



CURSO-TALLER II (Dinámica molecular)

• Introducción •

Este tutorial le guiará paso a paso para realizar una simulación de dinámica molecular de un complejo proteína-ligando, en el programa GROMACS v2022,¹ en Linux.

El sistema para simular consiste en la enzima *Death-Associated Protein Kinase 1* (DAPK1; <https://www.uniprot.org/uniprotkb/P53355/entry>),² en complejo con el ligando morina (RCSB PDB: <https://www.rcsb.org/structure/5AUY>).³

• Simulación de enzima DAPK1 en complejo con morina •

0. Descarga de documentos

- Descargar comprimido “Input” de la liga: <https://www.quimicafarmacaceuticafq.com/descargas>.
- Abrir una terminal.
- Mover el comprimido al directorio Documentos y descomprimir.

```
$ mv Descargas/02_Input.tar.gz Documentos/  
$ cd Documentos  
$ tar -zxvf 02_Input.tar.gz  
$ cd 02_Input
```

- Revise que se encuentra en el directorio indicado [instrucción: ls].

1. Visualización del complejo enzima-ligando

Las coordenadas en formato **.pdb** se pueden visualizar con el programa VMD⁴ con la instrucción:

```
$ vmd 5auyMRI.pdb
```

2. Parametrización del ligando y la enzima

Las coordenadas del ligando se pueden extraer del archivo **5auyMRI.pdb** con la instrucción:

```
$ grep "MRI" 5auyMRI.pdb > MRI.pdb
```

Nota: Los parámetros del ligando se pueden obtener de diferentes maneras, una de ellas es mediante el uso de servidores. Con el servidor ACPYPE (<https://www.bio2byte.be/acpype/>)⁵ se generan los parámetros necesarios a partir del archivo de coordenadas del ligando (archivo **MRI.itp**). El campo de fuerza utilizado es *General Amber Force Field 2* (GAFF2)⁶.

Las coordenadas de la proteína se pueden extraer del archivo **5auyMRI.pdb** con la instrucción:

```
$ grep -v "MRI" 5auyMRI.pdb > protein.pdb
```

Posteriormente se obtienen los parámetros de la proteína:

```
$ gmx pdb2gmx -f protein.pdb -o protein_processed.pdb -ff amber99sb -water tip3p -ignh -ter
```

Donde:

- f indica la ubicación y/o nombre del archivo de entrada que contiene las coordenadas iniciales.
- o indica la ubicación y/o nombre del archivo de salida que contiene las coordenadas procesadas.
- ff indica el campo de fuerza a utilizar, en este caso AMBER99SB.
- water indica el modelo de agua a utilizar, en este tutorial será TIP3P.
- ignh indica que se ignoren los átomos de hidrógeno en el archivo de entrada, si estuvieran presentes.
- ter indica que los extremos amino y carboxilo deben especificarse en su estado de protonación.

La instrucción, además de generar el archivo **protein_processed.pdb**, genera la topología de la proteína con el archivo **topol.top**, el cual contiene información necesaria y vínculos a archivos adicionales con extensión **.itp**.

Una vez parametrizada la proteína, se agregan las coordenadas del ligando en un nuevo archivo .pdb:

```
$ cat MRI.pdb protein_processed.pdb | grep -v "END" > 5auyMRI_processed.pdb
```

Dentro del archivo **topol.top**, escribir bajo la línea `#include "amber99sb.ff/forcefield.itp"`:

```
#include "MRI.itp"
```

Y posteriormente, en la parte final de **topol.top**, escribir en la sección `[molecules]`, antes de la línea `Protein_chain_A`, el texto:

```
MRI 1
```

Al finalizar la edición del **topol.top**, guardar el documento y salir. De esta manera, se indica en el archivo de topología que incluya los parámetros del ligando para las simulaciones.

3. Generación de la celda de simulación

Una vez generada la topología del sistema, se debe delimitar una celda de simulación con la instrucción:

```
$ gmx editconf -f 5auyMRI_processed.pdb -o 5auyMRI_box.pdb -c -d 1.0 -bt cubic
```

Donde:

- c indica que se centre las coordenadas en el centro de la celda.
- d indica que la pared de la celda se establezca a una distancia mínima de 1.0 nm del extremo de la proteína.
- bt indica que se construya una celda cúbica.

Esta función genera el archivo **5auyMRI_box.pdb**, el cual contiene la información de la geometría de la celda de simulación.

4. Solvatación del sistema

Una vez delimitada la celda, se solvata el sistema con el modelo TIP3P de molécula de agua:

```
$ gmx solvate -cp 5auyMRI_box.pdb -o 5auyMRI_solv.pdb -cs spc216.gro -p topol.top
```

Donde:

- cp es el archivo de entrada a solvatar (debe contener la información de la celda).
- o es la salida, el sistema solvatado.
- cs indica el archivo interno de GROMACS spc216.gro que contiene la información de las moléculas de agua para solvatar
Nota: no se requiere tener este archivo en la carpeta, se encuentra en las librerías del programa.
- p **topol.top** se usa para que se incluya el número de moléculas de agua agregadas en la topología.

5. Adición de iones y neutralización del sistema

El sistema debe poseer carga eléctrica total de cero para poder efectuar correctamente los cálculos de energía e interacciones. La adición de iones y la neutralización del sistema se obtienen en los dos pasos siguientes:

```
$ gmx grompp -f ions.mdp -c 5auyMRI_solv.pdb -o 5auyMRI_solv.tpr -p topol.top -maxwarn -1
```

Donde:

- f indica un archivo de parámetros **.mdp**, donde se especifican las condiciones de la simulación.
- c indica el archivo de coordenadas al cual se aplicará lo especificado en el **.mdp**.
- o es un archivo binario único de salida.
- p indica la topología del sistema para que el programa integre toda la información.

Nota: *gmx grompp* es la función de GROMACS que evalúa y compila toda la información del sistema para realizar las simulaciones. En esta etapa se reportan los errores e inconsistencias que puedan existir en el sistema. Esta función genera como salida un archivo binario con extensión **.tpr**, el cual contiene todos los datos necesarios para efectuar la simulación.

Una vez generado el archivo binario **5auyMRI_solv.tpr**, se procede a adicionar los iones:

```
$ gmx genion -s 5auyMRI_solv.tpr -o 5auyMRI_solv_ions.pdb -neutral -conc 0.15 -p topol.top
```

Donde:

- s indica el archivo binario **.tpr** de entrada.
- o indica el archivo de salida de coordenadas.
- neutral indica que la carga total del sistema sea neutra.
- conc indica la concentración de iones a añadir (en este caso 0.15 M).
- p indica que se incluyan los iones adicionados en la topología.

Para añadir iones (Na⁺ y Cl⁻, por defecto), la función preguntará la identidad del componente del sistema que se desea reemplazar. Se debe escoger el número índice de las moléculas de agua (SOL).

6. Minimización de energía potencial por optimización geométrica

Una vez que se tiene el sistema de DAPK1-morina solvatado y con iones, se puede proceder al paso de minimización de energía potencial, la cual se especifica en el archivo llamado **em.mdp**.

```
$ gmx grompp -f em.mdp -c 5auyMRI_solv_ions.pdb -o em.tpr -p topol.top -r 5auyMRI_solv_ions.pdb
```

Donde:

-r indica el archivo de referencia para las posiciones de los átomos.

Nota: La minimización de energía potencial y los equilibrios siguientes se realizan con restricciones en las posiciones atómicas, con el objetivo de no inducir un desplegamiento o desplazamiento considerable durante la estabilización de las variables mecánicas.

La simulación de este proceso se efectúa de la siguiente manera:

```
$ gmx mdrun -v -deffnm em
```

Donde:

-v indica que GROMACS imprima en pantalla el progreso de la simulación.

-deffnm indica que los nombres de los archivos de salida posean el mismo nombre que el **.tpr**.

Nota: La trayectoria de las simulaciones se guardan en archivos **.trr** o **.xtc**, y los valores de las variables mecánicas y termodinámicas en archivos **.edr**.

El perfil de la energía potencial a través del tiempo (y otras variables), se obtienen con la instrucción:

```
$ gmx energy -f em.edr -o potential.svg
```

Este perfil se puede visualizar gráficamente con:

```
$ xmgrace potential.svg
```

7. Simulación de un colectivo NVT: definición de temperatura y volumen en el sistema

Con el sistema optimizado, se pueden simular las condiciones necesarias para definir la temperatura y el volumen en la celda de simulación a través del colectivo isomolar-isocórico-isotérmico (NVT), cuyas especificaciones se encuentran en el archivo **nvt.mdp**. Las variables mecánicas necesarias, como la velocidad inicial de los átomos, se definen en esta simulación, que debe alcanzar un estado de equilibrio en sus variables mecánicas y termodinámicas para concluirlo.

En el archivo **nvt.mdp**, se debe crear un archivo índice que contenga todos los componentes del sistema, y al que se le pueden incorporar partes específicas de él, como grupos de átomos, residuos, etc. Los índices se crean con la instrucción:

```
$ gmx make_ndx -f em.tpr -o Index.ndx
```

La cual abre una consola para indicar grupos específicos en el sistema. Teclee la línea para crear el grupo *Protein_MRI*, que engloba la enzima y el ligando, con la sintaxis:

```
> 5 | r MRI
```

```
> q
```

El índice se guardará con el nombre **Index.ndx**

La generación del archivo **.tpr** necesario se realiza con:

```
$ gmx grompp -f nvt.mdp -c em.gro -o nvt.tpr -p topol.top -n Index.ndx -r 5auyMRI_solv_ions.pdb
```

Nuevamente, si no hay errores o inconsistencias, la función *grompp* genera el archivo binario **nvt.tpr**. La simulación de este colectivo se efectúa con:

```
$ gmx mdrun -v -deffnm nvt
```

Nota: Este proceso puede demorar dependiendo del tamaño del sistema, el tiempo de simulación solicitado y los recursos de la computadora para realizarlo.

Para visualizar el perfil de la temperatura y visualizar gráficamente, tecleé las instrucciones:

```
$ gmx energy -f nvt.edr -o temperature.xvg
```

```
$ xmgrace temperature.xvg
```

8. Simulación del colectivo NPT: definición de presión y densidad

Una vez que el sistema ha alcanzado el equilibrio térmico en el colectivo NVT, se puede establecer la presión, la cual se define durante el colectivo isomolar-isobárico-isotérmico (NPT) en el archivo **npt.mdp**.

```
$ gmx grompp -f npt.mdp -c nvt.gro -o npt.tpr -p topol.top -n Index.ndx -r 5auyMRI_solv_ions.pdb
```

```
$ gmx mdrun -v -deffnm npt
```

Para visualizar el perfil de la presión y visualizar gráficamente, tecleé las instrucciones:

```
$ gmx energy -f npt.edr -o pressure.xvg
```

```
$ xmgrace pressure.xvg
```

9. Producción de dinámica molecular del sistema

Una vez que las variables mecánicas y termodinámicas han alcanzado un estado de equilibrio, se procede a la simulación de la producción de dinámica molecular, especificada en el archivo **md.mdp**.

```
$ gmx grompp -f md.mdp -c npt.gro -o md.tpr -p topol.top -n Index.ndx
```

```
$ gmx mdrun -v -deffnm md
```

10. Introducción al análisis básico de una simulación de dinámica molecular

De manera general, los análisis fundamentales de una simulación de dinámica molecular incluyen evaluar los desplazamientos y movimientos promedio durante la trayectoria (**.trr** y **.xtc**). Estos análisis incluyen la raíz de la desviación cuadrática media (RMSD), la raíz de la fluctuación cuadrática media (RMSF), el radio de giro (Rg), estructura secundaria (SS), entre otros.

- **RMSD.** El perfil de RMSD de la cadena principal (átomos N-C α -C) de la proteína es de particular interés para evaluar desplazamientos importantes, de manera importante en regiones que tienen una estructura secundaria definida, y para analizar cambios conformacionales. El perfil temporal de RMSD para cualquier componente del sistema simulado se obtiene con la instrucción:

```
$ gmx rms -f md.xtc -s md.tpr -o rmsd_DAPK1_backbone.svg -n Index.ndx
```

- **RMSF.** El RMSF proporciona información sobre la magnitud de la movilidad de los átomos durante la simulación. Para cualquier componente del sistema simulado se obtiene con la instrucción:

```
$ gmx rmsf -f md.xtc -s md.tpr -o rmsd_DAPK1_backbone.svg -n Index.ndx -res
```

- **Rg.** El Rg es un indicador de la extensibilidad o compacidad de la proteína durante la simulación.

```
$ gmx gyrate -f md.xtc -s md.tpr -n Index.ndx
```

- **SS.** La estructura secundaria se calcula a partir de los valores de los ángulos diedros y formación de puentes de hidrógeno:

```
$ gmx do_dssp -f md.xtc -s md.tpr -o ss_DAPK1.xpm
```

Nota: El archivo de salida se encuentra en formato de imagen **.xpm** (X PixMap), y puede ser convertido al formato **.eps** (Encapsulated PostScript) y visualizarlo con las instrucciones:

```
$ gmx xpm2ps -f ss_DAPK1.xpm -o ss_DAPK1.eps
$ evince ss_DAPK1.eps
```

Referencias

1. H. Bekker, H.J.C. Berendsen, E.J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, and H. Keegstra *et al.*, "Gromacs: A parallel computer for molecular dynamics simulations"; pp. 252–256 in *Physics computing* **92**. Edited by R.A. de Groot and J. Nadrchal. World Scientific, Singapore, 1993 / Manual: <https://doi.org/10.5281/zenodo.3562512> / Sitio: <https://www.gromacs.org/>
2. The UniProt Consortium, "UniProt: the universal protein knowledgebase in 2021"; pp. D480-D489 in *Nucleic Acids Research* **49(D1)**.
3. T. Yokoyama, Y. Kosaka & M. Mizuguchi. "Structural Insight into the Interactions between Death-Associated Protein Kinase 1 and Natural Flavonoids"; pp. 7400–7408 in *Journal of Medicinal Chemistry* **58(18)**, 2015.
4. W. Humphrey, A. Dalke & K. Schulten. "VMD - Visual Molecular Dynamics"; pp. 33-38 in *Journal of Molecular Graphics* **14**, 1996.
5. A. W. Sousa da Silva, W. F. Vranken. "ACPYPE - AnteChamber PYthon Parser interfacE"; 367 in *BMC Research Notes* **5**, 2012.
6. H. Xibing, H. M. Viet, Y. Wei, L. Tai-Sung & W. Junmei. "A fast and high-quality charge model for the next generation general AMBER force field"; 114502 in *Journal of Chemical Physics* **153**, 2020.