

## **PRACTICA XVIII: SIMULACIÓN DE LA DINÁMICA MOLECULAR DE PROTEÍNAS. (Práctica en revisión)**

**Objetivo:** General conocer y aplicar las bases de la simulación de la dinámica molecular de una proteína.

### **Objetivos específicos:**

- Procesar archivo de coordenadas PDB de proteínas para generar archivos de estructuras de proteínas que puedan emplearse para el análisis de simulación de dinámica molecular.
- Realizar la solvatación de la proteína en esferas o cajas de agua.
- Llevar a cabo la dinámica de equilibrado de la proteína con NADM2.
- Analizar las trayectorias y comportamiento del sistema después de la dinámica molecular.

### **Dinámica molecular con NAMD**

En este tutorial se llevará a cabo una simulación de dinámica molecular para lograr el equilibrado de una molécula de proteína inmersa en solvente. Para este estudio se utilizará el software NAMD2 (Phillips *et al.*, 2005) para la simulación de la dinámica molecular y programa VMD-Visual Molecular Dynamics (Humphrey *et al.*, 1996). Ambos programas pueden obtenerse de la página del Theoretical and Computational Biophysics Group de la University of Illinois at Urbana-Champaign).

URL: <http://www.ks.uiuc.edu/>

Los programas son de uso libre para la comunidad académica pero se invita a los usuarios a registrarse en el sistema para la descarga de los mismos.

Los programas son multiplataforma y se dispone de versiones de ellos para MS-Windows, Mac OSX y Linux. Dado que la dinámica molecular es un proceso computacionalmente intensivo, existen versiones de estos programas que permiten explotar las características de procesamiento multicores, versiones para clusters y GPU-CUDA.

Algunas herramientas complementarias que pueden ser útiles para este tutorial son un editor de texto, herramientas para graficar datos (Excel, Gnuplot, Xmgrace, etc.) y software para la visualización de estructuras moleculares (RasMol, Jmol, Chimera).

Una vez instalados los programas se podrán seguir los pasos propuestos para este tutorial.

Para llevar a cabo la simulación de dinámica molecular se necesita:

1. El archivo PDB que almacena las coordenadas y/o velocidades de los átomos del sistema.

2. Un archivo de estructura de la proteína (archivo psf) el cual almacena información estructural de la proteína tales como los diversos tipos de interacciones de enlace.
3. Un archivo de parámetros de campos de fuerzas (force fields).
4. Un archivo de configuración que almacena los parámetros que requiere NAMD para llevar a cabo la simulación.

## **1. Obtención del archivo de coordenadas de la proteína.**

- a. Para este ejemplo se va a trabajar con la estructura de la proteasa del virus HIV-I. Una estructura de esta proteína, determinada por la técnica de cristalografía de rayos X puede obtenerse de la base de datos del Protein Data Bank (<http://www.pdb.org>) con la clave 1HPV. Al consultar esta estructura es conveniente revisar los datos relativos a la resolución de la proteína y a la presencia de ligando. La estructura 1HPV muestra a la proteína en la presencia del inhibidor Aprenavir (APV).
- b. Los archivos PDB son archivos de texto que contienen las coordenadas de los átomos de la estructura. Puede abrir estos archivo mediante algún editor de texto. El archivo contiene información respecto a la técnica experimental empleada, resolución y calidad de la estructura. Las coordenadas de los átomos de la estructura pueden consultarse en las sección "ATOM" del archivo. Observe que en esta sección se identifican los átomos y los aminoácidos a los que pertenecen, excepto los átomos de hidrógeno los cuales por lo general no se incluyen en estos archivos ya que estos no son detectados normalmente por la técnica de cristalografía de rayos X. Por otro lado, las estructuras pueden corresponder a proteínas oligoméricas por lo que pueden contener varias cadenas polipeptídicas. Cada cadena se identifica con una letra distinta.
- c. Identifique las coordenadas que pertenecen a las dos cadenas polipeptídicas presentes en esta estructura y verifique la longitud de las cadenas. Asimismo identifique las coordenadas correspondientes al inhibidor y otros átomos que pueden estar presentes en la estructura. Para este ejercicio solo se llevará a cabo la simulación de la dinámica molecular de la cadena "A" presente en esta estructura.
- d. La estructura de la cadena deseada puede extraerse previamente al ensayo con algún programa como RasMol/RasWin, Jmol o Chimera. También pueden emplearse las herramientas de VMD para extraer el archivo.

## **2. Generación del archivo de estructura de la proteína (psf).**

Un archivo PSF contiene información molecular necesaria para que se pueda aplicar un potencial de campo (Force Field). El archivo incluye datos respecto a la topología en la cual se definen los tipos de átomos, nombres, enlaces y cargas parciales que aparecen en cada residuo. Los parámetros del potencial de campo permiten estimar valores para interacciones entre átomos enlazados y no enlazados.

El archivo PSF será generado de manera automática utilizando la herramienta "Automatic PSF builder" de VMD para lo cual:

- a) Depositar en una carpeta el archivo PDB. (Nota: como en varias operaciones se necesitará tener acceso a esta carpeta mediante la consola de comandos, es

conveniente que esta carpeta se encuentre en un lugar fácilmente accesible para este fin).

- b) Ejecutar VMD y desde el menú file seleccionar “New molecule”. En el cuadro de diálogo deberá seleccionarse el archivo PDB con el cual se realizará el trabajo y entonces presionar el botón “Load”. La molécula deberá visualizarse ahora en la interfaz gráfica de VMD. Puede emplear las herramientas de representación menú “Graphics” para modificar la representación de la molécula. Por ejemplo, puede ser conveniente modificar los colores para presentar cada cadena polipeptídica empleando un color distinto.
- c) Desde el menú “Extensions” seleccionar la aplicación “Tk console”. La consola Tk es un interfaz tipo “UNIX” con la cual se pueden ejecutar algunos de los paquetes auxiliares de VMD. Para empezar es conveniente localizar desde esta terminal, la carpeta en la cual se tiene el archivo PDB de trabajo. De esta forma, todos los datos generados por VMD se guardarán ahora en esta carpeta.
- d) Seleccionar en la lista la molécula en estudio y posteriormente ingresar a “Extensions > Simulation > Automatic PSF builder”. Con esto se abrirá un cuadro de diálogo en el cual deberán configurarse varios parámetros: (i) Debe especificarse un archivo de topología este archivo contiene los datos del potencial CHARM necesarios para establecer la conectividad e interacciones entre los archivo de coordenadas PDB. A menudo pueden especificarse otros tipos de archivos de potencial dependiendo si la estructura es de proteína, ácidos nucleicos o mixta. Para este ejercicio se utiliza el archivo `top.all27.prot.lipid.inp`. Presionar el botón “Load input files”. (ii) En el segundo se debe seleccionar el tipo de molécula a generar, para este ejemplo se debe seleccionar “Protein” y luego presionar el botón “Guess and split chains using current selections”. (iii) En la ventana “Segments identified” aparecen listadas todas la cadenas polipeptídicas identificadas. En este ejemplo se va a utilizar la primera cada de la estructura (P1), para lo cual deben eliminarse de la lista las otras cadenas y enseguida presionar el botón “Create chain”. (iv) VMD genera las cadenas. Puede mostrar una alerta en cuanto a que no se emplearon parches para la identificación. Los parches son instrucciones complementarias que se emplean para identificar otros átomos que no fueron reconocidos en el análisis. Para este ejemplo no se agregarán parches. (v) Como resultado de este análisis se debieron generar tres archivos: un PDB con los datos de la cadena seleccionada, un archivo PSF y un archivo log con información detallada del proceso.

### **3. Solvatación de la proteína.**

Para llevar a cabo la dinámica molecular de la proteína es necesario considerar el solvente (incluso la presencia de iones en el medio). En esta sección se ilustran dos métodos distintos de solvatación: en una esfera de agua y en una caja de agua.

#### **3.1. Solvatación en una esfera de agua.**

- a) Es conveniente copiar los archivos `pdb` y `psf` obtenidos en el paso anterior, en una nueva carpeta. También en esta carpeta se debe copiar el script `wat_sphere.tcl`. El script es un archivo que contiene varios comandos que VMD ejecuta de manera

automática. Se debe editar la sección “set mol name” de este archivo con un editor de texto e incluir el nombre base de los archivos PDB y PSF (sin la extensión). El contenido del script describe varios de los parámetros que pueden ajustarse para la solvatación.

- b) Salir de la interfaz gráfica de VMD. A través de la consola se debe ingresar a la carpeta con los tres archivos necesarios para la solvatación y desde ella se ejecuta el comando:

```
vmd -dispdev text -e wat-sphere.tcl
```

Nota: El nombre del programa puede variar de una plataforma a otra e incluso puede ser necesario que se agregue la ruta del mismo a la variable de sistema \$PATH (ver apéndice).

Con esta instrucción se genera de manera automática la esfera de solvatación. Los archivos creados incorporan “ws” en el nombre. Deberá verificarse que no se presenten errores durante la ejecución del script.

Importante: Anotar los datos de las coordenadas del centro de masas de la esfera y del radio de la esfera ya que estos se utilizarán más tarde para el paso de equilibrio del sistema.

- c) Desde la interfaz gráfica de VMD abrir el archivo \*ws.pdf creado para visualizar a la molécula dentro de la esfera de solvatación.

### 3.2. Solvatación en una caja de agua

- a) Colocar los archivos pdb y psf obtenidos en la sección 4 en una carpeta separada.
- b) Ejecutar VMD y abrir la consola Tk. Desde la consola se debe ingresar a la carpeta conteniendo los archivos mencionado en el inciso anterior.
- c) Ejecutar los comandos

```
package require solvate
```

```
solvate 1HPV.psf 1HPV.pdb -t 5 -o 1HPV_wb
```

La primera instrucción llama al paquete “solvate” de VMD

En la segunda instrucción se deben especificar los nombres de los archivos psf y pdb (aquí se asume que son 1HPV.psf y 1HPV.pdb). La instrucción -t 5 permite establecer un borde de 5 Angstroms por lo menos de la caja de agua con respecto a la proteína (es un poco pequeño para una aplicación real, pero definir una más grande incrementa el tiempo de procesamiento).

- d) Al término de la ejecución del script deberá visualizarse a la molécula solvatada dentro de su caja de agua.

- e) Determinar las coordenadas mínimas y máxima de la caja de agua. Para esto ejecutar desde la consola TK las instrucciones:

```
set everyone [atomselect top all]
```

```
measure minmax $everyone
```

Con estas instrucciones el programa reporta las coordenadas de la caja de agua. Se deben registrar estos datos como en el caso del cálculo de la esfera, ya que se tomarán en cuenta para el proceso de equilibrado.

#### **4. Equilibrado del sistema.**

Las etapas anteriores son necesarias para preparar al sistema para la dinámica molecular. Existen diversos tipos de análisis que pueden efectuarse. En este ejemplo solamente se ilustrará como se lleva a cabo el proceso de equilibrado del sistema. Este proceso consiste en llevar a la molécula hacia un estado en el cual las moléculas de la proteína alcanzan el equilibrio junto con las moléculas de solvente u otras presentes en el sistema (por ejemplo, iones). Este equilibrio se juzga a través de la distribución de las velocidades, presiones, etc. del sistema y sobretodo porque la conformación de la proteína tiende a relajarse y estabilizarse (sin llegar necesariamente a el mínimo de menor energía).

##### **4.1 Simulación de equilibrio de la molécula en una esfera de agua.**

En esta simulación se ilustra como la molécula alcanza el equilibrio en un esfera de agua colocada en el vacío. Se considera esta situación asumiendo que la molécula se encuentra aislada de otras moléculas. Se dice entonces que las condiciones que delimitan el sistema no son periódicas.

- a) En la carpeta con los archivos de la esfera de agua se debe copiar el archivo con la configuración de la simulación de dinámica molecular en una esfera de agua `ws_eq.conf`.
- b) El archivo `ws_eq.conf` debe editarse con un editor de texto y modificar los datos de estructura (`structure`) y coordenadas (`coordinates`) para que correspondan con los nombres de los archivos empleados. Existe una amplia variedad de parámetros en este script que se emplean para configurar la dinámica molecular: tales como los parámetros de campo (`force-fields`), temperatura, pasos de la integración. El paso de integración (`timestep`) en este ejemplo, empleado para resolver el sistema, es de 2 femtosegundos (fs).
- c) En especial deben configurarse la condiciones de los límites de la esfera (Spherical Boundary Conditions). Se debe especificar el centro de la esfera (`sphericalBCcenter`) y se debe considerar el radio de la misma para especificar la distancia a la cual comienza a actuar el potencial de campo (`sphericalBCr1`).
- d) En la sección `run` se especifica el número total de pasos en los cuales se realiza el proceso de equilibrado. Si se especifica 2500, considerado que el tiempo del paso de

integración es de 2 fs, entonces el tiempo total de la simulación será de 5000 fs o bien 5 picosegundos (5 ps).

- e) Ejecutar la simulación con NAMD2 (la cual puede tomar un tiempo largo), para esto se debe ejecutar el comando:

```
namd2 ws_eq.conf > 1HPV_ws_eq.log
```

Nota: En equipos con múltiples procesadores, se puede aprovechar la característica de hyperthreading de NAMD2. Por ejemplo, para ejecutar NAMD en un sistema con 4 procesadores, se debe ejecutar NAMD2 con la instrucción:

```
namd2 +p4 ws_eq.conf > 1HPV_ws_eq.log
```

- f) NAMD genera varios archivos después de concluida la simulación. En particular el archivo con la extensión dcd contiene los datos de las trayectorias de los átomos y puede emplearse para visualizar la dinámica del sistema con VMD.

#### **4.1 Simulación de equilibrio de la molécula en una caja de agua.**

- a) En la carpeta con los datos de la caja de agua se debe copiar el archivo con la configuración de la simulación de dinámica molecular wb\_eq.conf.

- b) Como en el ejemplo de la esfera de agua, el archivo de configuración debe modificarse para que los datos de estructura y coordenadas, puedan localizar los archivos psf y pdb de la molécula en la caja de agua.

- c) Es importante definir las condiciones periódicas de los límites (Periodic Boundary Conditions). En el caso de una caja de agua, se está asumiendo que la molécula de la proteína está inmersa en una celda, que se replica varias veces. De esta forma se asume que la molécula no está aislada (como en el caso de la esfera de agua). Los datos de las coordenadas x, y, z de la caja, calculados cuando se estimó la caja de solvatación deben proporcionarse en el archivo de configuración se establecen como el origen de la celda (cell origin). Por otro lado, a partir de las condiciones x, y, z que representan los extremos más alejados de la caja, se deben estimar las dimensiones de los lados de las cajas ( $dx = x_{max} - x_{min}$ ,  $dy = y_{max} - y_{min}$  y  $dz = z_{max} - z_{min}$ ). Estos datos corresponden a los tres vectores "cellBasisVectors" que deben especificarse en esta sección.

- d) Ejecutar la simulación con NAMD2 (la cual puede tomar un tiempo largo), para esto se debe ejecutar el comando:

```
namd2 ws_eq.conf > 1HPV_ws_eq.log
```

Nota: En equipos con múltiples procesadores, se puede aprovechar la característica de hyperthreading de NAMD2. Por ejemplo, para ejecutar NAMD en un sistema con 4 procesadores, se debe ejecutar NAMD2 con la instrucción:

```
namd2 +p4 ws_eq.conf > 1HPV_ws_eq.log
```



f) NAMD genera varios archivos después de concluida la simulación. En particular el archivo con la extensión dcd contiene los datos de las trayectorias de los átomos y puede emplearse para visualizar la dinámica del sistema con VMD.

## **5. Visualización de las dinámicas y análisis.**

Las simulaciones anteriores generan diversos archivos. En particular el archivo log, contiene información detallada con respecto al progreso de la simulación. En este archivo pueden visualizarse las fluctuaciones de la energía y la temperatura del sistema. Los errores en la simulación se registran en este archivo y de esta manera el usuario podría examinarlos y modificar la configuración del análisis.

El archivo dcd contiene los datos de las trayectorias de los átomos y con esta información puede visualizarse la dinámica de la molécula en esta simulación.

Para visualizar la dinámica de la proteína, se debe cargar en VMD el archivo psf y pdb del sistema en su esfera o caja de solvatación y luego, dando click con el botón izquierdo del ratón sobre el nombre de la molécula, se despliega una opción para leer los datos de las trayectorias (Load data into Molecule). En la interfaz gráfica se puede desplegar la animación de la dinámica.

Por otro lado en la interfaz gráfica, en "Extensions > Analysis > NAMDplot" se puede desplegar una herramienta para visualizar el comportamiento de los valores de energía y temperatura, entre otros parámetros, del archivo log.

Diversas herramientas de VMD en la sección de extensiones, permiten analizar el comportamiento de la proteína, tales como la variación del RMSD, variación de la energía, etc.

## **Referencias Bibliográficas.**

Este tutorial esta basado en gran parte en el documento "NADM TUTORIAL de James Phillips y cols. que puede descargarse de la página WEB de NADM, en la sección de tutoriales (visite: Introductory NADM Tutoriales en <http://www.ks.uiuc.edu/Research/namd/>). En este documento podrá encontrar mayor información técnica y científica sobre este programa.

1. Phillips J.C. (2005): Scalable molecular dynamics with NAMD. Journal of Computational Chemistry. vol. 26, pp 1781–1802.
2. Humphrey, W., Dalke, A. y Schulten, K (1996): "VMD - Visual Molecular Dynamics", J. Mol. Graphics, vol. 14, pp. 33-38.
3. Stewart, A. A. y McCammon, J.A. (2006): Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem Rev. Vol. 106, pp: 1589–1615.
4. Wartha, F., Horn A. H. C., Meiselbach H y Stich H. (2005): Molecular Dynamics Simulations of HIV-1 Protease Suggest Different Mechanisms Contributing to Drug Resistance. J. Chem. Theory Comput. vol. 1, pp: 315-324.

# Apéndices.

## Apéndice 1: Notas adicionales a la instalación de NAMD y VMD.

Con frecuencia, los programas NAMD y VMD deben ser llamados desde la consola del sistema. Por tal motivo, es conveniente incluir la ruta de la carpeta donde se encuentran los archivos binarios de este sistema, a la variable de entorno \$PATH, que encontramos en sistemas operativos tanto Windows como Linux.

En sistemas Windows XP/Vista/7 esto puede hacerse ingresando a propiedades del sistema en el panel de control, Avanzadas, Variables de entorno. En el cuadro, en Variables de sistema, se debe listar la variable PATH. Esta debe editarse agregando al final la ruta del programa NAMD y VMD. Estas rutas deben separarse con comas de las anteriores y entrecomillarse si la ruta contiene espacios.

En los sistemas tipo UNIX, como Linux y MacOSX, en la carpeta principal de usuario debe existir un archivo .bash\_profile o .profile. Para visualizar estos archivos debe abrirse una terminal de sistema y utilizar un editor para abrir el archivo (el punto al principio del nombre del archivo significa que este se encuentra oculto). La ruta de los programas NAMD y VMD puede agregarse a la variable PATH.