el sistema se equilibró durante 1 ns. Para el análisis se utilizaron 50 ns de simulación.

Para analizar los movimientos de iones alrededor del soluto en una simulación de límite periódica, los iones que se alejan deben traducirse primero a la célula de simulación primaria. Estos cálculos de "imágenes" se realizaron con la utilidad ptraj del paquete AMBER. Se analizó qué porcentaje del espacio disponible en la caja de simulación fue muestreado efectivamente por contraiones durante el curso de la simulación. El volumen ocupado por el soluto se calculó numéricamente utilizando una cuadrícula con un espaciado de 0,2 Å y teniendo en cuenta los radios de van der Waals de los átomos de ADN (H, 1,0 Å; C, 1,7 Å; N, 1,5 Å; O, 1,4 Å; P, 1,9 Å) incrementado por el radio del ion (Na, 1,0 Å; K, 1,4 Å).

El volumen muestreado por los contraiones se calculó utilizando el mismo espaciado de cuadrícula y radio iónico. Un análisis del desplazamiento iónico con intervalos de tiempo de 1 ps, correspondiente a la frecuencia de ahorro de coordenadas, mostró que el desplazamiento medio fue de ~1 Å, aunque también se observaron desplazamientos de hasta 5 Å (excluyendo cualquier efecto causado por el procedimiento de imagen mencionado anteriormente). Las posiciones de iones no muestreadas debido al uso de pasos de 1 ps se tuvieron en cuenta mediante la generación de puntos intermedios. Uno de estos puntos se creó por interpolación lineal cuando la distancia entre dos puntos muestreados consecutivamente estaba entre 2,5 y 3 Å y se crearon dos puntos igualmente espaciados para distancias entre 3 y 5 Å.

Coefficientes de difusión

Los movimientos de los contraiones en presencia de ADN se caracterizaron por sus coeficientes de difusión (D) según la fórmula de Einstein:

$$\langle [X_{yd}(t_0 + \Delta t) - X_{yd}(t_0)]^2 \rangle = 6D \cdot \Delta t.$$

Esto requiere el cálculo de los desplazamientos cuadráticos medios de cada ion para una serie de intervalos de tiempo. El coeficiente de difusión viene dado simplemente por la pendiente de la gráfica de los desplazamientos cuadrados medios frente al tiempo. Los intervalos de tiempo utilizados en este cálculo variaron de 10 a 1000 ps en pasos de 10 ps. Para cada cálculo, el desplazamiento se promedió en todos los intervalos de tiempo a lo largo de la trayectoria.

Análisis de sitios de enlace

Se consideró que un ion estaba unido al soluto si su distancia de un átomo de soluto electronegativo era de <3.0 y 3.5 Å para Na y K, respectivamente. Este valor corresponde a un mínimo en la función de distribución radial de los iones alrededor de los átomos de soluto electronegativo, que separa un máximo agudo correspondiente a una unión directa y un máximo secundario correspondiente a interacciones mediadas por agua.

La ocupación del sitio se calculó para los sitios de unión a la base y a la columna vertebral en el ADN. Hay dos sitios de unión potenciales en el surco menor, Y(O2) y R(N3) (donde Y

= pirimidina y R = purina), y tres en el surco mayor, R(N7), G(O6) y T(O4). Para la columna vertebral, considerando los cinco átomos de oxígeno: los oxígenos aniónicos fosfato, O1P y O2P, los oxígenos estéricos fosfatados, O5' y O3' y el oxígeno del anillo de azúcar, O4'. Para cada uno de estos sitios, se calculó la ocupación, definida como un porcentaje del tiempo de simulación durante el cual un contraión estaba directamente unido. También se registró la media y las duraciones máximas de la consolidación. Para los sitios con las ocupaciones más altas, calculamos los tiempos de residencia de iones utilizando una función de correlación de tiempo estándar, con la distribución de tiempo de residencia (rtd) para un sitio de enlace dado definido como

$$rtd(\Delta t) = \sum_{t_0} \sum_t p_{yd}(t_0) \cdot p_{yd}(t_0 + \Delta t)$$

donde $p_i(t)$ es 1 si el ion i está presente en el tiempo t y de lo contrario 0. Los eventos cuando un ion fluctúa a través del límite dentro de 1 ps se ignoran. Los cálculos se realizaron con intervalos de tiempo que oscilaron entre 10 y 1000 ps. El tiempo de residencia efectivo (R) se calcula a partir del coeficiente de desintegración de una función exponencial ajustada a rtd:

$$\ln[rtd(\Delta t)/rtd(\Delta t_{\min})] = -R^{-1} \Delta t.$$

Influencia del contraión en la conformación del ADN

Con el fin de investigar el posible impacto de la unión iónica en la conformación del ADN, calculando tanto los parámetros helicoidales como los ángulos de flexión axial a lo largo de la simulación utilizando el programa Curves. Los anchos de ranura menores se midieron utilizando dos enfoques diferentes. Primero, calculando la distancia entre cadenas entre los átomos de fósforo P_i y P_{i+3} a través del surco menor (dPP). Esta distancia se redujo en 5,8 Å (suma de los radios de van der Waals de los dos grupos fosfato) para ser comparable con la convención adoptada por otros autores. En segundo lugar, utilizamos el ancho de ranura menor calculado por el programa Curves, que corresponde a la distancia más corta entre splines que pasan a través de los átomos de fósforo de la columna vertebral y se encuentran en un plano que pasa a través del vector pseudodyad del par de bases elegido. Las estructuras de ADN se clasificaron en grupos dependiendo de la presencia o ausencia de contraiones en lugares particulares alrededor del ADN, y se calculó el ancho promedio del surco y la desviación estándar para cada grupo. Se consideraron los siguientes grupos: (i) contraiones unidos en el surco menor asociado a un par de bases dado; ii) contraiones unidos en la ranura principal asociada a un par de bases determinado; iii) contraiones entre dos grupos fosfato a través del surco menor; y iv) sin contraiones consolidadas. Tenga en cuenta que las estructuras que satisfacen más de un criterio fueron raras y posteriormente se excluyeron de nuestro análisis.

Para encontrar estructuras pertenecientes al tercer grupo, se consideraron iones que caen a menos de $(2 + dPP/2)$ Å de los átomos de fósforo, P_i y P_{i+3} , lo que define el ancho de ranura menor local. Este enfoque define un volumen equidistante en forma de lente convexa de los grupos fosfato en la entrada del surco menor. También se probó un enfoque alternativo, donde el volumen de la lente es constante, y dio resultados muy similares. También calculamos la relación entre el ancho de ranura menor y la ocupación del sitio iónico para comparar nuestros resultados con los de Hamelberg y sus compañeros de trabajo. Esto implicó calcular el ancho

promedio de la ranura y las ocupaciones correspondientes de ranuras e iones fosfato para todos los segmentos de 10 ps durante el tiempo total de simulación.

Potencial electrostático molecular

El cálculo del potencial electrostático alrededor del oligómero de ADN se realizó con el programa Delphi. El programa se adaptó para calcular un valor promedio del potencial electrostático sobre el área accesible con disolvente asignada a un átomo dado. Se utilizó una cuadrícula de 145 puntos y se realizaron cálculos con radios de 1,2 o 1,4 Å. Las constantes dieléctricas internas y externas se establecieron en 2 y 80, respectivamente, y la fuerza iónica fue cero.

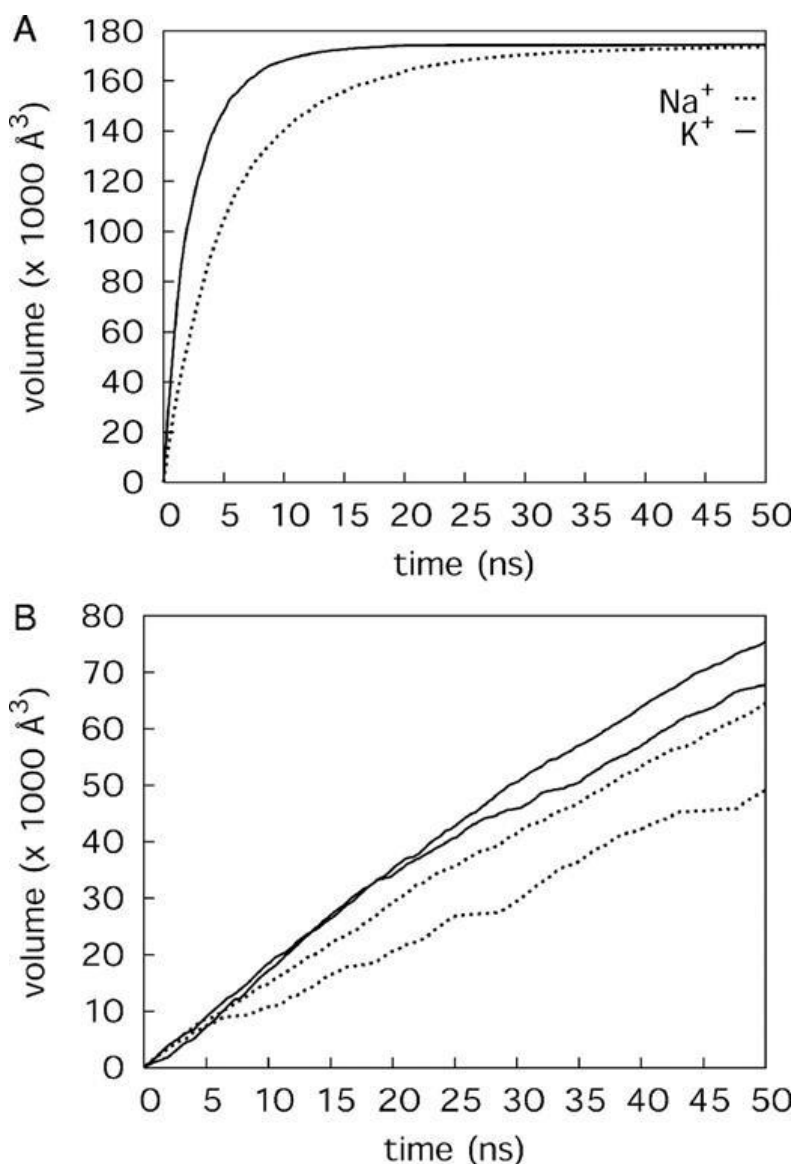


Fig. 15. Volumen muestreado por los contraiones Na (línea punteada) y K (línea sólida) durante el tiempo de simulación de 50 ns calculado para todos los iones (A) y para los iones más lentos y rápidos (B).

Los resultados muestran que el movimiento promediado de los contraiones en presencia de ADN sigue siendo difusivo, siendo el movimiento de K mucho más rápido que el de Na. Los

iones muestrean el espacio completo disponible alrededor del soluto durante el tiempo de simulación, pero los iones individuales muestrean solo alrededor de un tercio del espacio.

Los iones visitan preferentemente los sitios electronegativos alrededor del ADN, aunque la unión directa a las bases de ADN sigue siendo un evento bastante raro, con valores más altos de ocupación del sitio que no >13%. Todos los iones pasan algún tiempo en los surcos de ADN y los sitios más ocupados son visitados por muchos iones diferentes durante la simulación. Los sitios de unión preferidos dependen en gran medida de la naturaleza del ion: Na se une principalmente en el surco menor alrededor de T9 (O2), mientras que K se une en el centro del surco mayor. Curiosamente, el paso A3pT4, considerado como un probable bolsillo de unión en la ranura menor, no muestra una fuerte unión de iones. Esto indica que los sitios de unión iónica pueden ser sensibles a sus contextos de secuencia. La ocupación del sitio de fosfato es más uniforme a lo largo de la secuencia, pero los fosfatos cercanos a las bases con alta ocupación de iones generalmente exhiben una mayor ocupación fraccional. El Na y el K exhiben dinámicas de unión muy diferentes. Los tiempos de residencia para Na varían de 100 a 400 ps para sitios altamente ocupados con la duración de encuadración más larga que alcanza 1 ns. Por el contrario, K se une y se disocia con mayor frecuencia, lo que resulta en tiempos de residencia de <70 ps. Estas características, junto con su dominio en el entorno intracelular, hacen que K esté particularmente bien adaptado para simulaciones de ADN.

Utilizando varias técnicas de análisis, no se ha encontrado evidencia de acoplamiento de la unión iónica monovalente y los cambios en la geometría de la hélice local. Las características estructurales importantes del oligómero (como el surco menor estrecho en C8-G17) estuvieron presentes en simulaciones con contracción y entre diferentes grupos de estructuras. Por lo tanto, las propiedades estructurales del ADN parecen estar codificadas en su secuencia de bases. Sin embargo, se encontraron diferencias en la estructura del ADN dependiendo del tipo de ion monovalente utilizado en las simulaciones. Específicamente, las conformaciones no canónicas de α / γ de la columna vertebral aparecieron principalmente en la simulación con Na y fueron responsables de la torsión reducida y el surco menor más estrecho. Estas transiciones α/γ producen subestados conformacionales de larga duración en la escala de tiempo de nanosegundos, que requieren especial atención en estudios futuros. No se puede excluir que estos subestados conformacionales surjan como un artefacto del campo de fuerza. Sin embargo, podrían haber sido simplemente pasados por alto en estudios experimentales. La respuesta final a esta pregunta solo puede provenir de estudios experimentales especialmente diseñados de columnas vertebrales de ácidos nucleicos.

Conclusiones y recomendaciones:

Las MDS son muy importantes en el ámbito de la biología ya que nos ayudan a interactuar con estructuras a niveles atómicos, en nuestro caso fue con proteínas, y realizar experimentos y análisis sin necesidad de llevarlo a cabo en la vida real. En esta práctica pudimos ponerlas a prueba llevándolas a cabo con Python, además de llevar a cabo graficas que nos ayudan a hacer un análisis de dichas simulaciones, análisis de temas con las que ya habíamos trabajado anteriormente como lo son el RMSD y RMSF, además con ayuda de

una investigación pudimos ver como es que en la bioinformática son muy usadas las MDS y nos dan grandes aportes al campo biológico.

Bibliografía:

- Durrant, J., & McCammon, J. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1). <https://doi.org/10.1186/1741-7007-9-71>
- Hollingsworth, S., & Dror, R. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129-1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- Hospital A, Goñi JR, Orozco M, Gelpi J. Simulaciones de dinámica molecular: avances y aplicaciones. *Adv Appl Bioinform Chem*. 2015;8:37-47 <https://doi.org/10.2147/AABC.S70333>
- Varnai, P. (2004). DNA and its counterions: a molecular dynamics study. *Nucleic Acids Research*, 32(14), 4269-4280. <https://doi.org/10.1093/nar/gkh765>